

Article

Pathological, Morphological, Cytogenomic, Biochemical and Molecular Data Support the Distinction between *Colletotrichum cigarro comb. et stat. nov.* and *Colletotrichum kahawae*

Ana Cabral ^{1,2} , Helena G. Azinheira ^{1,2} , Pedro Talhinhos ^{1,2,*} , Dora Batista ^{1,2,3}, Ana Paula Ramos ^{1,4} , Maria do Céu Silva ^{1,2} , Helena Oliveira ¹ and Vítor Várzea ^{1,2}

- ¹ Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisbon, Portugal; anacabral@isa.ulisboa.pt (A.C.); hmga@edu.ulisboa.pt (H.G.A.); dccastro@fc.ul.pt (D.B.); pramos@isa.ulisboa.pt (A.P.R.); mariaceusilva@isa.ulisboa.pt (M.d.C.S.); heloliveira@isa.ulisboa.pt (H.O.); vitorvarzea@isa.ulisboa.pt (V.V.)
- ² Centro de Investigação das Ferrugens do Cafeeiro, Instituto Superior de Agronomia, Universidade de Lisboa, 2780-505 Oeiras, Portugal
- ³ Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisbon, Portugal
- ⁴ Laboratório de Patologia Vegetal “Veríssimo de Almeida”, Instituto Superior de Agronomia, Universidade de Lisboa, 1349-017 Lisbon, Portugal
- * Correspondence: ptalhinhos@isa.ulisboa.pt; Tel.: +351-213653249

Received: 26 March 2020; Accepted: 12 April 2020; Published: 14 April 2020



Abstract: The genus *Colletotrichum* has witnessed tremendous variations over the years in the number of species recognized, ranging from 11 to several hundreds. Host-specific fungal species, once the rule, are now the exception, with polyphagous behavior regarded as normal in this genus. The species *Colletotrichum kahawae* was created to accommodate the pathogens that have the unique ability to infect green developing coffee berries causing the devastating Coffee Berry Disease in Africa, but its close phylogenetic relationship to a polyphagous group of fungi in the *C. gloeosporioides* species complex led some researchers to regard these pathogens as members of a wider species. In this work we combine pathological, morphological, cytogenomic, biochemical, and molecular data of a comprehensive set of phylogenetically-related isolates to show that the Coffee Berry Disease pathogen forms a separate species, *C. kahawae*, and also to assign the closely related fungi, previously in *C. kahawae* subsp. *cigarro*, to a new species, *C. cigarro comb. et stat. nov.* This taxonomic clarification provides an opportunity to link phylogeny and functional biology, and additionally enables a much-needed tool for plant pathology and agronomy, associating exclusively *C. kahawae* to the Coffee Berry Disease pathogen.

Keywords: taxonomy; speciation; *Colletotrichum cigarro*; *Colletotrichum kahawae*; Coffee Berry Disease

1. Introduction

The genus *Colletotrichum* Corda comprises diverse plant pathogens, causing important diseases collectively known as anthracnose. The taxonomy of species in this genus has changed several times during the last decades. Since it was first described, and for over one century, a growing number of species was assigned to *Colletotrichum*. Von Arx [1] listed around 750 species, many of which had been described based on the host plant range rather than on morphology, and drastically reduced this number to 11 species based on morphological characteristics. Sutton [2] revised the genus and recognized 22 species, most of them polyphagous, based on morpho-cultural criteria. In this process, the species *Colletotrichum gloeosporioides* has played a central role in the genus *Colletotrichum*,

with numerous pathogens causing important diseases assigned to it. In fact, *C. gloeosporioides* was for a long time considered a species complex, frequently regarded as a dumping taxon for diverse *Colletotrichum* fungi and, in this sense, of little biological, taxonomical, or pathological meaning [3]. More recently, the taxonomy of *Colletotrichum* has witnessed major changes, comprising nearly 200 accepted species [4–6], the vast majority of which clustered into species complexes. The *C. gloeosporioides* species complex [7] is the widest, comprising at least 38 species [4].

Although clustering within the *C. gloeosporioides* species complex [8], the species *C. kahawae* Waller and Bridge was recognized as a separate taxon due to the unique feature of its members as being capable of infecting green coffee berries, thus causing the Coffee Berry Disease (CBD) [9]. This disease is restricted to Africa, where it represents the main constraint to sustainable production of Arabica coffee, due to yield losses up to 80% without chemical control [10–14]. The risk of its introduction on the main Arabica growing countries (in America and Asia) makes this a quarantine pathogen, referred to as a biological weapon in Australia [13–15].

Other *Colletotrichum* spp., namely other species from the *C. gloeosporioides* complex and *C. acutatum*, are frequently isolated from ripe coffee berries, but they are incapable of causing CBD [9,16,17]. In fact, *C. kahawae* was created to resolve the ambiguity caused by the inclusion of both CBD-causing and non-CBD causing isolates in the previously described species *C. coffeanum* Noak [9,16,18,19]. Additionally, *C. kahawae* isolates are unable to utilize citrate or tartrate as sole carbon sources, while the non-CBD causing isolates from coffee metabolized one or both [9]. However, a group of *Colletotrichum* isolates from multiple hosts and diverse geographic origins (unrelated to coffee) was shown to be indistinguishable from the CBD pathogens for six nuclear gene regions that are usually employed for taxonomic purposes, actin (*act*), calmodulin (*cal*), chitin synthase (*chs1*), glyceraldehyde-3-phosphate dehydrogenase (*gapdh*), manganese-superoxide dismutase (*sod2*), and β -tubulin 2 (*tub2*), along with the rDNA-ITS region [7], in spite of a clear differentiation provided by the mating type gene *mat1-2-1* (*mat1-2-1*), a fragment of DNA lyase *Apn2* [20] (*apn25L*), and glutamine synthetase (*gs*) [7], as well as the capacity to utilize either citrate or tartrate as a sole carbon source [7,9,21]. In this context, both CBD-causing and the phylogenetically close non-CBD causing isolates were all placed under *C. kahawae* [7], being created *C. kahawae* subsp. *kahawae* to accommodate the CBD-causing isolates and *C. kahawae* subsp. *cigarro* (as *C. kahawae* subsp. “*ciggaro*” B. Weir and P.R. Johnst.) for the phylogenetically close non-CBD causing isolates.

Subsequently, Doyle et al. [22] proposed two new species, *C. fructivorum* and *C. temperatum*, and epitypified the species *C. rhexiae*, all of them closely related to *C. kahawae*. These authors also reclassified an isolate from cranberry (CBS124.22) that was previously classified as *C. kahawae* subsp. *cigarro* [7], as *C. fructivorum*. Liu et al. [23] identified a new species, *C. jiangxiense*, among isolates from symptomatic and asymptomatic *Camellia sinensis* from China, phylogenetically closely related with CBD-causing isolates. Nevertheless, based on the pathogenicity tests and the pairwise homoplasy index test, these authors considered *C. jiangxiense* and CBD-causing isolates two independent species. Wang et al. [24] described, from diseased leaves of *Camellia sinensis* in China, a closely related species to *C. jiangxiense* and *C. kahawae* sensu lato, that was designated *C. wuxiense*. This later species can be distinguished from other species of the gloeosporioides complex using a concatenated ApMAT and *gs* gene tree [24]. More recently, Guarnaccia et al. [25] described *C. helleniense*, a species that is phylogenetically close to both subspecies of *C. kahawae* but clearly differentiated based on the sequence analysis of *cal*, *gapdh*, and *tub2* genes.

Host-shift speciation, a particular case of ecological speciation, is one of the main routes for the emergence of fungal pathogens [26,27]. An investigation on the origin, phylogeography, and evolution of the CBD-causing pathogen demonstrated that these fungi have undergone a recent speciation process via host-jump into coffee, exploiting the green coffee berries-ecological niche, departing from a background of a seemingly generalist group of fungi harmless to coffee berries [27]. While the divergence between these two groups is very recent (5600 years Before Present as an average estimate), genetic and biological evidences indicate the occurrence of an effective speciation event.

These evidences include significant and elevated differentiation indexes across all studied loci and a complete segregation of polymorphic sites [27]. This process may have also shaped populations in other *Colletotrichum* spp. [28,29].

The combination of immigrant unviability, a strong intrinsic barrier to gene flow arising from the impossibility of reproduction between organisms that cannot infect the same host, with a predominantly asexual behavior (preventing the occurrence of recombination), would have been rather effective in keeping populations separated during the early stages of divergence, thus creating pleiotropic interactions between local adaptation and reproductive patterns to speed up the speciation process [27]. While the genealogical concordance criteria alone are insufficient to recognize these entities as two separate species [7,27], biological and population genetic data show that CBD-causing isolates and non-CBD causing isolates represent ecologically distinct and isolated groups that have separated quite recently, and can therefore be currently regarded as distinct species [13,27].

Additionally, Pires et al. [30] reinforced evidences of differentiation between these two taxa considering the genome size expansion of CBD-causing isolates in comparison to the closely related non-CBD causing isolates.

In this work, we combine previously published data on evolutionary and population genetics [27], genome size estimation [30] and population genomics [14] with new analyses of phenotypic characters and molecular data (including novel loci putatively associated to pathogenicity and to sexuality) to sustain that *C. kahawae* subsp. *kahawae* and *C. kahawae* subsp. *cigarro* should be recognized as distinct species, respectively *C. kahawae* and *C. cigarro* comb. et stat. nov.

2. Results

2.1. Pathogenicity Tests

All 32 CBD-causing isolates were pathogenic to green coffee berries, scoring above two in the 0–5 disease severity scale, and originating CBD symptoms. The remaining nine isolates were not capable of causing CBD, with a final disease severity score of zero (see Supplementary Table S1). For all isolates, conidia germination and appressoria formation rates were above 50%.

2.2. Induction of Perithecia

The isolates associated with CBD were not able to produce fertile perithecia either in homothallic or heterothallic crosses, even under long incubation periods (up to three months). However, several perithecia were observed for isolate Ang67, but these perithecia were not able to differentiate asci and ascospores, remaining immature in all the assays (Figure 1, panels A–D). Among the isolates not associated with CBD, isolates ICMP 18539 and ICMP 12953 produced fertile perithecia profusely on the toothpicks and on the growth medium from one week after inoculation onwards in homothallic crosses (Figure 1, panels F,G). For the remaining non-CBD causing isolates, perithecia were not observed.

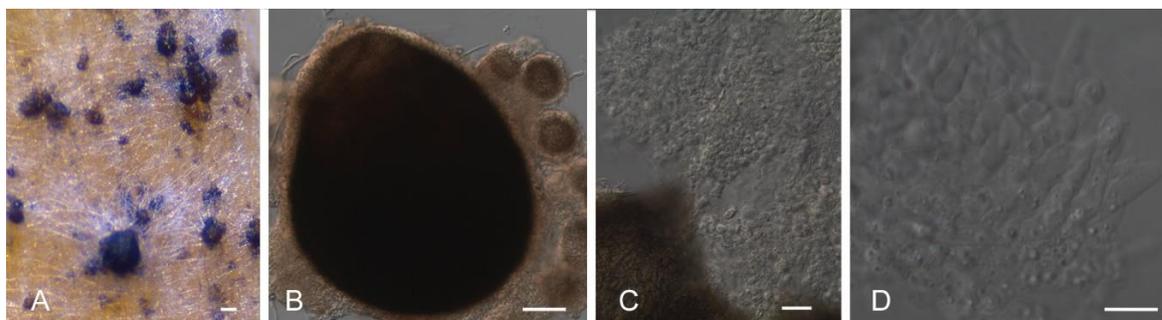


Figure 1. Cont.

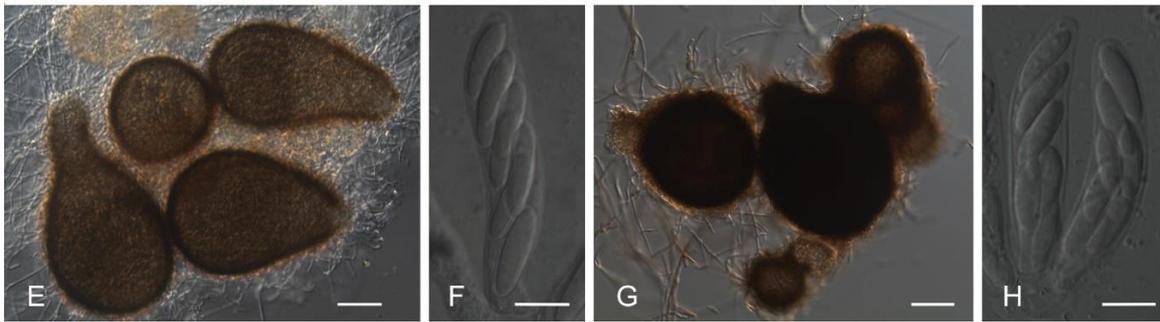


Figure 1. Differentiation of perithecia in *Colletotrichum* spp. *Colletotrichum kahawae* isolate Ang 67: (A,B), Perithecia observed after five weeks of incubation; (C,D), Perithecia oozing an undifferentiated mass. *Colletotrichum cigarro* isolates ICMP 18539 (E,F) and ICMP 12953 (G,H): E, G Perithecia; F, H Ascii and ascospores. Scale bars A = 100 μm ; B, E, G = 50 μm ; C = 20 μm ; D, F, H = 10 μm .

2.3. Substrate Use

All non-CBD causing isolates, *C. camelliae* and *C. gloeosporioides* isolates metabolized at least citric acid or ammonium tartrate as a sole carbon source. The CBD-causing isolates were unable to use any of these substrates (see Supplementary Figure S1 and Supplementary Table S1).

2.4. Morphology

Colony diameter at different temperatures was recorded (see Supplementary Table S2). At 30 °C, all CBD-causing isolates grew significantly less (1 to 15 mm after 7 days of incubation) than the non-CBD causing isolates (16 to 37 mm for the closely related isolates and 63 to 61 mm for the *C. gloeosporioides* isolates). At the remaining temperatures, CBD-causing isolates tended to grow slower than non-CBD causing isolates, although less discriminately.

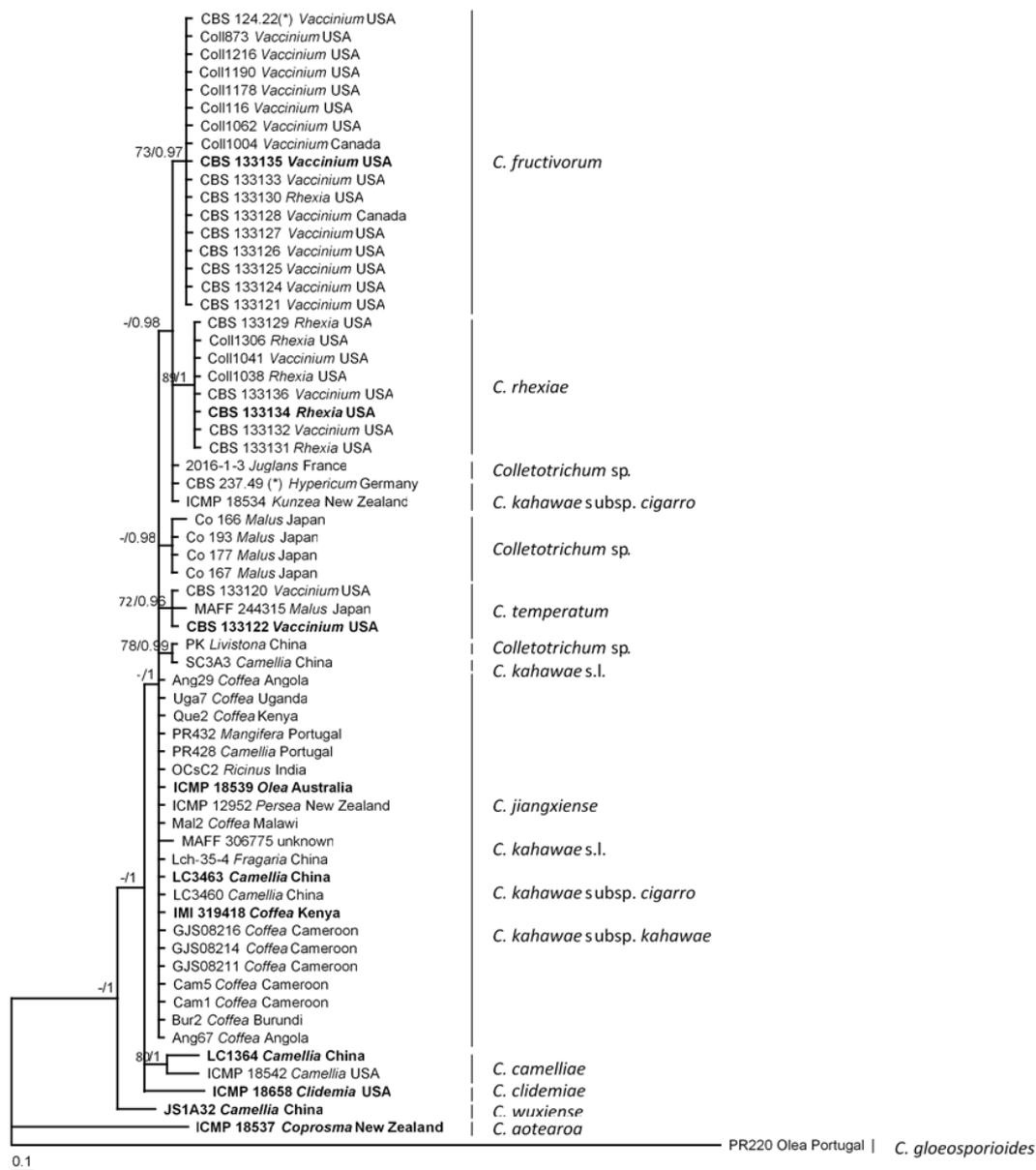
Conidia are cylindrical, straight with rounded ends. The average conidial length of CBD-causing isolates ranged between 11.7 and 18.1 μm , while that of non-CBD causing isolates ranged between 10.8 and 14.8 μm . The average conidial width ranged between 4.9 and 5.9 μm for the CBD-causing isolates and between 4.2 and 5.9 μm for non-CBD causing isolates. The length/width ratio ranged between 2.2 and 3.7 and between 2.1 and 3.4 for CBD-causing and non-CBD causing isolates respectively. Statistically significant differences were recorded between the set of CBD-causing isolates and the non-CBD causing isolates, with the former presenting longer (14.6 versus 13.2 μm) and wider (5.2 versus 5.1 μm) conidia, and with a larger length/width ratio (2.8 versus 2.7) (see Supplementary Table S3).

Hyphal appressoria are globose to fusiform, lobbed. The average appressorial length ranged between 9.9 to 11.8 μm for CBD-causing isolates and between 8.4 and 10.8 μm for non-CBD causing isolates, with a global average of 10.8 and 9.5 μm respectively, representing a statistically significant difference. The average appressorial width for CBD-causing isolates ranged between 5.9 and 7.7 μm (average 7.1 μm), while that for non-CBD causing isolates was of 5.2 to 7.6 μm (average 6.5 μm), again representing a statistically significant difference between both groups. The average length/width ratio ranged between 1.4 and 1.7 and between 1.2 and 2.0 for CBD-causing and non-CBD causing isolates respectively, with the appressoria of CBD-causing isolates more elongated (average 1.6) than those of non-CBD causing isolates (average 1.5), although with no significant differences.

2.5. Phylogenetic Analyses

The sequences for ApMAT, *tub2* and *gs* available in GenBank as *C. kahawae* sensu lato and sequences of isolates from related species, as *C. aotearoa*, *C. camelliae*, *C. clidemiae*, *C. fructivorum*, *C. jiangxiense*, *C. rhexiae*, *C. temperatum*, and *C. wuxiense* were retrieved and aligned with those obtained in the present study (see Supplementary Table S4). The ApMAT alignment has a length of 727 bases including alignment gaps and comprises 63 ingroup taxa. The *tub2* alignment has 599 bases including alignment gaps and 96 ingroup taxa, and the *gs* alignment has 877 bases including alignment gaps and comprises

40 ingroup taxa. The *C. gloeosporioides* isolate PR220 was used as outgroup. The tree topologies obtained by Bayesian consensus tree and Maximum Likelihood (ML) analyses were similar for ApMAT, *tub2*, and *gs* and therefore only Bayesian consensus trees are presented with bootstrap support values (BS) and posterior probability values (PP) near each node. Phylogenetic analyses of ApMAT identified monophyletic groups, each comprising isolates assigned to *C. aotearoa*, *C. camelliae*, *C. clidemiae*, *C. fructivorum*, *C. rhexiae*, *C. temperatum*, and *C. wuxiense* but is unable to separate *C. jiangxiense* from *C. kahawae* sensu lato (Figure 2, panel A). In the *gs* phylogeny, the CBD-causing isolates form a monophyletic group, and non-CBD causing isolates previously assigned to *C. kahawae* subsp. *cigarro* cluster in various groups (Figure 2, panel B). The *tub2* sequences were unable to resolve the species *C. fructivorum*, *C. jiangxiense*, *C. kahawae* subsp. *cigarro*, *C. kahawae* subsp. *kahawae*, and *C. rhexiae* (Figure 2, panel C).



A

Figure 2. Cont.



B

Figure 2. Cont.

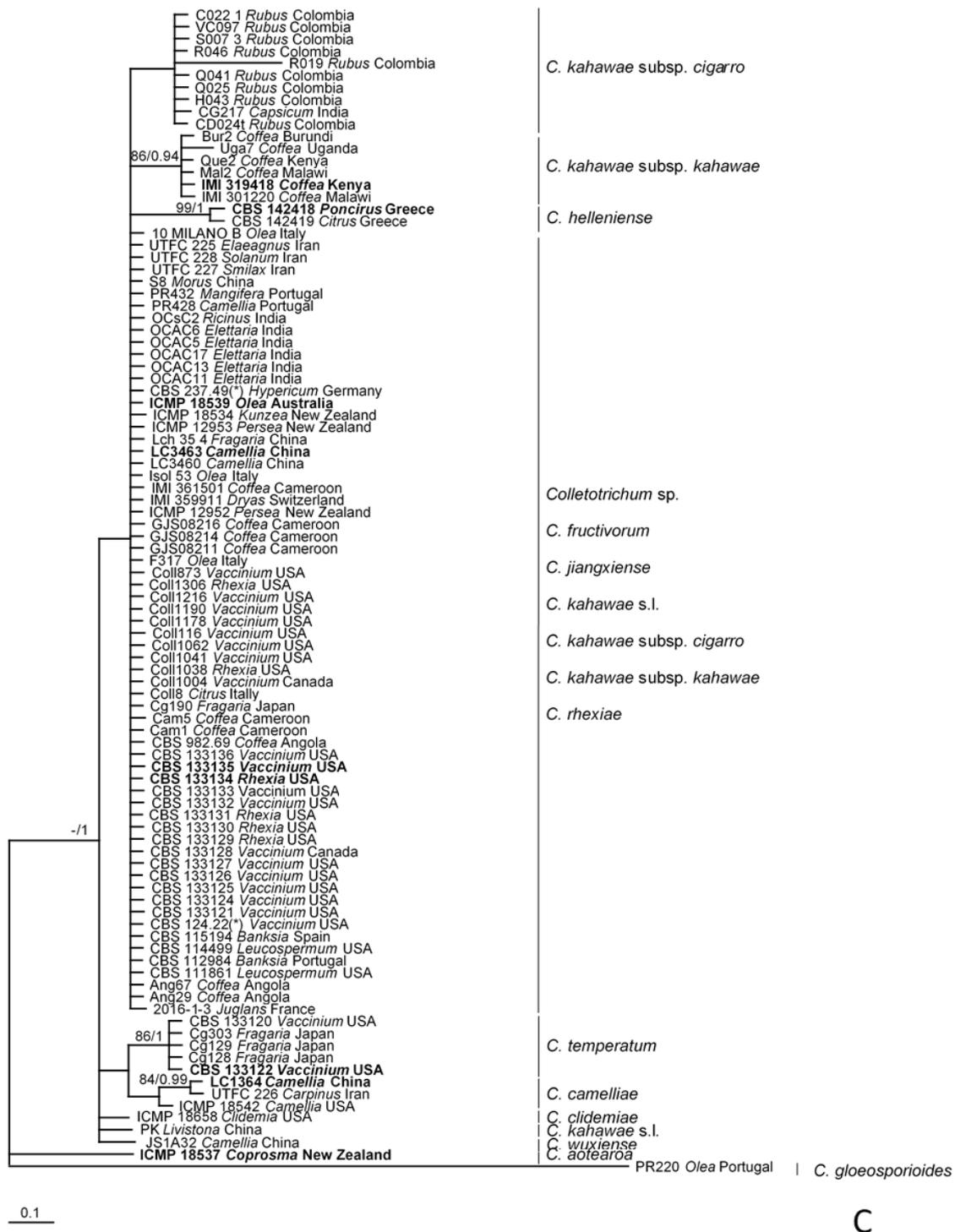


Figure 2. Fifty percent majority rule consensus tree from a Bayesian analysis based on the alignment the intergenic spacer between the 3' end of the *Apn2* gene and the mating type locus *MAT1-2-1* (*ApMAT*) (A), of partial glutamine synthetase gene (*gs*) (B) and the β -tubulin gene (*tub2*) (C) enabling the comparison of the isolates under study with those with similar sequences publicly available. The RAxML bootstrap support (≥ 70) values (BS) and Bayesian posterior probability (PP; ≥ 0.95) are displayed at the nodes (BS/PP). The tree was rooted to *Colletotrichum gloeosporioides* (PR220). The scale bar indicates the expected changes per site. Ex-type cultures are emphasized in bold. (*) ex-type or authentic culture of synonymized taxon.

The combined alignment of the nine loci (*apn25L*, *ApMAT*, *mat1-2-1*, *cas1*, *cellwall*, *siRNA*, *vosA*, *gs*, and *tub2*) comprises 8582 characters, including alignment gaps, from which 7656 characters were constant, 225 were parsimony-informative and 680 were variable but parsimony-uninformative. For each locus, the information of number of characters and parsimony informative sites, nucleotide substitution models, and score for the best tree of ML and Bayesian inference are listed in Table 1. The *C. gloeosporioides* isolate PR220 was used as outgroup. The topologies obtained by Bayesian consensus tree and ML analysis were identical for the single locus analysis, as well as in the nine gene-concatenated data.

Table 1. Phylogenetic Information on the loci used in this study.

Loci ¹	Taxa	Nucleotide Substitution Models	Chars	Constant	Parsimony-Informative	Parsimony-Uninformative	ML-In L ²	BI-In L ³
apMAT	64	K80	727	595	20	110	-1707.61	-1826.67
<i>gs</i>	47	HKY + G	877	751	39	68	-1941.59	-2021.35
<i>tub2</i>	96	HKY + G	599	541	19	31	-1180.52	-1289.96
apMAT	17	K80	725	614	12	98	-1528.87	-1588.01
<i>apn25L</i>	17	GTR	837	749	17	71	-1625.35	-1693.53
<i>cas1</i>	17	GTR + G	845	760	25	59	-1647.75	-1737.89
<i>cellwall</i>	17	GTR + I	1239	1112	50	73	-2481.99	-2521.34
<i>gs</i>	17	HKY + G	859	785	18	53	-1601.82	-1640.14
<i>mat1-2-1</i>	17	GTR + G	843	786	10	47	-1465.25	-1512.12
<i>siRNA</i>	17	HKY + G	1292	1112	52	125	-2839.28	-2880.50
<i>tub2</i>	17	HKY	597	548	8	33	-1067.93	-1100.61
<i>vosA</i>	17	HKY + G	1345	1190	33	121	-2723.89	-2760.43
combined	17		8582	7656	225	680	-17098.40	-17350.83

¹ apMAT—an intergenic spacer between the 3' end of the *Apn2* gene and the mating type gene *mat1-2-1*; *apn25L*—part of *Apn2* gene; *cas1*—apressorium specific protein; *cellwall*—cell wall protein; *gs*—glutamine synthetase; *mat1-2-1*—mating type gene; *siRNA*—a putative argonaute siRNA chaperone complex subunit; *tub2*— β -tubulin and *vosA*—developmental regulator; ² ML-In L—Likelihood score calculated in RAxML under GTRCAT model; ³ BI-In L—Estimated marginal likelihoods from MrBayes.

The phylogenetic analysis of single loci of *apn25L*, *mat1-2-1*, *cas1*, *cellwall*, *gs*, *siRNA*, and *vosA* showed that the CBD-causing isolates cluster as a monophyletic group (with MPBS $\geq 70\%$ and a Bayesian PP ≥ 0.95) and can be clearly separated from the non-CBD causing isolates, while *ApMAT* and *tub2* were unable to separate both groups of isolates. The *C. aotearoa* isolate was consistently placed out of the remaining. A similar situation occurred for the *C. camelliae* isolate except for *mat 1-2-1* and *gs* genes/regions. Isolates ICMP 18534 and CBS 237.49 formed a monophyletic group for all genes except for *gs*, *tub2*, and *vosA* (see Supplementary Figure S2). The analysis of the nine genes/regions concatenated data set depicts the CBD-causing isolates as a monophyletic group, along with four groups comprising non-CBD causing isolates. The group phylogenetically closest to that of the CBD-causing isolates contains the isolates ICMP 18534 and CBS 237.49, and the holotype of *C. kahawae* subsp. *cigarro* (ICMP 18539) clusters in a basal position (Figure 3).

The Bayesian Poisson tree processes method (BPTP) model estimated seven species in our dataset (see Supplementary Figure S3). The CBD-causing isolates are estimated to form a species with a PP of 0.83, while the non-CBD causing isolates are grouped in four hypothetical species, with PP ranging from 0.72 and 0.91.

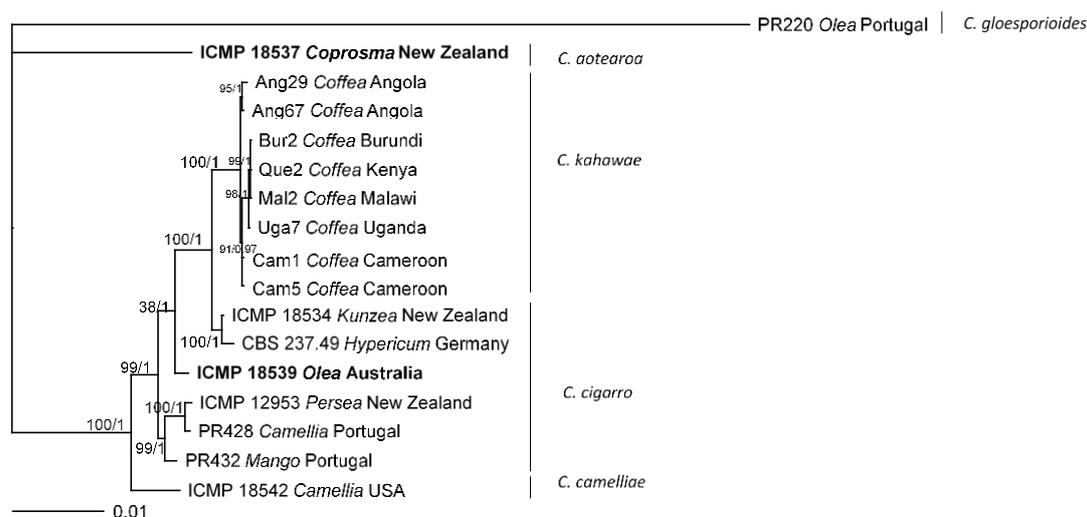


Figure 3. Fifty percent majority rule consensus tree from a Bayesian analysis based on a nine-loci combined dataset (*apn25L*, *ApMAT*, *mat1-2-1*, *cas1*, *cellwall*, *siRNA*, *vosA*, *gs*, and *tub2*) for isolates under study. The RAxML bootstrap support values (BS) and Bayesian posterior probability (PP) are displayed at the nodes (BS/PP). The tree was rooted with *Colletotrichum gloeosporioides* (PR220). The scale bar indicates the expected substitutions per site. Ex-type cultures are emphasized in bold. (*) ex-type or authentic culture of synonymized taxon.

2.6. In Silico Analyses of Proteins

The analysis of the predicted proteins obtained for *cas1* showed that the seven CBD-causing isolates studied have two amino acid differences at positions 176 (D or N) and 178 (I or V), respectively, when compared with the proteins predicted for the six non-CBD causing isolates (Table 2). The CAS1 protein is an ortholog of an appressorium specific protein, with a DUF3129 domain (protein of unknown function), belonging to family Egh16-like virulence factor (IPR021476). This protein is predicted as extracellular (prediction confirmed by SignalP, TargetP, TMHMM, and Phobius), with a signal peptide and a cleavage site between position 22 and 23. Predictions of phosphorylation and glycosylation showed no differences between proteins from non-CBD or CBD-causing isolates as well as in the secondary structure.

Concerning the protein coded by the *cellwall* gene, the differences of the amino acid composition at position 42 (L or Q), 79 (K or Q), and 243 (P or L), and a deletion between positions 181 to 183 and from positions 204 to 223 can clearly distinguish CBD-causing isolates from non-CBD causing isolates (Table 2). The in silico analysis of structure/function of this protein showed the presence of the PFAM domain PF12296 (Hydrophobic surface binding protein A HsbA) and a signal peptide with a cleavage site between position 17–18 suggesting its extracellular localization (prediction confirmed by SignalP, TargetP, TMHMM, and Phobius). The kinase localization positions were predicted by NETPHOS (see Supplementary Figure S4). It was observed that different phosphorylation sites were predicted in non-CBD causing isolates between positions 185–223 in the region that is absent from the CBD-causing isolates. The glycosylation predictor YinOYong suggested four glycosylation sites in non-CBD causing isolates located in positions 185, 210, 213 and 223 (see Supplementary Figure S5). These sites are absent in CBD-causing isolates. The structural analysis of those proteins showed several differences, namely the percentage of disorder regions, the solvent accessibility and the secondary structure (Figure 4 and see Supplementary Table S5). In the *VosA* protein (viability of spores A), three amino acid differences at positions 49 (M or K), 128 (R or G), and 203 (P or A) were observed between the predicted proteins of CBD-causing isolates and non-CBD causing isolates. Predictions of phosphorylation and glycosylation showed no differences between proteins of the two groups of isolates as well as in the secondary structure.

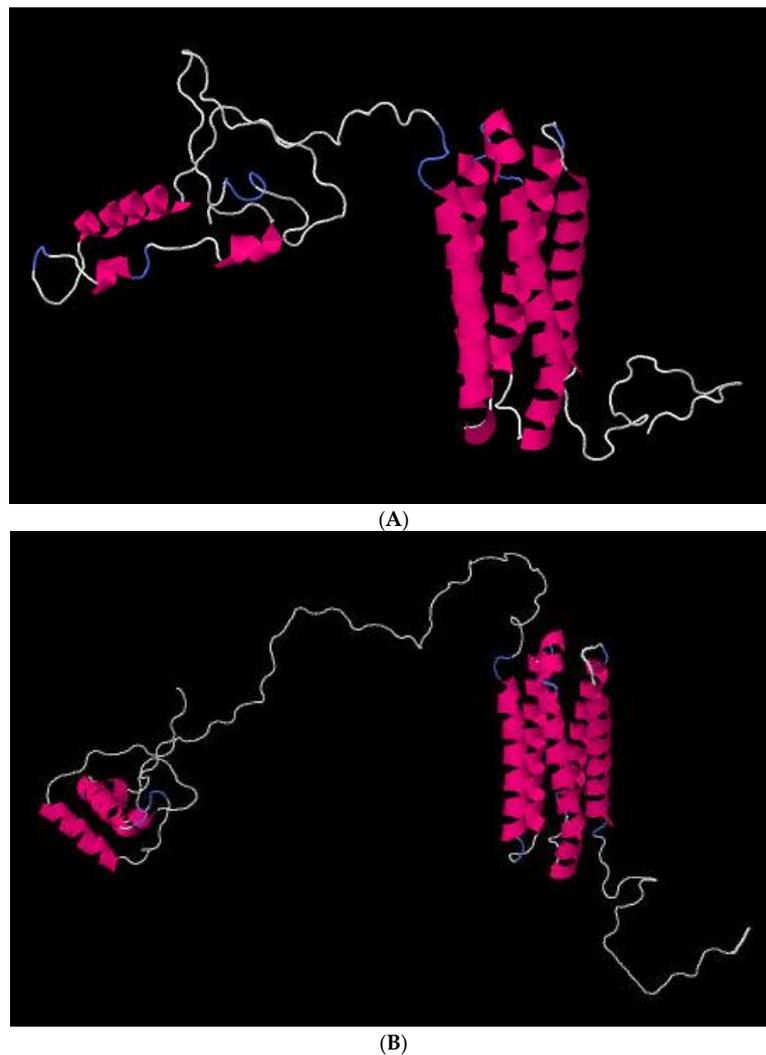


Figure 4. Molecular model (3D structure) of Cell wall protein of CBD-causing isolates (A) and non-CBD causing isolates (B) predicted by Phyre2. Jmol modelling software was used to visualize the predicted model [31].

For the siRNA protein, a putative argonaute siRNA chaperone complex subunit, ten amino acid differences were observed when comparing the 14 isolates listed in Table 2. However, when these isolates are grouped in CBD and non-CBD causing isolates, only one amino acid difference at position 194 (G to A) can distinguish both groups.

2.7. Genome Size

Genome size estimates ranged between 76.8 and 88.6 Mbp for CBD-causing isolates, and between 71.7 and 75.6 Mbp for non-CBD causing isolates. The genome size estimate for *Colletotrichum* sp. isolate PR428 is similar to that estimated for non-CBD causing isolates (72.9 Mbp). The *C. camelliae* isolate revealed the largest estimated genome size (88.7 Mbp) while the two *C. gloeosporioides* isolates have the smallest estimated genomes (average 70.3 Mbp). Statistical analysis performed on the genome size estimate of the isolates under study shows a significant difference between the CBD-causing isolates and non-CBD causing isolates. The average genome size of the CBD-causing isolates is 82.6 Mbp, while the group comprising non-CBD causing isolates previously assigned to *C. kahawae* subsp. *cigarro* shows an average genome size of 73.5 Mbp (see Supplementary Table S6).

2.8. Taxonomy

In this work, we have performed an analysis involving pathological, morphological, cytogenomic, biochemical, and molecular (new or re-assessed) traits that in combination shows that CBD-causing isolates can be clearly distinguished from the phylogenetically close non-CBD causing isolates. Based on the pathogenicity test on green coffee berries, metabolism of carbon sources, phenotypic characters such as growth rate at 30 °C, phylogeny, genome size, and population and evolutionary data from Silva et al. [27] and Vieira et al. [14], *Colletotrichum kahawae* subsp. *kahawae* revealed to form a separate species within the *Colletotrichum* genus and should be re-elevated to the species rank, *C. kahawae*. Consequently, the closely related fungi previously assigned to *C. kahawae* subsp. *cigarro* should also be raised to the species level.

Colletotrichum cigarro (B.S. Weir and P.R. Johnston) A. Cabral and P. Talhinhos, comb. et stat. nov. MycoBank MB 830326.

Basionym: *Colletotrichum kahawae* subsp. *cigarro* B.S. Weir and P.R. Johnst. [as ‘*ciggaro*’] Studies in Mycology 73:115-180. 2012.

Holotype: Australia, on *Olea europaea*, coll. V. Sergeeva UWS124, 1989, PDD 102232; ex-type culture ICMP 18539.

Weir et al. [7] provide a description.

Note: Weir et al. [7] used the term “*ciggaro*”, in an attempt to reflect the Portuguese word for “cigarette”. However, the correct term in Portuguese is “*cigarro*”, and this fact was already been corrected in Mycobank MB 626870.

The growth rate on PDA at 30 °C can be useful to discriminate both entities, since *C. cigarro* isolates (non-CBD causing) grow more after seven days cultivation on PDA (16 to 37 mm) than *C. kahawae* (1 to 15 mm). The genes *apn25L*, *cas1*, *cellwall*, *gs*, *mat1-2-1*, *siRNA*, and *vosA* clearly separate *C. kahawae* from *C. cigarro*. The isolates studied that were previously assigned to *C. kahawae* subsp. *cigarro* by Weir et al. [7] are grouped phylogenetically into four clusters. As a growing number of isolates has regularly been assigned to this group, in the future there may be a basis to further dissect this taxon as a putative cryptic species complex.

3. Discussion

The species *C. kahawae* was created to accommodate the CBD pathogen, found only in Africa, providing a taxonomic tool to plant pathologists to name and distinguish this pathogen from other *Colletotrichum* strains inhabiting coffee but incapable of causing this disease [9]. In mycological terms, however, this taxon has been shadowed by neighbouring fungi clustering in *C. gloeosporioides* sensu lato [8]. While resolving species limits and phylogenetic relationships within *C. gloeosporioides* sensu lato, Weir et al. [7] established numerous new species, but the differences between the CBD pathogen and closely related fungi were considered to be insufficient to recognize them as separate species. The CBD pathogen was placed along with non-CBD causing isolates into two subspecies within *C. kahawae*. This taxonomic framework led, in recent years, to the identification of “*C. kahawae*” (no subspecies given) in other hosts and regions [32–35], raising concern among plant pathologists dealing with coffee diseases, especially when the spread of *C. kahawae* out of Africa is so feared [13].

The taxonomy of *Colletotrichum* is no longer supported exclusively by morphology and host-specificity, since a polyphagous behaviour is common in most *Colletotrichum* species complexes [36]. The recent advances in molecular biology allowed new DNA-sequence based tools leading to more accurate phylogenetic relationships. The combination of molecular and pathological data has evidenced a very recent host jump-based speciation event leading to the Coffee Berry Disease pathogen, *C. kahawae* [14,27].

In this work, we provide a formal framework to differentiate *C. kahawae* from its close relatives with the establishment of a newly designated species, *C. cigarro*. For this we have performed analyses involving morphological, pathological, biochemical, cytogenomic, and molecular data with a comprehensive set of CBD-causing isolates and phylogenetically close non-CBD causing isolates.

Combining pathological and molecular data, Silva et al. [27] demonstrated a clear, albeit recent, speciation process leading to the differentiation of the CBD pathogen. In the present work, additional molecular data obtained for an enlarged set of relevant isolates further strengthen that divergence. This study revealed that, besides the genes *apn25L*, *mat1-2-1*, and *gs*, the genes *cas1*, *cellwall*, *siRNA*, and *vosA* are also able to discriminate between CBD and non-CBD causing isolates. The phylogenetic analysis performed demonstrate that CBD-causing isolates are grouped in a monophyletic clade while the non-CBD causing isolates (designated by Weir et al. [7] as *C. kahawae* subsp. *cigarro*) grouped in four hypothetical species estimated by the BPTP model, that are clearly different from CBD-causing isolates. Similar results were obtained by Vieira et al. [14] that identified 9160 SNPs, which completely differentiated CBD-causing isolates from non-CBD causing isolates, and also observed long tree branches within clades, comprising the non-CBD causing isolates, suggesting a highly divergent group.

Moreover, the analysis of predicted proteins for *cas1*, *vosA*, and *cellwall* genes revealed differences in the amino acid composition between the two groups of isolates. These differences may affect the function of proteins and may be at the origin of the different pathological behaviour of each group of isolates, namely the ability to infect green coffee berries.

The CAS1 protein belongs to a group of fungal proteins that are found in pathogenic fungi and may play important roles during the early infection stage and appressoria development [37]; VosA protein contains a velvet domain with a DNA binding motif that recognize nucleotide consensus in the promoters of key developmental regulatory genes controlling several processes including toxin production, cell wall formation and the development of resting or sexual fruiting bodies [38]. In *Aspergillus nidulans*, VosA is required for sporogenesis and trehalose biogenesis of conidia and ascospores [39].

More significant seems to be the differences in amino acid composition of the cellwall protein, where the structural analysis prediction showed differences in the percentage of disorder regions, solvent accessibility and secondary structure along with different number and localization of phosphorylation and glycosylation sites in CBD-causing and non-CBD causing isolates. Although such sites can be reversibly and dynamically modified by O-GlcNAc or phosphate groups during cell life [40], these differences could be functionally significant. The *cellwall* gene studied here is an orthologue of the gene ENH85888 from *Colletotrichum orbiculare* that belongs to the top 100 most highly expressed genes at one-day post inoculation [41], and the differences observed in the amino acid sequence between CBD and non-CBD causing isolates may affect their virulence and pathogenicity. It has been previously referred a higher diversity on secreted and nuclear localized proteins that could play an important role in adaptation to lineage-specific infection lifestyles [42].

Overall, the differences in protein structures could be related with the differences observed in the pathogenic behaviour of the CBD pathogen, since the major differences were reported in proteins that should be related with the early steps of the infection process and hypothetically with the ability to infect green coffee berries. The predicted differences in the secondary structure observed between proteins of CBD-causing and non-CBD causing organisms suggest that their function may be altered or even compromised. This may be related to the differences observed in the infection process. In fact, structural phylogenetics provide an important complement to sequence-based analyses [43].

Morphological differences between CBD and non-CBD causing isolates concerning conidia and appressoria are scarce, as previously noted by Weir et al. [7]. The growth rate on PDA at 30 °C, however, was useful to discriminate both entities. Furthermore, while none of the CBD-causing isolates was able to differentiate fertile perithecia, some of the non-CBD causing isolates produced fertile perithecia readily and profusely. Among the latter is the *C. cigarro* holotype, isolate ICMP 18539. Interestingly, one in 32 CBD-causing isolates were able to differentiate perithecia, although infertile. The difference in the capacity to differentiate fertile perithecia between CBD and closely related non-CBD causing isolates is newly reported here and could be related to the higher genetic diversity of the non-CBD causing isolates as compared to CBD-causing isolates and to the adoption of an asexual behaviour of the

latter in their adaptation to green coffee berries [27]. Additionally, the capacity to use citrate/tartrate as a sole carbon source [9] also remains as a distinctive trait between CBD and non-CBD causing isolates.

The ca. 8 Mbp genome size expansion between CBD and non-CBD causing isolates reported by Pires et al. [30] was corroborated by the results obtained in this study. Genome size variation is known to occur among species and species complexes in *Colletotrichum* [30] and could in fact reflect the effect of evolution on fungal genomes [44,45].

Altogether, these cytogenetic, molecular, morphological, biochemical, and pathological differences between the CBD and non-CBD causing isolates provide additional evidence to assign these entities into separate species. In fact, molecular data led recently to the differentiation of *C. camelliae*, *C. fructivorum*, *C. helleniense*, *C. jiangxiense*, *C. rhexiae*, *C. temperatum*, and *C. wuxiense* [22–25] from *C. kahawae* sensu lato. The populations that remained in *C. cigarro*, including the holotype ICMP 18539, are here recognized as separated from the CBD-causing pathogens, although further studies should be conducted with a larger collection of isolates from different hosts and locations to further dissect this taxon as a putative cryptic species complex.

4. Materials and Methods

4.1. Fungal Material

This study included monosporic isolates of *Colletotrichum* spp. associated with coffee and other hosts (Table 3). CBD-causing isolates from CIFIC (Centro de Investigação das Ferrugens do Cafeeiro, Instituto Superior de Agronomia, Universidade de Lisboa, Lisboa, Portugal) collection were obtained from green *Coffea arabica* berries exhibiting CBD symptoms comprising 10 African countries representing the three genetic groups that were previously identified [27]: Angola, Cameroon, and East Africa (including Burundi, Ethiopia, Kenya, Malawi, Rwanda, Tanzania, Uganda, and Zimbabwe). Non-CBD causing isolates were selected in order to represent taxa closely related to *C. kahawae*, including isolates described as *C. kahawae* subsp. *cigarro* by Weir et al. [7] and as *Colletotrichum* sp. UG1 by Silva et al. [27], and one unassigned *Colletotrichum* sp. isolate (PR428), along with two *C. gloeosporioides* sensu stricto isolates (PR220 and PT808) [46].

Table 3. List of *Colletotrichum* spp. isolates analysed. Isolates in bold-denote type strains. (*) = ex-type or authentic culture of synonymized taxon.

Isolate	Species	Host	Country, Region
Ang6	<i>C. kahawae</i>	<i>Coffea arabica</i>	Angola, Chianga
Ang29	<i>C. kahawae</i>	<i>C. arabica</i>	Angola, Ganda
Ang30	<i>C. kahawae</i>	<i>C. arabica</i>	Angola, Ganda
Ang67	<i>C. kahawae</i>	<i>C. arabica</i>	Angola, Ganda
Ang81	<i>C. kahawae</i>	<i>C. arabica</i>	Angola, Huambo
Bur2	<i>C. kahawae</i>	<i>C. arabica</i>	Burundi
Cam1	<i>C. kahawae</i>	<i>C. arabica</i>	Cameroon, Babadjou
Cam2	<i>C. kahawae</i>	<i>C. arabica</i>	Cameroon, Santa
Cam5	<i>C. kahawae</i>	<i>C. arabica</i>	Cameroon, Baham
Cam8	<i>C. kahawae</i>	<i>C. arabica</i>	Cameroon, Kumbo
Eti3	<i>C. kahawae</i>	<i>C. arabica</i>	Ethiopia, Sidamo
Eti9	<i>C. kahawae</i>	<i>C. arabica</i>	Ethiopia, Sidamo
Eti20	<i>C. kahawae</i>	<i>C. arabica</i>	Ethiopia
Mal2	<i>C. kahawae</i>	<i>C. arabica</i>	Malawi
Que2	<i>C. kahawae</i>	<i>C. arabica</i>	Kenya
Que42	<i>C. kahawae</i>	<i>C. arabica</i>	Kenya
Que48	<i>C. kahawae</i>	<i>C. arabica</i>	Kenya, Taita Taveta
Que72	<i>C. kahawae</i>	<i>C. arabica</i>	Kenya, Ruiru
Que82	<i>C. kahawae</i>	<i>C. arabica</i>	Kenya, Kitale
Que84	<i>C. kahawae</i>	<i>C. arabica</i>	Kenya, Mgunmuri

Table 3. Cont.

Isolate	Species	Host	Country, Region
Rua1	<i>C. kahawae</i>	<i>C. arabica</i>	Rwanda, Gicumbo
Tan2	<i>C. kahawae</i>	<i>C. arabica</i>	Tanzania, Mbinga
Tan13	<i>C. kahawae</i>	<i>C. arabica</i>	Tanzania, Mbinga
Uga2	<i>C. kahawae</i>	<i>C. arabica</i>	Uganda, Kapchorwa
Uga3	<i>C. kahawae</i>	<i>C. arabica</i>	Uganda, Kapchorwa
Uga5	<i>C. kahawae</i>	<i>C. arabica</i>	Uganda, Kapchorwa
Uga6	<i>C. kahawae</i>	<i>C. arabica</i>	Uganda, Kapchorwa
Uga7	<i>C. kahawae</i>	<i>C. arabica</i>	Uganda, Kapchorwa
Uga9	<i>C. kahawae</i>	<i>C. arabica</i>	Uganda, Kapchorwa
Zim1	<i>C. kahawae</i>	<i>C. arabica</i>	Zimbabwe, Hiton
Zim12	<i>C. kahawae</i>	<i>C. arabica</i>	Zimbabwe
Zim14	<i>C. kahawae</i>	<i>C. arabica</i>	Zimbabwe
CBS 237.49 (*), ICMP 17922, C1275.8	<i>C. cigarro</i> (syn. <i>Glomerella cingulata</i> var. <i>migrans</i>)	<i>Hypericum perforatum</i>	Germany
ICMP 12953, C1206.3	<i>C. cigarro</i>	<i>Persea americana</i>	New Zealand
ICMP 18534, C1252.12	<i>C. cigarro</i>	<i>Kunzea ericoides</i>	New Zealand
ICMP 18539, C1262.12	<i>C. cigarro</i>	<i>Olea europaea</i>	Australia
ICMP 18542, C1291, CG02g	<i>C. camelliae</i>	<i>Camellia</i> sp.	USA
PR432	<i>C. cigarro</i>	<i>Mangifera indica</i>	Portugal, Lisbon
PR428	<i>Colletotrichum</i> sp.	<i>Camellia japonica</i>	Portugal, Lisbon
PR220	<i>C. gloeosporioides</i> s.s.	<i>O. europaea</i>	Portugal, Tondela
PR808	<i>C. gloeosporioides</i> s.s.	<i>O. europaea</i>	Portugal, Silves

4.2. Pathogenicity Tests

The isolates employed in this study were reassessed for their pathogenicity to green coffee berries. Detached expanding green berries of *Coffea arabica* variety Caturra (CIFC 19/1) were inoculated as previously described [47], incubated under saturating humidity at 22 °C and symptoms were evaluated every day up to 14 days after inoculation. Ten green berries were used for each isolate. Symptoms were scored based on a 0 to 5 scale [30], where 0—no symptoms, 1—discrete lesions (necroses) less than 2 mm in length, 2—lesions occupying less than 25% of berry surface, 3—lesions occupying half of the berry surface, 4—lesion occupying the entire berry surface, and 5—lesion with abundant sporulation. Isolates responsible for final average disease severity scores of 2 or higher were considered as pathogenic, while the remaining were not pathogenic.

Conidia germination and appressoria formation rates were obtained in vitro (on glass slides) using the spore suspension employed in the inoculation assay. The slides were placed in a humidity box at 22 °C for 22 h. After this period, observations were made using a light microscope (Leitz Dialux 20, Stuttgart, Germany).

4.3. Induction of Perithecia

The ability of each isolate listed in Table 3 to form perithecia was screened. Two 3 mm plugs, obtained from the actively growing margin of monosporic 10-day-old PDA cultures, were placed on opposite sides of 60 mm diameter petri dishes containing 6 mL of minimal salt medium (1.6% agar). On the medium surface three autoclaved birch toothpicks were placed in a “N” configuration to provide a substrate for the sexual structures [48]. The plates were incubated at 22 °C under a 12 h/day photoperiod and the formation of perithecia was checked up to 12 weeks after inoculation. Each experiment comprised three plates per isolate. The experiment was conducted twice.

Additionally, heterothallic crosses were conducted between all CBD-causing isolates as previously described with three repetitions per cross. The experiment was conducted twice.

4.4. Phenotypic Characters

Conidia, appressoria, perithecia, asci, and ascospores were measured and described according to Weir et al. [7] from monosporic cultures grown on Synthetic Nutrient-poor Agar medium [49] (SNA) at 20 °C under white fluorescent light with a 12 h/day photoperiod. The observations were done using a Leica DM 2500 microscope (Stuttgart, Germany) with differential interference contrast illumination and the images were captured using a Leica DFC295 digital camera using the software Leica Application Suite (LAS) version 3.3.0. For each informative structure, 30 measurements were obtained. Measurements are presented as (minimum–) first quartile–medium–third quartile (–maximum).

Cardinal temperatures for fungal growth were assessed by inoculating 90 mm petri dishes containing PDA with a 3 mm diameter plug cut from the edge of an actively growing colony. Growth was determined after 7 days in two orthogonal directions. Experiments were conducted at 5 to 35 °C with 5 °C intervals, with three replicate plates per strain at each temperature.

Quantitative morphological data were compared using the Tukey Honest Significant Difference mean comparison test at 95% confidence (STATISTICA 8.0, StatSoft Inc., Tulsa, OK, USA).

4.5. Substrate Use

The capacity to utilize citric acid or ammonium tartrate as sole carbon sources was evaluated [9] in 90 mm diameter plates, using two replicates per each isolate. The experiment was repeated twice. Plates were incubated at 25 °C and medium color registered after 7 days as purple (substrate used) or yellow (substrate not used).

4.6. DNA Extraction, PCR Amplification, and Sequencing

For each of the selected isolates, total genomic DNA was extracted from mycelia grown in PDA plates [50].

Gene sequences were obtained in this study or retrieved from GenBank from eight nuclear gene regions and one intergenic spacer: part of *Apn2* gene (*apn25L*), an intergenic spacer between the 3' end of the *Apn2* gene and the mating type gene *mat1-2-1* (ApMAT), the mating type gene *mat1-2-1* (*mat1-2-1*), β -tubulin 2 (*tub2*) and glutamine synthetase (*gs*). Additionally, two genes were chosen from the *Colletotrichum orbiculare* transcriptome belonging to the top 100 most highly expressed genes at one day post inoculation [41], respectively: ENH76341, an appressorium specific protein (*cas1*); and ENH85888, a cell wall protein (*cellwall*). Additionally, two genes putatively involved in the sexual development [39,51] were used, a putative argonaute siRNA chaperone complex subunit (*siRNA*) and spore development regulator containing a velvet domain (*vosA*). Orthologous of each of the four genes were identified and sequenced for the isolates under study.

PCR amplifications were performed as previously described [20]. Primers used in this work are listed (see Supplementary Table S7). Sequencing in both directions was performed by StabVida (Portugal) and the assembly was done in the SeqMan module of DNASTAR software (Madison, WI, USA). All novel sequences were lodged in GenBank accession numbers MH346035 to MH346120, and alignments and phylogenetic trees in TREEBASE under reference TB2:S24331.

4.7. Phylogenetic Analyses

Sequences of the ApMAT, *gs*, and *tub2* previously available either from *Colletotrichum kahawae* sensu lato or from neighboring taxa, including *C. camelliae* Masee, *C. clidemiae* Weir and Johnst, *C. fructivorum* Doyle, Oudem, and Rehner, *C. helleniense* Guarnaccia and Crous, *C. jiangxiense* Liu and Cai, *C. rhexiae* Ellis and Everh., *C. temperatum* Doyle, Oudem, and Rehner and *C. wuxiense* Wang, Wang, and Yang [7,22–25] were download from Genbank (see Supplementary Table S4).

All the sequences obtained were used to performed two different analysis: one comparing all genes/regions sequenced for the isolates used in this study, and the other comparing these isolates

with others, not physically handled in this study, using sequences retrieved from the GenBank for a set of genes limited by availability (namely *gs*, *tub2*, and *ApMAT*).

The DNA sequences were aligned using MAFFT version 7 [52] and the alignments edited manually, if necessary, in MEGA7 [53]. The alignments for each locus were combined in a single file using the program SequenceMatrix 1.8 [54]. Prior to Bayesian inference, the best nucleotide substitution models for each locus were calculated in JMODELTEST 2.1.10 [55], with the following likelihood settings: number of substitution schemes = 3 (24 models), base frequencies (+F), proportion of invariable sites (+I), and rate variation among sites (+G) (nCat = 4). The models were selected according to the Akaike information criterion. MrBayes 3.2.6 [56] was used to perform the Bayesian analyses of the combined seven-loci dataset and individual locus data. The Markov Chain Monte Carlo sampling was set to 10 million generations, with two independent runs with four chains, one cold chain and three heated chains with a temperature of 0.2. The trees samples of the two cold chains were compared every 1000 generations and stopped when the average standard deviation of split frequencies fall below 0.01. Burn-in was set at 25% after which the likelihood values were stationary, and the remaining trees were used to calculate posterior probabilities. Trees from different runs were then combined and summarized in a majority rule 50% consensus tree. Maximum likelihood (ML) was implemented in the CIPRES Science Gateway V 3.357 [57] using RAXML-HPC 8 on XSEDE (8.2.9) [58] using the GTRCAT model and 1000 rapid bootstrap inferences.

Species boundaries were estimated using Bayesian Poisson tree processes method [59] (BPTP), using the nine-loci concatenated tree obtained in RAXML as input data. The calculations were conducted on the BPTP webserver, with 500,000 MCMC generations, with a thinning every 100 generations and a burn-in of 10%, the outgroup was excluded from the analysis. The convergence was checked by looking to the likelihood trace plot. The probability of each node to represent a species node was calculated using the maximum-likelihood solution.

4.8. Bioinformatic Prediction Tools

Gene predictions were performed in the program AUGUSTUS [60] training with the genome of *Fusarium graminearum* in order to compare the protein sequence from CBD-causing isolates and non-CBD causing isolates obtained for genes *cas1*, *cellwall*, *siRNA*, and *vosA*. The proteins obtained were aligned in MEGA7 [53], using MUSCLE and the amino acid differences were registered. The protein sequence analysis was done through a combination of INTERPRO [61], UNIPROT [62], HMMER [63], and PROSITE [64]. Secreted proteins were predicted using a battery of tools: SignalP (v4.1) [65], TargetP [66], and Phobius [67]. TMHMM [68] was used to exclude sequences with transmembrane domains. Predictions of serine, threonine, or tyrosine phosphorylation sites including generic and kinase specific were done by NETPHOS 3.1 [69]. Predictions for O- β -GlcNAc attachment sites were performed in by YinOYang [70] server. GPS-LIPID [71] was used as a predictor for protein lipid modification sites. To predict and analyze protein structure, function, and mutations, the Protein Model Portal [72] and the PHYRE2 [73] servers were used.

4.9. Genome Size

The genome size was estimated for isolates listed in Supplementary Table S6. A Flow Cytometry protocol using propidium iodide-stained nuclei [30] was followed using the ascomycete *Cenococcum geophilum* Fr. (isolate 844.1, 1C = 0.208 pg/203 Mbp [74,75]) as the internal standard.

5. Conclusions

In this work, *C. kahawae* subsp. *cigarro* is renamed *C. cigarro*, whereas *C. kahawae* subsp. *kahawae* is re-elevated to the species status, in which only the isolates causing the Coffee Berry Disease are accommodated. Over the last few decades, taxonomic criteria in *Colletotrichum* have moved from morphology to molecular phylogeny. In this work we have combined phylogenetic and pathological data, among others, to demonstrate that high divergence is found at many levels supporting the

recognition of distinct species, even though a remarkable genetic similarity shaped by a recent speciation event is still evident. Adding to the phylogenetic and biological relevance of these findings, the differentiation between *C. kahawae* and *C. cigarro* is of great importance for plant pathologists, plant breeders, and quarantine authorities to whom an accurate nomenclature is crucial to better distinguish CBD from non-CBD pathogens [4–13]. Since the causative agent of CBD, *C. kahawae*, is currently circumscribed to Africa, where annually determines heavy production losses, its spread around the world is a threat to the coffee production, leading for example the Australian and Chinese authorities to consider it as a quarantine pathogen. Resolving taxonomic ambiguities can be of relevance both to biology and agronomy, helping to improve food security [76], and distinguishing between the CBD pathogen, *C. kahawae*, and the cosmopolitan and polyphagous *C. cigarro* is such an example.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2223-7747/9/4/502/s1>, Figure S1: *Colletotrichum* species grown on 90 mm plates contained citric acid (left side of each letter) or ammonium tartrate (right side of each letter) as sole carbon sources, incubated at 25 °C and medium colour registered after 7 days as purple (substrate used) or yellow (substrate not used). Figure S2: Fifty percent majority rule consensus tree from a Bayesian analysis based single loci *apn25L* (part of *Apn2* gene), *ApMAT* (intergenic spacer between the 3' end of the *apn2* gene and the mating type gene *mat1-2-1*), *mat1-2-1* (mating-type gene), *gs* (intron 2 of the glutamine synthetase gene), *cas1* (appressorium specific protein), *cellwall* (cell wall protein), *siRNA* (a putative argonaute siRNA chaperone complex subunit), *vosA* (developmental regulator), and *tub2* (beta-tubulin) for isolates under study. Figure S3: Maximum likelihood solution generated by bPTP (a Bayesian implementation of the Poisson tree process model) based on the nine-locus concatenated tree obtained in RaXML. Figure S4: Graph illustrating the predictions of potential phosphorylation sites of cell wall protein of CBD-causing isolates and non-CBD causing isolates. Figure S5: O-GlcNAc and NetPhos predictions across the length of the sequences of cell wall protein of CBD-causing isolates (A) and non-CBD causing isolates. Table S1: Results of pathogenicity tests at 14 d after inoculation of the green coffee berries expressed as a disease severity score. Table S2: Colony diameter (mean ± standard deviation) measured seven days after inoculation of *Colletotrichum* isolates under study. Table S3: Conidia and appressoria size of *Colletotrichum* isolates. Table S4: Information about the *Colletotrichum* isolates used in phylogenetic analyses. Table S5: Summary of structural predictions of cell wall protein of Coffee Berry Disease (CBD)-causing isolates and non-CBD causing isolates. Table S6: Genome size (mean ± standard deviation) measured in *Colletotrichum* spp. propidium iodide-stained nuclei by Flow Cytometry using *Cenococcum geophilum* as internal standard. Table S7: PCR primers used in this study.

Author Contributions: Conceptualization, A.C., H.G.A., P.T. and V.V.; methodology, A.C., H.G.A. and P.T.; formal analysis, A.C., H.G.A. and P.T.; resources, A.P.R. and V.V.; writing—Original draft preparation, A.C., H.G.A. and P.T.; writing—review and editing, A.C., H.G.A., P.T., D.B., A.P.R., M.d.C.S., H.O. and V.V.; Supervision, H.O. and V.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Portuguese national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P. in the context of the Transitory Norm DL 57/2016/CP1382/CT0010 for Ana Cabral, DL 57/2016/CP1382/CT0023 for Pedro Talhinhos, and DL57/2016/CP1479/CT0002 for Dora Batista, and FCT Unit funding UIDB/04129/2020 (LEAF).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. von Arx, J.A. Die arten der gattung *Colletotrichum* Cda. *Phytopathol. Z.* **1957**, *29*, 413–468.
2. Sutton, B.C. *The Coelomycetes. Fungi Imperfecti with Pycnidia, Acervuli and Stromata*; CABI: Kew, UK, 1980.
3. Cannon, P.F.; Damm, U.; Johnston, P.R.; Weir, B.S. *Colletotrichum*—Current status and future directions. *Stud. Mycol.* **2012**, *73*, 181–213. [[CrossRef](#)]
4. Jayawardena, R.S.; Hyde, K.D.; Damm, U.; Cai, L.; Liu, M.; Li, X.H.; Zhang, W.; Zhao, W.S.; Yan, J.Y. Notes on currently accepted species of *Collet. Mycosphere* **2016**, *7*, 1192–1260. [[CrossRef](#)]
5. Baroncelli, R.; Talhinhos, P.; Pensec, F.; Sukno, S.A.; Le Floch, G.; Thon, M.R. The *Colletotrichum acutatum* species complex as a model system to study evolution and host specialization in plant pathogens. *Front. Microbiol.* **2017**, *8*, 2001. [[CrossRef](#)] [[PubMed](#)]
6. Marin-Felix, Y.; Groenewald, J.Z.; Cai, L.; Chen, Q.; Marincowitz, S.; Barnes, I.; Bensch, K.; Braun, U.; Camporesi, E.; Damm, U.; et al. Genera of phytopathogenic fungi: GOPHY 1. *Stud. Mycol.* **2017**, *86*, 99–216. [[CrossRef](#)] [[PubMed](#)]
7. Weir, B.; Johnston, P.R.; Damm, U. The *Colletotrichum gloeosporioides* species complex. *Stud. Mycol.* **2012**, *73*, 115–180. [[CrossRef](#)]

8. Sreenivasaprasad, S.; Brown, A.E.; Mills, P.R. Coffee berry disease pathogen in Africa: Genetic structure and relationship to the group species *Colletotrichum gloeosporioides*. *Mycol. Res.* **1993**, *97*, 995–1000. [[CrossRef](#)]
9. Waller, J.M.; Bridge, P.D.; Black, R.; Hakiza, G. Characterization of the coffee berry disease pathogen, *Colletotrichum kahawae* sp. nov. *Mycol. Res.* **1993**, *97*, 989–994. [[CrossRef](#)]
10. van der Vossen, H.A.M.; Walyaro, D.J. Additional evidence for oligogenic inheritance of durable host resistance to coffee berry disease (*Colletotrichum kahawae*) in Arabica coffee (*Coffea arabica* L.). *Euphytica* **2009**, *165*, 105–111. [[CrossRef](#)]
11. Loureiro, A.; Nicole, M.R.; Várzea, V.; Moncada, P.; Bertrand, B.; Silva, M.C. Coffee resistance to *Colletotrichum kahawae* is associated with lignification, accumulation of phenols and cell death at infection sites. *Physiol. Mol. Plant Pathol.* **2012**, *77*, 23–32. [[CrossRef](#)]
12. de Silva, D.D.; Crous, P.W.; Ades, P.K.; Hyde, K.D.; Taylor, P.W.J. Life styles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol. Rev.* **2017**, *31*, 155–168. [[CrossRef](#)]
13. Batista, D.; Silva, D.N.; Vieira, A.; Cabral, A.; Pires, A.S.; Loureiro, A.; Guerra-Guimarães, L.; Pereira, A.P.; Azinheira, H.; Talhinhos, P.; et al. Legitimacy and implications of reducing *Colletotrichum kahawae* to subspecies in plant pathology. *Front. Plant Sci.* **2017**, *7*, 2051. [[CrossRef](#)] [[PubMed](#)]
14. Vieira, A.; Silva, D.N.; Várzea, V.; Paulo, O.S.; Batista, D. Novel insights on colonization routes and evolutionary potential of *Colletotrichum kahawae*, a severe pathogen of *Coffea arabica*. *Mol. Plant Pathol.* **2018**, *19*, 2488–2501. [[CrossRef](#)] [[PubMed](#)]
15. Australia Group Common Control List Handbook—Volume II: Biological Weapons-Related Common Control Lists. Available online: <https://australiagroup.net/en/documents/Australia-Group-Common-Control-List-Handbook-Volume-II.pdf> (accessed on 6 March 2020).
16. Hindorf, H. *Colletotrichum* spp. isolated from *Coffea arabica* L. in Kenya. *Z. PflKrankh. PflSchutz.* **1970**, *77*, 328–331.
17. Prihastuti, H.; Cai, L.; Chen, H.; McKenzie, E.H.C.; Hyde, K.D. Characterisation of *Colletotrichum* species associated with coffee berries in northern Thailand. *Fungal Divers.* **2009**, *39*, 89–109.
18. Gibbs, J.N. Inoculum sources for coffee berry disease. *Ann. Appl. Biol.* **1969**, *64*, 515–522. [[CrossRef](#)]
19. Hindorf, H. *Colletotrichum*-Arten aus dem Kaffeeanbaugebiet von Kiambu in Kenia. *Z. PflKrankh. PflSchutz.* **1974**, *81*, 108–113.
20. Silva, D.N.; Talhinhos, P.; Várzea, V.; Cai, L.; Paulo, O.S.; Batista, D. Application of the Apn2/MAT locus to improve the systematics of the *Colletotrichum gloeosporioides* complex: An example from coffee (*Coffea* spp.) hosts. *Mycologia* **2012**, *104*, 396–409. [[CrossRef](#)]
21. Várzea, V.; Rodrigues, J.C.; Lewis, B. Distinguishing characteristics and vegetative compatibility of *Colletotrichum kahawae* in comparison with other related species from coffee. *Plant Pathol.* **2002**, *51*, 202–207.
22. Doyle, V.P.; Oudemans, P.V.; Rehner, S.A.; Litt, A. Habitat and host indicate lineage identity in *Colletotrichum gloeosporioides* s.l. from wild and agricultural landscapes in North America. *PLoS ONE* **2013**, *8*, e62394. [[CrossRef](#)]
23. Liu, F.; Weir, B.S.; Damm, U.; Crous, P.W.; Wang, Y.; Liu, B.; Wang, M.; Zhang, M.; Cai, L. Unravelling *Colletotrichum* species associated with *Camellia*: Employing ApMat and GS loci to resolve species in the *C. gloeosporioides* complex. *Persoonia* **2015**, *35*, 63–86. [[CrossRef](#)] [[PubMed](#)]
24. Wang, Y.-C.; Hao, X.-Y.; Wang, L.; Xiao, B.; Wang, X.-C.; Yang, Y.-J. Diverse *Colletotrichum* species cause anthracnose of tea plants (*Camellia sinensis* (L.) O. Kuntze) in China. *Sci. Rep.* **2016**, *6*, 35287. [[CrossRef](#)] [[PubMed](#)]
25. Guarnaccia, V.; Groenewald, J.Z.; Polizzi, G.; Crous, P.W. High species diversity in *Colletotrichum* associated with citrus diseases in Europe. *Persoonia* **2017**, *39*, 32–50. [[CrossRef](#)] [[PubMed](#)]
26. Giraud, T.; Gladieux, P.; Gavrillets, S. Linking the emergence of fungal plant diseases with ecological speciation. *Trends Ecol. Evol.* **2010**, *25*, 387–395. [[CrossRef](#)] [[PubMed](#)]
27. Silva, D.N.; Talhinhos, P.; Cai, L.; Manuel, L.; Gichuru, E.K.; Loureiro, A.; Várzea, V.; Paulo, O.S.; Batista, D. Host-jump drives rapid and recent ecological speciation of the emergent fungal pathogen *Colletotrichum kahawae*. *Mol. Ecol.* **2012**, *21*, 2655–2670. [[CrossRef](#)]
28. Crouch, J.A.; Tredway, L.; Clarke, B.; Hillman, B. Phylogenetic and population genetic divergence correspond with habitat for the pathogen *Colletotrichum cereale* and allied taxa across diverse grass communities. *Mol. Ecol.* **2009**, *18*, 123–135. [[CrossRef](#)]

29. Lardner, R.; Johnston, P.R.; Plummer, K.M.; Pearson, M.N. Morphological and molecular analysis of *Colletotrichum acutatum* sensu lato. *Mycol. Res.* **1999**, *103*, 275–285. [CrossRef]
30. Pires, A.S.; Azinheira, H.G.; Cabral, A.; Tavares, S.; Tavares, D.; Castro, M.; Várzea, V.; Silva, M.C.; Abranches, R.; Loureiro, J.; et al. Cytogenomic characterisation of *Colletotrichum kahawae*, the causal agent of Coffee Berry Disease, reveals diversity in minichromosome profiles and genome size expansion. *Plant Pathol.* **2016**, *65*, 968–977. [CrossRef]
31. Jmol: An Open-Source Java Viewer for Chemical Structures in 3D. Available online: <http://jmol.sourceforge.net/> (accessed on 1 October 2019).
32. Garibaldi, A.; Gilardi, G.; Puglisi, I.; Cacciola, S.O.; Gullino, M.L. First report of leaf spot caused by *Colletotrichum kahawae* on cultivated rocket (*Eruca sativa*) in Italy. *Plant Dis.* **2016**, *100*, 1240. [CrossRef]
33. Garibaldi, A.; Gilardi, G.; Franco-Ortega, S.; Gullino, M.L. First report of leaf spot caused by *Colletotrichum kahawae* on American Sweetgum (*Liquidambar styraciflua*) in Italy. *J. Plant Pathol.* **2016**, *95*, 36.
34. Coronado-Ruiz, C.; Avendaño, R.; Escudero-Leyva, E.; Conejo-Barboza, G.; Chaverri, P.; Chavarria, M. Two new cellulolytic fungal species isolated from a 19th-century art collection. *Sci. Rep.* **2018**, *8*, 1–9. [CrossRef]
35. Grammen, A.; Wenneker, M.; van Campenhout, J.; Pham, K.T.K.; van Hemelrijck, W.; Bylemans, D.; Geeraerd, A.; Keulemans, W. Identification and pathogenicity assessment of *Colletotrichum* isolates causing bitter rot of apple fruit in Belgium. *Eur. J. Plant Pathol.* **2019**, *153*, 253–269. [CrossRef]
36. Sharma, G.; Shenoy, B. *Colletotrichum* systematics: Past, present and prospects. *Mycosphere* **2016**, *7*, 1093–1102. [CrossRef]
37. Liang, X.; Wang, B.; Dong, Q.; Li, L.; Rollins, J.A.; Zhang, R.; Sun, G. Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. *PLoS ONE* **2018**, *13*, e0196303. [CrossRef] [PubMed]
38. Ahmed, Y.L.; Gerke, J.; Park, H.-S.; Bayram, Ö.; Neumann, P.; Ni, M.; Dickmanns, A.; Kim, S.C.; Yu, J.-H.; Braus, G.H.; et al. The Velvet Family of Fungal Regulators Contains a DNA-Binding Domain Structurally Similar to NF- κ B. *PLoS Biol.* **2013**, *11*, e1001750. [CrossRef]
39. Ni, M.; Yu, J.H. A novel regulator couples sporogenesis and trehalose biogenesis in *Aspergillus nidulans*. *PLoS ONE* **2007**, *2*, e970. [CrossRef] [PubMed]
40. Gupta, R.; Brunak, S. Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac. Symp. Biocomput.* **2002**, *7*, 310–322.
41. Gan, P.; Ikeda, K.; Irieda, H.; Narusaka, M.; O’Connell, R.J.; Narusaka, Y.; Takano, Y.; Kubo, Y.; Shirasu, K. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol.* **2013**, *197*, 1236–1249. [CrossRef] [PubMed]
42. Gan, P.; Narusaka, M.; Kumakura, N.; Tsushima, A.; Takano, Y.; Narusaka, Y.; Shirasu, K. Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol. Evol.* **2016**, *8*, 1467–1481. [CrossRef] [PubMed]
43. Lundin, D.; Poole, A.M.; Sjöberg, B.M.; Högbom, M. Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *J. Biol. Chem.* **2012**, *287*, 20565–20575. [CrossRef]
44. Spanu, P.D.; Abbott, J.C.; Amselem, J.; Burgis, T.A.; Soanes, D.M.; Stüber, K.; van Themaat, E.V.L.; Brown, J.K.M.; Butcher, S.A.; Gurr, S.J.; et al. Genome expansion and gene loss in powdery mildew fungi reveal tradeoffs in extreme parasitism. *Science* **2010**, *330*, 1543–1546. [CrossRef] [PubMed]
45. Tavares, S.; Ramos, A.P.; Pires, A.S.; Azinheira, H.G.; Caldeirinha, P.; Link, T.; Abranches, R.; Silva, M.C.; Voegelé, R.T.; Loureiro, J.; et al. Genome size analyses of Pucciniales reveal the largest fungal genomes. *Front. Plant Sci.* **2014**, *5*, 422. [CrossRef] [PubMed]
46. Talhinas, P.; Neves-Martins, J.; Oliveira, H.; Sreenivasaprasad, S. The distinctive population structure of *Colletotrichum* species associated with olive anthracnose in the Algarve region of Portugal reflects a host-pathogen diversity hot spot. *FEMS Microbiol. Lett.* **2009**, *296*, 31–38. [CrossRef] [PubMed]
47. Loureiro, A.; Guerra-Guimarães, L.; Lidon, F.C.; Bertrand, B.; Silva, M.C.; Várzea, V. Isoenzymatic characterization of *Colletotrichum kahawae* isolates with different levels of aggressiveness. *Trop. Plant Pathol.* **2011**, *36*, 287–293. [CrossRef]
48. Guerber, J.C.; Correll, J.C. Morphological description of *Glomerella acutata*, the teleomorph of *Colletotrichum acutatum*. *Mycologia* **2001**, *93*, 216–229. [CrossRef]

49. Nirenberg, H. Untersuchungen über die morphologische und biologische differenzierung in der *Fusarium*-Sektion *Liseola*. *Mitt. Biol. Bundesanst. Land Forstwirtschaft. Berl Dahl.* **1976**, *169*, 1–117.
50. Cenis, J.L. Rapid extraction of fungal DNA for PCR amplification. *Nucleic Acids Res.* **1992**, *20*, 2380. [[CrossRef](#)]
51. Lehr, N.A.; Wang, Z.; Li, N.; Hewitt, D.A.; López-Giráldez, F.; Trail, F.; Townsend, J.P. Differences among Three *Neurospora* Species Reveal Genes Required for Sexual Reproduction in *Neurospora crassa*. *PLoS ONE* **2014**, *9*, e110398. [[CrossRef](#)]
52. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **2017**, *20*, 1160–1166. [[CrossRef](#)]
53. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
54. Vaidya, G.; Lohman, D.J.; Meier, R. Sequence Matrix: Concatenation software for the fast assembly of multigene datasets with character set and codon information. *Cladistics* **2011**, *27*, 171–180. [[CrossRef](#)]
55. Darriba, D.; Taboada, G.L.; Doallo, R.; Posada, D. jModelTest 2: More models, new heuristics and parallel computing. *Nat. Methods* **2012**, *9*, 772. [[CrossRef](#)]
56. Ronquist, F.; Teslenko, M.; Van Der Mark, P.; Ayers, D.L.; Darling, A.; Höhna, S.; Larget, B.; Liu, L.; Suchard, M.A.; Huelsenbeck, J.P. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large modelspace. *Syst. Biol.* **2012**, *61*, 539–542. [[CrossRef](#)]
57. Miller, M.A.; Pfeiffer, W.; Schwartz, T. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Proceedings of the Gateway Computing Environments Workshop (GCE), New Orleans, LA, USA, 14 November 2010; Institute of Electrical and Electronics Engineers: Piscataway, NJ, USA, 2010.
58. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [[CrossRef](#)] [[PubMed](#)]
59. Zhang, J.; Kapli, P.; Pavlidis, P.; Stamatakis, A. A General Species Delimitation Method with Applications to Phylogenetic Placements. *Bioinformatics* **2013**, *29*, 2869–2876. [[CrossRef](#)] [[PubMed](#)]
60. Stanke, M.; Schoffmann, O.; Morgenstern, B.; Waack, S. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinf.* **2006**, *7*, 62. [[CrossRef](#)] [[PubMed](#)]
61. Mitchell, A.L.; Attwood, T.K.; Babbitt, P.C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S.D.; Chang, H.-Y.; El-Gebali, S.; Fraser, M.I.; et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2018**, *47*, D351–D360. [[CrossRef](#)]
62. The UniProt Consortium UniProt: The universal protein knowledgebase. *Nucleic Acids Res.* **2018**, *46*, 2699. [[CrossRef](#)]
63. Potter, S.C.; Luciani, A.; Eddy, S.R.; Park, Y.; Lopez, R.; Finn, R.D. HMMER web server: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W200–W204. [[CrossRef](#)]
64. de Castro, E.; Sigrist, C.J.A.; Gattiker, A.; Bulliard, V.; Langendijk-Genevaux, P.S.; Gasteiger, E.; Bairoch, A.; Hulo, N. ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **2006**, *34*, W362–W665. [[CrossRef](#)] [[PubMed](#)]
65. Petersen, T.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)] [[PubMed](#)]
66. Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* **2000**, *300*, 1005–1016. [[CrossRef](#)] [[PubMed](#)]
67. Käll, L.; Krogh, A.; Sonnhammer, E. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res.* **2007**, *35*, W429–W432. [[CrossRef](#)] [[PubMed](#)]
68. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [[CrossRef](#)] [[PubMed](#)]
69. Blom, N.; Gammeltoft, S.; Brunak, S. Sequence- and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **1999**, *294*, 1351–1362. [[CrossRef](#)]
70. Gupta, R. Prediction of Glycosylation Sites in Proteomes: From Post-Translational Modifications to Protein Function. Ph.D. Thesis, Technical University of Denmark (DTU), Lyngby, Denmark, 2001.
71. Xie, Y.; Zheng, Y.; Li, H.; Luo, X.; He, Z.; Cao, S.; Shi, Y.; Zhao, Q.; Xue, Y.; Zuo, Z.; et al. GPS-Lipid: A robust tool for the prediction of multiple lipid modification sites. *Sci. Rep.* **2016**, *6*, 28249. [[CrossRef](#)]
72. Haas, J.; Roth, S.; Arnold, K.; Kiefer, F.; Schmidt, T.; Bordoli, L.; Schwede, T. The Protein Model Portal—A comprehensive resource for protein structure and model information. *Database* **2013**, *2013*, bat031. [[CrossRef](#)]

73. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845–858. [[CrossRef](#)]
74. Bourne, E.C.; Mina, D.; Gonçalves, S.C.; Loureiro, J.; Freitas, H.; Muller, L.A.H. Large and variable genome size unrelated to serpentine adaptation but supportive of cryptic sexuality in *Cenococcum geophilum*. *Mycorrhiza* **2014**, *24*, 13–20. [[CrossRef](#)]
75. Talhinhas, P.; Tavares, D.; Ramos, A.P.; Gonçalves, S.; Loureiro, J. Validation of Standards Suitable for Genome Size Estimation of Fungi. *J. Microbiol. Methods* **2017**, *142*, 76–78. [[CrossRef](#)]
76. Crous, P.W.; Groenewald, J.Z.; Slippers, B.; Wingfield, M.J. Global food and fibre security threatened by current inefficiencies in fungal identification. *Phil Trans. R. Soc. B* **2016**, *371*, 20160024. [[CrossRef](#)] [[PubMed](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).