Mycologia, 104(2), 2012, pp. 396–409. DOI: 10.3852/11-145

Application of the *Apn2/MAT* locus to improve the systematics of the *Colletotrichum gloeosporioides* complex: an example from coffee (*Coffea* spp.) hosts

Diogo Nuno Silva

Centro de Investigação das Ferrugens do Cafeeiro (CIFC)/Instituto de Investigação Científica Tropical (IICT), Quinta do Marquês, 2784-505 Oeiras, Portugal, and Computational Biology and Population Genomics group, Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

Pedro Talhinhas

Vítor Várzea[1]

Centro de Investigação das Ferrugens do Cafeeiro (CIFC)/Instituto de Investigação Científica Tropical (IICT), Quinta do Marquês, 2784-505 Oeiras, Portugal

Lei Cai[2]

State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, No.10, North 4th Ring Road West, Beijing 100190, People's Republic of China

Octávio Salgueiro Paulo

Computational Biology and Population Genomics group, Centro de Biologia Ambiental, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal

Dora Batista

Centro de Investigação das Ferrugens do Cafeeiro (CIFC)/Instituto de Investigação Científica Tropical (IICT), Quinta do Marquês, 2784-505 Oeiras, Portugal

Abstract: To improve phylogenetic resolution of the *Colletotrichum gloeosporioides* species complex we developed and tested the performance of a new set of primers for the *Apn2/MAT* locus with a case study of 22 isolates. These were isolated mainly from coffee plants and represent six divergent and well characterized species within the *C. gloeosporioides* complex. Following previous studies on this locus, we have generated sequence data from an expanded region (> 4600 bp), revealing increased phylogenetic informativeness when compared to other commonly used markers such as ITS, β-tub2 and GS. Within the *Apn2/MAT* locus the ApMAT marker alone was almost as informative in terms of phylogenetic resolution as a seven-gene concatenated dataset. Our results further revealed that gene-tree discordance may come to be a common issue in resolving evolutionary relationships in the *C. gloeosporioides* complex, highlighting the importance of multilocus approaches. The use of state-of-the-art data analysis techniques and a highly informative dataset as employed here may abate this issue and hopefully assist in disentangling the *C. gloeosporioides* complex.

Key words: BEST (Bayesian estimate of species tree), gene-tree discordance, mating-type locus, phylogenetics

Introduction

Disentangling species complexes in the *Colletotrichum* genus will remain a considerable challenge for taxonomists in the years to come. Given the worldwide importance of this genus, mainly comprising plant pathogens, it is essential to accurately identify species and/or pathotypes to improve biosecurity and disease control (Johnston and Jones 1997, Freeman et al. 1998, Cai et al. 2009). Morphological, cultural and host-preference criteria have been the primary basis for species identification and delimitation. However, due to the unreliability and limited number of those diagnostic characteristics, studies based on such criteria have led to longstanding and unresolved taxonomic issues (Cannon et al. 2000, Afanador-Kafuri et al. 2002, Du et al. 2005, Hyde et al. 2009a). As Sutton (1992) stated, morphology alone is unlikely to provide enough information to improve systematics of species complexes, such as *C. gloeosporioides* (e.g. the graminicolous group, Du et al. 2005; the curved spore/dematium group, Damm et al. 2009).

C. gloeosporioides is regarded as the most challenging species complex to resolve, comprising the broadest host range of all *Colletotrichum* species (Du et al. 2005). Fungal strains from this complex have been reported from at least 1800 plant hosts and present such a wide range of morphological and pathogenic variation that the species name is of limited practical use (Hyde et al. 2009a). A major contribution for the enlargement of this complex was the revision of the genus by von Arx (1957) in which a wider spore range was established as a criterion for *C. gloeosporioides* identification (Sutton 1992). As a result, many *Colletotrichum* strains with similar cylindrical conidia were identified as *C. gloeosporioides*, lumping together potentially distinct taxa (Hindorf et al. 1970, Munaut et al. 1998, Freeman et al. 2000, Chakraborty et al. 2002, Xiao et al. 2004, Hyde et al. 2009b). Since

Submitted 11 May 2011; accepted for publication 8 Sep 2011.

[1] Corresponding author. E-mail: vitorvarzea@clix.pt

[2] E-mail: mrcailei@gmail.com

the revision by Sutton (1992) until the most recent major revision of the genus (Hyde et al. 2009a) most studies on *C. gloeosporioides* systematics (e.g. Abang et al. 2002, Lubbe et al. 2004, Xiao et al. 2004, Suzuki et al. 2010) addressed only specific hosts, often reclassifying isolates previously described as *C. gloeosporioides* into other *Colletotrichum* species (e.g. Gunnel and Gubler 1992, Sreenivasaprasad et al. 1994, Ramos et al. 2006). In fact, the exceptionally broad host range of *C. gloeosporiodes* hampers studies addressing the entire complex.

The recent epitypification of *C. gloeosporioides* set an important milestone for taxonomic studies of the *C. gloeosporioides* complex (Cannon et al. 2008). Ever since, the availability of living ex-type strains and sequence data has provided a solid reference basis to which isolates can be compared. However, the criteria for species delimitation have been subject of much debate throughout the years and *Colletotrichum* has been an example (Cannon et al. 2008). Until recently, single-locus phylogenies dominated molecular systematics of the *C. gloeosporioides* complex, despite their limited resolving ability (Sreenivasaprasad et al. 1993, 1996; Johnston and Jones 1997, Nguyen et al. 2009, Polashock et al. 2009). In particular, the internal transcribed spacers (ITS) flanking the 5.8 S ribosomal RNA is the most widely sequenced region in *Colletotrichum* phylogenetic studies. Consequently, ITS-based *C. gloeosporioides*-specific primers (Mills et al. 1992) have been widely used for identification of *Colletotrichum* fungi over the past 10–15 y. However there are serious concerns about the resolving power of ITS for closely related species (Cai et al. 2009, Crouch et al. 2009b). The limitations of ITS and other similar markers are already recognized, and thus the development and use of more informative loci has become increasingly necessary.

Molecular phylogenetic analyses using multilocus datasets are now providing an objective method to determine species boundaries—particularly by using the genealogical concordance phylogenetic species recognition (Taylor et al. 2000)—and also to better understand species' evolutionary relationships (Crouch et al. 2006, 2009a). The synergy of epitypification with the emergence of multilocus phylogenetic analyses provides an opportunity to unravel the species richness of the *C. gloeosporioides* complex. For instance studies focused on Arabica coffee plants (*Coffea arabica*) in Thailand and on cacao (*Theobroma cacao*) and other plant species in Panama were able to identify six new species from the *C. gloeosporioides* complex that represented phylogenetically distinct lineages from the ex-epitype strain of *C. gloeosporioides* (Prihastuti et al. 2009, Rojas et al. 2010).

Here we report the development and use of several primer sets for the *Apn2/MAT* locus to improve systematic resolution and knowledge of the *C. gloeosporioides* complex, using species inhabiting *Coffea* spp. As shown by Crouch et al. (2009a), sequence data from this locus allowed the successful resolution of the graminicolous *Colletotrichum* group. Rojas et al. (2010) have also amplified and sequenced an orthologous region in a phylogenetic study of *C. gloeosporioides* s.l., further revealing its usefulness. In the present study, we have further exploited this region for primer development and assessed the usefulness of the locus compared to other commonly used nuclear genes or regions, focusing on a well characterized group of species from coffee hosts. These included fairly divergent and well recognized species, such as the coffee berry disease (CBD) pathogen, *C. kahawae*, and other opportunistic pathogens characterized by Prihastuti et al. (2009) that occur on a number of other hosts as well as representatives of *C. gloeosporioides* s.s. We also applied a recent Bayesian data analysis technique for species tree reconstruction and compared it to the routinely used method of concatenation (Liu 2008).

MATERIALS AND METHODS

Fungal material and DNA extraction.—*Colletotrichum* spp. isolates used in this study is provided (TABLE I). Nineteen isolates were obtained from the collection maintained at CIFC/IICT (Portugal) and three from the Mae Fah Luang University collection (Thailand), representing six previously characterized species from the *C. gloeosporioides* complex, namely *C. fragariae*, *C. gloeosporioides*, *C. kahawae*, *C. asianum*, *C. siamense* and *C. fructicola*, including ex-type specimens for the three latter species. *C. kahawae* isolates were identified based on BLAST queries of the ITS and *β-tubulin 2* (β-tub2) sequences compared to the holotype (GenBank accession numbers FJ972608 [E value = 0, Max ident. = 100%] and FJ907432 [E value = 0, Max ident. = 100%]) and on their ability to cause CBD symptoms on detached green coffee berries. The latter was shown to be the most reliable method to identify true *C. kahawae* isolates because recent results revealed the existence of non-*C. kahawae* isolates with identical genetic sequences for several nuclear markers, including ITS and β-tub2 (data not shown). The *C. gloeosporioides* sampling comprised isolates that are similar to the ex-epitype strain at the ITS and β-tub2 loci (GenBank accession numbers EU371022 [E value = 0, Max ident. = 98%] and FJ907445 [E value = 0, Max ident. = 99%] respectively). Likewise, BLAST queries of the MAT1-2-1 HMG locus enabled the identification of five isolates as *C. fragariae* by comparison with the species epitype (GenBank accession number DQ002827 [E value = 9e-87, Max ident. 94%]). Except for isolates CR21, PT111 and PR220 (*C. gloeosporioides* s.s.), which were obtained from *Citrus limon* and *Olea europaea* (TABLE I), all other isolates

TABLE I. *Colletotrichum* isolates used in this study

Isolate	Species	Host	Origin	GenBank accession numbers								
				ITS	*b-tub2*	*GS*	*Apn15L*	*ApMAT*	*MAT1-2-1*	*MAT5L*	*Apn1Ex3*	*Apn13L*
Que2[a]	***C. kahawae***	*C. arabica*	Kenya	FR717712	FR719902	FR719921	FR719148	FR718818	FR719858	FR719880	FR718963	FR719127
Tan12	*C. kahawae*	*C. arabica*	Tanzania	FR717713	FR719903	FR719922	FR719149	FR718819	FR719859	FR719881	N/S	N/S
Mal2	*C. kahawae*	*C. arabica*	Malawi	FR717711	FR719901	FR719920	FR719147	FR718817	FR719857	FR719879	N/S	N/S
Bur2	*C. kahawae*	*C. arabica*	Burundi	FR717710	FR719900	FR719919	FR719146	FR718816	FR719856	FR719878	N/S	N/S
PT111	***C. gloeosporioides***	*Olea* sp.	Portugal	AJ749682	FR719888	FR719910	FR719129	FR718804	FR719845	FR719867	N/S	N/S
CR21[a]	*C. gloeosporioides*	*Citrus* sp.	Portugal	AJ300560	FR719887	FR719909	FR719128	FR718803	FR719844	FR719866	FR718956	FR719120
PR220	*C. gloeosporioides*	*Olea* sp.	Portugal	AJ749693	FR719889	FR719911	FR719130	FR718805	FR719846	FR719868	N/S	N/S
BDP-I 2[a]	***C. siamense***	*C. arabica*	Thailand	FR717707	FR719898	FJ972596	FR719143	FR718813	FR719854	FR719877	FR718961	FR719125
Mal5[a]	*C. siamense*[b]	*Coffea* sp.	Malawi	FR717709	FR719896	FR719918	FR719145	FR718815	FR719860	FR719875	FR718962	FR719126
Que77	*C. siamense*[b]	*Coffea* sp.	Kenya	FR717703	FR719895	FR719915	FR719141	FR718811	FR719852	FR719874	N/S	N/S
Bra9[a]	*C. siamense*[b]	*Coffea* sp.	Brazil	FR717702	FR719892	FR719914	FR719138	FR718808	FR719849	FR719871	FR718957	FR719121
Col1	*C. siamense*[b]	*Coffea* sp	Colombia	FR717705	FR719894	FR719917	FR719140	FR718810	FR719851	FR719873	N/S	N/S
Chi4[a]	*C. siamense*[b]	*Coffea* sp.	China	FR717704	FR719893	FR719916	FR719139	FR718809	FR719850	FR719872	FR718958	FR719122
BDP-I 4[a]	***C. asianum***	*C. arabica*	Thailand	FR717708	FR719899	FJ972595	FR719144	FR718814	FR719855	N/A	FR718959	FR719123
BDP-I 16[a]	***C. fructicola***	*C. arabica*	Thailand	FR717706	FR719897	FJ972599	FR719142	FR718812	FR719853	FR719876	FR718960	FR719124
Ang40	*C. fructicola*[b]	*Coffea* sp.	Angola	FR717695	FR719890	FR719912	FR719131	FR718806	FR719847	FR719869	N/S	N/S
Ang97	*C. fructicola*[b]	*Coffea* sp.	Angola	FR717699	FR719891	FR719913	FR719135	FR718807	FR719848	FR719870	N/S	N/S
Bra8	***C. fragariae***	*Coffea* sp.	Brazil	FR717701	FR719886	FR719908	FR719137	FR718802	FR719843	FR719865	N/S	N/S
Bra5[a]	*C. fragariae*	*Coffea* sp.	Brazil	FR717700	FR719885	FR719907	FR719136	FR718801	FR719842	FR719864	FR718955	FR719119
Ang52[a]	*C. fragariae*	*Coffea* sp.	Angola	FR717696	FR719882	FR719904	FR719132	FR718798	FR719839	FR719861	FR718954	FR719118
Ang91	*C. fragariae*	*Coffea* sp.	Angola	FR717698	FR719884	FR719906	FR719134	FR718800	FR719841	FR719863	N/S	N/S
Ang84	*C. fragariae*	*Coffea* sp.	Angola	FR717696	FR719883	FR719905	FR719133	FR718799	FR719840	FR719862	N/S	N/S

[a] Isolates included in the representative sample for sequencing of the full *Apn2/MAT* locus.

[b] Isolates received as *C. gloeosporioides* s.l.

N/S, Not sequenced

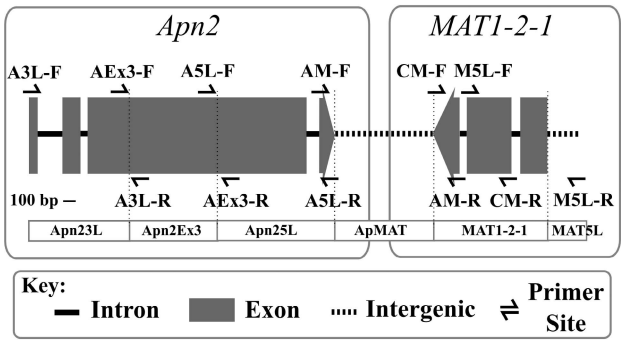

FIG. 1. *Apn2/MAT* locus map showing the position of the primers used in this study. The arrow-shaped end of the genes indicates the direction of transcription.

were obtained from *Coffea* spp. Isolates were revived on malt extract agar 1% with a bacterial inhibitor (KCNS, 50 mM) and grown 5–7 d. Isolates were grown in liquid media containing 3% malt extract and 0.5% Peptone 12–14 d at 25 C in the dark. DNA was extracted from freeze dried mycelia with the DNeasy plant Minikit (QIAGEN, Hilden, Germany) according to manufacturer's instructions.

Apn2/MAT *locus marker development strategy.*—A novel set of specific primers was developed to amplify a portion of the *Apn2/MAT* locus (FIG. 1), using *C. gloeosporioides* GenBank accession AY357890 as the main template. This genomic sequence spans 11 592 bp and includes the *Apn2/MAT* locus, which comprises two genes, *Apn2* (2244 bp) and *MAT1-2-1* (842 bp), connected by an intergenic region of 713 bp, and flanking regions. A representative sample of 10 isolates (TABLE I), comprising the six species, was used to amplify and sequence the full portion of this locus. Primers were designed with PerlPrimer 1.1.18 (Marshall 2004). The first two sets of specific primers, AM-F (5′- TCATTCTACG-TATGTGCCCG-3′)/AM-R (5′-CCAGAAATACACCGAA-CTTGC-3′) and M5L-F (5′-ATCTTTGCGGTAGAGAAT-GAAGG-3′)/M5L-R (5′-GACCCTTCTATGAACGAGCC-3′) were designed within and flanking the conserved HMG domain of *MAT1-2-1* (FIG. 1). These conditions were used for PCR amplification: 3 min at 94 C followed by 30 cycles of 45 s at 94 C, 45 s at 62 C and 1 min at 72 C, with a final extension of 7 min at 72 C. The amplification products were sequenced and used as templates along with the *C. gloeosporioides* accession, for further primer design. Based on these, primer pairs CM-F (5′-TCTACCTCATCGACG-CTGCT-3′)/CM-R (5′-CATGTGGGCAAAGGATGGC-3′) and A5L-F (5′-CAAGCGACGAAGTATACGAG-3′)/A5L-R (5′-GCATCACGGGAATAACTAGG-3′) were designed to amplify the remainder *MAT1-2-1* gene and the 5′ end of the *Apn2* gene respectively (FIG. 1). PCR conditions were identical, except for the use of a 66 C annealing temperature for the CM-F/CM-R pair. With the sequences obtained from the 5′ end of *Apn2* used as templates, primers AEx3-F (5′-CTCATCGGCACCTACAGC-3′)/AEx3-R (5′-CAAGATGTCTCCGAGCGTC-3′) and subsequently primers A3L-F (5′-TGACATGGAACGGTGAGTG-3′)/A3L-R (5′-TTCCAGTCCTCGACCGTCA-3′) were designed to amplify the remaining portion of the *Apn2* gene (FIG. 1). Amplification reactions were performed as follows: 3 min at 94 C followed by 30 cycles with 45 s at 94 C, touchdown annealing step for 45 s starting at 62 C and decreasing 0.5 C per cycle until stabilizing at 57 C during the remaining 20 cycles, 1 min at 72 C and a final extension of 7 min at 72 C. For all PCR reactions a total of 50 ng genomic DNA was used as template in a 15 μL PCR reaction mix containing 250 μM of dNTP mix (MBI Fermentas, Vilnius, Lithuania), 0.5–0.8 pmol each primer, 1 U DreamTaq DNA polymerase (MBI Fermentas) and 1× PCR buffer with 2 mM $MgCl_2$ (MBI Fermentas). The nomenclature of the molecular markers developed is as follows (from the 3′ of *Apn2* onward): Apn23L, with A3L-F/A3L-R; Apn2Ex3, with AEx3-F/AEx3-R; Apn25L, with A5L-F/A5L-R; ApMAT, with AM-F/AM-R; *MAT1-2-1* and MAT5L, with CM-F/M5L-R (FIG. 2). Sequences were deposited at GenBank under the accession numbers herein (TABLE I).

Molecular data.—Using the complete taxa sampling of this study (TABLE I), five nuclear genes and two intergenic regions were sequenced. From the *Apn2/MAT* locus (FIG. 1) the two intergenic regions (ApMAT and MAT5L) as well as the *MAT1-2-1* gene and the Apn25L gene fragment were selected, considering both their informative potential and sequence length. The ITS nuclear region (Brown et al. 1996) and two gene fragments from β-tub2 (O'Donnell and Cigelnik 1997) and *glutamine synthetase* (GS) (Prihastuti et al. 2009) were also analyzed with primers and PCR conditions previously described. PCR products were verified by electrophoresis on a 1.2% w/v agarose gel stained with ethidium bromide and purified with SureClean (Bioline, London, UK). Sequencing reactions were carried out with the BigDye 3.1 chemistry (Applied Biosystems, USA) on an ABI prism 310 automated sequencer. Amplicons were sequenced in both directions and chromatograms were manually verified for errors in Sequencher 4.0.5 (Gene Codes Corp., Ann Arbor, Michigan).

Data analysis.—Datasets for each marker were aligned in Mafft 6.717b (Katoh et al. 2009) with the L-INS-i method, followed by manual refinement in BioEdit 7.0.5.1. Using all studied loci, the nucleotide diversity (π) and its standard deviation were estimated for each selected marker in DNasp 5 (Rozas et al. 2003). In addition, to compare the polymorphisms between and among all used markers in the representative sample, a sliding window analysis of nucleotide diversity was performed with a window length of 50 bp and step of 25 bp. Sites with alignment gaps were not considered in the length of the windows, and all estimations were performed with the standard parameters of the program. Output graphics were produced on the R package 2.10.1 with the ggplot2 library.

For the phylogenetic analysis we adjusted the ApMAT marker sequence boundaries to separate intergenic from coding regions because these are likely to evolve according to different evolutionary models. In this way, given that both ends of the ApMAT original sequence were composed of the 3′ end of the *Apn2* (~ 46 bp) and *MAT1-2-1* genes (~ 16 bp) (FIG. 1), these fragments were cut and merged with the original Apn25L and MAT1-2-1 sequences, respectively. The phylogenetic performance of each individual marker

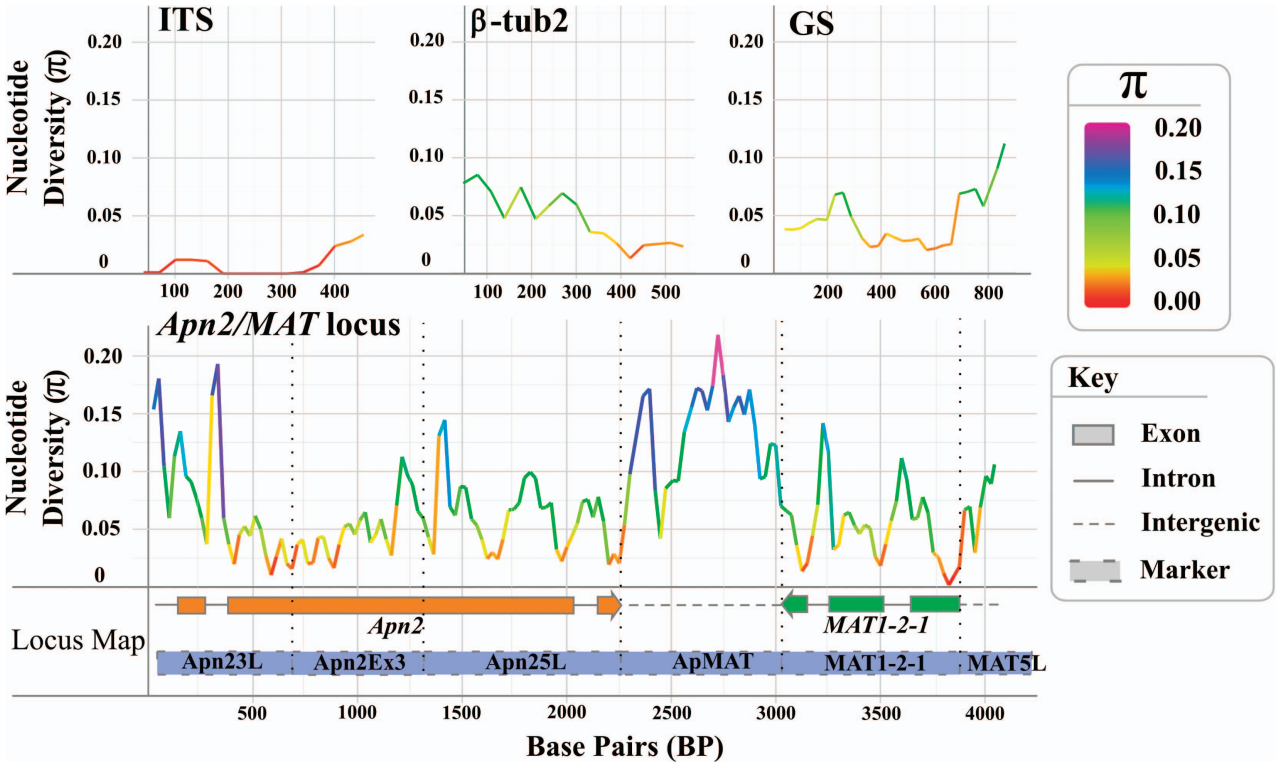

FIG. 2. Sliding window analysis of the nucleotide diversity (π) estimated from the molecular markers used, based on a 10 isolate representative sample. The window length is 50 bp and the step size is 25 bp. For the *Apn2/MAT* locus, below the sliding window plot, a schematic representation of gene and intergenic segments along their length is presented. In the *Apn2/MAT* locus plot the vertical dot lines represent the molecular markers' boundaries. Graphics were produced and adapted from DNasp 5 estimates with the R statistics package.

first was assessed with the online application PhyDesign (Lopez-Giraldez and Townsend 2011), which implements an empirical metric of phylogenetic informativeness (Townsend 2007). This provides a quantitative prediction of the utility and information that individual partitions contain to resolve phylogenetic questions. The required ultrametric tree was provided with the UPGMA method implemented in the program MEGA 5 (Tamura et al. 2011). Site rates for each partition were estimated with the HyPhy script, as recommended by the author. Phylogenetically informative values were calculated on a per-site basis to avoid biases from gene length and minimize the phylogenetic noise (Lopez-Giraldez and Townsend 2011).

To create concatenated matrices from the individual datasets the Concatenator 1.1.0 software was used (Pina-Martins and Paulo 2008). Three molecular marker combinations were assembled for subsequent comparative purposes and designated as: Apn-MAT dataset, comprising only *Apn2/MAT* locus makers, except *MAT1-2-1*; traditional dataset, combining ITS, β-tub2 and GS; and the total concatenated dataset. *MAT1-2-1* was not included in the Apn-MAT dataset to balance the sequence length of both Apn-MAT and traditional datasets, providing a more even comparison. Five *C. fragariae* isolates were used as outgroup taxa to root phylogenetic trees (Cai et al. 2009). Additional outgroup taxa could not be considered because orthologous regions to all our developed markers were unavailable in public databases, with the exception of *C. graminicola* whose sequences were too divergent for accurate and reliable alignments. Phylogenetic trees were constructed from the individual and combined analyses of the seven nuclear regions with maximum likelihood (ML) and a Bayesian framework with the Markov chain Monte Carlo (BMCMC) algorithm. Gaps were treated as missing data. The ML analysis was run in PAUP* 4.0d99 (Swofford 2003) with heuristic searches of 100 replicates with random sequence addition and a tree-bisection-reconnection (TBR) branch swapping algorithm. Nonparametric bootstrapping was conducted with 1000 pseudoreplicates with 10 random additions and TBR branch swapping. ModelTest 3.7 (Posada and Crandall 1998) was used to select the best fit model of nucleotide evolution under the Akaike information criterion (AIC) for each dataset. The BMCMC analysis was run in MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) with the optimal model selected under the AIC, as implemented in MrModelTest 2.3 (Nylander 2004), specified as prior for each partition. For the individual and combined datasets, Bayesian posterior probabilities were generated from 3×10^7 and 1×10^8 generations respectively, sampling every 1000th generation. The analysis was run three times with one cold and three incrementally heated Metropolis-coupled Monte Carlo Markov chains, starting from random trees. A total of 1×10^6 generations were discarded as burn-in. Trees were combined and summarized on a 50% majority rule consensus tree.

The total combined dataset also was analyzed with the species-tree approach implemented in BEST (Liu 2008) to incorporate the signals of each marker. BEST has been shown to deal with the common issue of deep coalescence in recently diverged species. It uses a hierarchical Bayesian approach to estimate the joint posterior distribution of multiple gene trees and is known to provide more accurate estimates than the concatenation method in some multilocus datasets (Edwards et al. 2007, Liu and Pearl 2007). The BEST analysis was run in MrBayes 2.3, and model priors for each partition were estimated as in the BMCMC analysis. Priors included an inverse gamma distribution (3, 0.003) for theta and a uniform distribution (0.2, 2) for gene mutation and were estimated as part of the analysis. A total of 2.5×10^9 generations were sampled every 2000 generations. Convergence and mixing were assessed for all parameters with Tracer 1.4 (Drummond and Rambaut 2007). The species tree was constructed from the combined runs.

Topology tests.—The current taxonomic relationship of the studied species also was assessed through topology tests on each relevant dataset. Alternative hypothesis were tested against the most recent hypothesis of relationships between our species of interest (Phoulivong et al. 2010) with the topological test of Shimodaira and Hasegawa (SH) (Shimodaira and Hasegawa 1999) as implemented in PAUP* 4.0d99 (Swofford 2003). Although this current taxonomy was not formally addressed, it provided an objective starting point for our analysis. First we tested the unconstrained phylogenies of each dataset that revealed discordant phylogenies to assess whether their branching order was significantly incongruent with the current taxonomy. In addition because *C. siamense* received only moderate support in the current taxonomy we also tested the alternative hypothesis of its monophyly to the second closest taxa, *C. asianum* (i.e. {[*C. siamense*, *C. asianum*], *C. fructicola*}, instead of the current taxonomy {[*C. siamense*, *C. fructicola*}, *C. asianum*]) for congruent datasets. One thousand replicates were performed by resampling the partial likelihoods for each site (RELL model).

RESULTS

Molecular marker analysis.—The newly developed set of primers allowed the amplification of ~ 4061 bp from the *Apn2/MAT* locus. Each primer pair produced partially overlapping fragments that in combination spanned all but the first 21 codons of *Apn2*, the intergenic region between *Apn2* and *MAT1-2-1*, the complete *MAT1-2-1* gene and a small intergenic region adjacent to the 3′ end of *MAT1-2-1* (FIG. 1). As in previous *Colletotrichum* studies, only one mating type (MAT) locus idiomorph, *MAT1-2-1*, was identified. Clean bands were reproducibly amplified by PCR for all the isolates using all primer sets, generating amplicons of ~ 740 bp for Apn23L, ~ 760 bp for Apn2Ex3, ~ 840 bp for Apn25L, ~ 910 bp for ApMAT and ~ 1040 bp for *MAT1-2-1* and MAT5L. An overview of the *Apn2/MAT* locus map is included herein (FIG. 2), along with the position and length of each developed marker, excluding overlapping segments. The ITS, β-tub2 and GS markers generated products of the expected size.

The complete set of markers was first sequenced for a representative sample of 10 isolates comprising the five studied species (*C. kahawae*, *C. gloeosporioides*, *C. asianum*, *C. fructicola*, *C. siamense*) plus the outgroup (*C. fragariae*). This preliminary dataset was subjected to a sliding window analysis to quantify nucleotide polymorphisms along the sequence of each marker and to provide a comparative exploratory analysis between markers (FIG. 2). The commonly used ITS region had the lowest overall nucleotide diversity ($\pi = 0.0077$) because it was mostly invariable. The intron rich β-tub2 ($\pi = 0.0436$) and GS ($\pi = 0.0433$) gene fragments presented a greater degree of nucleotide variation but still revealed regions of low polymorphism. The greatest nucleotide diversity was provided by the *Apn2/MAT* locus because even the most conserved region of this locus, Apn2Ex3, was more polymorphic ($\pi = 0.0467$) than β-tub2 and GS. The ApMAT intergenic region showed the highest nucleotide diversity ($\pi = 0.1282$), exceeding by more than twofold the diversity of β-tub2 or GS fragments. The *MAT1-2-1* gene ($\pi = 0.0512$) showed a considerable diversity in its introns ($\pi = 0.1135$), particularly the second one in the HMG region, in contrast with exon sequences ($\pi = 0.0419$). In the *Apn2* gene the nucleotide diversity was highest in the Apn23L marker region ($\pi = 0.0739$) due to the presence of two introns. The Apn25L region ($\pi = 0.0559$) also showed a considerable diversity, particularly along the exon sequence.

Based on this assessment, two gene fragments (Apn25L and *MAT1-2-1*) plus two intergenic regions (ApMAT and MAT5L) were selected for the multilocus phylogenetic analyses. A total of 22 sequences were obtained for each of the seven nuclear markers (ITS, β-tub2, GS, ApMAT, Apn25L, MAT1-2-1, MAT5L), producing ~ 4676 bp of sequence data for each studied isolate. According to the output of the PhyDesign web application, the index of phylogenetic power per base pair of each dataset varied widely, but all of our developed primers generated sequences with a superior resolution power throughout the phylogeny (FIG. 3). After visual inspection, the ApMAT sequence is predicted to be the most informative, followed by MAT5L, Apn25L, MAT1-2-1, GS and β-tub2, whose differences in phylogenetic informativeness were less evident. The ITS sequence had the lowest predicted resolution power. All of the newly developed markers revealed more information than the traditional markers, as determined with parsimony and nucleotide diversity statistics as comparative

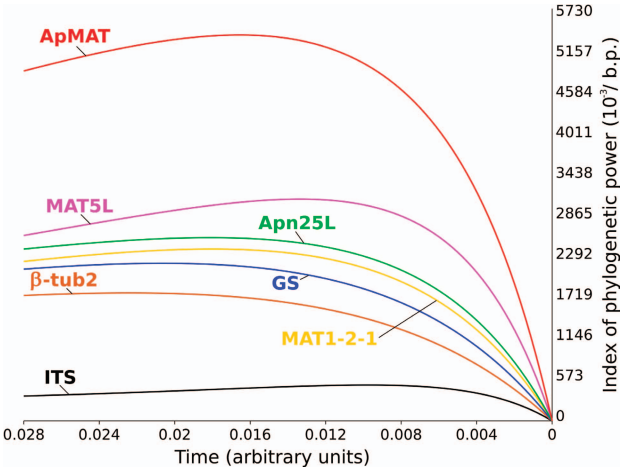

FIG. 3. Phylogenetic informativeness profiles on a per site basis of the seven molecular markers employed in this study throughout the phylogeny of the sampled isolates (in arbitrary units).

TABLE II. The length (bp), number and percentage of parsimony informative sites (PI;%PI), estimated nucleotide diversity (π) and model of sequence evolution (Model) for each dataset using the complete sampling

Data	Length	PI	%PI	π	Model
Individual datasets					
ITS	489	14	**2.86**	0.0077	TrNef + I
β-tub2	578	77	**13.32**	0.0436	TIM + G
GS	889	141	**15.86**	0.0433	K81uf + I
Apn25L	883	168	**19.03**	0.0559	TrN + G
ApMAT	772	268	**34.72**	0.1282	HKY + G
MAT1-2-1	843	144	**17.08**	0.0642	TIM + G
MAT5L	213	42	**19.72**	0.0512	TIM + I
Combined datasets					
ApnMAT	1886	478	**25.34**		
Traditional	1965	232	**11.81**		
Total	4676	801	**17.13**		

benchmarks (TABLE II). ApMAT, Apn25L, MAT5L and MAT1-2-1 were 35%, 19%, 20% and 17% parsimony informative, compared with 3%, 13% and 16% for ITS, β-tub2 and GS, respectively (TABLE II). Regarding the combined datasets, even excluding MAT1-2-1 to balance the sequence length of the two partially concatenated datasets, the Apn-MAT dataset exhibited 478 parsimony informative characters from a total of 1886 (25%), while the traditional markers contributed with only 232 parsimony informative characters from a total of 1965 (12%) (TABLE II). Altogether the total concatenated dataset presented 801 (17%) parsimony informative nucleotide characters with parsimony criteria.

Phylogenetic analysis.—We first constructed phylogenies from each of the individual markers (FIG. 4). The ML and BMCMC analyses always resulted in identical topologies for each dataset, although not all gene trees had the same resolution. For example, ITS and MAT5L provided such poorly resolved trees that no discernible topology could be recovered and thus were not analyzed individually (data not shown). In contrast, ApMAT provided the best resolving ability in which all species/group and even some intragroup nodes were distinguishable with high support in both analyses (FIG. 4a).

Most *C. gloeosporioides* s.l. isolates clustered with the ex-type strains of *C. siamense* and *C. fructicola* (TABLE I, FIG. 4). Generally species/groups were reciprocally monophyletic, although with varying degrees of support. An exception was the GS gene tree in which two isolates from the *C. siamense* group were part of a monophyletic group with *C. fructicola*. However further incongruities were found when comparing the relative taxonomic position of the ingroup species. Excluding *C. kahawae*, which was consistently revealed as the most basal lineage, all other species showed different taxonomic relationships in each of the independent loci analyzed due to significant gene-tree discordance. For example *C. siamense* was found to be part of the same monophyletic group with either *C. asianum* (FIG. 4a; bootstrap: 97, posterior probability [PP]: 1.00) or *C. fructicola* (FIG. 4d; bootstrap: 78, PP: 0.99). In the GS gene tree, although there was little support for the reciprocal monophyly of *C. siamense* isolates, the best ML and BMCMC trees supported a close relationship with *C. fructicola* rather than with *C. asianum*. The second most basal lineage identified also was subject of discordance because it was either *C. gloeosporioides* (FIG. 4a–d; bootstrap: 82–99; PP: 0.80–1.00) or *C. asianum* (FIG. 4e; bootstrap: 84; PP: 0.98). As it was clearly shown, these discordances among loci were well supported by both bootstrap and posterior probability values in each loci dataset and consequently, the phylogenetic relationships of all but one of our species were not obvious based on the topologies of gene trees alone.

Phylogenetic reconstructions were carried out for the three combined datasets to assess how these discordances affected species relationships and their node support in multilocus datasets and to compare degrees of resolution. Overall, both the traditional and Apn-MAT datasets provided almost completely resolved and highly supported trees but presented two different topologies in which *C. siamense* was again either reciprocally monophyletic with *C. asianum* (FIG. 5a; bootstrap: 95; PP: 1.00) or *C. fructicola* (FIG. 5b; bootstrap: 78; PP: 1.00). When all markers

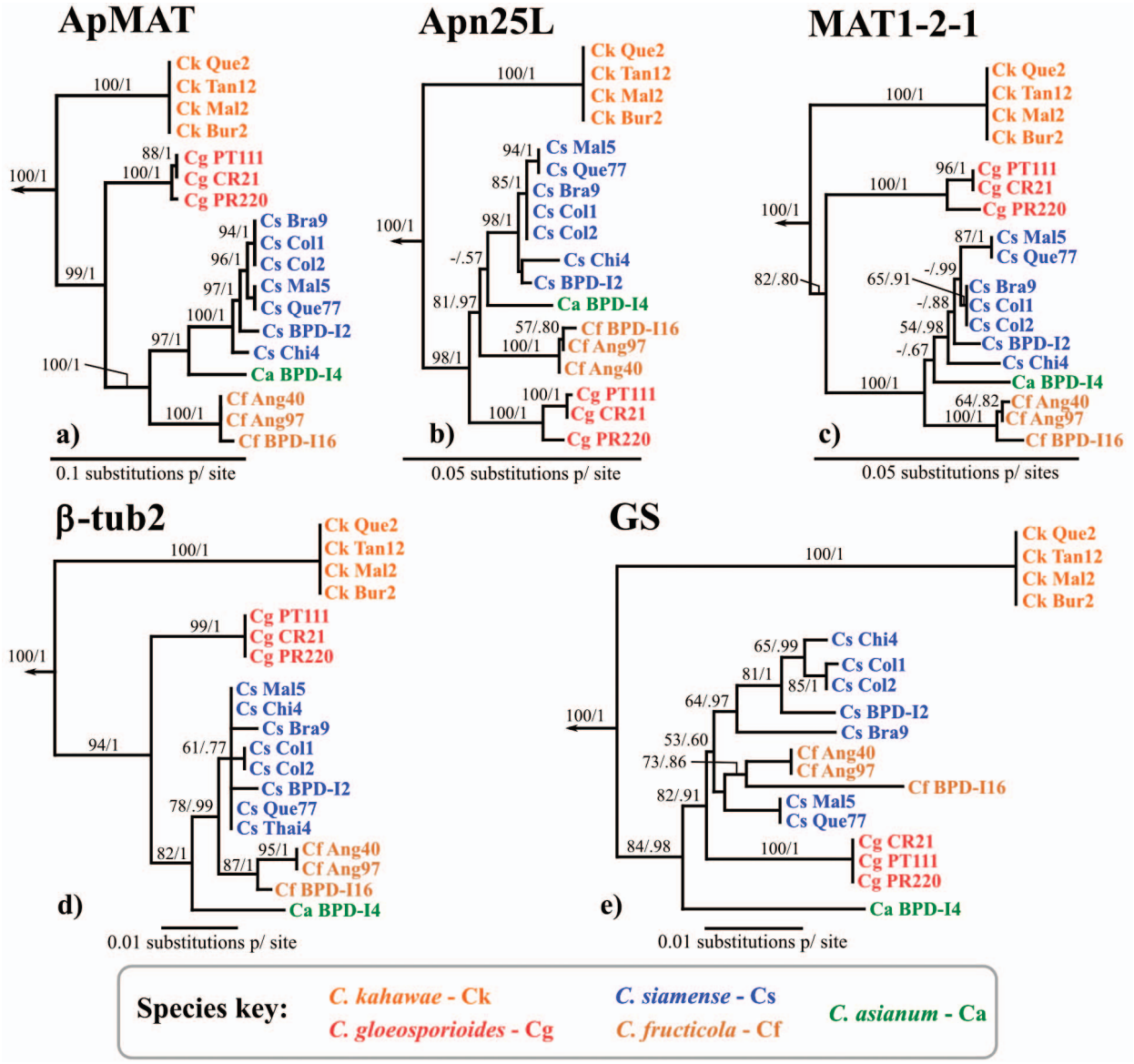

FIG. 4. Maximum likelihood gene trees (rooted with *C. fragariae*) for five individual datasets. The remaining ITS and MAT5L datasets provided a poor resolution and were not included. Bootstrap/posterior probability values are shown above branches. The key to species/groups codes and colors is provided at the bottom.

were combined, the prevailing phylogenetic relationships were congruent with the Apn-MAT dataset, supporting the reciprocal monophyly between *C. siamense* and *C. fructicola* (FIG. 5c; bootstrap: 97; PP: 1.00). Of interest, both the topology and node support of the single ApMAT marker were similar to the seven-gene concatenated dataset.

Species-tree approach and topology tests.—The current taxonomic and evolutionary relationships of the included species from the *C. gloeosporioides* complex were further assessed by two statistical methods. First, the species-tree approach of the BEST software was carried out because it has the useful property of accounting for discrepancies between gene trees and uses the coalescent theory to combine those gene trees to infer a species tree. The obtained tree shared the same topology as the Apn-MAT and total concatenated datasets, albeit with a much smaller support for the reciprocal monophyly between *C. siamense* and *C. asianum* (FIG. 6). Nonetheless the remaining relationships were highly supported.

For the topological tests we assumed the null hypothesis of the most recently reported species relationship (Phoulivong et al. 2010), which is identical to that exhibited by the traditional dataset (SH test, TABLE III). Except for β-tub2, all individual datasets were discordant, but only the branching

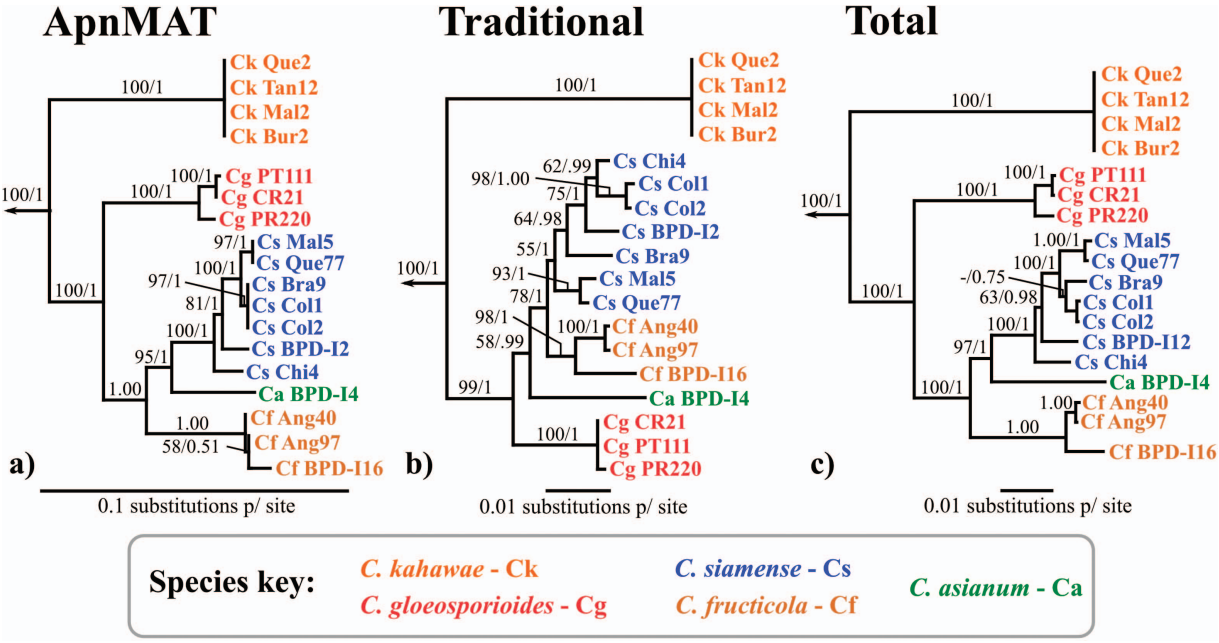

FIG. 5. Bayesian estimation of gene trees (rooted with *C. fragariae*) for the tree combined datasets: ApnMAT (ApMAT, Apn25L and MAT5L); traditional (ITS, β-tub2 and GS) and total (ApMAT, Apn25L, MAT5L, MAT1-2-1, ITS, β-tub2 and GS). Bootstrap/posterior probability values are branches. The key to species/groups codes and colors is at bottom.

order of ApMAT and MAT1-2-1 deviated significantly from the current taxonomic understanding (SH test, $P = 0.035$ and $P = 0.000$ respectively). Likewise, the combined Apn-MAT topology was significantly different (SH test, $P = 0.000$) but the deviation of the total combined dataset was only marginally significant (SH test, $P = 0.076$). We also tested whether the traditional dataset presented a significant deviation from the alternative scenario in which the true species relationship is similar to that recovered in the total dataset. Indeed, when the topology for this alternative scenario was constrained, it was not significantly different from the unconstrained phylogeny (SH test, $P = 0.143$).

TABLE III. Likelihood topology tests (SH) for each relevant dataset and hypothesis to be tested, based on the most recent taxonomic understanding of the studied species

Dataset	Hypothesis	Diff −ln L	*P*
Individual datasets			
Ap25L	Current[a]	2.73	0.259
ApMAT	Current[a]	12.93	0.035[c]
MAT1-2-1	Current[a]	80.91	0.000[c]
GS	Current[a]	5.02	0.132
Combined datasets			
ApnMAT	Current[a]	202.42	0.000[c]
Traditional	Current[a]	4.97	0.143
Total	Alternative[b]	25.71	0.076[d]

[a] (*C. asianum*, [*C. fruticola*, *C. siamense*]).
[b] (*C. frutícola*, [*C. asianum*, *C. siamense*]).
[c] Significance with $P < 0.05$.
[d] Significance with $P < 0.10$.

DISCUSSION

Resolving species complexes in *Colletotrichum*, particularly the *C. gloeosporioides* complex, is a demanding quest in which molecular systematic tools are becoming of the utmost importance for taxonomists. However, genomic regions are not equally informative and thus it is important to select the most useful loci to address this issue. The MAT locus has been employed with great success in phylogenetic studies of *Colletotrichum* (Du et al. 2005; Crouch et al. 2006, 2009a; Rojas et al. 2010) as well as in other genera, such as *Cochliobolus* (Turgeon 1998), *Fusarium* (O'Donnell et al. 2004) and *Leptosphaeria* (Voigt et al. 2005). In our work a newly developed set of primers enabled the analysis of an expanded region of the *Apn2/MAT* locus in relation to other studies exploring this locus (Du et al. 2005, Crouch et al. 2009a, Rojas et al. 2010). In fact, our primer sets were able to generate not only orthologous sequences to those already published (e.g. the 3′ end of the *Apn2* gene and intergenic region between *MAT1-2-1* [Crouch et al. 2009a, Rojas et al. 2010] or the intron in the HMG region of the *MAT1-2-1* gene [Du et al. 2005]) but also the remainder of the *Apn2* gene sequence as well as the complete *MAT1-2-1* gene

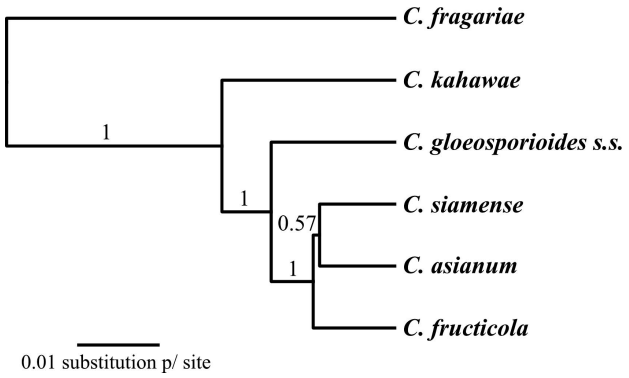

FIG. 6. Species tree of *C. gloeosporioides* complex inferred with the BEST software for the total combined dataset and rooted with *C. fragariae*. Values above branches indicate posterior probability support.

along with a portion of the 5′ intergenic region, covering ~ 2700 bp of novel sequence data from this locus. Overall, our analyses further demonstrated the potential of this genomic region to provide an outstanding improvement in phylogenetic informativeness to face the taxonomic challenges of the *C. gloeosporioides* complex.

Even though our analyses focused mostly on *Coffea* spp.-inhabiting species of the *C. gloeosporioides* complex, this sampling strategy proved valuable under the current taxonomic knowledge. First, the six species used in our study encompass relatively divergent lineages within the complex (4.5–10.1% sequence divergence). Therefore, despite probably not covering all of the putatively high species richness of the complex, they provide a wide range of sequence diversity, which is useful to assess the robustness of the developed primers and also enhances their potential applicability to other species in the *C. gloeosporioides* complex. Second, unlike the myriad undescribed and poorly known lineages in this complex, the species used here are well characterized and recognized by multilocus phylogenetic and morphological studies, which is paramount when investigating the value of the newly developed molecular tools for systematic purposes. Third, all these species were already used simultaneously in previous and independent phylogenetic studies, providing an objective starting point of their taxonomic relationships and allowing a much more meaningful comparison of the traditional gene-tree concatenation and the recent species-tree reconstruction methods.

Regarding their informative potential, all newly developed primer pairs generated sequences that were superior to ITS and to the two commonly used gene fragments, β-tub2 and GS, even though the latter had been useful and regarded as fairly informative (Talhinhas et al. 2005, Prihastuti et al. 2009, Rojas et al. 2010). The ITS sequence in particular, which is considered one of the most promising regions for fungal barcoding (Seifert 2009), performed poorly in our benchmark comparisons, being unable to distinguish taxonomic groups and recover their relationships. This also has been observed in other studies (Cai et al. 2009, Crouch et al. 2009b, Yang et al. 2009). As evidenced by the sliding window analysis of polymorphism and parsimony statistics, the intergenic region between the *Apn2* and *MAT1-2-1* genes was revealed to be the most variable segment and thus a promising marker for both taxonomic and population genetic studies in the complex. Intergenic regions are expected to be much more variable because they generally are not under functional constraints or any direct selective pressures (O'Donnell et al. 2004, Thomson et al. 2010), but for the same reason they are unsuitable regions for primer design over a broad range of species. However, the ApMAT intergenic segment is flanked by fairly conserved regions for which primers can be designed, making it an ideal marker for the *C. gloeosporioides* complex. To a lesser extent, both the introns from *MAT1-2-1* gene and the two *Apn2* gene extremities were variable and should also prove useful for systematic purposes, as reported in other *Colletotrichum* studies (Du et al. 2005, Crouch et al. 2006, 2009a; Rojas et al. 2010). Nevertheless, these markers are not barcode candidates, and attempts to use them outside the range of the *C. gloeosporioides* complex probably will fail due to their highly variable nature even in priming sites. For example, the markers developed by Crouch et al. (2009a) to amplify partial fragments of the *Apn2* gene and intergenic region in the same locus for the graminicolous *Colletotrichum* species could not be used in the isolates analyzed in this study (data not shown). In fact, after obtaining the homologous sequences from isolates of the *C. gloeosporioides* complex alignments revealed a high divergence between the two groups of species. However, these markers are much more suitable than barcode genes to accurately dissect and unravel the diversity and evolutionary relationships in a complex of closely related species. Moreover, as Cai et al. (2009) stated, the selection and decision of the most suitable barcode gene(s) can be made only in the process of species delimitation because only then we know which taxa need to be distinguished. In other words, we first need to fully resolve the *C. gloeosporioides* complex with a set of highly informative genes to adequately search and select a barcode gene.

In agreement with other studies in *Colletotrichum*, our study revealed the ubiquitous presence of only one MAT gene idiomorph, *MAT1-2-1* (Vaillancourt et al. 2000, Chen et al. 2002, Rodríguez-Guerra et al.

2005, García-Serrano et al. 2008). In the fungal kingdom MAT idiomorphs are known to code for transcription factors that regulate sexual development and recognition (Kronstad and Staben 1997). However their configuration in *Colletotrichum* species seems to deviate from the great majority of the ascomycetes, where a single MAT locus (MAT1) exists as two alternate idiomorphs (MAT1-1 and MAT1-2) (Turgeon 1998). Thus the exclusive presence of a MAT1-2 homolog in all species included in this work supports the idea that the genetic control underlying sexual reproduction in this genus does not conform to any known system in the entire fungal kingdom (Rodríguez-Guerra et al. 2005, Crouch and Beirn 2009).

In our phylogenetic analyses of the individual gene trees the ApMAT marker was singular in its ability to resolve species and even intraspecific relationships. Alone it provided essentially the same information and support as the concatenated tree of seven genes, a feat that was not achieved by any other single marker. The other markers had variable degrees of resolution and support, ranging from low to moderate, as is normally expected for nuclear gene markers when addressing recently diverged taxa (Thomson et al. 2010).

However, gene trees are only single realizations of the true species tree and they may not reflect the actual pattern of lineage splitting and divergence due to coalescent stochasticity (Degnan and Rosenberg 2009, Carstens and Dewey 2010). Therefore, it is useful to reconstruct individual gene trees to provide an overview of how each marker depicts species relationships. In this way, several incongruities were revealed by comparing the topology of each gene tree. For example, two isolates that were consistently shown to cluster with the *C. siamense* group (Mal5 and CCM5) were monophyletic with isolates from the *C. fructicola* group in the GS gene tree. Moreover, each analyzed locus provided with strong support a discordant species branching order with each other and the correct relationship of the studied *Colletotrichum* species was not clear. In fact only one gene tree (β-tub2) was congruent with the current taxonomic understanding of the sampled species whereas the ApMAT and *MAT1-2-1* gene trees revealed a significantly deviated phylogenetic signal. Even when different combinations of genes are concatenated these discordances remain significant, as evidenced by the ambiguous relationship of *C. siamense*, *C. fructicola* and *C. asianum* between the traditional and Apn-MAT datasets. There are several causes for discordance between gene trees in phylogenetic studies, including horizontal gene transfer (Andersson 2005), hybridization (Meng and Kubatko 2009) or incomplete lineage sorting (Degnan and Rosenberg 2009). To our knowledge this is the first report of gene-tree discordance among species of the *C. gloeosporioides* complex.

In this study, even though the total concatenated dataset provided a single resolved topology congruent with that of the combined Apn-MAT and individual ApMAT datasets, the concatenation method assumption of no heterogeneity between the embedded gene trees was violated. In addition, the longer and more informative markers, such as those from the *Apn2/MAT* locus, may bias tree topology toward themselves simply because their phylogenetic signal overwhelms the information contained in less informative markers (Knowles and Carstens 2007). Because the evolutionary history of species is widely considered to be mostly a stochastic process (Kingman 1982) we have applied a recent statistical framework that accounts for the discrepancies of gene trees from multiple unlinked loci and is specifically conceived to estimate the species tree. Using a Bayesian hierarchical model, the BEST program is able not only to accommodate gene trees heterogeneity but also to retrieve useful information from it and provide a more accurate estimate of the true species tree (Edwards et al. 2007). The resultant species-tree topology supported that of the total concatenated dataset, except for the *C. siamense* and *C. asianum* relationships. Although all other nodes were highly supported, there was little support for the reciprocal monophyly between *C. siamense* and *C. asianum* and the relationship of these two species plus *C. fructicola* was unresolved. This result was obtained in spite of using the large and highly informative sequence data of our analysis. Nonetheless, it is expected that the problems and difficulties derived from incomplete lineage sorting increase as the time since species divergence decreases (Degnan and Rosenberg 2009). Thus it seems likely that these species diverged recently and consequently their relationships may still not be easily resolved. Nonetheless, to further understand the processes that promote speciation and the evolution of phenotypic traits it will be essential to recognize and delimit nascent evolutionary lineages (Wiens 2007). In addition, it will be interesting to assess the effects of using different outgroup taxa in the phylogenetic inference as the current taxonomic knowledge of *Colletotrichum* increases and molecular tools improve.

Altogether this study demonstrates the considerable improvement that the *Apn2/MAT* locus may provide to the molecular systematics of the *C. gloeosporioides* complex. We report the development and application of a novel set of primers as valuable, promising tools to generate sequence data able to dissect species and even population relationships. Indeed, based on our data, an alternative view of the

evolutionary relationships of the studied species is hypothesized. Because we also showed that gene tree discordance may be a problem in recently diverged species of this complex we stress the fact that highly variable markers, such as the ApMAT, are much more likely to surpass those problems and successfully discern how species are phylogenetically related. It is hoped that this contribution may assist and stimulate further research on this challenging complex and eventually lead to a more comprehensive understanding of its structure, evolution and diversity.

ACKNOWLEDGMENTS

We thank Ana Vieira, Tiago Jesus and two anonymous referees for their contributions and comments during the elaboration of this manuscript. We also thank Francisco Pina-Martins for helping in the first steps of the data analysis. At CIFC/IICT we also appreciate the technical support provided by Sandra Sousa Emídio. Lei Cai acknowledges grants CAS-KSCX2-EW-J-6/NSFC31070020. This work was financially supported by the Foundation for Science and Technology (FCT) and IICT, Portugal.

LITERATURE CITED

Abang MM, Winter S, Green KR, Hoffmann P, Mignouna HD, Wolf GA. 2002. Molecular identification of *Colletotrichum gloeosporioides* causing yam anthracnose in Nigeria. Plant Pathol 51:63–71, doi:10.1046/j.0032-0862.2001.00655.x

Afanador-Kafuri L, Minz D, Maymon M, Freeman S. 2002. Characterization of *Colletotrichum* isolates from tamarillo, passiflora and mango in Colombia and identification of a unique species from the genus. Phytopathology 93:579–587, doi:10.1094/PHYTO.2003.93.5.579

Andersson JO. 2005. Lateral gene transfer in eukaryotes. Cel Mol Life Sci 62:1182–1197, doi:10.1007/s00018-005-4539-z

Brown A, Sreenivasaprasad S, Timmer L. 1996. Molecular characterization of slow-growing orange and key lime anthracnose strains of Colletotrichum from citrus as *C. acutatum*. Phytopathology 86:523–527.

Cai L, Hyde KD, Taylor PWJ, Weir BS, Waller J, Abang MM, Zhang JZ, Yang YL, Phoulivong S, Liu ZY, Prihastuti H, Shivas RG, McKenzie EHC, Johnston PR. 2009. A polyphasic approach for studying *Colletotrichum*. Fungal Divers 39:183–204.

Cannon PF, Bridge PD, Monte E. 2000. Linking the past, present and future of *Colletotrichum* systematic. In: Prusky D, Freeman S, Dickman M, eds. *Colletotrichum:* host specificity, pathology and host-pathogen interaction. St Paul, Minnesota: APS Press. p 1–20.

———, Buddie AG, Bridge PD. 2008. The typification of *Colletotrichum gloeosporioides*. Mycotaxon 104: 189–204.

Carstens BC, Dewey TA. 2010. Species delimitation using a combined coalescent and information-theoretic approach: an example from North American *Myotis* Bats. Syst Biol 59:400–414, doi:10.1093/sysbio/syq024

Chakraborty S, Fernandes C, Charchar MJD, Thomas M. 2002. Pathogenic variation in *Colletotrichum gloeosporioides* infecting *Stylosanthes* spp. in a center of diversity in Brazil. Phytopathology 92:553–562, doi:10.1094/PHYTO.2002.92.5.553

Chen F, Goodwin P, Khan A, Hsiang T. 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris*. Can J Microbiol 48:427–436, doi:10.1139/w02-034

Crouch J, Beirn LA. 2009. Anthracnose of cereals and grasses. Fungal Divers 39:19–44.

———, Clarke B, Hillman B. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turf grass and corn. Phytopathology 96:46–60, doi:10.1094/PHYTO-96-0046

———, ———, ———. 2009b. What is the value of ITS sequence data in *Colletotrichum* systematics and species diagnosis? A case study using the falcate-spored graminicolous *Colletotrichum* group. Mycologia 101: 648–656, doi:10.3852/08-231

———, Tredway L, Clarke B, Hillman B. 2009a. Phylogenetic and population genetic divergence correspond with habitat for the pathogen *Colletotrichum cereale* and allied taxa across diverse grass communities. Mol Ecol 18:123–135, doi:10.1111/j.1365-294X.2008.04008.x

Damm U, Woudenberg JHC, Cannon PF, Crous PW. 2009. *Colletotrichum* species with curved conidia from herbaceous hosts. Fungal Divers 39:45–87.

Degnan JH, Rosenberg NA. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol 24:332–340, doi:10.1016/j.tree.2009.01.009

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214, doi:10.1186/1471-2148-7-214

Du M, Schardl C, Nuckles E, Vaillancourt L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Collectotrichum* species complexes. Mycologia 97:641–658, doi:10.3852/mycologia.97.3.641

Edwards SV, Liu L, Pearl DK. 2007. High-resolution species trees without concatenation. Proc Natl Acad Sci USA 104:5936–5941, doi:10.1073/pnas.0607004104

Freeman S, Katan T, Shabi E. 1998. Characterization of *Colletotrichum* species responsible for anthracnose diseases of various fruits. Plant Dis 82:596–605, doi:10.1094/PDIS.1998.82.6.596

———, Minz D, Jurkevitch E, Maymon M, Shabi E. 2000. Molecular analyses of *Colletotrichum* species from almond and other fruits. Phytopathology 90:608–614, doi:10.1094/PHYTO.2000.90.6.608

García-Serrano M, Laguna EA, Rodríguez-Guerra R, Simpson J. 2008. Analysis of the MAT1-2-1 gene of *Colletotrichum lindemuthianum*. Mycoscience 49:312–317, doi:10.1007/s10267-008-0424-6

Gunnell PS, Gubler WD. 1992. Taxonomy and morphology of *Colletotrichum* species pathogenic to strawberry. Mycologia 84:157–165, doi:10.2307/3760246

Hindorf H. 1970. *Colletotrichum* spp. isolated from *Coffea arabica* L. in Kenya. Z Pflanzenkrankh Pflanzenschutz 77:328–331.

Hyde KD, Cai L, Cannon PF, Crouch JA, Crous PW, Damm U, Goodwin PH, Chen H, Johnston PR, Jones EBG, Liu ZY, McKenzie EHC, Moriwaki J, Noireung P, Pennycook SR, Pfenning LH, Prihastuti H, Sato T, Shivas RG, Tan YP, Taylor PWJ, Weir BS, Yang YL, Zhang JZ. 2009b. *Colletotrichum*—names in current use. Fungal Divers 39:147–182.

———, ———, McKenzie EHC, Yang Y, Zhang J, Prihastuti H. 2009a. *Colletotrichum*: a catalogue of confusion. Fungal Divers 39:1–17.

Johnston P, Jones D. 1997. Relationships among *Colletotrichum* isolates from fruit-rots assessed using rDNA sequences. Mycologia 89:420–430, doi:10.2307/3761036

Katoh K, Asimenos G, Toh H. 2009. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol 537: 39–64, doi:10.1007/978-1-59745-251-9_3

Kingman JFC. 1982. The coalescent. Stoch Process Appl 13: 235–248, doi:10.1016/0304-4149(82)90011-4

Knowles LL, Carstens BC. 2007. Delimiting species without monophyletic gene trees. Syst Biol 56:887–895, doi:10.1080/10635150701701091

Kronstad JW, Staben C. 1997. Mating type in filamentous fungi. Annu Rev Genet 31:245–276, doi:10.1146/annurev.genet.31.1.245

Liu L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. Bioinformatics 24:2542–2543, doi:10.1093/bioinformatics/btn484

———, Pearl DK. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst Biol 56:504–514, doi:10.1080/10635150701429982

Lopez-Giraldez F, Townsend JP. 2011. PhyDesign: an online application for profiling phylogenetic informativeness. BMC Evol Biol 11:152, doi:10.1186/1471-2148-11-152

Lubbe CM, Denman S, Cannon PF, Groenewald JZ, Lamprecht SC, Crous PW. 2004. Characterization of *Colletotrichum gloeosporioides* and similar species associated with anthracnose and dieback of Proteaceae. Mycologia 96:1268–1279, doi:10.2307/3762144

Marshall OJ. 2004. PerlPrimer: cross platform, graphical primer design for standard, bisulphite and real-time PCR. Bioinformatics 20:2471–2472, doi:10.1093/bioinformatics/bth254

Meng C, Kubatko LS. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor Pop Biol 75:35–45, doi:10.1016/j.tpb.2008.10.004

Mills PR, Sreenivasaprasad S, Brown AE. 1992. Detection and differentiation of *Colletotrichum gloeosporioides* isolates using PCR. FEMS Microbiol Lett 98:137–144, doi:10.1111/j.1574-6968.1992.tb05503.x

Munaut F, Hamaide N, Stappen JV, Maraite H. 1998. Genetic relationships among isolates of *Colletotrichum gloeosporioides* from *Stylosanthes* spp. in Africa and Australia using RAPD and ribosomal DNA markers. Plant Pathol 47: 641–648, doi:10.1046/j.1365-3059.1998.00287.x

Nguyen PTH, Pettersson OV, Olsson P, Liljeroth E. 2009. Identification of Colletotrichum species associated with anthracnose disease of coffee in Vietnam. Eur J Plant Pathol 127:73–87, doi:10.1046/j.1365-3059.1998.00287.x

Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Uppsala, Sweden: Evolutionary Biology Centre, Uppsala Univ.

O'Donnell K, Cigelnik E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol Phylogenet Evol 7:103–116, doi:10.1006/mpev.1996.0376

———, Ward T, Geiser D, Kistler HC, Aoki T. 2004. Genealogical concordance between the mating type locus and seven other nuclear genes supports formal recognition of nine phylogenetically distinct species within the *Fusarium graminearum* clade. Fungal Genet Biol 41:600–623, doi:10.1016/j.fgb.2004.03.003

Phoulivong S, Cai L, Chen H, McKenzie EHC, Abdelsalam K, Chukeatirote E, Hyde KD. 2010. *Colletotrichum gloeosporioides* is not a common pathogen on tropical fruits. Fungal Divers 44:33–43, doi:10.1007/s13225-010-0046-0

Pina-Martins F, Paulo OS. 2008. Concatenator: sequence data matrices handling made easy. Mol Ecol Resour 8: 1254–1255, doi:10.1111/j.1755-0998.2008.02164.x

Polashock JJ, Caruso FL, Oudemans PV, McManus PS, Crouch JA. 2009. The North American cranberry fruit rot fungal community: a systematic overview using morphological and phylogenetic affinities. Plant Pathol 58:1116–1127.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818, doi:10.1093/bioinformatics/14.9.817

Prihastuti H, Cai L, Chen H, McKenzie EHC, Hyde KD. 2009. Characterization of *Colletotrichum* species associated with coffee berries in northern Thailand. Fungal Divers 39:89–109.

Ramos AP, Merali Z, Talhinhas P, Sreenivasaprasad S, Oliveira H. 2006. Molecular and morphological characterisation of *Colletotrichum* species involved in citrus anthracnose in Portugal. Bull OILB/SROP 29: 317–326.

Rodríguez-Guerra R, Ramírez-Rueda MT, Cabral-Enciso M, García-Serrano M, Lira-Maldonado Z, Guevara-González RG, González-Chavira M, Simpson J. 2005. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. Mycologia 97: 793–803, doi:10.3852/mycologia.97.4.793

Rojas EI, Rehner SA, Samuels GJ, van Bael SA, Herre EA, Cannon P, Chen R, Pang J, Ruiwu W, Yaping Z, Peng YQ, Sha T. 2010. *Colletotrichum gloeosporioides* s.l. associated with *Theobroma cacao* and other plants in Panamá: multilocus phylogenies distinguish host-associated pathogens from asymptomatic endophytes. Mycologia 102:1318–1338, doi:10.3852/09-244

Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574, doi:10.1093/bioinformatics/btg180

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497, doi:10.1093/bioinformatics/btg359

Seifert K. 2009. Progress toward DNA barcoding of fungi. Mol Ecol Resour 9:83–89, doi:10.1111/j.1755-0998.2009.02635.x

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log likelihoods with applications to phylogenetic inference. Mol Biol Evol 16:1114–1116.

Sreenivasaprasad S, Brown AR, Mills P. 1993. Coffee berry disease pathogen in Africa: genetic structure and relationship to the group species *Colletotrichum gloeosporioides*. Mycol Res 87:995–1000, doi:10.1016/S0953-7562(09)80868-X

———, Mills PR, Brown AR. 1994. Nucleotide sequence of the rDNA spacer 1 enables identification of isolates of *Colletotrichum* as *C. acutatum*. Mycol Res 98:186–188, doi:10.1016/S0953-7562(09)80184-6

———, ———, Meehan BM, Brown A. 1996. Phylogeny and systematics of 18 *Colletotrichum* species based on ribosomal DNA spacer sequences. Genome 39:499–512, doi:10.1139/g96-064

Sutton BC. 1992. The genus *Glomerella* and its anamorph *Colletotrichum*. In: Bailey JA, Jeger MJ, eds. *Colletotrichum:* biology, pathology and control. Wallingford, UK: CAB International. p 1–26.

Suzuki T, Tanaka-Miwa C, Ebihara Y, Ito Y, Uematsu S. 2010. Genetic polymorphism and virulence of *Colletotrichum gloeosporioides* isolated from strawberry (*Fragaria* × *ananassa* Duchesne). J Gen Plant Pathol 76:247–253, doi:10.1007/s10327-010-0242-5

Swofford DL. 2003. PAUP* 4.0: phylogenetic analysis using parsimony (*and other methods). Sunderland, Massachusetts: Sinauer Associates.

Talhinhas P, Neves-Martins J, Oliveira H, Sreenivasaprasad S. 2005. Molecular and phenotypic analyses reveal association of diverse *Colletotrichum acutatum* groups and a low level of *C. gloeosporioides* with olive anthracnose. Appl Environ Microbiol 71:2987–2998, doi:10.1128/AEM.71.6.2987-2998.2005

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using likelihood, distance and parsimony methods. http://www.megasoftware.net/

Taylor J, Jacobson D, Kroken S, Kasuga T, Geiser D, Hibbett D, Fisher M. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet Biol 31: 21–32, doi:10.1006/fgbi.2000.1228

Thomson RC, Wang IJ, Johnson JR. 2010. Genome-enabled development of DNA markers for ecology, evolution and conservation. Mol Ecol 19:2184–2195, doi:10.1111/j.1365-294X.2010.04650.x

Townsend JP. 2007. Profiling phylogenetic informativeness. Syst Biol 2:222–231, doi:10.1080/10635150701311362

Turgeon B. 1998. Application of mating type gene technology to problems in fungal biology. Annu Rev Phytopathol 36:115–137, doi:10.1146/annurev.phyto.36.1.115

Vaillancourt LJ, Wang J, Hanau RM. 2000. Genetic regulation of sexual compatibility. In: Prusky D, Freeman S, Dickman MB, eds. *Colletotrichum* host specificity, pathology and host pathogen interaction. St Paul, Minnesota: APS Press. p 24–44.

Voigt K, Cozijnsen A, Kroymann J, Pöggeler S, Howlett B. 2005. Phylogenetic relationships between members of the crucifer pathogenic *Leptosphaeria maculans* species complex as shown by mating type (MAT1-2), actin and beta-tubulin sequences. Mol Phylogenet Evol 37:541–557, doi:10.1016/j.ympev.2005.07.006

von Arx JA. 1957. Die Arten der Gattung *Colletotrichum*. Phytopathol Z 29:413–468.

Wiens JJ. 2007. Species delimitation: new approaches for discovering diversity. Syst Biol 56:875–878, doi:10.1080/10635150701748506

Xiao C, MacKenzie S, Legard D. 2004. Genetic and pathogenic analyses of *Colletotrichum gloeosporioides* isolates from strawberry and noncultivated hosts. Phytopathology 94: 446–453, doi:10.1094/PHYTO.2004.94.5.446

Yang YL, Liu ZY, Cai L, Hyde KD, Yu ZN, McKenzie EHC. 2009. *Colletotrichum* anthracnose of Amaryllidaceae. Fungal Divers 39:123–149.