



Investigating species boundaries in *Colletotrichum*

Chitrabhanu S. Bhunjun^{1,2} · Chayanard Phukhamsakda^{3,4} · Ruvishika S. Jayawardena^{1,2} · Rajesh Jeewon⁵ · Itthayakorn Promputtha⁶ · Kevin D. Hyde^{1,6}

Received: 16 October 2020 / Accepted: 3 February 2021 / Published online: 13 March 2021
© MUSHROOM RESEARCH FOUNDATION 2021

Abstract

Colletotrichum is one of the most important plant pathogenic genera that is responsible for numerous diseases which can have a profound impact on the agricultural sector. Species delineation is difficult due to a lack of distinctive phenotypic variation. Therefore, in this study three different genomic approaches based on phylogenetic, evolutionary and coalescent-based methods are applied to establish robust species boundaries. The reliability of five different DNA barcodes was also assessed to provide further insights into species delineation. The ITS region can resolve the placement of taxa up to the species complex level. The *GAPDH* and *TUB2* markers are determined to be the most informative for most complexes. However, no single marker could discriminate between species in all complexes, therefore different molecular approaches based on multi-locus datasets are recommended. This is the first study to provide an estimated divergence time for all species complexes in *Colletotrichum*. The estimated divergence time for species complexes ranged between 4.8 to 32.2 MYA. Based on the high level of congruent results obtained from the different molecular approaches, a new species complex, the *Colletotrichum agaves* complex is introduced. This complex consists of five taxa which are characterised by the presence of straight or slightly curved conidia with obtuse apices. This study shows that coalescent approaches and multi-locus phylogeny are crucial to establish species boundaries in *Colletotrichum*. The taxonomic placement of three singleton taxa *Colletotrichum axonopodi*, *C. cariniferi* and *C. parallelophorum* is revised. We accept 248 species and provide recommendations regarding species boundaries in the graminicola–caudatum complex.

Keywords BEAST · General mixed yule coalescent method · *Glomerellaceae* · Multi-rate poisson tree process · *Sordariomycetes* · Taxonomy

✉ Kevin D. Hyde
kdhyde3@gmail.com

¹ Center of Excellence in Fungal Research, Mae Fah Luang University, Chiang Rai 57100, Thailand

² School of Science, Mae Fah Luang University, Chiang Rai 57100, Thailand

³ Institute of Plant Protection, College of Agriculture, Jilin Agricultural University, Changchun 130118, Jilin, People's Republic of China

⁴ Engineering Research Center of Chinese Ministry of Education for Edible and Medicinal Fungi, Jilin Agricultural University, Changchun 130118, Jilin, People's Republic of China

⁵ Department of Health Sciences, Faculty of Medicine and Health Sciences, University of Mauritius, Reduit, Mauritius

⁶ Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50200, Thailand

Introduction

Colletotrichum was introduced by Corda (1831) with *C. lineola* as the type species and it is the only member of *Glomerellaceae* (*Glomerellales*, *Sordariomycetes*) (Maharachchikumbura et al. 2016; Hyde et al. 2019, 2020b). *Colletotrichum* species are endophytes, pathogens and saprobes (Tao et al. 2013; Hyde et al. 2014; Jayawardena et al. 2016a, b; Rashmi et al. 2019). As endophytes, they are one of the most widely distributed genera and can produce a range of secondary metabolites (Moraga et al. 2019). As plant pathogens, species can cause anthracnose and postharvest fruit rots (Phoulivong et al. 2010), leading to yield loss especially in high-value crops worldwide. Some species are associated with a single host while others can infect multiple hosts, and this hinders effective disease management (da Silva et al. 2020). *Colletotrichum* species have also been known to cause human infections for example, *C. dematium*,

C. gloeosporioides and *C. graminicola* can cause Keratitis (Cano et al. 2004; Shivaprakash et al. 2011).

Accurate species delineation is vital especially among plant pathogenic genera and in studies focusing on biodiversity, conservation and evolution as well as for the establishment of quarantine measures. Morphology based delimitation of species is problematic due to overlapping morphological characters in asexual morphs and the lack of sexual morphs in this genus (Hyde et al. 2009; Cannon et al. 2012). The reliability of morphological characters is also biased due to variation in culture conditions such as the choice of synthetic media among different studies (Cai et al. 2009; Rojas et al. 2010; Weir et al. 2012; Doyle et al. 2013). The application of a standard medium such as potato dextrose agar (PDA) for morphological studies of *Colletotrichum* would minimize the phenotypic variability within isolates, thus potentially maximizing differences between species (Cannon et al. 2000). Several species have been wrongly named based only on their host's specificity (Cannon et al. 2012). However, our understanding of host specificity in *Colletotrichum* is limited as most studies have mainly focused on economically important crops or ornamental crops (Cannon et al. 2012). Accurate DNA-based identification of species is also hindered due to the lack of ex-type or authenticated sequences (Cai et al. 2009).

There is also no agreement on the barcodes that should be used for species identification as different markers have been used in different studies. Up to 13 different markers have been used for species delineation in *Colletotrichum* which is impractical in terms of time and cost (Damm et al. 2012a; Weir et al. 2012; Hyde et al. 2013; Crouch et al. 2014; Liu et al. 2016). Many of these markers provide little resolution for species delineation as the phylogenetic significance of a marker varies among complexes (Hyde et al. 2013). Over 900 epithets are listed in Index Fungorum (2020) under *Colletotrichum*, which comprises 247 accepted species (Jayawardena et al. 2020). The majority of the species are grouped into 14 species complexes namely acutatum, boninense, caudatum, dematium, destructivum, dracaenophilum, gigasporum, gloeosporioides, graminicola, magnum, orbiculare, orchidearum, spaethianum and truncatum (Jayawardena et al. 2016a; Marin-Felix et al. 2017; Damm et al. 2019). Cai et al. (2009) recommended a polyphasic approach including multi-loci sequence analyses, analyses of ecological, geographical and morphological data for better taxonomic resolution for this genus, however such polyphasic approaches can be difficult to apply. The identification of cryptic species based on the phylogenetic species concept can lead to an artificial increase in the number of taxa in a group. Therefore, advances in coalescent methods could represent a new approach for species delineation for cryptic species as they can provide insights in divergence and evolutionary relationships among species. Coalescent

methods can be based on a maximum likelihood or Bayesian function whereby species delimitation models are evaluated differently (Fujita et al. 2012). In the maximum likelihood approach, point estimates are used for the genealogies at each locus and for the population. The Bayesian approach incorporates genealogical uncertainty by estimating gene trees directly from sequence data for each locus.

Coalescent based methods have been commonly used for species delineation in the plant and animal kingdom, but rarely in fungal research. Coalescent based methods such as the multi-rate Poisson tree processes (mPTP) and the general mixed Yule-coalescent method (GMYC) were designed for single-locus analysis, but have successfully been applied to concatenated multi-locus datasets (Arrigoni et al. 2016; Nieto-Montes de Oca et al. 2017; Renner et al. 2017). These approaches can incorporate the process of lineage sorting and the presence of incongruent genomic regions into phylogenetic estimation procedures (Carstens and Knowles 2007). Coalescent based method computes the likelihood of a species tree from the probabilities of the individual gene trees which can be used to investigate causes of gene trees discordance (Liu et al. 2016). The species boundaries detected by coalescent based methods are considered as initial hypotheses and should therefore be validated using findings from for example the phylogenetic species concept. The GMYC method is based on the assumption that independent evolution leads to the appearance of distinct genetic clusters which are separated by longer internal branches (Barraclough et al. 2003). It delimits these clusters by finding the maximum likelihood solution for a model that combines diversification between species (based on a Yule model) and genealogical branching within species. The advantage of the GMYC framework is that it allows for statistical inference and hypothesis testing across the entire sampled clade (Fujisawa and Barraclough 2013). The advantage of the mPTP method is that it incorporates different levels of intraspecific genetic diversity derived from differences in the evolutionary history or sampling of each species (Kapli et al. 2017). We therefore aim to elucidate species and species complex boundaries within *Colletotrichum* by using phylogenetic analyses, evolutionary analyses coupled with coalescent-based methods including mPTP and GMYC as well as determine the phylogenetic significance of five different DNA barcodes for species delineation.

Materials and methods

Taxon sampling

The type species included in Jayawardena et al. (2020) are used as the starting point for this study. The dataset includes the genes actin (*ACT*), β -tubulin2 (*TUB2*), chitin

synthase (*CHS-1*), glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*), and the internal transcribed spacers (ITS) region of all the species that have been published until June 2020 (Table S1). These DNA regions were selected based on their previously reported usefulness and wider availability of genomic data in GenBank.

Phylogenetic analyses

Phylogenetic analyses were performed using seven datasets: (1) single genes (*ACT*, *CHS-1*, *GAPDH*, ITS, *TUB2*), (2) multi-locus dataset, (3) a larger dataset. The multi-locus dataset included only the type species and the larger dataset included up to two strains for each of the type species. The selected strains were from different geographic locations and different hosts. The sequences were aligned by MAFFT v. 7.036 (<https://mafft.cbrc.jp/alignment/server/>) using default settings. The alignments were manually improved using BioEdit v. 7.0.5.2 (Hall 1999) and AliView v. 1.26 (Larsson 2014).

Maximum likelihood, maximum parsimony and Bayesian inference methods were used for the phylogenetic reconstructions of the multi-locus and individual genes. Maximum likelihood analyses were performed by running 1000 pseudoreplicates using RAxML-HPC2 on XSEDE (8.2.8) (Stamatakis 2014) in the CIPRES Science Gateway platform (Miller et al. 2017). Maximum parsimony analyses were conducted using PAUP v.4.0b 10 (Swofford 2002) using the heuristic search option with 1000 bootstrap replicates. The Tree Length (TL), Consistency Indices (CI), Retention Indices (RI), Rescaled Consistency Indices (RC) and Homoplasy Index (HI) were calculated. The best-fitting substitution models were determined by jModelTest version 2.1.10 (Darriba et al. 2012). Bayesian inference analyses were conducted using MrBayes v. 3.2.2 (Ronquist et al. 2003). Six simultaneous Markov chains were run for 50,000,000 generations and trees were sampled every 1000th generation. The suitable burn-in phases were determined using Tracer version 1.7 (Rambaut et al. 2018) and were discarded. The remaining trees were used to calculate posterior probabilities in the majority rule consensus tree. FigTree v. 1.4 was used to view the resulting trees (Rambaut 2014) and the final layout was done with Adobe Illustrator CS v. 22.1 (Adobe Systems, USA).

Divergence time estimation

Divergence times were estimated in BEAST 2.6.2 (Bouckaert et al. 2019) based on the best fitting substitution models used for phylogenetic analyses. The dataset (ITS, *GAPDH*, *CHS-1*, *ACT* and *TUB2*) was partitioned and the XML file was prepared in BEAUTI 2.6.2. An uncorrelated relaxed clock model with lognormal distribution was used for the

fossil analysis. A Yule speciation process birth rate was used for the tree prior. The fossil *Protocolletotrichum deccanensis* represents an ancient lineage of extant *Colletotrichum* and it was used to constrain the common ancestor of *Colletotrichum* to the minimum age of 61 million years ago (MYA) (Kar et al. 2004; Hacquard et al. 2016; Samarakoon 2019). The analysis was carried out for 200 million generations and trees were sampled every 10,000th generations. The output was visualized in Tracer version 1.7 to ensure an effective sample size (ESS) greater than 200 (Drummond et al. 2012). Five independent analyses were performed to ensure congruence. The first 10% of the generated trees representing the burn-in phase were discarded. The remaining trees were combined using LogCombiner 2.6.2, summarized in TreeAnnotator 2.6.2 (BEAST packages) and visualized with FigTree v.1.4.0.

Multi-rate Poisson tree processes

The multi-rate Poisson tree processes (mPTP) incorporates different levels of intraspecific genetic divergences for each species (Kapli et al. 2017). The RAxML tree was used as the input file as this approach takes in a binary phylogenetic tree. The phylogenetic tree was rooted in mPTP v. 0.2.4 at *Monilochaetes camelliae* (BRIP 24607) and *Monilochaetes infuscans* (CBS 869.96) as the outgroup taxa.

The general mixed Yule-coalescent method

The GMYC method uses a likelihood approach, which is a combination of the neutral coalescent theory with Yule speciation model (Fujisawa and Barraclough 2013). It aims to detect shifts in branching rates between intra- and inter-specific relationships by comparing two models, the null and GMYC models. Under the null model, all the individuals belong to a single species or population (Fujisawa and Barraclough 2013). In the GMYC model, a Yule speciation and extinction model is applied whereas a coalescent process is applied to intraspecific relationships (Parnmen et al. 2012). If the GMYC model fits the data significantly better than the null model, the threshold can be used to estimate the number of species. The GMYC approach takes in an ultrametric and bifurcating tree file. The input tree was prepared from the output tree from the BEAST analysis with TreeAnnotator 2.6.2 using the maximum clade credibility method. The GMYC analyses were performed in R (R Core Team 2017) using the splits package v. 1.0-19 (Ezard et al. 2015). The analyses were performed using the single and multiple-threshold method. In the single threshold method, a single transition model is applied which assumes that all species have the same coalescent branching rate whereas in the multiple-threshold method variable transition is applied from coalescent to speciation across different clades (Monaghan

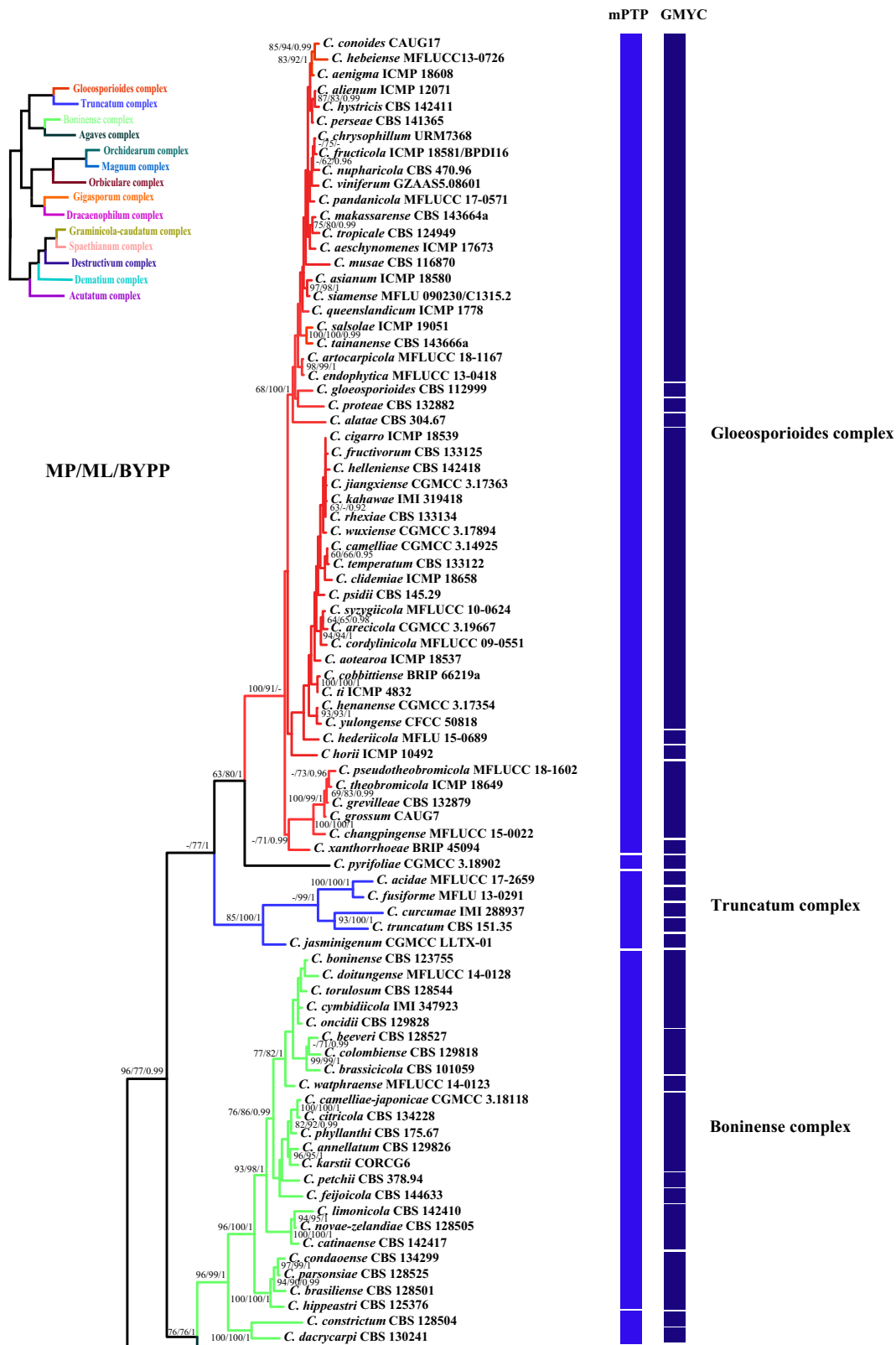


Fig. 1 One of the most parsimonious tree generated by maximum parsimony analysis of combined ITS, *GAPDH*, *CHS-1*, *ACT* and *TUB2* dataset of *Colletotrichum* type sequences. RAxML bootstrap support and maximum parsimony values $\geq 60\%$ (BT) as well as Bayesian posterior probabilities ≥ 0.90 (BYPP) are shown respectively near the nodes. The ex-type strains are in bold and the scale bar

indicates 80 changes. The tree is rooted with *Monilochaetes camelliae* (BRIP 24607) and *Monilochaetes infuscans* (CBS 869.96). The columns present the results of mPTP and GMYC based on the type sequences. The figure in the upper left corner represents the placement of all the complexes based on the maximum parsimony analysis. Each complex is represented by a different colour

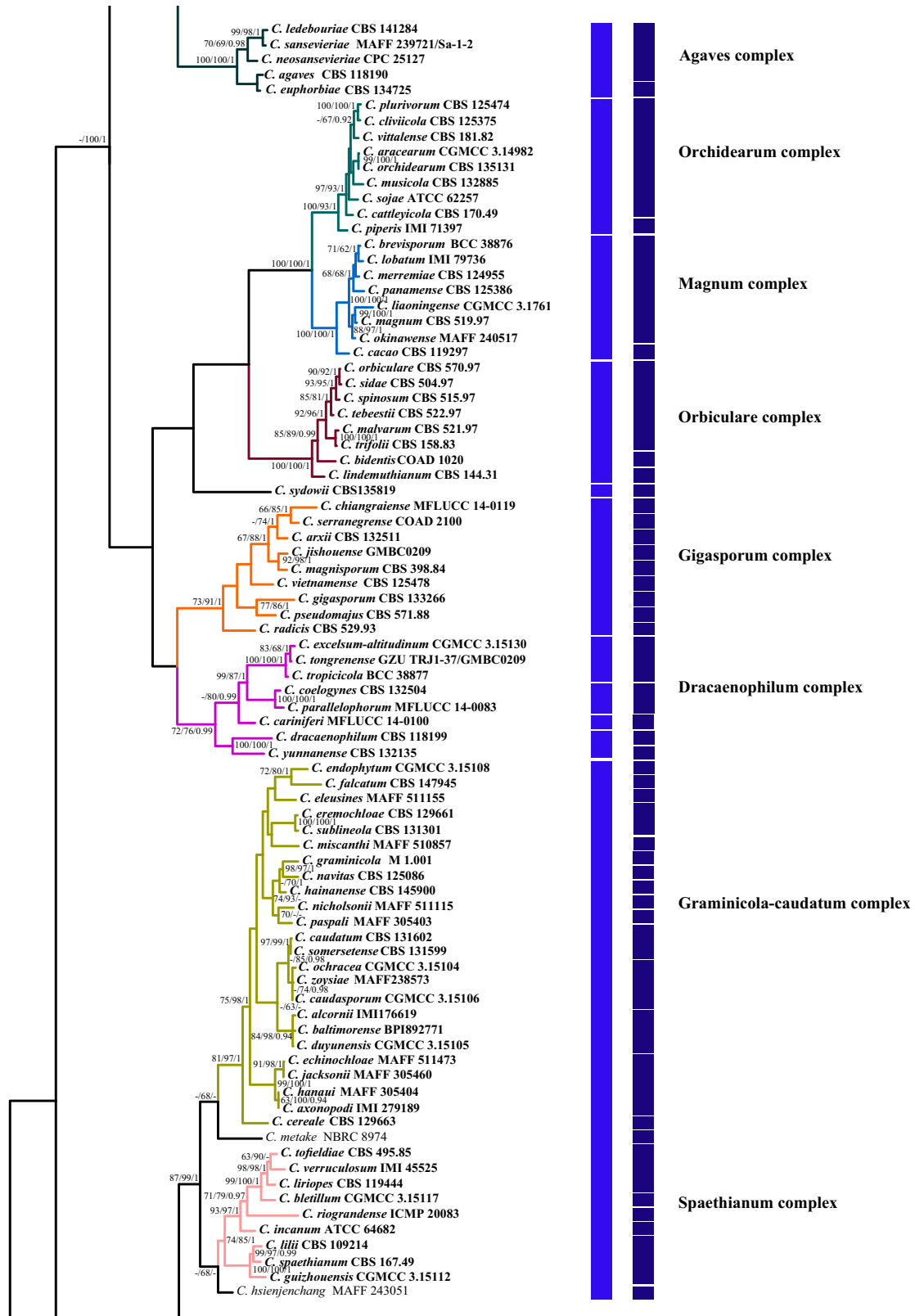


Fig. 1 (continued)

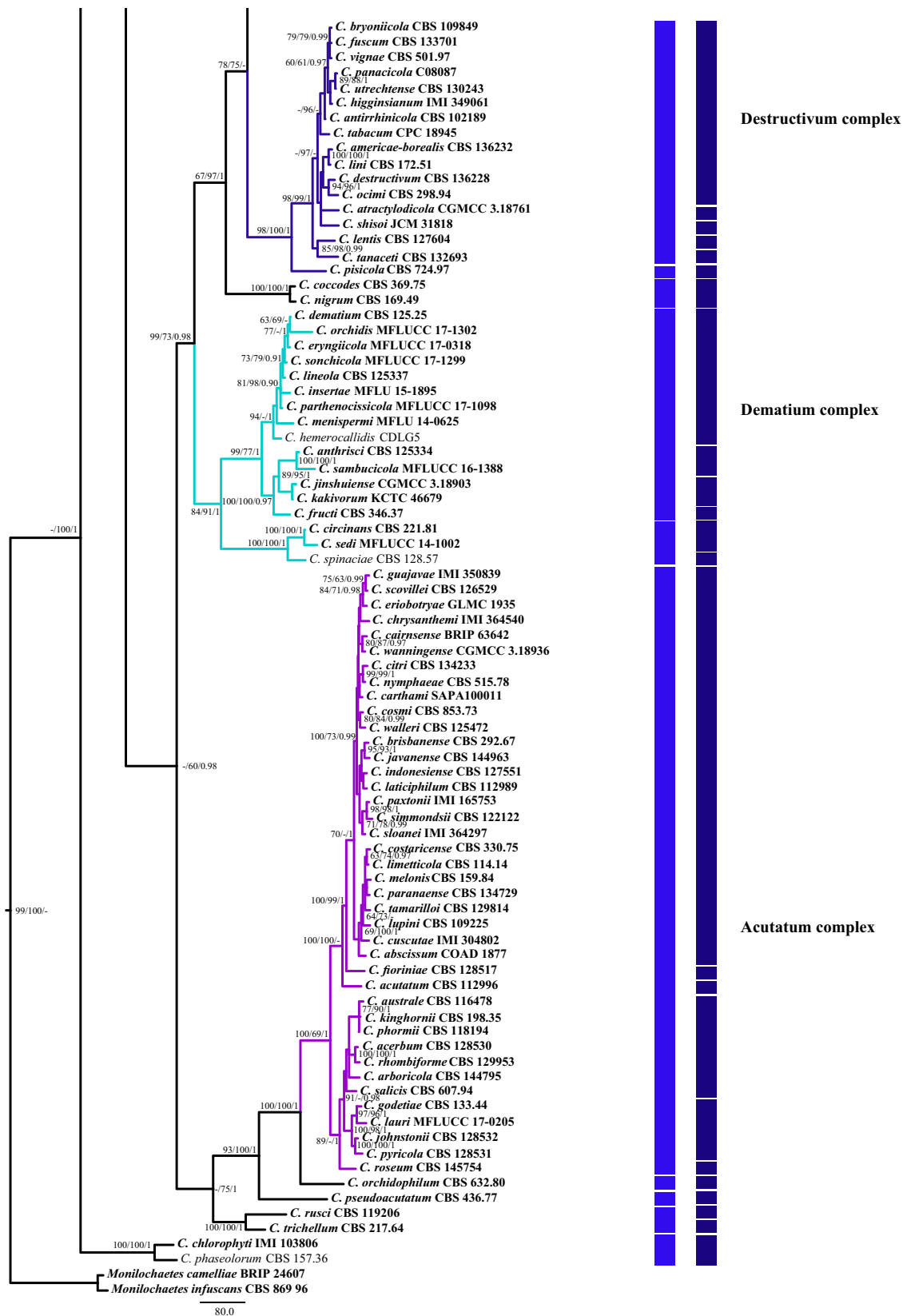


Fig. 1 (continued)

et al. 2009). All the analyses in this study (divergence time estimation, mPTP and GMYC) were repeated using the larger dataset.

Results

Phylogenetic analyses

The ITS dataset consisted of 250 taxa with 322/635 conserved sites, 289/635 variable sites and 212/635 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree out of 1000 (CI=0.375, RI=0.882, RC=0.331, HI=0.625, Tree Length=1260). The best scoring RAxML tree had a final likelihood value of -7658.697005. The general time reversible (GTR) model with a discrete gamma distribution plus invariant site (GTR + I + G) substitution model was implemented in the Bayesian analysis of the ITS dataset. The *GAPDH* dataset consisted of 217 taxa with 52/461 conserved sites, 382/461 variable sites and 317/461 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree out of 1000 (CI=0.314, RI=0.858, RC=0.270, HI=0.686, Tree Length=2723). The best scoring RAxML tree had a final likelihood value of -12,518.203342. The Hasegawa–Kishino–Yano (HKY) model with a discrete gamma distribution plus invariant site (HKY + I + G) substitution model was implemented in the Bayesian analysis of the *GAPDH* dataset. The *TUB2* dataset consisted of 235 taxa with 195/619 conserved sites, 365/619 variable sites and 307/619 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree out of 1000 (CI=0.268, RI=0.856, RC=0.230, HI=0.732, Tree Length=2857). The best scoring RAxML tree had a final likelihood value of -14,703.137100. A HKY + I + G model was implemented in the Bayesian analysis of the *TUB2* dataset. The *CHS-1* dataset consisted of 201 taxa with 155/282 conserved sites, 127/282 variable sites and 100/282 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree out of 1000 (CI=0.228, RI=0.813, RC=0.186, HI=0.772, Tree Length=871). The best scoring RAxML tree had a final likelihood value of -4786.387059. A GTR + I + G model was implemented in the Bayesian analysis of the *CHS-1* dataset. The *ACT* dataset consisted of 232 taxa with 72/340 conserved sites, 248/340 variable sites and 206/340 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree out of 1000 (CI=0.319, RI=0.862, RC=0.275, HI=0.681, Tree Length=1594). The best scoring RAxML tree had a final likelihood value of -8269.293999. A GTR + I + G model was implemented in the Bayesian analysis of the *ACT* dataset.

The multi-locus dataset (ITS, *GAPDH*, *CHS-1*, *ACT* and *TUB2*) consisted of 250 taxa with 796/2337 conserved sites, 1411/2337 variable sites and 1142/2337 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree (Fig. 1) out of 1000 (CI=0.282, RI=0.844, RC=0.238, HI=0.718, Tree Length=9953). The best scoring RAxML tree had a final likelihood value of -52,302.678794. With some exceptions, all three phylogenetic approaches resulted in similar tree topology and major clades based on the multi-locus dataset. The *gigasporum* and *dracaenophilum* complexes formed a sister clade to the *gloeosporioides/truncatum* and *agaves/boninense* complexes in the Bayesian Inference analysis (0.99 BYPP). In the maximum parsimony and maximum likelihood analysis, the *dracaenophilum* and *gigasporum* complexes are closely related to the *orbiculare/orchidearum/magnum* complexes (> 70% BT). Two taxa from the *graminicola* complex (*C. endophyllum* and *C. falcatum*) clustered together with strong support (1.00 BYPP), forming a basal clade to the *caudatum* complex in the Bayesian Inference analysis. The singleton *C. hsienjenchang* clustered with the singleton *C. metake* at the basal lineage of the *graminicola* complex in the maximum likelihood and Bayesian Inference analysis (68% MLBT/0.81 BYPP). In the maximum parsimony analysis, *C. metake* formed a sister taxon to the *graminicola* complex with low support (< 50% BT) and *C. hsienjenchang* formed a sister taxon to the *spaethianum* complex with low support (< 50% BT). The *caudatum* complex formed an inner clade in the *graminicola* complex based on all three phylogenetic analyses (98% MLBT/84% MPBT/1.00 BYPP). Five singleton taxa *C. agaves*, *C. euphorbiae*, *C. ledebou-riai*, *C. neosansevieriae* and *C. sansevieriae* formed a new complex in the maximum parsimony, maximum likelihood and the Bayesian inference analyses with strong support (100% MLBT/100% MPBT/1.00 BYPP). These species formed a sister clade to the *boninense* complex in all the phylogenetic analyses. The singleton *C. parallelophorum* (MFLUCC 14-0083) formed a sister taxon to *C. coelogyne*s in the *dracaenophilum* complex with strong support (100% MLBT/100% MPBT/1.00 BYPP). The singleton *C. cariniferi* (MFLUCC 14-0100) clustered in the *dracaenophilum* complex, forming a basal clade to *C. parallelophorum* and *C. coelogyne*s (80% MLBT/52% MPBT/0.99 BYPP). The singleton *C. axonopodi* (IMI279189) formed a sister taxon to *C. hanai* in the *graminicola* complex with strong support (100% MLBT/63% MPBT/0.94 BYPP).

The phylogenetic trees from the single genes were compared to the phylogenetic tree from the multi-locus dataset in terms of topology and support for species relationships. The ITS dataset resolved the placement of all the taxa up to complex level, except *C. cereale* and *C. orchidis* which clustered in the *spaethianum* and *destructivum* complexes respectively (Fig. S2). The *caudatum* complex formed an

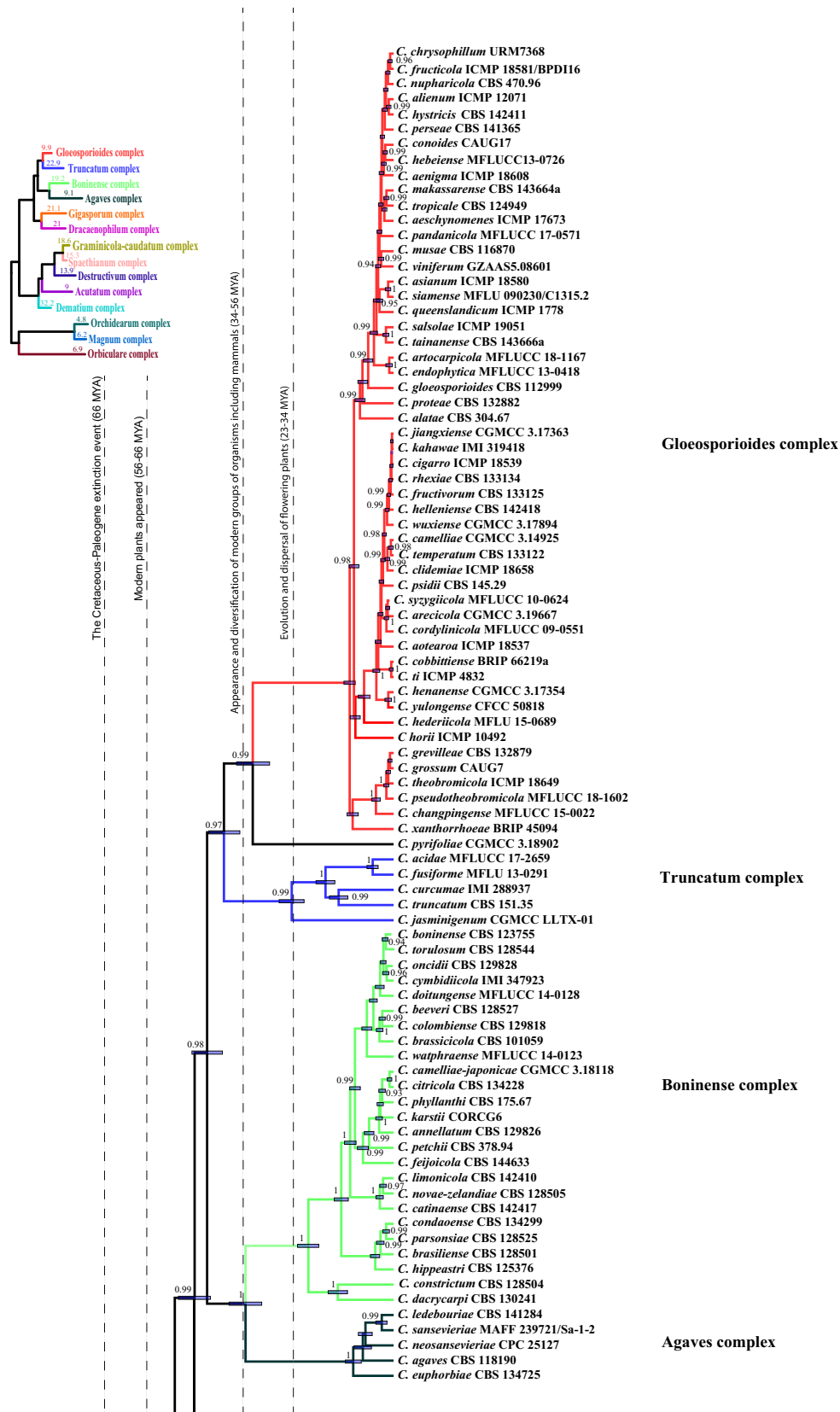


Fig. 2 Maximum clade credibility tree with divergence time estimates using BEAST. The divergence time is shown in million years and the numbers at the nodes indicate posterior probabilities (BYPP) for node support. Bars correspond to the 95% highest posterior density (HPD) intervals. The figure in the upper left corner represents the

placement of all the complexes based on the maximum clade credibility tree and the estimated crown ages for all the complexes are shown in MYA. Each complex is represented by a different colour. Geological time scales are given at the base, together with scale in MYA (Cohen et al. 2013)

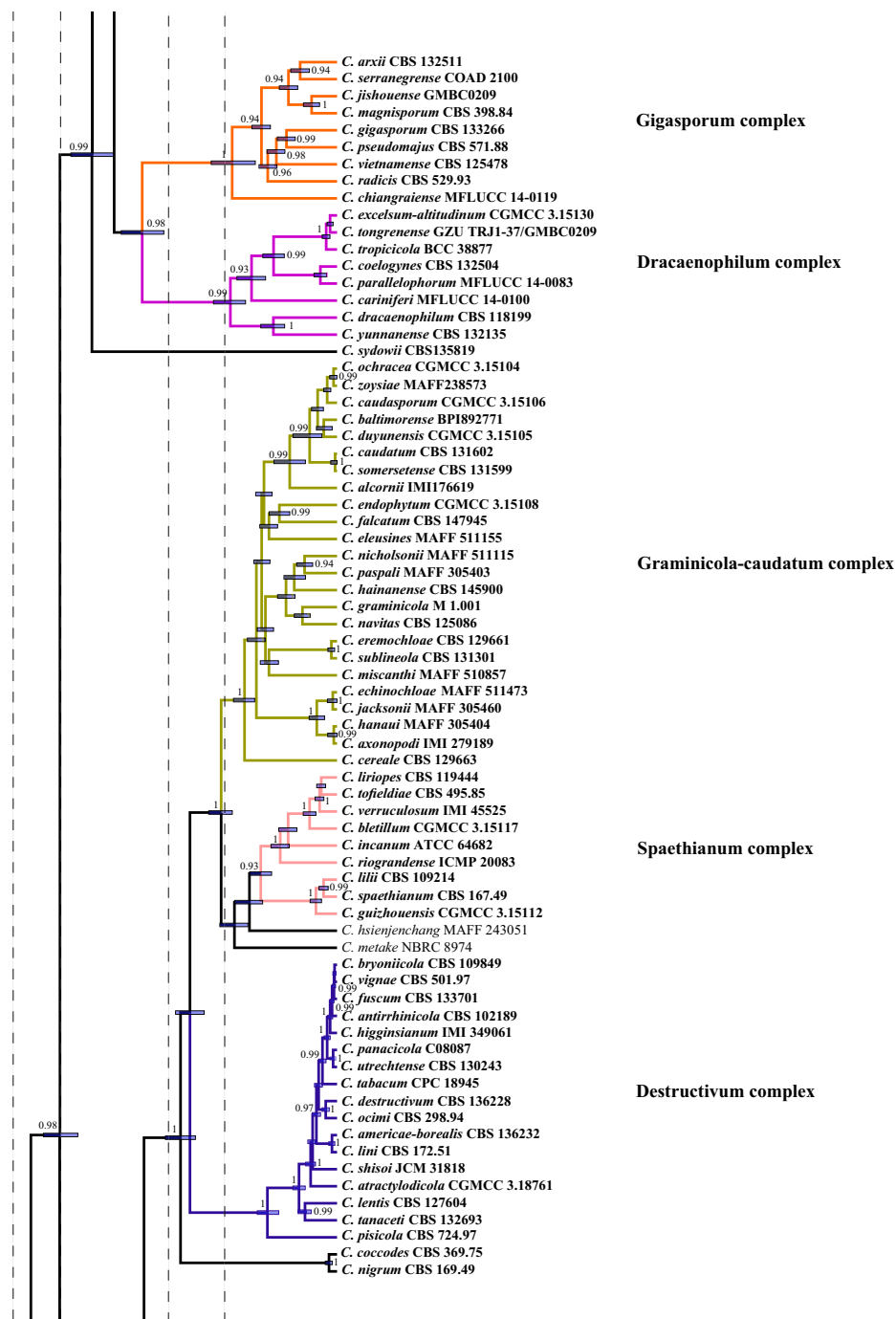


Fig. 2 (continued)

inner clade in the graminicola complex based on all three phylogenetic analyses of the ITS region (90% MLBT/83% MPBT/1.00 BYPP). The truncatum complex formed a sister clade to the boninense complex in the ITS dataset (95% MLBT/97% MPBT/1.00 BYPP) whereas the truncatum complex formed a sister clade to the gloeosporioides complex

in the multi-locus dataset (100% MLBT/85% MPBT/1.00 BYPP). Based on the phylogenetic tree derived from the ITS dataset, the dracaenophilum complex is closely related to the gigasporum complex in the maximum parsimony and the Bayesian inference analyses (1.00 BYPP) whereas it forms a sister clade to the orchidearum/magnum complexes in the



Fig. 2 (continued)

Table 1 Estimated crown and stem ages for all the complexes based on multi-locus and larger dataset. The median and the 95% Highest Posterior Density are provided in MYA

Complexes	Multi-locus dataset		Larger dataset	
	Crown age	Stem age	Crown age	Stem age
Acutatum	9 (7.7–10.3)	15.4 (13.4–17.6)	9.3 (8–10.7)	16.1 (13.7–18.7)
Agaves	9.1 (7.3–10.9)	33.3 (29.6–37)	9.5 (7.7–11.5)	34.5 (31–38.2)
Boninense	19.2 (16.8–21.6)	33.3 (29.6–37)	19.1 (16.9–21.5)	34.5 (31–38.2)
Graminicola-caudatum	18.6 (16.5–20.8)	23.3 (21–25.8)	18 (16.1–20.1)	23 (20.7–25.1)
Dematium	32.2 (28.2–36)	41.3 (38–44.6)	28 (24.2–32.8)	34.6 (31.1–38.6)
Destructivum	13.9 (11.7–16)	29.5 (26.7–32.4)	13.4 (11.4–15.6)	30 (27.1–33.2)
Dracaenophilum	21 (18.3–24.8)	39.1 (34.7–43.3)	21.4 (18.8–24.3)	35.5 (31.5–39.3)
Gigasporum	21.1 (16.4–25.3)	39.1 (34.7–43.3)	20.6 (17.12–24)	35.5 (31.5–39.3)
Gloeosporioides	9.9 (8.6–11)	31.7 (27.9–35.3)	10.25 (9–11.7)	36.1 (32.1–40.5)
Magnum	6.2 (5–7.3)	14.8 (12.7–17)	5.8 (4.7–7)	14.1 (12–16.1)
Orbiculare	6.9 (5.6–8.3)	55.8 (50.5–60.9)	6.7 (5.4–8.1)	48.8 (42.7–58)
Orchidearum	4.8 (3.8–5.9)	14.8 (12.7–17)	5.3 (4.3–6.6)	14.1 (12–16.1)
Spaethianum	15.3 (13–17.6)	17.5 (14.9–20.5)	16 (13.7–18.4)	20.7 (18.1–23.4)
Truncatum	22.9 (20–25.7)	38.2 (34.5–41.6)	23.7 (20.8–26.7)	40.6 (37.1–44)

maximum likelihood analysis (56% MLBT). A new complex with five taxa *C. agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosansevieriae* and *C. sansevieriae* was formed in the maximum parsimony, maximum likelihood and the Bayesian inference analyses based on the ITS dataset with strong support (100% MLBT/99% MPBT/1.00 BYPP). *Colletotrichum axonopodi* formed a sister taxon to *C. hanau* in the graminicola complex based on the phylogenetic tree derived from the ITS dataset (99% MLBT/72% MPBT/0.97 BYPP). *Colletotrichum parallelophorum* formed a sister taxon to *C. cariniferi* in the dracaenophilum complex with strong support only in the maximum likelihood analysis based on the ITS dataset (93% MLBT). The single marker *GAPDH* and *TUB2* recovered the highest number of species in the acutatum, boninense, dematium, destructivum, dracaenophilum, magnum and spaethianum complexes as monophyletic with strong support ($\geq 70\%$ BT/ ≥ 0.90 BYPP). The *TUB2* gene was the most informative in the gloeosporioides, graminicola and orchidearum complexes whereas *GAPDH* was the most informative in the gigasporum, orbiculare and truncatum complexes. The *ACT* gene was the next best marker in the graminicola and spaethianum complexes. The *GAPDH*, *ACT* and *TUB2* markers were all equally effective in species recognition in the caudatum and agaves complex.

Coalescent-based species delimitation methods

The mPTP method recovered 26 entities in the multi-locus dataset (Fig. 1). The acutatum, gigasporum, gloeosporioides, magnum, orbiculare, orchidearum and truncatum complexes were each recovered as distinct taxonomic entities. *Colletotrichum axonopodi*, *C. hsienjenchang*, *C. metake* as well as all the species from the caudatum, graminicola, and spaethianum complexes were recovered as conspecific.

The remaining complexes were split into several entities (boninense into 2 entities; dematium 2; destructivum 2; dracaenophilum into 3 entities). *C. agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosansevieriae* and *Colletotrichum sansevieriae* were all recovered as one entity. The singleton *C. coelogyne*s and *C. parallelophorum* from the dracaenophilum complex were recovered as conspecific.

The GMYC approach detected 97 entities based on the multi-locus dataset in the single and multiple threshold method (Fig. 1 and Fig. S3). The GMYC model provided a better fit to the ultra-metric tree based on the likelihoods for the null (L_0) and GMYC (L_{GMYC}) models (single threshold: L_0 : 396.29, L_{GMYC} : 403.0176, likelihood ratio (LR): 13.4552; multiple threshold: L_0 : 396.29, L_{GMYC} : 409.6819, LR: 26.78377). All the singletons were recovered as conspecific with few exceptions as detailed in Fig. 1. *Colletotrichum axonopodi* and all the species from the graminicola complex were recovered as one taxonomic unit. *Colletotrichum parallelophorum* and all species from the dracaenophilum complex were recovered as conspecific. Both the magnum and orchidearum complexes were divided into two distinct entities. The remaining complexes were divided into several entities (acutatum into 6 entities; boninense 10; caudatum 3; destructivum 6; dracaenophilum 4; dematium 6; gigasporum 9; gloeosporioides 9; graminicola 12; orbiculare 3; spaethianum 5; truncatum complex into 5 entities). *Colletotrichum ledebouriae* and *C. sansevieriae* were recovered as conspecific.

Divergence time estimation

The topology of the maximum clade credibility (MCC) tree from the BEAST analysis (Fig. 2) was generally congruent

with the result from the maximum likelihood, maximum parsimony and Bayesian inference analysis with some exceptions. The dracaenophilum and gigasporum complexes formed a sister clade to the agaves/boninense and gloeosporioides/truncatum complexes similar to the Bayesian Inference analysis (0.98 BYPP). Three species from the graminicola complex *C. eleusines*, *C. endophytum* and *C. falcatum* formed a basal clade to the caudatum complex. The caudatum complex formed an inner clade in the graminicola complex. *Colletotrichum axonopodi* formed a sister taxon to *C. hanau* in the graminicola complex with strong posterior probability (0.99 BYPP). *Colletotrichum parallelophorum* formed a sister taxon to *C. coelogyne*s in the dracaenophilum complex with strong posterior probability (1.00 BYPP). *Colletotrichum cariniferi* clustered in the dracaenophilum complex, forming a basal clade to *C. parallelophorum* and *C. coelogyne*s (0.93 BYPP).

The crown age of the acutatum complex at 9 MYA, boninense, complex at 19.2 MYA, caudatum complex was estimated at 9.5 MYA, dematium complex at 32.2 MYA, destructivum complex at 13.9 MYA, dracaenophilum complex at 21 MYA, gigasporum complex at 21.1 MYA, gloeosporioides complex at 9.9 MYA, magnum complex at 6.2 MYA, orbiculare complex at 6.9 MYA, orchidearum complex at 4.8 MYA, spaethianum complex at 15.3 MYA and the truncatum complex at 22.9 MYA (Table 1). *Colletotrichum agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosanseviera*e and *C. sanseviera*e formed a separate complex with a crown age of 9.1 MYA. The stem age of this complex and the boninense complex was estimated at 33.3 MYA. The caudatum and graminicola complex (including the singleton *C. axonopodi*) diverged at 18.6 MYA. The stem age of the gigasporum and dracaenophilum complex (including the singleton *C. parallelophorum* and *C. cariniferi*) was estimated at 39.1 MYA. The orchidearum and magnum complex diverged at 14.8 MYA. The orbiculare complex diverged from the orchidearum and magnum complex at 55.9 MYA.

Larger dataset results

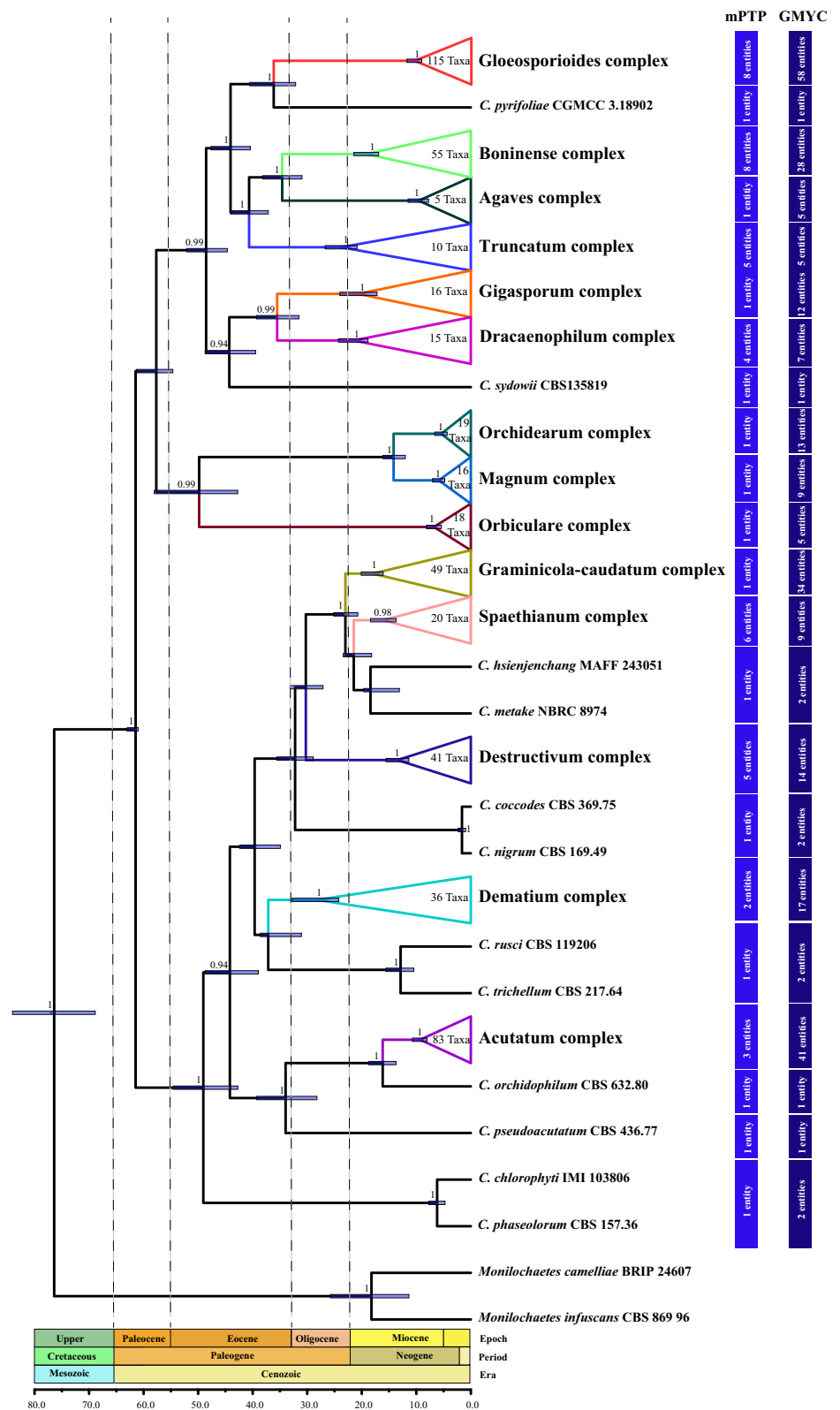
The larger concatenated dataset consisted of 555 taxa and sequences which did not cluster with their respective types based on all three phylogenetic analyses were removed. The final dataset consisted of 512 taxa with 824/2289 conserved sites, 1409/2289 variable sites and 1232/2289 were parsimony informative. The parsimony analysis of the dataset yielded one most parsimonious tree (Fig. S4) out of 1000 (CI=0.273, RI=0.920, RC=0.251, HI=0.727, Tree Length=10,334). The best scoring RAXML tree had a final likelihood value of -55,865.238686. With few exceptions, all three phylogenetic analyses based on the larger dataset resulted in clades with similar topology. The truncatum complex formed a sister clade to the gloeosporioides

complex in the maximum likelihood analysis and Bayesian Inference analysis (99% MLBT/1.00 BYPP), but it formed a sister clade to the boninense and new species complex in the maximum parsimony analysis (87% MPBT).

Phylogenetic trees derived from the larger dataset was similar to that from the multi-locus datasets, except for the gigasporum and dracaenophilum complexes, as well as for few taxa from the graminicola complex. In the maximum parsimony analysis, *C. hainanense* formed a basal clade to the caudatum complex (100% MPBT). In the maximum likelihood analysis and Bayesian Inference analysis, *C. eremochloae* and *C. sublineola* formed a basal clade to the caudatum complex (100% MLBT/1.00 BYPP). The taxonomic placement of the gigasporum and dracaenophilum complexes did not differ in the different phylogenetic analyses based on the larger dataset. *Colletotrichum hsienjenchang* clustered with *C. metake* in the maximum likelihood and the Bayesian inference analyses (51% MLBT/0.70 BYPP). In the maximum parsimony analysis, *C. metake* formed a sister taxon to the graminicola complex with low support and *C. hsienjenchang* formed a sister taxon to the spaethianum complex with low support. The caudatum complex formed an inner clade in the graminicola complex based on all phylogenetic analyses of the larger dataset (100% MLBT/87% MPBT/1.00 BYPP). *Colletotrichum agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosanseviera*e and *C. sanseviera*e formed a new complex in the maximum parsimony, maximum likelihood and the Bayesian inference analyses with strong support (100% MLBT/100% MPBT/1.00 BYPP). This complex formed a sister clade to the boninense complex in all phylogenetic analyses. *Colletotrichum parallelophorum* formed a sister taxon to *C. coelogyne*s species in the dracaenophilum complex (100% MLBT/100% MPBT/1.00 BYPP). *Colletotrichum cariniferi* clustered in the dracaenophilum complex with strong support in the Bayesian inference analysis, forming a basal clade to *C. parallelophorum* and *C. coelogyne*s (58% MLBT/0.98 BYPP) whereas *C. axonopodi* formed a sister taxon to *C. hanau* in the graminicola complex in all phylogenetic analyses (81% MLBT/0.86 BYPP).

The mPTP method delineated a higher number of entities (58 entities) in the larger dataset (Fig. 3) compared to the multi-locus dataset. Compared to the multi-locus dataset, mPTP did not recover *C. hsienjenchang*, *C. metake* and the spaethianum complex as part of the caudatum and graminicola complexes in the larger dataset. In the larger dataset, *C. coelogyne*s was not recovered as part of the dracaenophilum complex as compared to the multi-locus dataset. All the species in the gigasporum, magnum, orbiculare and orchidearum complexes were identified as distinct entities. The remaining complexes were divided into several entities (acutatum complex into 3 entities; boninense 8; dematium 2; destructivum 5; dracaenophilum 4; gloeosporioides 8; spaethianum 6; truncatum into 5). All the

Fig. 3 Maximum clade credibility tree of the larger dataset with divergence time estimates using BEAST. The node represents the calibration point and the divergence time are shown in million years. Numbers at the nodes indicate posterior probabilities (BYPP) for node support. Bars correspond to the 95% highest posterior density (HPD) intervals. Each complex is represented by a different colour and the number of taxa (including the type taxa) in each complex is indicated. The columns present the results of mPTP and GMYC based on the larger dataset. Geological time scales are given at the base, together with scale in MYA (Cohen et al. 2013)



species from the caudatum and graminicola complexes including *C. axonopodi* were recovered as conspecific. *Colletotrichum agaves*, *C. euphorbiae*, *C. ledebouriae*,

C. neosansevieriae and *C. sansevieriae* were identified as one taxonomic unit in the larger dataset. *Colletotrichum*

parallelophorum along with the dracaenophilum complex were recovered as conspecific.

The GMYC method detected 279 entities in the larger dataset with the single and multiple threshold method (Fig. 3 and Fig. S5), which is over twice the number of entities detected in the multi-locus dataset. In the larger dataset, *C. axonopodi* was not recovered as part of the graminicola complex and *C. parallelophorum* was not recovered as part of the dracaenophilum complex. In the larger dataset, the GMYC model provided a better fit to the ultra-metric tree based on the likelihoods for the null (L_0) and GMYC (L_{GMYC}) models (single threshold: L_0 : 1676.793, L_{GMYC} : 1688.774, likelihood ratio (LR): 23.96139; multiple threshold: L_0 : 1676.793, L_{GMYC} : 1707.259, LR: 60.93233). All the complexes were divided into several entities (acutatum into 41 entities; agaves 5; boninense 28; caudatum 10; dematium 17; destructivum 14; dracaenophilum 7; gigasporum 12; gloeosporioides 58; graminicola 24; magnum 9; orbiculare 5; orchidearum 13; spaethianum 9; truncatum into 5 entities). All the singleton species were recovered as conspecific.

The maximum clade credibility tree (MCC) tree based on the larger dataset resulted in similar estimated crown ages as the MCC tree for the multi-locus dataset of type species with the exception of the truncatum complex and few taxa from the graminicola complex. In the larger dataset, the truncatum complex formed a sister clade to the agaves/boninense complexes (1.00 BYPP) and the taxa from the graminicola complex (*C. endophytum*, *C. eremochloae*, *C. falcatum* and *C. sublineola*) formed a basal clade to the caudatum complex. The topology of the maximum clade credibility (MCC) tree from the BEAST analysis of the larger dataset (Fig. 3) was generally congruent with the result from the maximum likelihood, maximum parsimony and Bayesian inference analysis with some exceptions. The gigasporum and dracaenophilum complexes formed a sister clade to the gloeosporioides and boninense/sansevieriae/truncatum complexes in the MCC tree (0.99 BYPP). The caudatum complex formed an inner clade in the graminicola complex in the MCC tree. Three singleton taxa clustered in species complexes based on the MCC tree. *Colletotrichum parallelophorum* formed a sister taxon to *C. coelogyne* species in the dracaenophilum complex with strong posterior probability (1.00 BYPP). *Colletotrichum cariniferi* clustered in the dracaenophilum complex with strong support, forming a basal clade to *C. parallelophorum* and *C. coelogyne* (1.00 BYPP) and *C. axonopodi* formed a sister taxon to *C. hanau* in the graminicola complex (0.90 BYPP).

The estimated crown age based on the larger dataset was similar to the multi-locus dataset (Table 1). The crown age of the acutatum complex was estimated at 9.3 MYA, boninense complex at 19.1 MYA, caudatum complex at 5.8 MYA, dematium complex at 28 MYA, destructivum complex at 13.4 MYA, dracaenophilum complex at 21.4 MYA,

gigasporum complex at 20.6 MYA, gloeosporioides complex at 10.25 MYA, orbiculare complex at 6.7 MYA, orchidearum complex at 5.3 MYA, magnum complex at 5.8 MYA, spaethianum complex at 16 MYA and the truncatum complex at 23.7 MYA. *Colletotrichum agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosansevieriae* and *C. sansevieriae* formed a separate complex with a crown age of 9.5 MYA. The stem age of the dracaenophilum and gigasporum complex (including the singleton *C. cariniferi* and *C. parallelophorum*) was estimated at 35.5 MYA. The magnum and orchidearum complex diverged at 14.1 MYA. The orbiculare complex diverged from the magnum and orchidearum complex at 48.9 MYA.

Discussion

In this study, we evaluated the phylogenetic significance of five markers for species delineation within each species complex. The highest percentage of parsimony informative sites was recovered from the *GAPDH* gene (69%), followed by *ACT* (60%), *TUB2* (50%), the multi-locus dataset (49%), *CHS-1* (36%) and the ITS region (33%). Phylogeny based on the ITS marker effectively resolved the placement of all the taxa up to complex level, except *C. cereale* and *C. orchidis* which clustered in the spaethianum and destructivum complexes respectively. However, the placement of these two taxa can be resolved by using the multi-locus dataset. In the multi-locus dataset, *C. cereale* clustered in the graminicola complex and *C. orchidis* clustered in the dematium complex. The ITS region has been suggested to represent the universal barcode for fungi (Schoch et al. 2012) but with several concerns (Lücking et al. 2020) and a database of annotated ITS region of plant pathogens was published in UNITE (Nilsson et al. 2014).

Phylogenetic trees derived from the ITS, multi-locus and the larger dataset resulted in 14 strongly supported major clades, including a new complex, the agaves complex (Fig. 1, Fig. S2 and Fig. S4). This is in accordance to previous studies for 13 of the complexes (excluding the agaves complex) (excluding the agaves complex) (Cannon et al. 2012; Jayawardena et al. 2016a, 2020). The caudatum complex formed a strongly supported inner clade in the graminicola complex based on phylogenetic trees derived from the ITS, multi-locus dataset and the larger dataset which is also supported by previous studies (Jayawardena et al. 2016a, 2020). Based on phylogenetic trees derived from the ITS, multi-locus and larger dataset, three singleton taxa *C. axonopodi*, *C. cariniferi* and *C. parallelophorum* clustered in species complexes with strong support.

Five of the 13 markers that have been used in previous studies for species delineation in *Colletotrichum* were selected in this study. These markers were selected based

on the availability of data especially for the type species and these represent the minimal set of markers to accurately delineate species in any of the species complexes. The intergenic region of DNA lyase (*apn2*) and mating type (*MAT1-2-1*) genes (*ApMat*) and Glutamine synthetase (*GS*) genes, which are important markers for the gloeosporioides complex (Liu et al. 2015; Sharma et al. 2015) were not included due to the lack of data available for isolates from complexes other than the gloeosporioides complex. The *GAPDH* and *TUB2* were the most informative markers for seven species complexes which is in accordance with previous studies for the acutatum and destructivum complexes (Damm et al. 2012b, 2014). However, the performance of the most informative marker was not the same for all the complexes. The *GAPDH* gene for example was the most informative marker for the gigasporum, orbiculare and truncatum complexes. The *TUB2* marker was the most informative for the gloeosporioides, graminicola and orchidearum complexes. The *GAPDH*, *ACT* and *TUB2* markers were all equally informative for the caudatum and the new complex. However, the assessment of the phylogenetic significance of the different markers is hindered by the lack of data, for example, the *CHS-1* marker is not available for the caudatum complex and the *GAPDH* marker is missing for most of the species in the graminicola complex. Therefore, the phylogenetic significance of these markers could not be assessed.

The mPTP method has been used for species delineation in several organisms with limited phenotypic variation (Liu et al. 2016; Correa et al. 2017; Marki et al. 2018). In this study, mPTP detected 26 taxonomic units in the multi-locus dataset compared to 58 entities in the larger dataset. In the multi-locus dataset, all the species from the caudatum, graminicola, and spaethianum complexes as well as the singletons *Colletotrichum axonopodi*, *C. hsienjenschap*, *C. metake* were recovered as conspecific. In the larger dataset, all the species from the caudatum and graminicola complexes including *C. axonopodi* were recovered as conspecific. Five singleton taxa *C. agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosansevieriae* and *C. sansevieriae* were identified as one taxonomic unit in the multi-locus and the larger dataset. *Colletotrichum parallelophorum* was identified as part of the dracaenophilum complex in both datasets. In the larger dataset, *C. axonopodi* and the graminicola complex were identified as one entity. The results from mPTP therefore provides further support for the new *agaves* complex. It also supports the taxonomic placement of *C. axonopodi*, *C. parallelophorum* and the placement of the caudatum complex in the graminicola complex.

Coalescent based approaches such as GMYC have been crucial for identifying cryptic species in numerous organisms (Liu et al. 2016; Correa et al. 2017; Servis et al. 2020). The GMYC approach detected 97 entities in the multi-locus dataset compared to 279 entities in the larger dataset. In the

larger dataset, GMYC showed a tendency of over-splitting in some cases compared to phylogenetic and mPTP analyses. Over-splitting has been observed in some cases based on previous studies (Miralles et al. 2013; Zhang et al. 2013; Pentinsaari et al. 2016). Based on low congruence between the GMYC result of the larger dataset and the phylogenetic analyses, we refer to the GMYC results from the multi-locus only. *Colletotrichum axonopodi* and the graminicola complex were recovered as one taxonomic unit whereas *C. parallelophorum* and the dracaenophilum complex were recovered as conspecific and two taxa from the new complex were recovered as conspecific. The results from GMYC based on the smaller dataset were in accordance with phylogenetic and mPTP analyses, thereby supporting the placement of *C. axonopodi*, *C. parallelophorum* and two taxa from the new complex.

Our study provides a global evolutionary picture of all the species in *Colletotrichum* and provides an estimated divergence time of all the complexes. The topology of the MCC tree based on both datasets yielded similar results except for three species from the graminicola complex. The caudatum complex formed a strongly supported inner clade in the graminicola complex based on both datasets which is in accordance with phylogenetic analyses, mPTP and GMYC results. The divergence time estimates for the crown age of the complexes were similar in both datasets except for the caudatum complex. The MCC tree supported the phylogenetic placement of three singleton taxa (*C. axonopodi*, *C. cariniferi* and *C. parallelophorum*) within species complexes. The MCC tree based on both datasets strongly supported a new species complex with an estimated divergence time of 9.5 MYA. The MCC tree based on both datasets resulted in 14 strongly supported major clades including the new species complex. The estimated divergence time for the complexes varied from approximately 4.8 MYA in the orchidearum complex to 32.2 MYA in the dematium complex. The estimated divergence time of the new complex fits within the range of the divergence time of the complexes in *Colletotrichum*, which provides further support for the *agaves* complex.

The evolutionary estimates from this study were congruent with divergence estimates from previous studies. Hacquard et al. (2016) estimated the divergence time based on a penalized likelihood method of genome sequence data using the crown age of *Pezizomycotina*, *Sordariomycetes* and *Cordyceps-Metarhizium* divergence as the calibration points. They estimated that *C. incanum* and *C. tofieldiae* diverged approximately 8.8 MYA compared to 8–11 MYA in our study. Liang et al. (2018) estimated the divergence time of several complexes based on a penalized likelihood method of genome sequence data using the crown age of *Colletotrichum*, *Sordariomycetes*, and *Sordariomycetes-Leotiomyces* crown as the calibration points and the estimates were based

on four calibration schemes. In the gloeosporioides complex, *C. fructicola* and *C. gloeosporioides* diverged approximately 3.8–5.1 MYA compared to 4.6–8.6 MYA in our study. In the graminicola complex, *C. graminicola* and *C. sublineola* diverged approximately 11–15 MYA compared to 12–16 MYA in our study. In the spaethianum complex, *C. tofieldiae* and *C. incanum* diverged approximately 7.5–11 MYA compared to 8–11 MYA in our study. In the acutatum complex, *C. fioriniae*, *C. nymphaeae*, *C. salicis* and *C. simmondsii* diverged approximately 8–10.9 MYA compared to 7.7–10.3 MYA in our study. The most recent common ancestor (MRCA) of the acutatum, destructivum, graminicola and spaethianum complex was estimated at 37–49 MYA compared to 35–41 MYA in our study. The most recent common ancestor (MRCA) of the acutatum, graminicola, spaethianum and destructivum complex was estimated at 37–49 MYA compared to 35–41 MYA in our study. The MRCA of the graminicola and spaethianum complex was estimated at 19–26 MYA compared to 21–25 MYA in our study. The stem age of orbiculare complex was estimated at 47–60 MYA compared to 42–60 MYA in our study. Lelwala et al. (2019) estimated the divergence time based on a penalized likelihood method of genome sequence data using the crown age of *Leotiomycetes-Sordariomycetes*, crown age of *Sordariomycetes* and *Colletotrichum* crown as the calibration points. In the destructivum complex, *C. higginsianum* and *C. tanacetii* diverged around 9.9 MYA compared to 7.6 MYA in our study. In the gloeosporioides complex, *C. fructicola* and *C. gloeosporioides* diverged around 6.1 MYA compared to 7 MYA in our study. *Colletotrichum graminicola* and *C. sublineola* diverged around 15.8 MYA compared to 14.3 MYA in our study. In the acutatum complex, *C. fioriniae*, *C. nymphaeae*, *C. salicis* and *C. simmondsii* diverged around 10.9 MYA compared to 9 MYA in our study. The MRCA of the acutatum, destructivum, graminicola and spaethianum complex was estimated at 48.9 MYA compared to 38.7 MYA in our study. The MRCA of the destructivum, graminicola and spaethianum complex was estimated at 34.7 MYA compared to 29.5 MYA in our study.

Our results were therefore congruent with divergence estimates from previous studies which have used several calibration points, genome sequence data and a penalized likelihood method compared to the Bayesian analysis of five loci in our study (Hacquard et al. 2016; Liang et al. 2018; Lelwala et al. 2019). Achieving congruent results from different methodologies therefore provide reliable support for divergence time estimation in our study. The evolutionary tree was calibrated by setting the crown age of *Colletotrichum* at minimum 61 MYA based on the fossil *Protocolletotrichum deccanensis*. The estimated crown age of *Colletotrichum* in our study was 61–63 MYA which is in agreement to previous studies which were based on several calibration points (Hacquard et al. 2016; Samarakoon et al.

2019). Divergence time estimates can provide an insight in the evolutionary history of taxa, but at intraspecific divergence, it is still difficult to propose any taxonomic rearrangement due to overlap in estimated age.

Based on the resulting chronogram, the estimated crown age of *Colletotrichum* lies within the Paleocene epoch of the Paleogene period (61–63 MYA) which coincides with the period when modern angiosperm appeared (Coiffard et al. 2012; Phillips et al. 2019). The Cretaceous–Paleogene extinction event resulted in the disappearance of over 50% of plant species, thus causing a loss of plant species diversity (Wilf and Johnson 2004; McElwain and Punyasena 2007). The conditions of high humidity and reduced solar insolation after the extinction event favoured an increase of saprobic fungi which flourished on the detritus (Vajda and McLoughlin 2004). *Colletotrichum* species can occur as endophytes, pathogens and saprobes and can switch lifestyle from endophytes to pathogens as a result of host senescence or wounding (Hyde et al. 2009). For example, *C. fructicola* has been isolated as an endophyte and pathogen (Prihastuti et al. 2009a, b) with an estimated divergence time of 0.78 MYA. *Colletotrichum endophytica* has been isolated as an endophyte and saprobe (Manamgoda et al. 2013; Udayanga et al. 2013) with an estimated divergence time of 0.98 MYA. *Colletotrichum dematium* has been recorded as an endophyte, pathogen and saprobe (Damm et al. 2009) with an estimated divergence time of 1.49 MYA. This suggests that the endophytic phase is common to all species and other lifestyles could have evolved from this (Hyde et al. 2020a). The fossil *Protocolletotrichum* occurred as leaf spot disease during the Paleocene epoch of the Paleogene period which demonstrates that the pathogenic strategies of *Colletotrichum* were already well established (Kar et al. 2004).

The different complexes started to evolve at 28.2–36 MYA which lies within the Oligocene–Eocene epoch of the Paleogene period. The warm climate at the beginning of the Paleogene period favoured dense forest plants, then the humid subtropical conditions favoured an increase in the prevalence of several plants and grasses (Xing et al. 2014). As a result of evolution and dispersal of several plants including flowering plants during the Oligocene period (23–34 MYA), it can be hypothesised that *Colletotrichum* species diversified to adapt to various hosts. The Eocene epoch witnessed the appearance and diversification of several modern groups of organisms including mammals and the late Oligocene was linked to the expansion of grazing animals which could have facilitated the subsequent diversification of fungal species (Ring et al. 2020). The earliest primitive stone tools were dated at 3.3 MYA which could be attributed to hominin lineage (Lewis and Harmand 2016) and the earliest stone tools associated with *Homo* were estimated at 2.58 MYA (Braun et al. 2019). Stone tools were used for animal butchery as well as for plant processing at

2 MYA (Dominguez-Rodrigo 2009). At around 2 MYA, key adaptive behaviours were recorded in *Homo* including changes in body size and the ability of long-distance running (Bramble and Lieberman 2004). The new mobility allowed humans to explore new lands and allowed long range transport of materials and food (Potts 1991; Plummer 2004) as resources became scarce in their immediate environment. It can therefore be hypothesised that during this period the lack of food supply also led to the establishment of agricultural practices and monocropping although the earliest indication of cultivation is estimated at 23,000 years ago (Snir et al. 2015). Intensive disturbance to the environment caused by the onset of cultivation could be associated with the further diversification of *Colletotrichum* species.

The morphology of the singleton *C. axonopodi*, *C. cariniferi* and *C. parallelophorum* were compared to the graminicola and dracaenophilum complexes respectively. *Colletotrichum axonopodi* (IMI279189) was isolated from *Axonopus affinis* and it is morphologically similar to grass-associated taxa in *Colletotrichum* (Crouch et al. 2009). It is a pathogen which is characterised by falcate conidia. The presence of widely falcate conidia is a characteristic feature of the graminicola complex and most of its species are important pathogens (Damm et al. 2009). *Colletotrichum cariniferi* (MFLUCC 14-0100) was isolated as an endophyte from the stems of *Dendrobium cariniferum*. It has ellipsoidal to cylindrical conidia, with rounded ends when mature. *Colletotrichum parallelophorum* (MFLUCC 14-0083) was also isolated as an endophyte from *Dendrobium* sp. This species has smooth-walled, cylindrical conidia with rounded ends. The presence of cylindrical conidia with rounded ends is a characteristic feature of dracaenophilum complex (Damm et al. 2019). Therefore, the morphological characters of *C. axonopodi*, *C. parallelophorum* and *C. cariniferi* correspond to the graminicola and dracaenophilum complexes respectively.

The morphology of all the taxa in the new complex (*C. agaves*, *C. euphorbiae*, *C. ledebouriae*, *C. neosansevieriae* and *C. sansevieriae*) was also compared. *Colletotrichum sansevieriae* (MAFF239721) causes leaf blight of *Sansevieria trifasciata* (Nakamura et al. 2006). It forms straight, cylindrical, conidia with obtuse apex. *Colletotrichum agaves* is associated with foliar diseases of agave plants (Farr et al. 2006). This species has hyaline, cylindrical, straight or slightly curved conidia with broadly rounded apex. *Colletotrichum euphorbiae* was isolated from the leaves of *Euphorbia* sp. (Crous et al. 2013). It forms pale brown conidiogenous cells and hyaline to pale orange, cylindrical to clavate conidia with round end on one side and truncate end on the other. *Colletotrichum neosansevieriae* (CBS 139918) is associated with leaf spot disease of *Sansevieria trifasciata* (Crous et al. 2015). This species has hyaline conidiogenous cells and subcylindrical, straight to slightly curved, hyaline

conidia with obtuse apex. *Colletotrichum ledebouriae* (CBS 141284) is associated with anthracnose disease on *Ledebouria* (Crous et al. 2016). It forms hyaline to pale brown conidiogenous cells and hyaline, straight, subcylindrical conidia with obtuse apex. Therefore, the species in this complex are mostly pathogens with a wide range of hosts and they are characterised by the presence of straight or slightly curved conidia with obtuse apices. This species complex is named following the name of the species which was described first and therefore, this monophyletic group of five closely related species is named as the *Colletotrichum agaves* complex.

In this study, we compare the species boundaries in *Colletotrichum* using different genomic approaches including phylogenetic analyses, evolutionary analyses and coalescent-based methods. The coalescent and evolutionary methods provided congruent results to the phylogenetic analyses for most of the complexes. All analyses based on the ITS region, multi-locus dataset and larger dataset recovered the caudatum and graminicola complexes as one. Therefore, we recommend that these two complexes should be treated as one. Based on our findings, *C. axonopodi* is accepted in the graminicola complex whereas *C. parallelophorum* and *C. cariniferi* are accepted in the dracaenophilum complex. Based on congruent results from different molecular approaches, we accept 248 species and the vast majority are classified in 14 species complexes (40 species in acutatum complex, 25 in boninense, 17 in dematium, 17 in destructivum, 8 in dracaenophilum, 9 in gigasporum, 52 in gloeosporioides, 24 in graminicola-caudatum, 8 in magnum, 8 in orbiculare, 9 in orchidearum, 5 in sansevieriae, 9 in spathianum, 5 species in truncatum complex and 12 singleton taxa). We recommend multi-locus dataset for accurate species delineation in *Colletotrichum* as there is currently no single marker that can discriminate between species in all the complexes. Our study has identified the most informative marker for different complexes which must be included for species identification for each of the complexes. The implementation of coalescent based approaches supported our findings based on phylogenetic analyses. We therefore recommend the application of coalescent-based methodologies to delineate species in complex genus such as *Colletotrichum* as achieving congruent results across different genomic methods provides a robust framework to reliably elucidate species and species complex boundaries.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13225-021-00471-z>.

Acknowledgements We would like to thank the Thailand Research Fund, grant RDG6130001 entitled “Impact of climate change on fungal diversity and biogeography in the Greater Mekong Subregion”. Kevin D Hyde thanks Chiang Mai University for the award of a Visiting Professor. Dr Rajesh Jeewon would like to acknowledge the support from University of Mauritius and the Mauritius Research

and Innovative Council (MRC/RUN/1705). RS Jayawardena thanks the National Research Council of Thailand, grant for new researcher NRCT5-TRG630010-01, entitled “Biodiversity, taxonomy, phylogeny and evolution of *Colletotrichum* in northern Thailand”.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Arrigoni R, Berumen ML, Chen CA, Terraneo TI, Baird AH, Payri C, Benzoni F (2016) Species delimitation in the reef coral genera *Echinophyllia* and *Oxypora* (*Scleractinia*, *Lobophylliidae*) with a description of two new species. *Mol Phylogenet Evol* 105:146–159
- Barracough TG, Birky CW, Burt A (2003) Diversification in sexual and asexual organisms. *Evolution* 57:2166–2172
- Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D, De Maio N, Matschiner M (2019) BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 15:e1006650
- Bramble DM, Lieberman DE (2004) Endurance running and the evolution of *Homo*. *Nature* 432:345–352
- Braun DR, Aldeias V, Archer W, Arrowsmith JR, Baraki N, Campisano CJ, Deino AL, DiMaggio EN, Dupont-Nivet G, Engda B, Feary DA (2019) Earliest known Oldowan artifacts at >2.58 Ma from Ledi-Geraru, Ethiopia, highlight early technological diversity. *PNAS* 116:11712–11717
- Cai L, Hyde KD, Taylor PWJ, Weir B, Waller JM, Abang MM, Zang JC, Yang YL, Phouliyong S, Prihastuti ZY, 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* systematics. In: Dickman MB, Freeman S, Prusky D (eds) *Colletotrichum*: host specificity, pathology, and host pathogen interaction. APS Press, Minnesota
- Cannon PF, Damm U, Johnston PR, Weir BS (2012) *Colletotrichum* current status and future directions. *Stud Mycol* 73:181–213
- Cano J, Guarro J, Gené J (2004) Molecular and morphological identification of *Colletotrichum* species of clinical interest. *J Clin Microbiol* 42:2450–2454
- Carstens BC, Knowles LL (2007) Estimating species phylogeny from gene-tree probabilities despite incomplete lineage sorting: an example from *Melanoplus* grasshoppers. *Syst Biol* 56:400–411
- Cohen KM, Finney SC, Gibbard PL, Fan JX (2013) The ICS International Chronostratigraphic Chart. *Episodes* 36:199–204
- Coiffard C, Gomez B, Daviero-Gomez V, Dilcher DL (2012) Rise to dominance of angiosperm pioneers in European Cretaceous environments. *Proc Natl Acad Sci* 109:20955–20959
- Conrad ACI (1831) Die Pilze Deutschlands. In: Strum J (ed) *Deutschlands Flora in Abbildungen nach der Natur mit Beschreibungen*. Sturm Nürnberg 3:33–64
- Correa C, Vásquez D, Castro-Carrasco C, Zúñiga-Reinoso Á, Ortiz JC, Palma RE (2017) Species delimitation in frogs from South American temperate forests: the case of *Eupsophus*, a taxonomically complex genus with high phenotypic variation. *PLoS ONE* 12:e0181026
- Crouch JA (2014) *Colletotrichum caudatum* s.l. is a species complex. *IMA Fungus* 5:17–30
- Crouch JA, Clarke BB, White JF Jr, Hillman BI (2009) Systematic analysis of the falcate-spored graminicolous *Colletotrichum* and a description of six new species from warm-season grasses. *Mycologia* 101:717–732
- Crous PW, Wingfield MJ, Guarro J, Cheewangkoon R, van der Bank M, Swart WJ, Stchigel AM, Cano-Lira JF, Roux J, Madrid H, Damm U, Wood AR, Shuttleworth LA, Hodges CS, Munster M, de Jesús Y-M, Zúñiga-Estrada L, Cruywagen EM, de Hoog GS, Silvera C, Najafzadeh J, Davison EM, Davison PJ, Barrett MD, Barrett RL, Manamgoda DS, Minnis AM, Kleczewski NM, Flory SL, Castlebury LA, Clay K, Hyde KD, Maússe-Sitoe SN, Chen S, Lechat C, Hairaud M, Lesage-Meessen L, Pawłowska J, Wilk M, Śliwińska-Wyrzychowska A, Mętrak M, Wrzosek M, Pavlic-Zupanc D, Maleme HM, Slippers B, Mac Cormack WP, Archuby DI, Grünwald NJ, Telleria MT, Dueñas M, Martin MP, Marinowitz S, de Beer ZW, Perez CA, Gené J, Marin-Felix Y, Groenewald JZ (2013) Fungal Planet description sheets: 154–213. *Persoonia* 31:188–296
- Crous PW, Wingfield MJ, Guarro J, Hernández-Restrepo M, Sutton DA, Acharya K, Barber PA, Boekhout T, Dimitrov RA, Dueñas M, Dutta AK, Gené J, Gouliamova DE, Groenewald M, Lombard L, Morozova OV, Sarkar J, Smith MT, Stchigel AM, Wiederhold NP, Alexandrova AV, Antelmi I, Armengol J, Barnes I, Cano-Lira JF, Castañeda Ruiz RF, Contu M, Courteuisse PR, da Silveira AL, Decock CA, de Goes A, Edathodu J, Ercole E, Firmino AC, Fourie A, Fournier J, Furtado EL, Geering AD, Gershenson J, Giraldo A, Gramaje D, Hammerbacher A, He XL, Haryadi D, Khemmuks W, Kovalenko AE, Krawczynski R, Laich F, Lechat C, Lopes UP, Madrid H, Malysheva EF, Marin-Felix Y, Martin MP, Mostert L, Nigro F, Pereira OL, Picillo B, Pinho DB, Popov ES, Rodas Peláez CA, Rooney-Latham S, Sandoval-Denis M, Shivas RG, Silva V, Stoilova-Disheva MM, Telleria MT, Ullah C, Unsicker SB, van der Merwe NA, Vizzini A, Wagner HG, Wong PT, Wood AR, Groenewald JZ (2015) Fungal Planet description sheets: 320–370. *Persoonia* 34:167–266
- Crous PW, Wingfield MJ, Richardson DM, Le Roux JJ, Strasberg D, Edwards J, Roets F, Hubka V, Taylor PW, Heykoop M, Martin MP, Moreno G, Sutton DA, Wiederhold NP, Barnes CW, Carlavilla JR, Gené J, Giraldo A, Guarnaccia V, Guarro J, Hernández-Restrepo M, Kolařík M, Manjón JL, Pascoe IG, Popov ES, Sandoval-Denis M, Woudenberg JH, Acharya K, Alexandrova AV, Alvarado P, Barbosa RN, Baseia IG, Blanchette RA, Boekhout T, Burgess TI, Cano-Lira JF, Čmoková A, Dimitrov RA, Dyakov MY, Dueñas M, Dutta AK, Esteve-Raventós F, Fedosova AG, Fournier J, Gamboa P, Gouliamova DE, Grebenc T, Groenewald M, Hanse B, Hardy GE, Held BW, Jurjević Ž, Kaewgrajang T, Latha KP, Lombard L, Luangsa-Ard JJ, Lysková P, Mallátová N, Manimohan P, Miller AN, Mirabolfathy M, Morozova OV, Obodai M, Oliveira NT, Ordóñez ME, Otto EC, Paloi S, Peterson SW, Phosri C, Roux J, Salazar WA, Sánchez A, Sarria GA, Shin HD, Silva BD, Silva GA, Smith MT, Souza-Motta CM, Stchigel AM, Stoilova-Disheva MM, Sulzbacher MA, Telleria MT, Toapanta C, Traba JM, Valenzuela-Lopez N, Watling R, Groenewald JZ (2016) Fungal Planet description sheets: 400–468. *Persoonia* 36:316–458
- da Silva LL, Moreno HL, Correia HL, Santana MF, de Queiroz MV (2020) *Colletotrichum*: species complexes, lifestyle, and peculiarities of some sources of genetic variability. *Appl Microbiol Biotechnol* 104:1891–1904
- Damm U, Cannon PF, Woudenberg JH, Johnston PR, Weir BS, Tan YP, Shivas RG, Crous PW (2012a) The *Colletotrichum boninense* species complex. *Stud Mycol* 73:1–36
- Damm U, Cannon PF, Woudenberg JHC, Crous PW (2012b) The *Colletotrichum acutatum* species complex. *Stud Mycol* 73:37–113

- Damm U, O'Connell RJ, Groenewald JZ, Crous PW (2014) The *Colletotrichum destructivum* species complex - hemibiotrophic pathogens of forage and field crops. *Stud Mycol* 79:49–84
- Damm U, Sato T, Alizadeh A, Groenewald JZ, Crous PW (2019) The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. *Stud Mycol* 92:1–46
- Damm U, Woudenberg JHC, Cannon PF, Crous PW (2009) *Colletotrichum* species with curved conidia from herbaceous hosts. *Fungal Divers* 39:45–87
- Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772
- Domínguez-Rodrigo M (2009) Are all Oldowan sites palimpsests? If so, what can they tell us about Hominid carnivory? In: Hovers E, Braun DR (eds) *Interdisciplinary approaches to the Oldowan*. Springer, Amsterdam
- Doyle VP, Oudemans PV, Rehner SA, Litt A (2013) Habitat and host indicate lineage identity in *Colletotrichum gloeosporioides* s.l. from wild and agricultural landscapes in North America. *PLoS ONE* 8:e62394
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29:1969–1973
- Ezard T, Fujisawa T, Barraclough TG (2015) SPLITS: species' limits by threshold statistics. R package version 1:r45
- Farr DF, Aime MC, Rossman AY, Palm ME (2006) Species of *Colletotrichum* on *agavaceae*. *Mycol Res* 110:1395–1408
- Fujisawa T, Barraclough TG (2013) Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent approach: a revised method and evaluation on simulated data sets. *Syst Biol* 62:707–724
- Fujita MK, Leaché AD, Burbrink FT, McGuire JA, Moritz C (2012) Coalescent-based species delimitation in an integrative taxonomy. *Trends Ecol Evol* 27:480–488
- Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR, Weimann A, Damm U, Dallery JF, Hainaut M, Henrissat B (2016) Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun* 7:1–3
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Hyde KD, Cai L, McKenzie EHC, Yang YL, Zhang JZ, Prihastuti H (2009) *Colletotrichum*: a catalogue of confusion. *Fungal Divers* 39:1–17
- Hyde KD, Udayanga D, Manamgoda DS, Tedersoo L, Larsson E, Abarenkov K, Bertrand YJK, Oxelman B, Hartmann M, Kauserud H, Ryberg M, Kristiansson E, Nilsson RH (2013) Incorporating molecular data in fungal systematics: a guide for aspiring researchers. *Curr Res Environ Appl Mycol* 3:1–32
- Hyde KD, Nilsson RH, Alias SA, Ariyawansa HA, Blair JE, Cai L, de Cock AWAM, Dissanayake AJ, Glockling SL, Goonasekara ID, Gorczak M, Hahn M, Jayawardena RS, van Kan JAL, Laurence MH, Lévesque CA, Li XH, Liu JK, Maharachchikumbura SSN, Manamgoda DS, Martin FN, McKenzie EHC, McTaggart AR, Mortimer PE, Nair PVR, Pawłowska J, Rintoul TL, Shivas RG, Spies CFJ, Summerell BA, Taylor PWJ, Terhem RB, Udayanga D, Vaghefi N, Walther G, Wilk M, Wrzosek M XuJC, Yan JY, Zhou N (2014) One stop shop: backbone trees for important phytopathogenic genera. I. *Fungal Divers* 67:21–125
- Hyde KD, Tennakoon DS, Jeewon R, Bhat DJ, Maharachchikumbura SSN, Rossi W, Leonardi M, Lee HB, Mun HY, Houbraken J, Nguyen TTT, Jeon SJ, Frisvad JC, Dhanushka N, Wanasinghe DN, Luücking R, Aptroot A, Cáceres MES, Karunarathna SC, Hongsanan S, Phookamsak R, de Silva NI, Thambugala KM, Jayawardena RS, Senanayake IC, Boonmee S, Chen J, Luo ZL, Phukhamsakda C, Pereira OL, Abreu VP, Rosado AWC, Bart B, Randrianjohany E, Hofstetter V, Gibertoni TB, da Silva Soares AM, Plautz HL Jr, Sotão HMP, Xavier WKS, Bezerra JDP, de Oliveira TGL, de Souza-Motta CM, Magalhães OMC, Bundhun D, Harishchandra D, Manawasinghe IS, Dong W, Zhang SN, Bao DF, Samarakoon MC, Pem D, Karunarathna A, Lin CG, Yang J, Perera RH, Kumar V, Huang SK, Dayarathne MC, Ekanayaka AH, Jayasiri SC, Xiao YP, Konta S, Niskanen T, Liimatainen K, Dai YC, Ji XH, Tian XM, Mešić A, Singh SK, Phuthacharoen K, Cai L, Sorvongxay T, Thiagaraja V, Norphanphou C, Chaiwan N, Lu YZ, Jiang HB, Zhang JF, Abeywickrama PD, Aluthmuhandiram JVS, Brahmanage RS, Zeng M, Chethana T, Wei DP, Réblová M, Fournier J, Nekvindová J et al (2019) Fungal diversity notes 1036–1150: taxonomic and phylogenetic contributions on genera and species of fungal taxa. *Fungal Divers* 96:1–242
- Hyde KD, Jeewon R, Chen YJ, Bhunjun CS, Calabon MS, Jiang HB, Lin CG, Norphanphou C, Sysouphanthong P, Pem D, Tibpromma S, Zhang Q, Doilom M, Jayawardena RS, Liu JK, Maharachchikumbura SSN, Phukhamsakda C, Phookamsak R, Al-Sadi AM, Thongklang N, Wang Y, Gafforov Y, Jones EBG, Lumyong S (2020a) The numbers of fungi: is the descriptive curve flattening? *Fungal Divers* 103:219–271
- Hyde KD, Norphanphou C, Maharachchikumbura SSN, Bao DF, Bhat DJ, Boonmee S, Bundhun D, Calabon MS, Chaiwan N, Chen YJ, Chethana KWT, Dai DQ, Dayarathne MC, Devadatha B, Dissanayake AJ, Dissanayake LS, Doilom M, Dong W, Fan XL, Goonasekara ID, Hongsanan S, Huang SK, Jayawardena RS, Jeewon R, Jones EBG, Karunarathna A, Konta S, Kumar V, Lin CG, Liu JK, Liu N, Lu YZ, Luangsa-ard J, Lumyong S, Luo ZL, Marasinghe DS, McKenzie EHC, Niego AGT, Niranjana M, Perera RH, Phukhamsakda C, Rathnayaka AR, Samarakoon MC, Samarakoon SMBC, Sarma VV, Senanayake IC, Shang QJ, Stadler M, Tibpromma S, Wanasinghe DN, Wei DP, Wijayawardena NN, Xiao YP, Xiang MM, Yang J, Zeng XY, Zhang SN (2020b) Refined families of *Sordariomycetes*. *Mycosphere* 11:305–1059
- Index Fungorum (2020) <http://www.indexfungorum.org/names/names.asp>
- Jayawardena RS, Hyde KD, Damm U, Cai L, Liu M, Li XH, Zhang W, Zhao WS, Yan JY (2016a) Notes on currently accepted species of *Colletotrichum*. *Mycosphere* 7:1192–1260
- Jayawardena RS, Hyde KD, Jeewon R, Liu XH, Liu M, Yan JY (2016b) Why it is important to correctly name *Colletotrichum* species? *Mycosphere* 7:1076–1092
- Jayawardena RS, Hyde KD, Chen YJ, Papp V, Palla B, Papp D, Bhunjun CS, Hurdeal VG, Senwana C, Manawasinghe IS, Harischandra DL, Gautam AK, Avasthi S, Chuankid B, Goonasekara ID, Hongsanan S, Zeng XY, Liyanage KK, Liu N, Karunarathna A, Hapuarachchi KK, Luangharn T, Raspé O, Brahmanage R, Doilom M, Lee HB, Mei L, Jeewon R, Huanraluek N, Chaiwan N, Stadler M, Wang Y (2020) One stop shop IV: taxonomic update with molecular phylogeny for important phytopathogenic genera: 76–100 (2020). *Fungal Divers* 103:87–218
- Kapli P, Lutteropp S, Zhang J, Kobert K, Pavlidis P, Stamatakis A, Flouri T (2017) Multi-rate Poisson tree processes for single-locus species delimitation under maximum likelihood and Markov chain Monte Carlo. *Bioinformatics* 33:1630–1638
- Kar RK, Sharma N, Verma UK (2004) Plant pathogen *Protocolletotrichum* from a Deccan intertrappean bed (Maastrichtian), India. *Cretac Res* 25:945–950
- Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160–1166
- Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30:3276–3278
- Lelwala RV, Korhonen PK, Young ND, Scott JB, Ades PK, Gasser RB, Taylor PW (2019) Comparative genome analysis indicates high

- evolutionary potential of pathogenicity genes in *Colletotrichum tanacetii*. PLoS ONE 14:e0212248
- Lewis JE, Harmand S (2016) An earlier origin for stone tool making: implications for cognitive evolution and the transition to *Homo*. Philos Trans R Soc Lond B Biol Sci 371:20150233
- Liang X, Wang B, Dong Q, Li L, Rollins JA, Zhang R, Sun G (2018) Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. PLoS ONE 13:e0196303
- Liu F, Weir BS, Damm U, Crous PW, Wang Y, Liu B, Wang M, Zhang M, Cai L (2015) Unravelling *Colletotrichum* species associated with *Camellia*: employing ApMat and GS loci to resolve species in the *C. gloeosporioides* complex. Persoonia 35:63–86
- Liu F, Wang M, Damm U, Crous PW, Cai L (2016) Species boundaries in plant pathogenic fungi: a *Colletotrichum* case study. BMC Evol Biol 16:81
- Lücking R, Aime MC, Robbertse B, Miller AN, Ariyawansa HA, Aoki T, Cardinali G, Crous PW, Druzhinina IS, Geiser DM, Hawksworth DL, Hyde KD, Irinyi L, Jeewon R, Johnston PR, Kirk PM, Malosso E, May TW, Meyer W, Öpik M, Robert V, Stadler M, Thines M, Vu D, Yurkov AM, Zhang N, Schoch CL (2020) Unambiguous identification of fungi: where do we stand and how accurate and precise is fungal DNA barcoding? IMA Fungus 11:1–32
- Ma X, Nontachaiyapoom S, Jayawardena RS, Hyde KD, Gentekaki E, Zhou S, Qian Y, Wen T, Kang J (2018) Endophytic *Colletotrichum* species from *Dendrobium* spp. in China and Northern Thailand. MycoKeys 43:23–57
- Maharachchikumbura SSN, Hyde KD, Jones EBG, McKenzie EHC, Bhat JD, Dayarathne MC, Huang SK, Norphanphoun C, Senanayake IC, Perera RH, Shang QJ, Xiao Y, D'souza MJ, Hongsanan S, Jayawardena RS, Daranagama DA, Konta S, Goonasekara ID, Zhuang WY, Jeewon R, Phillips AJL, Abdel-Wahab MA, Al-Sadi AM, Bahkali AH, Boonmee S, Boonyuen N, Cheewangkoon R, Dissanayake AJ, Kang J, Li QR, Liu JK, Liu XZ, Liu ZY, Luangsa-ard JJ, Pang KL, Phookamsak R, Promputtha I, Suetrong S, Stadler M, Wen T, Wijayawardene NN (2016) Families of Sordariomycetes. Fungal Divers 79:1–317
- Manamgoda DS, Udayanga D, Cai L, Chukeatirote E, Hyde KD (2013) Endophytic *Colletotrichum* from tropical grasses with a new species *C. endophytica*. Fungal Divers 61:107–115
- Marin-Felix Y, Groenewald JZ, Cai L, Chen Q, Marincowitz S, Barnes I, Bensch K, Braun U, Camporesi E, Damm U, de Beer ZW, Dissanayake A, Edwards J, Giraldo A, Hernández-Restrepo M, Hyde KD, Jayawardena RS, Lombard L, Luangsa-ard J, McTaggart AR, Rossman AY, Sandoval-Denis M, Shen M, Shivas RG, Tan YP, van der Linde EJ, Wingfield MJ, Wood AR, Zhang JQ, Zhang Y, Crous PW (2017) Genera of phytopathogenic fungi: GOPHY 1. Stud Mycol 86:99–216
- Marki PZ, Fjeldså J, Irestedt M, Jönsson KA (2018) Molecular phylogenetics and species limits in a cryptically coloured radiation of Australo-Papuan passerine birds (*Pachycephalidae*: *Colluricincla*). Mol Phylogenet Evol 124:100–105
- McElwain JC, Punyasena SW (2007) Mass extinction events and the plant fossil record. Trends Ecol Evol 22:548–557
- Miller MA, Pfeiffer W, Schwartz T (2017) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In 2010 gateway computing environments workshop (GCE):1–8
- Miralles A, Vences M (2013) New metrics for comparison of taxonomies reveal striking discrepancies among species delimitation methods in *Madascincus* lizards. PLoS ONE 8:e68242
- Monaghan MT, Wild R, Elliot M, Fujisawa T, Balke M, Inward DJ, Lees DC, Ranaivosolo R, Eggleton P, Barraclough TG, Vogler AP (2009) Accelerated species inventory on Madagascar using coalescent-based models of species delineation. Syst Biol 58:298–311
- Moraga J, Gomes W, Pinedo C, Cantoral JM, Hanso JR, Carbún M, Garriso C, Durán-Patrón R, Collado IG (2019) The current status on secondary metabolites produced by plant pathogenic *Colletotrichum* species. Phytochem Rev 18:215–239
- Nakamura M, Ohzono M, Iwai H, Arai K (2006) Anthracnose of *Sansevieria trifasciata* caused by *Colletotrichum sansevieriae* sp. nov. J Gen Plant Pathol 72:253–256
- Nieto-Montes de Oca A, Barley AJ, Meza-Lázaro RN, GarcíaVázquez UO, Zamora-Abrego JG, Thomson RC, Leaché AD (2017) Phylogenomics and species delimitation in the knob-scaled lizards of the genus *Xenosaurus* (*Squamata*: *Xenosauridae*) using ddRAD-seq data reveal a substantial underestimation of diversity. Mol Phylogenet Evol 106:241–253
- Nilsson RH, Hyde KD, Pawłowska J, Ryberg M, Tedersoo L, Aas AB, Alias SA, Alves A, Anderson CL, Antonelli A, Arnold AE, Bahmann B, Bahram M, Bengtsson-Palme J, Berlin A, Branco S, Chomnunti P, Dissanayake A, Drenkhan R, Friberg H, Frøslev TG, Halwachs B, Hartmann M, Henricot B, Jayawardena R, Jumpponen A, Kausserud H, Koskela S, Kulik T, Liimatainen K, Lindahl BD, Lindner D, Liu JK, Maharachchikumbura S, Manamgoda D, Martinsson S, Neves MA, Niskanen T, Nyländer S, Pereira OL, Pinho DB, Porter TM, Queloz V, Riit T, Sánchez-García M, de Sousa F, Stefańczyk E, Tadych M, Takamatsu S, Tian Q, Udayanga D, Unterseher M, Wang Z, Wikee S, Yan J, Larsson E, Larsson KH, Köljalg U, Abarenkov K (2014) Improving ITS sequence data for identification of plant pathogenic fungi. Fungal Divers 67:11–19
- Parnmen S, Rangsiruji A, Mongkolsuk P, Boonpragob K, Nutakki A, Lumbsch HT (2012) Using Phylogenetic and coalescent methods to understand the species diversity in the *Cladia aggregata* complex (*Ascomycota*, *Lecanorales*). PLoS ONE 7:e52245
- Pentinsaari M, Vos R, Mutanen M (2016) Algorithmic single-locus species delimitation: effects of sampling effort, variation and non-monophyly in four methods and 1870 species of beetles. Mol Ecol Resour 17:393–404
- Phillips AJ, Hyde KD, Alves A, Liu JK (2019) Families in *Botryosphaeriales*: a phylogenetic, morphological and evolutionary perspective. Fungal Divers 94:1–22
- 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
- Plummer TW (2004) Flaked stones and old bones: biological and cultural evolution at the dawn of technology. Am J Phys Anthropol 125:118–164
- Potts R (1991) Why the oldowan? Plio-Pleistocene toolmaking and the transport of resources. J Anthropol Res 47:153–176
- Prihastuti H, Cai L, Chen H, McKenzie EHC, Hyde KD (2009b) Characterization of *Colletotrichum* species associated with coffee berries in Chiang Mai, Thailand. Fungal Divers 39:89–109
- R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna
- Rambaut A (2014) FigTree v1.4. Institute of Evolutionary Biology, University of Edinburgh, Edinburgh <http://tree.bio.ed.ac.uk/software/figtree/>
- Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 67:901
- Rashmi M, Kushveer JS, Sarma VV (2019) A worldwide list of endophytic fungi with notes on ecology and diversity. Mycosphere 10:798–1079
- Renner MA, Heslewood MM, Patzak SD, Schäfer-Verwimp A, Heinrichs J (2017) By how much do we underestimate species diversity of liverworts using morphological evidence? An example from Australasian *Plagiochila* (*Plagiochilaceae*: *Jungermannioipsida*). Mol Phylogenet Evol 107:576–593
- Ring SJ, Bocherens H, Wings O, Rabi M (2020) Divergent mammalian body size in a stable Eocene greenhouse climate. Sci Rep 10:3987

- Rojas EI, Rehner SA, Samuels GJ, Van Bael SA, Herre EA, Cannon P, Chen R, Pang J, Wang R, Zhang Y, Peng YQ, Sha T (2010) *Colletotrichum gloeosporioides* s.l. associated with *Theobroma cacao* and other plants in Panama: multilocus phylogenies distinguish pathogen and endophyte clades. *Mycologia* 102:1318–1338
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Samarakoon MC, Hyde KD, Hongsanan S, McKenzie EH, Ariyawansa HA, Promputtha I, Zeng XY, Tian Q, Liu JKJ (2019) Divergence time calibrations for ancient lineages of *Ascomycota* classification based on a modern review of estimations. *Fungal Divers* 96:285–346
- Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS* 109:6241–6246
- Servis JA, Reid BN, Timmers MA, Stergioula V, Naro-Maciel E (2020) Characterizing coral reef biodiversity: genetic species delimitation in brachyuran crabs of Palmyra Atoll, Central Pacific. *Mitochondrial DNA Part A* 25:1–2
- Sharma G, Pinnaka AK, Shenoy BD (2015) Resolving the *Colletotrichum siamense* species complex using *ApMat* marker. *Fungal Divers* 71:247–264
- Shivaprakash MR, Appannanavar SB, Dhaliwal M, Gupta A, Gupta S, Gupta A, Chakrabarti A (2011) *Colletotrichum truncatum*: an unusual pathogen causing mycotic keratitis and endophthalmitis. *J Clin Microbiol* 49:2894–2898
- Snir A, Nadel D, Groman-Yaroslavski I, Melamed Y, Sternberg M, Bar-Yosef O, Weiss E (2015) The origin of cultivation and proto-weeds, long before Neolithic farming. *PLoS ONE* 10:e0131422
- Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Swofford DL (2002) PAUP: phylogenetic analysis using parsimony, version 4.0 b10. Sinauer Associates, Sunderland
- Tao G, Liu ZY, Liu F, Gao YH, Cai L (2013) Endophytic *Colletotrichum* species from *Bletilla ochracea* (*Orchidaceae*), with description of seven new species. *Fungal Divers* 61:139–164
- Udayanga D, Manamgoda DS, Liu X-Z, Chukeatirote E, Hyde KD (2013) What are the common anthracnose pathogens of tropical fruits? *Fungal Divers* 61:165–179
- Vajda V, McLoughlin S (2004) Fungal proliferation at the Cretaceous-Tertiary boundary. *Science* 303:1489
- Weir B, Johnston PR, Damm U (2012) The *Colletotrichum gloeosporioides* species complex. *Stud Mycol* 73:115–180
- Wilf P, Johnson KR (2004) Land plant extinction at the end of the Cretaceous: a quantitative analysis of the North Dakota megafossil record. *Paleobiology* 30:347–368
- Xing Y, Onstein RE, Carter RJ, Stadler T, Peter Linder H (2014) Fossils and a large molecular phylogeny show that the evolution of species richness, generic diversity, and turnover rates are disconnected. *Evolution* 68:2821–2832
- Zhang J, Kapli P, Pavlidis P, Stamatakis A (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* 29:2869–2876