Mating-type locus rearrangement leads to shift from homothallism to heterothallism in

Citrus-associated Phyllosticta species

Desirrê Alexia Lourenço Petters-Vandresen^{a,c}, Bruno Janoski Rossi^a, Johannes Z.

Groenewald^e, Pedro W. Crous^e, Marcos Antonio Machado^b, Eva H. Stukenbrock^{c,d}, Chirlei

Glienke^a

^aLaboratório de Bioprospecção e Genética Molecular de Microrganismos, Postgraduate Program in Genetics. Department of Genetics, Federal University of Paraná (UFPR), Centro Politécnico, Jardim das Américas, 81531-990, Curitiba, Paraná State, Brazil.

^bCentro de Citricultura APTA Citros Sylvio Moreira – IAC, Campinas-SP, Brazil

^cEnvironmental Genomics, Max Planck Institute for Evolutionary Biology, Plön, Germany

^dEnvironmental Genomics, Christian-Albrechts University of Kiel, Kiel, Germany

^eWesterdijk Fungal Biodiversity Institute, Utrecht, Netherlands

Corresponding author: Chirlei Glienke

Corresponding author e-mail: ch.glienke@gmail.com

Corresponding author postal address: Laboratório de Bioprospecção e Genética Molecular de Microrganismos, Postgraduate Program in Genetics. Department of Genetics, Federal University of Paraná (UFPR), Centro Politécnico, Jardim das Américas, 81531-990, Curitiba, Paraná State, Brazil.

Declarations of interest: none

26 Abstract

27

28 Currently, eight *Phyllosticta* species are known to be associated with *Citrus* hosts,
 29 incorporating endophytic and pathogenic lifestyles. As sexual reproduction is a key factor
 30 involved in host-interaction, it could be related to the differences in lifestyle. To evaluate this
 31 hypothesis, we characterized the mating-type loci of six *Citrus*-associated *Phyllosticta* species
 32 from whole genome assemblies. Mating-type genes are highly variable in their sequence
 33 content, but the genomic locations and organization of the mating-type loci are conserved.
 34 *Phyllosticta citriasiana*, *P. citribraziliensis* and *P. paracitricarpa* are heterothallic, and *P.*
 35 *citrichinaensis* was confirmed to be homothallic. In addition, the *P. citrichinaensis* *MAT1-2*
 36 idiomorph occurs in a separate location from the mating-type locus. Ancestral state
 37 reconstruction suggests that homothallism is the ancestral thallism state in *Phyllosticta*, with a
 38 shift to heterothallism in *Phyllosticta* species that are pathogenic to *Citrus*. Moreover, the
 39 homothallic strategies of *P. capitalensis* and *P. citrichinaensis* result from independent
 40 evolutionary events. As the pathogenic species *P. citriasiana*, *P. citricarpa* and *P.*
 41 *paracitricarpa* are heterothallic and incapable of selfing, disease management practices
 42 focused in preventing the occurrence of sexual reproduction could assist in the control of
 43 Citrus Black Spot and Citrus Tan Spot diseases. This study emphasizes the importance of
 44 studying *Citrus-Phyllosticta* interactions under evolutionary and genomic perspectives, as
 45 these approaches can provide valuable information about the association between *Phyllosticta*
 46 species and their hosts, and also serve as guidance for the improvement of disease
 47 management practices.

48

49 **Keywords:** evolutionary transition, fungal pathogen, gene loss, transposable element, sexual
 50 reproduction

51

52 **1 Introduction**

53

54 Sexual reproduction plays a key role in the evolution of fungi and other eukaryotic
 55 organisms, by increasing genetic variation, purging deleterious mutations, and mediating the
 56 production of resistant sexual spores (Lee et al., 2010; Ni et al., 2011). The core features of
 57 sexual reproduction, such as ploidy changes, production of gametes via meiosis, cell-cell
 58 recognition and fusion, are conserved along all eukaryotic organisms (Heitman et al., 2013).
 59 Nevertheless, fungal species present diverse reproduction strategies to complete the sexual
 60 cycle (Ni et al., 2011). In some species, including *Coprinopsis cinerea* and *Schizophyllum*
 61 *commune*, sexual reproduction is regulated by multiple mating-types (Casselton and Kües,
 62 2007); at the other extreme some species are self-fertile, such as *Aspergillus nidulans*
 63 (Kronstad, 2007).

64 Sexual reproduction in fungi can be classified in two main strategies, namely
 65 heterothallism and homothallism. In Pezizomycotina, the differences between heterothallic
 66 and homothallic strategies arise from a bipolar mating-system (Ni et al., 2011), in which a
 67 single locus (*MAT1*) regulates mating processes in strains of the same species. Instead of
 68 alleles, the *MAT1* locus has two highly divergent versions (*MAT1-1* and *MAT1-2* idiomorphs)
 69 comprised of non-orthologous genes, and both are required for the successful completion of
 70 sexual reproduction (Wilken et al., 2017). In heterothallic species, individuals harbour only
 71 one *MAT1* idiomorph in their genomes, never both (Wilken et al., 2017). Thus, two
 72 individuals of opposite and compatible mating-types are required for sexual reproduction. On
 73 the other hand, the homothallic species possess both *MAT1* idiomorphs in their genomes.
 74 Therefore, each individual can reproduce with any other individual from the same species,
 75 including itself, because individuals are self-fertile (Nagel et al., 2018).

76 The *MAT1-1* idiomorph is characterized by a gene (*MAT1-1-1*) encoding a protein
 77 with an alpha-box domain (CDD: cl07856, pfam04769), while the *MAT1-2* idiomorph
 78 comprises a gene (*MAT1-2-1*) encoding a protein with a high mobility group domain (HMG
 79 box, CDD: cd01389) (Bihon et al., 2014). These mating-type genes are key components in
 80 sexual mating, and are transcription factors that control the sexual development and the
 81 downstream-regulation of the expression of mating-type specific genes (Kim et al., 2015). In
 82 addition to *MAT1-1-1* and *MAT1-2-1*, additional genes may exist in each idiomorph, and their
 83 nomenclature follows the system as originally described (Turgeon and Yoder, 2000; Wilken
 84 et al., 2017).

85 Even though the *MAT1* idiomorphs are not orthologous and show a high degree of
 86 variation among species, alignments of the sequences flanking the mating-type locus exhibit a
 87 high level of similarity between related species (Nagel et al., 2018). Sequence comparisons
 88 among Ascomycete species revealed that *MAT1-1* and *MAT1-2* flanking regions mostly
 89 encode the genes *SLA2*, *APN2*/DNA lyase, as well as *APC5* and *COX13* (Wilken et al., 2017).
 90 These genes show a higher degree of sequence conservation than *MAT1* idiomorphs. The
 91 presence of these flanking genes along several ascomycete species suggests a high degree of
 92 positional conservation for the mating-type locus in the genomes (Wilken et al., 2017).

93 The identification and characterization of the *MAT1* locus and adjacent regions is the
 94 first step to study and investigate the reproductive strategies of Ascomycete species (Wilken
 95 et al., 2017). Traditionally this has been done using molecular approaches, such as cloning
 96 and hybridization with specific probes (Glass et al., 1988), PCR with degenerate primers
 97 (Kerényi et al., 2004) and gene-walking approaches (Groenewald et al., 2006). Recently, with
 98 the advance of whole genome sequencing, BLAST analyses combined with gene prediction
 99 tools have been widely employed to identify and characterize the *MAT1* locus in different
 100 fungal species (Amorim et al., 2017; Bihon et al., 2014; Guarnaccia et al., 2019; Nagel et al.,

101 2018; Wang et al., 2016). For species in which the sexual structures have been described
 102 (Robinson and Natvig, 2019), characterization of the *MAT1* locus helps to determine if the
 103 species is homo- or heterothallic, information that is essential to design crossing experiments.
 104 Likewise, in species for which sexual structures are not described and for which experimental
 105 crossing is not possible, analyses of the *MAT* locus have been important understand the
 106 species' life cycle and sexual reproduction strategies (Linde et al., 2003; Waalwijk et al.,
 107 2002).

108 Insight into reproductive strategies of fungi have proven very important in population
 109 genetics and evolutionary studies of fungal pathogens. Sexual reproduction can play a key
 110 role in pathogen epidemiology, and determine the rate of evolution of important traits
 111 including fungicide resistance (Lopes et al., 2017; Meng et al., 2015). Although sexual
 112 structures are unknown for several species, population genetic analyses have provided
 113 evidence for the occurrence of sexual recombination in important plant pathogens and thereby
 114 provided essential knowledge regarding their epidemiology and evolutionary potential
 115 (Groenewald et al., 2006; McDonald and Linde, 2002; McDonald and McDermott, 1993)

116 Sexual reproduction has been studied in the heterothallic citrus pathogen *Phyllosticta*
 117 *citricarpa*, which causes the Citrus Black Spot (CBS) disease in sweet oranges and lemons.
 118 Ascospores of this fungus are responsible for infecting fruits and leaves and for long-distance
 119 dispersal of the pathogen (Kotzé, 1981). However, until recently, studies failed to induce
 120 sexual reproduction *in vitro* (Amorim et al., 2017; Baldassari et al., 2008; Wang et al., 2016).
 121 Successful production of ascospores under laboratory conditions was only possible after
 122 detailed analyses of the *MAT1* locus in *P. citricarpa*, which revealed the heterothallic nature
 123 of this species. Subsequently, PCR markers were developed to discriminate between *MAT1-1*
 124 and *MAT1-2* isolates (Amorim et al., 2017; Wang et al., 2016). Hereby, compatible isolates

125 could be identified for *in vitro* crossings experiments and ascospore production (Tran et al.,
126 2017).

127 Recent studies of the mating-type locus in *P. citricarpa* and a sister non-pathogenic
128 species *P. capitalensis* (Wang et al., 2016) provided the foundation for a first comparative
129 analysis of reproductive strategies in *Citrus*-associated *Phyllosticta* species. Although these
130 two species are closely related, they have distinct mating strategies. While *P. citricarpa* is
131 heterothallic, *P. capitalensis* is homothallic including both mating idiomorphs in the same
132 isolate.

133 Genome sequences of closely related *Phyllosticta* species are a valuable resource for
134 studying mating strategy evolution among species. To date there are eight *Phyllosticta* species
135 known to be associated with *Citrus*, either as pathogens, e.g. *P. citricarpa*, *P. paracitricarpa*,
136 *P. citrimaxima*, *P. citriasiana* and *P. citrichinaensis*, or as endophytes, e.g. *P.*
137 *citribraziliensis*, *P. capitalensis* and *P. paracapitalensis* (Glienke et al., 2011; Guarnaccia et
138 al., 2017; Wang et al., 2012; Wikee et al., 2013; Wulandari et al., 2009). The genomes of six
139 of these species have been sequenced (Guarnaccia et al., 2019; Rodrigues et al., 2019; Wang
140 et al., 2016). Comparative genome analyses combined with evidence for ascospore
141 production in pure cultures (Glienke et al., 2011; Guarnaccia et al., 2017; Wang et al., 2012)
142 provided evidence that *P. capitalensis*, *P. citrichinaensis* and *P. paracapitalensis* are
143 homothallic, while *P. citricarpa*, *P. citriasiana*, *P. citribraziliensis* are heterothallic (Amorim
144 et al., 2017; Guarnaccia et al., 2019; Wang et al., 2016).

145 This diversity in reproduction strategies, host-pathogen interactions, and lifestyles,
146 makes the *Citrus*-associated *Phyllosticta* species a valuable model system to study the
147 evolution of sexual mating strategies in fungal pathogens. Hereby, the main goal of this study
148 was to elucidate the evolution of *Citrus*-associated *Phyllosticta* species' mating strategies
149 using a detailed characterization of the *MAT* loci. We present thorough analyses of gene order

150 and sequence conservation along the mating-type locus for these *Phyllosticta* species, and
 151 make inferences about past shifts from homo- to heterothallism in the genus. We also discuss
 152 how sexual reproduction may have influenced the biology and evolution of *Citrus*-associated
 153 *Phyllosticta* species, and provide a foundation for further investigation of the evolutionary
 154 processes involved in these species' diverse associations with hosts.

155

156 **2 Materials and Methods**

157

158 *2.1 Genome data*

159

160 Genome data used in the analyses were obtained from public assemblies available
 161 from GenBank (Wang et al., 2020; Wang et al., 2016) and the JGI MycoCosm portal
 162 (Guarnaccia et al., 2019). We also conducted *de novo* assemblies of both unpublished and
 163 already published data (Table 1). The published genomes of *P. capitalensis* LGMF01 and *P.*
 164 *citricarpa* LGMF06 are highly fragmented (Rodrigues et al., 2019), with the full mating-type
 165 locus not assembled. Therefore, we generated *de novo* genome assemblies for these strains
 166 using published Illumina mate-pair read data.

167 For *P. citribraziliensis* LGMF08, for which a public assembly was not available,
 168 genomic DNA was extracted from seven-day-old cultures grown in liquid potato-dextrose
 169 medium (PD) at 200 rpm and 28 °C, employing the CTAB method (Doyle and Dickson,
 170 1987). Libraries were prepared with the Illumina Nextera DNA Sample Preparation Kit.
 171 Sequencing was performed using the Illumina HiSeq2500 platform, producing paired-end
 172 reads (250 bp), at the Laboratório Central de Tecnologias de Alto Desempenho em Ciências
 173 da Vida (LaCTAD), Campinas, Brazil.

174 For *P. capitalensis* LGMF01 and *P. citricarpa* LGMF06, libraries of the paired-end
 175 reads were processed with NxTrim (O’Connell et al., 2015) to remove Nextera adapters and
 176 generate mate-pair, paired-end, single-end and unknown libraries. The script
 177 “deinterleave_fastq.sh” (<https://gist.github.com/nathanhaigh/3521724>) was used to separate
 178 the reads from the mate-pair and paired-end libraries in “forward” and “reverse” files. For
 179 quality filtering, Trimmomatic v 0.38 (Bolger et al., 2014) was used to 1) trim bases at the
 180 start and end of the reads below a quality threshold of 25, 2) trim low quality segments using
 181 a 4 bp sliding window and a quality threshold of 15, and 3) discard reads shorter than 50 bp.
 182 For *P. citribraziliensis* LGMF08, the reads were filtered in Trimmomatic to remove Illumina
 183 adapters and trim for quality using the same parameters described before but discarding reads
 184 shorter than 90 bp.

185 FastQC v 0.11.8 (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) was
 186 employed to check the quality of reads after processing in Trimmomatic. *De novo* genome
 187 assemblies were generated with SPAdes v 3.13 (Bankevich et al., 2012), using default
 188 parameters. The final assemblies were evaluated with QUAST v 4.6.3 (Gurevich et al., 2013).
 189 Library and assembly statistics for the new assemblies are summarized in Table S1 and
 190 assemblies are available at Zenodo (<https://doi.org/10.5281/zenodo.3750350>).

191 For a better understanding on how the genomes compare to each other, and to ensure
 192 quality and consistency for the downstream analyses, we assessed the repeat content and
 193 genome completeness levels of all genomes evaluated in this study. Repetitive regions were
 194 annotated with the REPET package v 2.5 (Flutre et al., 2011; Quesneville et al., 2005). First,
 195 we identified repetitive elements using TEdenovo with default parameters, which generates a
 196 library of classified and non-redundant consensus repeat sequences. Then, this library was
 197 used for the annotation of the genomes with TEannot. Genome completeness was assessed

198 with BUSCO v 3 (Simão et al., 2015) using the Pezizomycotina lineage dataset (Waterhouse
199 et al., 2018).

200

201 2.2 Gene annotation, *MAT* locus identification and characterization of mating-type and 202 adjacent genes

203

204 Protein-coding genes were identified in *Phyllosticta* genome assemblies employing a
205 similar strategy to the one used in the Fungal Genome Annotation pipeline (Haas et al., 2011).
206 Briefly, GeneMark-ES v 4.46 (Ter-Hovhannisyan et al., 2008) was run in the self-training
207 mode to provide an initial set of *ab initio* predictions. Then, to obtain a set of RNA-Seq-based
208 predictions, RNA-Seq reads from the ex-type strains of the six *Phyllosticta* species analysed
209 (Guarnaccia et al., 2019) were mapped to the genomes of the strains from the same species
210 with HISAT2 v 2.1.0 (Kim et al., 2019). The mapping results were submitted to BRAKER1
211 (Hoff et al., 2016), which combines GeneMark-ET v 4.46 (Lomsadze et al., 2014) and
212 AUGUSTUS v 3.3 (Stanke et al., 2008). In addition, genome-guided transcript reconstruction
213 using Trinity v 2.8.4 (Grabherr et al., 2013; Haas et al., 2013) was performed. The generated
214 transcripts were used by PASA v 2.4.1 to obtain a set of transcriptome-based gene models.
215 Then, the EVidenceModeler software v 1.1.1 (Haas et al., 2008) was applied to generate the
216 final weighted consensus set of gene structures, combining the *ab initio* and the transcript-
217 based predictions. Gene predictions are available at Zenodo
218 (<https://doi.org/10.5281/zenodo.3750350>).

219 The *MAT* locus was identified in each genome by a local tBLASTx search with
220 BLAST v 2.2.26 applying the *Phyllosticta citricarpa* *MAT* idiomorphs and *APN2* genes
221 (Amorim et al., 2017) as queries. The scaffold and contigs that contained sequences with high
222 levels of similarity to the query sequences were extracted from the genomes, and the

corresponding nucleotide and amino acid sequences of each gene were extracted based on the previously generated gene predictions. To confirm the identities of the predicted genes, the translated amino acid sequences were queried against the NCBI non-redundant (nr) database with BLASTp. After confirmation, the gene annotations were further refined by comparison with the *MAT* loci from *Phyllosticta citricarpa* and some *Botryosphaeriaceae* species (Nagel et al., 2018). The sequences and annotations are deposited in Zenodo (<https://doi.org/10.5281/zenodo.3750350>). Functional domains in the annotated genes were determined using the Conserved Domain Search Service (CD-Search) from NCBI (Lu et al., 2020), using an e-value of 0.01 as threshold.

2.3 Characterization and comparison of mating-type loci organization

Mating-type and flanking gene sequence similarities among *Phyllosticta* species were evaluated with a pairwise BLASTn analysis with a maximum e-value cut-off of 0.0001. The regions were then used for a synteny analysis in EasyFig v 2.2.3 (Sullivan et al., 2011). Intergenic regions were evaluated for the presence of partial *MAT1-1-1*, *MAT1-1-8*, *MAT1-2-1* and *MAT1-2-5* sequences. Finally, we assessed the presence of partial fragments of the mating-type locus in the rest of the genome.

2.4 Primer design, PCR and sequencing

In addition to the primers previously designed to amplify fragments from the *MAT1-1-1* gene of *Phyllosticta* species (Guarnaccia et al., 2019), new primers were designed to confirm the presence of the LTR-*Copia* element in the *P. citrichinaensis* mating-type locus.

247 Primers were designed based on genomic sequences from strain CBS 130529 and
248 targeted the LTR-*Copia* element as well as the flanking gene *PIM*. The primers were tested
249 for performance characteristics such as internal structures, hairpins, and self- and hetero-
250 dimer formation using OligoIDTAAnalyzer v 3.1 (<https://eu.idtdna.com/calc/analyzer/>)
251 (Owczarzy et al., 2008). Melting temperatures were calculated with the ThermoFisher Tm
252 Calculator, to ensure primer compatibility. Primer sequences, annealing positions and
253 amplicon sizes are summarized in Supplementary Table S2.

254 A DNA sample from *P. citrichinaensis* for PCR amplification was extracted with
255 QIAGEN Genomic-tip 100/G (QIAGEN Benelux B.V., Venlo, Netherlands). PCR
256 amplifications were performed with Phusion High Fidelity PCR Master Mix (Invitrogen,
257 Carlsbad CA, USA). PCR conditions included an initial denaturation step of 98 °C for 30 s, 35
258 cycles of 98 °C for 10 s, annealing at 54 °C for 20 s, extension times of ~30 s per kilobase of
259 DNA, and a final extension step of 72 °C for 10 min. Amplicons were visualized on 0.8%
260 agarose gel electrophoresis using 1× TAE buffer stained with SYBR Safe (Invitrogen,
261 Carlsbad CA, USA), and compared to a GeneRuler 1 Kb Ladder (ThermoScientific, Waltham
262 MA, USA) for evaluation of size, quality and concentration. For sequencing, amplicons were
263 purified using illustra ExoProStar (GE Healthcare, Chicago IL, USA) and sequenced using
264 the BigDye Terminator Cycle Sequencing Kit v 3.1 (Applied Biosystems, Foster City CA,
265 USA). DNA sequences were obtained on an ABI 3130xl DNA sequencer (Applied
266 Biosystems, Foster City CA, USA) and the electropherograms were examined and manually
267 corrected when necessary using MEGA7 software (Kumar et al., 2016).

268

269 2.5 Phylogenetic analyses and ancestral state reconstruction

270

271 To provide a framework of species relationships for the synteny comparisons, we
 272 performed a multilocus analysis of the *ITS*, *TEF1*, *ACT* and *GAPDH* loci of *Citrus*-associated
 273 *Phyllosticta* species. Next, we investigated how the evolutionary history of mating-type and
 274 flanking genes relate to the evolutionary history of *Citrus*-associated species. To this end, we
 275 performed phylogenetic analyses of coding regions and amino acid sequences of *MAT1-1-1*,
 276 *MAT1-1-8*, *MAT1-2-1* and *MAT1-2-5*, as well as sequences from the flanking genes
 277 (*SAICARsyn*, *MCP*, *CIA30*, *CoxVIa*, *APN2*, *PIM* and *40S S9*), using *Diplodia seriata* as
 278 outgroup. Finally, to provide a backbone for an ancestral state reconstruction of thallism, we
 279 performed a new multilocus analysis of *ITS*, *TEF1*, *ACT* and *GAPDH*, with eight additional
 280 representative species of the genus *Phyllosticta* and using *Diplodia corticola* and
 281 *Neofusicoccum parvum* as outgroups. Accession codes of all sequences used in the
 282 phylogenetic analyses are summarized in Table S3.

283 Alignments were produced in MAFFT v 7 (Katoh and Standley, 2013), and manually
 284 adjusted when necessary with the MEGA7 software (Kumar et al., 2016). Phylogenetic
 285 analyses were performed in the CIPRES Science Gateway portal v 3.3 (Miller et al., 2012)
 286 using two different approaches: Maximum Likelihood (ML) analysis with Garli v 2.0
 287 (Zwickl, 2006) and Bayesian Inference (BI) analysis with MrBayes v 3.2.6 (Ronquist et al.,
 288 2012). Evolutionary models were tested using ModelTest-NG (Darriba et al., 2020) and the
 289 best-fit substitution model for each gene was selected based on the Akaike criterion
 290 (Supplementary Table S4).

291 For Maximum Likelihood analyses, evolution was simulated until likelihood scores
 292 converged. Nonparametric bootstrap analyses were performed with 1000 pseudoreplicates, in
 293 order to estimate the statistical support of the branches. Nodes with zero branch lengths were
 294 collapsed. Bootstrap trees were then compiled using SumTrees v 4.3.0 from the Dendropy v
 295 4.4.0 package (Sukumaran and Holder, 2010). For Bayesian Inference, analyses were

performed using two parallel runs and a sampling frequency set to every 10,000 generations; each of the two parallel runs had one cold and three heated chains, and were run until split frequencies were less or equal to 0.01. The first 25% of the generated trees were discarded as burn-in before calculating the 50% majority consensus trees for ML and posterior probability values for BI. The resulting trees from both analyses were outputted in FigTree v 1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) and their topologies were compared. We assessed possible incongruences and conflicts between clades with significant bootstrap/posterior probability support in both analyses, and compared their topologies to the ones presented in previously published phylogenies of *Phyllosticta* (Marin-Felix et al., 2019) and *Botryosphaeriales* (Phillips et al., 2018). Alignments and trees produced in this study are available in TreeBASE (study S25256, <http://purl.org/phylo/treebase/phylows/study/TB2:S25256?x-access-code=7ecdbf532816b871453448fa454336e0&format=html>).

The multilocus tree (*ITS*, *TEF1*, *ACT* and *GAPDH*) was used to map characters in the reconstruction of the ancestral thallic state in Mesquite v 3.61 (Maddison and Maddison, 2019). Character states were heterothallism (when only the *MAT1-1* or *MAT1-2* idiomorph was present in the genome), homothallism (when both mating-type idiomorphs were present in the genome, or when pure cultures were described to produce ascospores *in vitro*), or unknown. Besides the character states from *Phyllosticta* species, we also included the thallic states the outgroups, *Diplodia corticola* and *Neofusicoccum parvum* (Nagel et al., 2018).

316

317 **3 Results**

318

319 *3.1 Genome assemblies and gene annotation*

320

Our *de novo* assemblies for *P. capitalensis*, *P. citribraziliensis* and *P. citricarpa* resulted in genomes ranging from ~31Mb for *P. citribraziliensis* to ~32Mb for *P. capitalensis* and *P. citricarpa*, thereby being comparable to the other *Phyllosticta* genomes already available (Table 1). However, the assemblies revealed considerable differences in the extent of contig fragmentation: the new assemblies of *P. capitalensis* LGMF01 and *P. citricarpa* LGMF06 comprise 231 and 354 contigs, respectively, while the first published assemblies of these same strains included more than 11,000 contigs (Rodrigues et al., 2019).

An analysis of assembly completeness based on BUSCO demonstrates that our assemblies present genome completeness levels around 98%, comparable to the reference assemblies of the ex-type strains (Guarnaccia et al., 2019), regardless of being more fragmented (Table 1). This suggests that assembly fragmentation does not cause an increase in the number of missing protein coding genes. Instead, the unassembled fragments likely comprise repetitive sequences.

There is substantial variation in the repetitive content in *Phyllosticta* genomes. Some species present low TE content, such as *P. capitalensis* (0.14 to 1.54%), while other species present higher TE content, such as *P. citriasiana* strain CGM3.14344. The number of predicted genes varies between 9,083 (in *P. paracitricarpa* CPC 27169) and 11,217 genes (in *P. citricarpa* LGMF06) (Table 1).

3.2 Homothallism and heterothallism throughout the genus *Phyllosticta*

To enable a comparison of the mating-type locus structure in *Phyllosticta* species, we first assessed the presence of mating-type genes in the genomes. Genes with high similarity to mating-type genes of *P. citricarpa*, containing either the alpha-box domain (pfam04769) from the *MAT1-1-1* gene, or the high-mobility group (cd01389) domain from the *MAT1-2-1* gene, were found in all the analyzed *Phyllosticta* genomes.

347 The *MAT1-1-1* gene always included the alpha-box domain (Table 2) and co-occurred
348 with the *MAT1-1-8* gene, together comprising the *MAT1-1* idiomorph. On the other hand, the
349 *MAT1-2-1* gene always contained the HMG-box domain (Table 2), and co-occurred with the
350 *MAT1-2-5* gene, comprising the *MAT1-2* idiomorph.

351 Based on the presence of mating-type idiomorphs, it is possible to classify *Phyllosticta*
352 species as heterothallic (only *MAT1-1* or *MAT1-2* idiomorph in the genome) or homothallic
353 (when both idiomorphs occur in the genome). Based on our analyses of mating-type
354 idiomorphs we conclude that *P. capitalensis* and *P. citrichinaensis* are homothallic, and *P.*
355 *citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P. paracitricarpa* are heterothallic (Table 1),
356 All *P. citricarpa* strains (except strain CGMCC3.14348), *P. citriasiana* (CGMCC3.14344)
357 and *P. paracitricarpa* (CBS 141357) encode the *MAT1-2* idiomorph, while the *P. citriasiana*
358 strain CBS 120486 and the *P. citribraziliensis* strains encode the *MAT1-1* idiomorph (Table
359 1).

360

361 3.3 Mating-type locus conservation in *Citrus*-associated *Phyllosticta* species

362

363 To assess whether rearrangements in the mating-type locus are associated with
364 transitions in mating strategies, we investigated the organization and synteny of the locus
365 across the *Citrus*-associated *Phyllosticta* species. The analysis also included additional genes
366 upstream and downstream of the mating-type locus. This enabled syntenic comparisons not
367 only across the genus *Phyllosticta* (Phyllostictaceae), but also with the allied family
368 Botryosphaeriaceae (Nagel et al., 2018).

369 The upstream region (~12Kb in length) consists of a syntenic cluster of six genes
370 (*SAICARsyn-APN2* cluster), which are all protein coding and are present in the following
371 conserved order: Phosphoribosylaminoimidazole-succinocarboxamide synthetase
372 (*SAICARsyn*), Mitochondrial carrier protein (*MCP*), Complex I intermediate-associated

373 protein (*CIA30*), Anaphase-promoting complex subunit 5 (*APC5*), Cytochrome c oxidase
374 subunit Via (*CoxVia*) and DNA lyase (*APN2*). The downstream region (~5Kb in length)
375 comprises a gene coding a putative integral membrane protein (*PIM*) and the 40S ribosomal
376 protein S9.

377 In relation to the mating-type genes, their order and orientation is also conserved in the
378 *Citrus*-associated *Phyllosticta* species (Figure 1 and Figure S1). In the *MAT1-1* idiomorph of
379 heterothallic species, *MAT1-1-8* was always followed by *MAT1-1-1*, in opposite orientations
380 from each other. In the *MAT1-2* idiomorph, *MAT1-2-1* was always followed by *MAT1-2-5*,
381 with opposite orientations from each other. Furthermore, when comparing both idiomorphs,
382 *MAT1-1-1* and *MAT1-2-1* have opposite orientations. In the homothallic species *P.*
383 *capitalensis*, which presents both idiomorphs in the same genomic location, the *MAT1-1*
384 idiomorph was followed by the *MAT1-2* idiomorph, both between the *SAICARsyn-APN2*
385 cluster and *PIM*.

386 There are two different types of structure of the mating-type locus in the *Phyllosticta*
387 species analysed (Figure 1 and Figure S1). In the first one, present in most of the species, the
388 mating-type genes are localized between the upstream and downstream regions
389 aforementioned. In this arrangement, the idiomorph-specific region has a size ranging from
390 ~5Kb in heterothallic species *P. citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P.*
391 *paracitricarpa*, to ~9Kb in the homothallic species *P. capitalensis* (Figure 1). The second
392 type of structure comprises a different arrangement observed in the other homothallic species,
393 *P. citrichinaensis*. The *SAICARsyn-APN2* cluster was found in scaffold 9 of the genome
394 assembly, followed by the *MAT1-1* idiomorph, which was separated from *PIM* and 40S
395 protein S9 (*40S S9*) by a ~10Kb region rich in AT content (~70%), with no predicted genes.
396 Despite the absence of functional genes in this 10Kb region, degenerate domains of a reverse
397 transcriptase (pfam07727), integrase (cl21549) and Ty/Copia RNase HI (cd09272) were

found by CDD-Search, implying the insertion of a transposable element into the mating-type locus. Moreover, the *MAT1-2* idiomorph was found in a different scaffold (scaffold 6) (Figure 1 and Figure S1). In this scaffold, only the *MAT1-2-1* and *MAT1-2-5* genes were present: 5' and 3' flanking sequences were not similar to any of the six genes from the syntenic cluster. Upstream to the *MAT1-2* idiomorph location, there is a ~6Kb sequence identified as a TE fragment by REPET. However, the sequence is too degenerate to allow further characterization and recognition of conserved domains. Even though we could detect that the *MAT1-2* idiomorph was separated from the *MAT1-1* idiomorph and the *SAICARsyn-APN2* cluster, we were not able to determine if both idiomorphs are on the same or separate chromosomes.

3.4 *MAT1-2* strains present a truncated version of *MAT1-1-8* gene

Sequences showing similarity to the *MAT1-1-8* gene from *P. citricarpa* and *P. citriasiana* were observed in the intergenic region between *APN2* and *MAT1-2-1* genes in *P. citricarpa*, *P. citriasiana* and *P. paracitricarpa* *MAT1-2* strains (Figure S2). In the *MAT1-2* strains, *MAT1-1-8* sequences have a length of 1518 bp. Sequence comparisons with *MAT1-1-8* sequences of *MAT1-1* strains from the same species show 124 bp deletions in the 5' end of the gene in *MAT1-2* strains. Moreover, there are 16 nucleotide substitutions along the sequence in *P. citricarpa* and *P. paracitricarpa*. These differences would lead to a truncated version of *MAT1-1-8* in *MAT1-2* strains, suggesting that it might be a pseudogene. Apart from these *MAT1-1-8* fragments, no fragments from *MAT1-1-1*, *MAT1-2-1* and *MAT1-2-5* were found in the opposite idiomorph, or in different genomic locations, for all the *Phyllosticta* species analysed.

3.5 *MAT* genes follow the evolutionary history of *Phyllosticta*, but *PIM* gene shows unexpected relationships

For all the genes, both Bayesian inference and maximum likelihood phylogenetical analyses provided similar topologies, and therefore we present single trees for each gene with combined results from both approaches.

For the six genes in the flanking syntenic cluster (*SAICAR_{syn}*, *MCP*, *CIA30*, *APC5*, *CoxVIa* and *APN2*) (Figures S3 to S8), and for the mating-type genes (*MAT1-1-1*, *MAT1-1-8*, *MAT1-2-1*) (Figure 2, Figures S9 to S11), both for nucleotide and amino acid sequences, the trees are highly similar to the recognized species tree inferred from the multilocus phylogeny using ITS, *TEF1*, *ACT* and *GAPDH*, as well as previously published phylogeny (Marin-Felix et al., 2019) and phylogenome (Guarnaccia et al., 2019) of the genus *Phyllosticta*. The exceptions are *MAT1-2-5* (Figure S12), for which it was not possible to observe resolution at species level for *P. citricarpa*, *P. citriasiana* and *P. paracitricarpa*, and *40S S9* (Figure S13), which is highly conserved across the genus *Phyllosticta*.

However, the *PIM* gene was unexpectedly incongruent with the species tree (Figure 3, Figure S14). The *PIM* sequences of *P. citricarpa* strain CGM3.14348 and *P. citriasiana* strain CBS 120486 (both *MAT1-1* strains) group together in a separate clade, instead of grouping with the *PIM* sequences from the other isolates of their respective species. Furthermore, the species tree inferred from *PIM* sequences shows a different relationship for *P. citribraziliensis* and *P. citrichinaensis*. The multilocus species tree (ITS+*TEF1*+*ACT*+*GAPDH*) (Figure 1) shows *P. citribraziliensis* being most closely related to *P. citrichinaensis*, and both are in a separate clade from *P. citriasiana*, *P. citricarpa* and *P. paracitricarpa*. However, in the tree inferred from *PIM* sequences, *P. citribraziliensis* is not most closely related to *P. citrichinaensis* as expected, but groups in the same clade as *P. citriasiana* and *P. citricarpa*.

448 *MAT1-1* strains. This suggests that there are two different versions of the *PIM* gene, and that
 449 these versions show some level of idiomorph specificity: for heterothallic species, *PIM* genes
 450 group according to the idiomorph of the strain. As for the homothallic species, for *P.*
 451 *capitalensis* the *PIM* gene is closely related to the *PIM* from *MAT1-2* strains. This could be
 452 related to the mating-type locus organization in *P. capitalensis* where *PIM* is located right
 453 next to the *MAT1-2* idiomorph genes. Interestingly, the *P. citrichinaensis* *PIM* also groups
 454 with the *PIM* from *MAT1-2* strains, even though the *MAT1-2* idiomorph is in another genomic
 455 location.

456

457 3.6 The ancestral mating type locus of the genus *Phyllosticta* was homothallic

458

459 To assess the direction of transition between homothallic and heterothallic forms we
 460 performed an ancestral state reconstruction analysis. Hereby we find evidence for
 461 homothallism being the ancestral form in the genus *Phyllosticta* (Figure 4). Homothallism is
 462 present in several species from different clades, while heterothallism is restricted to a clade
 463 comprised mostly of *Citrus*-associated species.

464 Moreover, the reconstruction suggests that homothallism of *P. capitalensis* and *P.*
 465 *citrichinaensis* does not originate from the same evolutionary event. This scenario is also
 466 supported by the fact that they differ in the mating-type locus organization: in *P. capitalensis*,
 467 both idiomorphs are present in the same locus, whereas in *P. citrichinaensis* only the *MAT1-1*
 468 is present at the main mating-type locus, and *MAT1-2* occurs in another genomic region.

469

470 3.7 Mating-type genes present a different rate of sequence evolution

471

472 To investigate if the mating-type genes present a different rate of evolution from genes
 473 which are not associated with reproduction processes, we compared the degree of

conservation, both at nucleotide and amino acid level, for the mating-type and flanking genes. Nucleotide and amino acid sequences of the *MAT1-1-1*, *MAT1-1-8*, *MAT1-2-1* and *MAT1-2-5* genes are less conserved than the sequences from the flanking genes (Table 3). The similarity level for mating-type genes ranges from 64.89% to 75.97% for nucleotide and from 53.19% to 61.94% for amino acid sequences. In contrast, for the flanking genes, the similarity level is higher, ranging from 69.30% to 93.21% for nucleotide and from 70.22% to 98.95% for amino acid sequences. Interestingly, despite being a flanking gene, the similarity level at the *PIM* gene is lower than the mating-type genes: 53.07% for nucleotide and 43.45% for amino acid sequences. This low similarity level is explained by the differences presented by the *PIM* sequences from *P. citribraziliensis*, *P. citriasiana* and *P. citricarpa* *MAT1-1* strains, when compared to the rest of the *PIM* sequences. This reinforces the suggestion of two versions of the *PIM* gene being present, as seen by the phylogenetic analysis (Figure 3).

On the other hand, despite the high variation at sequence level, we observed a high degree of conservation in terms of gene structure, for both the mating-type and adjacent genes. This conservation is observed at the size, number and position of exons and introns. One of the few exceptions would be the *MAT1-1-8* gene in *P. citricarpa* and *P. citriasiana* *MAT1-1* strains, in which the first intron is larger (~267 bp) than the introns from its homologues in *P. capitalensis*, *P. citribraziliensis* and *P. citrichinaensis* (~48 bp).

Discussion

In our study we performed a detailed and comparative characterization of the mating-type locus in *Citrus*-associated *Phyllosticta* species. Moreover, we provide important resources for future comparative and evolutionary studies in this group of fungal species, in

the form of new robust whole genome assemblies and gene annotations. Here, our comparative analyses demonstrate examples on how specific rearrangements in the overall conserved mating-type locus of *Phyllosticta* species may lead to important biological changes, such as the shift from homothallism to heterothallism.

Following the usual BLAST-based approach employed as the starting point in studies on mating-type loci in fungal species, we could identify the presence of *MAT* idiomorphs in all genomes analysed. Besides the essential *MAT1-1-1* and *MAT1-2-1* genes, which per se are the characteristic features of the canonical *MAT* loci in Ascomycetes, we demonstrate the presence of additional genes in the loci of both idiomorphs. A similar situation was also found in genera of the *Botryosphaeriaceae*, which is a family sister to *Phyllostictaceae* (Botryosphaeriales, Dothideomycetes). In *Botryosphaeriaceae* species such as *D. sapinea* (Bihon et al., 2014), *Neofusicoccum* spp. (Lopes et al., 2017), *Botryosphaeria dothidea*, *Macrophomina phaseolina* and *Lasiodiplodia* spp. (Nagel et al., 2018), *MAT1-1-8* and *MAT1-2-5* co-occur with *MAT1-1-1* and *MAT1-2-1*, respectively.

MAT1-1-8 was previously referred to as *MAT1-1-4* in *P. citricarpa* (Amorim et al., 2017) and *oml1* (Wang et al., 2016) in *P. citricarpa* and *P. capitalensis*. *MAT1-1-4* was originally described in *Pyrenopeziza brassicae* as a mating-type gene with homology to metallothionein proteins (Wilken et al., 2017). These proteins are usually associated with the regulation of cellular growth and homeostatic control of heavy metal levels, suggesting that *MAT1-1-4* expression is induced by the senescence of host leaves, a situation in which there is an accumulation of heavy metal ions. This would be particularly important for the reproduction of *Py. brassicae*, which produces sexual structures in senescing tissue of the host (Wilken et al., 2017). *MAT1-1-8* was previously assigned as *MAT1-1-4* due to 31% of sequence identity to the *MAT1-1-4* gene described in *Diplodia sapinea* (Bihon et al., 2014). However, a recent revision of the mating-type genes nomenclature suggests that the gene

524 from *D. sapinea* does not meet the required criteria to be classified as *MAT1-1-4* (Wilken et
525 al., 2017). Firstly, there is the low sequence similarity to the original *MAT1-1-4* described in
526 *Py. brassicae* (below 20%) and secondly, there is no information regarding the existence of a
527 possible functional metallothionein domain in the *D. sapinea* gene. Given this, the authors
528 propose that both are likely different genes, and that *MAT1-1-4* from *D. sapinea* be renamed
529 as *MAT1-1-8* instead, allowing for future investigation and determination of the *MAT1-1-8*
530 gene product and its biological function (Wilken et al., 2017). Thus, we propose here to adapt
531 the *Phyllosticta* mating-type nomenclature to accommodate this change.

532 Likewise, *MAT1-2-5* was previously referred to as *MAT1-2-9* in *P. citricarpa* and *P.*
533 *capitalensis* (Wang et al., 2016), due to the use of *MAT1-2-5* as a name for an idiomorph-
534 specific *COX13* version in *Coccidioides* species (Mandel et al., 2007). Nomenclature of this
535 gene was also revised and corrected, with the proposition of retaining *MAT1-2-5* for *P.*
536 *citricarpa* and *D. sapinea*, and referring to *Coccidioides* genes as alternative versions of
537 *COX13* instead (Wilken et al., 2017), which we follow accordingly. As is the case for *MAT1-*
538 *1-8*, further studies on the *MAT1-2-5* biological function and involvement in sexual
539 reproduction are still required.

540 Regarding the overall mating-type locus organization, the presence of the synteny
541 cluster *SAICARsyn-APN2* upstream to the mating-type idiomorphs, in this orientation, is also
542 observed in different *Botryosphaeriaceae* species, such as *Lasiodiplodia pseudotheobromae*,
543 *L. gonubiensis*, *Botryosphaeria dothidea* and *Macrophomina phaseolina* (Nagel et al., 2018).
544 This high degree of similarity spanning from *Phyllosticta* across the whole order
545 *Botryosphaeriales* is consistent with the frequent descriptions of positional conservation for
546 the mating-type locus, commonly associated to genes such as *APN2* (Wilken et al., 2017).

547 However, despite the high level of overall organization for the mating-type locus, a
548 remarkable rearrangement was observed for *P. citrichinaensis*, as one of the idiomorphs

occurs in a separate genomic location. Comparable to this different kind of arrangement, previous studies have documented the presence of mating-type genes separate from the main mating-type locus, e.g. in *Aspergillus nidulans* (Galagan et al., 2005), *Eutiarosporella darliae* (Thynne et al., 2017), *Neurospora sublineolata* (Gioti et al., 2012), *Neofusicoccum australe* and *N. luteum* (Nagel et al., 2018). Moreover, in *Cercospora beticola*, apart from the main complete mating-type locus, fragments of *MAT1-1-1* and *MAT1-2-1* are dispersed throughout the genome (Bolton et al., 2014), something we did not observe for the *Citrus*-associated *Phyllosticta* species. Interestingly, the mechanisms involved in the remodelling of the mating-type locus in the aforementioned species differ. For example, in the heterothallic species *C. beticola* the *MAT* fragments do not contain introns, which implies that they likely originated from RNA transcripts that integrated themselves in the genome (Bolton et al., 2014). On the other hand, in the homothallic species *E. darliae* the *MAT1-1-1* gene, which is separated from the main *MAT* locus, still retains the intron sequences, as well as flanking fragments of non-coding DNA. In this case, it has been proposed that a DNA breakage or transposition event is responsible for the excision of *MAT1-1-1* from an ancestor heterothallic strain, which was later integrated in the *E. darliae* genome (Thynne et al., 2017). For *N. sublineolata*, the presence of the repetitive element *nsubGypsy* inside the mating-type locus suggests that the rearrangement of this locus happened through a TE-mediated translocation (Gioti et al., 2012). However, despite the several possible mechanistic differences observed in these species, one common outcome of the mating-type locus remodelling is a shift in thallism state, usually from heterothallism to homothallism. Studies in Ascomycetes show that the change from heterothallism is a frequent event (Nagel et al., 2018; Thynne et al., 2017), and may occur multiple times within one genus (Gioti et al., 2012). Moreover, it has been proposed that shifts from heterothallism to homothallism could have been a driver of speciation among pathogenic *Botryosphaeriaceae* species (Thynne et al., 2017).

574 Our results suggest that what happened in the genus *Phyllosticta* was the opposite: a
575 shift from homothallism to heterothallism. Due to the basal placement of *P. capitalensis* in the
576 phylogenetic analyses (Figure 4; also see Wikee et al., 2013), we hypothesize that the mating-
577 type locus organization for the homothallic ancestor of the *Phyllosticta* genus would likely
578 resemble what is observed for *P. capitalensis*, with both *MAT* idiomorphs occurring in the
579 same locus. From this homothallic ancestor, we propose two models to explain the mating-
580 type locus organization for the homothallic species *P. citrichinaensis* and for the heterothallic
581 species *P. citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P. paracitricarpa* (Figure 5).

582 In *P. citrichinaensis*, we hypothesize a model in which the mating-type locus
583 remodelling was mediated by TEs (Figure 5a). First, an insertion of a transposable element
584 between *MAT1-1* and *MAT1-2* idiomorphs occurred. Then, this event was followed either by a
585 recombination event between the TE inserted in the mating-type locus and another TE in a
586 different genome location (e.g. the degenerate TE found ~6Kb upstream the *MAT1-2*
587 idiomorph location), or solely by the mobilization of the TE. The consequence of this
588 translocation would be the mating-type locus organization that is observed in *P.*
589 *citrichinaensis*, in which *MAT1-2* is separated from the rest of the mating-type locus. This
590 model also explains the phylogenetic placement of the *P. citrichinaensis* *PIM* together with
591 the *PIM* genes of *MAT1-2* strains (Figure 3): even though there is no physical association
592 between the *PIM* gene and *MAT1-2* genes in this species anymore, the *MAT1-2* idiomorph
593 was closely located to the *PIM* gene in the *P. citrichinaensis* ancestor (Figure 5a). The
594 involvement of TEs in the remodelling of the mating-type locus has already been proposed in
595 other species such as *Neurospora sublineolata* (Gioti et al., 2012), *Chrysosporthe*
596 *austroafricana* (Kanzi et al., 2019) and *Sclerotinia trifoliorum* (Xu et al., 2016), either
597 through translocations, gene loss and unidirectional mating-type switching. Moreover, TEs
598 are known to impact the genome evolution of several fungal species by promoting

599 chromosomal rearrangements (Castanera et al., 2016). Even though the TE content in
600 *Phyllosticta* species seems to be low, it is possible that some of these TEs may be involved in
601 important rearrangements, such as the one we describe here for the mating-type locus. In this
602 sense, further characterization on the TE dynamics in the *Phyllosticta* species may provide
603 interesting insights.

604 For the heterothallic species *P. citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P.*
605 *paracitricarpa*, we propose a model based on gene loss (Figure 5b). In different individuals
606 from a homothallic ancestor, partial deletions of either the *MAT1-1* or *MAT1-2* idiomorphs
607 occurred, creating heterothallic strains incapable of selfing. In the *MAT1-1* strains, the whole
608 *MAT1-2* idiomorph was lost whereas, in the *MAT1-2* strains, only the *MAT1-1-1* gene was
609 completely lost, as a truncated version of *MAT1-1-8* still exists. This presence of partial
610 *MAT1-1-8* in *MAT1-2* strains is not unusual, and is already reported in literature for some
611 other *Botryosphaeriales* species, such as *Diplodia pinea* (Bihon et al., 2014), *Lasiodiplodia*
612 *theobromae*, *L. gonubiensis*, *Macrophomina phaseolina*, *Neofusicoccum australe* and *N.*
613 *luteum* (Nagel et al., 2018). Moreover, it has been proposed that these partial gene fragments
614 in the mating-type locus of heterothallic strains can arise either from heterothallic ancestors
615 that underwent unequal crossing over/recombination events, or from homothallic ancestors
616 that experienced two independent deletion events (Nagel et al., 2018), such as what we
617 hypothesize for the genus *Phyllosticta*. In this scenario, it is possible to expect some level of
618 homologous recombination due to the considerable sequence similarity between the *MAT1-1-*
619 *8* gene and the partial fragments in the *MAT1-2* idiomorph in *Phyllosticta* species. This
620 recombination would reduce the beginning of the effective idiomorphic region (Nagel et al.,
621 2018) to just part of the *MAT1-1-8* gene.

622 Nevertheless, an opposite effect is observed for the *PIM* gene which includes two
623 idiomorph-specific versions (Figures 1 and 3). This diversity suggests that *PIM* is being

624 incorporated into the idiomorphic region not only in *P. citricarpa* (Wang et al., 2016), but
 625 also in the other heterothallic species. The expansion of the idiomorphic region has already
 626 been observed for other fungal species, even including genes that are not involved in the
 627 mating-type determination (Branco et al., 2017; Wang et al., 2016). To some extent, the
 628 expansion of the recombination suppression region observed in the heterothallic *Phyllosticta*
 629 species is consistent with proposed models of evolution of *MAT* loci towards sex
 630 chromosomes, which occurs through successive increases of the recombination suppression
 631 regions (evolutionary strata) (Branco et al., 2017).

632 The incorporation of the *PIM* gene to the idiomorphic region is also evidenced by the
 633 low degree of sequence conservation, comparable to what is observed for the mating-type
 634 genes. Mating-type genes usually evolve at faster rates than other sequences in the genome
 635 (Turgeon, 1998), contrasting with the conservation observed at gene structure and genomic
 636 levels for the mating-type locus. This diversity in evolutionary rates reinforces the importance
 637 of evaluating the genome evolution of the *Citrus*-associated *Phyllosticta* species across
 638 different levels.

639 Our characterization of the mating-type locus in the *Citrus*-associated *Phyllosticta*
 640 species also provides foundations for further functional studies on the sexual reproduction
 641 processes in these species. As mentioned before, *MAT1-1-8* and *MAT1-2-5* genes still lack
 642 information regarding their biological products and functions (Wilken et al., 2017). Moreover,
 643 little is known about expression of mating-type genes in *Phyllosticta* species, especially in the
 644 context of the interaction with the hosts. In this sense, functional approaches focusing on gene
 645 expression and characterization (Kim et al., 2015; Wilson et al., 2020) are required to
 646 determine the exact contribution of *MAT1-1-1*, *MAT1-1-8*, *MAT1-2-1* and *MAT1-2-5* to sexual
 647 reproduction in *Phyllosticta*.

648 The investigation of the sexual reproduction in *Phyllosticta* species also provides
649 valuable insights in the context of the Citrus Black Spot and the Citrus Tan Spot diseases, by
650 allowing an evaluation of ascospore pathogenicity and its role in the disease cycle. Regarding
651 *P. citricarpa*, ascospore production and pathogenicity were described on leaves of Troyer
652 seedlings and fruits of Murcott tangor (Tran et al., 2018, 2017). Moreover, *P. citricarpa*
653 isolates with different genotypes vary in their aggressiveness, implying that the genetic
654 diversity generated by sexual reproduction has epidemiological significance and should be
655 taken into consideration (Tran et al., 2018). Regarding the other pathogenic species, further
656 investigation on ascospore pathogenicity and the influence of genetic diversity in disease
657 development are still required. For *P. citriasiana*, pathogenicity assays were not conducted
658 since its description (Wulandari et al., 2009), and for *P. paracitricarpa*, only conidia were
659 evaluated until now (Guarnaccia et al., 2017). As both species are heterothallic, similar
660 strategies to the ones employed in *P. citricarpa* (mating-type PCR and crossings with
661 opposite isolates) (Amorim et al., 2017; Guarnaccia et al., 2019; Tran et al., 2017; Wang et
662 al., 2016) may allow the *in vitro* production of ascospores and their evaluation in future. As
663 recombination increases genetic diversity and can generate new combinations of virulence
664 alleles, disease management practices that prevent the occurrence of sexual reproduction
665 could assist in the control of *P. citriasiana*, *P. citricarpa* and *P. paracitricarpa*, similar to
666 other recombining pathogens (McDonald and Linde, 2002; McDonald and Mundt, 2016). The
667 successful implementation of such practices relies on the knowledge about the mating-type
668 distribution in different populations, reinforcing the need of a broader sampling of isolates
669 from these *Phyllosticta* species (Guarnaccia et al., 2019).

670 Finally, interesting insights on the link between speciation and shifts between
671 homothallism and heterothallism may become available from further investigations on
672 mating-type loci from other species of *Phyllosticta*, especially the homothallic ones. Possible

673 candidates would be species from the *P. musarum* and *P. abieticola* clades, to assess if their
 674 mating-type locus is similar to *P. capitalensis*, which would be expected based on our results
 675 from the ancestral state reconstruction (Figure 4). Another possibility is to investigate if the
 676 maintenance of homothallism provides lifestyle benefits to these species, given the ease of
 677 spread and sexual propagation to new locations provided by the unselective mating in
 678 homothallic individuals (Nagel et al., 2018; Thynne et al., 2017). Moreover, for the
 679 heterothallic species, it would be interesting to assess which advantages arise from the shift to
 680 heterothallism, and, especially, how they relate to the evolutionary potential of the pathogenic
 681 species, such as *P. citricarpa*. Such possibilities highlight the importance of addressing the
 682 *Citrus-Phyllosticta* interactions under an evolutionary and genomic perspective in future, in a
 683 whole genome context.

684

685 4 Conclusions

686

687 In summary, our study shows the mating-type locus in *Phyllosticta* as a very dynamic
 688 region, with different possible rearrangements that lead to changes in reproduction strategies.
 689 Our results demonstrate that although mating-type genes show high sequence variability, the
 690 mating-type locus organization is well conserved along *Citrus*-associated *Phyllosticta* species,
 691 contrasting evolutionary processes at gene and genomic level. Reconstruction of the ancestral
 692 thallism in *Phyllosticta* suggests that homothallism is the ancestral state, with a transition to
 693 heterothallism in some of the *Citrus*-associated *Phyllosticta* species mediated by gene losses.
 694 Moreover, homothallism in *P. capitalensis* and *P. citrichinaensis* results from different
 695 evolutionary events, as the remodeling of the mating-type locus in *P. citrichinaensis* could
 696 have happened through a TE-mediated translocation. Idiomorphic-specific versions of the
 697 *PIM* gene provide evidence of an expansion of the idiomorphic region in the heterothallic

species. As a whole, knowledge on mating-type evolution in *Citrus*-associated *Phyllosticta* species offers new perspectives in the study of *Citrus-Phyllosticta* interactions, especially under evolutionary and genomic approaches, and may provide valuable information to guide disease management practices.

Acknowledgements

Authors are grateful to Doreen Landermann and Mieke Starink-Willemse for technical assistance, to Paulo J. C. dos Santos for the support in generating *P. citribraziliensis* read data, and Primrose Boynton for her feedback on the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, which provided a PhD scholarship to Desirê A. L. Petters-Vandresen; it was also supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) with the INCT-Citros grant 465440/2014-2 and the research grant 424738/2016-3 to Chirlei Glienke, and by FAPESP with the grant 2014/50880-0, allowing the generation of some of the *Phyllosticta* genomes used in this study.

References

- Amorim, R., Savi, D.C., Ferreira-Maba, L., Aluizio, R., Goulin, E.H., Takita, M.A., Machado, M.A., Glienke, C., 2017. *MAT* gene idiomorphs suggest a heterothallic sexual cycle in the citrus pathogen *Phyllosticta citricarpa*. *Eur. J. Plant Pathol.* 147, 325–337.
<https://doi.org/10.1007/s10658-016-1005-8>
- Baldassari, R.B., Wickert, E., de Goes, A., 2008. Pathogenicity, colony morphology and diversity of isolates of *Guignardia citricarpa* and *G. mangiferae* isolated from *Citrus*

723 spp. Eur. J. Plant Pathol. 120, 103–110. <https://doi.org/10.1007/s10658-007-9182-0>

724 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin,
725 V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A. V., Sirotkin, A. V.,
726 Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: A new genome
727 assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19,
728 455–477. <https://doi.org/10.1089/cmb.2012.0021>

729 Bihon, W., Wingfield, M.J., Slippers, B., Duong, T.A., Wingfield, B.D., 2014. *MAT* gene
730 idiomorphs suggest a heterothallic sexual cycle in a predominantly asexual and important
731 pine pathogen. Fungal Genet. Biol. 62, 55–61. <https://doi.org/10.1016/j.fgb.2013.10.013>

732 Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina
733 sequence data. Bioinformatics 30, 2114–2120.
734 <https://doi.org/10.1093/bioinformatics/btu170>

735 Bolton, M.D., de Jonge, R., Inderbitzin, P., Liu, Z., Birla, K., Van de Peer, Y., Subbarao, K.
736 V., Thomma, B.P.H.J., Secor, G.A., 2014. The heterothallic sugarbeet pathogen
737 *Cercospora beticola* contains exon fragments of both *MAT* genes that are homogenized
738 by concerted evolution. Fungal Genet. Biol. 62, 43–54.
739 <https://doi.org/10.1016/j.fgb.2013.10.011>

740 Branco, S., Badouin, H., Rodríguez De La Vega, R.C., Gouzy, J., Carpentier, F., Aguileta, G.,
741 Siguenza, S., Brandenburg, J.T., Coelho, M.A., Hood, M.E., Giraud, T., 2017.
742 Evolutionary strata on young mating-type chromosomes despite the lack of sexual
743 antagonism. Proc. Natl. Acad. Sci. USA. 114, 7067–7072.
744 <https://doi.org/10.1073/pnas.1701658114>

745 Casselton, L.A., Kües, U., 2007. The Origin of Multiple Mating Types in the Model
746 Mushrooms *Coprinopsis cinerea* and *Schizophyllum commune*, in: Heitman, J.,
747 Kronstad, J.W., Taylor, J.W., Casselton, L.A. (Eds.), Sex in Fungi: Molecular

748 Determination and Evolutionary Implications. ASM Press, Washington DC, pp. 283–
749 300. <https://doi.org/10.1128/9781555815837.ch17>

750 Castanera, R., López-Varas, L., Borgognone, A., LaButti, K., Lapidus, A., Schmutz, J.,
751 Grimwood, J., Pérez, G., Pisabarro, A.G., Grigoriev, I. V., Stajich, J.E., Ramírez, L.,
752 2016. Transposable Elements versus the Fungal Genome: Impact on Whole-Genome
753 Architecture and Transcriptional Profiles. PLoS Genet. 12, 1–27.
754 <https://doi.org/10.1371/journal.pgen.1006108>

755 Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T., 2020.
756 ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein
757 Evolutionary Models. Mol. Biol. Evol. 37, 291–294.
758 <https://doi.org/10.1093/molbev/msz189>

759 Doyle, J.J., Dickson, E.E., 1987. Preservation of plant samples for DNA restriction
760 endonuclease analysis. Taxon 36, 715–722. <https://doi.org/10.2307/1221122>

761 Flutre, T., Duprat, E., Feuillet, C., Quesneville, H., 2011. Considering transposable element
762 diversification in de novo annotation approaches. PLoS One 6.
763 <https://doi.org/10.1371/journal.pone.0016526>

764 Galagan, J.E., Calvo, S.E., Cuomo, C., Ma, L.J., Wortman, J.R., Batzoglou, S., Lee, S.I.,
765 Baştürkmen, M., Spevak, C.C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scacciocchio, C.,
766 Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G.H., Draht, O., Busch, S.,
767 D’Enfert, C., Bouchier, C., Goldman, G.H., Bell-Pedersen, D., Griffiths-Jones, S.,
768 Doonan, J.H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E.U., Archer, D.B.,
769 Peñalva, M.Á., Oakley, B.R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida,
770 M., Nierman, W.C., Denning, D.W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R.,
771 Miller, B., Dyer, P., Sachs, M.S., Osmani, S.A., Birren, B.W., 2005. Sequencing of
772 *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature

773 438, 1105–1115. <https://doi.org/10.1038/nature04341>

774 Gioti, A., Mushegian, A.A., Strandberg, R., Stajich, J.E., Johannesson, H., 2012.

775 Unidirectional evolutionary transitions in fungal mating systems and the role of

776 transposable elements. *Mol. Biol. Evol.* 29, 3215–3226.

777 <https://doi.org/10.1093/molbev/mss132>

778 Glass, N., Vollmer, S., Staben, C., Grotelueschen, J., Metzenberg, R., Yanofsky, C., 1988.

779 DNAs of the two mating-type alleles of *Neurospora crassa* are highly dissimilar.

780 *Science* . 241, 570–573. <https://doi.org/10.1126/science.2840740>

781 Glienke, C., Pereira, O.L., Stringari, D., Fabris, J., Kava-Cordeiro, V., Galli-Terasawa, L.,

782 Cunningham, J., Shivas, R.G., Groenewald, J.Z., Crous, P.W., 2011. Endophytic and

783 pathogenic *Phyllosticta* species, with reference to those associated with Citrus Black

784 spot. *Persoonia* 26, 47–56. <https://doi.org/10.3767/003158511X569169>

785 Grabherr, M.G., Hass, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X.,

786 Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A.,

787 Rhind, N.R., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N.,

788 Regev, A., 2013. Trinity: reconstructing a full-length transcriptome without a genome

789 from RNA-Seq data. *Nat. Biotechnol.* 29, 644–652.

790 <https://doi.org/10.1038/nbt.1883>. Trinity

791 Groenewald, M., Groenewald, J.Z., Harrington, T.C., Abeln, E.C.A., Crous, P.W., 2006.

792 Mating type gene analysis in apparently asexual *Cercospora* species is suggestive of

793 cryptic sex. *Fungal Genet. Biol.* 43, 813–825. <https://doi.org/10.1016/j.fgb.2006.05.008>

794 Guarnaccia, V., Gehrman, T., Silva Junior, G.J., Fourie, P.H., Haridas, S., Vu, D.,

795 Spatafora, J., Martin, F.M., Robert, V., Grigoriev, I. V., Groenewald, J.Z., Crous, P.W.,

796 2019. *Phyllosticta citricarpa* and sister species of global importance to Citrus. *Mol. Plant*

797 *Pathol.* 20, 1619–1635. <https://doi.org/10.1111/mpp.12861>

798 Guarnaccia, V., Groenewald, J.Z., Li, H., Glienke, C., Carstens, E., Hattingh, V., Fourie, P.H.,
799 Crous, P.W., 2017. First report of *Phyllosticta citricarpa* and description of two new
800 species, *P. paracapitalensis* and *P. paracitricarpa*, from citrus in Europe. Stud. Mycol.
801 87, 161–185. <https://doi.org/10.1016/j.simyco.2017.05.003>

802 Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: Quality assessment tool for
803 genome assemblies. Bioinformatics 29, 1072–1075.
804 <https://doi.org/10.1093/bioinformatics/btt086>

805 Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger,
806 M.B., Eccles, D., Li, B., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N.,
807 Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc,
808 R.D., Friedman, N., Regev, A., 2013. *De novo* transcript sequence reconstruction from
809 RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8,
810 1494–1512. <https://doi.org/10.1038/nprot.2013.084>

811 Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Robin, C.R.,
812 Wortman, J.R., 2008. Automated eukaryotic gene structure annotation using
813 EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 9,
814 1–22. <https://doi.org/10.1186/gb-2008-9-1-r7>

815 Haas, B.J., Zeng, Q., Pearson, M.D., Cuomo, C.A., Wortman, J.R., 2011. Approaches to
816 fungal genome annotation. Mycology 2, 118–141.
817 <https://doi.org/10.1080/21501203.2011.606851>

818 Heitman, J., Sun, S., James, T.Y., 2013. Evolution of fungal sexual reproduction. Mycologia
819 105, 1–27. <https://doi.org/10.3852/12-253>

820 Hoff, K.J., Lange, S., Lomsadze, A., Borodovsky, M., Stanke, M., 2016. BRAKER1:
821 Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and
822 AUGUSTUS. Bioinformatics 32, 767–769.

823 <https://doi.org/10.1093/bioinformatics/btv661>

824 Kanzi, A.M., Steenkamp, E.T., Van der Merwe, N.A., Wingfield, B.D., 2019. The mating
825 system of the *Eucalyptus* canker pathogen *Chrysosporthe austroafricana* and closely
826 related species. Fungal Genet. Biol. 123, 41–52.
827 <https://doi.org/10.1016/j.fgb.2018.12.001>

828 Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7:
829 Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.
830 <https://doi.org/10.1093/molbev/mst010>

831 Kerényi, Z., Moretti, A., Waalwijk, C., Oláh, B., Hornok, L., 2004. Mating type sequences in
832 asexually reproducing *Fusarium* species. Appl. Environ. Microbiol. 70, 4419–23.
833 <https://doi.org/10.1128/AEM.70.8.4419-4423.2004>

834 Kim, D., Paggi, J.M., Park, C., Bennett, C., Salzberg, S.L., 2019. Graph-based genome
835 alignment and genotyping with HISAT2 and HISAT-genotype. Nat. Biotechnol. 37,
836 907–915. <https://doi.org/10.1038/s41587-019-0201-4>

837 Kim, H.K., Jo, S.M., Kim, G.Y., Kim, D.W., Kim, Y.K., Yun, S.H., 2015. A Large-Scale
838 Functional Analysis of Putative Target Genes of Mating-Type Loci Provides Insight into
839 the Regulation of Sexual Development of the Cereal Pathogen *Fusarium graminearum*.
840 PLoS Genet. 11, 1–32. <https://doi.org/10.1371/journal.pgen.1005486>

841 Kotzé, J.M., 1981. Epidemiology and Control of Citrus Black Spot in South Africa. Plant Dis.
842 65, 945. <https://doi.org/10.1094/PD-65-945>

843 Kronstad, J.W., 2007. Self-Fertility: The Genetics of Sex in Lonely Fungi. Curr. Biol. 17,
844 R843–R845. <https://doi.org/10.1016/j.cub.2007.08.002>

845 Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics
846 Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. 33, 1870–1874.
847 <https://doi.org/10.1093/molbev/msw054>

848 Lee, S.C., Ni, M., Li, W., Shertz, C., Heitman, J., 2010. The Evolution of Sex: a Perspective
849 from the Fungal Kingdom. *Microbiol. Mol. Biol. Rev.* 74, 298–340.
850 <https://doi.org/10.1128/MMBR.00005-10>

851 Linde, C.C., Zala, M., Ceccarelli, S., McDonald, B.A., 2003. Further evidence for sexual
852 reproduction in *Rhynchosporium secalis* based on distribution and frequency of mating-
853 type alleles. *Fungal Genet. Biol.* 40, 115–125. [https://doi.org/10.1016/S1087-](https://doi.org/10.1016/S1087-1845(03)00110-5)
854 [1845\(03\)00110-5](https://doi.org/10.1016/S1087-1845(03)00110-5)

855 Lomsadze, A., Burns, P.D., Borodovsky, M., 2014. Integration of mapped RNA-Seq reads
856 into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* 42, 1–8.
857 <https://doi.org/10.1093/nar/gku557>

858 Lopes, A., Phillips, A.J.L., Alves, A., 2017. Mating type genes in the genus *Neofusicoccum*:
859 Mating strategies and usefulness in species delimitation. *Fungal Biol.* 121, 394–404.
860 <https://doi.org/10.1016/j.funbio.2016.08.011>

861 Lu, S., Wang, J., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., Gwadz, M.,
862 Hurwitz, D.I., Marchler, G.H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M.,
863 Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2020. CDD/SPARCLE: the
864 conserved domain database in 2020. *Nucleic Acids Res.* 48, D265–D268.
865 <https://doi.org/10.1093/nar/gkz991>

866 Maddison, W.P., Maddison, D.R., 2019. Mesquite: a modular system for evolutionary
867 analysis. Version 3.61 <http://www.mesquiteproject.org>

868 Mandel, M.A., Barker, B.M., Kroken, S., Rounsley, S.D., Orbach, M.J., 2007. Genomic and
869 Population Analyses of the Mating Type Loci in *Coccidioides* Species Reveal Evidence
870 for Sexual Reproduction and Gene Acquisition. *Eukaryot. Cell* 6, 1189–1199.
871 <https://doi.org/10.1128/ec.00117-07>

872 Marin-Felix, Y., Hernández-Restrepo, M., Wingfield, M.J., Akulov, A., Carnegie, A.J.,

873 Cheewangkoon, R., Gramaje, D., Groenewald, J.Z., Guarnaccia, V., Halleen, F.,
874 Lombard, L., Luangsa-ard, J., Marinowitz, S., Moslemi, A., Mostert, L., Quaadvlieg,
875 W., Schumacher, R.K., Spies, C.F.J., Thangavel, R., Taylor, P.W.J., Wilson, A.M.,
876 Wingfield, B.D., Wood, A.R., Crous, P.W., 2019. Genera of phytopathogenic fungi:
877 GOPHY 2. Stud. Mycol. 92, 47–133. <https://doi.org/10.1016/j.simyco.2018.04.002>
878 McDonald, B.A., Linde, C., 2002. Pathogen population genetics, evolutionary potential, and
879 durable resistance. Annu. Rev. Phytopathol. 40, 349–379.
880 <https://doi.org/10.1146/annurev.phyto.40.120501.101443>
881 McDonald, B.A., McDermott, J.M., 1993. Population Genetics of Plant Pathogenic Fungi.
882 Bioscience 43, 311–319. <https://doi.org/10.2307/1312063>
883 McDonald, B.A., Mundt, C.C., 2016. How knowledge of pathogen population biology
884 informs management of *Septoria tritici* blotch. Phytopathology.
885 <https://doi.org/10.1094/PHYTO-03-16-0131-RVW>
886 Meng, J.W., Zhu, W., He, M.H., Wu, E.J., Duan, G.H., Xie, Y.K., Jin, Y.J., Yang, L.N.,
887 Shang, L.P., Zhan, J., 2015. Population genetic analysis reveals cryptic sex in the
888 phytopathogenic fungus *Alternaria alternata*. Sci. Rep. 5, 1–10.
889 <https://doi.org/10.1038/srep18250>
890 Miller, M.A., Pfeiffer, W., Schwartz, T., 2012. XSEDE '12: Proceedings of the 1st
891 Conference of the Extreme Science and Engineering Discovery Environment: Bridging
892 from the eXtreme to the campus and beyond. Article No.: 39 Pages 1–8.
893 <https://doi.org/10.1145/2335755.2335836>
894 Nagel, J.H., Wingfield, M.J., Slippers, B., 2018. Evolution of the mating types and mating
895 strategies in prominent genera in the *Botryosphaeriaceae*. Fungal Genet. Biol. 114, 24–
896 33. <https://doi.org/10.1016/j.fgb.2018.03.003>
897 Ni, M., Feretzaki, M., Sun, S., Wang, X., Heitman, J., 2011. Sex in Fungi. Annu. Rev. Genet.

898 45, 405–430. <https://doi.org/10.1146/annurev-genet-110410-132536>

899 O’Connell, J., Schulz-Trieglaff, O., Carlson, E., Hims, M.M., Gormley, N.A., Cox, A.J.,
900 2015. NxTrim: Optimized trimming of Illumina mate pair reads. *Bioinformatics* 31,
901 2035–2037. <https://doi.org/10.1093/bioinformatics/btv057>

902 Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J.A., McQuisten, K.A., Almabrazi, H.G.,
903 Pedersen, K.F., Lin, Y., Garretson, J., McEntaggart, N.O., Sailor, C.A., Dawson, R.B.,
904 Peek, A.S., 2008. IDT SciTools: a suite for analysis and design of nucleic acid
905 oligomers. *Nucleic Acids Res.* 36, 163–169. <https://doi.org/10.1093/nar/gkn198>

906 Phillips, A.J.L., Hyde, K.D., Alves, A., Liu, J.-K., 2018. Families in *Botryosphaeriales*: a
907 phylogenetic, morphological and evolutionary perspective. *Fungal Divers.* 94, 1–22.
908 <https://doi.org/10.1007/s13225-018-0416-6>

909 Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M.,
910 Anxolabehere, D., 2005. Combined Evidence Annotation of Transposable Elements in
911 Genome Sequences. *PLoS Comput. Biol.* 1, e22.
912 <https://doi.org/10.1371/journal.pcbi.0010022>

913 Robinson, A.J., Natvig, D.O., 2019. Diverse members of the Xylariales lack canonical
914 mating-type regions. *Fungal Genet. Biol.* 122, 47–52.
915 <https://doi.org/10.1016/j.fgb.2018.12.004>

916 Rodrigues, C.M., Takita, M.A., Silva, N.V., Ribeiro-Alves, M., Machado, M.A., 2019.
917 Comparative genome analysis of *Phyllosticta citricarpa* and *Phyllosticta capitalensis*,
918 two fungi species that share the same host. *BMC Genomics* 20, 554.
919 <https://doi.org/10.1186/s12864-019-5911-y>

920 Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget,
921 B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian
922 Phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61,

923 539–542. <https://doi.org/10.1093/sysbio/sys029>

924 Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E. V., Zdobnov, E.M., 2015.

925 BUSCO: Assessing genome assembly and annotation completeness with single-copy

926 orthologs. *Bioinformatics* 31, 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>

927 Stanke, M., Diekhans, M., Baertsch, R., Haussler, D., 2008. Using native and syntenically

928 mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637–

929 644. <https://doi.org/10.1093/bioinformatics/btn013>

930 Sukumaran, J., Holder, M.T., 2010. DendroPy: A Python library for phylogenetic computing.

931 *Bioinformatics* 26, 1569–1571. <https://doi.org/10.1093/bioinformatics/btq228>

932 Sullivan, M.J., Petty, N.K., Beatson, S.A., 2011. Easyfig: A genome comparison visualizer.

933 *Bioinformatics* 27, 1009–1010. <https://doi.org/10.1093/bioinformatics/btr039>

934 Ter-Hovhannisyan, V., Lomsadze, A., Chernoff, Y.O., Borodovsky, M., 2008. Gene

935 prediction in novel fungal genomes using an ab initio algorithm with unsupervised

936 training. *Genome Res.* 18, 1979–1990. <https://doi.org/10.1101/gr.081612.108>

937 Thynne, E., McDonald, M.C., Solomon, P.S., 2017. Transition from heterothallism to

938 homothallism is hypothesised to have facilitated speciation among emerging

939 *Botryosphaeriaceae* wheat-pathogens. *Fungal Genet. Biol.* 109, 36–45.

940 <https://doi.org/10.1016/j.fgb.2017.10.005>

941 Tran, N.T., Miles, A.K., Dietzgen, R.G., Dewdney, M.M., Zhang, K., Rollins, J.A., Drenth,

942 A., 2017. Sexual Reproduction in the Citrus Black Spot Pathogen, *Phyllosticta*

943 *citricarpa*. *Phytopathology* 107, 732–739. <https://doi.org/10.1094/phyto-11-16-0419-r>

944 Tran, N.T., Miles, A.K., Smith, M.W., Dietzgen, R.G., Drenth, A., 2018. Pathogenicity of

945 *Phyllosticta citricarpa* Ascospores on *Citrus* spp. *Plant Dis.* 102, 1386–1393.

946 <https://doi.org/10.1094/pdis-08-17-1331-re>

947 Turgeon, B.G., 1998. Application of Mating Type Gene Technology To Problems in Fungal

948 Biology. Annu. Rev. Phytopathol. 36, 115–137.
 949 <https://doi.org/10.1146/annurev.phyto.36.1.115>

950 Turgeon, B.G., Yoder, O.C., 2000. Proposed nomenclature for mating type genes of
 951 filamentous ascomycetes. Fungal Genet. Biol. 31, 1–5.
 952 <https://doi.org/10.1006/fgbi.2000.1227>

953 Waalwijk, C., Mendes, O., Verstappen, E.C.P., de Waard, M.A., Kema, G.H.J., 2002.
 954 Isolation and Characterization of the Mating-Type Idiomorphs from the Wheat Septoria
 955 Leaf Blotch Fungus *Mycosphaerella graminicola*. Fungal Genet. Biol. 35, 277–286.
 956 <https://doi.org/10.1006/fgbi.2001.1322>

957 Wang, M., Liu, B., Ruan, R., Zeng, Y., Luo, J., Li, H., 2020. Genomic Sequencing of
 958 *Phyllosticta citriasiana* Provides Insight Into Its Conservation and Diversification With
 959 Two Closely Related *Phyllosticta* Species Associated With Citrus. Front. Microbiol. 10,
 960 1–9. <https://doi.org/10.3389/fmicb.2019.02979>

961 Wang, N.Y., Zhang, K., Huguet-Tapia, J.C., Rollins, J.A., Dewdney, M.M., 2016. Mating
 962 type and simple sequence repeat markers indicate a clonal population of *Phyllosticta*
 963 *citricarpa* in Florida. Phytopathology 106, 1300–1310. [https://doi.org/10.1094/phyto-12-](https://doi.org/10.1094/phyto-12-15-0316-r)
 964 15-0316-r

965 Wang, X., Chen, G., Huang, F., Zhang, J., Hyde, K.D., Li, H., 2012. *Phyllosticta* species
 966 associated with citrus diseases in China. Fungal Divers. 52, 209–224.
 967 <https://doi.org/10.1007/s13225-011-0140-y>

968 Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G.,
 969 Kriventseva, E. V., Zdobnov, E.M., 2018. BUSCO Applications from Quality
 970 Assessments to Gene Prediction and Phylogenomics. Mol. Biol. Evol. 35, 543–548.
 971 <https://doi.org/10.1093/molbev/msx319>

972 Wikee, S., Lombard, L., Nakashima, C., Motohashi, K., Chukeatirote, E., Cheewangkoon, R.,

973 McKenzie, E.H.C., Hyde, K.D., Crous, P.W., 2013. A phylogenetic re-evaluation of
974 *Phyllosticta* (*Botryosphaeriales*). *Stud. Mycol.* 76, 1–29.
975 <https://doi.org/10.3114/sim0019>

976 Wilken, P.M., Steenkamp, E.T., Wingfield, M.J., de Beer, Z.W., Wingfield, B.D., 2017.
977 Which *MAT* gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature
978 reconsidered. *Fungal Biol. Rev.* 31, 199–211. <https://doi.org/10.1016/j.fbr.2017.05.003>

979 Wilson, A.M., Wilken, P.M., van der Nest, M.A., Wingfield, M.J., Wingfield, B.D., 2020.
980 The novel *Huntia omanensis* mating gene, *MAT1-2-7*, is essential for ascomatal
981 maturation. *Fungal Genet. Biol.* 137, 103335. <https://doi.org/10.1016/j.fgb.2020.103335>

982 Wulandari, N., To-Anun, C., Hyde, K., Duong, L., De Gruyter, J., Meffert, J., Groenewald, J.,
983 Crous, P., 2009. *Phyllosticta citriasiana* sp. nov., the cause of Citrus tan spot of *Citrus*
984 *maxima* in Asia. *Fungal Divers.* 34, 23–39.

985 Xu, L., Jardini, T.M., Chen, W., 2016. Direct repeat-mediated DNA deletion of the mating
986 type *MAT1-2* genes results in unidirectional mating type switching in *Sclerotinia*
987 *trifoliorum*. *Sci. Rep.* 6, 1–10. <https://doi.org/10.1038/srep27083>

988 Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large
989 biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation,
990 The University of Texas at Austin.
991 <https://doi.org/http://www.zo.utexas.edu/faculty/antisense/garli/Garli.html>
992

Table 1 – Genome and strain information of *Phyllosticta* spp. evaluated in this study

Species	Strain ^a	Host	Origin	Thallism	Mating-type	Mating-locus location	Genome size (bp)	Contigs	GC content	Repeat content	Predicted genes	BUSCO ^b	Assembly
<i>P. capitalensis</i>	CBS 128856	<i>Stanhopea</i> sp.	Brazil	Homothallic	Both	Scaffold 10	32461131	14	54.58%	1.54%	9977	97.9%	(Guarnaccia et al., 2019)
	Gm33	<i>Citrus sinensis</i>	USA	Homothallic	Both	Contigs 513 and 1011	32454403	1341	54.56%	0.49%	10183	94.4%	(Wang et al., 2016)
	LGMF 01	<i>Citrus latifolia</i> (leaf)	Brazil	Homothallic	Both	Contig 24	32606250	231	54.48%	0.15%	9953	98%	This study
<i>P. citriasiana</i>	CBS 120486	<i>Citrus maxima</i> (fruit)	Thailand	Heterothallic	MAT1-1	Scaffold 34	32696106	133	51.56%	8.32%	9282	98.1%	(Guarnaccia et al., 2019)
<i>P. citribraziliensis</i>	CGMCC 3.14344	<i>Citrus</i> sp.	China	Heterothallic	MAT1-2	Unitig 22	34225214	92	51.42%	14.50%	9291	97.9%	(Wang et al., 2020)
	CBS 100098	<i>Citrus</i> sp. (leaf)	Brazil	Heterothallic	MAT1-1	Scaffold 7	31670975	32	54.17%	6%	9574	98%	(Guarnaccia et al., 2019)
	LGMF 08	<i>Citrus</i> sp. (leaf)	Brazil	Heterothallic	MAT1-1	Contig 61	31002620	563	54.34%	4.83%	9941	97.9%	This study
<i>P. citricarpa</i>	CBS 127454	<i>Citrus limon</i>	Australia	Heterothallic	MAT1-2	Scaffold 46	28952665	152	54.60%	1.30%	9108	96.1%	(Guarnaccia et al., 2019)
	CPC 27913	<i>Citrus sinensis</i> (leaf litter)	Malta	Heterothallic	MAT1-2	Scaffold 20	32267666	82	52.56%	8.07%	9352	97.6%	(Guarnaccia et al., 2019)
	Gc 12	<i>Citrus sinensis</i>	USA	Heterothallic	MAT1-2	Contigs 1345, 2477, 3787	31127197	5748	53.05%	5.93%	9472	94.5%	(Wang et al., 2016)
	CGMCC 3.14348	-	China	Heterothallic	MAT1-1	Contigs 4070 and 4071	32007210	6716	52.57%	7.94%	9768	87.5%	GenBank Assembly: GCA_000382785.1

	LGMF 06	<i>Citrus sinensis</i>	Brazil	Heterothallic	MAT1-2	Contig 51	32021677	354	52.68%	0.64%	11217	97.8%	This study
<i>P. citrichinaensis</i>	CBS 130529	<i>Citrus maxima</i> (leaf)	China	Homothallic	Both	Scaffolds 6 and 9	29162704	25	55.07%	2.99%	9131	97.5%	(Guarnaccia et al., 2019)
<i>P. paracitricarpa</i>	CPC 27169	<i>Citrus limon</i> (leaf litter)	Greece	Heterothallic	MAT1-2	Scaffold 48	29529839	134	54.35%	2.45%	9083	96.2%	(Guarnaccia et al., 2019)

^a Type strains included in the analysis are indicated in **bold**. Culture collection abbreviations: CBS = Westerdijk Fungal Biodiversity Institute, Utrecht, Netherlands; CPC = Collection of Pedro W. Crous, held at the Westerdijk Fungal Biodiversity Institute; CGMCC = China General Microbiological Culture Collection Center, Chinese Academy of Sciences, Beijing, China; LGMF = Laboratório de Bioprospecção e Genética Molecular de Microrganismos, Universidade Federal do Paraná, Paraná, Brazil.

^b BUSCO completeness assessed with the Pezizomycotina dataset

N/A: not available or not applicable to the present study.

Table 2 - Conserved protein domains present in mating-type and flanking genes in *Phyllosticta* species

Species	Domain position in the amino acid sequences								
	<i>SAICARsyn</i> (<i>pfam01259</i>)	<i>MCP</i> (<i>pfam00153</i>)	<i>CIA30</i> (<i>pfam08547</i>)	<i>APC5</i> (<i>pfam12862</i>)	<i>CoxVIa</i> (<i>pfam02046</i>)	<i>APN2</i> (<i>cd09088</i>)	<i>MAT1-1-1</i> (<i>pfam04769</i>)	<i>MAT1-2-1</i> (<i>cd01389</i>)	<i>40S S9</i> (<i>PTZ00155</i>)
<i>P. capitalensis</i>	5–305	71–166, 175–266, 302–379	44–219	7–159, 257– 347	15–128	5–322	94–271	191–267	2–179
<i>P. citriasiana</i>	6–306	71–166, 175–266, 287–379	44–219	7–159, 256– 346	15–128	4–321	101–274	181–257	2–179
<i>P. citribraziliensis</i>	5–305	71–166, 175–266, 287–379	44–219	89–158, 255– 345	15–128	4–321	137–271	-	2–179
<i>P. citricarpa</i>	6–306	71–166, 175–266, 287–379	44–219	7–159, 256– 346	15–128	4–321	100–274	181–257	2–179
<i>P. citrichinaensis</i>	15–305	71–166, 175–266, 287–379	44–219	89–158, 255– 346	15–128	4–321	112–265	180–256	2–179

		287–379		345					
		71–166, 175–266,		7–159, 256–					
<i>P. paracitricarpa</i>	6–306		44–219		15–128	4–321	-	181–257	2–179
		287–379		346					

Gene abbreviations: *SAICARsyn*: Phosphoribosylaminoimidazole-succinocarboxamide Synthetase; *MCP*: Mitochondrial carrier protein; *CIA30*: Complex I intermediate-associated protein 30; *APC5*: Anaphase-promoting complex subunit 5; *CoxVIa*: Cytochrome c oxidase subunit VIa, *APN2*: DNA lyase; *40S S9*: 40S ribosomal protein S9.

Note: No conserved domains are expected for *MAT1-1-8*, *MAT1-2-5* and putative integral membrane protein (*PIM*) genes.

Table 3 – Nucleotide and amino acid conservation of mating-type and flanking genes in *Citrus*-associated *Phyllosticta* species

Gene	Nucleotide		Amino acid	
	<i>Conserved^a / total sites (bp)</i>	<i>Sequence similarity (%)^b</i>	<i>Conserved^a / total sites (aa)</i>	<i>Sequence similarity (%)^b</i>
<i>SAICARsyn</i>	693/890	77.87%	233/304	76.64%
<i>MCP</i>	891/1149	77.55%	342/382	89.53%
<i>CIA30</i>	573/698	82.09%	199/235	84.68%
<i>APC5</i>	1621/2339	69.30%	547/779	70.22%
<i>CoxVIa</i>	346/392	88.27%	108/130	83.08%
<i>APN2</i>	1386/1914	72.41%	479/642	74.61%
<i>MAT1-1-1</i>	1169/1727	67.69%	328/589	55.69%
<i>MAT1-1-8</i>	621/957	64.89%	175/329	53.19%
<i>MAT1-2-1</i>	949/1296	73.23%	275/444	61.94%
<i>MAT1-2-5</i>	708/932	75.97%	191/312	61.22%
<i>PIM</i>	1003/1890	53.07%	292/672	43.45%
<i>40S S9</i>	535/574	93.21%	189/191	98.95%

^a Identical sites in the alignment of *Citrus*-associated *Phyllosticta* species, excluding gaps

^b Sequence similarity based in the proportion of identical sites in relation to the total number of sites in the alignment

Gene abbreviations: *SAICARsyn*: Phosphoribosylaminoimidazole-succinocarboxamide Synthetase; *MCP*: Mitochondrial carrier protein; *CIA30*: Complex I intermediate-associated protein 30; *APC5*: Anaphase-promoting complex subunit 5; *CoxVIa*: Cytochrome c oxidase subunit VIa, *APN2*: DNA lyase; *PIM*: putative integral membrane protein; *40S S9*: 40S ribosomal protein S9

Figure Captions

Figure 1 – Pairwise comparison of mating-type and flaking genes between *Phyllosticta* species. The black horizontal lines represent nucleotide genomic sequences, and colour-coded arrows represent coding sequences in forward (right-oriented arrow) or reverse (left-oriented arrow) strand. Boxes between the genomic sequences represent the pairwise similarity based on BLASTn analyses, coded from red (most similar) to blue (least similar). Gene abbreviations: *SAICARsyn*: Phosphoribosylaminoimidazole-succinocarboxamide Synthetase; *MCP*: Mitochondrial carrier protein; *CIA30*: Complex I intermediate-associated protein 30; *APC5*: Anaphase-promoting complex subunit 5; *CoxVia*: Cytochrome c oxidase subunit Via, *APN2*: DNA lyase; *PIM*: putative integral membrane protein; *40S S9*: 40S ribosomal protein S9. Phylogenetical relationships between *Phyllosticta* species, based in a multilocus dataset of the internal transcribed spacers and intervening 5.8S nrDNA (*ITS*), translation elongation factor 1-alpha (*TEF1*), actin (*ACT*), and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) sequences are indicated by the cladogram on the left. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Scale bar indicates the number of substitutions per nucleotide.

Figure 2 - Bayesian Inference tree of *MAT1-1-1* and *MAT1-2-1* nucleotide sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure 3 – Bayesian Inference tree of a multilocus dataset (*ITS*+*TEF1*+*ACT*+*GAPDH*) and *PIM* nucleotide sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 are indicated with an asterisk (*) to the left of the nodes. The *PIM* tree is rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide. Conflicts between the species tree based on the multilocus dataset and the tree based on *PIM* are highlighted in blue.

Figure 4 – Ancestral state reconstruction of thallism states in *Phyllosticta* species. Characters were mapped onto a multilocus phylogenetic tree based on *ITS*, *TEF1*, *GAPDH* and *ACT* sequences. Thallism states are colour-coded according to the legend next to the tree. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 are indicated with an asterisk (*) to the left of the nodes. The tree is rooted to *Diplodia corticola* and *Neofusicoccum parvum*. Scale bar indicates the number of substitutions per nucleotide.

Figure 5 – Mating-type locus rearrangement in Citrus-associated *Phyllosticta* species. (A) Schematic representation of TE-mediated remodeling in *P. citrichinaensis*. 1) Insertion of a TE between *MAT1-1* and *MAT1-2* idiomorphs, in the ancestor with a *P. capitalensis*-like *MAT* locus. 2) Mobilization of the *MAT1-2* idiomorph could have happen either by: 2a) the recombination of the TE in the mating-type locus with a TE in a different genomic location, 2b) the transposition of the TE in the mating-type locus to a different genomic location, moving the flanking sequence containing the *MAT1-2* idiomorph. 3) This rearrangement results in unlinked *MAT* idiomorphs in *P. citrichinaensis*. (B) Schematic representation of the homothallic to heterothallic shift in *P. citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P. paracitricarpa*. 1) Partial deletions of *MAT1-1* and *MAT1-2* idiomorphs in separated strains from the ancestor with a *P. capitalensis*-like *MAT* locus, creating heterothallic individuals. 2) Degeneration of the remaining *MAT1-1-8* gene in the *MAT1-2* strains, and expansion of the idiomorphic region until the *PIM* gene in *MAT1-1* strains. 3) These changes result in heterothallism in *P. citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P. paracitricarpa*, and idiomorphic-specific versions of *PIM*. The black horizontal lines represent nucleotide genomic sequences, and colour-coded arrows represent coding sequences in forward (right-oriented arrow) or reverse (left-oriented arrow) strand. Gene abbreviations: *APN2*: DNA lyase; *PIM*: putative integral membrane protein.

Supplementary Figures Captions

Figure S1 – Pairwise comparison of mating-type and flanking genes between *Phyllosticta* species, including all strains evaluated in this study. The black horizontal lines represent nucleotide genomic sequences, and colour-coded arrows represent coding sequences in forward (right-oriented arrow) or reverse (left-oriented arrow) strand. Boxes between the genomic sequences represent the pairwise similarity based on BLASTn analyses, coded from red (most similar) to blue (least similar). Gene abbreviations: *SAICAR_{syn}*: Phosphoribosylaminoimidazole-succinocarboxamide Synthetase; *MCP*: Mitochondrial carrier protein; *CIA30*: Complex I intermediate-associated protein 30; *APC5*: Anaphase-promoting complex subunit 5; *CoxVIa*: Cytochrome c oxidase subunit Via, *APN2*: DNA lyase; *PIM*: putative integral membrane protein; *40S S9*: 40S ribosomal protein S9. Phylogenetical relationships between *Phyllosticta* species, based in a multilocus dataset of the internal transcribed spacers and intervening 5.8S nrDNA (*ITS*), translation elongation factor 1-alpha (*TEF1*), actin (*ACT*), and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) sequences are indicated by the cladogram on the left. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Scale bar indicates the number of substitutions per nucleotide.

Figure S2 – Pairwise comparison between mating-type genes and flanking intergenic regions showing homology of *MAT1-1-8* gene fragments in *MAT1-2* strains of *Phyllosticta citriasiana*, *P. citricarpa* and *P. paracitricarpa*. The black horizontal lines represent nucleotide genomic sequences, and colour-coded arrows represent coding sequences in forward (right-oriented arrow) or reverse (left-oriented arrow) strand. Boxes between the genomic sequences represent the pairwise similarity based on BLASTn analyses, coded from red (most similar) to blue (least similar).

Figure S3 - Bayesian Inference tree of *SAICAR_{syn}* (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S4 - Bayesian Inference tree of MCP (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S5 - Bayesian Inference tree of CIA30 (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S6 - Bayesian Inference tree of APC5 (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S7 - Bayesian Inference tree of CoxVIa (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S8 - Bayesian Inference tree of APN2 (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S9 - Bayesian Inference tree of MAT1-1-I (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP)

above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S10 - Bayesian Inference tree of *MAT1-1-8* (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S11 - Bayesian Inference tree of *MAT1-2-1* (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S12 - Bayesian Inference tree of *MAT1-2-5* (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S13 - Bayesian Inference tree of *40S S9* (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

Figure S14 - Bayesian Inference tree of *PIM* (a) nucleotide (b) and amino acid sequences of *Phyllosticta* species. ML bootstrap support values above 95% and Bayesian posterior probability values (PP) above 0.9 indicated with an asterisk (*) to the left of the nodes. Both trees are rooted to *Diplodia seriata*. Scale bar indicates the number of substitutions per nucleotide.

MAT1-1-1

MAT1-2-1

Multilocus**(ITS+EF+ACT+GAPDH)****PIM**

a) TE-mediated remodelling in *P. citrichinaensis*

Step 01: Insertion of a TE between *MAT1-1* and *MAT1-2* idiomorphs

Step 02: Mobilization of *MAT1-2* idiomorph either by:
 - Recombination of the TE in mating-type locus with a TE in different genomic location
 - Transposition of the TE in the mating-type locus to a different genomic location

Outcome: Unlinked *MAT* idiomorphs in *P. citrichinaensis*

b) Homothalic shift to heterothallic

Step 01: Partial deletions of *MAT1-1* and *MAT1-2* idiomorphs in different strains, creating heterothallic strains

Step 02: Degeneration of *MAT1-1-8* in the *MAT1-2* strain and expansion of idiomorphic region to *PIM* in the *MAT1-1* strains

Outcome: Heterothallism in *P. citriasiana*, *P. citribraziliensis*, *P. citricarpa* and *P. paracitricarpa*, with idiomorphic-specific versions of *PIM*