OXFORD G3 Genes | Genomes | Genetics

G3, 2021, jkab187
DOI: 10.1093/g3journal/jkab187
Advance Access Publication Date: 17 June 2021
Fungal Genetics and Genomics

Unique patterns of mating pheromone presence and absence could result in the ambiguous sexual behaviors of *Colletotrichum* species

Andi M. Wilson [1], Ruvini V. Lelwala [1,*], Paul W. J. Taylor [2], Michael J. Wingfield [1] and Brenda D. Wingfield [1]

[1]Department of Biochemistry, Genetics & Microbiology, Forestry & Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria 0083, South Africa and [2]School of Agriculture and Food, Faculty of Veterinary and Agriculture Sciences, The University of Melbourne, Parkville, VIC 3010, Australia

*Corresponding author: University of Pretoria, Corner of Lunnon/Herold Streets, Hatfield, Pretoria 0083, South Africa. Email: andi.wilson@fabi.up.ac.za .
Communicating editor: J. Dunlap

Abstract

Colletotrichum species are known to engage in unique sexual behaviors that differ significantly from the mating strategies of other filamentous ascomycete species. For example, most ascomycete fungi require the expression of both the *MAT1-1-1* and *MAT1-2-1* genes to induce sexual reproduction. In contrast, all isolates of *Colletotrichum* harbor only the *MAT1-2-1* gene and yet, are capable of recognizing suitable mating partners and producing sexual progeny. The molecular mechanisms contributing to mating types and behaviors in *Colletotrichum* are, however, unknown. A comparative genomics approach analyzing 35 genomes, representing 31 *Colletotrichum* species and two *Verticillium* species, was used to elucidate a putative molecular mechanism underlying the unique sexual behaviors observed in *Colletotrichum* species. The existence of only the *MAT1-2* idiomorph was confirmed across all species included in this study. Comparisons of the loci harboring the two mating pheromones and their cognate receptors revealed interesting patterns of gene presence and absence. The results showed that these genes have been lost multiple, independent times over the evolutionary history of this genus. These losses indicate that the pheromone pathway no longer plays an active role in mating type determination, suggesting an undiscovered mechanism by which mating partner recognition is controlled in these species. This further suggests that there has been a redirection of the underlying genetic mechanisms that regulate sexual development in *Colletotrichum* species. This research thus provides a foundation from which further interrogation of this topic can take place.

Keywords: *Colletotrichum*; sexual reproduction; filamentous fungi; mating pheromones; cognate receptors; gene loss; ancestral state reconstruction

Introduction

The genetic factors that control sexual reproduction in filamentous ascomycetes are largely conserved. This conservation is particularly notable when considering the mating-type (*MAT1*) locus and its associated genes (Kronstad and Staben 1997). Most described species require the expression of genes from both the *MAT1-1* and *MAT1-2* idiomorphs in order to sexually reproduce. Those that harbor both sets of genes within a single genome are considered homothallic and are self-fertile (Yun *et al.* 1999). Alternatively, those that possess either *MAT1-1-1* or *MAT1-2-1* are typically obligate outcrossers that require an opposite mating type partner in order to engage in sexual reproduction (Turgeon and Yoder 2000; Dyer *et al.* 2016). These species are thus heterothallic and self-sterile.

The genes present at the *MAT1* locus are directly responsible for controlling the expression of the mating pheromones in ascomycete fungi (Zhang *et al.* 1998; Bobrowicz *et al.* 2002). In the case of most heterothallic species, the *MAT1-1-1* gene positively influences the expression of the **α**-factor pheromone, while the *MAT1-2-1* gene is responsible for **a**-factor pheromone expression (Zhang *et al.* 1998; Shen *et al.* 1999; Bobrowicz *et al.* 2002; Wilson *et al.* 2018). Thus, despite the presence of both pheromone genes in the genome, they are expressed in a mating type-dependent manner. Consequently, the expression of these pheromones allows for individuals of opposite mating type to recognize one another as such, subsequently enabling physical interaction and mating (Bolker and Kahmann 1993) . This means that while mating type is determined genetically by the *MAT1* locus, it is determined physiologically by the pheromone that is being expressed.

Species of *Colletotrichum* do not exhibit either of these two typical sexual strategies and instead display ambiguous sexual behaviors that are difficult to classify. Sexual reproduction has been shown to occur in numerous *Colletotrichum* species (Vaillancourt *et al.* 2000; Guerber *et al.* 2003; Souza *et al.* 2010; Barimani *et al.* 2013), all of which harbor only the *MAT1-2-1* gene, while a *MAT1-1-1* homolog has never been identified (Vaillancourt *et al.* 2000; Rodríguez-Guerra *et al.* 2005; Menat *et al.* 2012). Consequently, all *Colletotrichum* species that have been studied to date are of a single mating type- MAT1-2. Despite this, species such as *Colletotrichum acutatum* are obligate outcrossers,

Received: February 18, 2021. **Accepted:** May 18, 2021

with heterothallic-like behavior occurring between two isolates of the same mating type (Damm *et al.* 2012). This has led to the development of various hypotheses to describe the ambiguous mating behaviours observed in *Colletotrichum* species.

One hypothesis considering mating strategies in *Colletotrichum* has been derived from studies on species including *Colletotrichum gloeosporioides* and *Colletotrichum fructicola*, and describes a unique form of heterothallism (Menat *et al.* 2012; Liang *et al.* 2021). This system, which has been referred to as "unbalanced heterothallism," is thought to have evolved from a previously homothallic ancestor that has accumulated mutations in genes responsible for self-fertility (Wheeler 1954). Consequently, many different mating identities exist and two individuals are sexually compatible if they harbor complementary mutations. Another hypothesis is that mating identity and cross fertility is determined by two, unlinked loci, rather than the single *MAT* locus as is seen in most other heterothallic species. Although this has been shown to be true in *Colletotrichum graminicola*, the actual loci have not yet been identified (Vaillancourt and Hanau 1991; Vaillancourt *et al.* 2000). In contrast, *Colletotrichum lentis* (formerly known as *Colletotrichum truncatum* or *Colletotrichum acutata*) appears to have a classical bipolar mating system that is not linked to the *MAT1* locus (Banniza *et al.* 2018) and has not yet been linked to any other genes.

To further complicate our understanding of the sexual behaviors in *Colletotrichum*, species such *C. gloeosporioides* (Weir *et al.* 2012), *C. fructicola* (Liang *et al.* 2021), *Colletotrichum fioriniae* (Guerber *et al.* 2003), and *C. graminicola* (Politis 1975; Vaillancourt and Hanau 1991) have been described as both homothallic and heterothallic. In these species, some isolates are capable of sexual reproduction when cultured alone, while other isolates require a mating partner in order to produce sexual offspring. In *C. fructicola*, this isolate-dependent behavior is thought to be facilitated by a plus-to-minus switch, with both plus and minus strains being capable of homothallic mating but preferentially undergoing heterothallic mating during periods of hyphal interaction (Liang *et al.* 2021). However, this system is not well understood, and the presence of homothallic and heterothallic isolates within a single species has made it difficult to elucidate mating type, sexual identity, and mating compatibility in these species. A further complicating factor is that the genus has undergone many taxonomic revisions (Hyde *et al.* 2009; Marin-Felix *et al.* 2017; Bhunjun *et al.* 2021), making it difficult to assign mating strategies to fungi that have been renamed and reclassified many times.

Colletotrichum is divided into 14 currently recognized species complexes (Marin-Felix *et al.* 2017; Damm *et al.* 2019; Bhunjun *et al.* 2021) that accommodate around 220 species (O'Connell *et al.* 2012; Marin-Felix *et al.* 2017; Damm *et al.* 2019). The genus includes important plant pathogens, some of which are recognized as among the top 10 fungal plant pathogens worldwide (Dean *et al.* 2012). Many *Colletotrichum* species are hemibiotrophic and have complex life cycles, switching from biotrophic to necrotrophic or endophytic states within the same life cycle (De Silva *et al.* 2017). These fungi cause a disease known as anthracnose, and infect a diverse range of hosts, including grasses such as maize (Sukno *et al.* 2008) and sorghum (Pande *et al.* 1991), fruits such as strawberry (Eastburn and Gubler 1990) and citrus (Rhaiem and Taylor 2016), perennial plants such as pyrethrum (Lelwala *et al.* 2019), as well as flowering plants like the vanilla orchid (Charron *et al.* 2018). Symptoms of infection are similarly diverse and include necrotic spots on many different parts of the plant, rotting of the crown and stem as well as seedling blight and brown blotching (Marin-Felix *et al.* 2017). It is thus important to understand the sexual behaviors of *Colletotrichum* species, the impact and importance that sexual reproduction plays in their lifecycles and the potential effect ascospore production has on their ability to spread and cause disease.

The aim of this study was to characterize the sexual ambiguity exhibited by *Colletotrichum* species using a comparative genomics approach. We confirmed the absence of the *MAT1-1-1* gene and the presence of the *MAT1-2-1* gene at the *MAT1* loci of 31 species across this genus, representing many of the recognized species complexes. In addition, we showed that these species exhibit a unique pattern of gene loss associated with genes of the pheromone response pathway, including the pheromone factors and their cognate receptors. We have shown that these genes have been lost independently, multiple times and we suggest that this could represent a redirection of the underlying genetic mechanisms that regulate the unique mating strategies observed in these fungi.

Methods

Genomes used

This study used genomes from species that represent nine of the 14 recognized *Colletotrichum* species complexes (Table 1). A total of 25 genomes and their CDS annotations were downloaded from the National Centre for Biotechnology Information (NCBI). This assemblage of genomes included 23 *Colletotrichum* species, and two *Verticillium* species that were used as outgroup species in the various analyses conducted. An additional 10 *Colletotrichum* genomes were downloaded from the NCBI and were annotated using the online version of AUGUSTUS (Stanke and Waack 2003), with default settings and *Verticillium longisporum* gene models. These 10 additional genomes represented only eight species because three different isolates of *C. truncatum* were included- *C. truncatum* 1, *C. truncatum* 2 (formerly *Colletotrichum capsici*), and *C. truncatum* 3. Thus, in total, the 35 genomes represented 31 *Colletotrichum* species and two *Verticillium* species.

Phylogenetic analyses

The six gene regions generally accepted as barcoding genes for *Colletotrichum* were used to generate a species tree (Damm *et al.* 2019). These regions included *act*, *chs-1*, *GAPDH*, *his-3*, *sod-2*, and *tub* (Supplementary Table S1). Where possible, these sequences were retrieved from Genbank. In the species for which these data were not publicly available, the homologous regions were mined from their respective genomes using the local BLASTn function in CLC Main Workbench V12.0.3. These sequences and the alignments are presented in the supplementary data (Supplementary File S1).

The gene regions were aligned using default parameters in MAFFT V7 (Katoh and Standley 2013). The alignments were subsequently subjected to model testing, using MrModelTest2 V2.4 (Nylander 2004). Thereafter, the alignments were concatenated into a single dataset and analyzed as such. Bayesian inference (BI) analyses were conducted using MrBayes V3.2.7 (Ronquist *et al.* 2012) using the best substitution model identified by MrModelTest2. The analysis was run for 1 million generations (ensuring average standard deviation of split frequencies ≤ 0.01), with 10 parallel runs and four chains. Trees were sampled every 100 generations and 25% of the sampled trees were discarded. Posterior probabilities were then calculated from the remaining trees. The relevant data have been deposited into TreeBASE (Accession number: 28260). The tree was edited in FigTree V1.4.4 (Rambaut 2018) and used in the analyses described below,

Table 1 Genomes used in this study

Species	Strain	Accession number
Acutatum species complex		
C. nymphaeae	KY 613	JABGME000000000.1
C. simmondsii	CBS 122122	JFBX00000000.1
C. fioriniae	HC 89	PNFH00000000.1
C. salicis	CBS 607.94	JFFI00000000.1
C. godetiae	C184	LZRM00000000.1
C. acutatum	KC05	LUXP00000000.1
Graminicola species complex		
C. graminicola	M1.001	ACOD00000000.1
C. sublineola	CgSI1	MQVQ00000000.1
C. falcatum	Cf671	LPVI00000000.1
Spaethianum species complex		
C. tofieldiae	CBS 168.49	LFHQ00000000.1
C. incanum	MAFF 238712	JTLR00000000.1
Destructivum species complex		
C. higginsianum	IMI 349063	LTAN00000000.1
C. shisoi	PG-2018a	PUHP00000000.1
C. tanaceti	BRIP 57314	PJEX00000000.1
No assigned complex		
C. orchidophilum	IMI 309357	MJBS00000000.1
C. coccodesa	NJ-RT1	LECQ00000000.1
C. chlorophyti	NTL11	MPGH00000000.1
Gloeosporioides species complex		
C. fructicola	CGMCC 3.17371	SSNE00000000.1
C. asianum	ICMP 18580	WOWK00000000.1
C. camelliaea	CcLH18	JAATWK000000000.1
C. gloeosporioides	Cg-14	QFRH00000000.1
Boninense species complex		
C. karstiia	CkLH20	JAATWM000000000.1
C. sansevieriaea	Sa-1-2	NJHP00000000.1
Truncatum species complex		
C. truncatum (1)	MTCC 3414	NBAU00000000.2
C. truncatum (2) *(previously C. capsici)*	KLC.C-4	JAATLN000000000.1
C. truncatum (3)	CMES1059	VUJX00000000.1
Orchidearum species complex		
C. musicolaa	LFN0074	WIGM00000000.1
C. plurivoruma	LFN00145	WIGO00000000.1
C. sojaea	LFN0009	WIGN00000000.1
Orbiculare species complex		
C. sidae	CBS 815.97	QAPF00000000.1
C. orbiculare	MAFF 240422	AMCV00000000.2
C. trifolii	543-2	RYZW00000000.1
C. spinosum	CBS 515.97	QAPG00000000.1
Verticillium (Outgroup)		
V. tricorpus	MUCL 9792	JPET00000000.1
V. dahliae	VdLs.17	ABJE00000000.1

including ancestral state reconstruction. It was also used to produce Figures 2, 3, and 6.

Gene identification and annotation

MAT *genes*

MAT1-1-1 CDS sequences and their translated protein sequences from a variety of Sordariomycetes were used in a local BLASTn and tBLASTn searches against all 33 *Colletotrichum* genomes in CLC Main Workbench V8. These included sequences from *Neurospora crassa* (M33876.1), *Trichoderma reesei* (FJ599756.1), *Fusarium graminearum* (AF318048.1), *Verticillium fungicola* (AB258383.1), and *Podospora anserina* (X64194.1).

The complete *MAT1-2-1* CDS sequence from *C. lentis* (KX932470.1) was used as a local BLASTn query against all 33 *Colletotrichum* genomes in CLC Main Workbench V8. This allowed for the identification of the *MAT1-2-1* genes as well as the *MAT1* loci in these species. Genes flanking the *MAT1-2-1* gene were used as BLASTn queries against the NCBI non-redundant nucleotide database. These results were used to determine the span and gene content of the *MAT1* locus as well as the genes that flank the locus. Genes present at the *MAT1* locus were translated and subjected to functional domain identification using the NCBI's Conserved Domains Database (Marchler-bauer *et al.* 2015).

In an effort to ensure that all of the possible *MAT* genes were annotated, the entire *MAT* locus from each of the *Colletotrichum* and *Verticillium* genomes was also subjected to a systematic, *de novo* annotation. The region, which included the two flanking genes, *APN2* and *SLA2*, was annotated using Augustus and Fgenesh, using gene models from *V. longisporum* and *Verticillium hemipterigena*, respectively. Default settings were used for both analyses. Manual annotation was conducted in the few cases where a particular gene was not predicted by either program but where significant similarity allowed for manual gene identification.

Pheromone genes

For the identification of the α-factor pheromone, the *Colletotrichum higginsianum ccg-4* mRNA sequence (XM_018306306.1) and its predicted protein sequence were used as queries in local BLASTn and

tBLASTn searches against all 35 genomes. In species where the gene had not been previously annotated, the BLAST result was used to manually annotate the region likely encoding the pheromone.

Where no usable BLAST results were obtained, a microsynteny approach was used to locate the pheromone gene. Multiple genes found in the 5' and 3' regions flanking the α-factor pheromone in the *C. higginsianum* genome were extracted and translated into their predicted proteins. These gene and protein sequences were then used as queries in local BLASTn and tBLASTn searches against the remaining genomes. Where relevant, the intergenic regions were subjected to manual identification and annotation.

The predicted protein sequences of the α-factor pheromones were analyzed using ExPASy ProtScale (Gasteiger *et al.* 2005) and Phobius (Krogh *et al.* 2007) to confirm that they possessed the two functional domains associated with the α-factor pheromone- an N-terminal hydrophobic region as well as a signal peptide. The sequences were also manually searched for the KEX2 cleavage sites (KR) which flank repeats of an 11 aa region that represents the mature pheromone sequence (Martin *et al.* 2011).

The **a**-factor pheromone genes were also identified via a microsynteny-based approach, using the *N. crassa* **a**-factor pheromone and its flanking genes. The **a**-factor pheromone gene in *N. crassa* is flanked by a cyanase gene (XP_011394492.1) and a hypothetical, EBP2 superfamily protein (XP_960774.1). The predicted protein sequences of these two genes were used as queries in local tBLASTn searches against all 35 genomes. As above, the intergenic regions were subjected to manual identification and annotation. All predicted full-length protein sequences were confirmed to have the defining feature of the **a**-factor pheromone- the C-terminal CaaX domain. This domain is comprised of a cysteine residue, followed by two aliphatic amino acids and an amino acid of any type (Martin *et al.* 2011).

Pheromone receptor genes

The **aα**-factor (*preB*, XM_018304390.1) as well as the **a**-factor (*preA*, XM_018297638.1) pheromone receptor mRNA sequences and their corresponding protein sequences from *C. higginsianum* were used in various local BLASTn and tBLASTn searches against all 35 genomes.

Ancestral state reconstruction

In order to determine the ancestral state of the presence or absence of the pheromones and their receptor genes, ancestral state reconstruction was conducted using the unordered, parsimony model in Mesquite V3.61 (Maddison and Maddison 2011). Absence (0) or presence (1) of the genes was mapped onto the BI species tree produced above.

RIP analysis

All 35 whole genomes were subjected to analysis using The RIPper, a web-based tool that identifies regions in the genome that have been subjected to repeat induced point (RIP) mutations (Van Wyk *et al.* 2019). Default parameters were used for all analyses; including those for window and slide size, as well as the composite, product, and substrate values. The average genome-wide GC content and RIP percentages as well as the number of Long RIP Affected Regions (LRAR) and their average GC content were calculated per genome.

Results

The *MAT1* locus is conserved across the genus

In all the *Colletotrichum* species considered, the *MAT1* locus was flanked by the DNA lyase (*APN2*) and cytoskeleton assembly control (*SLA2*) genes (Figure 1). In all cases, the *MAT1* locus harbored the *MAT1-2-1* gene and a previously undescribed secondary *MAT1-2* gene. The locus orientation was also highly conserved, with the *MAT1-2-1* gene found directly downstream of the *APN2* gene (±280–1 100 nt). The only exception was *Colletotrichum chlorophyti*, where the locus is inverted such that the *MAT1-2-1* gene was instead found directly downstream from the *SLA2* gene (±1 290 nt).

All of the *Colletotrichum MAT1-2-1* genes encoded a protein possessing the HMG (high mobility group) box domain, which is the functional domain associated with this protein in all ascomycete fungi. The secondary *MAT1-2* gene present at this locus had not been previously characterized or named and did not show any similarity to other previously described *MAT1-2*-associated genes. It also did not possess a recognizable functional domain. It has thus been provided with the name *MAT1-2-15* in accordance with accepted *MAT* gene nomenclature (Wilken *et al.* 2017). Consistent with all previously published research, the *MAT1-1-1* gene was not identified in any of the *Colletotrichum* genomes considered in this study; neither at the *MAT1* locus nor elsewhere in the genome.

The α-pheromone gene has undergone multiple, independent loss events

The α-pheromone was identified in 17 of the 31 *Colletotrichum* species and both *Verticillium* species considered in this study (Supplementary Table S2). All of the identified α-pheromone genes encoded proteins that exemplified the expected structure of the ascomycete α-pheromone factor. In this regard, they possessed a hydrophobic region and signal peptide at the N-terminal. Furthermore, they harbored numerous KEX2 processing sites (KR) as well as one or more repeats of the mature, 11 aa pheromone factor (Supplementary Table S2).

The gene encoding the α-pheromone was found at a fairly conserved location within the genomes of these species, although multiple rearrangements had taken place at this locus (Figure 2). The region downstream of the pheromone gene was highly conserved, with a DUF1640 domain-containing protein present in all species considered. Various other genes encoding known, uncharacterized, and hypothetical proteins (Figure 2, Supplementary Table S3) were found at this locus in species that do not harbor a gene encoding for the pheromone.

In contrast to the conserved nature of the downstream region, the genes found upstream of the α-pheromone varied amongst species (Figure 2). An integral membrane protein as well as a major facilitator superfamily transporter were present in all species considered, but their orientation and location differed between the various species complexes. The presence of a FAD binding domain-containing protein was also conserved at this locus, although it had been lost in *Colletotrichum orchidophilum* and in the species accommodated in the acutatum species complex.

The ancestral state reconstruction analyses showed that the α-pheromone was likely present in the ancestor of *Colletotrichum* and *Verticillium* (Figure 3). While this reconstruction showed seven loss events, investigation into the genetic imprints left by these losses indicates more independent loss events. Thus, at least 10 genetic events have occurred at the α-pheromone locus during the evolutionary trajectory of *Colletotrichum*, each of which

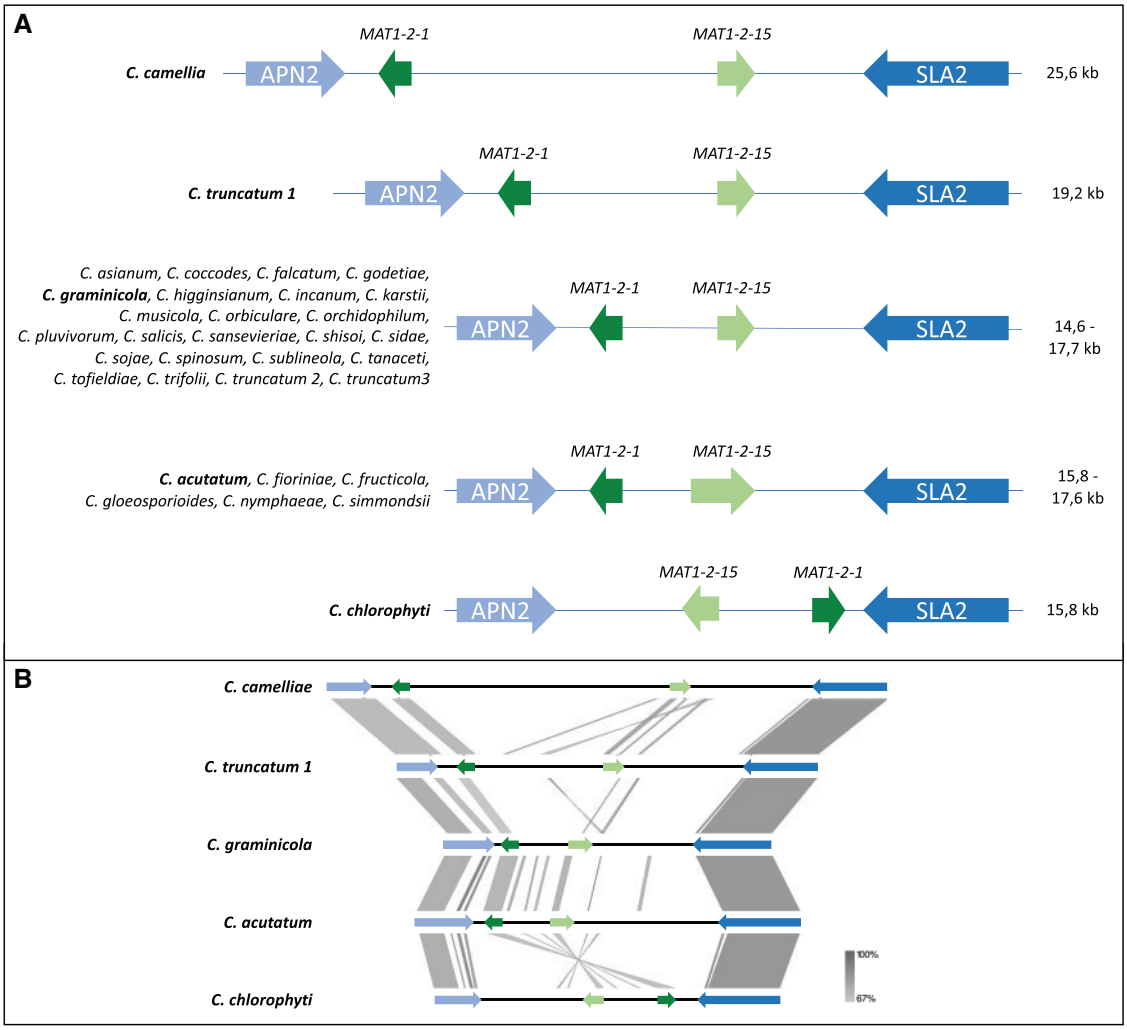

Figure 1 The different *MAT1* locus configurations of *Colletotrichum* species. In all of the species considered, the locus is flanked by the *APN2* and *SLA2* genes, a common configuration in Pezizomycotina species (A). The first configuration is found in *Colletotrichum camelliae*, which harbors a greatly expanded *MAT1* locus (25.6 kb). The second configuration is that of *C. truncatum* 1, which is also expanded (19.2 kb). Most of the remaining species exhibit the third or fourth configuration, with the major difference being the position of the *MAT1-2-15* gene's start codon. Lastly, the *MAT* locus of *C. chlorophyti* has been inverted. Representative species of these configurations were used to generate synteny maps as indicated in bold text (B). There is a high level of similarity in the flanking genes, as well as within the *MAT1-2* genes themselves. However, the rest of the locus is not well conserved. These diagrams are not drawn exactly to scale but do illustrate the sizes of the loci relative to one another.

resulted in the loss of the α-pheromone (Figure 4). These events included small changes such as point mutations disrupting the start codon or introducing a premature stop codon. In addition, larger translocations, insertions and/or deletions also occurred and resulted in completely different genes being encoded at this locus.

The first predicted loss event occurred at the split between the orbiculare species complex and all other species of *Colletotrichum*. In all four species considered from this complex, the α-factor pheromone was absent, and instead, two genes encoding hypothetical proteins are present (Figures 2 and 4). Despite this, the locus was not remarkably different in length compared to the other species. There is also no significant similarity between the region encoding the α-pheromone in any of the other *Colletotrichum* species and those of the orbiculare complex. It is thus likely that a translocation removed the α-factor pheromone gene and its immediate flanking sequences and replaced these with the two hypothetical protein genes.

The second loss event occurred within the truncatum species complex (Figure 3). While the *C. truncatum* 1 and *C. truncatum* 3 genomes harbored a gene encoding the α-pheromone, the *C. truncatum* 2 (formerly *C. capsici*) genome did not. In fact, in *C. truncatum* 2, genes encoding the FAD-linked oxidoreductase, major facilitator transporter, and the DUF1640 domain-containing protein were not present on the same contig and are flanked by other genes. It is thus not clear how the gene was lost, but it could have been a large rearrangement or translocation coupled with a deletion.

The loss of the α-pheromone gene in *Colletotrichum coccodes* represented the third loss of this gene during the evolution of *Colletotrichum* (Figure 3). While the locus was conserved between *C. coccodes* and species that do harbor the gene, a point mutation has disrupted the start codon of the putative *C. coccodes* α-factor pheromone gene (Figure 4). The start codon was mutated from ATG to ACG and would likely not support translation of the protein. Furthermore, the region downstream of the disrupted start codon did not code for any recognizable mature repeats, despite harboring a number of potential KEX2 cutting sites. It is thus likely that this gene is undergoing pseudogenization.

A fourth loss event has occurred in *Colletotrichum incanum* which is accommodated in the spaethianum complex (Figure 3).

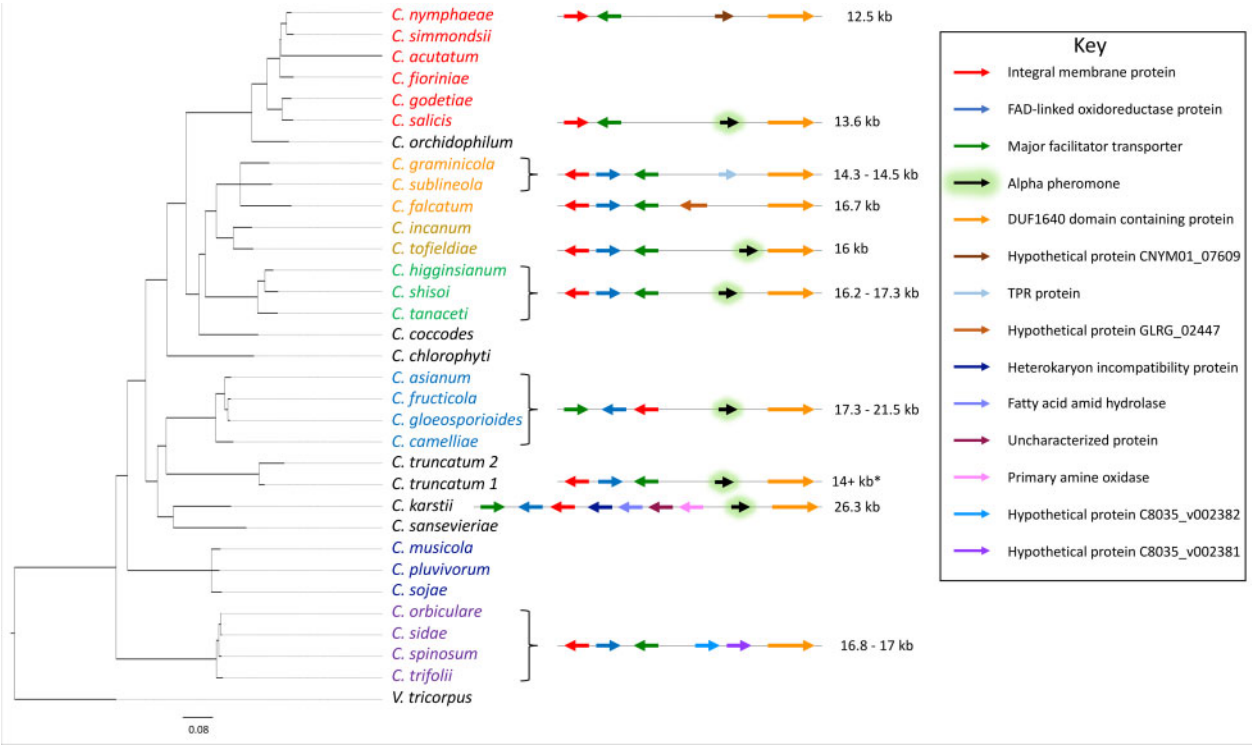

Figure 2 The diversity found at the **α**-factor pheromone locus in the *Colletotrichum* genus. Representative gene maps illustrate the differences in gene content, position, and orientation at this locus. Arrows of the same color represent the gene as indicated in the key. These gene maps are not drawn to scale, but the length of the illustrated regions are indicated to the right of each map. Species clades are indicated in the following colors: acutatum (red), graminicola (orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), orchidearum (dark purple), and orbiculare (light purple). Refer to Supplementary Table S3 for details regarding gene identities and accession numbers. The phylogeny illustrated here was derived from the phylogenetic analyses performed in this study. *The length of the *C. truncatum* 1 locus is indicated at 14+ kb as the locus is found on a scaffold that is not complete.

The pheromone region (from integral membrane protein to DUF1640 domain-containing protein) was almost 20 kb in *C. incanum* and only just over 16 kb in *Colletotrichum tofieldiae*. Despite the much smaller locus, an intact **α**-pheromone gene was identified from the *C. tofieldiae* locus. Interestingly, the encoded protein was much smaller than that of the other *Colletotrichum* species, at only 91 aa long (compared to the others 191–335 aa). Furthermore, this protein harbors only a single mature pheromone repeat, compared to the two to seven repeats in the other species (Supplementary Table S2). It was particularly relevant that the terminal 43 nt at the 3' region of the *C. tofieldiae* **α**-pheromone gene were well conserved in the *C. incanum* (Figure 5). However, the 5' region of this gene was not conserved between the two species and no definable start codon was present in *C. incanum*. It is therefore possible that there has been an insertion or translocation into the *C. incanum* locus that has replaced the 5' region of **α**-pheromone in this species.

The **α**-pheromone gene was also lost by species residing in the graminicola species complex (Figure 3). In these species, an entirely different gene was present at this locus. In both *C. graminicola* and *Colletotrichum sublineola*, a gene encoding a tetratricopeptide (TTP) protein was present, while this locus in *Colletotrichum falcatum* encoded a hypothetical protein (Figure 4). Similar to the loss in the orbiculare complex species, it is probable that a translocation resulted in the deletion of the **α**-pheromone factor gene and the insertion of new sequences.

Further gene loss events occurred within the acutatum species complex (Figures 3 and 4). The first of these events occurred in *Colletotrichum godetiae*, where a single nucleotide polymorphism (SNP) lead to a nonsense mutation. The putative protein has thus been truncated to only 43 aa. The next two loss events occurred in *C. fioriniae* and *C. acutatum* and both involved the disruption of the start codon, which had been mutated to ATA and ACG, respectively. The final loss event left no genetic imprint and was likely a similar event to the loss of the gene in the graminicola and orbiculare species complexes.

The **α**-factor pheromone flanking genes in *Colletotrichum sansevieriae* were identified at the ends of two different contigs that could not be joined. Therefore, an intact pheromone was not detected in this species. However, the genes as well as the existing intergenic regions are highly conserved between *C. sansevieriae* and *Colletotrichum karstii*, suggesting that the pheromone is likely to present.

The a-factor pheromone exhibits less extreme patterns of gene loss

The gene encoding the **a**-factor pheromone was identified in a total of 22 *Colletotrichum* species and both *Verticillium* species (Supplementary Table S4) and was found in the same genomic location in all of them. The gene found directly upstream of the pheromone encodes a cyanate hydratase protein (*CHP*), while the gene directly downstream of the pheromone was eukaryotic rRNA processing protein (*EBP2*). In many of the ten species where the **a**-factor pheromone has been lost, these flanking genes occurred next to one another in the same orientation as those that retained the pheromone gene.

The defining features of the **a**-factor pheromone are its short length and the presence of a highly conserved C-terminal CaaX

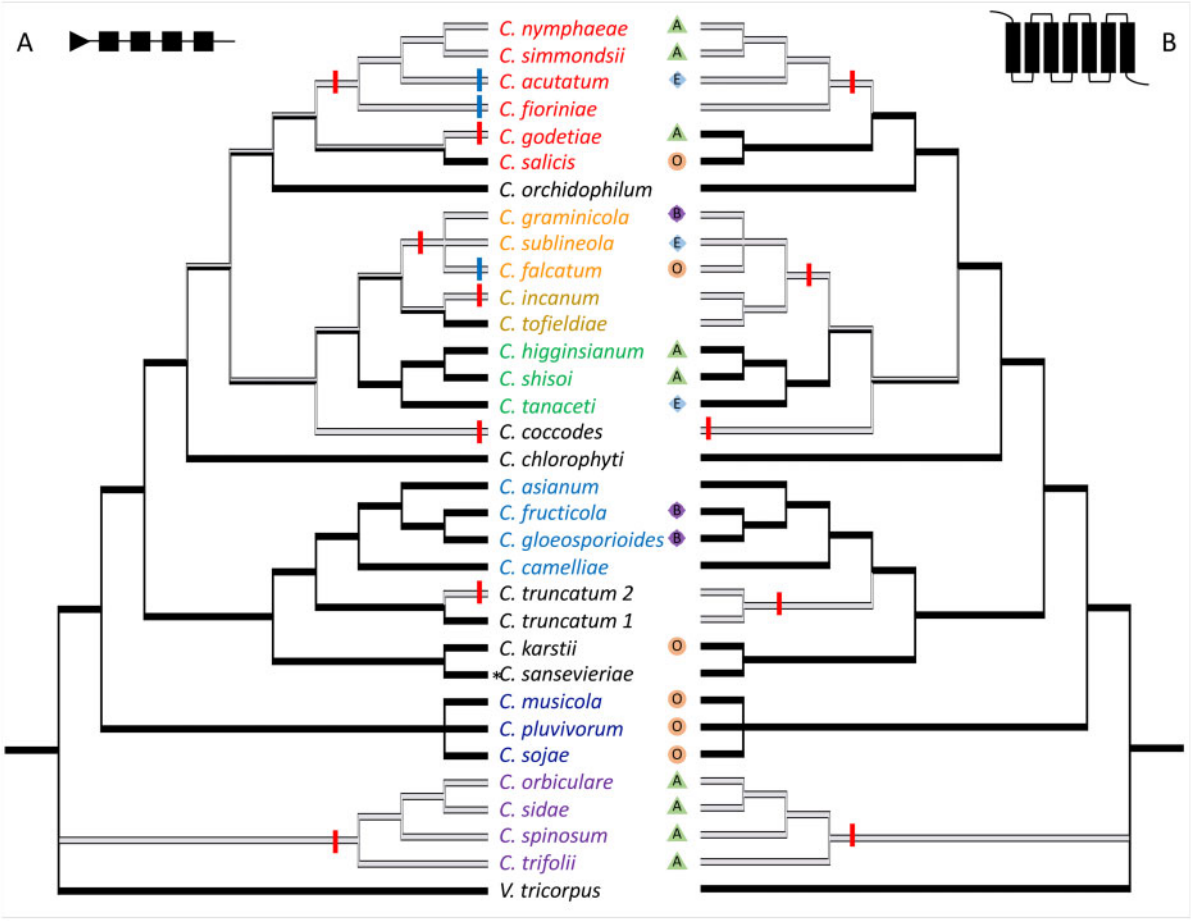

Figure 3 Ancestral state reconstruction of the presence/absence of the **α**-factor pheromone (A) and **α**-factor receptor (B) genes. Black lines represent the presence of the gene, while gray lines represent the absence of the gene. Red bars signify putative gene loss according to ancestral state reconstruction. Further genetic analysis showed that certain loss events are comprised of multiple independent losses and these are indicated by blue bars. Sexual capability/strategy is indicated next to species names where they are known: Asexual (A in a green triangle), heterothallic (E in a blue diamond), homothallic (O in an orange circle) and both heterothallic and homothallic (B in a purple shape). Species clades are indicated in the following colors: acutatum (red), graminicola (orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), orchidearum (dark purple), and orbiculare (light purple). The phylogeny illustrated here was derived from the phylogenetic analyses performed in this study. *This gene could not be identified due to the flanking genes being present at the ends of different contigs that could not be joined.

domain. In the *Colletotrichum* species, the sequence of this conserved domain was either CVIL or CVVM (Supplementary Table S4). Although the genes all encode short proteins, there was some variation in the size of the **a**-factor pheromone protein. For example, *Colletotrichum musicola* and *Colletotrichum plurivorum* produced the longest proteins at 109 aa, while species belonging to the orbiculare species complex produced a protein of only 35 aa.

A parsimony-based ancestral state reconstruction showed that the **a**-factor pheromone was present in the ancestor of *Colletotrichum* and *Verticillium*, but that it has subsequently been lost as the species of *Colletotrichum* have evolved (Figure 6). Although the ancestral state reconstruction showed a total of four loss events, the genetic imprints left during these losses suggest that at least eight independent deletion events have occurred. These events are similar to those that occurred for the **α**-factor pheromone.

The first loss of the **a**-factor pheromone occurred in *Colletotrichum sojae* (Figure 7). Although the *C. sojae* locus showed high levels of similarity to this locus in other species residing in the orchidearum species complex, a 7 nt deletion in the coding sequence resulted in a frameshift mutation. The mutation disrupted the CaaX domain and stop codon. Therefore, any protein encoded by this putative pheromone gene would likely not function as required.

The next **a**-factor pheromone loss occurred in *C. truncatum* 2 (Figure 6). While the gene encoding the upstream flanking gene, CHP, is present, the downstream flanking gene is unidentifiable in the *C. truncatum* 2 genome. This locus has thus been entirely disrupted in *C. truncatum* 2, leading to the loss of the **a**-factor pheromone as well. Similarly, no gene encoding this pheromone could be found in *C. chlorophyti* either, representing the third loss of this gene in *Colletotrichum* (Figure 6). In this case, the up- and downstream flanking genes are conserved, yet the **a**-factor pheromone could not be identified in this region or elsewhere in the genome.

The loss of the **a**-factor pheromone gene in *C. coccodes* represented the fourth loss of this gene in *Colletotrichum* species (Figures 6 and 7). In this case, a mutation disrupted the start codon of the gene, changing the start codon from ATG to AGG. The remaining gene sequence was similar to that of closely related species but does not encode for a protein with the terminal CaaX domain.

This pheromone was also lost in the acutatum species complex (Figure 6). However, genetic analysis showed that at least four different deletion events have taken place in this species complex (Figure 7). The first of these occurred in *C. godetiae* and *Colletotrichum salicis*, where a disruption of a start codon was

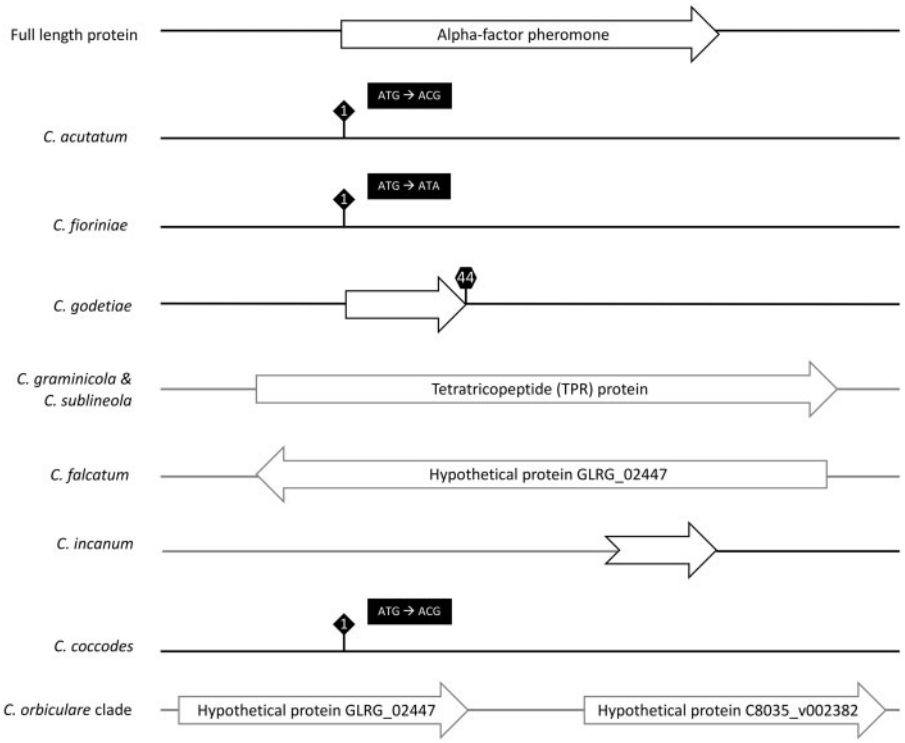

Figure 4 Genetic events that resulted in loss of the **α**-factor pheromone in *Colletotrichum* species clades. In *C. acutatum*, *C. fioriniae* and *C. coccodes*, an SNP has disrupted the start codon, shown by ◆. In *C. godetiae*, a premature stop codon was introduced at amino acid position 44, shown by ■. Only the 3' region of the **α**-factor gene can be identified in *C. incanum*. In various other species, entirely different genes are present at this locus, likely arriving via translocation and replacing the **α**-factor pheromone. These diagrams are not drawn to scale.

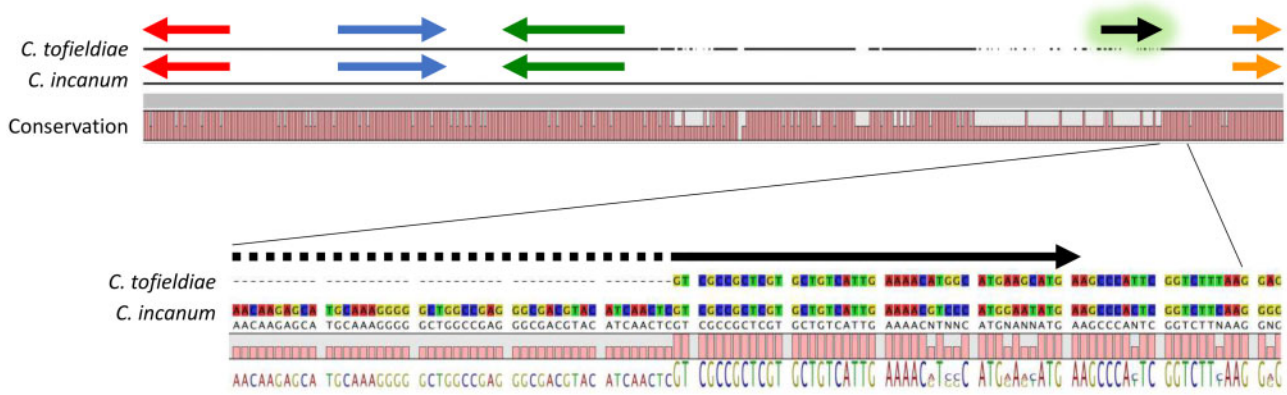

Figure 5 The loss of the **α**-factor pheromone gene in *C. incanum*. Although sequence with high similarity to the 3' region of the gene in *C. tofieldiae* can be found in *C. incanum*, the 5' region of the gene is unidentifiable. In this region in *C. incanum*, it appears that an insertion occurred, displacing part of the **α**-factor pheromone as there is no homologous region at this locus in *C. tofieldiae*. Gene arrow colors are the same as those in Figure 2.

responsible for the loss of this gene. In this case, the start codon was mutated from ATG to GCG and has thus undergone at least two mutation events. Interestingly, a gene sequence similar to that of other species' **a**-factor pheromone was detectable after the disrupted start codon and a CaaX domain can be identified. Although there was an alternative start codon, it was found halfway through the gene sequence. Furthermore, there was no stop codon encoded after the CaaX domain and thus any potentially translated protein would be very different from a typical **a**-factor pheromone.

The second loss to occur in the acutatum species complex involves the **a**-factor pheromone gene of *C. acutatum* (Figure 7). In this case, a point mutation has introduced a premature stop codon into the gene sequence and would prematurely terminate translation of the putative protein before the terminal CaaX domain, producing a protein of only 31 aa. Interestingly, the third loss involved a similar mutation and occurred in many of the other species residing in the acutatum species complex (Figure 7). In these species, the premature stop codon occurred at position 35. Lastly, the gene was lost in *Colletotrichum simmondsii* due to another disruption of the start codon, which was mutated from ATG to ACG (Figure 7).

The orbiculare species complex should be mentioned although this gene has not been entirely lost in these species. The **a**-factor pheromone locus in these species has undergone a deletion of 15 bp that corresponds to the first five amino acids of the **a**-factor pheromone (Figure 7). An alternative start codon was available, and since no functional domain is known from the N-

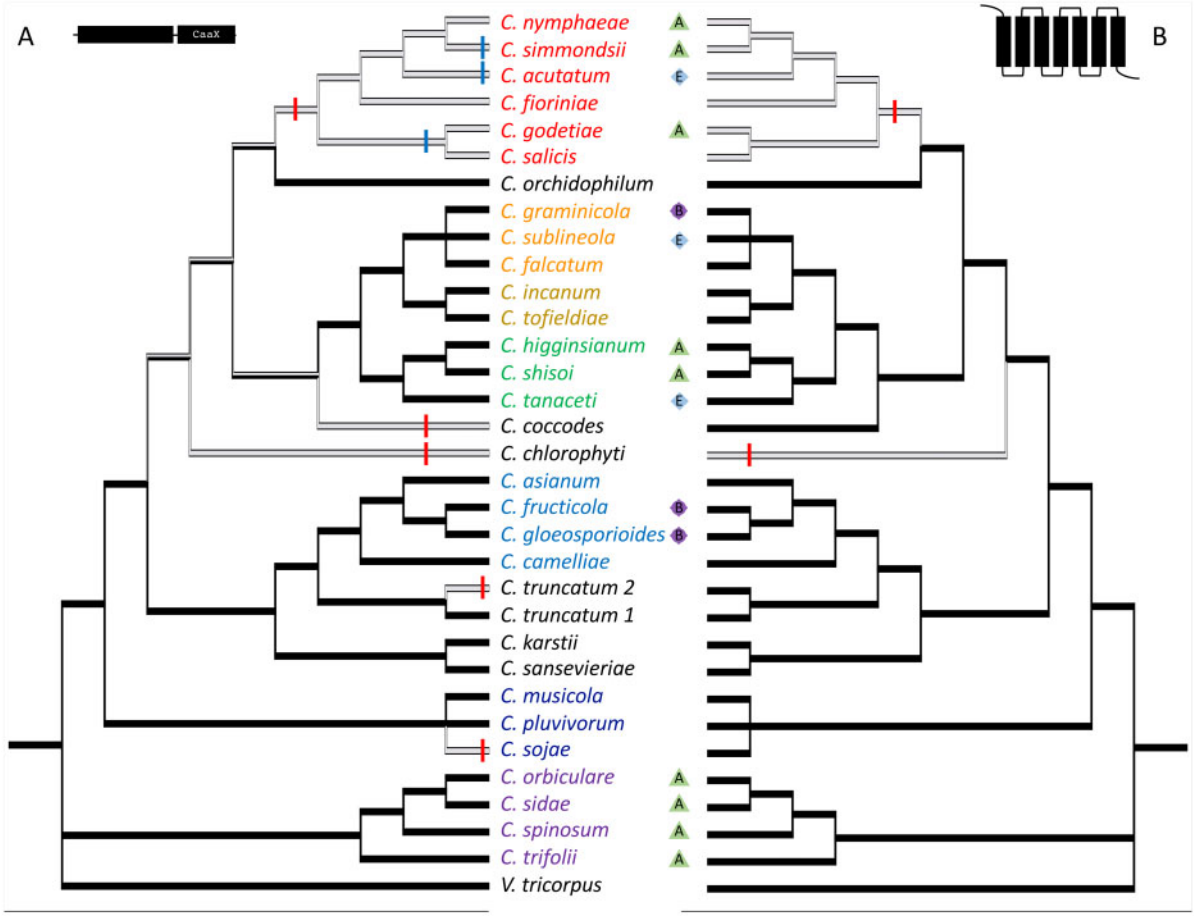

Figure 6 Ancestral state reconstruction of the presence/absence of the **a**-factor pheromone (A) and **a**-factor receptor (B) genes. Black lines represent the presence of the gene, while gray lines represent the absence of the gene. Red bars signify putative gene loss according to ancestral state reconstruction analysis. Further genetic analysis showed that certain loss events are comprised of multiple independent losses and these are indicated by blue bars. Sexual capability/strategy is indicated next to species names where they are known: Asexual (A in a green triangle), heterothallic (E in a blue diamond), homothallic (O in an orange circle), and both heterothallic and homothallic (B in a purple shape). Species clades are indicated in the following colors: acutatum (red), graminicola (orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), orchidearum (dark purple), and orbiculare (light purple). The phylogeny illustrated here was derived from the phylogenetic analyses performed in this study.

terminal of this protein and the C-terminal CaaX domain is still intact, this protein was likely to act as a functional pheromone. However, this would need to be confirmed experimentally.

The cognate receptors display similar patterns of presence/absence as their pheromones

In a similar manner to their respective pheromones, genes encoding the pheromone receptors could be identified in the genomes of only a subset of all *Colletotrichum* species. In almost all cases, the presence (or absence) of a pheromone gene correlated to the presence (or absence) of the gene encoding the pheromone's cognate receptor. The receptors, therefore, showed similar patterns of gene presence and absence as their pheromones (Figures 3 and 6). However, there were a few exceptions to this pattern.

With regards to the **α**-factor pheromone and its cognate receptor (*preB*), the exceptions included: *C. truncatum* 1 and *C. tofieldiae*, where the pheromone was present while the receptor was absent, and *C. godetiae*, where the pheromone was absent while the receptor was present (Figure 3). When the **a**-factor and its cognate receptor (*preA*) are considered, three exceptions exist. These were found in *C. sojae*, *C. coccodes*, and *C. truncatum* 2 which, while harboring an intact pheromone receptor, do not possess the corresponding pheromone (Figure 6).

In species where the pheromone receptors were absent, these genes had been lost in similar ways to the pheromone genes. This was particularly true in the case of the **α**-factor receptor (*preB*), where disruptions to the start codon and the presence of in-frame, premature stop codons were common across the genus (Figure 8). Interestingly, these genes have also experienced small to large indel events, a few of which resulted in frameshift mutations, which lead to premature stop codons. Similar genetic events could not be identified at the locus harboring the **a**-factor receptor (*preA*), and it is thus likely that these genes were lost via translocation, insertion, and/or deletion events that left no identifiable trace in the genome.

Colletotrichum species show highly varied levels of genome-wide RIP

Almost all of the genomes considered in this study showed some level of RIP (Table 2), which varied greatly from 0.93% (*C. nymphaea*) to 56.02% (*Colletotrichum trifolii*). The only species that displayed no genomic evidence of RIP was *C. sublineola*. In general, species residing in the same species complexes showed similar RIP profiles.

The overall GC content of the *Colletotrichum* genomes was fairly consistent, at around 50% in most of the genomes considered. However, in species where there was significant RIP (>30%),

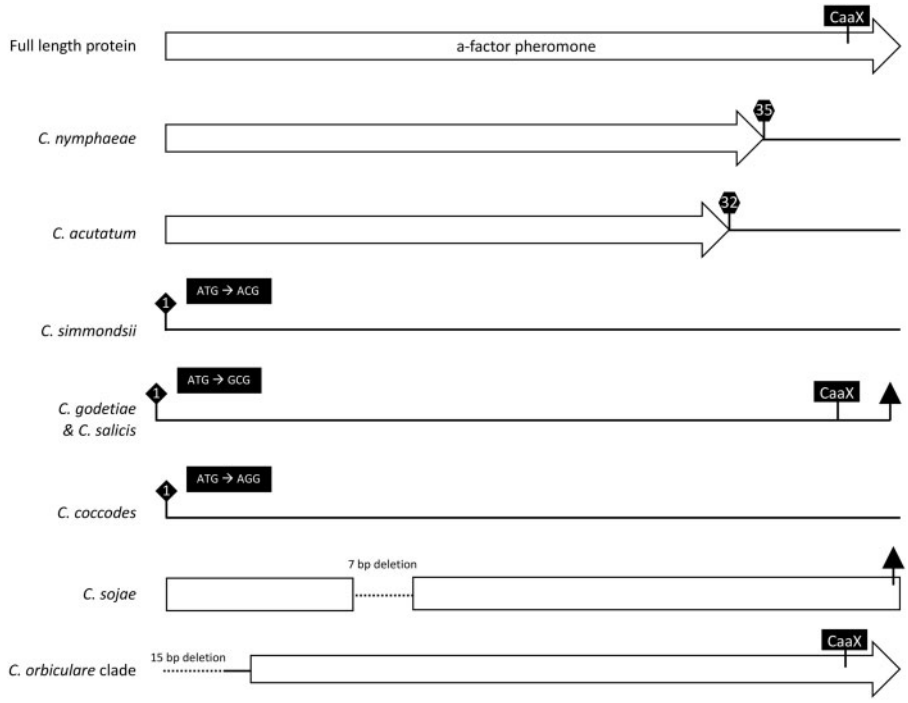

Figure 7 Genetic events that resulted in loss of the **a**-factor pheromone in *Colletotrichum* species clades. A premature, in-frame stop codon truncates the protein in *C. nymphaeae* before the functional CaaX domain, as shown by ■. Other *C. acutatum* clade species and *C. coccodes* have disrupted start codons as indicated by ◆. *C. godetiae* and *C. salicis* as well as *C. sojae* are missing a stop codon, as indicated by ▲. Interestingly, the *C. orbiculare* clade species have a 15 bp deletion in the 5' region of the gene, abolishing the start codon that is shared by many of the other species. However, an alternative start codon is present and thus a slightly shorter protein harboring the CaaX domain can likely be produced. These diagrams are not drawn to scale.

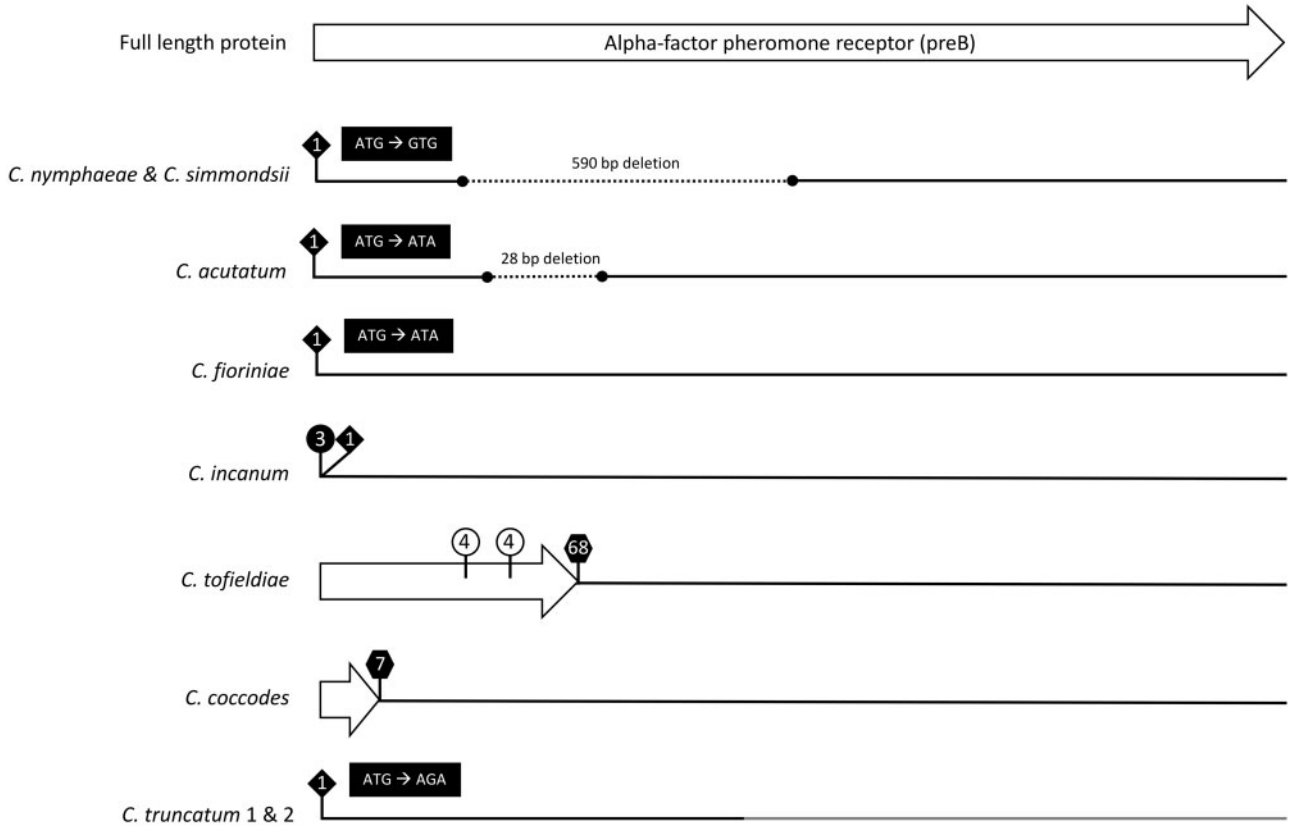

Figure 8 Genetic events that disrupted the **α**-factor pheromone receptor (*preB*) gene. In many of the species in which this gene has been disrupted, a SNP has abolished the start codon, as shown by ◆. In others, small indels resulted in a frameshift mutation that subsequently produce an in-frame stop codon as shown by ■. The number in the hexagon refers to the amino acid position at which the stop codon occurs. Deletions are shown by ■ while insertions are shown by ●. The number inside the circle refers to the number of nucleotides inserted or deleted. These diagrams are not drawn to scale.

Table 2 The RIP profiles and GC content of the genomes used in this study

Species	RIP %	GC content	No of LRAR	GC content of LRARs
Acutatum species complex				
C. nymphaeae	0.93	52.75	942	30.11
C. simmondsii	3.97	51.73	26	24.44
C. fioriniae	1.42	52.49	11	35.53
C. salicis	0.99	52.78	3	38.52
C. godetiae	2.89	52.09	61	26.97
C. acutatum	2.71	52.25	45	25.89
Graminicola species complex				
C. graminicola	18.09	48.42	545	25.15
C. sublineola	0	53.11	0	0
C. falcatum	10.01	52.78	230	27.76
Spaethianum species complex				
C. tofieldiae	4.8	52.99	110	28.12
C. incanum	4.64	51.9	117	33.34
Destructivum species complex				
C. higginsianum	3.27	54.41	117	25.89
C. shisoi	35.92	46.26	284	28.8
C. tanaceti	24.1	49.19	674	31.1
No assigned complex				
C. orchidophilum	6.22	51.06	92	18.58
C. coccodes	1.31	53.79	0	–
C. chlorophyti	10.85	49.8	214	25.8
Gloeosporioides species complex				
C. fructicola	1.37	53.2	39	19.7
C. asianum	14.32	49.37	485	23.58
C. camelliae	10.76	50.01	361	20.9
C. gloeosporioides	2.61	51.98	107	19.7
Boninense species complex				
C. karstii	3.75	52.69	101	23.4
C. sansevieriae	11.53	50.74	38	34.88
Truncatum species complex				
C. truncatum (1)	3.9	49.73	62	27.49
C. truncatum (2) *(previously C. capsici)*	5.19	48.26	123	26.72
C. truncatum (3)	2.85	50.12	122	35.63
Orchidearum species complex				
C. musicola	5.72	54.94	133	34.87
C. plurivorum	0.99	55.84	15	24.47
C. sojae	1.02	55.9	14	30.3
Orbiculare species complex				
C. sidae	47.37	37.96	613	19.62
C. orbiculare	49.53	36.47	1067	17.91
C. trifolii	56.02	35.24	1045	20.85
C. spinosum	45.32	38.7	588	19.36
Verticillium (Outgroup)				
V. tricorpus	2.12	57.24	32	32.38
V. dahliae	2.59	55.34	24	28.9

the GC content was considerably lower. For example, members of the orbiculare species complex had some of the highest levels of RIP (45–56%) and, correspondingly, the lowest GC levels (35–38%).

Data availability

The accession numbers of the genomes used in this study are detailed in Table 1, while the accession numbers of the DNA sequences used for the phylogenetic analyses are included in Supplementary Table S1. In addition, Supplementary File S1 contains the sequence data used in the phylogenetic analyses done in this study. The tree files from these analyses were also uploaded to TreeBASE and can be found under the accession number: 28260. Supplemental Material available at figshare: https://doi.org/10.25387/g3.14672052.

Discussion

It has long been apparent that the sexual behaviors of *Colletotrichum* species are unique amongst filamentous ascomycetes (Politis 1975; Vaillancourt *et al.* 2000; Menat *et al.* 2012). These species exhibit ambiguous sexual preferences, with individual isolates in some species having different sexual strategies, as is found in *C. fructicola*, *C. gloeosporioides*, and *C. graminicola* (Vaillancourt and Hanau 1991; Vaillancourt *et al.* 2000; Menat *et al.* 2012; Liang *et al.* 2021). In this study, we have shown that genes encoding the mating pheromones and their cognate receptors displayed an interesting pattern of presence and absence. This might explain the ambiguous and unpredictable nature of sexual reproduction in *Colletotrichum*.

The only mating-type genes identified in the 31 *Colletotrichum* species considered in this study were associated with the *MAT1-2* idiomorph. While the *MAT1-2* locus of these species is comparable to that in other ascomycetes, the complete absence of *MAT1-1*-associated genes in this genus was the first genetic indicator suggesting that sexual reproduction in *Colletotrichum* is different to that of other filamentous ascomycetes (Vaillancourt *et al.* 2000; Rodríguez-Guerra *et al.* 2005; Menat *et al.* 2012; O'Connell *et al.* 2012). The absence of any *MAT1-1*-associated genes suggests that mating type or sexual identity is no longer genetically determined by the *MAT1* locus in these species. While the *MAT1-2-1* and

MAT1-2-15 genes probably remain important for regulating sexual development, it is unlikely that they confer mating type as they do in other species.

The loss of the *MAT1-1* idiomorph and its associated genes from all species of *Colletotrichum* does not appear to have affected the sexual capabilities of a number of these species, which continue to exhibit various forms of homothallic and heterothallic behavior. These species can thus be described as unisexual, as only a single mating type is required for sexual development (Roach *et al.* 2014). The ambiguous sexual behaviors observed in this genus thus might partly be explained by evolution toward unique sexual strategies that better suit these species and the environments in which they occur. Similar transitions have been observed in other species that exhibit unisexuality, which is thought to be a response to environments where locating a suitable mating partner is difficult (Wang and Lin 2011; Wilson *et al.* 2015).

The sexual ambiguity and unpredictable nature of mating strategies in species of *Colletotrichum* can also be linked to the patterns of gene loss associated with the pheromone factors and their cognate receptors. Only 14 *Colletotrichum* species harbored genes encoding both pheromone factors. Of the remaining species, 11 harbored only one of the pheromones and a further 6 possessed neither. This is in stark contrast to many other studied filamentous ascomycete fungi (Zhang *et al.* 1998; Shen *et al.* 1999; Pöggeler 2000; Bobrowicz *et al.* 2002; Turina *et al.* 2003; Mayrhofer and Pöggeler 2005; Mayrhofer *et al.* 2006; Lee *et al.* 2008; Martin *et al.* 2011; Wilson *et al.* 2018) and strongly suggests that mating identity and preference is not under the control of the *MAT1* locus and the pheromone pathway in these species. In addition, both homothallism and heterothallism have been documented in these species, behaviors that cannot be linked to a specific pheromone absence/presence pattern. This provides further evidence that the presence and absence of *MAT* genes and the relevant pheromone pathway genes is not predictive of sexual capabilities or behaviors in *Colletotrichum* species.

Ancestral state reconstruction and further genetic analysis showed that both pheromone factors have been lost multiple times during the evolutionary trajectory of *Colletotrichum*. These results show that the **α**-factor was lost at least ten independent times, while the **a**-factor was lost at least eight times. While the genetic mechanisms underlying each loss event varied, the outcome was consistently an inability to produce mating pheromone factors. These concerted loss events strongly suggest that there has been a redirection of the underlying genetic mechanisms that regulate mating type and sexual compatibility in *Colletotrichum* species. While the pheromone factors facilitate mating partner recognition in other heterothallic species, the same cannot be true in the *Colletotrichum* species.

With a few notable exceptions, the pheromone genes had similar patterns of presence or absence as their cognate receptors. The exceptions include a number of species that either possess the pheromone and not the receptor or *vice versa*. The functions of these genes are thus unclear as the absence of a pheromone would make the receptor redundant. It is possible that the genes have since been co-opted for a different function, potentially unrelated to mating. Given the notoriety of *Colletotrichum* species as plant pathogens, it is possible that these proteins play a role in virulence or plant host interaction. This kind of functional diversification has been documented in *Fusarium oxysporum*, where the **α**-factor receptor is intimately involved in the chemotrophic attraction of the fungus to its host (Turrà *et al.* 2015). There are also examples in *Colletotrichum* where the pheromone is present, while its cognate receptor is absent. In this case, the absence of the receptor would make it impossible for these species to recognize the pheromone.

An attractive interpretation of the pheromone presence/absence patterns exhibited in *Colletotrichum* is that these genes may be implicated in the process of unbalanced heterothallism. It might be envisaged that individuals in the populations of each species could harbor one of the two pheromones and that mating identity and compatibility are determined by these loci. However, the results of this study do not support that hypothesis. Individual isolates residing in the gloeosporioides clade and in which unbalanced heterothallism was first described (Wheeler 1954; Liang *et al.* 2021), harbor both the **a**- and **α**-factor pheromones and their receptors. The same is true for a number of other species of *Colletotrichum*. In addition, there are some species in which neither pheromone was identified. Thus, presence or absence of one or the other pheromone does not explain the sexual dimorphism in *Colletotrichum*. Furthermore, the presence/absence patterns are highly conserved within clades. It is consequently unlikely that in all these cases, isolates chosen for whole-genome sequencing would coincidentally exhibit the same patterns of absence or presence as one another in each species complex. In addition, a recent study in *C. fructicola* (Liang *et al.* 2021) found that sexually compatible isolates express both mating pheromones and exhibit no sequence differences at these loci, thus supporting the results presented here. It is unlikely that the *MAT* genes or pheromone genes play a role in unbalanced heterothallism and thus the genes that are involved in this process remain to be identified.

The diversity in the mature pheromone peptide sequences of the *Colletotrichum* **α**-pheromone factor was particularly notable. In species such as those of *Fusarium* and *Neurospora*, the mature pheromone peptide is highly conserved (Martin *et al.* 2011). This is even true for *N. crassa* and *Sordaria macrospora* which, despite being accommodated in different genera, produce identical **α**-factor pheromones. This is in contrast to the high level of diversity seen in the **α**-factor pheromones of *Colletotrichum* species. This is true both of the sequence and number of mature pheromone repeats encoded by the gene. For example, *C. higginsianum* and *Colletotrichum tanaceti* reside in the same species complex and yet encode two and three mature repeats respectively and share no mature pheromone repeats. Furthermore, *Colletotrichum asianum* produces an **α**-factor protein harboring a total of six mature repeats, made up of four different amino acid sequences. This suggests that this region is rapidly evolving and may contribute to the unpredictable nature of mating in this genus.

Despite sexual cycles not being apparent in many of the *Colletotrichum* species considered in this study, RIP profiles suggest that sexual reproduction is more common in the genus than is reflected by known sexual states. RIP is a process that is thought to exclusively occur during the pre-meiotic phase of sexual reproduction in filamentous ascomycete fungi and may thus be predictive of the sexual cycle (Hane *et al.* 2015). Importantly, RIP may indicate that sexual reproduction has taken place in the past and may consequently not be predictive of sexual recombination actively occurring in an extant population. Some *Colletotrichum* species, particularly those belonging to the orbiculare species complex, have surprisingly high RIP values; up to 56% in *C. trifolii*. This is significantly higher than other reports of genome-wide RIP, including those for *N. crassa* (15%) and various *Fusarium* species (up to 6.4%) (van Wyk *et al.* 2019). This is particularly interesting given that the sexual cycle has not been observed in any of the four orbiculare clade representatives included in this study. The significance of these high RIP levels deserves further investigation as it may have relevant implications in the evolutionary history and future of this genus.

The three *C. truncatum* genomes considered in this study displayed various interesting differences that are worth noting. The *C. truncatum* 1 and *C. truncatum* 2 isolates were both isolated from *Capsicum annuum*, while *C. truncatum* 3 was isolated from soybean. Despite this, both *C. truncatum* 1 and 3 harbored both mating pheromones as well as the **a**-factor receptor, while *C. truncatum* 2 harbored neither pheromone. In fact, the loci at which the pheromone genes are typically found in other *Colletotrichum* species were entirely disrupted in *C. truncatum* 2, with significant rearrangement and gene loss events having taken place. However, when the *MAT1* locus is considered, *C. truncatum* 1 is different from the other two isolates, with a significantly longer locus harboring a 4Kb insertion. This is of significance because *C. truncatum* 2 was originally described as *C. capsici* but has since been reduced to synonymy with *C. truncatum*. These results may provide support for treating *C. capsici* as a discrete species, and they also provide preliminary evidence of a *C. truncatum* × *C. capsici* hybrid.

Conclusions

The ambiguous sexual behaviors exhibited by species of *Colletotrichum* are potentially the result of a unique pattern of gene presence and absence. Unlike most other studied filamentous ascomycetets, both the pheromones and their receptors are not consistently present in the genomes of species in this genus. This has implications for how these fungi can sexually reproduce, particularly in the manner in which they recognize potential mating partners. Future research should include transcriptomic analyses to determine the expression patterns of the genes that are present, particularly in isolates known to be sexually compatible.

Acknowledgments

The authors thank both Dr Tuan Duong and Dr Deon de Jager who both provided significant assistance with various analyses important to the outcomes of this study.

Funding

Funding for this study was provided by the South African Department of Science and Innovation's South African Research Chair Initiative (SARChI), through the research chair of Prof. BD Wingfield. Additional funding was provided by Botanical Resources Australia—Agricultural Services, Pty. Ltd. The work was also supported by a post-doctoral grant awarded to Dr AM Wilson from the University of Pretoria, South Africa; while Dr RV Lelwala received Melbourne International Fee Remission and Melbourne International Research Scholarships from the University of Melbourne, Australia.

Conflicts of interest

None declared.

Literature cited

Banniza S, Warale R, Menat J, Cohen-Skali A, Armstrong-Cho C, *et al.* 2018. The long path to understanding the host–pathogen interactions of *Colletotrichum lentis* on lentil. Can J Plant Pathol. 40: 199–209.

Barimani M, Pethybridge SJ, Vaghefi N, Hay FS, Taylor PWJ. 2013. A new anthracnose disease of pyrethrum caused by *Colletotrichum tanaceti* sp. nov. Plant Pathol. 62:1248–1257.

Bhunjun CS, Phukhamsakda C, Jayawardena RS, Jeewon R, Promputtha I, *et al.* 2021. Investigating species boundaries in *Colletotrichum*. Fungal Divers. 107:107–127.

Bobrowicz P, Pawlak R, Correa A, Bell-Pedersen D, Ebbole DJ. 2002. The *Neurospora crassa* pheromone precursor genes are regulated by the mating type locus and the circadian clock. Mol Microbiol. 45:795–804.

Bolker M, Kahmann R. 1993. Sexual pheromones and mating responses in fungi. Plant Cell. 5:1461.

Charron C, Hubert J, Minatchy J, Wilson V, Chrysot F, *et al.* 2018. Characterization of *Colletotrichum orchidophilums*, the agent of black spot disease of vanilla. J Phytopathol. 166:525–531.

Damm U, Cannon PF, Woudenberg JHC, Crous PW. 2012. The *Colletotrichum acutatum* species complex. Stud Mycol. 73:37–113.

Damm U, Sato T, Alizadeh A, Groenewald JZ, Crous PW. 2019. The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. Stud Mycol. 92:1–46.

Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Pietro AD, *et al.* 2012. The Top 10 fungal pathogens in molecular plant pathology. Mol Plant Pathol. 13:414–430.

De Silva DD, Crous PW, Ades PK, Hyde KD, Taylor PWJ. 2017. Life styles of *Colletotrichum* species and implications for plant biosecurity. Fungal Biol Rev. 31:155–168.

Dyer PS, Inderbitzin P, Debuchy R. 2016. Mating-type structure, function, regulation and evolution in the Pezizomycotina. In: Wendland J, editor. Growth, Differentiation and Sexuality. Cham, Switzerland: Springer International Publishing. p. 351–385.

Eastburn DM, Gubler W. 1990. Strawberry Anthracnose: detection and survival of *Colletotrichum acutatum* in soil. Plant Dis. 74:161.

Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, *et al.* 2005. Protein identification and analysis tools on the ExPASy Server. In: J Walker, editor. The Proteomics Protocols Handbook. Totowa, NJ: Springer: Humana Press Inc. p. 571–608.

Guerber JC, Liu B, Correll JC, Johnston PR. 2003. Characterization of diversity in *Colletotrichum acutatum sensu lato* by sequence analysis of two gene introns, mtDNA and intron RFLPs, and mating compatibility. Mycologia. 95:872–895.

Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP. 2015. Repeat-induced point mutation: A fungal-specific, endogenous mutagenesis process. In: M van den Berg, K Maruthachalam, editors. Genetic Transformation Systems in Fungi. Cham, Switzerland: Springer International Publishing. p. 55–68.

Hyde K, Cai L, McKenzie E, Yang YL, Zhang JZ, *et al.* 2009. *Colletotrichum*: a catalogue of confusion. Fungal Divers. 39:1–17.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30:772–780.

Krogh A, Sonnhammer ELL, Ka L. 2007. Advantages of combined transmembrane topology and signal peptide prediction—the Phobius web server. Nucleic Acids Res. 35:W429–W432.

Kronstad JW, Staben C. 1997. Mating type in filamentous fungi. Annu Rev Genet. 31:245–276.

Lee J, Leslie JF, Bowden RL. 2008. Expression and function of sex pheromones and receptors in the homothallic ascomycete *Gibberella zeae*. Eukaryot Cell. 7:1211–1221.

Lelwala RV, Scott JB, Ades PK, Taylor PWJ. 2019. Population structure of *Colletotrichum tanaceti* in Australian Pyrethrum reveals high evolutionary potential. Phytopathology. 109:1779–1792.

Liang X, Yao L, Kong Y, Li B, Hao X, *et al.* 2021. Molecular dissection of perithecial mating line development in *Colletotrichum fructicola*, a

species with a nontypical mating system featured by plus-to-minus switch and plus-minus mediated sexual enhancement. Appl Environ Microbiol. 87:e0047421.
Maddison WP, Maddison DR. 2011. Mesquite: A Modular System for Evolutionary Analysis. Version 3.61 http://www.mesquiteproject.org.
Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, *et al.* 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43:D222–226.
Marin-Felix Y, Groenewald JZ, Cai L, Chen Q, Marincowitz S, *et al.* 2017. Genera of phytopathogenic fungi: GOPHY 1. Stud Mycol. 86: 99–216.
Martin SH, Wingfield BD, Wingfield MJ, Steenkamp ET. 2011. Causes and consequences of variability in peptide mating pheromones of ascomycete fungi. Mol Biol Evol. 28:1987–2003.
Mayrhofer S, Pöggeler S. 2005. Functional characterization of an α-factor-like *Sordaria macrospora* peptide pheromone and analysis of its interaction with its cognate receptor in *Saccharomyces cerevisiae*. Eukaryot Cell. 4:661–672.
Mayrhofer S, Weber JM, Pöggeler S. 2006. Pheromones and pheromone receptors are required for proper sexual development in the homothallic ascomycete *Sordaria macrospora*. Genetics. 172: 1521–1533.
Menat J, Cabral AL, Vijayan P, Wei Y, Banniza S. 2012. *Glomerella truncata*: another *Glomerella* species with an atypical mating system. Mycologia. 104:641–649.
Nylander JAA. 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, *et al.* 2012. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nat Genet. 44:1060–1065.
Pande S, Mughogho L, Bandyopadhyay R, Karunakarr R. 1991. Variation in pathogenicity and cultural characteristics of sorghum isolates of *Colletotrichum graminicola* in India. Plant Dis. 75: 778–783.
Pöggeler S. 2000. Two pheromone precursor genes are transcriptionally expressed in the homothallic ascomycete *Sordaria macrospora*. Curr Genet. 37:403–411.
Politis DJ. 1975. The identity and perfect state of *Colletotrichum graminicola*. Mycologia. 67:56–62.
Rambaut A. 2018. FigTree. Distrib. by author. http://tree.bio.ed.ac.uk/software/figtree/.
Rhaiem A, Taylor PWJ. 2016. *Colletotrichum gloeosporioides* associated with anthracnose symptoms on citrus, a new report for Tunisia. Eur J Plant Pathol. 146:219–224.
Roach KC, Feretzaki M, Sun S, Heitman J. 2014. Unisexual reproduction. In: T Friedmann, JC Dunlap, SF Goodwin, editors. Advances in Genetics. Waltham: Academic Press. p. 255–305.
Rodríguez-Guerra R, Ramírez-Rueda M-T, Cabral-Enciso M, García-Serrano M, Lira-Maldonado Z, *et al.* 2005. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. Mycologia. 97:793–803.
Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, *et al.* 2012. Mrbayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 61:539–542.
Shen WC, Bobrowicz P, Ebbole DJ. 1999. Isolation of pheromone precursor genes of *Magnaporthe grisea*. Fungal Genet Biol. 27:253–263.
Souza EA, Camargo OA, Pinto JM. 2010. Sexual recombination in *Colletotrichum lindemuthianum* occurs on a fine scale. Genet Mol Res. 9:1759–1769.
Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics. 19:ii215–ii225.
Sukno SA, García VM, Shaw BD, Thon MR. 2008. Root infection and systemic colonization of maize by *Colletotrichum graminicola*. Appl Environ Microbiol. 74:823–832.
Turgeon BG, Yoder OC. 2000. Proposed nomenclature for mating type genes of filamentous ascomycetes. Fungal Genet Biol. 31: 1–5.
Turina M, Prodi A, Van Alfen NK. 2003. Role of the *Mf1-1* pheromone precursor gene of the filamentous ascomycete *Cryphonectria parasitica*. Fungal Genet Biol. 40:242–251.
Turrà D, Ghalid ME, Rossi F, Pietro AD. 2015. Fungal pathogen uses sex pheromone receptor for chemotropic sensing of host plant signals. Nature. 527:521–524.
Vaillancourt L, Du M, Wang J, Rollins J, Hanau R. 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia. 92:430–435.
Vaillancourt LJ, Hanau RM. 1991. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. Phytopathology. 81:530–534.
Wang L, Lin X. 2011. Mechanisms of unisexual mating in *Cryptococcus neoformans*. Fungal Genet Biol. 48:651–660.
Weir BS, Johnston PR, Damm U. 2012. The *Colletotrichum gloeosporioides* species complex. Stud Mycol. 73:115–180.
Wheeler H. 1954. Genetics and evolution of heterothallism in *Glomerella*. Phytopathology. 44:342–345.
Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD. 2017. Which *MAT* gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. Fungal Biol Rev. 31:199–211.
Wilson AM, Godlonton T, van der Nest MA, Wilken PM, Wingfield MJ, *et al.* 2015. Unisexual reproduction in *Huntiella moniliformis*. Fungal Genet Biol. 80:1–9.
Wilson AM, van der Nest MA, Wilken PM, Wingfield MJ, Wingfield BD. 2018. Pheromone expression reveals putative mechanism of unisexuality in a saprobic ascomycete fungus. PLoS One. 13: e0192517.
Van Wyk S, Harrison CH, Wingfield BD, De Vos L, Van Der Merwe NA, *et al.* 2019. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. PeerJ. 7: e7447.
van Wyk S, Wingfield BD, De Vos L, van der Merwe NA, Santana QC, *et al.* 2019. Repeat-induced point mutations drive divergence between *Fusarium circinatum* and its close relatives. Pathogens. 8: 298.
Yun SH, Berbee ML, Yoder OC, Turgeon BG. 1999. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. Proc Natl Acad Sci USA. 96:5592–5597.
Zhang L, Baasiri RA, Van Alfen NK. 1998. Viral repression of fungal pheromone precursor gene expression. Mol Cell Biol. 18: 953–959.

Communicating editor: J. Dunlap