Unique Patterns of Mating Pheromone Presence and Absence could Result in the Ambiguous Sexual Behaviours of Colletotrichum Species

Andrea Melissa Wilson (✉ andi.wilson@fabi.up.ac.za)

FABI, University of Pretoria <https://orcid.org/0000-0002-3239-0045>

RV Lelwala

University of Melbourne Faculty of Veterinary and Agricultural Sciences

PWJ Taylor

University of Melbourne Faculty of Veterinary and Agricultural Sciences

MJ Wingfield

FABI, University of Pretoria

BD WINGFIELD

FABI, University of Pretoria

Research article

Keywords: Colletotrichum, Sexual reproduction, Filamentous fungi, Mating pheromones, Cognate receptors, Gene loss, Ancestral state reconstruction

Posted Date: November 24th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-112807/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: *Colletotrichum* species are known to engage in unique sexual behaviours that differ significantly from the mating strategies of other filamentous ascomycete species. Most ascomycete fungi require the expression of both the *MAT1-1-1* and *MAT1-2-1* genes to regulate mating type and induce sexual reproduction. In contrast, all isolates of *Colletotrichum* are known to harbour 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 behaviours in *Colletotrichum* are thus unknown.

Results: A comparative genomics approach analysing genomes from 47 *Colletotrichum* isolates was used to elucidate a putative molecular mechanism underlying the unique sexual behaviours observed in *Colletotrichum* species. The existence of only the *MAT1-2* idiomorph was confirmed across all species included in this study. Comparisons at the loci harbouring the two mating pheromones and their cognate receptors revealed interesting patterns of gene presence and absence as well as gene loss. The results also showed that these genes have been lost multiple times over the evolutionary history of this genus.

Conclusion: The multiple losses of the pheromone genes in these species suggest strong selection against the typical mating strategies seen in other species. This further suggests that these pheromones no longer play a role in mating type determination and that the species of this genus have undiscovered mechanisms by which to control mating type and mating partner recognition. This research thus provides a base from which further interrogation of this topic can take place.

Background

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 (1). Most described species require the expression of genes from both the *MAT1-1* and *MAT1-2* idiomorphs in order to sexually reproduce. Those that harbour both sets of genes within a single genome are considered homothallic and are self-fertile (2). Alternatively, those that possess either the *MAT1-1-1* or *MAT1-2-1* gene are typically obligate outcrossers (heterothallic) that require an opposite mating type partner in order to engage in sexual reproduction (3, 4).

The genes present at the *MAT1* locus are directly responsible for controlling the expression of the mating pheromones in ascomycete fungi (5, 6). 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 α -factor pheromone expression (5–8). 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 (9). Mating pheromones are also expressed by

certain homothallic species, often with both being expressed by a single isolate (10, 11), consistent with the presence of both mating genes.

Species of *Colletotrichum* do not undergo either of the typical sexual strategies and rather exhibit ambiguous sexual behaviours that are difficult to classify. Sexual reproduction has been shown to occur in numerous *Colletotrichum* species (12–15), all of which harbour only the *MAT1-2-1* gene, while a *MAT1-1-1* homolog has never been identified (14, 16, 17). 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, with heterothallic-like behaviour occurring between two isolates of the same mating type (18).

Studies on *Colletotrichum* species such as *Colletotrichum gloeosporioides* have suggested that a unique type of heterothallism could explain the distinctive sexual behaviour of these species (16). The 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 (19). Consequently, many different mating identities exist and two individuals are sexually compatible if they harbour complementary mutations. This form of mating is thus theoretically capable of producing both self-fertile and self-sterile progeny (20). This is not, however, the case for all species of *Colletotrichum*. In contrast, research on *Colletotrichum graminicola* has showed that cross fertility in this species is determined by two-unlinked loci (14, 21), while *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 (22). Whether these unique mating behaviours could be influenced by the mating pheromone pathway has not previously been considered.

To further complicate our understanding of the sexual behaviours in *Colletotrichum*, species such as *C. gloeosporioides* (23), *Colletotrichum fiorinae* (13) and *Colletotrichum graminicola* (21, 24) have been described as both homothallic and heterothallic. In these species, some isolates are capable of sexual reproduction when cultured in isolation, while other isolates require a mating partner in order produce sexual offspring. A further complicating factor is that the genus has undergone many taxonomic revisions (25, 26), making it difficult to assign mating strategies to fungi that have been renamed and reclassified many times.

The *Colletotrichum* genus is divided into 14 currently recognized species complexes (26, 27) which accommodate around 220 species (26–28). The genus includes important plant pathogens, some of which are recognized as among the top 10 fungal plant pathogens worldwide (29). Many *Colletotrichum* species are hemibiotrophic and have complex life cycles, switching from biotrophic to necrotrophic or endophytic states within the same life cycle (30). These fungi cause a disease known as anthracnose, and infect a diverse range of hosts, including grasses such as maize (31) and sorghum (32), fruits such as strawberry (33), and citrus (34), perennial plants such as pyrethrum (35), as well as flowering plants like the vanilla orchid (36). 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 (26). It is thus crucial that we understand the sexual behaviors of *Colletotrichum* species, the impact and importance that sexual reproduction plays in their lifecycles and the role of ascospore production 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 45 species across this genus, representing most of the recognized species complexes. Additionally, 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.

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 (Fig. 1). In all cases, the *MAT1* locus harboured the *MAT1-2-1* gene and a secondary *MAT1-2* gene. The locus orientation was also highly conserved, with the *MAT1-2-1* gene found directly downstream of the *APN2* gene. 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.

All of the *Colletotrichum* *MAT1-2-1* genes encoded a protein possessing the HMG (high mobility group) box domain- 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 *MAT1-2-13* in accordance with accepted *MAT* gene nomenclature (37). 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 18 of the 45 *Colletotrichum* species and both *Verticillium* species considered in this study (Table S1). 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 single peptide at the N-terminal. Furthermore, they harboured numerous KEX2 processing sites (KR) as well as one or more repeats of the mature, 11 aa pheromone factor (Table S1).

The gene encoding the α -pheromone was found at a fairly conserved location within the genome of these species, although multiple rearrangements had taken place at this locus (Fig. 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 (Fig. 2, Table S2) were also found in species that do not harbour 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 (Fig. 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 for the different 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* as well as all the species 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* (Fig. 3). At least 10 independent genetic events have occurred at the α -pheromone locus during the evolutionary trajectory of *Colletotrichum*, each of which resulted in the loss of the α -pheromone (Fig. 4). These events included small changes such as point mutations disrupting the start codon or introducing a premature stop codon. Additionally, 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 residing in this complex, the α -factor pheromone was absent, and instead, two genes encoding hypothetical proteins are present (Fig. 2, Fig. 4). Despite this, the locus was not remarkably different in length. 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 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 (Fig. 3). While the *C. truncatum* 1 and *C. truncatum* 3 genomes harboured a gene encoding the α -pheromone, the *C. truncatum* 2 (formerly *Colletotrichum capsici*) genome did not. In fact, in *C. truncatum* 2, genes encoding for the FAD-linked oxidoreductase and major facilitator transporter proteins were not present on the same contig and are flanked by other genes. Additionally, neither the α -factor pheromone gene nor the DUF1640 domain containing protein gene could be identified in the *C. truncatum* 2 genome. 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* (Fig. 3). While the locus was conserved between *C. coccodes* and closely related species such as *Colletotrichum higginsianum* that do harbour the gene, a point mutation has disrupted the start codon of the putative *C. coccodes* α -factor pheromone gene (Fig. 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 harbouring 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 (Fig. 3). 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 harbours only a single mature pheromone repeat, compared to the two to seven repeats in the other species (Table S1). It was particularly relevant that the terminal 43 nt at the 3' region of the *C. tofieldiae* α -pheromone gene were well conserved the *C. incanum* (Fig. 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 removed the 5' region of α -pheromone in this species.

The α -pheromone gene was also lost by species residing in the caudatum and graminicola species complexes (Fig. 3). In these species, an entirely different gene was present at this locus. In almost all of the species, a gene encoding a tetratricopeptide (TTP) protein was present, while this locus in *Colletotrichum falcatum* encoded a hypothetical protein (Fig. 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 (Fig. 3, Fig. 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. fiorinae* 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 caudatum, graminicola and orbiculare species complexes. Interestingly, the genes found at this locus were not conserved, and suggesting that multiple disruptions have occurred at this locus.

The α -factor pheromone flanking genes in *Colletotrichum sansevieriae* were identified at the ends of two different contigs that could not be joined. Therefore, a 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 present (Fig. 3).

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

The gene encoding the a-factor pheromone was identified in a total of 28 *Colletotrichum* species and both *Verticillium* species 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 most of the 17 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 is its short length and the presence of a highly conserved C-terminal CaaX domain. In the *Colletotrichum* species, the sequence of this conserved domain was either CVIL or CVVM (Table S3). 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 pluvivorum* 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 (Fig. 7). Although the ancestral state reconstruction showed a total of four loss events, the genetic imprints left during these losses suggests 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* (Fig. 6). Although the *C. sojae* locus showed high levels of similarity to this locus in other species residing in the orcidearum species complex, a 7 nt deletion in the gene-encoding region 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 (Fig. 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* (Fig. 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 (Fig. 6, Fig. 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.

The final pheromone loss occurred when the acutatum species complex branched off from the rest of the genus (Fig. 6). However, genetic analysis showed that at least four different deletion events have taken place in this species complex (Fig. 7). The first of these occurred in *C. godetiae*, *Colletotrichum phormii*

and *Colletotrichum salicis*, where a disruption of a start codon was 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 pheromones 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* (Fig. 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 (Fig. 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 (Fig. 7).

The orbiculare species complex is worth mentioning here although this gene has not been entirely lost in these species. However, the a-factor pheromone locus in these species has undergone a deletion of 15 bp that correspond to the first five amino acids of the a-factor pheromone (Fig. 7). An alternative start codon was available, and since no functional domain is known from the N-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 pheromone receptors display similar patterns of presence and absence as their pheromone factors

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 a gene encoding the pheromone's cognate receptor. The receptors, therefore, showed similar patterns of gene presence and absences as their pheromones (Fig. 3, Fig. 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 *Colletotrichum tofieldiae*, where the pheromone was present while the receptor was absent, and *C. godetiae*, where the pheromone was absent while the receptor was present (Fig. 3). When the a-factor and its cognate receptor (*preA*) are considered, three exceptions exist. *C. sojae*, *C. coccodes* and *C. truncatum* 2, while harbouring an intact pheromone receptor, do not possess the corresponding pheromone's gene (Fig. 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

(Fig. 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 harbouring the α -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 0.44% (*Colletotrichum caudatum*) to 56.02% (*Colletotrichum trifolii*). The only species that displayed no genomic evidence of RIP was *Colletotrichum sublineola*. In general, species residing in the same species complexes showed similar RIP profiles.

Table 1
The RIP profiles and GC content of the genomes used in this study

Species	RIP %	GC Content	# of LRAR	GC Content of LRARs
acutatum species complex				
<i>C. nymphaeae</i>	0.93	52.75	942	30.11
<i>C. simmondsii</i>	3.97	51.73	26	24.44
<i>C. costarricense</i>	5.06	51.35	118	23.24
<i>C. paranaense</i>	2.11	52.23	34	23.55
<i>C. lupini</i>	19.6	47.58	879	26.72
<i>C. tamarilloi</i>	5.51	51.2	175	22.59
<i>C. cuscutae</i>	41.1	41.34	2037	25.99
<i>C. melonis</i>	2.41	52.15	51	24.01
<i>C. abscissum</i>	5.93	51.11	180	23.49
<i>C. florinae</i>	1.42	52.49	11	35.53
<i>C. phormii</i>	1.36	52.74	25	42.47
<i>C. salicis</i>	0.99	52.78	3	38.52
<i>C. godetiae</i>	2.89	52.09	61	26.97
<i>C. acutatum</i>	2.71	52.25	45	25.89
caudatum species complex				
<i>C. somersetense</i>	14.57	50.05	339	26.41
<i>C. caudatum</i>	0.44	54.95	0	0
<i>C. zoysiae</i>	3.99	53.76	152	25.13
graminicola species complex				
<i>C. navitas</i>	18.31	48.63	246	24.82
<i>C. graminicola</i>	18.09	48.42	545	25.15
<i>C. sublineola</i>	0	53.11	0	0
<i>C. eremochloae</i>	3.35	52.54	66	25.67
<i>C. falcatum</i>	10.01	52.78	230	27.76
<i>C. cereale</i>	9.9	52.29	230	29.04
spaethianum species complex				

Species	RIP %	GC Content	# of LRAR	GC Content of LRARs
<i>C. tofieldiae</i>	4.8	52.99	110	28.12
<i>C. incanum</i>	4.64	51.9	117	33.34
destructivum species complex				
<i>C. higginsianum</i>	3.27	54.41	117	25.89
<i>C. shiso</i>	35.92	46.26	284	28.8
<i>C. tanacet</i>	24.1	49.19	674	31.1
No assigned complex				
<i>C. orchidophilum</i>	6.22	51.06	92	18.58
<i>C. coccodes</i>	1.31	53.79	0	-
<i>C. chlorophyti</i>	10.85	49.8	214	25.8
gloeosporioides species complex				
<i>C. fruticola</i>	1.37	53.2	39	19.7
<i>C. asianum</i>	14.32	49.37	485	23.58
<i>C. camelliae</i>	10.76	50.01	361	20.9
<i>C. gloeosporioides</i>	2.61	51.98	107	19.7
boninense species complex				
<i>C. karstii</i>	3.75	52.69	101	23.4
<i>C. sansevieriae</i>	11.53	50.74	38	34.88
truncatum species complex				
<i>C. truncatum</i> (1)	3.9	49.73	62	27.49
<i>C. truncatum</i> (2) (previously <i>C. capsici</i>)	5.19	48.26	123	26.72
<i>C. truncatum</i> (3)	2.85	50.12	122	35.63
orcidearum species complex				
<i>C. musicola</i>	5.72	54.94	133	34.87
<i>C. pluvivorum</i>	0.99	55.84	15	24.47
<i>C. soj</i>	1.02	55.9	14	30.3
orbiculare species complex				
<i>C. sidae</i>	47.37	37.96	613	19.62

Species	RIP %	GC Content	# of LRAR	GC Content of LRARs
<i>C. orbiculare</i>	49.53	36.47	1067	17.91
<i>C. trifolii</i>	56.02	35.24	1045	20.85
<i>C. spinosum</i>	45.32	38.7	588	19.36
Verticillium (Outgroup)				
<i>V. tricorpus</i>	2.12	57.24	32	32.38
<i>V. dahliae</i>	2.59	55.34	24	28.9

Table 2
Genomes used in this study

Species	Strain ^a	Database	Accession Number/ Project ID ^b	Reference/Citation
acutatum species complex				
<i>C. nymphaeae</i>	SA-01	JGI	-	N/A
<i>C. simmondsii</i>	CBS 122122	JGI	-	N/A
<i>C. costarricense</i>	IMI 309622	JGI	-	N/A
<i>C. paranaense</i>	IMI 384185	JGI	-	N/A
<i>C. lupini</i>	CBS 109225	JGI	-	N/A
<i>C. tamarilloi</i>	CBS 129955	JGI	-	N/A
<i>C. cuscutae</i>	IMI 304802	JGI	-	N/A
<i>C. melonis</i>	CBS 134730	JGI	-	N/A
<i>C. abscissum</i>	IMI 504890	JGI	-	N/A
<i>C. fiorinae</i>	MH 18	JGI	Project ID: 1006306	N/A
<i>C. phormii</i>	CBS 102054	JGI	Project ID: 1060996	N/A
<i>C. salicis</i>	CBS607.94	JGI	-	N/A
<i>C. godetiae</i>	CBS 193.32	JGI	Project ID: 1060998	N/A
<i>C. acutatum</i>	CBS 112980	JGI	Project ID: 1061000	N/A
caudatum species complex				
<i>C. somersetense</i>	CBS 131599	JGI	Project ID: 1043133	N/A
<i>C. caudatum</i>	CBS 131602	JGI	Project ID: 1006149	N/A
<i>C. zoysiae</i>	MAFF 235873	JGI	Project ID: 1043139	N/A

Species	Strain ^a	Database	Accession Number/ Project ID ^b	Reference/Citation
graminicola species complex				
<i>C. navitas</i>	CBS 125086	JGI	-	N/A
<i>C. graminicola</i>	M1.001	JGI	-	N/A
<i>C. sublineola</i>	CBS 131301	JGI	Project ID: 1006165	N/A
<i>C. eremochloae</i>	CBS 129661	JGI	Project ID: 1043109	N/A
<i>C. falcatum</i>	MAFF 306170	JGI	Project ID: 1043115	N/A
<i>C. cereale</i>	CBS 129662	JGI	Project ID: 1043097	N/A
spaethianum species complex				
<i>C. tofieldiae</i>	CBS 168.49	JGI	-	N/A
<i>C. incanum</i>	MAFF 238712	JGI	-	N/A
destructivum species complex				
<i>C. higginsianum</i>	IMI 349063	NCBI	LTAN000000000.1	(28)
<i>C. shisoi</i>	PG-2018a	NCBI	PUHP000000000.1	(59)
<i>C. tanacetii</i>	BRIP 57314	NCBI	PJEX000000000.1	(35)
No assigned complex				
<i>C. orchidophilum</i>	IMI 309357	NCBI	MJBS000000000.1	(60)
<i>C. coccodes</i> ^c	NJ-RT1	NCBI	LECQ000000000.1	USDA-ARS
<i>C. chlorophyti</i>	NTL11	NCBI	MPGH000000000.1	(61)
gloeosporioides species complex				
<i>C. fructicola</i>	CGMCC 3.17371	NCBI	SSNE000000000.1	(62)
<i>C. asianum</i>	ICMP 18580	NCBI	WOWK000000000.1	(63)

Species	Strain ^a	Database	Accession Number/ Project ID ^b	Reference/Citation
<i>C. camelliae</i> ^c	CcLH18	NCBI	JAATWK000000000.1	Central South University of Forestry & Technology
<i>C. gloeosporioides</i>	23	JGI	Project ID: 1006302	N/A
boninense species complex				
<i>C. karstii</i> ^c	CkLH20	NCBI	JAATWM000000000.1	Central South University of Forestry & Technology
<i>C. sansevieriae</i> ^c	Sa-1-2	NCBI	NJHP000000000.1	(64)
truncatum species complex				
<i>C. truncatum</i> (1) ^c	MTCC 3414	NCBI	NBAU000000000.2	(65)
<i>C. truncatum</i> (2) (previously <i>C. capsici</i>) ^c	KLC.C-4	NCBI	JAATLN000000000.1	KL University
<i>C. truncatum</i> (3) ^c	CMES1059	NCBI	VUJX000000000.1	(66)
orcidearum species complex				
<i>C. musicola</i> ^c	LFN0074	NCBI	WIGM000000000.1	(66)
<i>C. pluvivorum</i> ^c	LFN00145	NCBI	WIGO000000000.1	(66)
<i>C. sojae</i> ^c	LFN0009	NCBI	WIGN000000000.1	(66)
orbiculare species complex				
<i>C. sidae</i>	CBS 815.97	NCBI	QAPF000000000.1	(67)
<i>C. orbiculare</i>	MAFF 240422	NCBI	AMCV000000000.2	(67)
<i>C. trifolii</i>	543-2	NCBI	RYZW000000000.1	(67)
<i>C. spinosum</i>	CBS 515.97	NCBI	QAPG000000000.1	(67)
Verticillium (Outgroup)				
<i>V. tricorpus</i>	MUCL 9792	NCBI	JPET000000000.1	(68)

Species	Strain ^a	Database	Accession Number/ Project ID ^b	Reference/Citation
<i>V. dahliae</i>	VdLs.17	NCBI	ABJE000000000.1	(69)
^a Isolate information is not always available from NCBI or JGI				
^b Genomes from the JGI where project numbers are not yet available are indicated as -				
^c Genomes which needed to be annotated using AUGUSTUS				

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%), 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%).

Discussion

It has long been apparent that the sexual behaviors of species of *Colletotrichum* are unique amongst filamentous ascomycetes (14, 16, 24). These species exhibit ambiguous sexual preferences, with individual isolates in some species engaging in different sexual strategies (eg: *C. fioriniae* (13, 38)). 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 45 *Colletotrichum* species considered in this study were associated with the *MAT1-2* locus. The *MAT1-2-1* gene was shown to be highly conserved across all the species studied, with the HMG box domain consistently present. A second gene was also found to be consistently present at this locus and encoded an unknown protein with no recognizable domains. This gene has thus been named *MAT1-2-13* in accordance with the proposed *MAT* gene nomenclature system (4, 37). While the *MAT1-2* locus of these species is comparable to other ascomycetes, the complete absence of *MAT1-1*-associated genes in this genus was the first genetic indicator suggesting sexual reproduction in *Colletotrichum* is different to that of other filamentous ascomycetes (14, 16, 17, 28).

The sexual ambiguity and unpredictable nature of mating choices made by species of *Colletotrichum* can be linked to the patterns of gene loss associated with the pheromone factors and their cognate receptors. Of the *Colletotrichum* species considered, only 15 harboured both pheromone factors. Of the remaining species, 17 harboured only one of the pheromones and a further 13 possessed neither. This is in stark contrast to many other studied filamentous ascomycete fungi (5–8, 11, 39–43) and strongly suggests that mating identity and preference is not under the control of the *MAT1* locus and the pheromone pathway in these species. Additionally, both homothallism and heterothallism have been documented in these species, further suggesting that the presence and absence of *MAT* genes as well as the relevant

pheromone pathway genes is not predictive of sexual capabilities or behaviours in *Colletotrichum* species as is true for other ascomycetes.

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 the same- an inability to produce mating pheromone factors. These concerted loss events strongly imply some form of selective pressure against the typical mating pathways associated with sexual reproduction in other filamentous ascomycetes. For example, evolution towards unique sexual strategies such as unisexuality are thought to occur in environments where finding a suitable mating partner is extremely difficult (44, 45). The same may be true in *Colletotrichum*, where the *MAT1-1* idiomorph has been lost and thus a different manner of sexual reproduction has needed to evolve.

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 coopted 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 (46). 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.

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 (39). This is even true for *Neurospora 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 tanacetii* 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 harbouring a total six mature repeats, made up of four different amino acid sequences. This suggests that this region is rapidly evolving and contributes 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 thus may be predictive of the sexual cycle (47).

Importantly, RIP may indicate that sexual reproduction has taken place in the past and thus may not be predictive of sexual recombination actively occurring in an extant population. Interestingly, certain *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 content, including those for *N. crassa* (15%) and various *Fusarium* species (up to 6.4%)(48). The significance of these high RIP levels needs to be further investigated as it may have noteworthy implications for the evolutionary history and future of this genus.

The three *C. truncatum* genomes used in this study showed some interesting differences that are worth noting here. 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 harboured both mating pheromones as well as the a-factor receptor, while *C. truncatum* 2 harboured neither the pheromones nor the receptor. 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 harbouring a 4Kb insert. This is of significance as *C. truncatum* 2 was 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 also provides preliminary evidence of a *C. truncatum* x *C. capsici* hybrid.

Conclusions

The ambiguous sexual behaviours 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 throughout this genus. This has implications for how these fungi are able to 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.

Methods

Genomes used

This study used genomes from species that represent 10 of the 14 currently recognized *Colletotrichum* species complexes. A total of 39 genomes and their CDS annotations were downloaded from the National Centre for Biotechnology Information (NCBI) and the Joint Genomes Institute (JGI). This assemblage of genomes included 37 *Colletotrichum* species, and two *Verticillium* species that were used as outgroup species in the various analyses conducted (Table 1). An additional 10 *Colletotrichum* genomes were downloaded from the NCBI and were annotated using the online version of AUGUSTUS (49), with default settings and *Verticillium longisporum* gene models (Table 1). These 10 additional genomes represented only eight species because three different isolates of *C. truncatum* were included-

C. truncatum 1, *C. truncatum* 2 (formerly *C. capsici*) and *C. truncatum* 3. Thus, in total, the 49 genomes represented 45 *Colletotrichum* species and two *Verticillium* species.

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 47 *Colletotrichum* genomes. These included sequences of *N. 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 *Colletotrichum lentis* (KX932470.1) was used as a local BLASTn query against all 47 *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 this locus were translated and subjected to functional domain identification using the NCBI's Conserved Domains Database (50).

Pheromone genes

For the identification of the α -factor pheromone, the *C. higginsianum* *ccg-4* mRNA sequence (XM_018306306.1) and its predicted protein sequence were used as queries in local BLASTn and tBLASTn searches against the 49 genomes. In species where the gene had not been annotated in the original genome source, the BLAST result was used to manually annotate the region likely encoding the pheromone.

Where no significant 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 possible, the intergenic regions were subjected to manual identification and annotation of the pheromone gene.

The predicted protein sequences of the α -factor pheromones were analyzed using ExPASy ProtScale (51) and Phobius (52) 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 (39).

The α -factor pheromone genes were also identified via the microsynteny-based approach, using the *N. crassa* α -factor pheromone and its flanking genes. The α -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 49 genomes. As above, the intergenic regions were subjected to manual identification and annotation of the pheromone gene. 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 (39).

Pheromone receptor genes

The **a**-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 49 genomes.

Phylogenetic Analyses

The six gene regions generally accepted as barcoding genes for *Colletotrichum* were used to generate a species tree (26). These regions included *act*, *chs-1*, *GAPDH*, *his-3*, *sod-2* and *tub* (Table S4). Where possible, these sequences were retrieved from Genbank- ensuring that the sequences available were generated from the same isolate used to produce the whole genome sequences. In the species for which these data was 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 (Additional file 5).

The gene regions were aligned using default parameters in MAFFT V7 (53). The alignments were subsequently subjected to model testing, using MrModelTest2 V2.4 (54). Thereafter, the alignments were concatenated into a single dataset and analyzed as such. Bayesian inference (BI) analyses were conducted using MrBayes V3.2.7 (55) 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. Trees were then edited in FigTree V1.4.4 (56). The relevant data has been deposited into TreeBASE (Submission ID: S27079, Reviewer access URL:

<http://purl.org/phylo/treebase/phylows/study/TB2:S27079?x-access-code=e5acf2c9708260dccac58f728765e161&format=html>).

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 (57). Absence (0) or presence (1) of the genes was mapped onto the BI species tree produced above.

RIP Analysis

All 49 whole genomes were subjected to analysis using The RIPper (<http://theripper.hawk.rocks>), a web-based tool that identifies regions in the genome that have been subjected to repeat induced point (RIP) mutations (58). 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.

List Of Abbreviations

aa: amino acids; APN2: DNA lyase; bp: base pairs; CHP: cyanate hydratase protein; EBP2: eukaryotic rRNA processing protein; HMG: high mobility group; *MAT*: mating type; nt: nucleotides; *preA*: a-factor receptor; *preB*: a-factor receptor; RIP: repeat induced point; SLA2: cytoskeleton assembly control; SNP: single nucleotide polymorphism; TTP: tetratricopeptide

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available in the published article and its supplementary information files, in the NCI's GenBank repository (<https://www.ncbi.nlm.nih.gov/genbank/>) or in the TreeBASE repository (<https://www.treebase.org/treebase-web/home.html>). Details pertaining to each dataset's availability can be found within the article's text.

Competing interests

The authors declare that they have no competing interests.

Funding

Funding for this study was provided by the South African Department of Science and Innovation's South African Research Chair Initiative (SARChI), with which Prof. BD Wingfield holds a chair in Fungal Genomics. Additional funding was provided by Botanical Resources Australia (BRA) – Agricultural Services, Pty. Ltd as well as the University of Pretoria (UP), South Africa. The work was also supported by a post-doctoral grant awarded to Dr AM Wilson by UP; while Dr RV Lelwala received Melbourne International Fee Remission and Melbourne International Research Scholarships from the University of Melbourne, Australia. While no wet laboratory research was performed for this study, access to computers, servers, analysis software, and internet was essential. The funding received from SARChI, UP and BRA was used to cover these costs. The funding bodies had no role in the study design, data analysis, result interpretation or writing of the manuscript.

Authors' contributions

All authors contributed significantly to the study and have approved the manuscript for submission. Each listed author also agreed to be personally accountable for their own contributions and to ensure that questions related to the accuracy or integrity of the work are appropriately investigated.

AMW and BDW were responsible for the conceptualization of the project. AMW was responsible for data acquisition, methodology, formal analysis, data visualization and writing of the first draft. RVL, PWJT, BDW and MJW contributed to the interpretation and visualization of the data and participated substantially in preparing the manuscript for submission. PWJT, BDW and MJW were responsible the acquisition of funding that made this study possible.

Acknowledgements

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

References

1. Kronstad JW, Staben C. Mating type in filamentous fungi. *Annu Rev Genet.* 1997;31(1):245–76.
2. Yun SH, Berbee ML, Yoder OC, Turgeon BG. Evolution of the fungal self-fertile reproductive life style from self-sterile ancestors. *Proc Natl Acad Sci.* 1999;96(10):5592–7.
3. Dyer PS, Inderbitzin P, Debuchy R. Mating-type structure, function, regulation and evolution in the Pezizomycotina. In: *Growth, Differentiation and Sexuality.* 3rd ed. Springer International Publishing; 2016. p. 351–85.
4. Turgeon BG, Yoder OC. Proposed nomenclature for mating type genes of filamentous ascomycetes. *Fungal Genet Biol.* 2000;31(1):1–5.

5. Bobrowicz P, Pawlak R, Correa A, Bell-Pedersen D, Ebbole DJ. The *Neurospora crassa* pheromone precursor genes are regulated by the mating type locus and the circadian clock. *Mol Microbiol.* 2002;45(3):795–804.
6. Zhang L, Baasiri RA, Van Alfen NK. Viral repression of fungal pheromone precursor gene expression. *Mol Cell Biol.* 1998;18(2):953–9.
7. Shen WC, Bobrowicz P, Ebbole DJ. Isolation of pheromone precursor genes of *Magnaporthe grisea*. *Fungal Genet Biol.* 1999;27(2–3):253–63.
8. Wilson AM, van der Nest MA, Wilken PM, Wingfield MJ, Wingfield BD. Pheromone expression reveals putative mechanism of unisexuality in a saprobic ascomycete fungus. *PLoS One.* 2018;13(3):e0192517.
9. Bolker M, Kahmann R. Sexual Pheromones and Mating Responses in Fungi. *Plant Cell.* 2007;5(10):1461.
10. Pöggeler S, Kück U. Identification of transcriptionally expressed pheromone receptor genes in filamentous ascomycetes. *Gene.* 2001;280(1–2):9–17.
11. Lee J, Leslie JF, Bowden RL. Expression and function of sex pheromones and receptors in the homothallic ascomycete *Gibberella zeae*. *Eukaryot Cell.* 2008;7(7):1211–21.
12. Barimani M, Pethybridge SJ, Vaghefi N, Hay FS, Taylor PWJ. A new anthracnose disease of pyrethrum caused by *Colletotrichum tanacetii* sp. nov. *Plant Pathol.* 2013;62(6):1248–57.
13. Guerber JC, Liu B, Correll JC, Johnston PR. Characterization of diversity in *Colletotrichum acutatum sensu lato* by sequence analysis of two gene introns, mtDNA and intron RFLPs, and mating compatibility. *Mycologia.* 2003;95(5):872–95.
14. Vaillancourt L, Du M, Wang J, Rollins J, Hanau R. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. *Mycologia.* 2000;92(3):430–5.
15. Souza EA, Camargo OA, Pinto JM. Sexual recombination in *Colletotrichum lindemuthianum* occurs on a fine scale. *Genet Mol Res.* 2010;9(3):1759–69.
16. Menat J, Cabral AL, Vijayan P, Wei Y, Banniza S. *Glomerella truncata*: Another *Glomerella* species with an atypical mating system. *Mycologia.* 2012;104(3):641–9.
17. Rodríguez-Guerra R, Ramírez-Rueda M-T, Cabral-Enciso M, García-Serrano M, Lira-Maldonado Z, Guevara-González RG, et al. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. *Mycologia.* 2005;97(4):793–803.
18. Damm U, Cannon PF, Woudenberg JHC, Crous PW. The *Colletotrichum acutatum* species complex. *Stud Mycol.* 2012;73:37–113.
19. Wheeler H. Genetics and evolution of heterothallism in *Glomerella*. *Phytopathology.* 1954;44:342–5.
20. Olive LS. On the Evolution of Heterothallism in Fungi. *Am Nat.* 1958;92(865):233–51.
21. Vaillancourt LJ, Hanau RM. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. *Phytopathology.* 1991;81(5):530–4.

22. Banniza S, Warale R, Menat J, Cohen-Skali A, Armstrong-Cho C, Bhadauria V. The long path to understanding the host–pathogen interactions of *Colletotrichum lentis* on lentil. *Can J Plant Pathol*. 2018;40(2):199–209.
23. Weir BS, Johnston PR, Damm U. The *Colletotrichum gloeosporioides* species complex. *Stud Mycol*. 2012;73:115–80.
24. Politis DJ. The identity and perfect state of *Colletotrichum graminicola*. *Mycologia*. 1975;67(1):56–62.
25. Hyde K, Cai L, McKenzie E, ... YY-F, 2009 U. *Colletotrichum*: A catalogue of confusion. *Fungal Divers*. 2009;39:1–17.
26. Marin-Felix Y, Groenewald JZ, Cai L, Chen Q, Marinowitz S, Barnes I, et al. Genera of phytopathogenic fungi: GOPHY 1. *Stud Mycol*. 2017;86:99–216.
27. Damm U, Sato T, Alizadeh A, Groenewald JZ, Crous PW. The *Colletotrichum dracaenophilum*, *C. magnum* and *C. orchidearum* species complexes. *Stud Mycol*. 2019;92:1–46.
28. O'Connell RJ, Thon MR, Hacquard S, Amyotte SG, Kleemann J, Torres MF, et al. Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. *Nat Genet*. 2012;44(9):1060–5.
29. Dean R, Van Kan JAL, Pretorius ZA, Hammond-Kosack KE, Di Pietro A, Spanu PD, et al. The Top 10 fungal pathogens in molecular plant pathology. *Mol Plant Pathol*. 2012;13(4):414–30.
30. De Silva DD, Crous PW, Ades PK, Hyde KD, Taylor PWJ. Life styles of *Colletotrichum* species and implications for plant biosecurity. *Fungal Biol Rev*. 2017;31(3):155–68.
31. Sukno SA, García VM, Shaw BD, Thon MR. Root infection and systemic colonization of maize by *Colletotrichum graminicola*. *Appl Environ Microbiol*. 2008;74(3):823–32.
32. Pande S, Mughogho L, Bandyopadhyay R, Karunakarr R. Variation in pathogenicity and cultural characteristics of sorghum isolates of *Colletotrichum graminicola* in India. *Plant Dis*. 1991;75:778–83.
33. Eastburn DM, Gubler W. Strawberry Anthracnose: Detection and survival of *Colletotrichum acutatum* in soil. *Plant Dis*. 1990;74(2):161.
34. Rhaïem A, Taylor PWJ. *Colletotrichum gloeosporioides* associated with anthracnose symptoms on citrus, a new report for Tunisia. *Eur J Plant Pathol*. 2016;146(1):219–24.
35. Lelwala R V., Scott JB, Ades PK, Taylor PWJ. Population structure of *Colletotrichum tanacetii* in Australian pyrethrum reveals high evolutionary potential. *Phytopathology*. 2019;109(10):1779–92.
36. Charron C, Hubert J, Minatchy J, Wilson V, Chrysot F, Gerville S, et al. Characterization of *Colletotrichum orchidophilum*s, the agent of black spot disease of vanilla. *J Phytopathol*. 2018;166(7–8):525–31.
37. Wilken PM, Steenkamp ET, Wingfield MJ, de Beer ZW, Wingfield BD. Which *MAT* gene? Pezizomycotina (Ascomycota) mating-type gene nomenclature reconsidered. *Fungal Biol Rev*. 2017;31(4):199–211.

38. Marcelino J, Giordano R, Gouli S, Gouli V, Parker BL, Skinner M, et al. *Colletotrichum acutatum* var. *florinae* (teleomorph: *Glomerella acutata* var. *florinae* var. nov.) infection of a scale insect. *Mycologia*. 2008;100(3):353–74.
39. Martin SH, Wingfield BD, Wingfield MJ, Steenkamp ET. Causes and consequences of variability in peptide mating pheromones of ascomycete fungi. *Mol Biol Evol*. 2011;28(7):1987–2003.
40. Pöggeler S. Two pheromone precursor genes are transcriptionally expressed in the homothallic ascomycete *Sordaria macrospora*. *Curr Genet*. 2000;37(6):403–11.
41. Mayrhofer S, Pöggeler S. 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*. 2005;4(4):661–72.
42. Mayrhofer S, Weber JM, Pöggeler S. Pheromones and pheromone receptors are required for proper sexual development in the homothallic ascomycete *Sordaria macrospora*. *Genetics*. 2006;172(3):1521–33.
43. Turina M, Prodi A, Van Alfen NK. Role of the *Mf1-1* pheromone precursor gene of the filamentous ascomycete *Cryphonectria parasitica*. *Fungal Genet Biol*. 2003;40(3):242–51.
44. Wang L, Lin X. Mechanisms of unisexual mating in *Cryptococcus neoformans*. *Fungal Genet Biol*. 2011;48(7):651–60.
45. Wilson AM, Godlonton T, van der Nest MA, Wilken PM, Wingfield MJ, Wingfield BD. Unisexual reproduction in *Huntiaella moniliformis*. *Fungal Genet Biol*. 2015;80:1–9.
46. Turrà D, El Ghalid M, Rossi F, Di Pietro A. Fungal pathogen uses sex pheromone receptor for chemotropic sensing of host plant signals. *Nature*. 2015;527(7579):521–4.
47. Hane JK, Williams AH, Taranto AP, Solomon PS, Oliver RP. Repeat-induced point mutation: A fungal-specific, endogenous mutagenesis process. In: van den Berg M, Maruthachalam K, editors. *Genetic transformation systems in fungi*. Springer International Publishing; 2015. p. 55–68.
48. van Wyk S, Wingfield BD, De Vos L, van der Merwe NA, Santana QC, Steenkamp ET. Repeat-induced point mutations drive divergence between *Fusarium circinatum* and its close relatives. *Pathogens*. 2019;8(4).
49. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003;19(2):215–25.
50. Marchler-bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. *Nucleic Acids Res*. 2015;43:D222-226.
51. Gasteiger E, Hoogland C, Gattiker A, Duvaud S, Wilkins MR, Appel RD, et al. Protein identification and analysis tools on the ExPASy Server. In: Walker J, editor. *The Proteomics Protocols Handbook*. Totowa, NJ: Springer: Humana Press Inc.; 2005. p. 571–608.
52. Krogh A, Sonnhammer ELL, Ka L. Advantages of combined transmembrane topology and signal peptide prediction – the Phobius web server. *Nucleic Acids Res*. 2007;35:W429–32.

53. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
54. Nylander JAA. MrModeltest v2. Distrib by author. 2004;
55. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–42.
56. Rambaut A. FigTree. Distrib by author. 2018;
57. Maddison WP, Maddison DR. Mesquite: A modular system for evolutionary analysis. V2.7. Distrib by author. 2011;
58. Van Wyk S, Harrison CH, Wingfield BD, De Vos L, Van Der Merwe NA, Steenkamp ET. The RIPper, a web-based tool for genome-wide quantification of Repeat-Induced Point (RIP) mutations. *PeerJ.* 2019;2019(8):1–18.
59. Gan P, Tsushima A, Hiroshima R, Narusaka M, Takano Y, Narusaka Y, et al. *Colletotrichum shioi* sp. nov., an anthracnose pathogen of *Perilla frutescens* in Japan: Molecular phylogenetic, morphological and genomic evidence. *Sci Rep.* 2019;9(1):1–13.
60. Baroncelli R, Sukno SA, Sarrocco S, Cafà G, Le Floch G, Thon MR. Whole-genome sequence of the orchid anthracnose pathogen *Colletotrichum orchidophilum*. *Mol Plant-Microbe Interact.* 2018;31(10):979–81.
61. Gan P, Narusaka M, Tshushima A, Narusaka Y, Takano Y, Shirasu K. Draft genome assembly of *Colletotrichum chlorophyti*, a pathogen of herbaceous plants. *Genome Announc.* 2017;5:e01733-16.
62. Liang X, Cao M, Li S, Kong Y, Rollins JA, Zhang R, et al. Highly contiguous genome resource of *Colletotrichum fructicola* generated using long-read sequencing. *Mol Plant-Microbe Interact.* 2020;33(6):790–3.
63. Meng Y, Gleason ML, Zhang R, Sun G. Genome sequence resource of the wide-host-range anthracnose pathogen *Colletotrichum siamense*. *Mol Plant-Microbe Interact.* 2019;32(8):931–4.
64. Nakamura M, Fujikawa T, Nakamori D, Iwai H. Draft genome sequence of *Colletotrichum sansevieriae* Sa-1–2, the anthracnose pathogen of *Sansevieria trifasciata*. *Data Br.* 2018;18:691–5.
65. Rao S, Nandineni MR. Genome sequencing and comparative genomics reveal a repertoire of putative pathogenicity genes in chilli anthracnose fungus *Colletotrichum truncatum*. *PLoS One.* 2017;12:1-13.
66. Rogério F, Bouffleur TR, Ciampi-Guillardi M, Sukno SA, Thon MR, Massola Júnior NS, et al. Genome sequence resources of *Colletotrichum truncatum*, *C. plurivorum*, *C. musicola*, and *C. sojae*: Four species pathogenic to soybean (*Glycine max*). *Phytopathology.* 2020;110(9):1497–9.
67. Gan P, Tsushima A, Narusaka M, Narusaka Y, Takano Y, Kubo Y, et al. Genome sequence resources for four phytopathogenic fungi from the *Colletotrichum orbiculare* species complex. *Mol Plant-Microbe Interact.* 2019;32(9):1088–90.
68. Seidl MF, Faino L, Shi-Kunne X, Van Den Berg GCM, Bolton MD, Thomma BPHJ. The genome of the saprophytic fungus *Verticillium tricorpus* reveals a complex effector repertoire resembling that of its

pathogenic relatives. Mol Plant-Microbe Interact. 2015;28(3):362–73.

69. Klosterman SJ, Subbarao K V., Kang S, Veronese P, Gold SE, Thomma BPHJ, et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. PLoS Pathog. 2011;7(7).

Figures

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 harbours a greatly expanded MAT1 locus. The second configuration is that of *C. truncatum*, which is also expanded, but not to the extent of *C. camelliae*. Most of the remaining species exhibit the third or fourth configuration, with the major difference being the position of the MAT1-2-13 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 (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 to scale.

A

B

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 harbours a greatly expanded MAT1 locus. The second configuration is that of *C. truncatum*, which is also expanded, but not to the extent of *C. camelliae*. Most of the remaining species exhibit the third or fourth configuration, with the major difference being the position of the MAT1-2-13 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 (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 to scale.

Figure 2

The diversity found at the α -factor pheromone locus in the *Colletotrichum* genus. Representative gene maps are shown to illustrate the differences in gene content, position and orientation at this locus. These diagrams are not drawn to scale. Species clades are indicated in the following colours: *acutatum* (red), *caudatum* (dark orange), *graminicola* (light orange), *spaethianum* (gold), *destructivum* (green), *gloeosporioides* (dark blue), *boninense* (light blue), *truncatum* (purple), *orcidearum* (pink) and *orbiculare* (maroon). Refer to Table S2 for details regarding gene identities and accession numbers.

Figure 2

The diversity found at the α -factor pheromone locus in the *Colletotrichum* genus. Representative gene maps are shown to illustrate the differences in gene content, position and orientation at this locus. These diagrams are not drawn to scale. Species clades are indicated in the following colours: acutatum (red), caudatum (dark orange), graminicola (light orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), boninense (light blue), truncatum (purple), orcidarium (pink) and orbiculare (maroon). Refer to Table S2 for details regarding gene identities and accession numbers.

Figure 3

Ancestral state recombination of the presence/absence of the α -factor pheromone (A) and -factor receptor (B) genes. Black lines represent the presence of the gene, while grey lines represent the absence of the gene. Red bars signify putative gene loss according to ancestral state recombination. Further genetic analysis showed that certain loss events are comprised of multiple independent losses and are indicated by blue bars. The asterisks indicates that this gene could not be identified due to the flanking genes being present at the ends of different contigs that could not be joined. Species clades are indicated in the following colours: acutatum (red), caudatum (dark orange), graminicola (light orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), boninense (light blue), truncatum (purple), orcidarium (pink) and orbiculare (maroon).

Figure 3

Ancestral state recombination of the presence/absence of the α -factor pheromone (A) and -factor receptor (B) genes. Black lines represent the presence of the gene, while grey lines represent the absence of the gene. Red bars signify putative gene loss according to ancestral state recombination. Further genetic analysis showed that certain loss events are comprised of multiple independent losses and are indicated by blue bars. The asterisks indicates that this gene could not be identified due to the flanking genes being present at the ends of different contigs that could not be joined. Species clades are indicated in the following colours: *acutatum* (red), *caudatum* (dark orange), *graminicola* (light orange), *spaethianum* (gold), *destructivum* (green), *gloeosporioides* (dark blue), *boninense* (light blue), *truncatum* (purple), *orcidarium* (pink) and *orbiculare* (maroon).

Figure 4

Genetic events that resulted in loss of the α -factor pheromone in *Colletotrichum* species clades. In *C. acutatum*, *C. fiorinae* 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 4

Genetic events that resulted in loss of the α -factor pheromone in *Colletotrichum* species clades. In *C. acutatum*, *C. fiorinae* 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 in this locus in *C. tofieldiae*. Gene arrow colours are the same as those from Fig. 2.

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 in this locus in *C. tofieldiae*. Gene arrow colours are the same as those from Fig. 2.

Figure 6

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.

Figure 6

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.

Figure 7

Ancestral state recombination 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 grey 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 are indicated by blue bars. Species clades are indicated in the following colours: acutatum (red), caudatum (dark orange), graminicola (light orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), boninense (light blue), truncatum (purple), orcidearum (pink) and orbiculare (maroon).

Figure 7

Ancestral state recombination 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 grey 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 are indicated by blue bars. Species clades are indicated in the following colours: acutatum (red), caudatum (dark orange), graminicola (light orange), spaethianum (gold), destructivum (green), gloeosporioides (dark blue), boninense (light blue), truncatum (purple), orcidearum (pink) and orbiculare (maroon).

Figure 8

Genetic events that resulted in loss of the a-factor pheromone in *Colletotrichum* species clades. Many species of the *C. acutatum* clade have premature in frame stop codons that truncate the protein 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*, *C. phormi* 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 at the 5' end 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 harbouring the CaaX domain can likely be produced. These diagrams are not drawn to scale.

Figure 8

Genetic events that resulted in loss of the a-factor pheromone in *Colletotrichum* species clades. Many species of the *C. acutatum* clade have premature in frame stop codons that truncate the protein 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*, *C. phormi* 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 at the 5' end 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 harbouring the CaaX domain can likely be produced. These diagrams are not drawn to scale.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFile1SuppTable1.xlsx](#)
- [AdditionalFile1SuppTable1.xlsx](#)

- [AdditionalFile2SuppTable2.xlsx](#)
- [AdditionalFile2SuppTable2.xlsx](#)
- [AdditionalFile3SuppTable3.xlsx](#)
- [AdditionalFile3SuppTable3.xlsx](#)
- [AdditionalFile4SuppTable4.xlsx](#)
- [AdditionalFile4SuppTable4.xlsx](#)
- [AdditionalFile5GeneRegions.txt](#)
- [AdditionalFile5GeneRegions.txt](#)