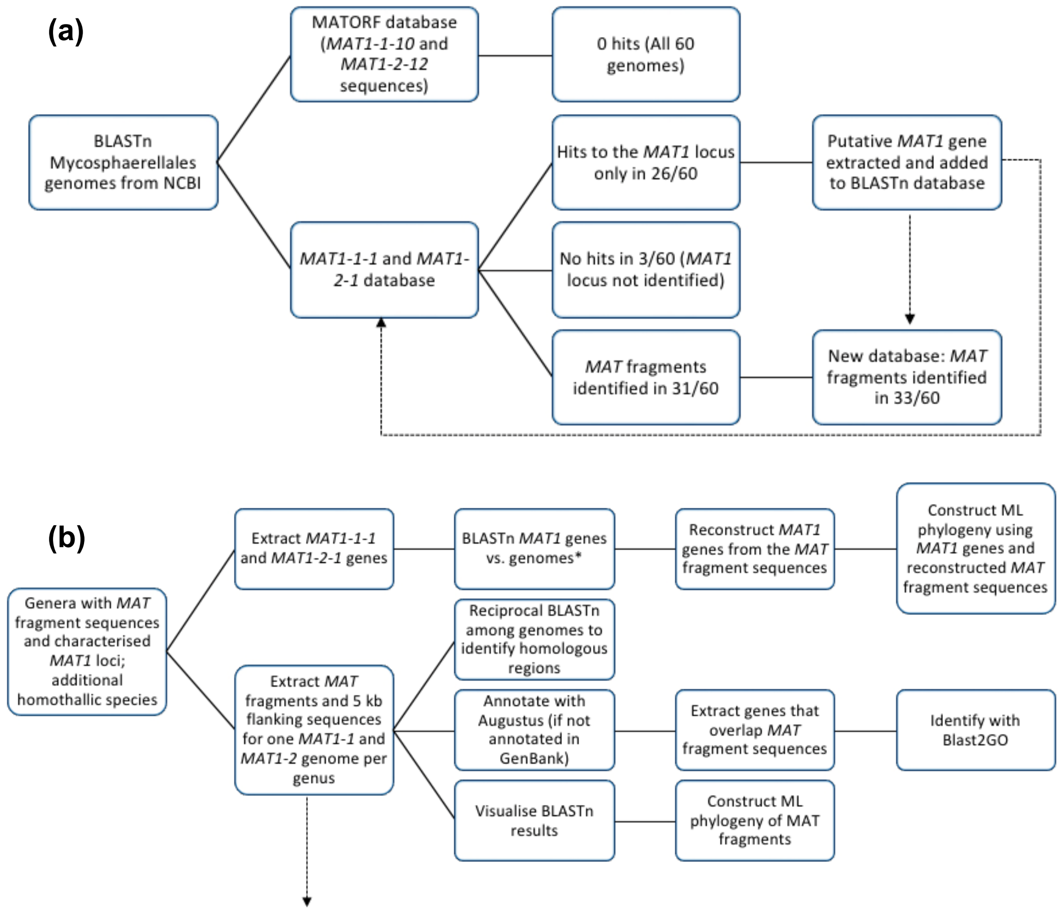
Supplementary Figures

Figure S1



Genus	***MAT1-1* genome**	***MAT1-2* genome**
Cercospora	*C. beticola* 09-40	*C. berteroae* CBS538.71
Dothistroma	*D. pini* CBS 116487	*D. septosporum* NZE10
Nothophaeocryptopus	*N. gaeumannii* CBS 267.37 (homothallic)	
Pallidocercospora	*P. crystallina* XJS617 (homothallic)	
Passalora	Not available	*Pa. fulva* CBS 131901
Pseudocercospora	*Ps. musae* CBS 116634	*Ps. eumusae* CBS 114824
Teratosphaeria	*T. zuluensis* CBS 119470	*T. destructans* CMW 44962
	T. nubilosa CBS 116005 (homothallic)	

* A *MAT1-1-1* and *MAT1-2-1* gene was reconstructed from each genome. In species for which the *MAT1-1-1* or *MAT1-2-1* gene sequence is not known the genome was the *MAT1* gene of the closest available relative (as shown by Fig. 1) was used as query in the BLASTn analysis.

Figure S1 Workflow for the (a) identification and (b) analysis of *MAT* fragments in the Mycosphaerellales genomes.

Figure S2

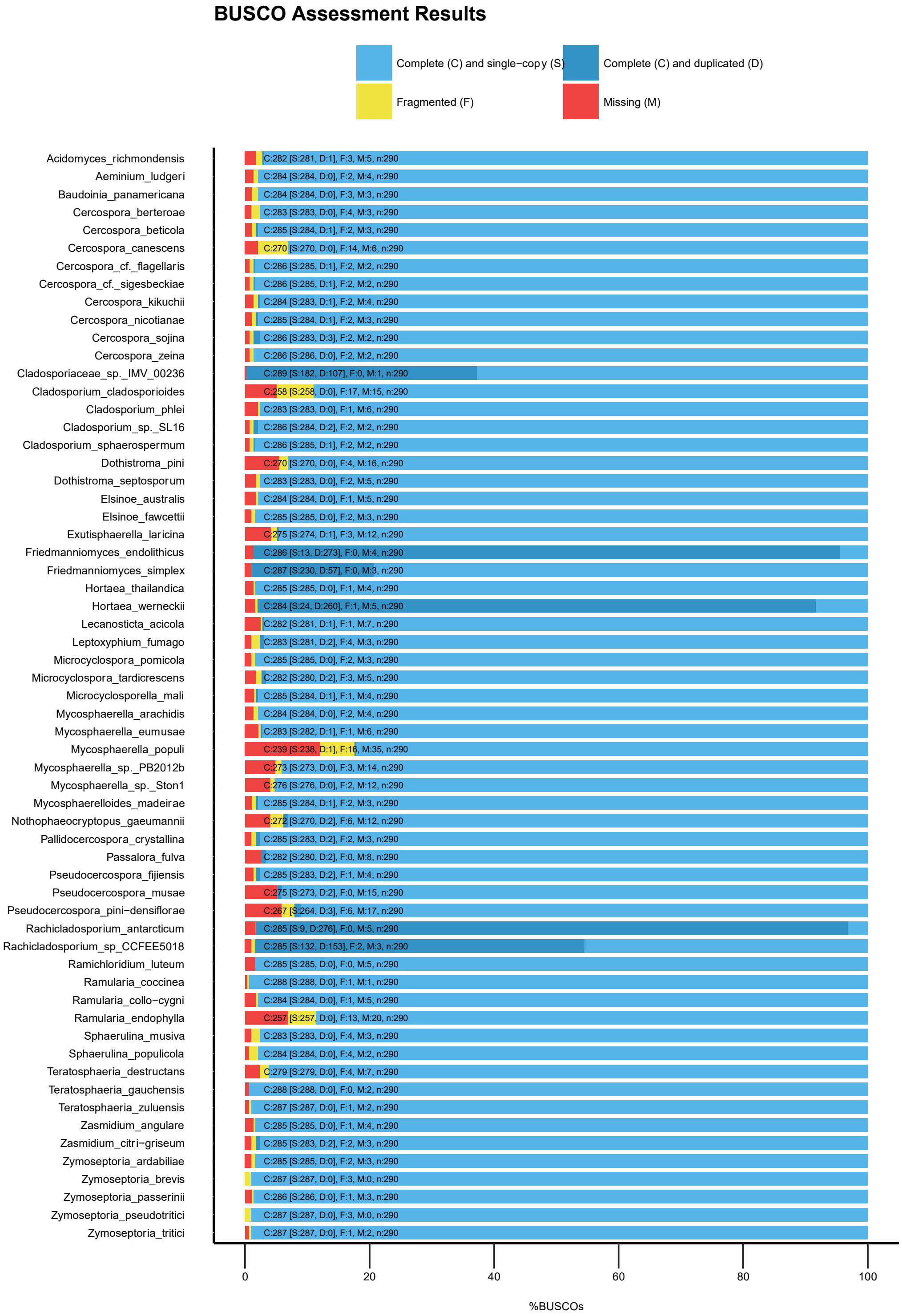


Figure S2 Histogram depicting the "Fungi *odb9*" BUSCO orthologs identified in each genome. As per the legend, orthologs are classified as being "complete and single-copy", "complete and duplicated", "fragmented" or "missing".

Figure S3a

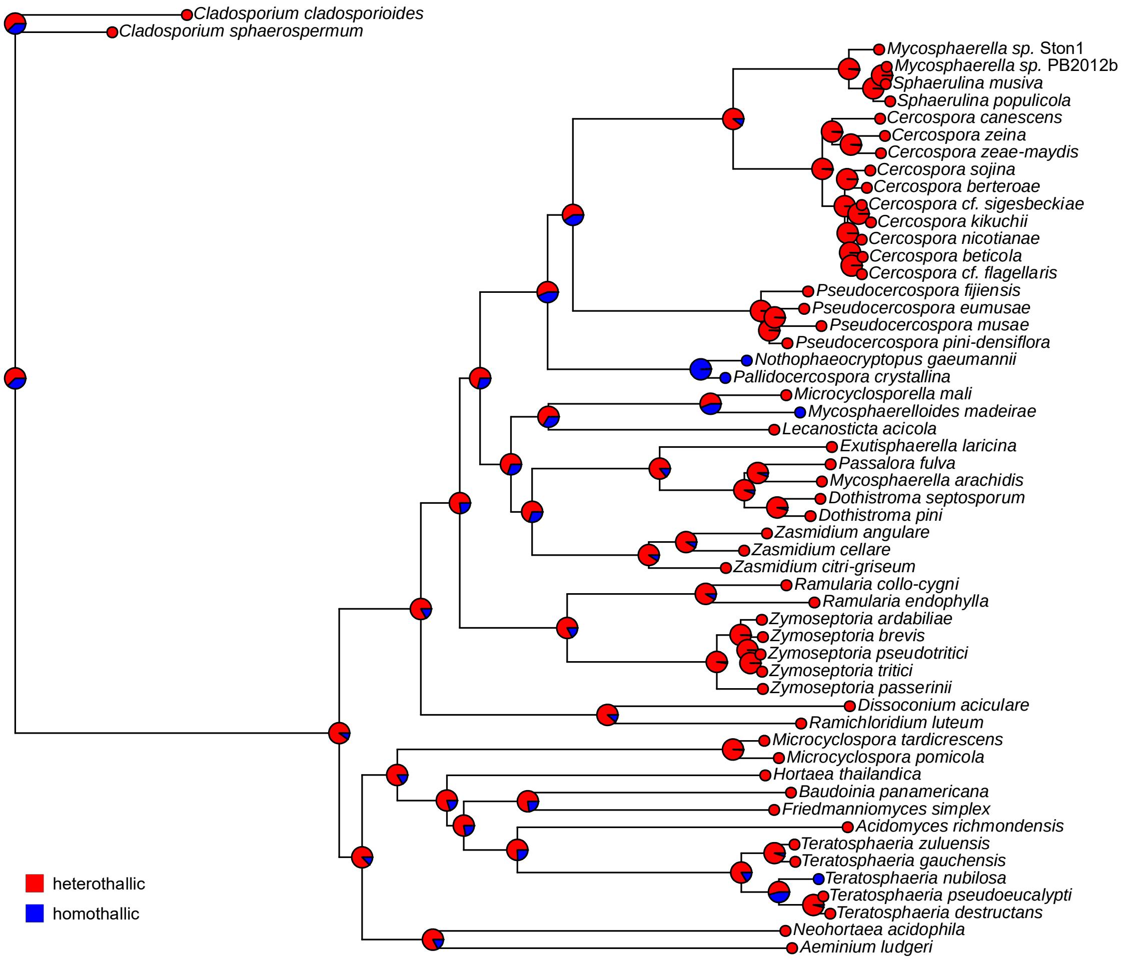

Cladosporium cladosporioides
Cladosporium sphaerospermum
Mycosphaerella sp. Ston1
Mycosphaerella sp. PB2012b
Sphaerulina musiva
Sphaerulina populicola
Cercospora canescens
Cercospora zeina
Cercospora zeae-maydis
Cercospora sojina
Cercospora berteroae
Cercospora cf. sigesbeckiae
Cercospora kikuchii
Cercospora nicotianae
Cercospora beticola
Cercospora cf. flagellaris
Pseudocercospora fijiensis
Pseudocercospora eumusae
Pseudocercospora musae
Pseudocercospora pini-densiflora
Nothophaeocryptopus gaeumannii
Pallidocercospora crystallina
Microcyclosporella mali
Mycosphaerelloides madeirae
Lecanosticta acicola
Exutisphaerella laricina
Passalora fulva
Mycosphaerella arachidis
Dothistroma septosporum
Dothistroma pini
Zasmidium angulare
Zasmidium cellare
Zasmidium citri-griseum
Ramularia collo-cygni
Ramularia endophylla
Zymoseptoria ardabiliae
Zymoseptoria brevis
Zymoseptoria pseudotritici
Zymoseptoria tritici
Zymoseptoria passerinii
Dissoconium aciculare
Ramichloridium luteum
Microcyclospora tardicrescens
Microcyclospora pomicola
Hortaea thailandica
Baudoinia panamericana
Friedmanniomyces simplex
Acidomyces richmondensis
Teratosphaeria zuluensis
Teratosphaeria gauchensis
Teratosphaeria nubilosa
Teratosphaeria pseudoeucalypti
Teratosphaeria destructans
Neohortaea acidophila
Aeminium ludgeri
🟥 heterothallic
🟦 homothallic

Figure S3b

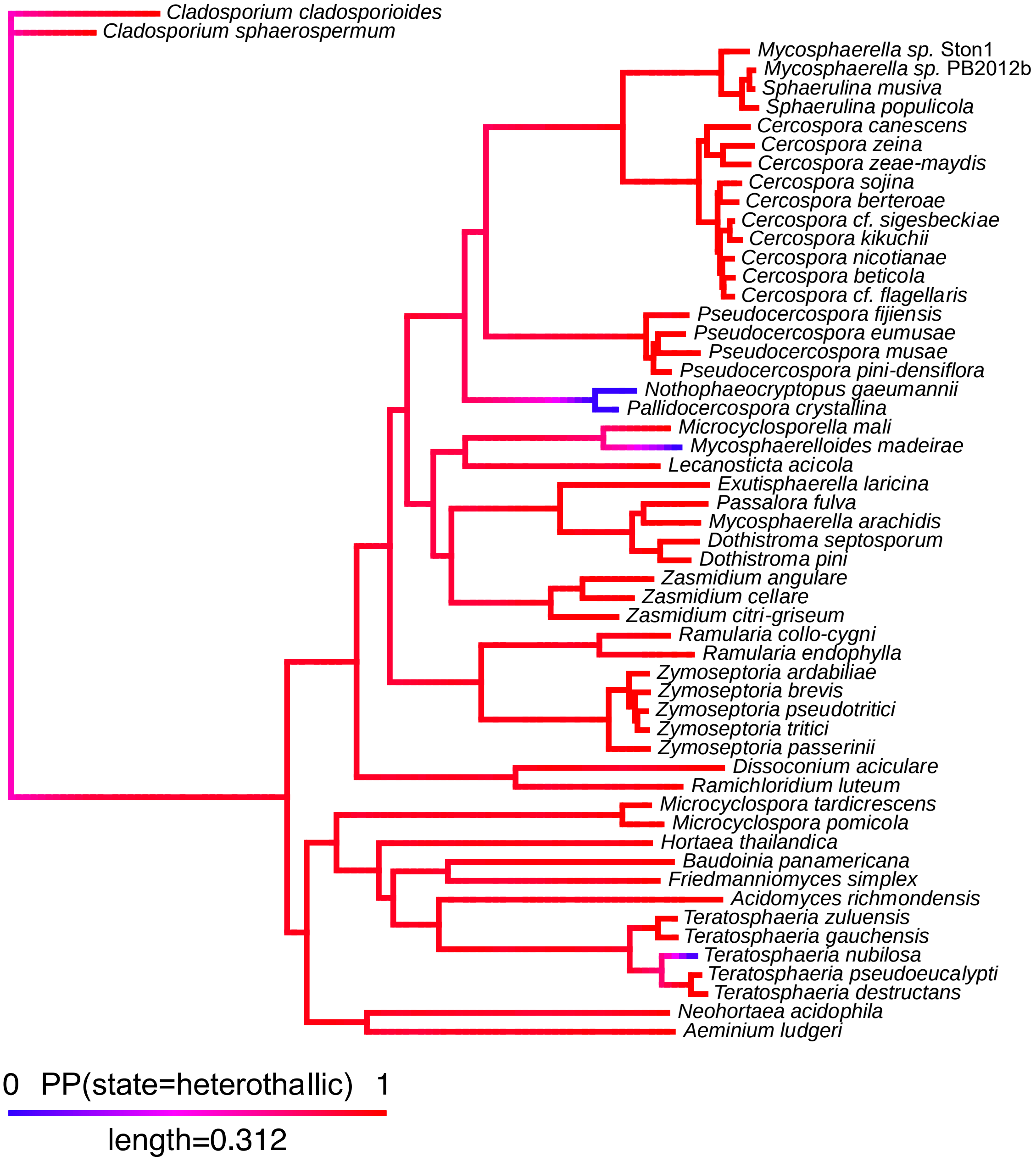
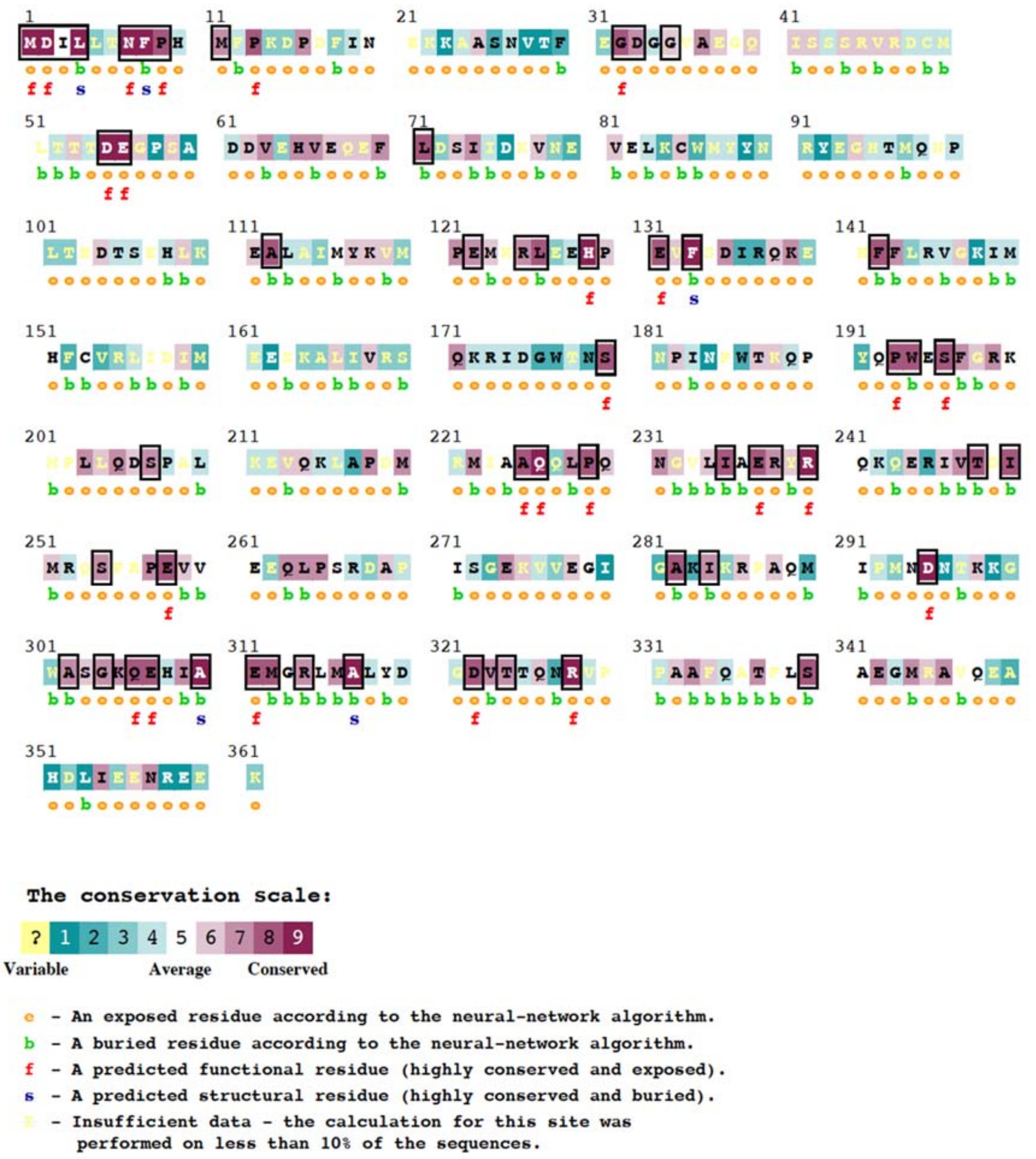


Figure S3 Estimation of ancestral thallic states using (a) a continuous-time Markov chain (Mk) model and (b) stochastic character mapping. Red represents heterothallism and blue homothallism. In (a), pie charts at internal nodes show the empirical Bayesian posterior probabilities for each state, whereas posterior probabilities for the heterothallic state are visualised as a density map in (b).

Figure S4a



The conservation scale:

? 1 2 3 4 5 6 7 8 9

Variable Average Conserved

e - An exposed residue according to the neural-network algorithm.

b - A buried residue according to the neural-network algorithm.

f - A predicted functional residue (highly conserved and exposed).

s - A predicted structural residue (highly conserved and buried).

X - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure S4b


The conservation scale:

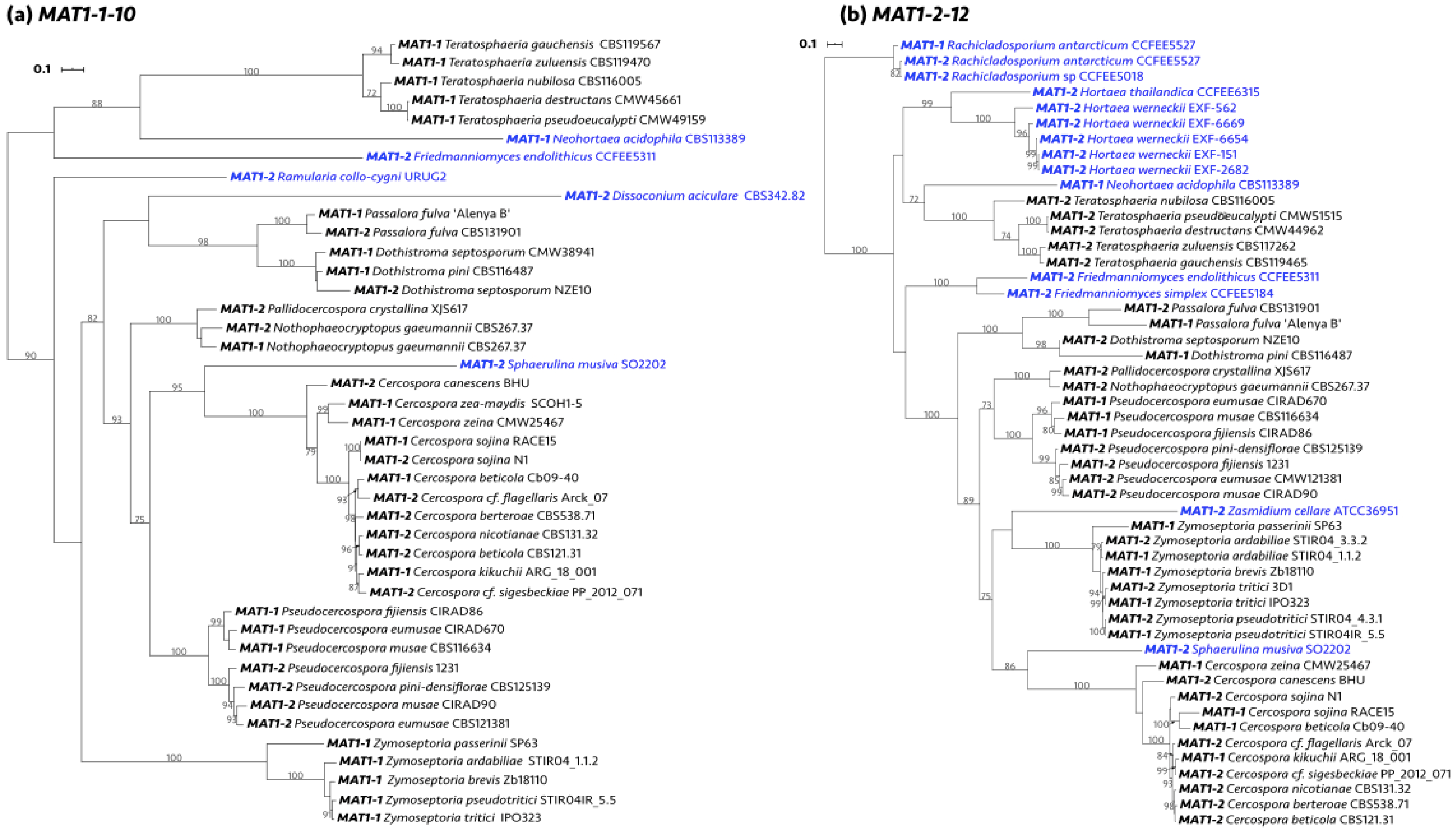
? 1 2 3 4 5 6 7 8 9

Variable Average Conserved

- o - An exposed residue according to the neural-network algorithm.
- b - A buried residue according to the neural-network algorithm.
- f - A predicted functional residue (highly conserved and exposed).
- s - A predicted structural residue (highly conserved and buried).
- ▨ - Insufficient data - the calculation for this site was performed on less than 10% of the sequences.

Figure S4 Conserved residues of the hypothetical (a) MAT1-1-10 and (b) MAT1-2-12 proteins based on sequences derived from the *MAT1-1* and *MAT1-2* idiomorphs, respectively. Conservation was predicted in ConSeq using all available sequences from genera with characterised *MAT1* loci, although only the *Pseudocercospora eumusae* sequence is shown. Conservation scores are indicated according to the colour legend, with 9 being most conserved. Black boxes denote residues that are also highly conserved (conservation scale 8 and 9) in the alignments of sequences derived from both the *MAT1-1* and the *MAT1-2* idiomorphs. In the MAT1-2-12 sequence, lines (—) denote residues predicted to be under purifying selection, whereas asterisks (*) show those predicted to be under diversifying selection in the *MAT1-1* copies of *MAT1-2-12* (see Supplementary File 1 and main text for details).

Figure S5



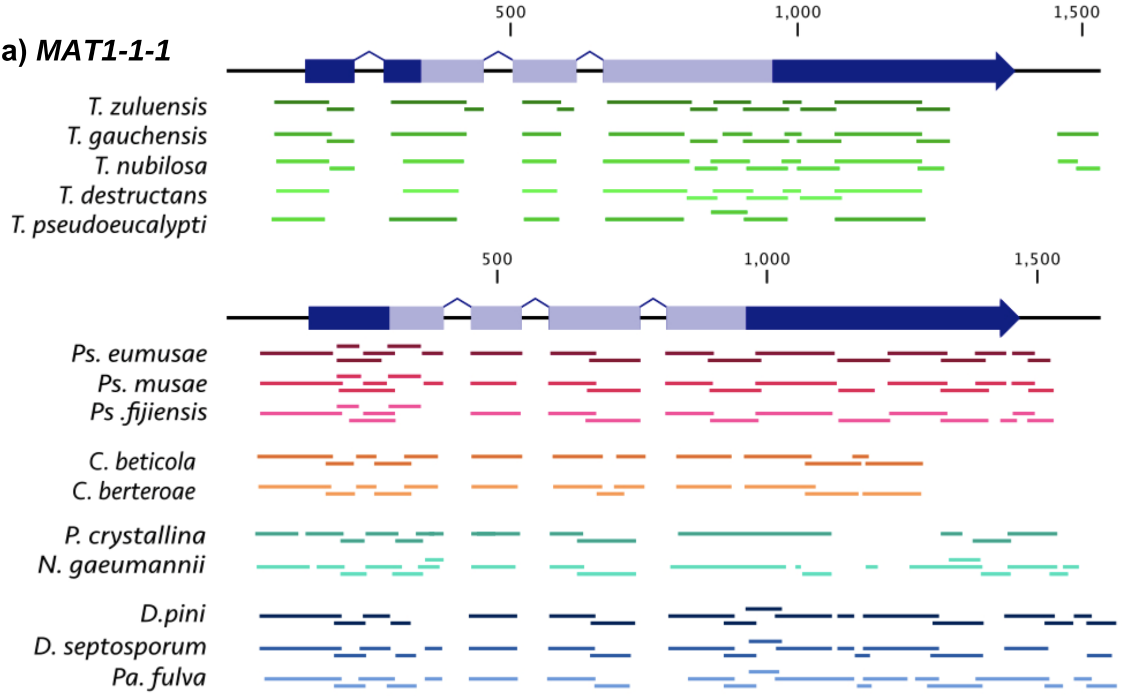
(a) *MAT1-1-10*

(b) *MAT1-2-12*

Figure S5 Phylogeny of *MAT1-1-10* (a) and *MAT1-2-12* (b) coding sequences, including the sequences of Mycosphaerellales homologs identified from GenBank via BLASTp and hmmsearch (blue). *MAT1-1* and *MAT1-2* indicates the idiomorph in which the copy of *MAT1-1-10* or *MAT1-2-12* was identified. The *MAT1-1-10* tree was rooted on the branch leading to the Teratosphaeriaceae genes, whereas the *MAT1-2-12* tree was rooted using the *Rachicladosporium* (Cladosporiales) genes.

Figure S6

(a) *MAT1-1-1*



(b) *MAT1-2-1*

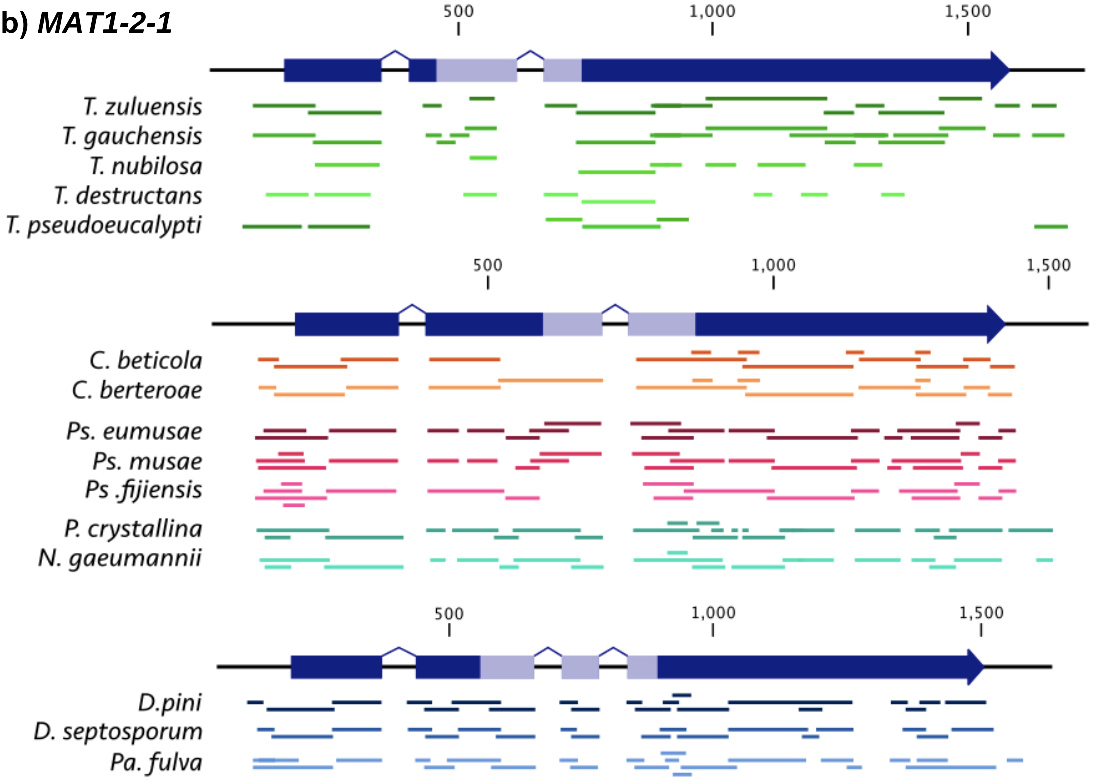
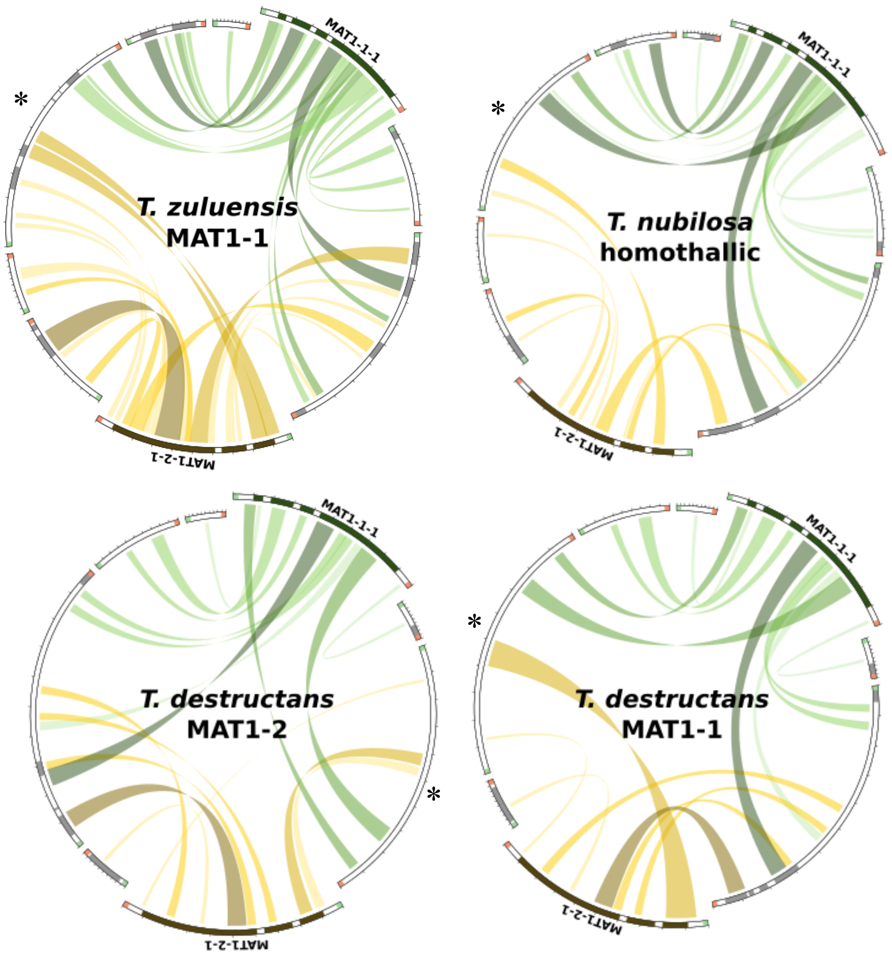
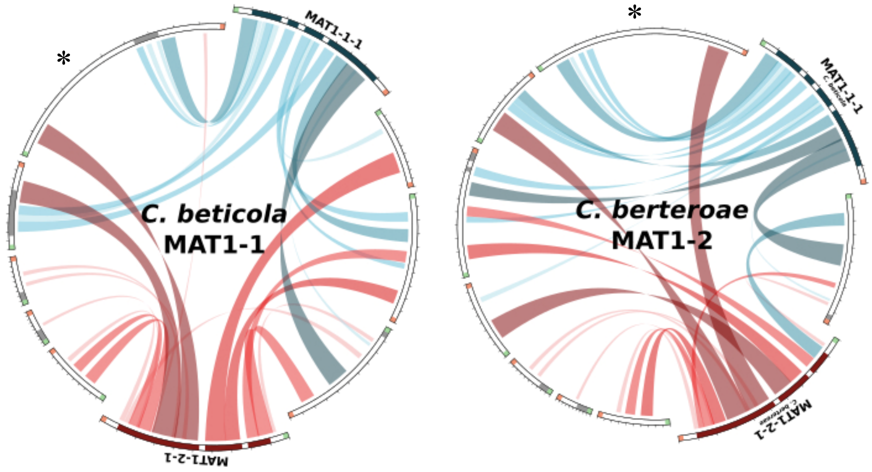


Figure S6 BLASTn alignments of the *MAT* fragment sequences of *Cercospora*, *Dothistroma*, *Nothophaeocryptopus, Passalora, Pseudocercospora, Pallidocercospora* and *Teratosphaeria* species against their *MAT1-1-1* (a) and *MAT1-2-1* (b) genes. Dark blue arrows represent *MAT* coding sequences and light regions within the arrows indicate the position of the conserved MAT alpha_HMG box domain (IPR006856) in *MAT1-1-1* and the HMG-box domain (IPR036910) in *MAT1-2-1*. For each species, both the *MAT1-1* and *MAT1-2* fragments originated from the same genome.

Figure S7

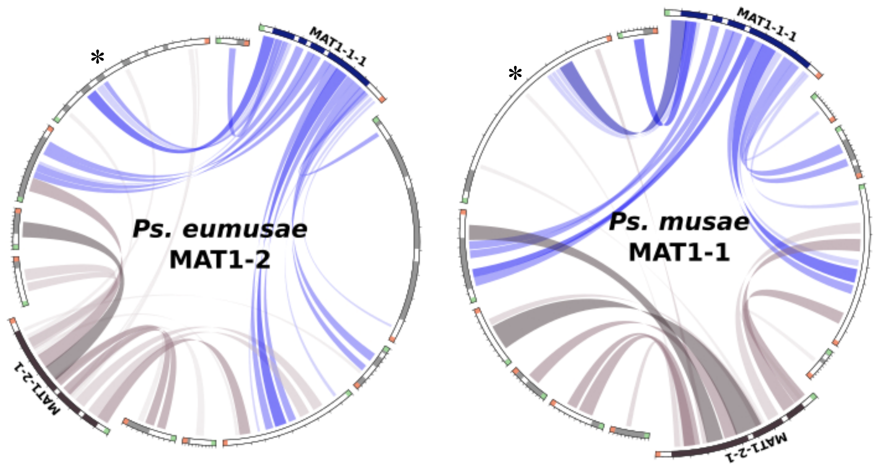


(a) ***Teratosphaeria***

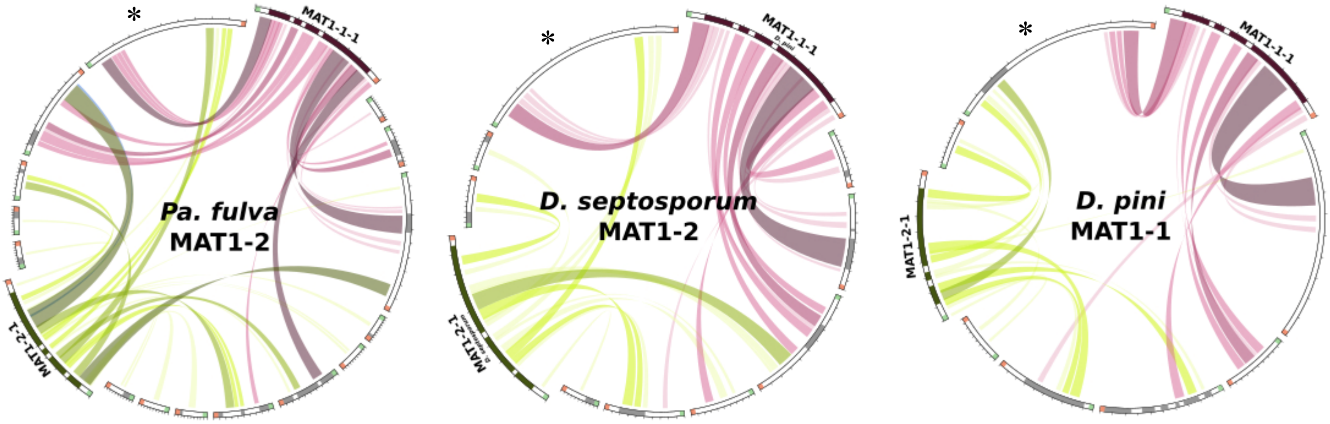


(b) ***Cercospora***

(c) *Pseudocercospora*



(d) *Passalora* and *Dothistroma*



(e) *Nothophaeocryptopus* and *Pallidocercospora*

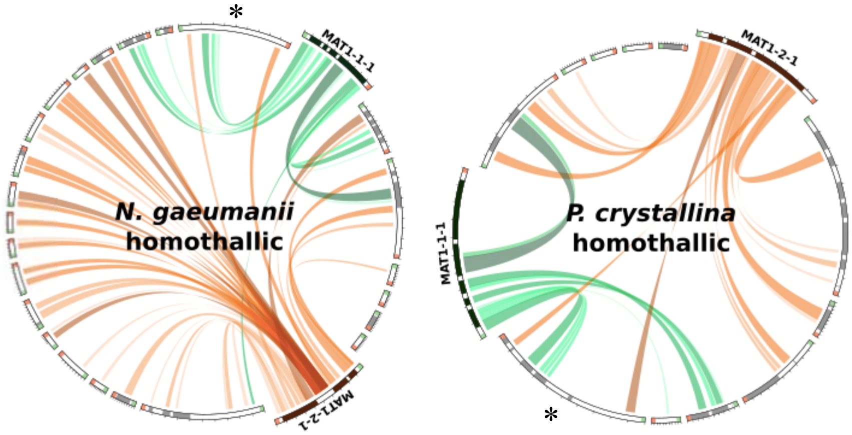


Figure S7 Circoletto plots illustrating the presence of *MAT* fragment sequences in the genomes of representative Mycosphaerellaceae and Teratosphaeriaceae species. Each plot shows the two *MAT1* genes (offset), originating from the same isolate in homothallic species and from different isolates in heterothallic species. Coloured ribbons connect regions of the *MAT1* genes (query) to their corresponding BLASTn hits (*MAT* fragment sequences) on contigs in the genome. All plots are based on contigs from a single genome. Darker coloured ribbons represent lower (i.e. stronger) e-values. The coding regions of the *MAT1* genes are indicated as dark boxes, grey boxes represent predicted ORFs in the genomic regions and white spaces represent introns or non-coding sequence. The 5' to 3' ends are read in the clockwise direction. All alignments have e-values < $1x10^{-5}$ and nucleotide identities > 80%. Asterisks indicate a fragment region that is homologous among all species and used in Fig. S8.

Figure S8

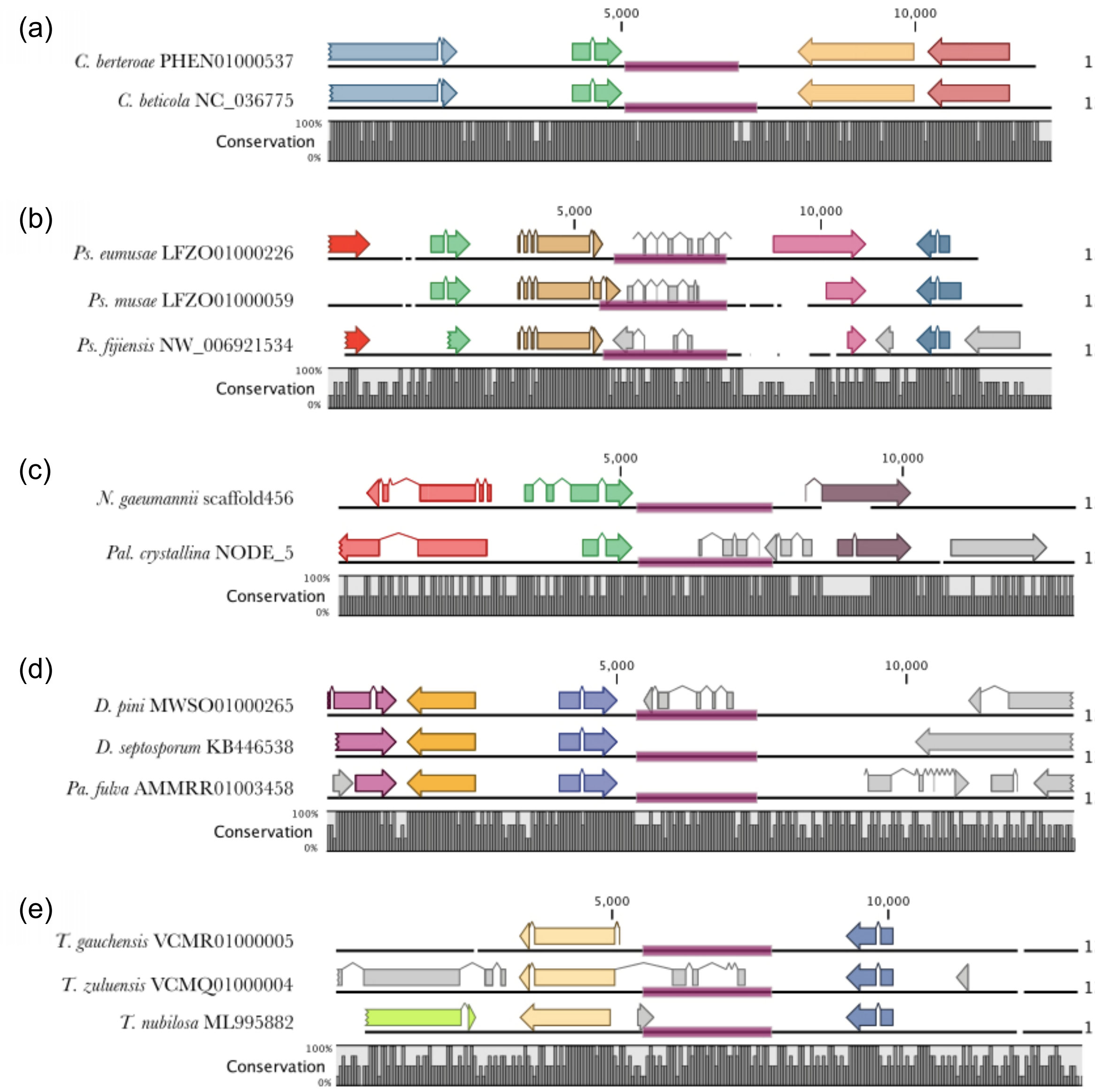


Figure S8 Synteny around the largest region containing *MAT* fragment sequences (pink block) identified in the examined Mycosphaerellaceae and Teratosphaeriaceae genomes. Synteny was visible only within genera: (a) *Cercospora*, (b) *Pseudocercospora*, (e) *Teratosphaeria*; or between closely related species: (c) *Nothophaeocryptopus gaeumannii* and *Pallidocercospora crystallina* and (d) *Dothistroma* and *Passalora fulva*. Colours indicate homologous genes, whereas gene predictions without homologs are shown in gray. *Cercospora* (a), *Pseudocercospora* (b) and *N. gaeumannii* and *P. crystallina* (c) had putative "cell division control (Cdc31)" genes (green) upstream of the region containing *MAT* fragment sequences. Codes after species names represent contig numbers. The *MAT* fragment region used in this figure is also marked in Fig. S5

Figure S9

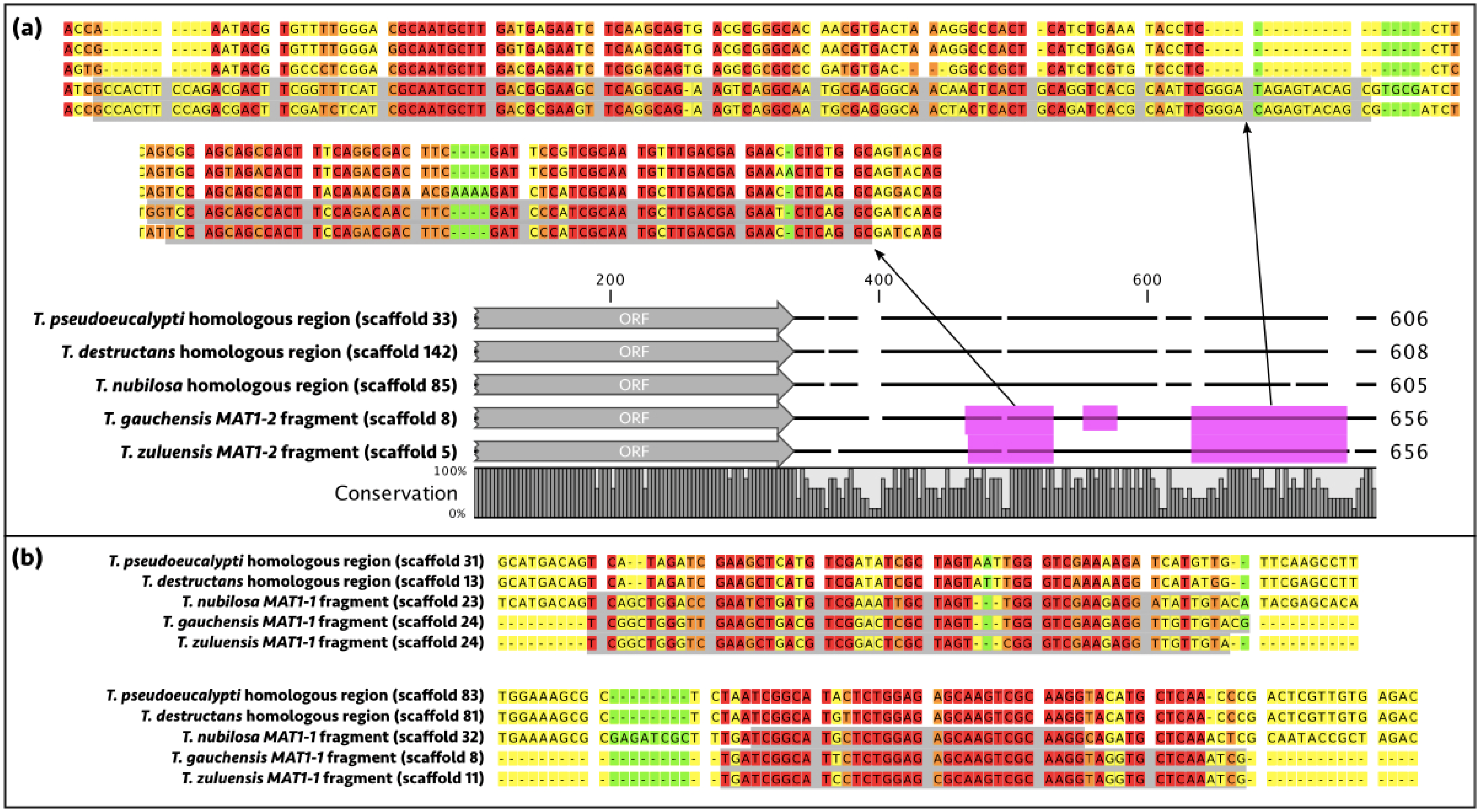

(a)
ACCA------ ----AATACG TGTTTTGGGA CGCAATGCTT GATGAGAATC TCAAGCAGTG ACGCGGGCAC AACGTGACTA AAGGCCCACT -CATCTGAAA TACCTC---- ---------- -------CTT
ACCG------ ----AATACG TGTTTTGGGA GGCAATGCTT GGTGAGAATC TCAAGCAGTG ACGCGGGCAC AACGTGACTA AAGGCCCACT -CATCTGAGA TACCTC---- ---------- -------CTT
AGTG------ ----AATACG TGCCCTCGGA CGCAATGCTT GACGAGAATC TCGGACAGTG AGGCGCGCCC GATGTGAC-- --GGCCCGCT -CATCTCGTG TCCCTC---- ---------- -------CTC
ATCGCCACTT CCAGACGACT TCGGTTTCAT CGCAATGCTT GACGGGAAGC TCAGGCAG-A AGTCAGGCAA TGCGAGGGCA ACAACTCACT GCAGGTCACG CAATTCGGGA TAGAGTACAG CGTGCGATCT
ACCGCCACTT CCAGACGACT TCGATCTCAT CGCAATGCTT GACGCGAAGT TCAGGCAG-A AGTCAGGCAA TGCGAGGGCA ACTACTCACT GCAGATCACG CAATTCGGGA CAGAGTACAG CG----ATCT
AGCGC AGCAGCCACT TTCAGGCGAC TTC----GAT TCCGTCGCAA TGTTTGACGA GAAC-CTCTG GCAGTACAG
AGTGC AGTAGACACT TTCAGACGAC TTC----GAT TCCGTCGCAA TGTTTGACGA GAAAACTCTG GCAGTACAG
AGTCC AGCAGCCACT TACAAACGAA ACGAAAAGAT CTCATCGCAA TGCTTGACGA GAAC-CTCAG GCAGGACAG
GGTCC AGCAGCCACT TCCAGACAAC TTC----GAT CCCATCGCAA TGCTTGACGA GAAT-CTCAG GCGATCAAG
ATTCC AGCAGCCACT TCCAGACGAC TTC----GAT CCCATCGCAA TGCTTGACGA GAAC-CTCAG GCGATCAAG
200
400
600
T. pseudoeucalypti homologous region (scaffold 33)
T. destructans homologous region (scaffold 142)
T. nubilosa homologous region (scaffold 85)
T. gauchensis MAT1-2 fragment (scaffold 8)
T. zuluensis MAT1-2 fragment (scaffold 5)
ORF
ORF
ORF
ORF
ORF
606
608
605
656
656
100%
Conservation
0%
(b)
T. pseudoeucalypti homologous region (scaffold 31)
T. destructans homologous region (scaffold 13)
T. nubilosa MAT1-1 fragment (scaffold 23)
T. gauchensis MAT1-1 fragment (scaffold 24)
T. zuluensis MAT1-1 fragment (scaffold 24)
GCATGACAGT CA--TAGATC GAAGCTCATG TCGATATCGC TAGTAATTGG GTCGAAAAGA TCATGTTG-- TTCAAGCCTT
GCATGACAGT CA--TAGATC GAAGCTCATG TCGATATCGC TAGTATTTGG GTCAAAAAGG TCATATGG-- TTCGAGCCTT
TCATGACAGT CAGCTGGACC GAATCTGATG TCGAAATTGC TAGT---TGG GTCGAAGAGG ATATTGTACA TACGAGCACA
---------T CGGCTGGGTT GAAGCTGACG TCGGACTCGC TAGT---TGG GTCGAAGAGG TTGTTGTACG ----------
---------T CGGCTGGGTC GAAGCTGACG TCGGACTCGC TAGT---CGG GTCGAAGAGG TTGTTGTA-- ----------
T. pseudoeucalypti homologous region (scaffold 83)
T. destructans homologous region (scaffold 81)
T. nubilosa MAT1-1 fragment (scaffold 32)
T. gauchensis MAT1-1 fragment (scaffold 8)
T. zuluensis MAT1-1 fragment (scaffold 11)
TGGAAAGCG C--------T CTAATCGGCA TACTCTGGAG AGCAAGTCGC AAGGTACATG CTCAA-CCCG ACTCGTTGTG AGAC
TGGAAAGCG C--------T CTAATCGGCA TGTTCTGGAG AGCAAGTCGC AAGGTACATG CTCAA-CCCG ACTCGTTGTG AGAC
TGAAAAGCG CGAGATCGCT TTGATCGGCA TGCTCTGGAG AGCAAGTCGC AAGGCAGATG CTCAAACTCG CAATACCGCT AGAC
--------- ---------- -TGATCGGCA TTCTCTGGAG AGCAAGTCGC AAGGTAGGTG CTCAAATCG- ---------- ----
--------- ---------- -TGATCGGCA TCCTCTGGAG CGCAAGTCGC AAGGTAGGTG CTCAAATCG- ---------- ----

Figure S9 Homologous regions in the genomes of five *Teratosphaeria* species where BLASTn identified *MAT* fragment sequences in some species, but not in others. *MAT* fragments are indicated by grey and purple boxes in the zoomed in nucleotide alignments and alignment overview, respectively. (a) A region homologous across all five genomes and containing a conserved predicted open reading frame (ORF), in which *MAT1-2* fragments were identified in *T. gauchensis* and *T. zuluensis*, but not in *T. destructans*, *T. pseudoeucalypti* and *T. nubilosa*. (b) Two *MAT1-1* fragments identified in *T. gauchensis*, *T. zuluensis* and *T. nubilosa*, but not in *T. destructans* and *T. pseudoeucalypti*. The regions flanking these fragments are not conserved in *T. gauchensis* and *T. zuluensis* and have been excluded from the alignment. Colours in the zoomed in alignments indicate conservation: red = 100%; orange = 80%; yellow = 40-60%; green = 20%.