

RESEARCH ARTICLE

# Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses

Xiaofei Liang<sup>1</sup>, Bo Wang<sup>1</sup>, Qiuyue Dong<sup>1</sup>, Lingnan Li<sup>1</sup>, Jeffrey A. Rollins<sup>2</sup>, Rong Zhang<sup>1\*</sup>, Guangyu Sun<sup>1\*</sup>

**1** State Key Laboratory of Crop Stress Biology in Arid Areas and College of Plant Protection, Northwest A&F University, Yangling, Shaanxi Province, China, **2** Department of Plant Pathology, University of Florida, Gainesville, United States of America

\* [rongzh@nwsuaf.edu.cn](mailto:rongzh@nwsuaf.edu.cn) (RZ); [sgy@nwsuaf.edu.cn](mailto:sgy@nwsuaf.edu.cn) (GS)



**OPEN ACCESS**

**Citation:** Liang X, Wang B, Dong Q, Li L, Rollins JA, Zhang R, et al. (2018) Pathogenic adaptations of *Colletotrichum* fungi revealed by genome wide gene family evolutionary analyses. PLoS ONE 13 (4): e0196303. <https://doi.org/10.1371/journal.pone.0196303>

**Editor:** Sabrina Sarrocco, Università degli Studi di Pisa, ITALY

**Received:** January 3, 2018

**Accepted:** April 10, 2018

**Published:** April 24, 2018

**Copyright:** © 2018 Liang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The *Colletotrichum fructicola* 1104-7 genome assembly generated in this study was deposited at GenBank under accession number MVNS00000000.

**Funding:** This work was supported by Chinese Universities Scientific Fund (Z109021712, Z109021610), the National Science Foundation of China (31601595), the General Financial Grant from the China Postdoctoral Science Foundation (2016M592844), the China Agriculture Research System (CARS-28) and the USDA National Institute

## Abstract

The fungal genus *Colletotrichum* contains hemibiotrophic phytopathogens being highly variable in host and tissue specificities. We sequenced a *C. fructicola* genome (1104–7) derived from an isolate of apple in China and compared it with the reference genome (Nara\_gc5) derived from an isolate of strawberry in Japan. Mauve alignment and BlastN search identified 0.62 Mb lineage-specific (LS) genomic regions in 1104–7 with a length criterion of 10 kb. Genes located within LS regions evolved more dynamically, and a strongly elevated proportion of genes were closely related to non-*Colletotrichum* sequences. Two LS regions, containing nine genes in total, showed features of fungus-to-fungus horizontal transfer supported by both gene order collinearity and gene phylogeny patterns. We further compared the gene content variations among 13 *Colletotrichum* and 11 non-*Colletotrichum* genomes by gene function annotation, OrthoMCL grouping and CAFE analysis. The results provided a global evolutionary picture of *Colletotrichum* gene families, and identified a number of strong duplication/loss events at key phylogenetic nodes, such as the contraction of the detoxification-related RTA1 family in the monocot-specializing graminicola complex and the expansions of several ammonia production-related families in the fruit-infecting gloeosporioides complex. We have also identified the acquisition of a RbsD/FucU fucose transporter from bacterium by the *Colletotrichum* ancestor. In sum, this study summarized the pathogenic evolutionary features of *Colletotrichum* fungi at multiple taxonomic levels and highlights the concept that the pathogenic successes of *Colletotrichum* fungi require shared as well as lineage-specific virulence factors.

## Introduction

The *Colletotrichum* genus is genetically diverse, comprising over 100 Ascomycota fungal species grouped into 10 major species complexes or species sensu lato [1, 2]. *Colletotrichum* species are also overwhelmingly successful phytopathogens, causing anthracnose foliar blight or

of Food and Agriculture, Hatch project 1005726. The funders had no role in the design of the study or the collection, analysis or interpretation of the data or the writing of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

fruit/stem rot on more than 3,000 plant species [3], and generating large economic losses on crops, vegetables, and fruit trees worldwide. While many *Colletotrichum* species are phytopathogenic, some interact with plants as endophytes, live freely as saprobes, or exhibit more than one lifestyle. In rare cases, *Colletotrichum* spp have been known to cause opportunistic animal infections [4, 5].

Most *Colletotrichum* pathogens penetrate the plant cuticle using melanized appressoria. Upon penetration, they differentiate infectious hyphae which spread intercellularly and/or intracellularly, and pass through biotrophic and necrotrophic infection phases sequentially [6]. Host interaction style varies among pathogen species, host organ/tissue types and plant developmental stages [7]. For instance, the biotrophic phase of *C. higginsianum* is limited to the first invaded epidermal cell whereas that of *C. graminicola* is present both at the advanced lesion margin and in the central colonization areas [3]. Species belonging to the gloeosporioides and actutaum complexes cause post-harvest fruit rots, in which the pathogens actively penetrate young fruit, persist quiescently for months, and reinitiate colonization when fruit begin to mature [8]. Host senescence or wounding can trigger the switch from quiescent endophyte to pathogenic colonizer [9–11].

Given the diverse taxonomic lineages and plant-interaction styles, it is difficult to assign a genus-wide representative pathogen model for study purposes, as knowledge gained from one pathosystem may not be directly transferred to another [7]. Yet, these variations may manifest through a unified mechanistic principle, where host defense levels and pathogen ‘stealth’ strategies together shape the interaction type (endotroph, biotroph, or necrotroph) and the time points at which phase shifts occur [7]. The entire genus may share both conserved and novel virulence factors tailored in lineage-specific manners for host/tissue adaptation. Identifying these virulence factors and characterizing their evolution are critical for *Colletotrichum* disease control and for better understanding the fundamental mechanisms of host-pathogen interaction. Comparative genomics thorough genome sampling both in and outside of the genus is an approach with high potential to identify virulence factors.

Currently, a dozen *Colletotrichum* genomes are publicly available [3, 12–18]. These representatives of the genus belong to six independent species complexes encompassing different plant-interaction styles (including endophytes, monocot and dicot foliar pathogens, and fruit pathogens). These genomes have been analyzed either separately or in combination to identify genomic features associated with host-adaptive evolutions [15–17], which concordantly reveal that *Colletotrichum* species may tailor their plant cell wall degrading enzymes (PCWDEs) and proteinases in accordance with their own infection styles. Thus the contents of these genomes are more likely to be grouped based on host range similarity rather than phylogenetic relatedness [15, 16]. *Colletotrichum* genomes are also known to be enriched with enzymes catalyzing secondary metabolite biosynthesis, many of which show phase-specific expressions during infection [3].

*Colletotrichum fructicola* is a recently established species belonging to the economically important gloeosporioides species complex. It is globally distributed and has a very broad host range, including over 50 plant species distributed in eight different families [7]. Diseases caused by *C. fructicola* are important economic concerns on many crops such as strawberry, apple, pear, and oil tea. On apple, natural *C. fructicola* isolates show pathogenic variation related to tissue/cultivar specificities [19], indicating that this broad host range species is made up of individual host-limited forms. A *C. fructicola* strain isolated from strawberry, Nara-gc5, has been genome sequenced [13], providing a reference for gene function studies and genome comparison purposes.

In this study, we sequenced a *C. fructicola* strain isolated from an apple Glomerella leaf spot lesion in China and performed gene content comparison encompassing 13 *Colletotrichum*

and 11 non-*Colletotrichum* genomes. The objectives of this study were several fold: first, by comparing representative *Colletotrichum* genomes with non-*Colletotrichum* genomes, we expected to identify genomic features conserved across the entire *Colletotrichum* genus, e.g. gene functions being genus-specific or expanded prior to the genus divergence; second, by characterizing gene content variation of different *Colletotrichum* species complexes, we expected to identify factors related to host adaptations among distinctive *Colletotrichum* lineages; third, to compare the intraspecific gene content variation between the two *C. fructicola* genomes derived from isolates of different hosts.

## Materials and methods

### Fungal isolate, sequencing, assembling and annotations

The *C. fructicola* 1104–7 isolate was obtained from an apple *Glomerella* leaf spot lesion in a private orchard in Hebei Province, China. Its *C. fructicola* species identity was confirmed by multi-locus concatenation phylogeny. The leaf sample was collected with the permission of the orchard owner. Pathogenicity test demonstrated that the isolate could cause apple bitter rot (ABR) and *Glomerella* leaf spot (GLS). The isolate was self-fertile and produced the *Glomerella* teleomorph in culture. Its morphological characteristics and sexual behavior fit the ‘plus’ strain descriptions [20, 21]. The isolate was cultured on potato dextrose agar and preserved as a 15% glycerol conidial stock at -80°C, and was deposited in the Agricultural Culture Collection of China (ACCC) under the accession number ACCC39328. Genomic DNA was extracted with freshly-collected mycelia from a 4-day potato dextrose broth shake culture (150 rpm, room temperature) based on a modified cetyl trimethylammonium bromide (CTAB) procedure [22]. Genome sequencing was performed with an Illumina HiSeq 2000 platform at the Novogene Genomic Sequencing Center, Beijing, China. The mean insertion size of sequencing libraries was 350 bp and the sequencing strategy was 100-bp pair-ends. Raw reads were trimmed with an in house perl script to remove low quality reads ( $N > 10\%$ , or  $sQ \leq 5$ ) and reads with adaptor contamination. Clean reads were then *de novo* assembled using the AbySS assembler version 1.3.5 [23], with a Kmer value of 50. GapFiller version 2.0 [24] was used to further fill gaps and generate scaffolds. The generated genome assembly was deposited at GenBank under accession no. MVNS00000000.

Repetitive DNA elements were predicted with a combination of RepeatMasker version 4.0.5 and RepeatModeler version 1.0.8. To predict gene structures, Augustus version 3.1 [25], SNAP version 2013-11-29 [26], GeneMark-ES version 2.3c [27], and MAKER2 version 2.31.8 [28] were used in combination. Augustus and SNAP were trained with gene models of the JGI *Glomerella cingulate* 23 strain (<http://genome.jgi.doe.gov/programs/fungi/index.jsf>), GeneMark-ES was self-trained. Prediction results of Augustus, GeneMark-ES, and protein models of *G. cingulate* 23 were combined for a final MAKER2 integration. Predicted genes were functionally annotated with the Blast2GO software [29], putative functions were assigned based on BLASTP search against a local NCBI nr database (release date: 2016-09-01). Predicted transcript sequences and gene annotations were deposited as supplemental information. BUSCO version 1.2 [30] was used to evaluate the completeness of genome assembly and gene predictions. Genome alignment of 1104–7 and Nara\_gc5 was performed with Mauve software version 2.4.0 [31], and single nucleotide polymorphism sites (SNPs) were extracted with the SNP-sites software [32].

### Phylogenomic analysis and gene family evolution

Predicted proteins encoded by a total of 24 fungal genomes (accessions listed in Table A in [S1 File](#)) were filtered (removing those containing less than 70 aa), and clustered into orthologous

groups (Table B in S2 File) by OrthoMCL version 2.0.9 with an inflation value of 1.5 [33]. Single copy ortholog groups were then extracted for phylogenomic tree construction. Independent ortholog groups were aligned with MAFFT version 7 (<http://mafft.cbrc.jp/alignment/server>) and the conserved sites were extracted and concatenated with Gblocks version 0.91b [34]. Based on the concatenated dataset, a maximum-likelihood (ML) phylogenetic tree was constructed with RAxML version 8.1.1 [35] using the LG+G+I model chosen by ProtTest version 3.4 [36] with the bootstrap value set as 1,000.

Based on the ML dendrogram generated above, a calibrated species tree was constructed with the r8s software version 1.7 [37], analyses were based on penalized likelihood method and the TN algorithm. The *Colletotrichum* crown, Sordariomycetes crown, and Sordariomycetes-Leotiomycetes crown were chosen as calibration points [17, 38], predictions from a combination of four calibration schemes and three smoothing factors were compared to estimate divergence ranges. CAFE program version 3.1 was used for gene family expansion/contraction analysis [39], a universal lambda value (maximum likelihood value of the birth & death parameter) was assumed, and the best value was obtained by iterative calculations. Families showing significant size variance were identified based on 1,000 random samples and a p-value cutoff of 0.01, deviated branches were further identified based on the Viterbi algorithm in CAFE with a p-value cutoff of 0.05.

### Gene function predictions

Putative protein domains were identified by querying against a local Interproscan database (Jones et al. 2014). SMURF (<http://jcv.org/smurf/index.php>) was used to predict putative secondary metabolite genes and clusters with the default parameters except that terpene cyclases (TCs) were identified by Hmmscan in HMMer version 3.0 [40] using the PFAM domain PF03936 (e-value, 1E-03). Candidate transcription factors (TFs) were identified with Hmmscan based on reported TF domains [41] with a cut-off e-value of 1E-03. Candidate cytochrome P450s (P450s) were identified by Hmmscan with PFAM domain PF00067 (cut-off e-value, 1E-03), and further classified into families and subfamilies following BLASTP against all named fungal CYPs (<http://blast.uthsc.edu/>). For family/subfamily assignment, the international cytochrome P450 nomenclature criteria were followed (i.e. P450s showing >40% identity were assigned to the same family) [42]. Candidate transporters were identified based on the TransportTP server (<http://bioinfo3.noble.org/transporter/>) with an e-value threshold of 1E-05. Candidate *Colletotrichum*-genus-specific genes were identified by BLASTP search against a local NCBI fungal database excluding *Colletotrichum* sequences (cut-off e-value, 1E-05).

Secretomes were identified using a procedure similar to that previously reported [43], in which SignalP version 4.1 [44], TMHMM Server version 2.0 [45], GPI-SOM [46] and WoLF PSORT [47] were run sequentially. Putative proteases were identified and classified by BLASTP querying against the MEROPS database (<http://merops.sanger.ac.uk/>) with a cut-off e-value of 1E-04, sequences containing mutated active sites or incomplete domains were removed. Carbohydrate utilizing enzymes were identified and classified based on BLASTP search against carbohydrate-active enzyme (CAZY) database ([www.cazy.org](http://www.cazy.org)) with a cut-off e-value of 1E-03. Functional enrichment tests were performed with FUNRICH version 2.1.2 [48].

## Results

### General features of the *Colletotrichum fructicola* 1104–7 genome

In total, 5.8 Gb pair-ended clean reads were assembled into 686 scaffolds with a total length of 57.1 Mb. The assembly size was similar to other *Colletotrichum* genomes such as the *C. fructicola* Nara\_gc5 strain (55.6 Mb), *C. gloeosporioides* (53.2 Mb), *C. graminiicola* (57.4 Mb) and *C.*

*higginsum* (53.4 Mb). The longest scaffold was 1.8 Mb and the  $N_{50}$  length was 339 kb. The average GC content was 53.2% and approximately 2.7% of the assembly consisted of repeat elements. Based on Mauve progressive alignment (Min LCB Weight = 250, Match Seed Weight = 15), 95.01% (54.3 Mb) of the 1104–7 genome could be aligned with Nara\_gc5 (length > 500 bp), among which 50.2% (27.2 Mb) were in blocks longer than 100 kb and 97.4% (52.9 Mb) were in blocks longer than 10 kb, the aligned sequences shared 98.7% nucleotide identity and the average SNP frequency was 0.26%.

An integrative *ab-initio* approach predicted 17,827 protein-encoding genes. Among the 17,827 putative proteins, 92.46% (16,483) had at least one BLASTP hit in a local NCBI non-redundant (nr) database (e-value cut-off 1E-05), 52.9% (9,430), 64.3% (11,457), 58.1% (10,349), and 72.8% (12,973) could be annotated based on Gene Ontology (GO), Clusters of Orthologous Groups (COG), Kyoto Encyclopedia of Genes and Genomes (KEGGs) and PFAM respectively. In BUSCO analysis, 96.8% of the fungal core genes had hits as ‘complete’ and 90.5% had hits as ‘complete and single-copy’, demonstrating completeness of the annotation. Based on an independent project (Liang et al., unpublished), 84.8% (15,114) of the 1104–7 predicted genes contained at least five RNA-seq reads among a total of ~65 million tags (sequenced samples included conidia, *in vitro* appressoria, cellophane infectious hyphae, and infected plant).

### Gene content variation between the two *Colletotrichum fruticola* genomes

The 1104–7 genome was compared with the other publicly available *C. fruticola* genome, Nara\_gc5 (GenBank accession: ANPB00000000.1). To minimize annotation pipeline-related variation, the Nara\_gc5 assembly was re-annotated with the same parameters as 1104–7, 17,844 gene models were predicted in total.

Based on OrthoMCL clustering of the two genomes, 980 genes were specific to 1104–7 (unclustered or clustered only with proteins from the same genome), among which 65.3% (640) had RNA-seq evidence support (Liang et al., unpublished data), 616 (62.9%) had significant NCBI nr BlastP hit (e-value cut-off 1E-05) and 146 (14.9%) contained PFAM domains. Top enriched PFAM functions were related to DNA transposition (hAT family protein, gag, Tc5 transposase), apoptosis (caspase, NACHT), DNA binding (helix-turn-helix, zinc knuckle), protein-protein interaction (ankyrin repeats), binding (ferritin-like, CFEM) and aspartyl protease (Table 1). 1,128 genes were specific to Naga\_gc5, 708 (62.8%) had significant BlastP hits and 286 (25.4%) contained PFAM domains, top enriched functions were related to heterokaryon incompatibility, DNA transposition (DDE endonuclease; MULE transposase), protein kinase and patatin-like phospholipase activities.

Fungal lineage-specific (LS) genomic regions are often enriched with genes mediating host interactions and niche adaptations, we therefore identified and analyzed LS regions in 1104–7 and Nara\_gc5. Long (> 10 kb) and unaligned DNA blocks were identified by performing Mauve alignment, their lineage specificities were further confirmed by genome BlastN search. In total, 0.62 Mb LS regions were identified in 1104–7 (distributed on 32 contigs, containing 118 genes), 0.33 Mb LS regions were identified in Nara\_gc5 (distributed on 20 contigs, containing 72 genes). In 1104–7 and Nara\_gc5, 61.9% (73) and 39.4% (28) isolate-specific proteins had significant BlastP hits in a local NCBI non-redundant (nr) database (e-value cut-off 1E-05, coverage > 50%), the ratios were much lower compared with the genome backgrounds (approximately 90% for both). Interestingly, in 1104–7 and Nara\_gc5, 21.1% (25) and 8.4% (6) of isolate-specific genes respectively, had only non-*Colletotrichum* homologs (BLASTP e-value cut-off 1E-05, query coverage > 50%), or were more similar to non-*Colletotrichum* sequences than to *Colletotrichum* sequences (BLASTP e-values for best hits differed by at least 1E+10

**Table 1. Top enriched PFAM domains in OrthoMCL-defined isolate-specific genes in *C. fructicola* 1104–7 and Nara-gc5.**

| PFAM     | Annotation                            | Number | Fold Enrichment | B-H P-value <sup>a</sup> |
|----------|---------------------------------------|--------|-----------------|--------------------------|
| 1104–7   |                                       |        |                 |                          |
| PF05699  | HAT family C-terminal dimerisation    | 9      | 49              | 7.6E-14                  |
| PF03732  | Retrotransposon gag protein           | 8      | 44              | 1.7E-11                  |
| PF00656  | Caspase domain                        | 8      | 37              | 2.0E-10                  |
| PF05225  | Helix-turn-helix, Psq domain          | 8      | 34              | 4.4E-10                  |
| PF00098  | Zinc knuckle                          | 7      | 25              | 2.2E-07                  |
| PF13646  | HEAT repeats                          | 6      | 29              | 6.9E-07                  |
| PF12796  | Ankyrin repeats (3 copies)            | 17     | 5               | 1.4E-06                  |
| PF03221  | Tc5 transposase DNA-binding domain    | 5      | 25              | 2.5E-05                  |
| PF13650  | Aspartyl protease                     | 4      | 35              | 3.9E-05                  |
| PF13668  | Ferritin-like domain                  | 4      | 25              | 0.0003                   |
| PF05730  | CFEM domain                           | 6      | 5               | 0.002                    |
| PF05729  | NACHT domain                          | 8      | 11              | 0.008                    |
| Nara_gc5 |                                       |        |                 |                          |
| PF11702  | Protein of unknown function (DUF3295) | 3      | 32              | 0.006                    |
| PF06985  | Heterokaryon incompatibility protein  | 17     | 3               | 0.007                    |
| PF10551  | MULE transposase domain               | 2      | 40              | 0.02                     |
| PF01734  | Patatin-like phospholipase            | 4      | 15              | 0.02                     |
| PF13358  | DDE superfamily endonuclease          | 2      | 40              | 0.03                     |
| PF00069  | Protein kinase domain                 | 11     | 3               | 0.03                     |

<sup>a</sup>B-H: Benjamini-Hochberg adjusted

<https://doi.org/10.1371/journal.pone.0196303.t001>

fold). The frequencies of genes with such characteristics were only 1% in control groups made up of randomly-selected genes (Table 2, type III + type IV). As a comparison, the frequencies of genes having only *Colletotrichum* hits (Table 2, type II) were similar between LS and control groups in both 1104–7 (9% vs. 12.7%) and Nara\_gc5 (10% vs 15.5%). Thus, genes located within LS regions evolve more dynamically, and a strongly elevated proportion of genes are

**Table 2. BlastP hit characteristics of randomly-chosen genes and genes located in lineage-specific (LS) genomic regions against the NCBI nr database.**

|                       | 1104–7 Ref <sup>1</sup> | 1104–7 LS <sup>2</sup> | Nara_gc5 Ref | Nara_gc5 LS |
|-----------------------|-------------------------|------------------------|--------------|-------------|
| Type I <sup>3</sup>   | 160 (80% <sup>8</sup> ) | 33 (28%)               | 160 (80%)    | 11 (15.5%)  |
| Type II <sup>4</sup>  | 18 (9%)                 | 15 (12.7%)             | 20 (10%)     | 11 (15.5%)  |
| Type III <sup>5</sup> | 1 (0.5%)                | 22 (18.6%)             | 1 (0.5%)     | 4 (5.6%)    |
| Type IV <sup>6</sup>  | 1 (0.5%)                | 3 (2.5%)               | 1 (0.5%)     | 2 (2.8%)    |
| Type V <sup>7</sup>   | 20 (10%)                | 45 (38%)               | 18 (9%)      | 43 (61%)    |
| Total                 | 200                     | 118                    | 200          | 71          |

<sup>1</sup>Ref, genes randomly chosen from the genome

<sup>2</sup>LS, genes located in lineage-specific (LS) regions

<sup>3</sup>Type I, conserved genes having significant BlastP hits (e-value cut-off 1E-05, query coverage > 50%) both in and out of the *Colletotrichum* genus; e-value ratios for best BlastP hits ( $E_{in}/E_{out}$ ) ≤ 1E+10.

<sup>4</sup>Type II, genes having significant BlastP hits only in the *Colletotrichum* genus.

<sup>5</sup>Type III, genes having significant BlastP hits only outside of the *Colletotrichum* genus.

<sup>6</sup>Type IV, genes having better BlastP hit outside of the *Colletotrichum* genus ( $E_{in}/E_{out}$  > 1E+10).

<sup>7</sup>Type V, no BlastP hit found.

<sup>8</sup>%, Relative percentage.

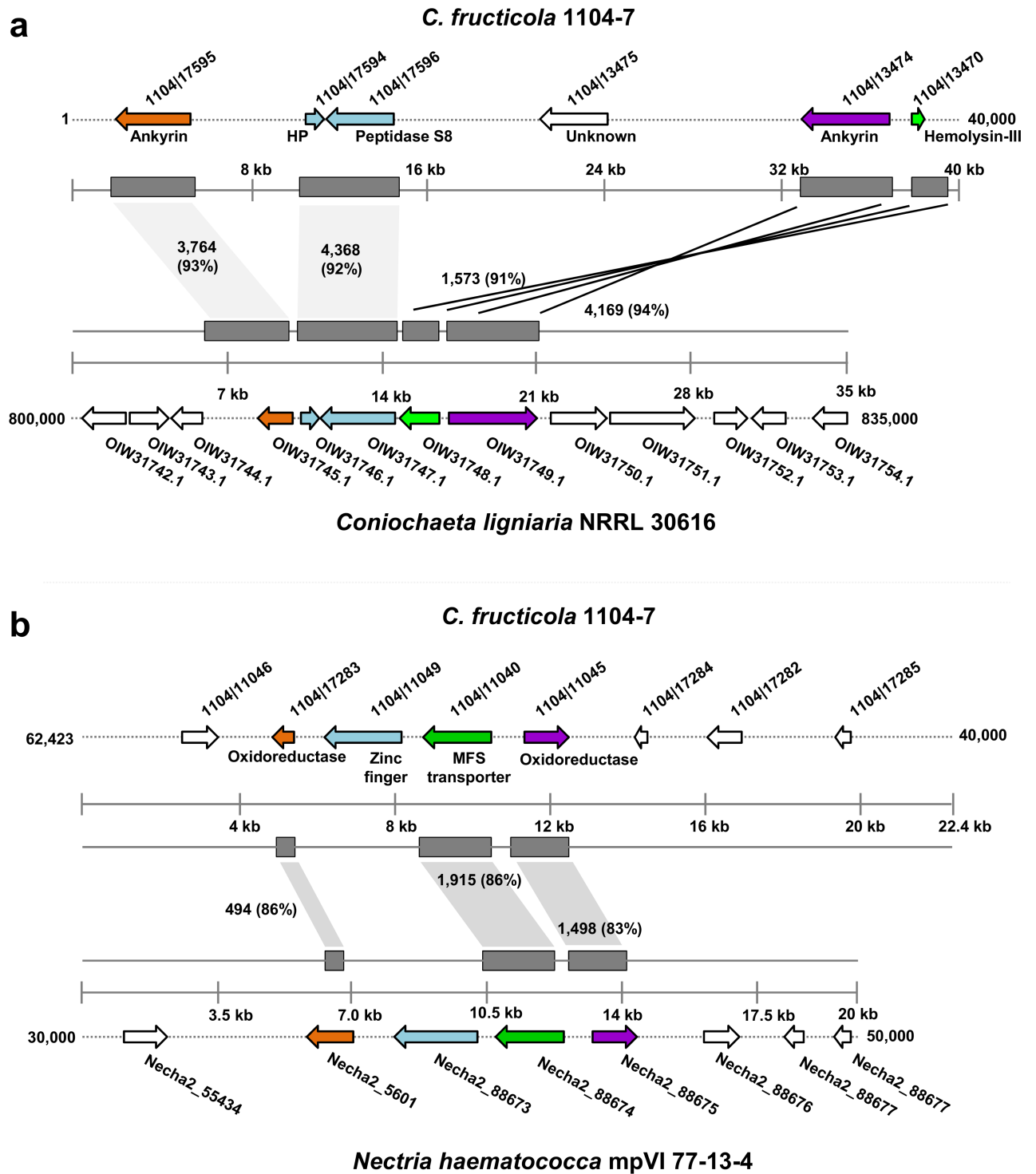
<https://doi.org/10.1371/journal.pone.0196303.t002>

closely related to non-*Colletotrichum* sequences. Phylogenetically, many non-*Colletotrichum* related LS genes were deeply-rooted with poor bootstrap support, making it difficult to infer gene evolutionary histories (data not shown). However, two putative fungus-to-fungus horizontal transfer events (HGTs), involving nine LS genes in total, were identified among non-*Colletotrichum* related genes in the 1104–7 genome. The two HGTs were supported by both gene order collinearity (Fig 1) and gene phylogeny patterns (S1 Fig). The first HGT cluster contained five genes, among which were two ankyrin proteins, one serine peptidase, one hemolysin-III domain protein, and one hypothetical protein, the cluster genes were most closely related to genes from *Nectria haematococca* and *Coniochaeta ligniaria*, and the gene orders among the three species were collinear, nucleotide identities for aligned DNA blocks reached over 90%. The second HGT cluster probably functions in secondary metabolism as it contained two oxidoreductases, one MFS transporter, and one zinc finger transcription factor. The cluster genes were most closely related to genes from *N. haematococca* and the two gene clusters were collinear, nucleotide identities for aligned DNA blocks were over 80%.

### Divergences and overall gene gain and loss patterns among *Colletotrichum* lineages

OrthoMCL clustering identified 1,212 core single-copy ortholog groups among the 24 compared *Colletotrichum* and non-*Colletotrichum* genomes. A maximum-likelihood (ML) phylogenomic tree was constructed based on their concatenated alignment. On the ML tree, all branches received 100% bootstrap value support. Lineage divergence times were then estimated in r8s, for which the combined effects of three smoothing factors (1, 100, 1,000), and four calibration schemes were tested (Table C in S2 File), the results were presented in Fig 2 and Table C in S2 File. The two *C. fructicola* strains, 1104–7 and Nara\_gc5, diverged approximately 1.3 million years (My) ago whereas *C. fructicola* and *C. gloeosporioides* diverged approximately 4.5 My ago. The gloeosporioides complex includes two phylogenetic clades, Musae and Kahawae [49], the fact that both *C. fructicola* and *C. gloeosporioides* belong to the Musae clade precluded origin estimation for the gloeosporioides complex. Origins for the other three complexes (graminicola, spaethianum, and acutatum) were similar, ranging between 9.0 and 13 My ago. The gloeosporioides and acutatum complexes, two pathogen groups commonly associated with post-harvest fruit infections, diverged by at least 47 My (the shortest divergence estimation for gloeosporioides and orbiculare complexes).

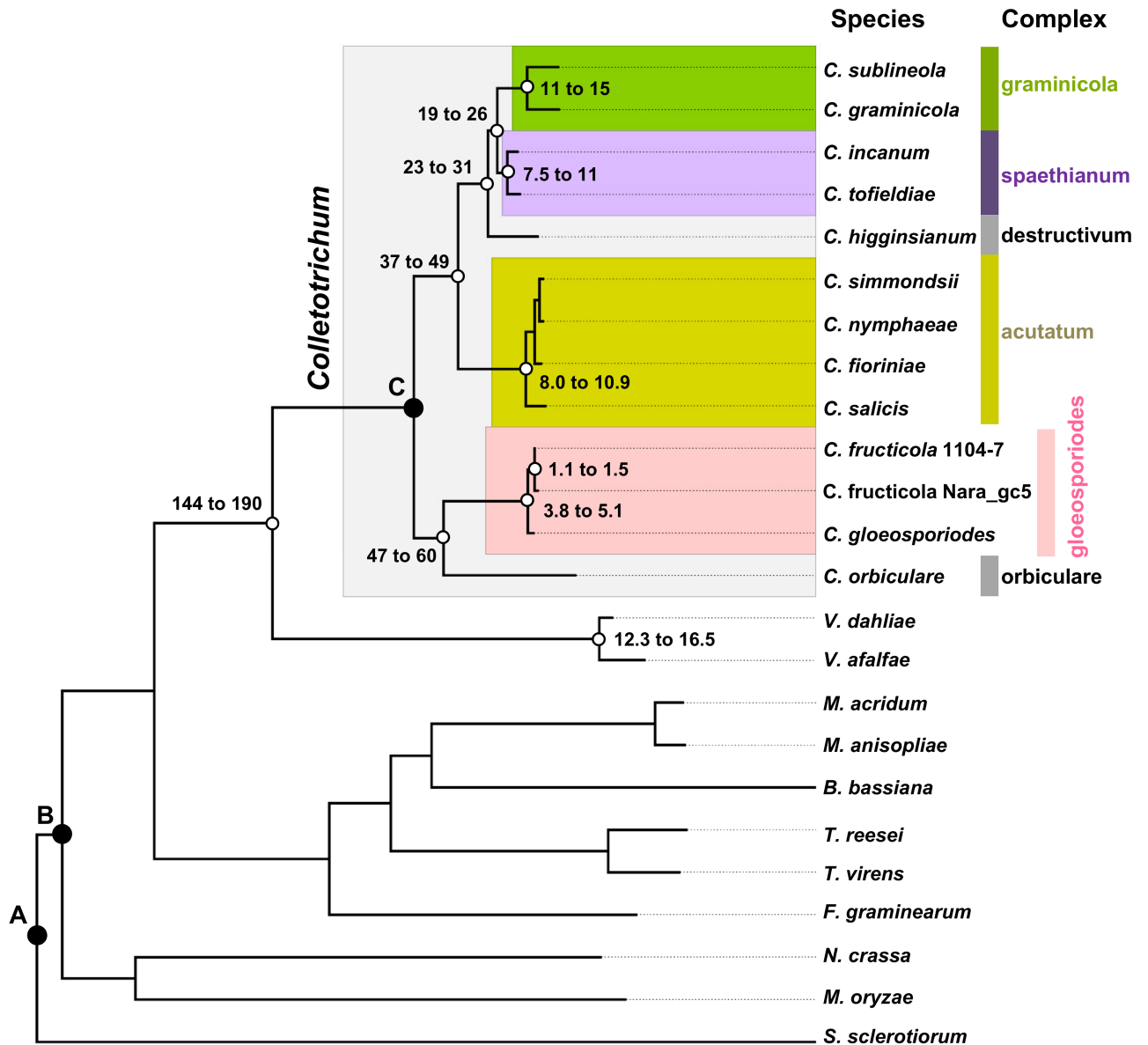
CAZs, secreted proteases, secondary metabolite synthetases, cytochrome P450s, transporters, and small secreted proteins (SSPs) are known virulence factors in fungi. Putative genes belonging to these functional categories were identified from the compared genomes via a custom prediction pipeline. In general, *Colletotrichum* genomes contained more virulence genes compared with non-*Colletotrichum* genomes (S2–S7 Figs), with the enrichments of CAZs, cytochrome P450s, transporters, and SSPs being marked. From a total of 4,596 families (defined either based on PFAM domain or annotated functional category), CAFE based analysis of gene gain and loss patterns identified 454 families evolving in a non-random birth and death manner at a 0.01 family-wise significance level. For these families, the expected expansions/contractions and the corresponding Viterbi *p*-values were calculated for individual branches. Five branches closely related to *Colletotrichum* evolution were examined in greater detail (Fig 3A). These branches contained the most recent common ancestor (MRCA) of Glomerellales (node 1), the *Colletotrichum* MRCA (node 2), the graminicola complex MRCA (node 3), the *acutatum* complex MRCA (node 4), and the gloeosporioides complex MRCA (node 5). At a family-wide significance threshold of 0.05, 208 non-redundant families showed significant expansions/contractions (Table D in S2 File). The overall gene gain and loss



**Fig 1. The two putatively fungus-to-fungus horizontally-transferred gene clusters present in the lineage-specific regions of the 1104-7 genome.** Syntenic DNA blocks (identified based on Blast search) are in dark grey boxes, genes are in arrowheads, orthologous genes are in the same color, alignment length and nucleotide percentage identity (in bracket) are also shown. Maximum likelihood based phylogenetic trees of the HGT genes are shown in *S1 Fig*.

<https://doi.org/10.1371/journal.pone.0196303.g001>



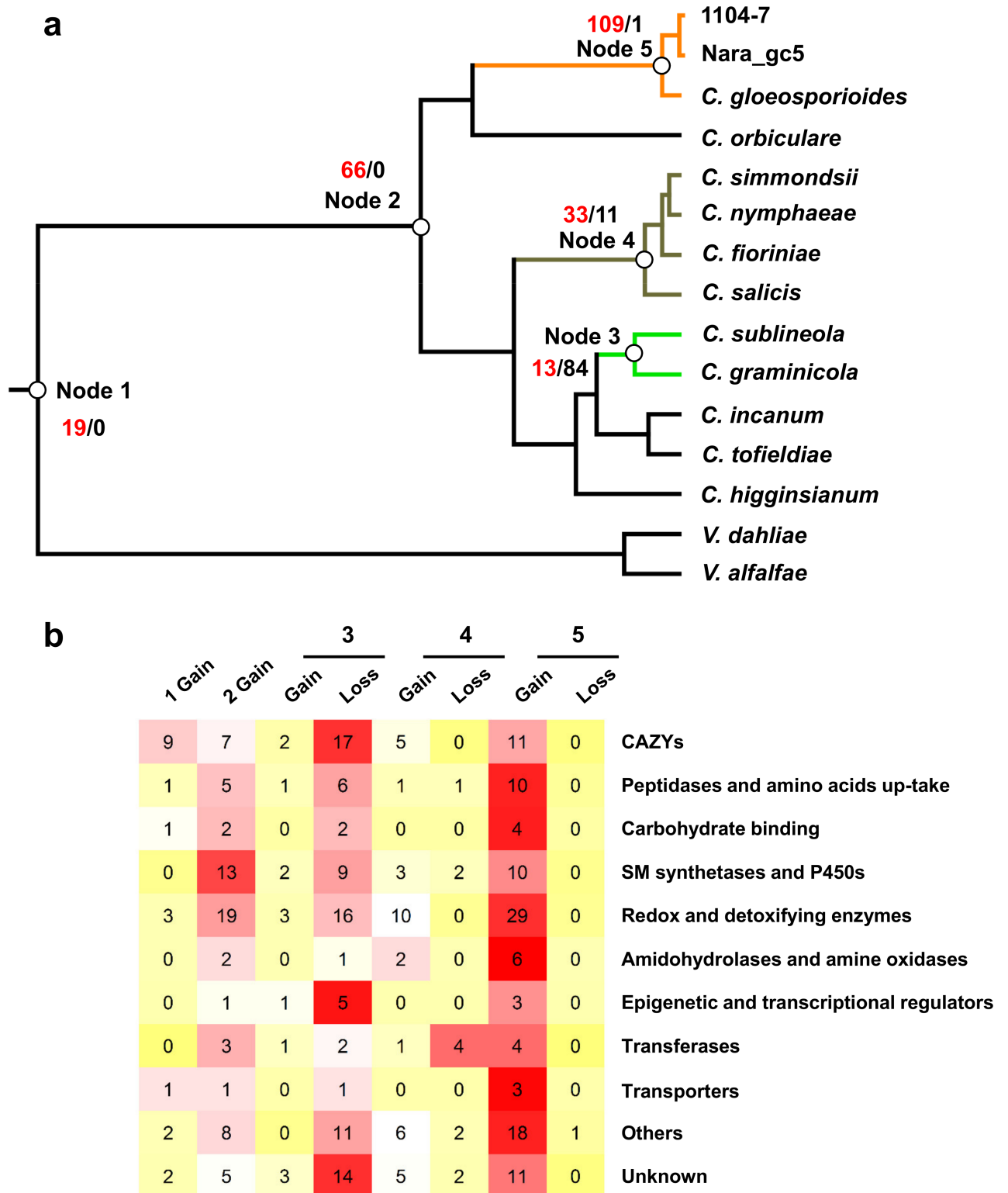


**Fig 2. Maximum-likelihood phylogenetic tree constructed from 1,212 single-copy core genes and divergence time estimation using r8s analysis.** A, B and C are calibration points, divergence times are shown in million years, the ranges were calculated based on estimations with different combinations of smoothing factors and calibration schemes (see Table C in S2 File for detail).

<https://doi.org/10.1371/journal.pone.0196303.g002>

patterns associated with these five nodes are shown in Fig 3. Consistent with previous reports [15, 16], GH43, AA7, and NLPs were strikingly expanded at the acutatum complex MRCA.

In general, the graminicola complex MRCA (node 3) was dominated by gene loss whereas the Glomerellales MRCA (node 1), the *Colletotrichum* MRCA (node 2), and the gloeosporioides complex MRCA (node 5) were dominated by gene gains (Fig 3A). A large number of gene families being expanded at the Glomerellales MRCA (node 1) were CAZyS, or more specifically ones related to pectin degradation. Other nodes were characterized by different expansion/contraction patterns with families experiencing significant size changes related to secondary metabolism, P450s, oxidoreductases, and detoxifications among others (Fig 3B).



**Fig 3. Gene gain and loss patterns at major five nodes of the *Colletotrichum* phylogeny.** (a) Number of families significantly expanded (red) or contracted (CAFE analysis, family  $P < 0.01$ , Viterbi  $P < 0.05$ ). (b) Functional categories of the families significantly expanded or contracted at indicated nodes.

<https://doi.org/10.1371/journal.pone.0196303.g003>

## Gene family evolution prior to *Colletotrichum* and *Verticillium* divergence

At the Glomerellales MRCA (node 1), 19 families were significantly expanded (Viterbi  $P < 0.05$ ). Interesting, many of these families were functionally related to degrading pectins (PL1, PL3, GH28, GH78, GH88, GH43, CBM67), celluloses or hemicelluloses (GH43, AA3, and AA9). Thus, the Glomerellales MRCA evolution involves a strong expansion of plant cell wall degrading enzymes (PCWDEs).

*Colletotrichum* genomes are known to be enriched with PCWDEs [15, 16], we further examined major PCWDE-related CAZY families to gain a global insight into their evolutions (Fig 4). Gene family expansions were obvious with both the Glomerellales MRCA (node 1) and the *Colletotrichum* MRCA (node 2), each containing seven significantly expanded families, suggesting that the elevated PCWDE content in *Colletotrichum* was due to stepwise expansions. Within the *Colletotrichum* genus, the gloeosporioides complex showed obvious CAZY gains whereas the graminicola complex showed obvious CAZY losses, which were consistent with previous reports [15, 16].

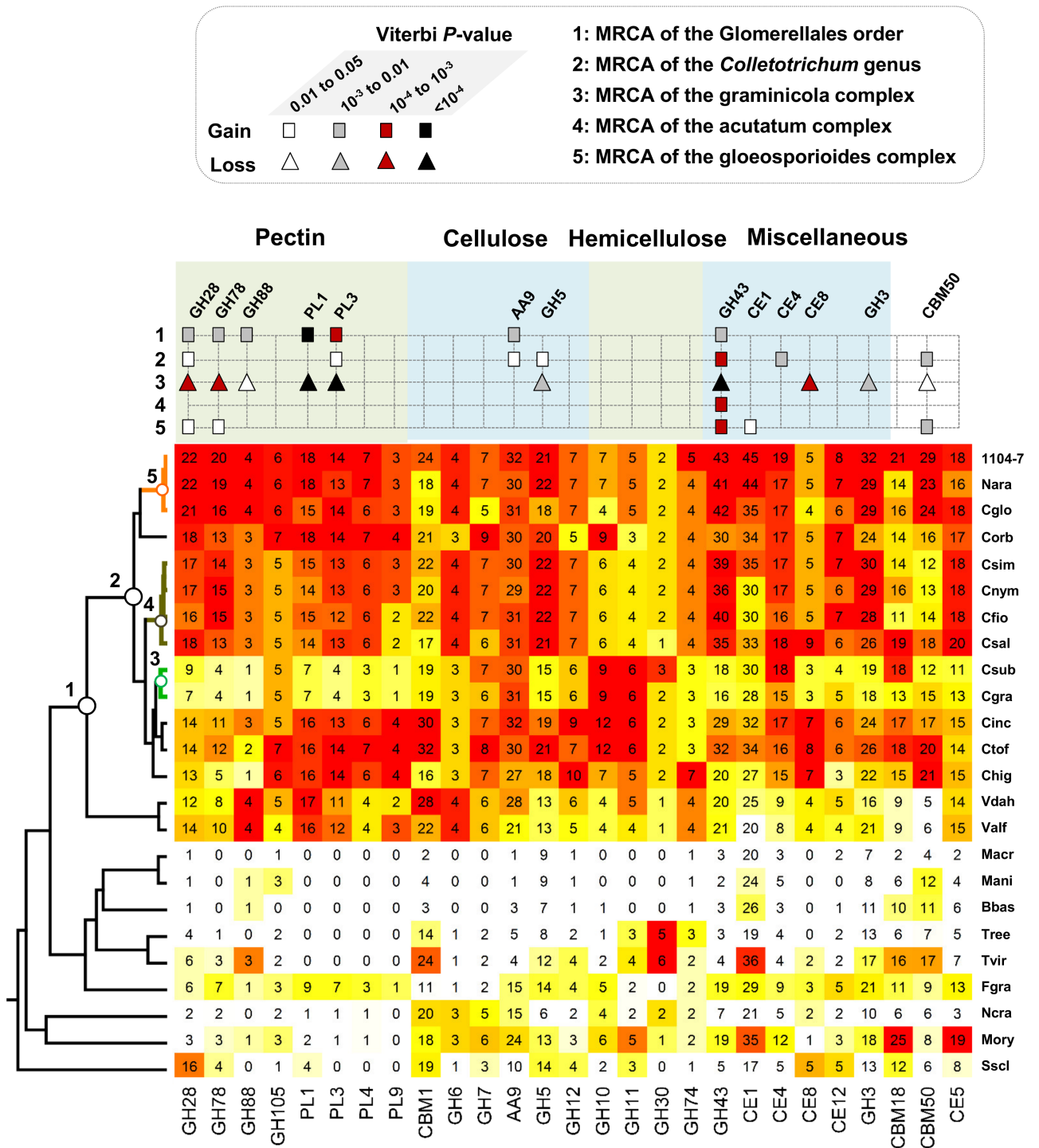
## Gene family evolution at the *Colletotrichum* MRCA

At the *Colletotrichum* MRCA (node 2), 66 families were significantly expanded (Viterbi  $P < 0.05$ , Fig 5, Table D in S2 File). The most strongly-expanded family (Viterbi  $P = 1E-06$ ) contained a PF11807 domain. While most PF11807 proteins are functionally unknown, the *Ustilaginoidea virens* ustYa and ustYb participate in the biosynthesis of the ribosomal peptide-derived toxin UstiloxinB [50], and the *Talaromyces islandicus* CctP functions in synthesizing the NRPS mycotoxin cyclochlorotine [51]. The second and fourth most strongly-expanded families were CYP68 and CYP65, two groups of cytochrome P450s being also related to secondary metabolite biosynthesis. CYP62, CYP5080, CYP552, as well as PKSs, and DMATs were also strongly expanded (Viterbi  $P < 0.01$ ). Moreover, the expansion extent of berberine bridge enzymes (BBEs, PF08031), a family of flavin-dependent oxidoreductases critical for isoquinoline alkaloid biosynthesis [52, 53], ranked 11<sup>th</sup> in expansion significance among all families. These results together supported a strong diversification in secondary metabolite production at the *Colletotrichum* MRCA.

Redox enzymes may contribute toward fungal pathogenesis in multiple ways, such as oxidative breakdown of cellulose and hemicellulose, synthesizing toxins, and counteracting plant-derived phenolic compounds. Tyrosinase (PF00264), type II peroxidase (PF01328), and GMC oxidoreductase (PF00732) were all strongly expanded at the *Colletotrichum* MRCA (Viterbi  $P < 0.01$ ).

Protein families being strongly expanded at the *Colletotrichum* MRCA also included ones functioning in peptide degradation (e.g. x-pro dipeptidyl-peptidase, subtilase), nutrient uptake (e.g. OPT oligopeptide transporter, cytosine/purine permease), transcriptional regulation (e.g. NmrA-like protein), and chitin binding (e.g. CBM50) among others. Worthy to note, PF00135 (carboxylesterases) and PF07519 (tannase and feruloyl esterase activities), two detoxification-related families, were also strongly expanded (Viterbi  $P = 1E-04$  and  $2E-04$  respectively). Carboxylesterase detoxifies xenobiotics (toxins or drugs) in animals [54]. Tannase degrades tannins, a group of plant defense related phenolic compounds [55] whereas feruloyl esterases facilitate xylan and pectin degradation [56].

OrthoMCL clustering identified three protein families showing *Colletotrichum* lineage-specific loss (present in all 11 compared non-*Colletotrichum* genomes, but none of the 13 *Colletotrichum* genomes). All three families were made up of single-copy orthologs, including one putative  $Ca^{2+}$ /calmodulin-dependent protein kinase (CAMK, corresponding to XP\_003717191 in *Magnaporthe oryzae*), one CofD\_Yvck family protein (XP\_003717966 in *M.*



**Fig 4. Content variation of CAZY families with plant cell wall degrading activity or known to be important for plant pathogen interactions.** Species abbreviations: Nara, *Colletotrichum fructicola* Nara\_gc5; Cglo, *C. gloeosporioides*; Corb, *C. orbiculare*; Csim, *C. simmondsii*; Cnym, *C. nymphaeae*; Cfio, *C. fiorinae*; Csal, *C. salicis*; Csub, *C. sublineola*; Cinc, *C. incanum*; Ctof, *C. tofieldiae*; Chig, *C. higginsianum*; Vdah, *Verticillium dahliae*; Valf, *V. alfalfae*; Macr, *Metarhizium acridum*; Mani, *M. anisopliae*; Bbas, *Beauveria bassiana*; Tree, *Trichoderma reesei*; Fgra, *Fusarium graminearum*; Ncra, *Neurospora crassa*; Mory, *Magnaporthe oryzae*; Sscl, *Sclerotinia sclerotiorum*.

<https://doi.org/10.1371/journal.pone.0196303.g004>

*oryzae*) and one lacking any function-indicative signature (XP\_003715556 in *M. oryzae*). The CAMK gene lacks distinct ortholog in *S. cerevisiae* and no obvious phenotype was observed

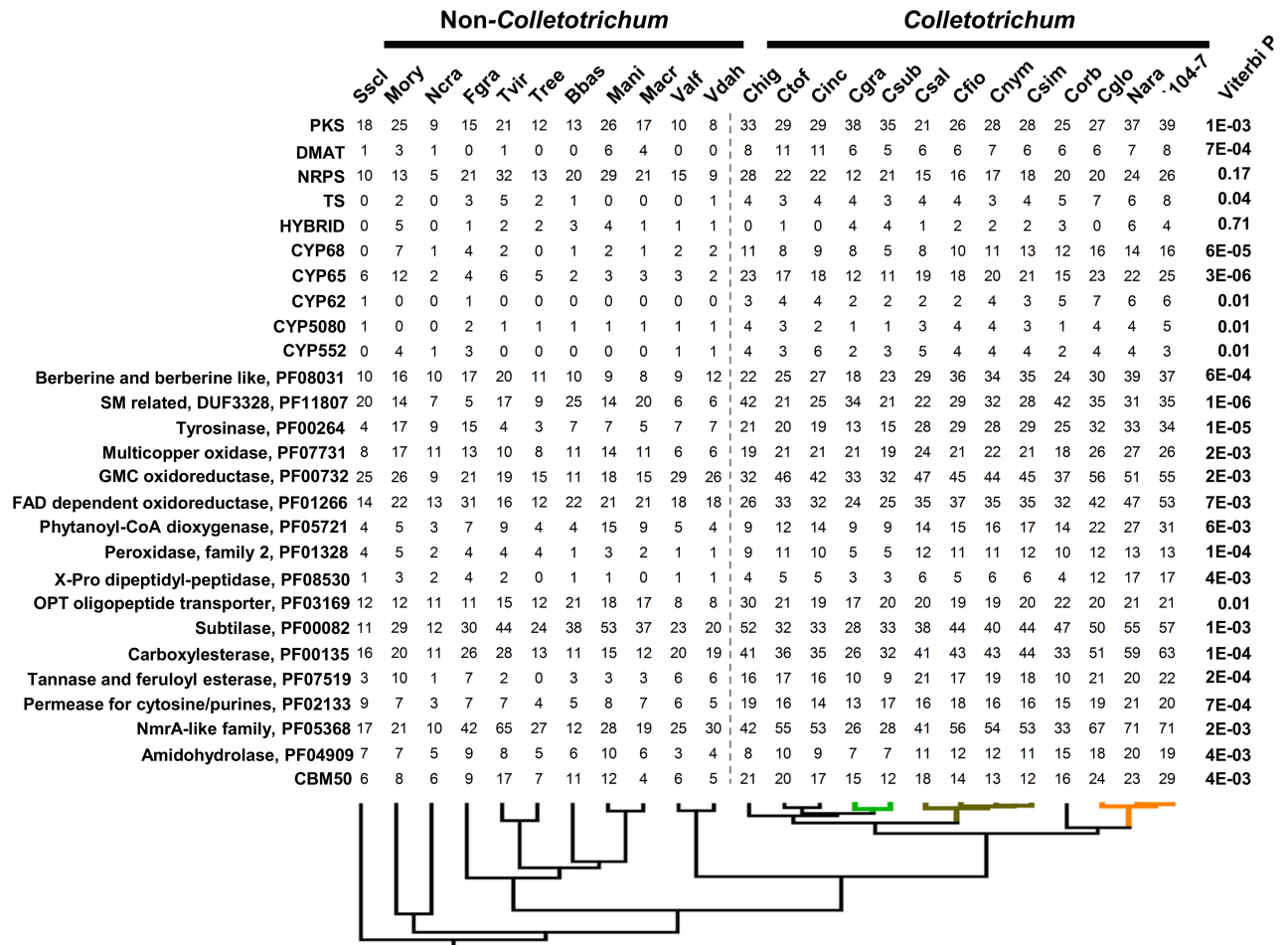


Fig 5. Copy number differences of selected gene families (defined based on PFAM or functional predictions) between non-*Colletotrichum* and *Colletotrichum* species. For each family, the Viterbi P value calculated with CAFE is shown on right.

<https://doi.org/10.1371/journal.pone.0196303.g005>

with the gene deletion mutant in *Fusarium graminearum* (FGSG\_05549) [57]. CofD\_Yvck family protein is related to carbon metabolism, but no fungal gene has been characterized.

### Genes families being specifically conserved among *Colletotrichum* genomes and *Colletotrichum* genus-specific SSPs

Based on OrthoMCL clustering, 260 families were identified to be *Colletotrichum*-specific among compared genomes and contained proteins from all 13 *Colletotrichum* genomes (Table E in S2 File). These genus core families contained members known or putatively important for plant infection, especially for appressorium functions, such as CAP22 [58], CAS1-like proteins [59], CFEMs [60], putative cutinase and ligninase. Four families were made of *Colletotrichum* genus-specific SSPs (defined by NCBI nr BLASTP, e-value cut-off 1E-05), which included the previously identified *C. higginsianum* effector candidates EC2 and EC65 [61], and one CFEM domain protein.

Based on queries of a local installation of the NCBI fungal database, we identified 939 *Colletotrichum* genus-specific SSPs. These proteins contained a predicted secretion signal, were less than 300 aa, and lacked a BlastP hit (E-value cutoff = 1E-05) in other fungal species. 29 genus-specific SSPs contained recognizable PFAM domains (eight domains in total, Table 3).

**Table 3. PFAM domains contained by *Colletotrichum* small secreted proteins which lack significant BlastP hit (e-value cut-off 1E-05, query coverage > 50%) outside the genus.**

| PFAM ID | Annotation   | Representative proteins <sup>1</sup>   |
|---------|--|--|
| PF12296 | HsbA, hydrophobic surface binding protein A        | EQB52112.1 ( <i>C. gloeosporioides</i> , 1E-11)<br>ENH89122.1 ( <i>C. orbiculare</i> , 7E-11)  |
| PF09792 | Ubiquitin 3 binding protein But2 C-terminal domain | ENH78092.1 ( <i>C. orbiculare</i> , 1.6E-05)   |
| PF08881 | CVNH domain  | KDN63891.1 ( <i>C. sublineola</i> , 2.5E-05)<br>KDN62312.1 ( <i>C. sublineola</i> , 8.7E-11)<br>KZL65396.1 ( <i>C. tofieldiae</i> , 3E-09)     |
| PF14856 | Hce2, putative necrosis-inducing factor            | XP_007602516.1 ( <i>C. fioriniae</i> , 3E-11)<br>KZL66113.1 ( <i>C. tofieldiae</i> , 2E-12)<br>EQB50157.1 ( <i>C. gloeosporioides</i> , 5E-13) |
| PF05730 | CFEM   | XP_007285601.1 ( <i>C. fructicola</i> , 1.4E-09)<br>ENH76065.1 ( <i>C. orbiculare</i> , 1.2E-09)   |
| PF01822 | WSC, a putative carbohydrate binding domain        | XP_007598049.1 ( <i>C. fioriniae</i> , 1E-06)<br>KXH31806.1 ( <i>C. simmondsii</i> , 5.6E-08)<br>KXH62552.1 ( <i>C. nymphaeae</i> , 4.2E-08)   |
| PF06766 | Fungal hydrophobin                                 | KZL63596.1 ( <i>C. incanum</i> , 8.3E-05)<br>ENH81598.1 ( <i>C. orbiculare</i> , 6.6E-09)  |
| PF00024 | PAN domain   | XP_007279807.1 ( <i>C. fructicola</i> , 2.5E-05)   |

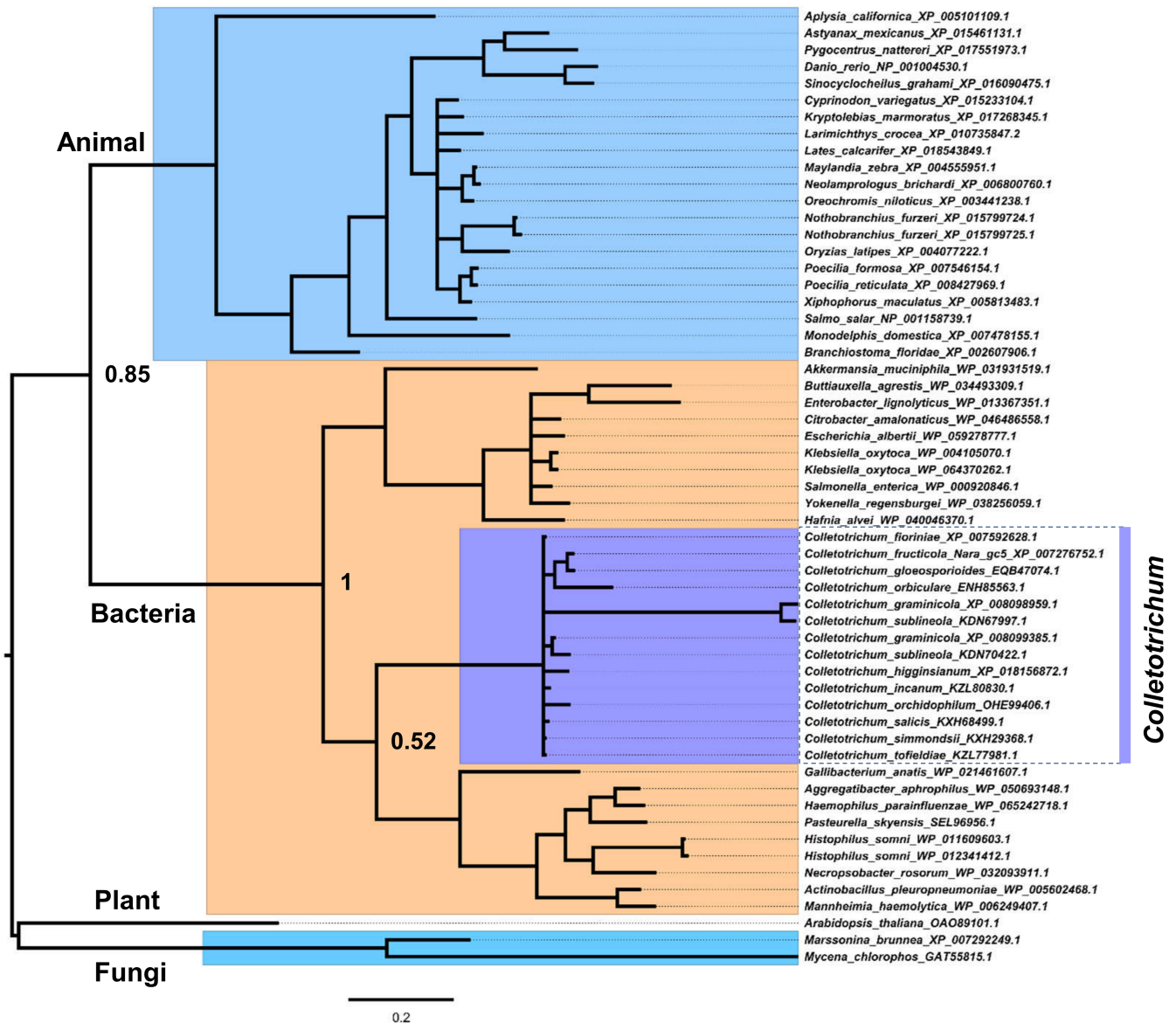
<sup>1</sup>Representative proteins, each GenBank accession is followed by a parenthesis showing the species name and PFAM domain hit E-value.

<https://doi.org/10.1371/journal.pone.0196303.t003>

PF14856 (Hce2) corresponds to the *Cladopsorium fulvum* Ecp2 effector which contains a necrosis-inducing activity [62]. PF05730 (CFEM) is functionally associated with fungal pathogenesis. PF08881 (CVNH), PF01822 (WSC), and PF00024 (PAN domain) are related to protein-oligosaccharide interactions. PF12296 (HsbA) and PF06766 (Hydrophobin2) are related to hydrophobic surface binding.

### Horizontal transfer of a RbsD/FucU fucose transporter from bacterium to the *Colletotrichum* ancestor

InterProScan search (cutoff E-value, 1e-04) and manual inspection identified tens of PFAM domains specific to the *Colletotrichum* genus among compared genomes and being present in more than one species. The co-occurrence of these domains in different genomes made it unlikely that their presence was due to DNA contaminations. BlastP searches showed that most protein homologous of these proteins distributed sporadically among fungi, or were specific to the *Colletotrichum* genus, making it hard to predict their evolutionary histories. However, one family, RbsD/FucU fucose transporter (PF05025), showed strong signatures of bacteria-to-fungi transfer. This RbsD/FucU fucose transporter (PF05025) family is conservatively present among all compared *Colletotrichum* genomes. In the NCBI nr database, the *Colletotrichum* proteins had homologs in diverse bacteria and animal species, but had no homolog across the fungal kingdom (BlastP, cutoff  $P = 1E-05$ ). Phylogenetically, the *Colletotrichum* proteins formed a monophyletic clade nested within bacterial lineages with strong statistical support (Fig 6). Such combined patterns of taxonomic distribution and phylogenetic topology supported bacteria-derived gene gain by the *Colletotrichum* ancestor. The genus-wide conservation of this family indicates its importance for lineage-specific adaptations. L-fucose is a major constituent of N-linked glycans, which distribute widely on the cell surfaces of microbes, plants and animals, L-fucose is also abundant in soil and can be used as the sole carbon source by several groups of microorganisms [63]. The acquisition of the RbsD/FucU fucose transporter may benefit *Colletotrichum* species in natural nutrient competition.



**Fig 6. Putative bacteria-to-*Colletotrichum* horizontal transfer of the RbsD/FucU fucose transporter family (PF05025).** Bayesian phylogenetic tree was constructed with the best fungal, bacterial, animal and plant BLASTP hits of the *Colletotrichum* RbsD/FucU fucose transporters in the NCBI nr database. The tree was constructed with MrBayes, WAG+G substitution model,  $5 \times 10^6$  mcmc generations, sample frequency = 1000, first 25% discarded as burn-in, numbers indicate posterior probabilities.

<https://doi.org/10.1371/journal.pone.0196303.g006>

### Gene family evolution related to species complex diversification

Species in the graminicola species complex contain a strongly reduced set of pectin-degrading enzymes associated with monocot host adaptation [3, 16]. In this study, we showed that a number of gene families functioning beyond pectin degradation were also reduced (Fig 3, Fig 7, Table D in S2 File). Among these families, Fn3-like protein (PF06280), NmrA-like protein (PF05368), and RTA1 (PF04479) showed the strongest reductions. The Fn3 domain is frequently found in streptococcal C5a peptidases (SCP) and adhesin/invasion proteins [64]. NmrA-like proteins are related to transcriptional regulation. The RTA1 protein family (PF04479) contains export proteins transporting antimicrobial compounds such as sphingoid

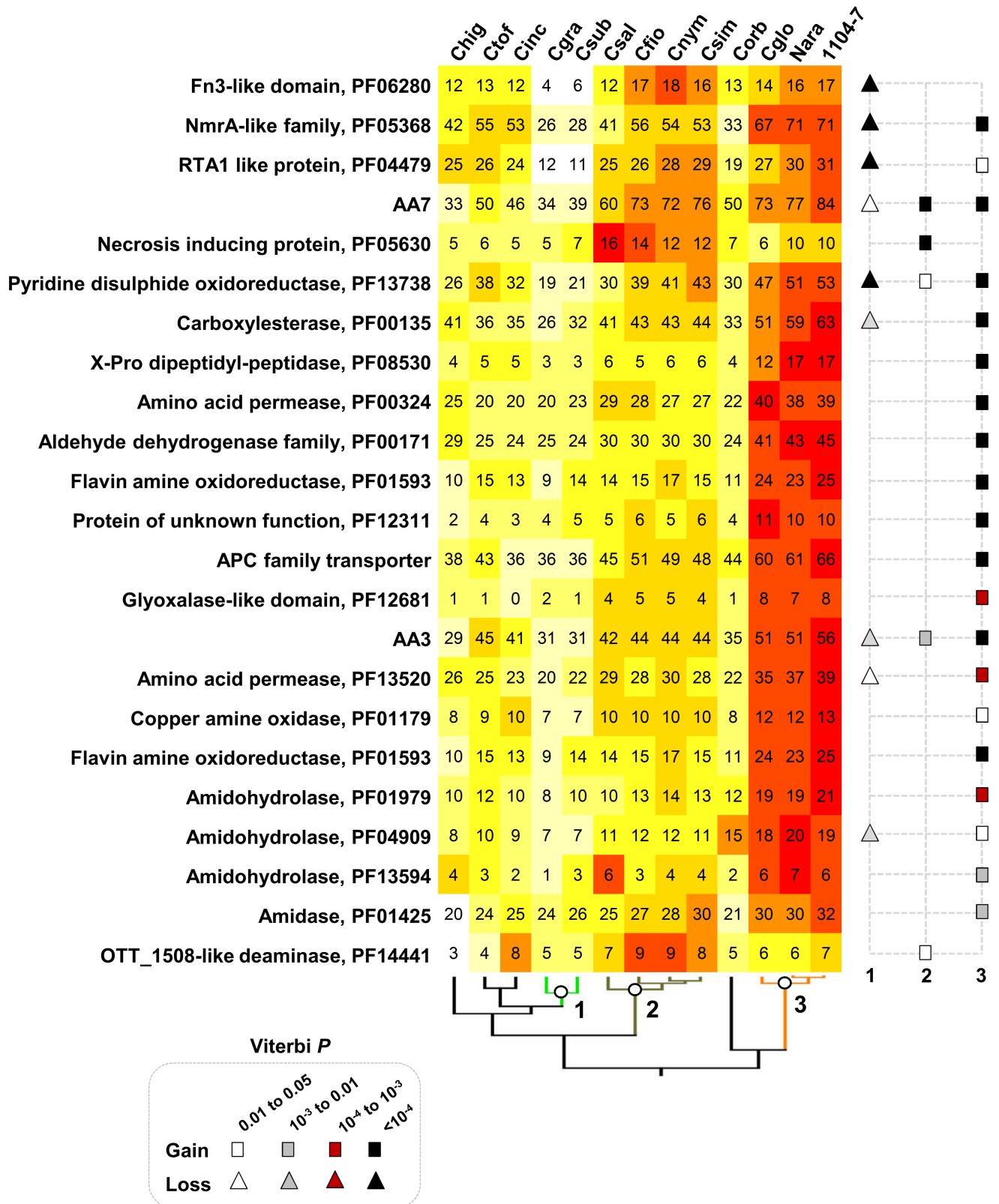


Fig 7. Evolution of gene families related to species complex diversification.

<https://doi.org/10.1371/journal.pone.0196303.g007>



bases and 7-aminocholesterol. Overexpression of RTA1 proteins confer drug or toxin resistance in yeast [65].

At the gloeosporioides species complex MRCA, the main functional categories of expanded gene families included redox and detoxifying enzymes, CAZyS, peptidases and amino acid transporters (Fig 3). The strong expansions observed with one putative peptidase family (PF08530) and two putative amino acid transporter families (PF00324, PF13520) indicated an improved capacity of the gloeosporioides complex to utilize protein-derived nutrients. Moreover, among the five non-redundant amidohydrolase families and two amine oxidase families catalyzing ammonia production, four amidohydrolase families (PF01979, PF13594, PF01425, PF04909) and two oxidase families (PF01593, PF01179) were significantly expanded (Viterbi  $P < 0.05$ ), indicating an improved capacity to produce ammonia.

## Discussion

*Colletotrichum* species are genetically diverse and cause diseases on a wide range of plant species. Although differing considerably in host specificity and symptom appearance, most pathogens infect as hemibiotrophs, subverting host defense reactions first, and initiating host killing and host cell wall degradations thereafter. These phenomena support a universal infection strategy and perhaps underlying molecular mechanisms [6,7]. On the other hand, the considerable variation of plant-interaction style (host and tissue specificity, symptom appearance) implies the importance of lineage-specific adaptations [6,7]. Combined efforts in genomic and transcriptomic research have provided key insights into *Colletotrichum* fungi evolution. For instance, compared with other fungi, *Colletotrichum* genomes are markedly rich with pathogenicity-related genes including PCWDEs, proteases, SM biosynthetic enzymes, secreted effectors [3,7]. During pathogenesis, these genes express dynamically to fulfill stage-specific pathogenic functions [3, 13]. Moreover, the gain and loss of PCWDE protein families have been indicated to be important in shaping their host specificities [15,16].

In this study, we systematically compared the gene content variation across 13 *Colletotrichum* and 11 non-*Colletotrichum* genomes. Pathogenicity-related genes were annotated, classified, and compared; in addition, marked expansion/contraction events at key phylogenetic nodes were identified based on CAFE analysis. These results provided a global evolutionary picture of *Colletotrichum* gene families (summarized in Fig 8).

## Evolutionary dynamics of virulence-related gene families at the *Colletotrichum* MRCA

A range of gene families showed very strong expansions at the *Colletotrichum* MRCA. These include berberine bridge enzyme and PF11807 related to SM biosynthesis; type II peroxidase, tyrosinase and multicopper oxidase families related to oxidoreduction; carboxylesterase and tannase related to detoxification; OPT oligopeptide transporter and cytosine/purine permeases related to transport. Moreover, OrthoMCL analysis identified a range of core *Colletotrichum* genus-specific protein families with putative virulence roles including necrosis-induction (Hce2), signaling (CFEM), protein-oligosaccharide interactions (CVNH, WSC, PAN), and appressorium development (CAP22, CAS1), these genes are specific to *Colletotrichum* and conservatively present in all compared *Colletotrichum* genomes, and may thus be important for *Colletotrichum* infection. We have also identified three lineage-specific losses and one bacterial-derived horizontal transfer event at the *Colletotrichum* MRCA, demonstrating that lineage-specific gene loss and horizontal transfer have also contributed to *Colletotrichum* evolution.

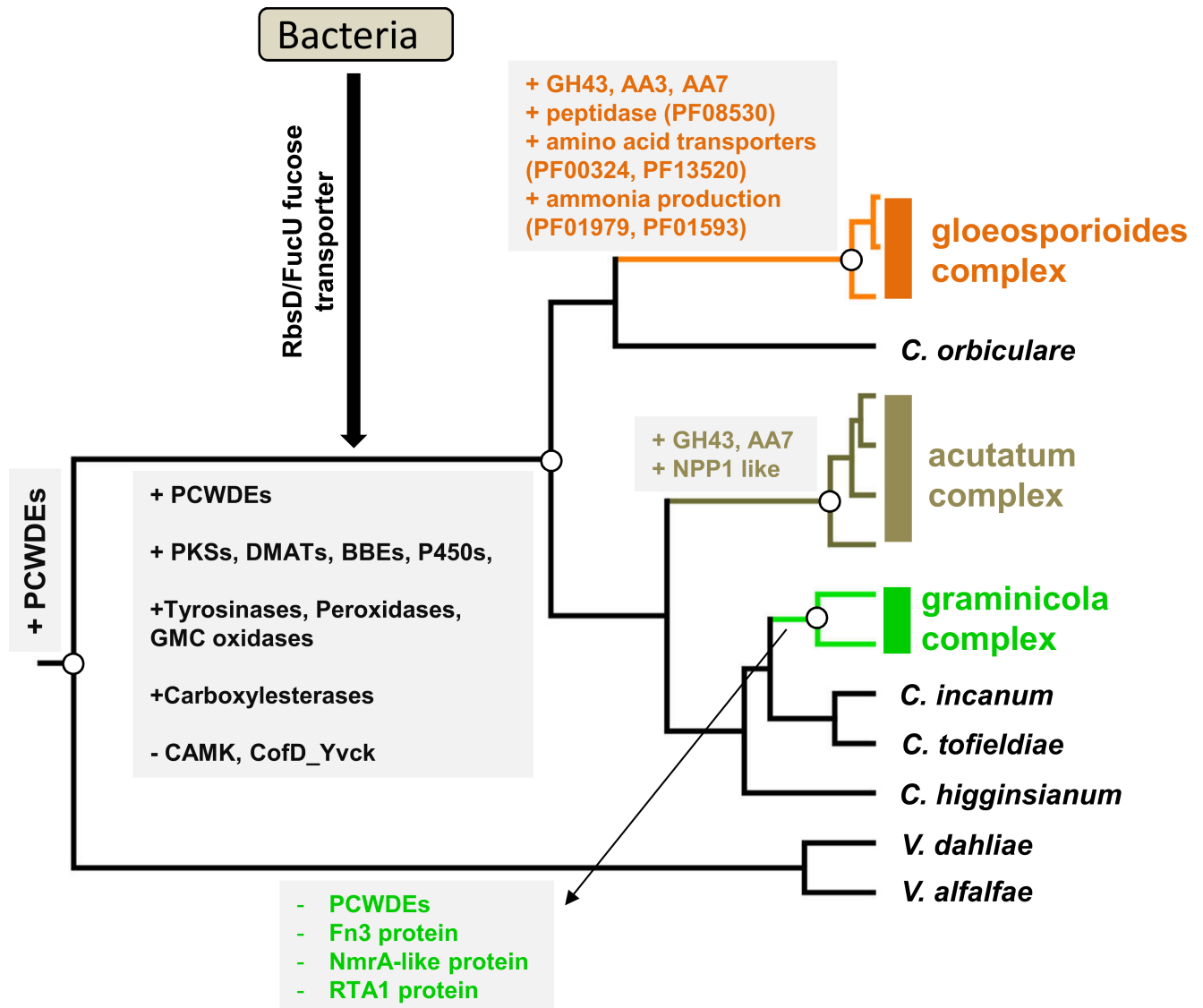


Fig 8. A summary representation of the important evolutionary events of *Colletotrichum* gene families.

<https://doi.org/10.1371/journal.pone.0196303.g008>

*Colletotrichum* and *Verticillium* are related phytopathogens in the Glomerellales order, the former belongs to Glomerellaceae whereas the later belongs to Plectosphaerellaceae. Differing from *Colletotrichum* pathogens which mainly colonize leaves and fruits, *Verticillium* pathogens mainly colonize the plant root and vascular system. On the phylogenetic tree, the enrichment of pectinases was observed with both *Colletotrichum* and *Verticillium*, whereas many SM genes (e.g., synthetases, P450s, transporters), redox and detoxification-related enzymes are specifically enriched with *Colletotrichum*. Thus, these two categories of virulence factors appear to have different evolutionary histories although all being strongly expanded in *Colletotrichum*. The co-enrichment of pectin-degrading enzymes in *Colletotrichum* and *Verticillium* could be due to either single duplication prior to divergence or recent duplications related to independent adaptations. Plectosphaerellaceae family contains pathogenic genera such as *Plectosphaerella* and *Gibellulopsis* in addition to *Verticillium* [66], analyzing these genomes will be critical to understand PCWDE evolution in the Glomerellales.

## Evolutionary dynamics of virulence-related gene families among different *Colletotrichum* lineages

Among the 13 compared isolates, *C. sublineola* and *C. graminicola* specialize on monocot plants whereas other isolates specialize on dicot plants. In addition, species belonging to the acutatum complex and the gloeosporioides complex are more commonly observed to colonize fruits. Previous studies have reported a reduced set of pectin-degrading enzymes in *C. graminicola* and an elevated set of plant cell wall degrading enzymes in the acutatum and gloeosporioides complexes [3, 15, 16]. Our systemic CAFE analysis of gene family size evolution confirmed these results. More importantly, we identified a range of additional gene families showing gain or loss patterns relevant to such lineage-specific pathogenic adaptations.

*Colletotrichum* species are 'alkaline' fungi, accumulating high-level of ammonia both in culture and during plant infection, which is reportedly important for fungal infections [3, 67]. Two protein families with putative amidohydrolase activities (PF01979, PF04909) were significantly expanded at the *Colletotrichum* MRCA. Moreover, these two families together with four additional families related to ammonia production were further expanded at the gloeosporioides complex MRCA, suggesting a stepwise improvement in ammonia-producing potential. In the gloeosporioides complex, a deamination-related glutamate dehydrogenase plays significant roles in ammonia production, and the enzymatic activity requires amino acids as substrate [67]. In this study, the flavin containing amine oxidoreductase family (PF01593), which catalyzes ammonia production by oxidizing monoamines and polyamines [68], showed strong expansion in the gloeosporioides complex. *Colletotrichum* species belonging to the gloeosporioides complex are well-known fruit-infecting pathogens, their host fruit tissues are generally acidic in pH and these pathogens can modulate host local pH to promote infection [67, 69]. The expansion of flavin containing amine oxidoreductase might thus represent a virulence-relevant adaptation strategy in terms of pH regulation.

Another important protein family related to lineage-specific pathogenic adaptation is RTA1, which showed strong size reduction in the monocot-specializing graminicola complex. The family size was on-average one half that of other *Colletotrichum* species. As limited information is known regarding the biological functions of RTA1 proteins in filamentous fungi, it is difficult to interpret the significance of its reduction. Yet, in yeast, RTA1 overexpression confers drug or toxin tolerance [65], indicating a potential function of detoxifying monocot-relevant defense compounds.

## The evolution of lineage-specific genes in *C. fructicola*

*C. fructicola* has a broad host range, however pathogenicity test indicates that this species might encompass individual host-limited forms [19]. In this study, we compared the genomes of 1104-7 and Nara\_gc5, two *C. fructicola* isolates derived from different hosts. The two genome assemblies were similar in size (57.1 Mb vs 55.6 Mb), shared 98.7% nucleotide identity in the alignment regions, up to 52.9 Mb of the 1104-7 genome were in > 10 kb alignment blocks when comparing with Nara\_gc5. Thus, from a whole genome perspective, 1104-7 and Nara\_gc5 were highly similar. By applying the same gene prediction pipeline to the 1104-7 and Nara\_gc5 assemblies, their gene content variations could be compared in a non-biased manner. Interestingly, although similar total gene models were predicted (17,827 vs 17,844), OrthoMCL clustering identified approximately 1,000 isolate-specific genes in each genome, many of which may represent true genes based on the finding that over 60% of these genes had significant NCBI BlastP hits and that approximately 65% of the genes in 1104-7 had RNA-seq support.

Many fungal plant pathogen genomes can be classified into conserved core regions and plastic variable regions [70–72]. A plastic and fast-evolving subgenome is beneficial for deriving new host adaptations by elevating intraspecific diversification [70–72]. Although the biological traits of 1104–7 and Nara\_gc5 have not been compared side by side, it is likely that the observed gene content variations are related to local adaptations. A plausible explanation for the high-degree of genome nucleotide identity and the existence of large numbers of isolate-specific genes would be that the *C. fructicola* genome encompasses subregions evolving at different speeds. To further dissect the intraspecific genomic variation among the two *C. fructicola* isolates, we identified and examined the evolutionary characteristics of genes located in lineage-specific (LS) regions in both genomes. With a length criterion of 10 kb, 0.62 Mb LS regions were identified in 1104–7 whereas 0.33 Mb LS regions were identified in Nara\_gc5. Genes located within the LS regions are highly dynamic from an evolutionary perspective. Based on Blast queries, an elevated proportion of genes have no hit or are more closely related to non-*Colletotrichum* sequences than to *Colletotrichum* sequences. Moreover, two gene clusters showing strong signatures of fungus-to-fungus horizontal transfer were identified from the 1104–7 LS genomic regions. The putative functions of genes on the two clusters include serine protease, hemolysin-III protein known to function in membrane toxicity [73], as well as enzymes catalyzing secondary metabolite biosynthesis, all of which are virulence-related. While the host specificities of the two *C. fructicola* isolates have not been directly compared, the presence of virulence-related genes at the plastic subgenomic regions do support lineage-specific adaptations. In the *C. gloeosporioides* species complex, a strain-wide presence-absence polymorphism pattern of conditionally dispensable chromosomes (CDCs) has been observed [74], CDCs can transfer among strains even though direct evidence supporting their roles in pathogenicity transfer is lacking [75, 76]. In the future, determining whether the *C. fructicola* LS genomic DNAs identified in this study represent CDC and are virulence related will be of significant interest.

## Supporting information

**S1 Fig. Maximum likelihood (ML) based phylogenies of genes in the 1104–7 HGT1 and HGT2 clusters.** Maximum likelihood (ML) based phylogenies of genes in the 1104–7 HGT1 and HGT2 clusters. For each gene (red color), best non-*Colletotrichum* BlastP hits (black nodes) and best *Colletotrichum* hits (green nodes) were retrieved from NCBI nr database, aligned for ML tree construction in RAxML 8.1.1. The best amino acid substitution models (shown for each tree) were identified with ProtTest3. Bootstrap values (based on 1,000 replicates) are indicated for major nodes.

(PDF)

**S2 Fig. Carbohydrate-active enzyme (CAZY) content variation among compared genomes.** Carbohydrate-active enzyme (CAZY) content variation among compared genomes. GH, glycoside hydrolase; GT, glycoside transferase; PL, polysaccharide lyases; CE, carbohydrate esterase; CBM, carbohydrate-binding modules; AA, auxiliary activities.

(PDF)

**S3 Fig. Variation of secreted proteases among compared genomes.** A, aspartic type; M, metallo type; S, serine type.

(PDF)

**S4 Fig. Variation of secondary metabolite synthetases among compared genomes.** DMAT, dimethylallyl tryptophan transferase; NRPS, nonribosomal peptide synthase; PKS, polyketide

synthase; TS, terpene synthase; HYBRID, NRPS-PKS hybrid.  
(PDF)

**S5 Fig. Variation of cytochrome P450s among compared genomes.**  
(PDF)

**S6 Fig. Variation of transporter genes among compared genomes.**  
(PDF)

**S7 Fig. Variation of small secreted protein (SSP) content among compared genomes.** SSPs are defined as proteins containing predicted secretion signals and being less than 300 aa. CSSPs, cysteine-rich SSPs (cysteine% > 3%); NCSSPs, non cysteine-rich SSPs (cysteine% ≤ 3%).  
(PDF)

**S1 File. The gene annotations and predicted protein sequences of the *C. fructicola* 1104–7 genome.**  
(RAR)

**S2 File.** Table A to E.  
(XLSX)

## Acknowledgments

We would like to thank the anonymous reviewers for their kind and helpful comments on the original manuscript.

## Author Contributions

**Conceptualization:** Xiaofei Liang, Jeffrey A. Rollins, Rong Zhang, Guangyu Sun.

**Data curation:** Bo Wang, Qiuyue Dong.

**Formal analysis:** Xiaofei Liang.

**Funding acquisition:** Xiaofei Liang.

**Investigation:** Xiaofei Liang, Bo Wang, Qiuyue Dong, Lingnan Li.

**Software:** Bo Wang.

**Supervision:** Jeffrey A. Rollins, Rong Zhang, Guangyu Sun.

**Writing – original draft:** Xiaofei Liang.

**Writing – review & editing:** Xiaofei Liang, Jeffrey A. Rollins, Guangyu Sun.

## References

1. Cannon PF, Damm U, Johnston PR, Weir BS. *Colletotrichum*—current status and future directions. *Stud Mycol.* 2012; 73:181–213. <https://doi.org/10.3114/sim0014> PMID: 23136460
2. Liu F, Cai L, Crous PW, Damm U. The *Colletotrichum gigasporum* species complex. *Persoonia.* 2014; 33:83–97. <https://doi.org/10.3767/003158514X684447> PMID: 25737595
3. 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):1060–1065. <https://doi.org/10.1038/ng.2372> PMID: 22885923
4. Manire CA, Rhinehart HL, Sutton DA, Thompson EH, Rinaldi MG, Buck JD et al. Disseminated mycotic infection caused by *Colletotrichum acutatum* in a Kemp's ridley sea turtle (*Lepidochelys kempi*). *J Clin Microbiol.* 2002; 40(11):4273–4280. <https://doi.org/10.1128/JCM.40.11.4273-4280.2002> PMID: 12409409

5. Winter RL, Lawhon SD, Halbert ND, Levine GJ, Wilson HM, Daly MK. Subcutaneous infection of a cat by *Colletotrichum* species. *J Feline Med Surg*. 2010; 12(10):828–830. <https://doi.org/10.1016/j.jfms.2010.07.005> PMID: 20813572
6. Perfect SE, Hughes HB, O'Connell RJ, Green JR. *Colletotrichum*: a model genus for studies on pathology and fungal–plant interactions. *Fungal Genet Biol*. 1999; 27(2–3):186–198. <https://doi.org/10.1006/fgbi.1999.1143> PMID: 10441444
7. Crouch J, O'Connell R, Gan P, Buiate E, Torres MF, Beirn L et al: The Genomics of *Colletotrichum*. In: Dean AR, Lichens-Park A, Kole C, editors. *Genomics of Plant-Associated Fungi: Monocot Pathogens*. Berlin, Germany: Springer Berlin Heidelberg; 2014: p. 69–102.
8. Prusky D, Alkan N, Mengiste T, Fluhr R. Quiescent and necrotrophic lifestyle choice during postharvest disease development. *Annu Rev Phytopathol*. 2013; 51:155–176. <https://doi.org/10.1146/annurev-phyto-082712-102349> PMID: 23682917
9. Prusky D, Lichter A. Activation of quiescent infections by postharvest pathogens during transition from the biotrophic to the necrotrophic stage. *FEMS Microbiol Lett*. 2007; 268(1):1–8. <https://doi.org/10.1111/j.1574-6968.2006.00603.x> PMID: 17227463
10. Hyde KD, Cai L, McKenzie E, Yang Y, Zhang J, Prihastuti H. *Colletotrichum*: a catalogue of confusion. *Fungal Divers*. 2009; 39:1–17.
11. Hiruma K, Gerlach N, Sacristán S, Nakano Ryohei T, Hacquard S, Kracher B et al. Root endophyte *Colletotrichum tofieldiae* confers plant fitness benefits that are phosphate status dependent. *Cell*. 2016; 165(2):464–474. <https://doi.org/10.1016/j.cell.2016.02.028> PMID: 26997485
12. Alkan N, Meng XC, Friedlander G, Reuveni E, Sukno S, Sherman A et al. Global aspects of pacC regulation of pathogenicity genes in *Colletotrichum gloeosporioides* as revealed by transcriptome analysis. *Mol Plant-Microbe Interact*. 2013; 26(11):1345–1358. <https://doi.org/10.1094/MPMI-03-13-0080-R> PMID: 23902260
13. Gan P, Ikeda K, Irieda H, Narusaka M, O'Connell RJ, Narusaka Y et al. Comparative genomic and transcriptomic analyses reveal the hemibiotrophic stage shift of *Colletotrichum* fungi. *New Phytol*. 2013; 197(4):1236–1249. <https://doi.org/10.1111/nph.12085> PMID: 23252678
14. Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR. Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. *Genome Announce*. 2014; 2(3): e00540–14. <https://doi.org/10.1128/genomeA.00540-14> PMID: 24926053
15. Baroncelli R, Amby DB, Zapparata A, Sarrocco S, Vannacci G, Le Floch G, et al. Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*. *BMC Genomics*. 2016; 17(1):555.
16. Gan P, Narusaka M, Kumakura N, Tsushima A, Takano Y, Narusaka Y. Genus-wide comparative genome analyses of *Colletotrichum* species reveal specific gene family losses and gains during adaptation to specific infection lifestyles. *Genome Biol Evol*. 2016; 8(5):1467–1481. <https://doi.org/10.1093/gbe/evw089> PMID: 27189990
17. Hacquard S, Kracher B, Hiruma K, Münch PC, Garrido-Oter R, Thon MR et al. Survival trade-offs in plant roots during colonization by closely related beneficial and pathogenic fungi. *Nat Commun*. 2016; 7:11362. <https://doi.org/10.1038/ncomms11362> PMID: 27150427
18. Zampounis A, Pigné S, Dallery J-F, Wittenberg AHJ, Zhou S, Schwartz DC, et al. Genome sequence and annotation of *Colletotrichum higginsianum*, a causal agent of crucifer anthracnose disease. *Genome Announce*. 2016; 4(4): e00821–16. <https://doi.org/10.1128/genomeA.00821-16> PMID: 27540062
19. Rockenbach MF, Velho AC, Gonçalves AE, Mondino PE, Alaniz SM, Stadnik MJ. Genetic structure of *Colletotrichum fruticola* associated to apple bitter rot and Glomerella leaf spot in Southern Brazil and Uruguay. *Phytopathology*. 2016; 106(7):774–781. <https://doi.org/10.1094/PHYTO-09-15-0222-R> PMID: 27019063
20. Barcelos QL, Pinto JMA, Vaillancourt LJ, Souza EA. Characterization of *Glomerella* strains recovered from anthracnose lesions on common bean plants in Brazil. *PLoS One*. 2014; 9(3):e90910. <https://doi.org/10.1371/journal.pone.0090910> PMID: 24633173
21. Edgerton CW. Plus and minus strains in the genus *Glomerella*. *Am J Bot*. 1914; 1(5):244–254.
22. Tai TH, Tanksley SD. A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue. *Plant Mol Biol Rep*. 1990; 8(4):297–303.
23. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. *Genome Res*. 2009; 19(6):1117–1123. <https://doi.org/10.1101/gr.089532.108> PMID: 19251739
24. Nadalin F, Vezzi F, Policriti A. GapFiller: a de novo assembly approach to fill the gap within paired reads. *BMC Bioinformatics*. 2012; 13(14):S8.

25. Stanke M, Waack S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*. 2003; 19(suppl\_2):ii215–ii225.
26. Korf I. Gene finding in novel genomes. *BMC Bioinformatics*. 2004; 5:59. <https://doi.org/10.1186/1471-2105-5-59> PMID: 15144565
27. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res*. 2008; 18(12):1979–1990. <https://doi.org/10.1101/gr.081612.108> PMID: 18757608
28. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*. 2011; 12:491. <https://doi.org/10.1186/1471-2105-12-491> PMID: 22192575
29. Conesa A, Götz S. Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*. 2008;619832. <https://doi.org/10.1155/2008/619832> PMID: 18483572
30. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31(19):3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
31. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One*. 2010; 5(6):e11147. <https://doi.org/10.1371/journal.pone.0011147> PMID: 20593022
32. Page AJ, Taylor B, Delaney AJ, Soares J, Seemann T, Keane JA, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. *Microb Genom*. 2016; 2(4): e000056. <https://doi.org/10.1099/mgen.0.000056> PMID: 28348851
33. Li L, Stoeckert CJ, Roos DS. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res*. 2003; 13(9):2178–2189. <https://doi.org/10.1101/gr.1224503> PMID: 12952885
34. Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*. 2000; 17(4):540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334> PMID: 10742046
35. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*. 2006; 22(21):2688–2690. <https://doi.org/10.1093/bioinformatics/btl446> PMID: 16928733
36. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011; 27(8):1164–1165. <https://doi.org/10.1093/bioinformatics/btr088> PMID: 21335321
37. Sanderson MJ. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*. 2003; 19(2):301–2. PMID: 12538260
38. Beimforde C, Feldberg K, Nylinder S, Rikkinen J, Tuovila H, Dörfelt H, et al. Estimating the phanerozoic history of the Ascomycota lineages: Combining fossil and molecular data. *Mol Phylogenet Evol*. 2014; 78:386–398. <https://doi.org/10.1016/j.ympev.2014.04.024> PMID: 24792086
39. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*. 2006; 22(10):1269–1271. <https://doi.org/10.1093/bioinformatics/btl097> PMID: 16543274
40. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research*. 2013; 41(12):e121. <https://doi.org/10.1093/nar/gkt263> PMID: 23598997
41. Todd RB, Zhou M, Ohm RA, Leeggangers HA, Visser L, de Vries RP. Prevalence of transcription factors in ascomycete and basidiomycete fungi. *BMC Genomics*. 2014; 15(1):214.
42. Chen W, Lee M-K, Jefcoate C, Kim S-C, Chen F, Yu J-H. Fungal cytochrome P450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. *Genome Biol Evol*. 2014; 6(7):1620–1634. <https://doi.org/10.1093/gbe/evu132> PMID: 24966179
43. Xu C, Chen H, Gleason ML, Xu J-R, Liu H, Zhang R, et al. *Peltaster fruticicola* genome reveals evolution from an invasive phytopathogen to an ectophytic parasite. *Sci Rep*. 2016; 6:22926. <https://doi.org/10.1038/srep22926> PMID: 26964666
44. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat methods*. 2011; 8(10):785–786. <https://doi.org/10.1038/nmeth.1701> PMID: 21959131
45. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol*. 2001; 305(3):567–580. <https://doi.org/10.1006/jmbi.2000.4315> PMID: 11152613

46. Fankhauser N, Mäser P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics*. 2005; 21(9):1846–1852. <https://doi.org/10.1093/bioinformatics/bti299> PMID: 15691858
47. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ. WoLF PSORT: protein localization predictor. *Nucleic Acids Res*. 2007; 35:W585–587. <https://doi.org/10.1093/nar/gkm259> PMID: 17517783
48. Pathan M, Keerthikumar S, Ang C-S, Gangoda L, Quek CYJ, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*. 2015; 15(15):2597–2601. <https://doi.org/10.1002/pmic.201400515> PMID: 25921073
49. Weir BS, Johnston PR, Damm U. The *Colletotrichum gloeosporioides* species complex. *Stud Mycol*. 2012; 73(1):115–180. <https://doi.org/10.3114/sim0011> PMID: 23136459
50. Ye Y, Minami A, Igarashi Y, Izumikawa M, Umemura M, Nagano N, et al. Unveiling the biosynthetic pathway of the ribosomally synthesized and post-translationally modified peptide ustiloxin B in filamentous fungi. *Angew Chem Int Ed*. 2016; 55(28):8072–8075.
51. Schafhauser T, Kirchner N, Kulik A, Huijbers MME, Flor L, Caradec T, et al. The cyclochlorotine mycotoxin is produced by the nonribosomal peptide synthetase CctN in *Talaromyces islandicus* ('*Penicillium islandicum*'). *Environ Microbiol*. 2016; 18(11):3728–3741. <https://doi.org/10.1111/1462-2920.13294> PMID: 26954535
52. Daniel B, Wallner S, Steiner B, Oberdorfer G, Kumar P, van der Graaff E, et al. Structure of a berberine bridge enzyme-like enzyme with an active site specific to the plant family Brassicaceae. *PLoS One*. 2016; 11(6):e0156892. <https://doi.org/10.1371/journal.pone.0156892> PMID: 27276217
53. Baccile JA, Spraker JE, Le HH, Brandenburger E, Gomez C, Bok JW, et al. Plant-like biosynthesis of isoquinoline alkaloids in *Aspergillus fumigatus*. *Nat Chem Biol*. 2016; 12(6):419–424. <https://doi.org/10.1038/nchembio.2061> PMID: 27065235
54. Redinbo MR, Potter PM. Keynote review: mammalian carboxylesterases: from drug targets to protein therapeutics. *Drug Discovery Today*. 2005; 10(5):313–325. [https://doi.org/10.1016/S1359-6446\(05\)03383-0](https://doi.org/10.1016/S1359-6446(05)03383-0) PMID: 15749280
55. Scalbert A: Antimicrobial properties of tannins. *Phytochemistry*. 1991; 30(12):3875–3883.
56. Dilokpimol A, Mäkelä MR, Aguilar-Pontes MV, Benoit-Gelber I, Hildén KS, de Vries RP. Diversity of fungal feruloyl esterases: updated phylogenetic classification, properties, and industrial applications. *Biotechnol Biofuels*. 2016; 9(1):231.
57. Wang C, Zhang S, Hou R, Zhao Z, Zheng Q, Xu Q, et al. Functional analysis of the kinome of the wheat scab fungus *Fusarium graminearum*. *PLOS Pathog*. 2011; 7(12):e1002460. <https://doi.org/10.1371/journal.ppat.1002460> PMID: 22216007
58. Hwang CS, Kolattukudy PE. Isolation and characterization of genes expressed uniquely during appressorium formation by *Colletotrichum gloeosporioides* conidia induced by the host surface wax. *Mol Gen Genet*. 1995; 247(3):282–294. PMID: 7770033
59. Xue C, Park G, Choi W, Zheng L, Dean RA, Xu J-R. Two novel fungal virulence genes specifically expressed in appressoria of the rice blast fungus. *The Plant Cell*. 2002; 14(9):2107–2119. <https://doi.org/10.1105/tpc.003426> PMID: 12215509
60. Kulkarni RD, Kelkar HS, Dean RA. An eight-cysteine-containing CFEM domain unique to a group of fungal membrane proteins. *Trends Biochem Sci*. 2003; 28(3):118–121. [https://doi.org/10.1016/S0968-0004\(03\)00025-2](https://doi.org/10.1016/S0968-0004(03)00025-2) PMID: 12633989
61. Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, Themaat EVL, Does HC. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog*. 2012; 8:e1002643. <https://doi.org/10.1371/journal.ppat.1002643> PMID: 22496661
62. Laugé R, Joosten MHAJ, Van den Ackerveken GFJM, Van den Broek HWJ, De Wit PJGM. The in planta-produced extracellular proteins ECP1 and ECP2 of *Cladosporium fulvum* are virulence factors. *Mol Plant-Microbe Interact*. 1997; 10(6):725–734.
63. Gunn FJ, Tate CG, Henderson PJF. Identification of a novel sugar-H<sup>+</sup> symport protein, FucP, for transport of L-fucose into *Escherichia coli*. *Mol Microbiol*. 1994; 12(5):799–809 PMID: 8052131
64. Brown CK, Gu ZY, Matsuka YV, Purushothaman SS, Winter LA, Cleary PP, et al. Structure of the streptococcal cell wall C5a peptidase. *Proc Natl Acad Sci U S A*. 2005; 102:18391–18396. <https://doi.org/10.1073/pnas.0504954102> PMID: 16344483
65. Soustre I, Letourneux Y, Karst F. Characterization of the *Saccharomyces cerevisiae* RTA1 gene involved in 7-aminocholesterol resistance. *Curr Genet*. 1996; 30(2):121–125. PMID: 8660468
66. Hirooka Y, Kawaradani M, Sato T. Description of *Gibellulopsis chrysanthemi* sp. nov. from leaves of garland chrysanthemum. *Mycol Prog*. 2014; 13(1):13–19.



67. Miyara I, Shafran H, Davidzon M, Sherman A, Prusky D. pH regulation of ammonia secretion by *Colletotrichum gloeosporioides* and its effect on appressorium formation and pathogenicity. *Mol Plant-Microbe Interact.* 2010; 23(3):304–316. <https://doi.org/10.1094/MPMI-23-3-0304> PMID: 20121452
68. Schilling B, Lerch K. Cloning, sequencing and heterologous expression of the monoamine oxidase gene from *Aspergillus niger*. *Mol Gen Genet.* 1995; 247(4):430–438. PMID: 7770050
69. Prusky D, McEvoy JL, Leverentz B, Conway WS. Local modulation of host pH by *Colletotrichum* species as a mechanism to increase virulence. *Mol Plant-Microbe Interact.* 2001; 14(9):1105–1113. <https://doi.org/10.1094/MPMI.2001.14.9.1105> PMID: 11551075
70. Dong S, Raffaele S, Kamoun S. The two-speed genomes of filamentous pathogens: waltz with plants. *Curr Opin Genet Dev.* 2015; 35:57–65. <https://doi.org/10.1016/j.gde.2015.09.001> PMID: 26451981
71. Faino L, Seidl MF, Shi-Kunne X, Pauper M, van den Berg GCM, Wittenberg AHJ, et al. Transposons passively and actively contribute to evolution of the two-speed genome of a fungal pathogen. *Genome Res.* 2016; 26(8):1091–1100. <https://doi.org/10.1101/gr.204974.116> PMID: 27325116
72. Plissonneau C, Stürchler A, Croll D. The evolution of orphan regions in genomes of a fungal pathogen of wheat. *mBio.* 2016; 7(5): e01231–16. <https://doi.org/10.1128/mBio.01231-16> PMID: 27795389
73. Vesper SJ, Jo Vesper M. Possible role of fungal hemolysins in sick building syndrome. *Adv Appl Microbiol.* 2004; 55:191–213. [https://doi.org/10.1016/S0065-2164\(04\)55007-4](https://doi.org/10.1016/S0065-2164(04)55007-4) PMID: 15350795
74. He C, Masel AM, Irwin JAG, Kelemu S, Manners JM. Distribution and relationship of chromosome-specific dispensable DNA sequences in diverse isolates of *Colletotrichum gloeosporioides*. *Mycol Res.* 1995; 99(11):1325–1333.
75. Masel A, He C, Poplawski AM, Irwin J, Manners J. Molecular evidence for chromosome transfer between biotypes of *Colletotrichum gloeosporioides*. *Mol Plant-Microbe Interact.* 1996; 9(5):339–348.
76. He C, Rusu AG, Poplawski AM, Irwin JAG, Manners JM. Transfer of a supernumerary chromosome between vegetatively incompatible biotypes of the fungus *Colletotrichum gloeosporioides*. *Genetics.* 1998; 150(4):1459–1466. PMID: 9832523