Evolutionary history of the cytochrome P450s from *Colletotrichum* species and prediction of their putative functional roles during host-pathogen interactions

Jossue Ortiz-Álvarez[1,2], Sioly Becerra[1], Riccardo Baroncelli[1,3], César Hernández-Rodríguez[4], Serenella A. Sukno[1*] and Michael R. Thon[1*]

Abstract

The genomes of species belonging to the genus *Colletotrichum* harbor a substantial number of cytochrome P450 monooxygenases (CYPs) encoded by a broad diversity of gene families. However, the biological role of their CYP complement (CYPome) has not been elucidated. Here, we investigated the putative evolutionary scenarios that occurred during the evolution of the CYPome belonging to the *Colletotrichum* Graminicola species complex (s.c.) and their biological implications. The study revealed that most of the CYPome gene families belonging to the Graminicola s.c. experienced gene contractions. The reductive evolution resulted in species restricted CYPs are predominant in each CYPome of members from the Graminicola s.c., whereas only 18 families are absolutely conserved among these species. However, members of CYP families displayed a notably different phylogenetic relationship at the tertiary structure level, suggesting a putative convergent evolution scenario. Most of the CYP enzymes of the Graminicola s.c. share redundant functions in secondary metabolite biosynthesis and xenobiotic metabolism. Hence, this current work suggests that the presence of a broad CYPome in the genus *Colletotrichum* plays a critical role in the optimization of the colonization capability and virulence.

Keywords *Colletotrichum* Graminicola species complex, P450 enzymes, CYPome, Evolution, Plant-pathogenic fungi

*Correspondence:
Serenella A. Sukno
ssukno@usal.es
Michael R. Thon
mthon@usal.es
[1] Institute for Agrobiotechnology Research (CIALE), Department of Microbiology and Genetics, University of Salamanca, Villamayor, Salamanca, Spain
[2] Present Address: Programa "Investigadoras e Investigadores por México" Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCyT), Mexico City, México
[3] Department of Agricultural and Food Sciences, University of Bologna, Bologna, Italy
[4] Laboratorio de Biología Molecular de Bacterias y Levaduras, Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, Ciudad de Mexico, México

Background

Species belonging to the genus *Colletotrichum* are plant-pathogen agents that cause anthracnose disease in many crops around the world [1, 2]. Thus, *Colletotrichum* is included as one of the most important plant-pathogenic fungi, because these species may provoke losses up to 100% in many food crops, generating serious economic repercussions [3–5]. In particular, the Graminicola species complex (Graminicola s. c.) comprehends a well-defined monophyletic clade comprised of host-specific species [6]. This complex is shaped by plant pathogenic fungal species such as *Colletotrichum graminicola* on maize, *Colletotrichum falcatum* on sugarcane, *Colletotrichum sublineola* on sorghum and

Colletotrichum eremochloae on cultivated turfgrasses [1, 6–8].

Many *Colletotrichum* species exhibit a multistage hemibiotrophic infection strategy employing a set of several mechanical and enzymatic mechanisms during plant invasion [2, 9, 10] including Carbohydrate-Active Enzymes (CAZymes), proteolytic enzymes, necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs), effector protein candidates and secondary metabolites with phytotoxic activity, among others [11–13]. These virulence factors are important for fungal penetration and development in the host tissues, suppression of the host immune system, as well as cell and tissue destruction [2, 14, 15].

The superfamily of cytochrome P450 enzymes (CYPs) are multi-extended heme-thiolate enzymes that are found in all biological domains [16, 17]. The cytochrome P450 enzymes (CYPs), participate in cellular processes such as biosynthesis of secondary metabolites, utilization of compounds as sole carbon and energy sources, and cellular detoxification, among others [18–23]. CYPs are particularly widely expanded among filamentous fungi, conferring the capability to live in diverse habitats [24, 25].

Plant-pathogenic fungi commonly possess a great diversity of CYPs classified in multiple protein families. In fact, plant pathogenic fungi often possess larger numbers and diversity of CYPs than other fungi. For instance, *Magnaporthe oryzae*, *Cryphonectria parasitica*, *Aspergillus flavus*, *Botrytis cinerea* and *Grosmannia clavigera* each harbor more than 50 CYPs [26]. In general, fungal CYPs fulfill several functions during plant-host interaction, having important roles during fungal development and virulence [27, 28]. Numerous CYP families play parts in mycotoxin biosynthesis pathways [29–32]. Also, CYPs are implied in conidia germination and have critical roles in the ergosterol and hormone-like biosynthesis pathways during cell growth and reproduction [33–35]. Additionally, CYPs confer cell detoxification against antimicrobial compounds secreted by the plant host [19, 36, 37].

Some members of the genus *Colletotrichum* possess a large arsenal of *CYP* genes encoding P450 enzymes, comprising in some cases approximately 1% of all gene content encoded in their genomes [38]. For instance, more than 200 *CYP* genes have been recognized in *Colletotrichum higginsianum* and *Colletotrichum simmondsii* [11, 39]. Usually, paralogous *CYP* genes provide functional redundancy during both biotrophic and necrotrophic stages of infection [14], and, although some paralogous *CYP* genes display an apparent subfunctionalization, there are no experimental data available that support this asseveration assertion yet.

Currently, the exhaustive exploration of the function of the CYPs of *Colletotrichum* has not been unexplored. However, the genomic resources that are now available for many *Colletotrichum* species provide an opportunity to allow the formulation of new hypotheses focused on explaining the biological significance of the Cytochrome P450 complement CYPome in the genus *Colletotrichum*. For this work, we used as model the genomic and proteomic data from *C. graminicola*, a fungal maize pathogen that provokes severe losses in the Americas and Europe [3, 40, 41] as well as other species belonging to Graminicola s. c. [13]. Based on comparative genomics, phylogenetic, and transcriptomic analyses, we performed an exploration of the evolutionary scenarios involved in the evolution of CYPs in the Graminicola species complex (s. c.), with the aim of inferring the putative functional roles of these protein families during the fungal infection into the plant, as well as the biological significance of the wide expansion of these enzyme families.

Results

Distribution and frequency of the CYPs in *Colletotrichum* genome projects

A total of 150–250 amino acid sequences were identified as members belonging to P450 superfamily in each of the *Colletotrichum* genomes (Fig. 1a; Table S1). From all species belonging to the Graminicola s.c., *C. sublineola*, *C. graminicola*, and *Colletotrichum somersetense* possess the greatest variety of CYPs, however, unlike other species, the Graminicola s. c. exhibited a low content of CYPs (< 150 copies) (Fig. 1a; Table S1). In contrast, *Colletotrichum orbiculare*, *Colletotrichum tofieldiae* and *Colletotrichum fructicola* genomes encode 170, 190 and 250 P450 copies respectively, perhaps associated with adaptation to several plant-physiological conditions (Fig. 1a; Table S1). A total of 98 CYP families were identified from the eight members from the Graminicola s. c. tested. Among them, 63 CYP families were singletons, which were uniquely found in one of the eight species without any paralogs. In addition, 21 families harbored only two copies, 10 families possess among 3–5 copies, and four CYP families correspond to multicopy gene families with more than 5 copies into the eight species genomes (Fig. 1b; Table S2). The CYP65, CYP68, CYP526 and CYP570 families exhibited the largest content of paralog copies, harboring between 5–18 copies (Fig. 1b; Table S2). Likewise, *C. graminicola*, *C. somersetense* and *C. zoysiae*, exhibited the most diversity of CYP families (Fig. 1b). In contrast, a reduced content of CYP families were observed in *Colletotrichum falcatum* and *Colletotrichum navitas*. The comparative analysis revealed that 18 CYP families are shared among the Graminicola s.
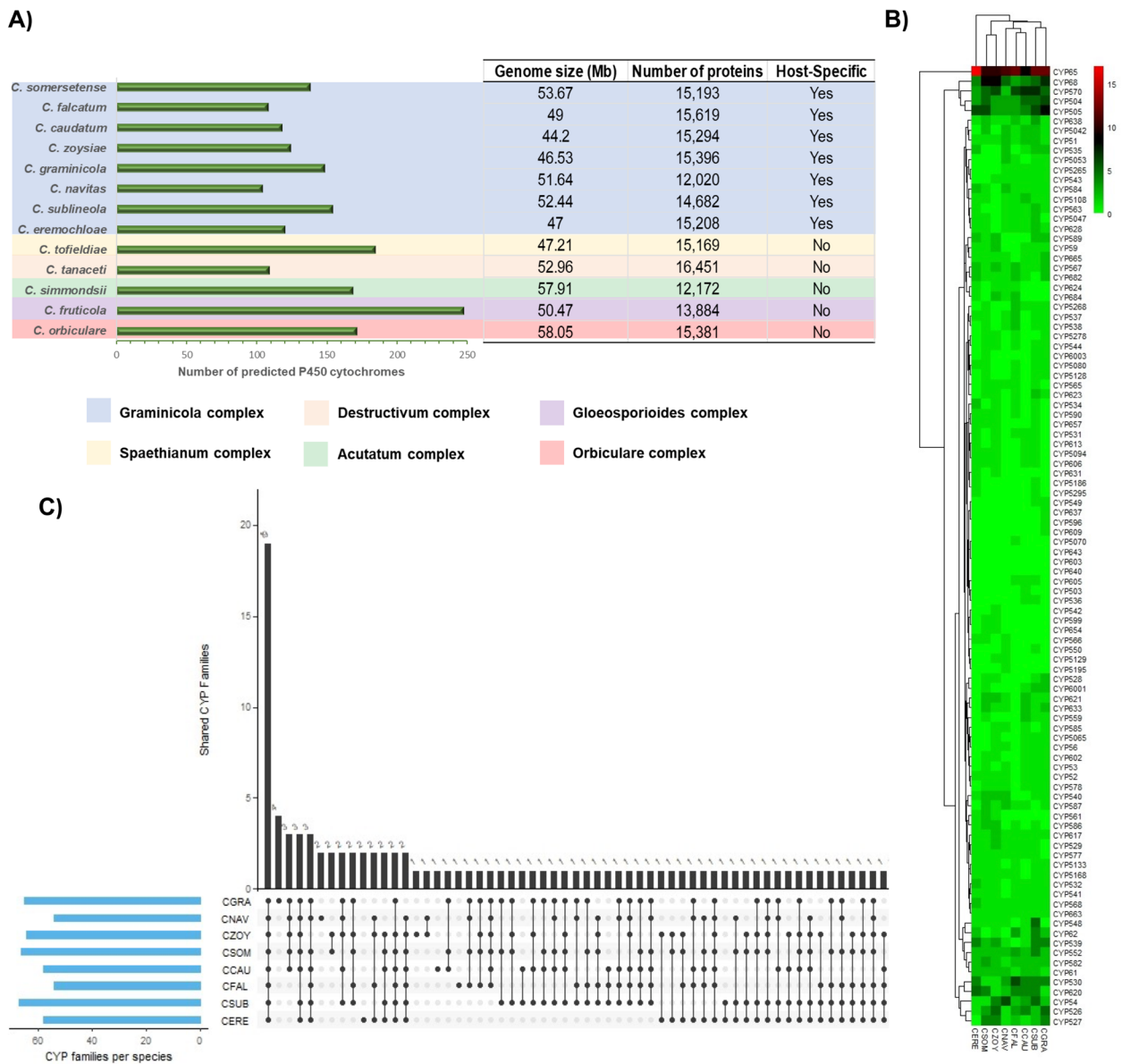
Ortiz-Álvarez *et al. BMC Genomics*   (2024) 25:56

Page 3 of 19

Fig. 1 Comparative analysis of the CYPome harbored in the *Colletotrichum* spp. genomes. **A** Comparison of the genome size, protein containing and their correlation with their host-specificity. Background colors highlight the representative species belonging to each complex. **B** Hierarchical clustering comparison of the p450 enzymes that conform the CYPome of each member belonging to Graminicola s. c. The number and classification of each CYP family was visualized employing a heatmap created with the "pheatmap" package included in R program. The color scale in the bar (black, green, and red) represents the protein abundance in each CYP family. **C** Determination of the number of each CYP family shared in each fungal species from the Graminicola s.c. The figure was created by using the R program. Black bars and black points represent the total number of CYP families shared. The blue bars represent the total CYPs contained in each organism. Abbreviations: CCAU, *C. caudatum*; CERE, *C. eremochloae*; CFAL, *C. falcatum*; CGRA, *C. graminicola*; CNAV, *C. navitas*; CSOM, *C. somersetense*; CSUB, *C. sublineola*; CZOY, *C. zoysiae*; CORB, *C. orbiculare*; CFRU, *C. fructicola*; CTOF, *C. tofieldiae*; CTAN, *C. tanaceti*; CSIM, *C. simmondsii*

c. The CYP families corresponding to CYP52, CYP54, CYP507, CYP530, CYP531, CYP58, CYP559, CYP65 contain the largest number of paralogs in the analyzed genomes. Nonetheless, the CYPome of the Graminicola s. c. was shaped by clans containing non-duplicated CYP families, meanwhile only a few families contain large numbers of paralogs (Fig. 1b; Table S2). Most of the expanded CYP families (> 3 copies), such as CYP65, CYP68, CYP526 and CYP570 were present in all fungal species evaluated (Fig. 1c; Table S3). Some CYP families were found to be restricted to the Graminicola s. c.; for instance, *C. graminicola* harbored 3 species-specific

Ortiz-Álvarez *et al. BMC Genomics*   (2024) 25:56

Page 4 of 19

CYP families, while the rest of the members harbored one species-specific CYP family (Fig. 1c; Table S3). The classification of the CYPs into functional clans, according to the pipeline described by the Fungal Cytochrome P450 Database (FCDB) revealed that most of them were predicted to be involved in xenobiotic and secondary metabolism (Table 1).

Phylogenetic relationships between paralogous CYP families

The phylogenetic divergence in paralogous CYP's was analyzed employing the 10 highly expanded CYP families. We analyzed the evolutionary changes and origin of the paralogous members of CYPs using as models the families CYP504, CYP505, CYP526, CYP527, CYP530, CYP535, CYP552, CYP570, CYP65, and CYP68, because the biological functions of these families have been well described and characterized in other plant pathogenic fungal species [25, 28]. The phylogeny displayed numerous orthologous groups clustered in the phylogeny (Fig. 2). In general, the CYP families were gathered as orthologous groups that possess a consistent with the phylogenetic relationships among members of the Graminicola s. c. Also, it was observed that CYP65, CYP68, CYP505, CYP552 belonging to *C. falcatum* and *C. eremochloae* presented additional duplication events with a putative recent appearance, hence they were cataloged as orthologous/paralogous groups, although this putative duplication event is not in all the phylogenetic tree (Fig. 2). In fact, de distribution of the several orthologous groups belonging to a same families could be associated to an ancient paralogous duplication event (Fig. 2). On the other hand, paralogs from internal duplications were not observed in the rest of Graminicola s.c (Fig. 2). Most of the CYP displayed short branch lengths compared with their clustered orthologs (Fig. 2). Also, the phylogenetic tree showed that CYP65, CYP527, CYP570, CYP535 and CYP552 families are closely related, forming a unique evolutionary clade. On the other hand, CYP504 was recognized as the oldest family in the phylogeny, due to its localization near the root (Fig. 2). The CYPome belonging to the closely related fungi used as external group (*Neurospora crassa*, *Verticillium dahliae* and *Verticillium albo-atrum*) exhibited orthologs paralogous CYP family tested in this analysis. Nevertheless, these species displayed few paralogous CYP's: *N. crassa* clustered two CYP52; *V. dahliae* clustered two CYP526, two CYP68, three CYP570 and two CYP65; *V. albo-atrum* clustered two CYP526 and four CYP570 (Fig. 2).

Expansion and contraction patterns of CYP families among Graminicola s.c.

An analysis of gene family evolution was carried out using the software CAFE. In general, the analysis revealed that the Graminicola s. c. clade exhibits a greater number of family contractions than the number of family expansions, except for *C. sublineola* CBS 129661 where the number of expansions were predominant (Fig. 3; Table S4). In fact, all species except *C. sublineola* CBS 129661 displayed more families experiencing contractions than the family's undergoing expansion. In *C. graminicola*, the number of expansions and contractions was maintained in the same proportion, unlike that of *C. navitas*, where the gene loss is indeed more dominant than duplications. Other species that kept a similar proportion of expansions/contractions were *C. somersetense* and *C. zoysiae*. Regarding to the *CYP* gain and loss, we observed that the family expansion occurred families CYP5070, CYP635, CYP637, and CYP654 (Table S6). The branches that contain the common ancestors of the Graminicola s. c. clade also presented more predominance of gene family contractions and no expansions in some nodes particularly in the branch belonging to the ancestor of *C. graminicola* and *C. navitas*, and the branch corresponding to the common ancestor of *C. eremochloae* and *C. sublineola* strains, the proportion of expansions/contractions indicate a greater amount of gene losses than gene gains (Fig. 3; Table S4).

Diversity and conservation of the motif sites

In general, the sequence logos revealed moderate conservation of the motif sites (Fig. 4). The motif AGXTTXX associated with oxygen binding displayed ~30% of amino acid variations in their most conserved sequence sites located at positions 1, 2 and 6. Approximately, 50% of amino acid variations in this motif were observed at position 3, 4 and 7. Position 5 in motif AGXTTXX was the most conserved motif displaying only a replacement in CYP68 and CYP570 (Fig. 4). The EXXR motif responsible for the stabilization of heme pocket site exhibited conservative amino acids at position 1 and 4, while the positions 2 and 3 displayed also conservative variabilities, with exception of position 3 in CYP68 family, which presented evidence of amino acids with non-shared chemical structure properties L and Q [42]. Likewise, most of the positions at PER sites were maintained well conserved, except for position 2, which displayed high variability in all families. Particularly, CYP68 showed non-conservation in the PER site (Fig. 4). In general, the heme-binding site (FXXGXRXCXG) displayed conservation at position 1, 4, 8 and 10, however, the rest of the positions have high variability (Fig. 4).

Table 1 SMB clusters detected in the genome project of *C. graminicola* strain M1.001. The analysis was performed by using antiSMASH software and carrying out a deep exploration of the genomic resources harbored in NCBI and Mycocosm database

CYP Clan	CYP Family	Putative functions
CYP54	CYP503, CYP54, CYP599, CYP602	Secondary metabolism
CYP504	CYP504	Xenobiotic metabolism
CYP5042	CYP5042	Non determined
CYP505	CYP505, CYP541	Xenobiotic metabolism
CYP507	CYP527, CYP535, CYP570	Xenobiotic metabolism
CYP530	CYP5065, CYP530, CYP663, CYP665	Xenobiotic metabolism
CYP531	CYP5080, CYP631	Xenobiotic metabolism
CYP546	CYP5053	Non determined
CYP547	CYP5070, CYP582, CYP617	Secondary metabolism
CYP58	CYP5094, CYP542, CYP552, CYP682	Secondary metabolism
CYP51	CYP51	Primary metabolism
CYP5108	CYP5108	Non determined
CYP52	CYP52, CYP538, CYP539, CYP584, CYP585	Xenobiotic metabolism
CYP526	CYP526, CYP638	Secondary metabolism
CYP528	CYP528	Non determined
CYP53	CYP53	Xenobiotic metabolism
CYP531	CYP531, CYP532, CYP536	Xenobiotic metabolism
CYP534	CYP534	Non determined
CYP537	CYP537, CYP577	Xenobiotic metabolism
CYP540	CYP540	Primary metabolism
CYP529	CYP543	Non determined
CYP544	CYP544	Non determined
CYP548	CYP548	Xenobiotic metabolism
CYP549	CYP549	Non determined
CYP550	CYP550, CYP633	Secondary metabolism
CYP559	CYP559, CYP606, CYP623	Non determined
CYP56	CYP56	Primary metabolism
CYP65	CYP561, CYP563, CYP565, CYP567, CYP568, CYP65	Secondary metabolism
CYP566	CYP566	Non determined
CYP578	CYP578	Secondary metabolism
CYP59	CYP586, CYP587, CYP59	Secondary metabolism
CYP589	CYP589	Non determined
CYP590	CYP590	Non determined
CYP68	CYP596, CYP68	Secondary metabolism
CYP605	CYP605	Secondary metabolism
CYP609	CYP609	Non determined
CYP61	CYP61	Primary metabolism
CYP613	CYP613	Secondary metabolism
CYP62	CYP62, CYP684	Non determined
CYP533	CYP620, CYP621	Xenobiotic metabolism
CYP624	CYP624	Non determined
CYP574	CYP628	Secondary metabolism
CYP637	CYP637	Non determined
CYP653	CYP653	Secondary metabolism
CYP657	CYP657	Primary metabolism

CYP's did not have match with any clan harbored in FCPD: CYP5047, CYP5128, CYP5129, CYP5133, CYP5168, CYP5183, CYP5195, CYP5278, CYP5295, CYP6001, CYP6003
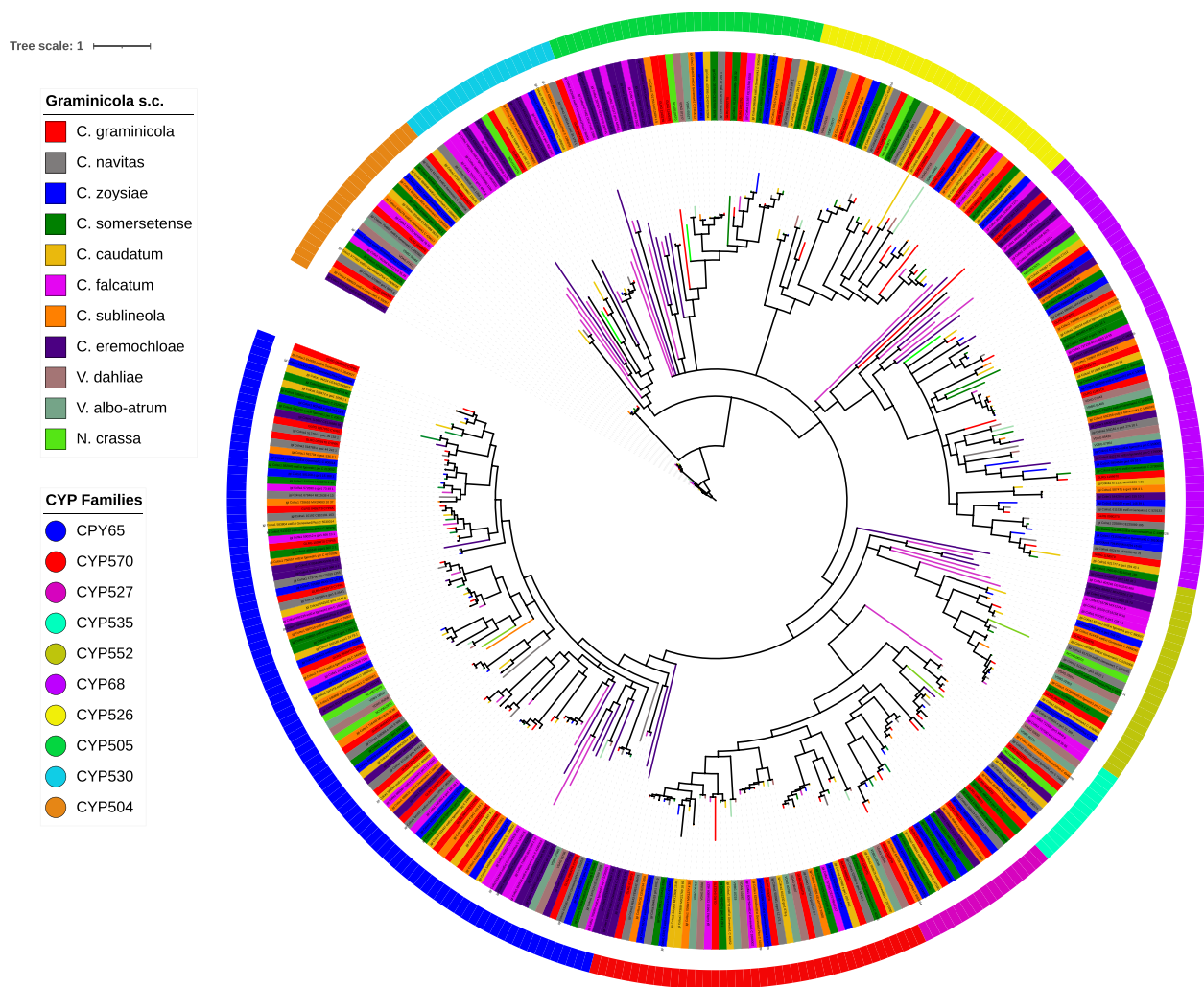
Ortiz-Álvarez *et al. BMC Genomics*   (2024) 25:56

Page 6 of 19

Fig. 2 Bayesian phylogenetic tree constructed with amino acid sequences from representative multi-extended paralogous CYP families. The LG + I + G evolutionary model was selected for the phylogenetic reconstruction. The scale bar represents the number of substitutions per site

Expression of the CYPome during the infection stages in *C. graminicola* M1.001

To identify differentially expressed *CYP* genes during the main infection stages of infection, we reanalyzed the previously published RNA-seq data by focusing only on the genes associated to CYPome [14]. We determined a subset of 58 Differentially Expressed Genes (DEGs) that are the common among the three proposed pairwise comparisons, these three results follow the same comparisons of the O'Connell's analysis (Fig. 5a and b). As shown in Fig. 5a, 3 DEGs were detected in common in the three pairwise comparison GLRG_01187 (CYP527F2); GLRG_01934 (CYP5047A2); GLRG_01817 (CYP68X1). Two of these are up-regulated while the GLRG_01817 gene is repressed (Fig. 5c). Interestingly, the genes GLRG_02160 (CYP5065A2), GLRG_09843 (CYP504B10), GLRG_02897 (CYP621A2), GLRG_09375

(CYP529A2) were suppressed in biotrophic phase and the same genes were up-regulated in the necrotrophic phase, but these 4 genes were not differentially expressed when compared in the appressorium and necrotrophic phase (Fig. 5c). Currently, research on the virulence role of the P450 genes is poorly studied, but these results suggest that these 4 genes may be involved in virulence, playing a key role in the necrotrophic phase of this hemibiotrophic fungus. A total of 26 up-regulated DEGs and 8 DEGs down-regulated were detected at the Necrotrophic Phase (NP) vs Biotrophic Phase (BP) comparison, from a total of 13 DEGs in relation to secondary metabolism, 12 DEGs in xenobiotic metabolism, and 9 expressed genes which were annotated as non-determined for a *CYP* specific response (Fig. 5d). In the case of NP vs Appressorium phase (PA) comparison, 44 DEGs were detected, as well as 19 down-regulated and 25 up-regulated, which
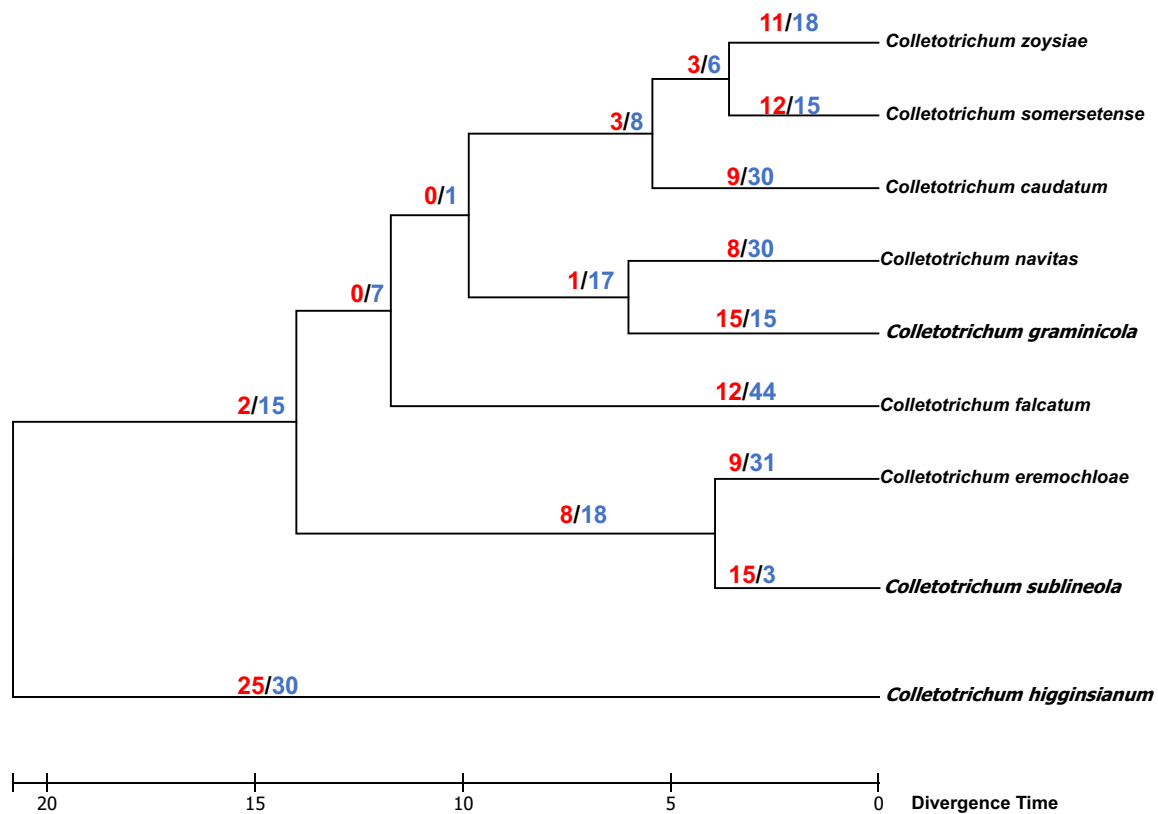
Fig. 3 Gene gain and loss patterns in species belonging to *Colletotrichum* Graminicola s. c. Phylogenomic tree was constructed based on distance-methods by using 1765 single-copy orthologous genes. Divergence times were estimated with MEGAX. Numbers at the nodes represent expansion (red) and contraction (blue) events

22 *CYP* genes were related with secondary metabolism, and 11 DEGs with xenobiotic metabolism, but 11 *CYP* genes were classified as non-determined for any category. In the compassion of BP vs AP 24 DEGs were found, 15 down-regulated and 9 up-regulated, of which, 12 were related with secondary metabolism, and 6 DEGs with xenobiotic metabolism, 6 with no associated metabolism category. (Fig. 5d). The genes related to secondary metabolism were more predominant with a total of 28 genes, within this category. Some DEG's *CYPs* were recognized as members of Polyketide synthase (PKS) and PKS-Non-Ribosomal Peptide Synthetase (NRPS) (HYBRID) clusters (Fig. 5e). Two *CYPs* located into the HYBRID clusters and one *CYP* located in a PKS cluster displayed up-regulation at the NP vs PA, while two *CYPs* cataloged as members of HYBRID and PKS cluster respectively during the NP vs BP were up-regulated, and all *CYPs* classified into Dimethylallyl tryptophan synthase -like cluster (DMAT) were defined as down-regulated genes in all conditions. On the other hand, all *CYPs* harbored in Secondary Metabolite Gene Clusters (SMGCs) displayed down-regulation when BP vs AP conditions were evaluated. This first summary of RNA-Seq data [14] in three

stages of disease development showed the involvement of only *CYP* genes during pathogenicity of *C. graminicola* M1.001 thus explaining the role of *CYP* families in plant-pathogen interaction.

Structural evolution of the paralogs CYP enzymes of *C. graminicola* M1.001

We performed an analysis to evaluate the hypothetical structural evolution using a 3D-structure-based phylogeny. The evolutionary structural relationship among paralogous CYP enzymes belonging to *C. graminicola* is shown in Fig. 6. In the phylogeny, almost all paralogous CYP families clustered in a highly similarity topology compared with the amino-acid sequence-based phylogeny (Figs. 2 and 6). Nevertheless, some paralogs belonging CYP527 and CYP570 clustered with other paralogs displaying a putative structural evolutionary convergence. Thereby, most of the paralogous CYP members show a common folding origin, but CYP527 and CYP570 experienced polyphyletic structural evolution (Fig. 6). Additionally, the Structural-Based Dendrogram (SBD) topology reflects that of the CYP families, and there are

Fig. 4 Sequence logos from conserved motive sites harbored in representative paralogous CYP families. Alignments were performed with MUSCLE v3.8.3 software. The consensus logos images were generated by using WebLogo webserver

some paralogous enzymes of the same family that exhibit differences in the structural domain architectures (Fig. 6).

Diversity of CYP families harbored in SMGCs and syntenic conservation

A total of 30 *CYP* genes were identified inside the SMGCs recognized in the genome project of *C. graminicola* M1.001 (Table 2). Most of the *CYP* genes were recognized as paralogous members belonging to *CYP65, CYP68, CYP526, CYP527, CYP539, CYP570* and *CYP682* families. Four *CYP65* genes were detected into SMGCs

clusters (Table 2). PKS clusters exhibited that harbor 11 *CYP* genes, while one and two *CYPs* were present in the siderophore and NRPS-like clusters respectively (Table 2). On the other hand, the number of CYPs in DMAT and HYBRID clusters correspond to three and five genes respectively (Table 2). All paralogous *CYPs* of *C. graminicola* M1.001 were harbored in SMGCs that display an evident loss of synteny. Besides, the genomic arrangements of the clusters indicate that several *CYP* paralogous members are in regions alongside genes not associated with SMGCs (Fig. 7).

(See figure on next page.)
Fig. 5 Transcriptome analysis for differentially expressed genes in three stage developmental stages of infection in *C. graminicola* M1.001 associated with CYPome. **A** Venn diagram shows common and unique genes for all three comparisons. **B** Volcano Plots show *p-value* and Fold Change for each pairwise comparison with ID CYP for each differential expressed gene, where red is up-regulated, and blue is down-regulated. **C** Heatmap with cluster showing the pattern of expression in the three pairwise comparisons, CYP genes are clustered, and the color gradient is the Fold Change value, where green is up-regulated, and red is down-regulated. **D** UpSetR displays are grouped differential expressed genes for all pairwise comparisons with an intersection about the metabolism (Secondary metabolism, Xenobiotic metabolism, and non-determined metabolism). NP = 60 h post-infection (hpi), in necrotrophic phase. BP = 40hpi, in biotrophic phase. PA = 22 hpi, in appressorium phase
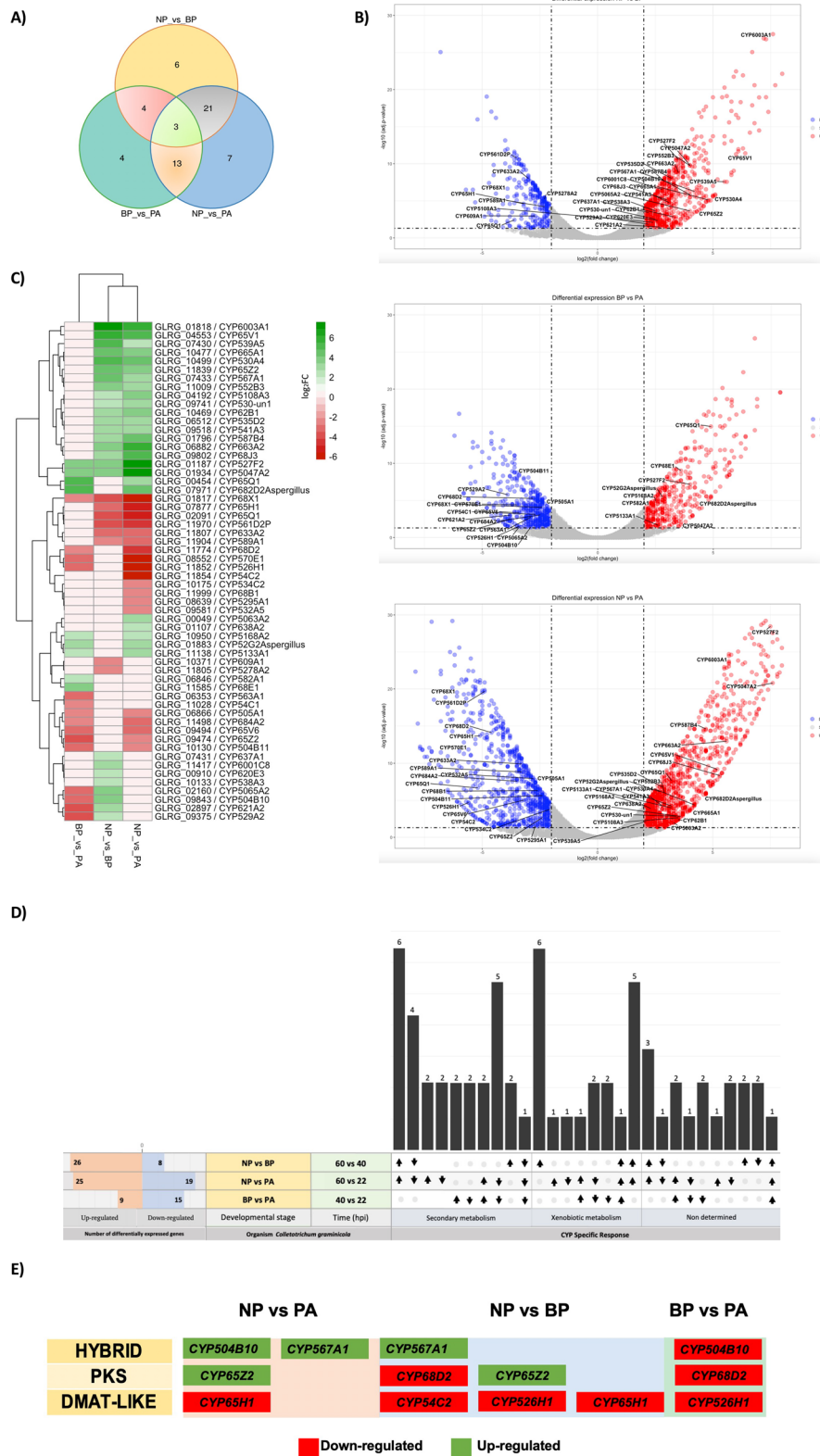
Fig. 5 (See legend on previous page.)

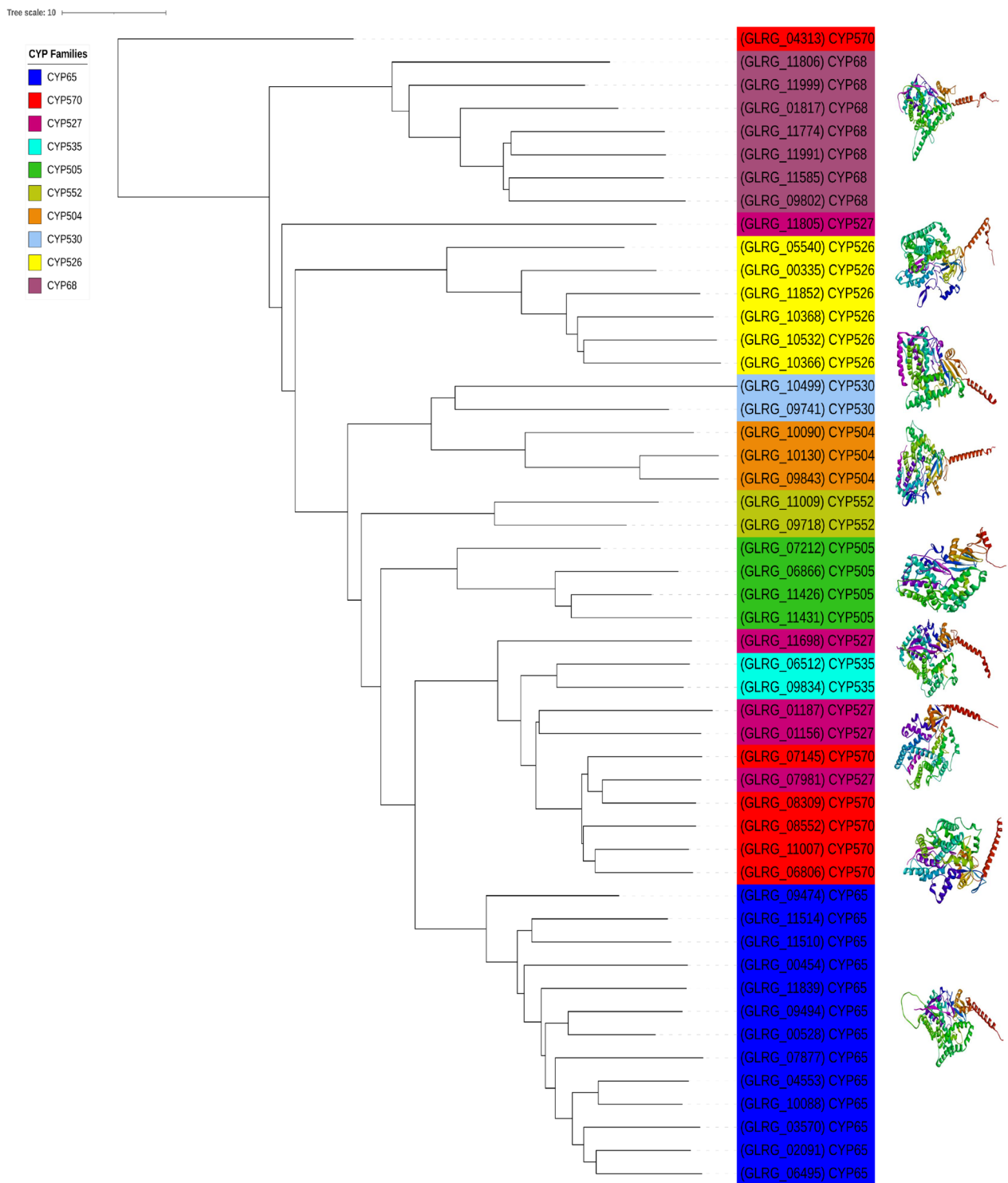**Fig. 6** Phylogeny constructed with the hypothetical three-dimensional (3D) structures of representative paralogous CYP families in *C. graminicola* M1.001. The phylogenetic tree represents the evolutionary relationships among CYP families based on 3D-structural similarities. The phylogeny was constructed employing hypothetical three-dimensional structures generated with Alphafold2 software and computed with DALI software

Table 2 Hypothetical functional assignation of CYP families. The assignment was performed based on the analysis described by [25]

Gene ID	CYP gene	SM cluster	CLUSTER ID	Source
GLRG_09843T0	CYP504B10	HYBRID	Colgr1.5	Mycocosm/AntiSMASH
GLRG_10366T0	CYP526H1	HYBRID	Colgr1.7	Mycocosm/AntiSMASH
CLGR_10368T0	CYP526A1	HYBRID	Colgr1.7	Mycocosm/AntiSMASH
CLGR_10371T0	CYP609A1	HYBRID	Colgr1.7	Mycocosm/AntiSMASH
GLRG_07433T0	CYP567A1	HYBRID	Colgr.1.46	Mycocosm/AntiSMASH
GLRG_11698T0	CYP527F1	NRPS-like	Colgr1.52	Mycocosm/AntiSMASH
GLRG_01187	CYP527F2	NRPS-like	NW_007361007	NCBI
GLRG_08552	CYP570E1	Siderophore	NW_007361046	NCBI
GLRG_11852T0	CYP526H1	DMAT-like	NW_007361152	NCBI
GLRG_11854T0	CYP54C2	DMAT-like	NW_007361152	NCBI
GLRG_07877T0	CYP65H1	DMAT-like	Colgr.45	Mycocosm/AntiSMASH
GLRG_10319T0	CYP585A3P	PKS	Colgr1.6	Mycocosm/AntiSMASH
GLRG_11880T0	CYP589A1	PKS	Colgr1.9	Mycocosm/AntiSMASH
GLRG_11881T0	CYP539C1	PKS	Colgr1.9	Mycocosm/AntiSMASH
GLRG_10534T0	CYP682B3	PKS	Colgr1.10	Mycocosm/AntiSMASH
GLRG_10536T0	CYP596B1	PKS	Colgr1.10	Mycocosm/AntiSMASH
GLRG_11562T0	CYP606B2	PKS	Colgr1.19	Mycocosm/AntiSMASH
GLRG_09074T0	CYP68B1	PKS	Colgr1.21	Mycocosm/AntiSMASH
GLRG_03513T0	CYP665A1	PKS	Colgr1.26	Mycocosm/AntiSMASH
GLRG_11834T0	CYP539C1	PKS	Colgr1.30	Mycocosm/AntiSMASH
GLRG_11839T0	CYP65Z2	PKS	Colgr1.31	Mycocosm/AntiSMASH
GLRG_11774T0	CYP68D2	PKS	Colgr1.40	Mycocosm/AntiSMASH
GLRG_10532T0	CYP526A1	PKS-like	NW_007361066	NCBI
GLRG_10534T0	CYP682B3	PKS-like	NW_007361066	NCBI
GLRG_07145T0	CYP570C4	PKS-like	NW_007361033	NCBI
GLRG_08309	CYP570B1	PKS-like	NW_0077361042	NCBI
GLRG_11426	CYP505C1	PKS-like	NW_007361091	NCBI
GLRG_09718	CYP552B1	PKS-like	NW_007361055	NCBI
GLRG_00454	CYP65Q1	PKS-like	NW_007361005	NCBI
GLRG_11514	CYP65AK2	PKS-like	NW_007361094	NCBI

Discussion

Although *Colletotrichum* species exhibit a large CYPome shaped by many families [38], assigning the specific biological roles to each member is a research challenge. According to our analysis, some members from *Colletotrichum* genus have larger CYPome in terms of the number of CYP elements than other plant pathogenic fungi such as *Cryphonectria parasitica* (70 CYP members), *M. oryzae* (77 CYP members) or *Aspergillus* sp. (57–92 CYP members), whose CYPomes are well documented in the fungal CYP database (http://p450.riceblast.snu.ac.kr/). However, in terms of gene gain and loss, Graminicola s. c. experienced a reductive evolution in several families of its CYPome, but the presence of a large number of CYPs is derived from an extensive appearance of paralogs in certain CYP families, and the maintenance of several CYP families. This information suggests that the presence of a

broad CYPome possibly plays a critical role upon the optimization of the colonization capability and virulence in plant pathogenic fungi, resulting in its significant importance for the adaptation of pathogens during plant invasion [43, 44].

The size of CYPome from Graminicola s. c. is reduced in comparison with the number of CYP enzymes identified in *Colletotrichum* species from Spaethianum, Acutatum, Gloeosporioides and Orbiculare s. c., despite displaying a proteome with a similar or even larger size than the proteomes from the species belonging to the complexes described above. Hence, it is suggested that CYPome of the members of the Graminicola s. c. were subjected to gene losses during their evolutionary history. In addition, the members of this complex suffered several gene losses in other families associated with virulence such as lineage specific candidate effectors and CAZY enzymes, indicating that the contraction
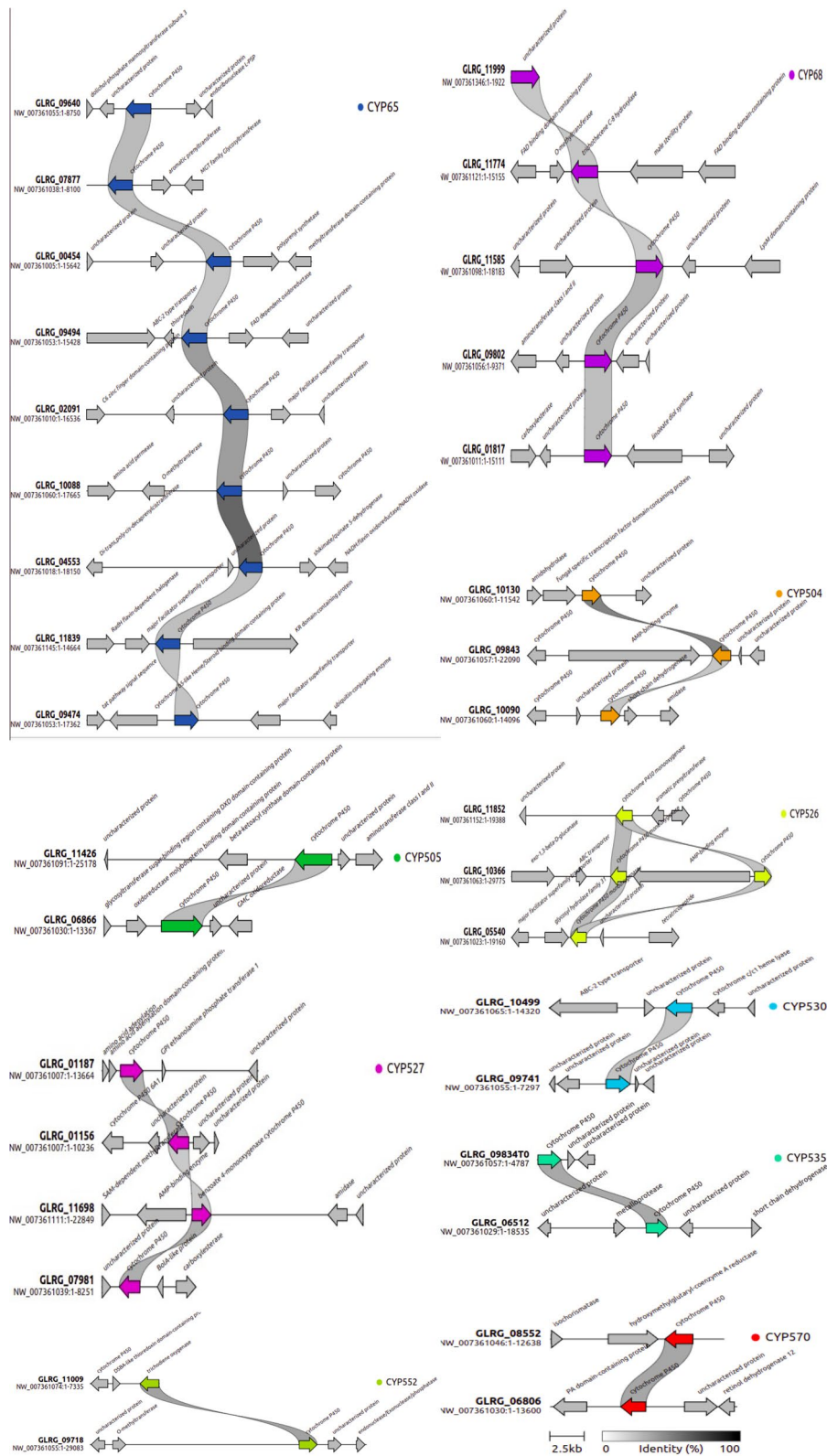
Fig. 7 Synteny conservation of CYP genes harbored in Secondary Metabolite Gene Clusters (SMGC) in *C. graminicola* M1.001. Legends at the bottom represent the functional assignation of the genetic elements found in the clusters. Synteny analysis was performed by using the MAUVE program included in the Geneious Prime software

Ortiz-Álvarez *et al. BMC Genomics*   (2024) 25:56

Page 13 of 19

of the CYPome may be associated mainly to host range and specificity [11, 13, 39]. Although there is non-specific gene family expansion associated with host range, regularly eudicot infecting species have a higher overall number of gene families than monocot infecting species [13]. According to our results, eudicot-infecting *Colletotrichum* species*,* for instance *C. fructicola* possess very expanded CYPomes in terms of number of CYP. Therefore, our results are consistent with the hypothesis monocot pathogens have lower diversity of CYPome encoding genes than eudicot infecting *Colletotrichum* species [13]. Moreover, these fungi present a broad expansion of their virulence related gene families [39]. Earlier studies also affirmed that protein gains and losses are associated with their host range and specificity [11, 45].

Unfortunately, most of these families lack experimental evidence to confirm their functions. Nevertheless, according to the CYP classification [25, 46], we predicted functions based on similarity and functional domain content (Table S5), where most of the CYPome of members belonging to the Graminicola s.c. might be involved in the synthesis of secondary metabolites, since, based on previous experimental evidence, several families present in the genomes of these fungi participate in xenobiotic metabolism [47]. In other fungi, for instance *Fusarium* and *Trichoderma* spp., several members of the CYPome are associated in cell detoxification and production of fungal compounds [28, 48]. We propose that the putative function of the CYPome of the Graminicola s. c. members is associated with the biosynthesis of toxins or chemical modification of compounds secreted by the host.

In *Trichoderma* spp. the expanded CYPome enabled their survival in their respective habitats [48]. The genus *Fusarium* is shaped by several species of plant pathogens that also exhibit a saprophytic stage. The presence of a wide CYPome in *Fusarium* spp. is not only associated with virulence, since there are CYPs whose function is closely related with sexual spore development [49]. In the case of *Colletotrichum* spp., its CYPome may exhibit functional roles resembles to of *Trichoderma* spp. and *Fusarium* spp., thereby the evolution of CYPome of *Colletotrichum* might be influenced by the selection pressures exercised during their adaptation as pathogens. The expansion of CYPome may allow them to be more efficient during their invasion of the host.

The CYPome of members of the Graminicola s. c. are composed mostly of single-copy families. The multi-copy families, even though they represent a low percentage of the total CYPome, are very expanded by most species. In fact, the most expanded CYPs families are conserved into all Graminicola s. c. Most of the CYPome display families that cluster only unique copies or duplicated genes, nonetheless these families were not present in all the analyzed species. In fact, various single-copies CYP families may be catalogued as species-restricted families. Species-restricted CYP families have been reported in Basidiomycete biotrophic plant pathogens, where these genes may fulfill an important role in the host invasion [50]. Also, the presence of species-restricted families is very common in pluricellular organisms, for instance as water strides, wherein these gene families have allowed several evolutionary advantages such as a better locomotion [51]. The presence of species-restricted genes confers adaptive advantages in the organisms for phenotypic novelty [52, 53]. Our findings suggest that, although the maintenance of multi-copy families indicate that they are critically essential for Graminicola s. c members, the presence of species-restricted families suggest not only specific gene loss events but also that the retained CYPs are especially important for their lifestyle. The species restricted CYPs play new functions in the *Colletotrichum* Graminicola s.c.

The paralogous CYP families employed for this study possess the greatest number of orthologous and paralogous members; thus, it is important to inquire about the most accurate biological background that prompted their expansions. CYP504 family has important roles in xenobiotic metabolism in *Trichoderma* species, as well as CYP504 enzymes in *Aspergillus nidulans* participates in the degradation of phenylacetate as sole carbon source [48, 54]. CYP505 family are membrane-associated cytochromes capable of hydroxylate fatty acid [55]. The expression of CYP505 allows plant-pathogen fungi, for instance *Fusarium oxysporum* to employ oxidized fatty acids for inactivating plant defense system [56]. The members CYP526, CYP65 and CYP68 families underlie the biosynthetic pathway of trichothecene mycotoxins in *Fusarium* sp. [28, 57]. The CYP527, CYP530, CYP535 and CYP570 families have displayed putative roles in xenobiotics metabolism [25, 48]. The presence of members belonging to the CYP552 family confers fungal protection against toxic compounds secreted by the host, however, CYP552 cytochromes also may underlie secondary metabolites pathways [25, 58].

All paralogous CYPs families described above are highly expanded and conserved among Graminicola s. c members, which means that they already were present in the common ancestral *Colletotrichum* species, and their expansion may be a consequence of multiple ancestral and modern duplications that occurred both in the common ancestor and after the speciation event. However, there is a small number of species-restricted paralogous detected in the CYPs of study. Hence, the events of duplications in multi-protein CYPs were complemented by a set of gain and loss protein processes. This evolutionary

Ortiz-Álvarez *et al. BMC Genomics*   (2024) 25:56

Page 14 of 19

scenario has already been well documented in some P450 cytochromes, such as members of CYP52 family in Saccharomycetales yeast [59]. In fact, the CYPs contraction was more pronounced in some members of the Graminicola s. c. The gene loss events occurred in several *Colletotrichum* species already described in earlier studies have mentioned the loss of orthologous and paralogous members into the multigene CYPs [28, 45]. On the other hand, the phylogenetic distribution suggests that CYP504 evolved early, while other families such as CYP65, CYP570, CYP527, CYP535 and CYP552 are closely related, and their origin is more recent.

Remarkably, in *C. graminicola*, some members belonging to distinct paralogous CYP families display different phylogenetic relationships at tertiary structure level, since those paralogous CYP's display structural similarities among them, indicating a putative event of convergent/parallel evolution. This evolutionary event detected in CYPs belonging to *Colletotrichum* is not an unusual phenomenon, rather it is quite common among the P450 superfamily members [16, 60]. For instance, CYP384 and CYP2J19, two enzymes from *Tetranychus kanzawai* and canaries respectively, do not share a common lineage, however both families have similar folding and thus enzymatic activity [61]. This example of convergent evolution enabled the early diverging CYPs to have similar functions. This phenomenon has been identified in P450s of the CYP52 family, which have affinity by a wide range of alkanes and fatty acids but maintain some functional moieties [59]. Is important to emphasize that.

The presence of a high number of paralogous proteins harbored into CYPs is associated mainly with evolutionary process for adapting to niches frequently exposed to several environmental and nutritional conditions [59]. It is still unclear why certain CYPs in *Colletotrichum* and other fungal species are highly expanded. In mammals, the expansion of the opsin gene family, which encodes a protein that is located in photoreceptor cells of primates, allows them to distinguish a wide spectrum of wavelengths [62, 63]. Hence, the high fate of duplication in many CYP families may contribute as an adaptation mechanism for increasing or improvement of their virulence. Nonetheless, it is of utmost importance to inquire which is the biological role of each paralogous. Exploring the transcriptional profiles of the CYPome based on raw data obtained previously [14] by using as model *C. graminicola* M1.001, 59 *CYP* genes belong to *C. graminicola* showed upregulation and downregulation profiles in the different stages of infections. In fact, most of this cluster of *CYP* genes were overexpressed during necrotrophic phase, while various *CYPs* were repressed in the biotrophic phase. These findings suggest that *CYPs* may play a fundamental role in the cellular and tissue

destruction of the plant host by *C. graminicola*. However, several paralogous, members of CYP65, CYP68, CYP535 and CYP527 display similar expression levels, suggesting that many paralogous are still functionally redundant, although signals of subfunctionalization also might be associated, however, is necessary to obtain more experimental data for confirming this asseveration. Perhaps, the robustness event, a phenomenon that explain the maintaining of redundant paralogous genes in an organism [64, 65], may be responsible of the putative redundancy of orthologous and paralogous members of the CYPome with the purpose of maintaining the normal functions of the fungal cells in the presence of several perturbations during the stages of their lifestyle.

In this work we also examined the genomic context and changes in the motif sequences of the CYPome by using as model multi-protein CYP families. In general, in each CYPs there is a low conservation of the signature features of the motif sites. In particular, the motif sequence FXXGXRXCXG is highly variable, indicating great variations into the catalytic pocket. Likewise, the variations observed in other motif sequences reflect the CYPome belonging to Graminicola s. c have accumulated several mutations along their evolutionary history, by exposition to several pressure selection conditions. These variations could be the result of the adaptation of each CYP for interacting with several substrates, as it was observed into the CYP52 family [59]. On the other hand, genomic rearrangements in the region of the main paralogous P450 families indicate an evident loss of synteny. A probably mechanisms associated with the random location in the genome of the paralogous *CYPs* could be mediated by transposable elements [63, 66], however, more additional experimental information is necessary for confirming this hypothesis. Although some *CYPs* are in SMGCs, most of them are in regions whose genetic elements do not provide clear information about their biological roles.

Conclusions

The evolutionary analysis of the CYPome of Graminicola s. c provides new insights into the putative environmental phenomena involved in the expansion and contraction events observed in the CYPome of *Colletotrichum*. Even though this study allowed to inquire that CYPome may be an essential role in the biosynthesis of secondary metabolites, antifungal compounds, or confers protection against plant defense mechanisms during the interaction the host, it is necessary to confirm this through further laboratory experiments. CYPome of Graminicola s.c. suffered reductive evolution, expanding single gene CYP families and

Table 3 List of *Colletotrichum* spp. genome resources used in this study

Abbr.	Strain	Organisms	Species complex	Host	Origin	Accession Nº	Database	Bioproject	Reference
CCAU	CBS 131602	*C. caudatum*	Graminicola	*Sorghastrum nutans*	USA	JAHLTY000000000.1	NCBI	PRJNA262368	[13]
CERE	CBS 129661	*C. eremochloae*	Graminicola	*Eremochloa ophiuroides*	USA	JAH-MUH000000000.1	NCBI	PRJNA262442	[13]
CFAL	MAFF 306170	*C. falcatum*	Graminicola	*Saccharum officinarum*	Japan	JAHLKA000000000.1	NCBI	PRJNA262221	[13]
CGRA	M1.001	*C. graminicola*	Graminicola	*Zea mays*	USA	ACOD00000000.1	NCBI	PRJNA37879	[14]
CNAV	CBS 125086	*C. navitas*	Graminicola	*Panicum virgatum*	USA	JAHLJV000000000.1	NCBI	PRJNA262371	[13]
CSOM	CBS 131599	*C. somersetense*	Graminicola	*Sorghastrum nutans*	USA	JAHMAF000000000.1	NCBI	PRJNA262441	[13]
CSUB	CBS 131301 v1.0	*C. sublineola*	Graminicola	*Sorghum bicolor*	Burkina Fasso	-	JGI	-	[13, 66]
CZOY	MAFF 235873	*C. zoysiae*	Graminicola	*Zoysia tenuifolia*	Japan	JAHMAF000000000.1	NCBI	PRJNA262217	[13]
CORB	MAFF 240422	*C. orbiculare*	Orbiculare	*Cucumis sativus*	Japan	AMCV00000000.1	NCBI	PRJNA171217	[13, 44]
CFRU	Nara gc5	*C. fructicola*	Gloeosporioides	*Fragaria x ananassa*	Japan	ANPB00000000.1	NCBI	PRJNA171218	-
CTOF	CBS 168.49	*C. tofieldiae*	Spaethianum	*Lupinus polyphyllus*	Germany	LFHQ00000000.1	NCBI	PRJNA286723	[67]
CTAN	BRIP5731	*C. tanaceti*	Destructivum	*Tanacetum cinerariifolim*	Australia	PJEX00000000.1	NCBI	PRJNA421029	[68]
CSIM	CBS 122122	*C. simmondsii*	Acutatum	*Carica papaya*	Australia	JFBX00000000.1	NCBI	PRJNA239224	[13]

conserving few families containing multiple paralogous. In fact, we suggest that CYPome suffered several gene loss events after an extensive duplication event. On the other hand, although the CYPome of Graminicola s.c. suffered some depurative events, some CYP families generated many paralogs to compensate for the possible negative effects of the gene losses. Therefore, the CYPome in our organisms of study is result of an evolutive process for conferring a better fitness in the members of Graminicola s. c.

Methods
Genome project data
Proteomes from genome projects of 8 species/strains from the Graminicola s.c. were used for this study [13, 14, 45, 67–69]. One representative genome project belonging to the *Colletotrichum* species complex (s.c.) closely related to Graminicola s.c. was chosen for some comparative analysis (Table 3).

Annotation and classification of the CYP proteins
The annotation and classification pipeline of the CYPome was carried out based on a two-step procedure. First, the annotation step was performed by using Hidden Markov Models with HmmerBuild and HmmerSearch programs

included in the HMMER v3.3 package [70]. The seed alignment of domain PF00067 deposited in the Pfam protein family database (www.pfam.xfam.org) was used for protein annotation in the selected fungal proteomes, applying an E value = 10^{-4} as a threshold for the selection of the positive hits. The classification procedure was performed with the BLASTP tool comparing the positive hits against the complete CYP sequence list deposited into the Fungal Cytochrome P450 Database (FCDP) [26]. The family assignation of the CYP enzymes was performed based on the highest similarity percentage (at least 40%) exhibited by each hit during the BLAST analysis, according to the classification parameters established as previously referred [71, 72].

Phylogenetic analysis of representative multiprotein CYP families
Annotated CYP enzymes were aligned with MUSCLE v. 3.8.31, included in the program SeaView version 4.7 [73, 74]. The selection of the best evolutionary model was computed with ProtTest 3 [75]. A Bayesian MCMC-based phylogenetic tree was constructed with MrBayes version 3.2.7a [76], by using LG + I + G evolutionary model. The BMCMC was run for 3×10^{6} generations and sampling every 100 generations. The posterior probabilities

threshold was over 75%. The phylogenetic tree was edited with the web server iTOL v3 [77].

Gene family evolution

CAFE software version 4.2.1 was employed for the gene family expansions/contractions analysis [78]. As part of the analysis, the deduced proteomes of *Graminicola s.c.* were classified into orthogroups with OrthoFinder v0.4 [79]. All single-copy protein orthogroups were selected for the phylogenomic tree reconstruction based on the software described above. The dendrogram generated was converted to an ultrametric tree with MEGAX software with a penalized likelihood method and a TN algorithm [80]. Calibration of the tree was performed using the divergence times of 15.78 My between *C. graminicola* and *C. sublineola* estimated previously [69]. The corresponding branch that clusters with *C. higginsianum* was used as an outgroup group. For running CAFE analysis, a lambda value (maximum likelihood value of the birth-death parameter) of 0.0395877 was assumed. The gene families with representative size variance were detected employing 1,000 random, number of threads = 10, and a *p-value* cutoff ≤ 0.01. The branches with a significant evolutionary value were identified based on the Viterbi algorithm with a *p-value* cutoff of 0.05. The analysis of expansions and contractions for each CYP family were obtained by using CAFE.

Conservation of motif sequences analysis

Recognition and analysis of the motif sites in the multiprotein CYP sequences alignment obtained by MUSCLE v3.8.3 was performed using GeneDoc (http://nrbsc.org/gfx/genedoc). The residues assigned in the motif sites were reserved for analysis. Consensus logos corresponding to motif sequences were generated and visualized with the WebLogo webserver, plotting a stack of amino acids for each position [81, 82].

Differential expression analysis of the *C. graminicola* M1.001 transcriptome

The expression profile analysis of the CYPome of *C. graminicola* during the host-pathogen interaction was carried out employing the *C. graminicola* M1.001 pathosystem expression data obtained by O'Connell et al. (2012). Illumina RNA sequencing data for three developmental stages were selected corresponding to the following conditions: 22 h post-inoculation- *in planta* appressoria (PA), 40 h post infection (hpi) - biotrophic phase (BP), and 60 hpi – necrotrophic phase (NP). Differentially Expressed *CYP* genes from the total set of Differentially Expressed Genes (DEGs) were selected

for each pairwise comparison: NP vs PA, NP vs BP, BP vs PA. From Supplementary Table 14 of O'Connell et al. [14], we extracted normalized read counts. Differential expression analysis for each pairwise comparison was performed using the DESeq2 R package v.1.28.1 [83]. In this study, genes with a log2 fold-change ≥ 2 or ≤ -2, and a *p-value* < 0.05, were considered as DEGs, of which only the *CYP* genes were used in downstream analysis. Shared DEG *CYP*s were visualized by plotting a Venn Diagram generated with the Intervene web tool [84]. The R package pheatmap 1.0.12 was used for showing and comparing the number of each *CYP* family differentially expressed by the different infection stages [85]. Volcano plots were created with the following criteria: *p-value* 0.05 and ≤ $\log_2$foldchange☐ = 2 by using the R package ggplot2 3.3.2 [86]. Intervene's UpSetR module was used for preparing a general graphic UpSetR with all intersection of three pairwise comparisons. The DEG *CYP* gene information was used for recreating a schematic representation which infers the putative Secondary Metabolite Gene Clusters (SMGCs) expressed in the different infection stages.

Structure-based dendrogram (SBD) construction of CYP enzymes of *C. graminicola* M1.001

The SBD was obtained according to the following: the hypothetical three-dimensional structures (TDS) belonging to the paralogous CYP's corresponding to the families CYP65, CYP68, CYP504, CYP505, CYP526, CYP527, CYP530, CYP535, CYP552 and CYP570 were predicted for this analysis, by using Alphafold2 (AF) included in the ColabFold software [87, 88]. The modelling of TDS with AF was conducted employing 100 representative structures deposited in the Alphafold Database. The sequence alignments and template generations were generated using MMseqs2 and HHsearch included in ColabFold [89, 90]. The rankest TDS suggested by AF were selected for SBD reconstruction. The SBD was generated by using DALI software [91]. The dendrogram was visualized and edited with iTOL. Representative TDSs were visualized with Discovery Studio 2020 Client [92].

Synteny analysis and assignment of the paralogous CYP genes to Secondary Metabolite Gene Clusters (SMGCs) in *C. graminicola* M1.001

AntiSMASH version1.2.2 [93], and the MycoCosm web portal (https://mycocosm.jgi.doe.gov/) were used to identify SMGCs. BLASTP was employed for manual classification of *CYP* genes into SMGCs. Synteny analysis was explored in 10 representative *CYP* families encoded in the genome of *C. graminicola*. The nearest neighbors clustered in the genomic context suggested for the Gene

Ortiz-Álvarez *et al. BMC Genomics*   (2024) 25:56

Page 17 of 19

Tool deposited in NCBI were used for the analysis. The analysis of synteny among clusters was complemented also was performed with Clinker v.0.028 software [94].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-023-09858-5.

Additional file 1: Table S1. Number of P450 cytochromes harbored in representative species belonging to *Colletotrichum* genus.

Additional file 2: Table S2. Abundance of gene copies present in each CYP family in species belonging to *Colletotrichum* graminicola complex.

Additional file 3: Table S3. Number of shared CYP families among species belonging to *Colletotrichum* Graminicola s.c.

Additional file 4: Table S4. Quantitative values corresponding to expansions and contractions occurred during the evolution of the CYPome in species belonging to *Colletotrichum* Graminicola s.c. The analysis was performed by using CAFE version 4.2.1, employing a Viterbi algorithm with a *p*-value cutoff of 0.05.

Additional file 5: Table S5. Functional assignation of CYP families. Highlight colors represent the assignation of the function in each CYP family.

Additional file 6: Table S6. Gene gains and loss for each CYP family that have occurred during the evolution of the CYPome in species belonging to the Graminicola s.c. The analysis was performed using CAFE version 4.2.1.

Authors' contributions

Conceptualization, J.O.A., M.R.T; methodology, J.O.A; S.B., analysis, J.O.A., S.B., validation of results, R.B., S.A.S., M.R.T.; writing-original draft preparation, J.O.A., S.B., writing-review and editing, R.B., S.A.S., C.H.R., M.R.T., supervision, S.A.S., M.R.T., funding S.A.S., M.R.T.

Declarations

Competing interests

The authors declare no competing interests.

References

1. Cannon PF, Damm U, Johnston PR, Weir BS. Colletotrichum - current status and future directions. Stud Mycol. 2012;73:181–213.
2. 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;27:186–98.
3. Bergstrom GC, Nicholson RL. The biology of corn anthracnose: knowledge to exploit for improved management. Plant Dis. 1999;83:596–608.
4. 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;13:414–30.
5. Talhinhas P, Sreenivasaprasad S, Neves-Martins J, Oliveira H. Molecular and phenotypic analyses reveal association of diverse Colletotrichum acutatum groups and a low level of C. gloeosporioides with olive anthracnose. Appl Environ Microbiol. 2005;71:2987–98.
6. Crouch JA, Beirn LA. Anthracnose of cereals and grasses. Fungal Divers. 2009;39:19–44.
7. Talhinhas P, Baroncelli R. Colletotrichum species and complexes: geographic distribution, host range and conservation status. Fungal Divers. 2021;110:109–98.
8. Guevara-Suarez M, Cárdenas M, Jiménez P, Afanador-Kafuri L, Restrepo S. Colletotrichum species complexes associated with crops in Northern South America: a review. Agronomy. 2022;12(3):548.
9. Swinburne T. Colletotrichum: Biology, Pathology and Control, eds J. A. Bailey & M. J. Jeger. xii 388 pp. Wallingford: CAB International (1992).£60.00 or $114.00 (hardback). ISBN 0 85198 756 7. J Agric Sci. 1993;121(1):136–7. https://doi.org/10.1017/S0021859600076929.
10. Inoue Y, Phuong Vy TT, Singkaravanit-Ogawa S, Zhang R, Yamada K, Ogawa T, et al. Selective deployment of virulence effectors correlates with host specificity in a fungal plant pathogen. New Phytol. 2023;238:1578–92.
11. 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;17:555.
12. Scharf DH, Heinekamp T, Brakhage AA. Human and plant fungal pathogens: the role of secondary metabolites. PLoS Pathog. 2014;10(1):e1003859.
13. Baroncelli R, Cobo-Díaz JF, Benocci T, Peng M, Battaglia E, Haridas S, et al. Genome evolution and transcriptome plasticity associated with adaptation to monocot and eudicot plants in *Colletotrichum* fungi. bioRxiv. 2022:2022.09.22.508453.
14. 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;44:1060–5.
15. Oliveira Silva AD, Aliyeva-Schnorr L, Wirsel SGR, Deising HB. Fungal pathogenesis-related cell wall biogenesis, with emphasis on the maize anthracnose fungus Colletotrichum graminicola. Plants. 2022;11(7):849.
16. Nelson DR. A world of cytochrome P450s. Philos Trans R Soc Lond B Biol Sci. 2013;368:20120430.
17. Werck-Reichhart D, Feyereisen R. Cytochromes P450: a success story. Genome Biol. 2000;1:REVIEWS3003.
18. Anzenbacher P, Anzenbacherová E. Cytochromes P450 and metabolism of xenobiotics. Cell Mol Life Sci. 2001;58:737–47.
19. Crešnar B, Petrič S. Cytochrome P450 enzymes in the fungal kingdom. Biochim Biophys Acta. 2011;1814:29–35.
20. Iida T, Sumita T, Ohta A, Takagi M. The cytochrome P450ALK multigene family of an n-alkane-assimilating yeast, Yarrowia lipolytica: cloning and characterization of genes coding for new CYP52 family members. Yeast Chichester Engl. 2000;16:1077–87.
21. Iwama R, Kobayashi S, Ishimaru C, Ohta A, Horiuchi H, Fukuda R. Functional roles and substrate specificities of twelve cytochromes P450 belonging to CYP52 family in n-alkane assimilating yeast Yarrowia lipolytica. Fungal Genet Biol. 2016;91:43–54.
22. Mizutani M. Impacts of diversification of cytochrome P450 on plant metabolism. Biol Pharm Bull. 2012;35:824–32.
23. Ortiz-Álvarez J, Vera-Ponce de León A, Trejo-Cerro O, Vu HT, Chávez-Camarillo G, Villa-Tanaca L, et al. Candida pseudoglaebosa and Kodamaea ohmeri are capable of degrading alkanes in the presence of heavy metals. J Basic Microbiol. 2019;59:792–806.
24. Chen W, Lee MK, Jefcoate C, Kim SC, Chen F, Yu JH. Fungal cytochrome P450 monooxygenases: their distribution, structure, functions, family expansion, and evolutionary origin. Genome Biol Evol. 2014;6:1620–34.
25. Moktali V, Park J, Fedorova-Abrams ND, Park B, Choi J, Lee Y-H, et al. Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. BMC Genomics. 2012;13:525.

26. Park J, Lee S, Choi J, Ahn K, Park B, Park J, et al. Fungal cytochrome P450 database. BMC Genomics. 2008;9:402.

27. Casadevall A. Determinants of virulence in the pathogenic fungi. Fungal Biol Rev. 2007;21:130–2.

28. Shin J, Kim J-E, Lee Y-W, Son H. Fungal cytochrome P450s and the P450 complement (CYPome) of Fusarium graminearum. Toxins. 2018;10:112.

29. Chettri P, Ehrlich KC, Cary JW, Collemare J, Cox MP, Griffiths SA, et al. Dothistromin genes at multiple separate loci are regulated by AflR. Fungal Genet Biol. 2013;51:12–20.

30. Hannemann F, Bichet A, Ewen KM, Bernhardt R. Cytochrome P450 systems–biological variations of electron transport chains. Biochim Biophys Acta. 2007;1770:330–44.

31. Siewers V, Viaud M, Jimenez-Teja D, Collado IG, Gronover CS, Pradier J-M, et al. Functional analysis of the cytochrome P450 monooxygenase gene bcbot1 of Botrytis cinerea indicates that botrydial is a strain-specific virulence factor. Mol Plant-Microbe Interact. 2005;18:602–12.

32. Zhang D-D, Wang X-Y, Chen J-Y, Kong Z-Q, Gui Y-J, Li N-Y, et al. Identification and characterization of a pathogenicity-related gene VdCYP1 from Verticillium dahliae. Sci Rep. 2016;6:27979.

33. Mellado E, Garcia-Effron G, Buitrago MJ, Alcazar-Fuoli L, Cuenca-Estrella M, Rodriguez-Tudela JL. Targeted gene disruption of the 14-alpha sterol demethylase (cyp51A) in Aspergillus fumigatus and its role in azole drug susceptibility. Antimicrob Agents Chemother. 2005;49:2536–8.

34. Sharma M, Sengupta A, Ghosh R, Agarwal G, Tarafdar A, Nagavardhini A, et al. Genome wide transcriptome profiling of Fusarium oxysporum f sp. ciceris conidial germination reveals new insights into infection-related genes. Sci Rep. 2016;6:37353.

35. Van Bogaert INA, Groeneboer S, Saerens K, Soetaert W. The role of cytochrome P450 monooxygenases in microbial fatty acid metabolism. FEBS J. 2011;278:206–21.

36. Coleman JJ, White GJ, Rodriguez-Carres M, Vanetten HD. An ABC transporter and a cytochrome P450 of Nectria haematococca MPVI are virulence factors on pea and are the major tolerance mechanisms to the phytoalexin pisatin. Mol Plant-Microbe Interact. 2011;24:368–76.

37. Lah L, Podobnik B, Novak M, Korošec B, Berne S, Vogelsang M, et al. The versatility of the fungal cytochrome P450 monooxygenase system is instrumental in xenobiotic detoxification. Mol Microbiol. 2011;81:1374–89.

38. 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;197:1236–49.

39. Liang X, Wang B, Dong Q, Li L, Rollins JA, Zhang R, et al. Pathogenic adaptations of Colletotrichum fungi revealed by genome wide gene family evolutionary analyses. PLoS One. 2018;13:e0196303.

40. Cuevas-Fernández FB, Robledo-Briones AM, Baroncelli R, Trkulja V, Thon MR, Buhinicek I, et al. First report of Colletotrichum graminicola causing maize anthracnose in Bosnia and Herzegovina. Plant Dis. 2019;103:3281.

41. Rogério F, Baroncelli R, Cuevas-Fernández FB, Becerra S, Crouch J, Bettiol W, et al. Population genomics provide insights into the global genetic structure of *Colletotrichum graminicola*, the causal agent of maize anthracnose. mBio. 2023;14:e02878–22.

42. Dewangan Y, Berdimurodov E, Verma DK. Chapter 1 - Amino acids: classification, synthesis methods, reactions, and determination. In: Verma C, Verma DK, editors. Handbook of biomolecules. Paris: Elsevier; 2023. p. 3–23.

43. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, et al. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One. 2008;3:e2300.

44. Li L, Zhu X-M, Zhang Y-R, Cai Y-Y, Wang J-Y, Liu M-Y, et al. Research on the molecular interaction mechanism between plants and pathogenic fungi. Int J Mol Sci. 2022;23(9):4658.

45. Gan P, Narusaka M, Kumakura N, Tsushima A, Takano Y, Narusaka Y, et al. 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;8:1467–81.

46. Durairaj P, Hur J-S, Yun H. Versatile biocatalysis of fungal cytochrome P450 monooxygenases. Microb Cell Factories. 2016;15:125.

47. Moraga J, Gomes W, Pinedo C, Cantoral JM, Hanson JR, Carbú M, et al. The current status on secondary metabolites produced by plant pathogenic Colletotrichum species. Phytochem Rev. 2019;18:215–39.

48. Chadha S, Mehetre ST, Bansal R, Kuo A, Aerts A, Grigoriev IV, et al. Genome-wide analysis of cytochrome P450s of Trichoderma spp.: annotation and evolutionary relationships. Fungal Biol Biotechnol. 2018;5:12.

49. Fischer GJ, Keller NP. Production of cross-kingdom oxylipins by pathogenic fungi: an update on their role in development and pathogenicity. J Microbiol Seoul Korea. 2016;54:254–64.

50. Qhanya LB, Matowane G, Chen W, Sun Y, Letsimo EM, Parvez M, et al. Genome-wide annotation and comparative analysis of cytochrome P450 monooxygenases in basidiomycete Biotrophic plant pathogens. PLoS One. 2015;10:e0142100.

51. Santos ME, Le Bouquin A, Crumière AJJ, Khila A. Taxon-restricted genes at the origin of a novel trait allowing access to a new environment. Science. 2017;358:386–90.

52. Johnson BR. Taxonomically restricted genes are fundamental to biology and evolution. Front Genet. 2018;9:407.

53. Bomblies K, Peichel CL. Genetics of adaptation. Proc Natl Acad Sci. 2022;119:e2122152119.

54. Mingot JM, Peñalva MA, Fernández-Cañón JM. Disruption of phacA, an Aspergillus nidulans gene encoding a novel cytochrome P450 monooxygenase catalyzing phenylacetate 2-hydroxylation, results in penicillin overproduction. J Biol Chem. 1999;274:14545–50.

55. Kitazume T, Takaya N, Nakayama N, Shoun H. Fusarium oxysporum fattyacid subterminal hydroxylase (CYP505) is a membrane-bound eukaryotic counterpart of Bacillus megaterium cytochrome P450BM3. J Biol Chem. 2000;275:39734–40.

56. Minerdi D, Sadeghi SJ, Pautasso L, Morra S, Aigotti R, Medana C, et al. Expression and role of CYP505A1 in pathogenicity of Fusarium oxysporum f. sp. lactucae. Biochim Biophys Acta Proteins Proteomics. 2020;1868:140268.

57. Kimura M, Tokai T, Takahashi-Ando N, Ohsato S, Fujimura M. Molecular and genetic studies of fusarium trichothecene biosynthesis: pathways, genes, and evolution. Biosci Biotechnol Biochem. 2007;71:2105–23.

58. Castell-Miller CV, Gutierrez-Gonzalez JJ, Tu ZJ, Bushley KE, Hainaut M, Henrissat B, et al. Genome assembly of the fungus Cochliobolus miyabeanus, and transcriptome analysis during early stages of infection on American Wildrice (Zizania palustris L.). PLoS One. 2016;11:e0154122.

59. Ortiz-Álvarez J, Becerra-Bracho A, Méndez-Tenorio A, Murcia-Garzón J, Villa-Tanaca L, Hernández-Rodríguez C. Phylogeny, evolution, and potential ecological relationship of cytochrome CYP52 enzymes in Saccharomycetales yeasts. Sci Rep. 2020;10:10269.

60. Christ B, Xu C, Xu M, Li F-S, Wada N, Mitchell AJ, et al. Repeated evolution of cytochrome P450-mediated spiroketal steroid biosynthesis in plants. Nat Commun. 2019;10:3206.

61. Wybouw N, Kurlovs AH, Greenhalgh R, Bryon A, Kosterlitz O, Manabe Y, et al. Convergent evolution of cytochrome P450s underlies independent origins of keto-carotenoid pigmentation in animals. Proc Biol Sci. 2019;286:20191039.

62. Bowmaker JK. Evolution of colour vision in vertebrates. Eye Lond Engl. 1998;12(Pt 3b):541–7.

63. De Grassi A, Lanave C, Saccone C. Genome duplication and gene-family evolution: the case of three OXPHOS gene families. Gene. 2008;421:1–6.

64. Diss G, Ascencio D, DeLuna A, Landry CR. Molecular mechanisms of paralogous compensation and the robustness of cellular networks. J Exp Zoolog B Mol Dev Evol. 2014;322:488–99.

65. Ghose DA, Przydzial KE, Mahoney EM, Keating AE, Laub MT. Marginal specificity in protein interactions constrains evolution of a paralogous family. Proc Natl Acad Sci. 2023;120:e2221163120.

66. Dujon B. Yeast evolutionary genomics. Nat Rev Genet. 2010;11:512–24.

67. Baroncelli R, Sanz-Martín JM, Rech GE, Sukno SA, Thon MR. Draft genome sequence of Colletotrichum sublineola, a destructive pathogen of cultivated sorghum. Genome Announc. 2014;2:e00540–e614.

68. 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.

69. Lelwala RV, Korhonen PK, Young ND, Scott JB, Ades PK, Gasser RB, et al. Comparative genome analysis indicates high evolutionary potential of pathogenicity genes in Colletotrichum tanaceti. PLoS One. 2019;14:e0212248.

70. Johnson LS, Eddy SR, Portugaly E. Hidden Markov model speed heuristic and iterative HMM search procedure. BMC Bioinformatics. 2010;11:431.

71. Ngcobo PE, Nkosi BV, Chen W, Nelson DR, Syed K. Evolution of cytochrome P450 enzymes and their redox partners in archaea. Int J Mol Sci. 2023;24(4):4161.
72. Harris KL, Thomson RES, Gumulya Y, Foley G, Carrera-Pacheco SE, Syed P, et al. Ancestral sequence reconstruction of a cytochrome P450 family involved in chemical defense reveals the functional evolution of a promiscuous, xenobiotic-metabolizing enzyme in vertebrates. Mol Biol Evol. 2022;39:msac116.
73. Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics. 2004;5:1–19.
74. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27:221–4.
75. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinforma Oxf Engl. 2011;27:1164–5.
76. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001;17:754–5.
77. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–245.
78. Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol Biol Evol. 2013;30:1987–97.
79. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238.
80. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9.
81. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: a sequence logo generator. Genome Res. 2004;14:1188–90.
82. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 1990;18:6097–100.
83. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.
84. Khan A, Mathelier A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. BMC Bioinformatics. 2017;18:287.
85. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. Bioinforma Oxf Engl. 2012;28:573–80.
86. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2016.
87. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. Nat Methods. 2022;19:679–82.
88. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. Nat Struct Mol Biol. 2022;29:1056–67.
89. Mirdita M, Steinegger M, Söding J. MMseqs2 desktop and local web server app for fast, interactive sequence searches. Bioinformatics. 2019;35:2856–8.
90. Fidler DR, Murphy SE, Courtis K, Antonoudiou P, El-Tohamy R, Ient J, et al. Using HHsearch to tackle proteins of unknown function: a pilot study with PH domains. Traffic Cph Den. 2016;17:1214–26.
91. Holm L, Laiho A, Törönen P, Salgado M. DALI shines a light on remote homologs: one hundred discoveries. Protein Sci. 2023;32:e4519.
92. Biovia DS. Discovery studio modeling environment. Release; 2017.
93. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, et al. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. Nucleic Acids Res. 2011;39 Web Server issue:339–46.
94. Gilchrist CLM, Chooi Y-H. clinker & clustermap.js: automatic generation of gene cluster comparison figures. Bioinformatics. 2021;37:2473–5.

Publisher's Note