

RESEARCH ARTICLE

Open Access



Gene family expansions and contractions are associated with host range in plant pathogens of the genus *Colletotrichum*

Riccardo Baroncelli^{1*}, Daniel Buchvaldt Amby², Antonio Zapparata³, Sabrina Sarrocco³, Giovanni Vannacci³, Gaétan Le Floch¹, Richard J. Harrison⁴, Eric Holub⁵, Serenella A. Sukno⁶, Surapareddy Sreenivasaprasad⁷ and Michael R. Thon^{6*}

Abstract

Background: Many species belonging to the genus *Colletotrichum* cause anthracnose disease on a wide range of plant species. In addition to their economic impact, the genus *Colletotrichum* is a useful model for the study of the evolution of host specificity, speciation and reproductive behaviors. Genome projects of *Colletotrichum* species have already opened a new era for studying the evolution of pathogenesis in fungi.

Results: We sequenced and annotated the genomes of four strains in the *Colletotrichum acutatum* species complex (CAsc), a clade of broad host range pathogens within the genus. The four CAsc proteomes and secretomes along with those representing an additional 13 species (six *Colletotrichum* spp. and seven other Sordariomycetes) were classified into protein families using a variety of tools. Hierarchical clustering of gene family and functional domain assignments, and phylogenetic analyses revealed lineage specific losses of carbohydrate-active enzymes (CAZymes) and proteases encoding genes in *Colletotrichum* species that have narrow host range as well as duplications of these families in the CAsc. We also found a lineage specific expansion of necrosis and ethylene-inducing peptide 1 (Nep1)-like protein (NLPs) families within the CAsc.

Conclusions: This study illustrates the plasticity of *Colletotrichum* genomes, and shows that major changes in host range are associated with relatively recent changes in gene content.

Keywords: CAZyme, Plant pathogen, Anthracnose, Fungal genomics, *Colletotrichum* spp.

Background

Plant pathogenic fungi exhibit remarkable differences in the number and diversity of hosts they are able to colonize and/or infect. Based on their host range, phytopathogenic fungi can be categorised as specialists infecting a single plant or a small group of closely related plants (narrow host range), generalists associated with a wide variety of plants in diverse environments (broad host range), and transitional species capable of infecting

a limited range of plants (intermediate host range). What is remarkable is the existence of plant pathogens manifesting these host range categories in the same phylogenetic lineage or different lineages within a single genus as exemplified by the globally important fungal genus *Colletotrichum* [1, 2]. Host range shifts are also intricately linked to speciation and are potentially driven by changes in lifestyle [2, 3]. Understanding the molecular determinants of the host range alternations has major implications in global food security including crop disease management, and control of pathogen introductions into new environments.

Colletotrichum species exhibit endophytic and/or pathogenic associations with a wide variety of herbaceous and woody plants in tropical, subtropical and temperate climates in natural and agricultural ecosystems [1, 2]. The

* Correspondence: riccardobaroncelli@gmail.com; mthon@usal.es

¹Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (LUBEM), University of Western Brittany, Technopôle Brest-Iroise, 29280 Plouzané, France

⁶Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), University of Salamanca, Campus de Villamayor, C/Del Duero, 12, 37185, Villamayor, Salamanca, Spain

Full list of author information is available at the end of the article



economic impact of crop-losses caused by *Colletotrichum* pathogens has been well recognized [1, 4]. Recent multi-locus phylogenetic studies of the genus *Colletotrichum* led to the identification of at least 10 major clades such as acutatum, gloeosporioides and boninense including at least 28, 22 and 17 species, respectively [2]. *Colletotrichum* species identified within and among these major clades or lineages exhibit remarkable differences in their host range. Within the *C. acutatum* species complex (CAsc), species such as *C. nymphaeae*, *C. simmondsii* and *C. fioriniae* display broad host range, *C. salicis* an intermediate range of woody hosts [5], and *C. lupini* a narrow host range for lupins [6, 7]. A similar pattern can be found among species belonging to the *C. gloeosporioides* and *C. boninense* species complexes. Conversely, the *C. orbiculare* complex includes species with a narrow host range [8–11]. The trajectory of evolution of specialists and generalists in *Colletotrichum* pathogens, and how this change is mirrored in the genomic architecture of various species remain to be addressed.

Since the first genome sequences of phytopathogenic fungi became available, researchers have been analyzing gene content to find associations that may explain the differences in fungal lifestyles [12] and varying patterns are beginning to emerge. Some studies have suggested that differences in gene family size are more strongly associated with phylogenetic relatedness than lifestyle [13]. In contrast, other studies have found a larger number of secreted enzymes in pathogens compared to non-pathogens, and also in necrotrophic and hemibiotrophic fungi compared to biotrophs [14–18]. These studies suggest that specific patterns of gene content may be associated with the adaptation of diverse fungal lifestyles.

In this manuscript, we report the genome sequences of four *Colletotrichum* species representing the diversity within the CAsc, and the comparative analysis with the genome sequences of species representing narrow, intermediate and broad host ranges from other major clades/lineages. We studied differences in gene content between the species by focusing our analyses on two classes of proteins known to have roles in plant - pathogen interactions: secreted proteins and enzymes responsible for secondary metabolite biosynthesis. Comparative genomics revealed contractions of gene families encoding carbohydrate active (CAZymes) and proteolytic enzymes specifically within host-specific species as well as lineage specific expansions with the CAsc. We also found an expansion of necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs) and a contraction of Lineage Specific Effector protein Candidates (LSECs) within CAsc. Based on these patterns, we hypothesize that in broad host range *Colletotrichum* species particularly within CAsc, LSECs have reduced roles in plant interactions and that they may instead rely on CAZymes, proteases and NLPs for host colonization. This study demonstrates the utility of higher

resolution sampling for comparative genomic studies in the filamentous fungal pathogens.

Methods

Evolutionary relationships in *Colletotrichum* spp.

A phylogeny of the genus *Colletotrichum* was constructed based on publicly available DNA sequences of four nuclear loci: part of the ribosomal RNA gene cluster (rRNA) (ITS1-5.8S-ITS2), beta tubulin (β TUB) partial sequence, actin (ACT) partial sequence and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) partial sequence. A complete list of the sequences used is reported in Additional file 1: Table S1. All the sequences were aligned using MAFFT 7 [19] and the multiple sequence alignments were exported to MEGA 6.06 [20] where best-fit substitution models were calculated for each separate sequence dataset. The concatenated alignment (ITS, TUB2, ACT and GAPDH) was performed with Geneious 8.1.4. A Markov Chain Monte Carlo (MCMC) algorithm was used to generate phylogenetic trees with Bayesian probabilities using MrBayes 3.2.1 [21]. Models of nucleotide substitution for each gene determined by MEGA 6.06 [20] were included for each locus. The analyses of four MCMC chains were run from random trees for 5000000 numbers of generations and sampled every 100 generations.

Fungal strains

C. simmondsii - CBS 122122 also known as BRIP 28519; HKUCC 10928; ICMP 17298; KACC 43258). This strain was collected during May of 1987 by L.M. Coates in Queensland, Australia from infected fruit tissues of papaya [sn: *Carica papaya*]. The strain has been nominated as the holotype of the species [5].

C. fioriniae [22] - IMI 504882 also known as PJ7 was isolated by Peter R. Johnston from infected strawberry [sn: *Fragaria x ananassa*] fruit in the Auckland area, New Zealand in 1988 [23]. The strain has been used as a reference strain for phylogenetic analyses of the CAsc and for mating tests and pathogenicity assays [24]. Heterothallic mating capability of this strain has been demonstrated in laboratory experiments [24].

C. nymphaeae - IMI 504889 also known as SA01 collected by T. Sundelin in June 2000 in Denmark in a production field on the island of Falster from infected strawberry plants of cv. Kimberly [25]. The strain has been used as a model to study *Colletotrichum*/strawberry interaction with an integrated 'omics approach. *C. nymphaeae* is the most wide distributed species of CAsc and is one of the most important pathogen on different crops such strawberry and olive [26, 27]. A sexual state has not been reported.

C. salicis - CBS 607.94 isolated in 1994 by H.A. van der Aa from infected leaf tissue of *Salix* sp. in the *Salix* Forest near Blocq van Kuffeler, Netherlands. Originally

epitype of *Sphaeria salicis*, now designated, culture ex-epitype for *C. salicis* [5]. Synonymous of *Glomerella miyabeana* isolates belonging to this species are homothallic [23, 28].

DNA extraction

Genomic DNA was extracted based on a modified cetyltrimethylammonium-bromide (CTAB) procedure [29]. The mycelium (250 mg) was ground under liquid nitrogen using a presterilized chilled mortar and pestle. The resultant powder was mixed with 15 ml of a preheated solution (60 °C) containing 10 % CTAB, 2 M Tris-CI (pH 8.0), 0,5 M EDTA, 1.4 M NaCl and 0,5 % 2-mercaptoethanol. After incubation for 30 min at 60 °C, proteins were removed twice with 15 ml volume of chloroformisoamyl alcohol 24:1 (v/v). The aqueous phase was transferred to a clean tube, and the nucleic acids were precipitated with 0.6 volume of cold 2-propanol. After 2-hour incubation at room temperature, the samples were centrifuged for 2 min at 460 g. The pellet was washed twice with 66 % (v/v) EtOH and 34 % of 0.1 M NaCl. Tubes were centrifuged at 1500 g for 10 min, Washing buffer (supernatant) was removed and pellet were air dried in the fume hood (approximately 1 h). The pellet was resuspended in one ml of AB, left for few minutes, centrifuged for 5 min and supernatant (DNA) saved and pellet discarded.

Genome sequencing and assembly

Fifty microliters samples of genomic DNA (20 µg/µl) in AB buffer quantified with PicoGreen on a Qubit 2.0 Fluorometer were used for library preparation and sequencing. The samples were sequenced on Illumina GAI and MiSeq instruments using different library preparation and sequencing kits (Table 1).

For the samples sequenced on the Illumina GAI, genomic libraries with an average insert size of 260 bp were constructed using TruSeq™ RNA and DNA sample preparation kits (Illumina Inc.). Library

preparation and sequencing of the 70 base, paired end libraries was carried out at the School of Life Sciences of the University of Warwick. The 50 base, paired end libraries were prepared and sequenced at the Wellcome Trust Sanger Institute. The 250 base paired end MiSeq libraries were prepared using the Nextera DNA Sample Prep Kit and sequenced at the NIAB-EMR (East Malling, Kent, UK).

The genome assemblies were performed using Velvet 1.2.10 [30] after optimizing k-mer values within the range of 21 to 69. The k-mer found to give the best results based on N50 statistics was used to produce the final assemblies (39 for *C. salicis*, 31 for *C. fioriniae* and *C. simmondsii* and 65 for *C. nymphaeae*). Only scaffolds longer than 200 bp and with k-mer coverage higher than 10X were retained in the final genomic assembly. The overall genome coverage (Table 1) was estimated using the peak on the k-mer frequency distribution curve reported by Jellyfish 2.1.3 [31]. The contigs corresponding to the mitochondrial genome (mtDNA) and the ribosomal RNA encoding gene cluster were identified by BLASTN searches using the *C. graminicola* mitochondrial genome as the query sequence. Each mitochondrial genome and rRNA cluster were assembled into one scaffold and removed from further analyses. The completeness of the assembly was assessed using BUSCO v1.2 [32].

Gene structure annotation

The MAKER2 annotation pipeline [33] was used to annotate the CAsc genome. Two different sets of assembled sequences from transcriptomic samples were used as EST evidence. The first library belonging to *C. fioriniae* conidial germination stage and available in GenBank (EST: LIBEST_024551) the second library belonging to *C. nymphaeae* SA-01 during plant interaction (kindly provided by Birgit Jensen, Department of Plant and Environmental Sciences, University of Copenhagen, Denmark).

Protein evidence included fungal proteins downloaded from UniProtKB/SwissProt (release 2013_12) (Uniprot [34]). Three different *ab initio* gene annotation programs were trained for use with MAKER2. GeneMark-ES [35] was self-trained for each of the four genomes. SNAP [36] was trained following the protocol included with the documentation using transcriptome sequences. AUGUSTUS [37] was used with gene model belonging to closely related organism *Fusarium graminearum*.

Putative functions were assigned to the annotations by using BLASTP [38] to identify homologs in a database constructed of proteins from the UniProt database (release 2013_12) from *Neurospora crassa*, *Magnaporthe oryzae*, *Fusarium oxysporum*, *Fusarium graminearum*, *Verticillium alfalfae*, *Verticillium dahliae*, *C. graminicola*, and *C. higginsianum*.

Table 1 Summary of sequence data sets generated for CAsc genome sequencing

	GAI		MiSEQ		Tot N bp	Coverage
	70 bp	50 bp	250 bp			
<i>C. fioriniae</i> IMI 504882	6.84E + 08	1.75E + 09	-		2.44E + 09	43.03
<i>C. simmondsii</i> CBS 122122	1.97E + 09	-	-		1.97E + 09	33.33
<i>C. salicis</i> CBS 607.94	2.89E + 08	6.83E + 08	1.05E + 09		2.02E + 09	36.40
<i>C. nymphaeae</i> IMI 504889	-	-	2.35E + 09		2.35E + 09	42.08

Comparative genomics

The genomes of twelve additional fungal species belonging to the Sordariomycetes, including all publicly available *Colletotrichum* spp. were also included the analyses (Table 2).

The proteomes were clustered OrthoFinder v0.4 [39] and the clusters were analysed with Mirlo (<https://github.com/mthon/mirlo>) to identify the five most phylogenetically informative single copy gene families. The five families were aligned with MAFFT 7 [19] and then concatenated. A substitution model and its parameter values were selected using ProtTest 3.4 [40]. A phylogenetic tree was reconstructed using Bayesian MCMC analysis constructed from the alignment based on the concatenated alignment under the WAG + I evolutionary model and the gamma distribution calculated using four rate categories and homogeneous rates across the tree. The posterior probabilities threshold was selected as over 75 % (Fig. 1).

Functional annotation

Prediction and analyses of secretomes

Proteins that are transported out of the cell and into the extracellular space were identified with WoLF-PSORT [41] as described previously [17]. The final set of putative secreted proteins were scanned with InterProScan [42] using RunIprScan 1.1.0 (<http://michaelrthon.com/runiprscan/>) to identify protein domain signatures and relative changes in selected genomes.

Secreted carbohydrate active enzymes (CAZymes)

We used the Hmmscan program in the HMMER 3.0 package [43] to search each of fungal predicted proteomes with the family-specific HMM profiles of CAZymes downloaded from dbCAN database 2.0 [44] with cut-off E-value of 1E-3 as queries. The primary results were processed and checked manually: overlapping matches removed and cut-off E-value adjusted based on comparison with the manually curated *C. graminicola* CAZome. For each CAZy class, the number of enzyme modules and the families they belong to are reported.

Secreted proteases

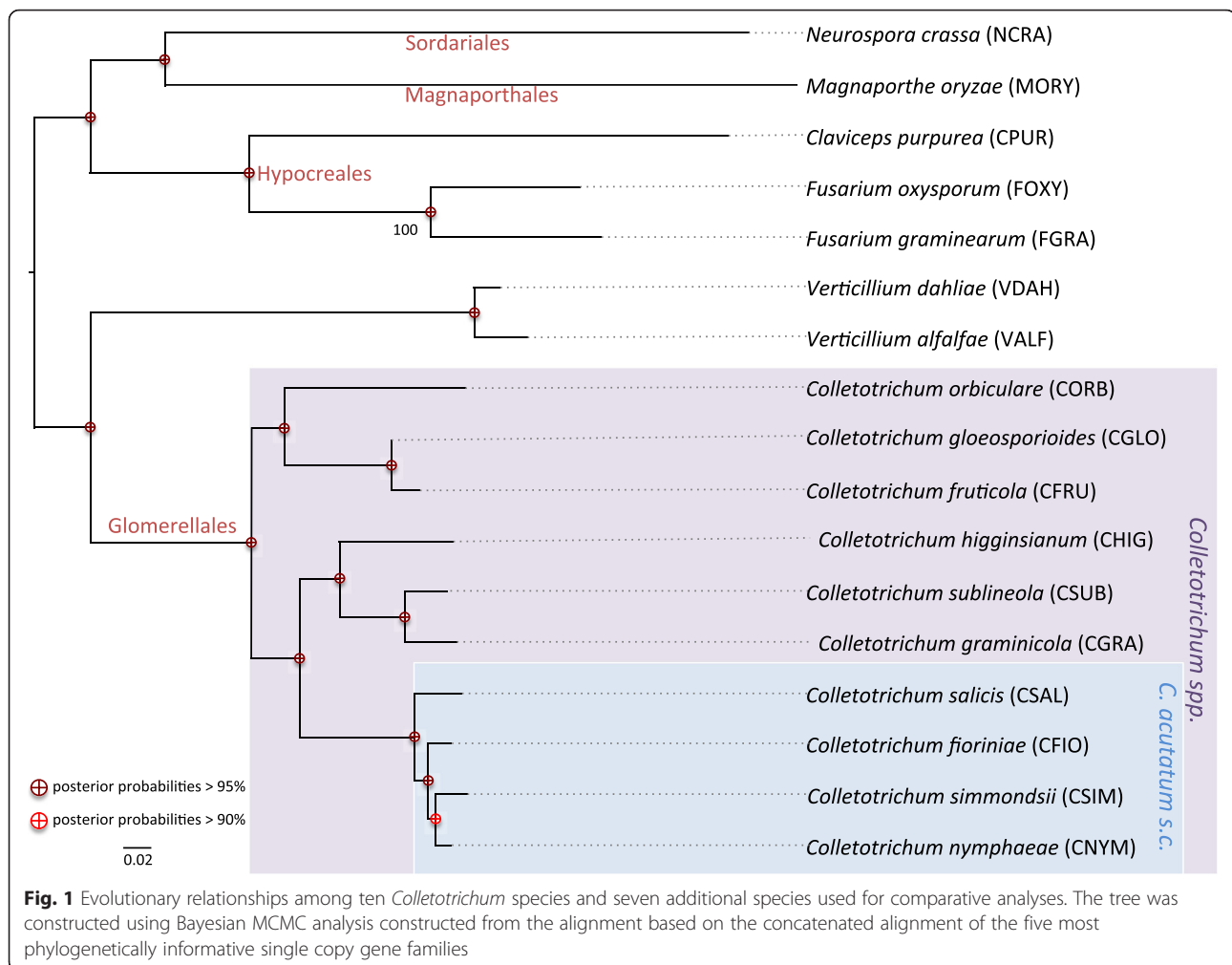
Predicted proteins were classified as proteases by querying the MEROPS database 10.0 [45] using a BLASTp cut-off E-value of 1E-5. Sequences with similarity to protease domains but with mutated active sites and incomplete protease domains were further excluded as proteases.

Necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs)

The CaNLP genes were identified in the *Colletotrichum* spp. genomes by searching a six-frame translation on the conserved GHRHDWE motif [46]. RunIprScan was also used to scan predicted secretomes to identify genes encoding NLPs (IPR008701/PF05630).

Table 2 List of species used in this study. In bold are highlighted the four *Colletotrichum acutatum* species presented in this work

Strain	Organisms	Abbr.	Host	Origin	Accession N°	Project N°	Ref.
IMI 504889	<i>C. nymphaeae</i>	CNYM	<i>Fragaria x ananassa</i>	Denmark	JEMN00000000.1	PRJNA237763	-
CBS 122122	<i>C. simmondsii</i>	CSIM	<i>Carica papaya</i>	Australia	JFBX00000000.1	PRJNA239224	-
IMI 504882	<i>C. fiorinia</i>	CFIO	<i>Fragaria x ananassa</i>	New Zealand	JARH00000000.1	PRJNA233987	[22]
CBS 607.94	<i>C. salicis</i>	CSAL	<i>Salix sp.</i>	Netherlands	JFFI00000000.1	PRJNA238477	-
M1.001	<i>C. graminicola</i>	CGRA	<i>Zea mays</i>	USA	ACOD00000000.1	PRJNA37879	[17]
TX430BB	<i>C. sublineola</i>	CSUB	<i>Sorghum bicolor</i>	USA	JMSE00000000.1	PRJNA246670	[100]
IMI 349063	<i>C. higginsianum</i>	CHIG	<i>Brassica rapa</i>	Trinidad & Tobago	CACQ00000000.2	PRJNA47061	[17]
Nara gc5	<i>C. fruticola</i>	CFRU	<i>Fragaria x ananassa</i>	Japan	ANPB00000000.1	PRJNA171218	[65]
Cg-14	<i>C. gloeosporioides</i>	CGLO	<i>Persea americana</i>	Israel	AMYD00000000.1	PRJNA176412	[101]
MAFF 240422	<i>C. orbiculare</i>	CORB	<i>Cucumis sativus</i>	Japan	AMCV00000000.1	PRJNA171217	[65]
VaMs.102	<i>V. alfalfae</i>	VALF	<i>Medicago sativa</i>	USA	ABPE00000000.1	PRJNA51263	[102]
VdLs.17	<i>V. dahliae</i>	VDAH	<i>Lactuca sativa</i>	USA	ABJE00000000.1	PRJNA28529	[102]
PH1	<i>F. graminearum</i>	FGRA	<i>Triticum sp.</i>	USA	AACM00000000.2	PRJNA13839	[103]
FOL 4287	<i>F. oxysporum</i>	FOXY	<i>Solanum lycopersicum</i>	Spain	AAXH00000000.1	PRJNA18813	[103]
20.1	<i>Cl. purpurea</i>	CPUR	<i>Secale cereale</i>	Germany	CAGA00000000.1	PRJEA76493	[47]
70-15	<i>M. oryzae</i>	MORY	<i>Oryza sativa</i>	French Guiana	AACU00000000.3	PRJNA16061	[12]
or74a	<i>N. crassa</i>	NCRA	no pathogen	USA	AABX00000000.3	PRJNA13841	[104]



Identification of lineage specific effector protein candidates (LSECs)

We defined LSECs as proteins with no (detectable) homology to any other protein. We also exclude proteins with conserved domains, as a conserved domain may imply shared ancestry. We identified LSECs by performing BLAST searches to the GenBank nr BLAST database using an e-value threshold of $1e-5$. Proteins with homology to proteins from other members of the same genus but not to other genera were termed 'genus-specific.' Those that had no homology to any other protein either within or outside of the same genus were termed 'species-specific.'

Secondary metabolites related genes and clusters

BLASTp and RunIprScan were used to manually identify genes encoding enzymes that are signatures of backbone genes and secondary metabolite (SM) gene in the Ascomycota: nonribosomal peptide synthetases (NRPS; IPR010071, IPR006163, IPR001242), polyketide synthases (PKS; IPR013968), DMATS-family aromatic prenyltransferases (IPR017795, Pfam PF11991), and terpene

synthases/cyclases (IPR008949) [47]. AntiSMASH version 1.2.2 [48] was downloaded and run locally on all genomes analysed in order to identify secondary metabolite gene clusters. The predicted clusters were visually evaluated for conservation of synteny by examining whether similar block of genes related to SM biosynthesis are orthologs in the other species. In cases where a gene in the species in which the cluster was identified no longer had orthologs in the other species, we inferred a break in synteny.

Results

Phylogenetic analyses

To better understand the evolutionary relationships among species within the genus *Colletotrichum*, we performed a phylogenetic analysis using four loci obtained from 133 isolates (Additional file 2: Figure S1). It is now well established that the genus *Colletotrichum* consists of ten major monophyletic clades that are referred to as species complexes [2, 49], all of which are present and well supported by our phylogenetic analysis. Although species complexes have no taxonomic value, they do

provide a useful summary of knowledge to link evolutionarily related species within a genus.

The number of clades may increase as several other groups have been identified in the analyses and existing clade with very few species have already been described. For example, the truncatum clade includes one major species (*C. truncatum*) and two poorly-known species (*C. curcuma* and *C. jasminigenum*); the second species has been removed from the analyses because the only existing isolate seems to be originated by hybridization event between one isolate of *C. truncatum* and one belonging to the CGsc (based on data available on GenBank).

The CAsc and CGsc have the highest number of species among the species complexes of *Colletotrichum* and they appear to be evolutionary very distant despite the similarities in morphology and host range. Both of these species complexes contain broad host range species but representative genome sequences are only available for the CGsc. Therefore, we selected four isolates from the CAsc for whole genome shotgun sequencing. The four isolates chosen represent the genetic diversity of the CAsc and are also commonly used in research laboratories as references for evolutionary analyses, phylogenetics and pathogenicity assays.

Genome sequencing, assembly and gene prediction

Genome sequencing was performed with the Illumina Genome Analyzer Ix or MiSeq sequencers and assembled into scaffolds with Velvet 1.2.10 [30]. Contigs

representing the mitochondrial genomes were identified with BLAST and removed from the assemblies. The nuclear genome assembly sizes ranged from 48 to 50 Mb (Table 3), which are comparable to the other sequenced *Colletotrichum* genomes. Benchmarking Universal Single-Copy Orthologs (BUSCO v. 1.2; [32]) was used to provide an estimate of assembly completeness. According to this analysis, the assemblies for the CAsc genomes cover from 99.79 to 99.93 % of the total gene space, which is comparable to other sequenced fungal genomes (Table 3). Gene models were annotated with MAKER2. Secreted proteins were identified using WoLF-PSORT [41] (Table 3).

Genome-wide analysis of gene content

We scanned the annotated proteins of the CAsc genomes with RunIprScan to identify conserved functional domains and families described in the Interpro database. We compared their InterPro (IPR) domain content to those of other *Colletotrichum* species and a representative set of closely related fungi.

We performed hierarchical clustering of the IPR terms of each species (Additional file 3: Table S2) to identify genome-wide patterns of functional domain content. Hierarchical clustering of the CAsc isolates with 13 other fungi belonging to *Colletotrichum* and other spp. revealed that the CAsc isolates clustered with the two CGsc isolates. This result was unexpected since it has been previously shown that gene family content in fungi is more closely associated with phylogeny than lifestyle

Table 3 Assembly and gene prediction information of *Colletotrichum* spp. genomes and of other fungi chosen as out-group

Abbr.	N° Scaffolds	Assembly Size	max contig	N50	N90	BUSCO complete	BUSCO partial	Proteome size	Secretome size	% of secreted proteins
CNYM	1884	50	494648	91051	19816	98.54 %	99.79 %	14404	2244	15.58 %
CSIM	929	50	802711	292136	87006	99.17 %	99.93 %	13884	2211	15.92 %
CFIO	1096	49	596408	137254	38253	99.10 %	99.93 %	13759	2200	15.99 %
CSAL	2776	48	217493	46166	10883	98.75 %	99.79 %	13783	2126	15.42 %
CGRA	654	52	1824042	579194	37593	99.17 %	99.86 %	12006	1650	13.74 %
CSUB	1625	47	423147	70717	13454	99.03 %	99.86 %	12699	1820	14.33 %
CHIG	10235	49	49362	6147	2360	87.48 %	98.40 %	16172	2136	13.21 %
CFRU	1241	56	493961	112809	28004	94.92 %	99.37 %	15463	2356	15.24 %
CGLO	4537	53	128447	25337	7246	99.30 %	97.43 %	15736	2376	15.10 %
CORB	526	91	2513218	449288	92779	99.03 %	99.86 %	13479	2149	15.94 %
VALF	26	33	4782674	2315232	1129871	84.49 %	96.31 %	10220	1310	12.82 %
VDAAH	52	34	2667998	1273651	499255	96.52 %	99.10 %	10535	1383	13.13 %
FGRA	31	36	8931406	5350016	2732284	98.75 %	99.79 %	13321	1542	11.58 %
FOXY	114	61	4358238	1976106	500264	97.91 %	99.51 %	17701	1935	10.93 %
CPUR	191	32	977986	433221	110809	98.96 %	99.44 %	8823	871	9.87 %
MORY	30	42	4429722	2890137	851181	98.96 %	99.65 %	11054	1832	16.57 %
NCRA	20	41	9798893	6000761	4218384	98.96 %	99.86 %	9907	1040	10.50 %

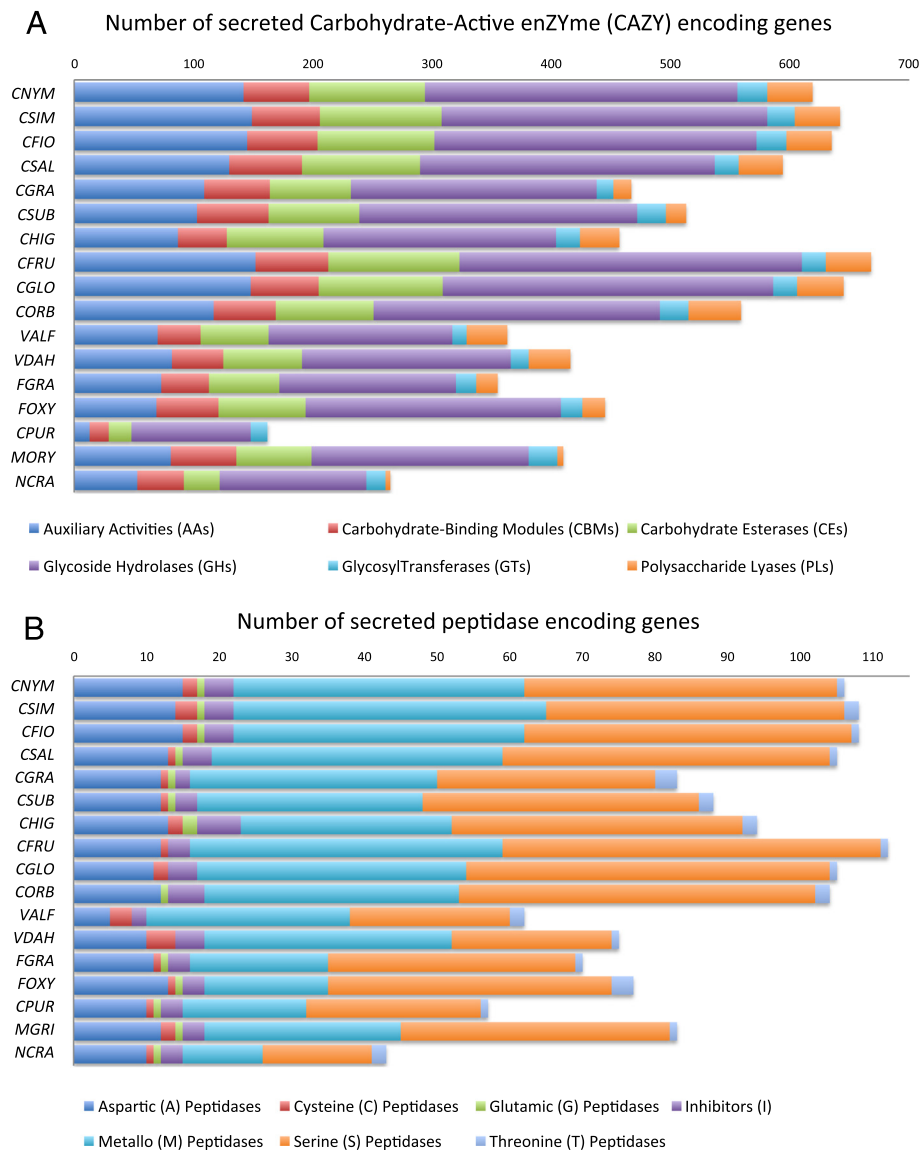


Fig. 3 a Distribution of secreted enzymes belonging to each CAZY (Carbohydrate Active enZYmes) class identified in the genomes used in this study. The legend on the bottom reports the designation of enzyme classes: Glycoside Hydrolases (GH), Glycosyl Transferases (GT), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), Auxiliary Activities (AA). The legend on the bottom reports the designation of enzyme classes: Glycoside Hydrolases (GH), Glycosyl Transferases (GT), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), Auxiliary Activities (AA). **b** Extracellular secreted protease homologs classified according to the MEROPS database 10.0 [45] in *Colletotrichum* and other fungal genomes used in this study

families in CAsc and CGsc species (Additional file 5: Figure S2). Based on this new heatmap, we selected the six families showing the largest number of genes in CAsc and CGsc for phylogenetic analysis (Additional file 5: Figures S8–S13). Each gene tree can be divided into clades that reflect the phylogenetic relationships of the species (red for CAsc and blue for other *Colletotrichum* spp., Additional file 5: Figures S8–S13), indicating that in many cases, the gene duplications leading to the CAZyme content present in these fungi preceded the speciation events. Many of the clades that

contain CAsc gene copies lack one or more other *Colletotrichum* species, indicating gene loss in these species.

All of the fungal genomes studied here encode similar numbers of glycosyltransferases (GTs), which are genes that are involved in basal activities of the fungal cell. These results are consistent with those of O’Connell et al. [17]. The families with the highest copy numbers in members of CAsc and CGsc compared to the others are genes encoding carbohydrate esterases (CEs) that catalyze the de-O or de-N-acylation of substituted

saccharides, genes encoding enzymes that hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety (GHs). The large reserve of sugar-cleaving enzymes is further extended by polysaccharide lyases (PLs) enzymes encoded by the *Colletotrichum* species belonging to the CAsc and CGsc. PLs are more overrepresented in those pathogens capable of infecting dicotyledonous such as members of CAsc and CGsc as well as *C. higginsiaum* and *C. orbiculare* compared to the monocotyledonous pathogens *C. graminicola* and *C. sublineola*.

Among the most expanded gene families in CAsc and CGsc are those involved in the degradation of pectin such as GH53, GH78, GH28, GH105, CE8, CE12, PL3 and PL1 and those involved in degradation of xyloglucan and cellulose such as GH1, GH12 and GH7 (Additional file 5: Figure S2). Families encoding xylan and pectin degrading enzymes such as GH39 and GH43 are also exhibiting higher copy numbers. This last family shows on average twice the number of genes in CAsc and CGsc compared to other organisms analyzed. A different distribution is found in the carbohydrate-binding modules family 50 (CBM50 or LysM). The number of LysM motif containing proteins varies considerably in closely related species. The number of LysM encoding genes is between 7 and 13 in the CAsc and 8 and 16 in the two genomes of CGsc.

Auxiliary Activities (AA) is another class of genes not belonging to carbohydrate-active enzymes but instead linked to biomass degradation and microbe-plant interaction (e.g., involved in lignin breakdown). All *Colletotrichum* species analyzed encode a large diversity of AA genes. Also in this case species belonging to the CAsc and CGsc encode the highest number of genes belonging to this class. The AA class is a widespread group of catalytic modules involved in plant cell wall degradation and a former classification dedicated to fungal ligninolytic enzymes [52]. Particularly overrepresented are the AA1, AA7 and AA3 families (Additional file 5: Figure S2). The categorized AA1 enzymes are multicopper oxidases that use diphenols and associated constituents as donors with oxygen as the acceptor. AA3 is the glucose-methanol-choline (GMC) oxidoreductases family and are flavoproteins containing a flavin-adenine dinucleotide (FAD)-binding domain, proteins belonging to this sub-family act as cellobiose dehydrogenases, aryl alcohol oxidase, glucose 1-oxidase, alcohol oxidase and pyranose 2-oxidase. AA7 are glucooligosaccharide oxidases that oxidize the reducing end glycosyl residues of oligosaccharides linked by alpha- or beta-1,4 bonds and glucose. Identified AA7 enzymes are potentially implicated in the biotransformation or detoxification of lignocellulosic compounds [52]. The large number of AA genes encoded by species belonging to the CAsc and

CGsc reflect the high ecological diversity of those organisms compared to the other species, in fact CAsc and CGsc have an impressive wide range of hosts that included trees.

Secreted proteases

Several families of proteases were also identified as having higher copy number in the genome wide heat map analysis of Fig. 2. Therefore, to further investigate the differences in protease content between species, we curated protease gene families by performing BLASTP searches to the MEROPS database (Additional file 6: Table S4). In the CAsc genomes, approx. 28 % of putative secreted proteins were assigned to protease encoding gene families (Fig. 3b). The genomes of CAsc, CGsc and *C. orbiculare* contain the highest number of secreted peptidases compared to the other *Colletotrichum* species and other spp. (Additional file 5: Figure S3).

Among the largest protease gene families in CAsc are those showing similarities with aspartic, metallo and serine peptidase families A01, M35 and S10 (Additional file 5: Figure S3) which are characterized (along with C13, G01, and S09) by the ability to efficiently digest proteins in inhospitable environments such as the extracellular matrix [53]. Family M43 cytophagalyzin zinc-dependent metallo-protease was also among the most overrepresented; however, very little information is available concerning a protective role of these genes in plant pathogens. We constructed phylogenetic trees for these families (Additional file 5: Figures S4–S7) and, as is the case with the CAZy families, we found that these families typically contain evidence of lineage specific duplications in the CAsc and CGsc species, as well as evidence of extensive gene loss in the other species.

Colletotrichum pathogens, including members of the CAsc, alkalize the host tissue during their infection [54, 55]. Different protease families often favor different pH ranges. The two most overrepresented, metallo and serine peptidases (Fig. 3b), are most often considered highly active under alkaline or neutral conditions. Interestingly, secreted serine proteases have been shown to play a central role in both pathogenic and mutualistic relationships and are putatively involved in cell wall degradation [56–58]. However, aspartic proteases make up the third most expanded protease gene family, which are optimally active under acidic conditions such as the environments of the most common fruits it infects (e.g. strawberry, blueberry, apple, citrus, etc.). These data suggest that CAsc are equipped with proteases for various pH environmental conditions and perhaps important to certain stages of the infection.

All of the CAsc genomes contain a member of the plant-like subtilisin family (CFIO01_03718; CNYMO1_02854; CSAL01_01351; CSIM01_08242) described by Armijos Jaramillo et al. [59, 60] as originating by a

horizontal gene transfer event (HGT) and postulated to have a role in pathogenicity in the *C. graminicola*/maize interaction. This evidence reinforces the hypothesis that the HGT event occurred in an ancestor of the *Colletotrichum* lineage and was subsequently maintained, though natural selection in all extant species.

Necrosis and ethylene-inducing peptide 1 (Nep1)-like proteins (NLPs)

NLPs trigger leaf necrosis and immunity associated responses exclusively in dicotyledonous plants [61]. NLPs are effectors that boost pathogen virulence during host colonization by disintegration of the plasma membrane. We identified NLPs by searching for the conserved motif GHRHDWE [61] in a six-frame translation of the genome sequences. In addition, we searched for proteins with the IPR term IPR008701. NLPs were particularly overrepresented in the CAsc and all four species have nearly twice as many NLPs as the other *Colletotrichum* spp. and other species included in this study. A phylogenetic reconstruction shows that most of the *Colletotrichum* spp. NLPs form a single clade that shows a rapid expansion of the gene family within *Colletotrichum* and three lineages of CAsc specific gene duplications (Fig. 4).

Lineage Specific Effector Candidates (LSECs)

LSECs were identified as secreted proteins that have no homology to any other protein (species-specific) or that have homology to proteins from other members of the same genus (genus-specific). The predicted LSECs (Fig. 5) share properties that are consistent with known fungal protein effectors. They are small proteins, having an average length of 191 amino acids over all of the species analyzed whereas the average length of all proteins is 460 amino acids. In addition, they are cysteine-rich, with cysteine making up 2.43 % of the amino acid content. In LSECs of less than 200 amino acids, the cysteine content is even more pronounced, with those proteins having 3.38 % cysteine content and LSECs less than 100 amino acids in length having 5.6 %. An average of 174 LSECs were identified in the CAsc genomes, of those 48 have been identified as species-specific and 126 as genus-specific. Variation within the species complex is evident with *C. fioriniae* encoding the highest number: 201 LSECs (128 genus-specific and 73 species-specific) and *C. salicis* the lowest number: 156 LSECs (119 genus-specific and 37 species-specific).

The definitions of species-specific and genus-specific LSECs are sensitive to the presence or absence of closely related species in the GenBank as is demonstrated by the lack of genus-specific LSECs in *Magnaporthe oryzae* and *Claviceps purpurea* (Fig. 5). No additional annotated genomes of *Magnaporthe* or *Claviceps* spp. were available in GenBank at the time this study was completed.

However, since all member of the same genus have approximately the same evolutionary distance to other genera, we reasoned that within species comparisons of the overall numbers of LSECs among different *Colletotrichum* spp. should reflect their relative importance of LSECs in each species. The LSECs content of the CAsc genomes are comparable to the content of the *C. graminicola* (150) and *C. sublineola* (182) genomes but are much lower than *C. higginsianum* (333), the *Colletotrichum* genome with the highest number of LSECs.

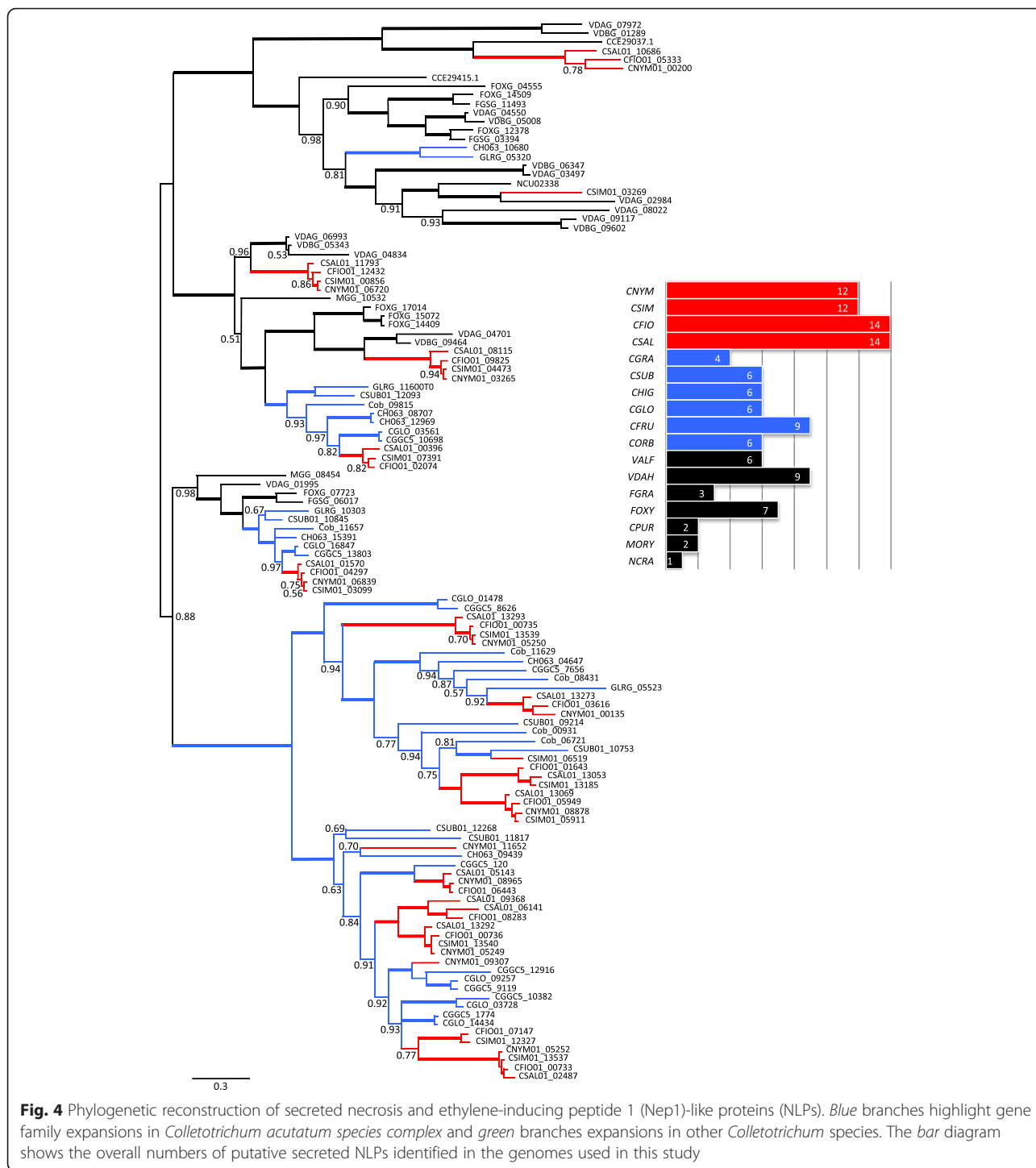
Secondary metabolite synthesis capacity

Fungi produce an enormous array of secondary metabolites, which may serve as signalling molecules and toxins against microorganisms (antimicrobials), plants (phyto-toxins) or animals and humans (mycotoxins) [62, 63]. Precursor genes are required for the biosynthesis of SMs and are often located inside fungal gene clusters. In addition, tailoring enzymes such as cytochrome P450 monooxygenases (P450s), also located inside the gene clusters, form the final products [64]. *Colletotrichum* spp. have a large diversity of putative SM-related genes (Fig. 6a and b) and biosynthetic gene clusters (Fig. 6c and Additional file 7: Table S5) compared to other plant pathogenic fungi and thus hold the genetic potential to generate diverse SMs. The P450s are overrepresented in several *Colletotrichum* spp. (Fig. 6b).

Putative polyketide synthase (PKS) clusters and “backbone” genes within *Colletotrichum* spp. and *M. oryzae* are found in substantially higher numbers than in other ascomycetes (Fig. 6a). The PKS genes of *Colletotrichum* spp. are highly active and important to appressorium-mediated host penetration [17, 65–67]. The type 1 PKSs commonly responsible for the biosynthesis of macrolides are one of most abundant clusters within CAsc. and other *Colletotrichum* spp.

The number of terpenoid synthases and relative gene clusters in members of the CAsc are comparable to the other *Colletotrichum* spp. (Fig. 6a). Neither natural products nor hybrid gene clusters with terpenoid and NRPS origin are commonly seen in nature, but are predicted in the anti-SMASH analysis and only in *C. nymphaeae* and *C. simondsii*. In addition, the more common hybrid gene cluster t1PKS-terpene (e.g. meroterpenoids) was found in 7 out of 10 *Colletotrichum* spp. genomes. The terpenoid clusters and their PKS hybrids are potential candidates for synthesizing respectively, antimicrobial triterpenoids (ergosterols and derivatives) and phytotoxic meroterpenoids (i.e. colletotrichin and derivatives) by different *Colletotrichum* spp. [68–70].

Higher numbers of non-ribosomal peptide synthase (NRPS) genes and clusters are found within species of the genus *Colletotrichum* (Fig. 6a). Candidate genes and clusters potentially involved in the production of



phytotoxic ferricrocin (e.g. siderophore) [71] and different diketopiperazines with phytotoxic and antimicrobial activities [72, 73] secreted by *C. gloeosporioides* are found within *Colletotrichum* spp. genomes. The non-ribosomal peptide mycosporine biosynthetic cluster has been characterized in fungi as well as other organisms [74]. Although *Colletotrichum* spp. have been shown to produce

mycosporine with antimicrobial and phytotoxic activities [75, 76] no mycosporine biosynthetic clusters were identified in *Colletotrichum* spp. genomes in this work.

The precursor dimethylallyl tryptophan synthases (DMATS) within *Colletotrichum* spp. are found in lower numbers compared to the other precursors (Fig. 6a) but are overrepresented in *Colletotrichum* spp.

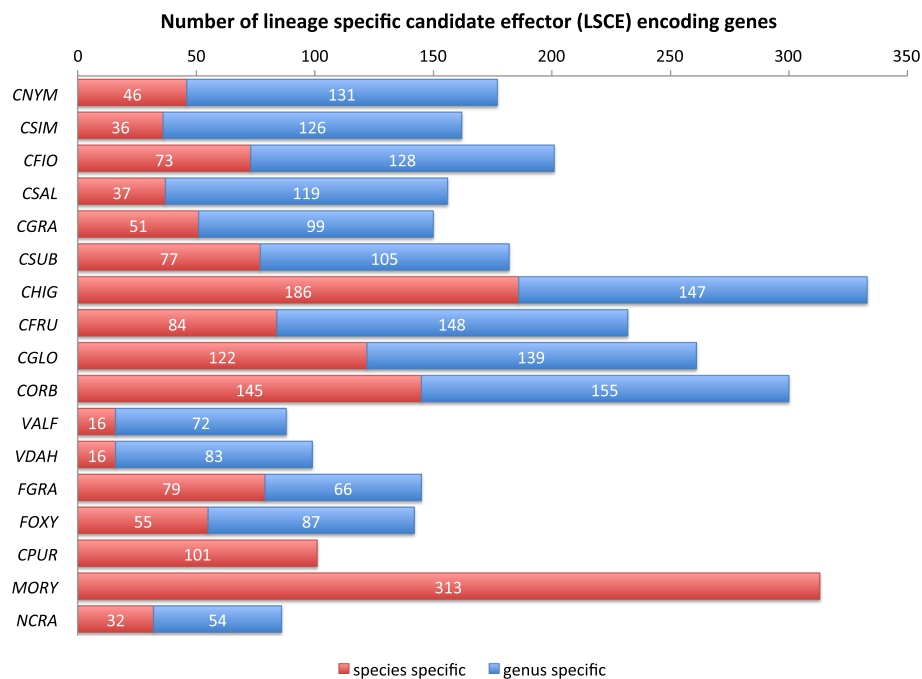


Fig. 5 Lineage Specific Effector protein Candidates (LSECs) identified in *Colletotrichum* and other fungal genomes used in this study, based on no-BLAST sequence similarity with proteins predicted in other species (red) or other genera (green)

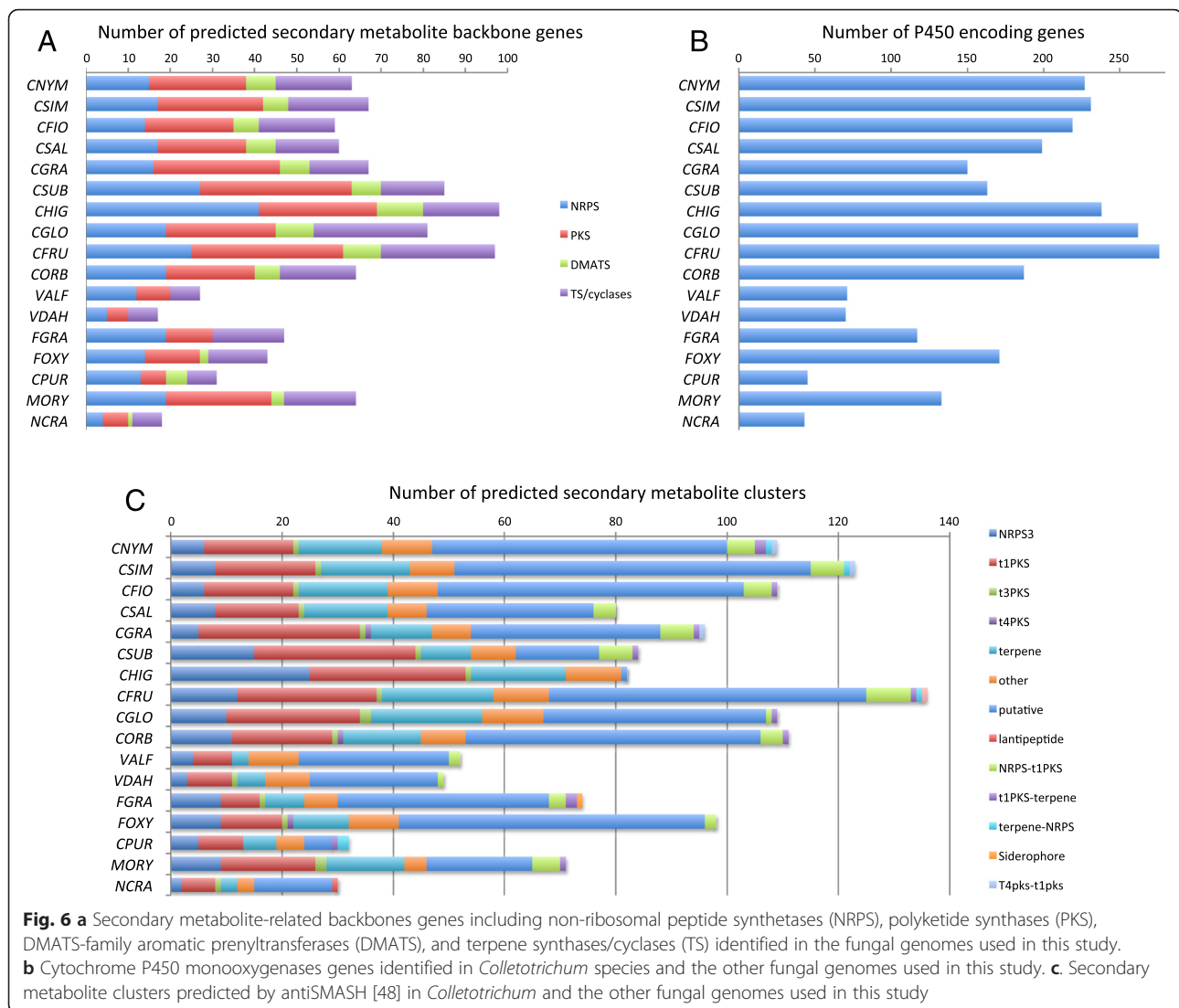
Discussion

The availability of genome sequences of six species of *Colletotrichum* with differing host ranges and belonging to five different species complexes provided an opportunity to study the evolution of host range in the genus. We sequenced the genomes of four species belonging to the CAsc, a group of typically broad host range pathogens and compared gene content among all of the genomes. In order to investigate changes in gene content between host-specific and broad host range pathogens we focus our study on two classes of molecules known to have a direct role in fungus/plant interaction. We characterized genes involved in secondary metabolite biosynthesis and genes encoding secreted proteins. Generally, pathogens characterized by a hemibiotrophic lifestyle such as *Colletotrichum* spp. and *Magnaporthe oryzae*, secrete a higher percentage of proteins extracellularly [18, 77]. Hierarchical clustering of gene family and functional domain assignments revealed overrepresentation of CAZy and protease gene families with both the CAsc and CGsc with respect to the other fungal genomes analyzed. Members of these species complexes are broad host range pathogens, suggesting that the higher number in CAZy and protease diversity may be associated with the ability to infect multiple host species. Hierarchical clustering of the 13 studied species resulted in a cluster comprised of CAsc and CGsc species, two distantly related species complexes. This result highlights the similarity in both secretomes and whole proteomes of these species complexes and suggests that their

gene family content, especially their repertoires of CAZymes and peptidases are the product of recent, lineage specific expansions of these families independently in each species complex. Interestingly, phylogenetic analyses of the CAZyme and peptidase families revealed that, in contrast to our expectations, gene loss in other *Colletotrichum* species is as important, if not more important force driving the evolution of gene family size.

A recent work focused on *Colletotrichum* comparative genome analyses [78] reported that the content secreted proteins, CAZy and proteases of members from the CAsc and CGsc were more similar to each other despite their phylogenetic distance. Analyses carried out by these authors reveal that *Colletotrichum* species have tailored profiles of their carbohydrate-degrading enzymes according to their infection lifestyles. Phylogenetic analyses revealed lineage-specific expansions of GH43 and S10 members within the CAsc and CGsc, with duplications of specific genes within the respective lineages. However, close examination of the phylogenies of Gan et al. [78] reveals gene losses in lineages outside of the CAsc and CGsc, providing further support of the view that gene loss, in addition to gene duplication is an important component of the adaptation of gene families.

The comparative analysis of fungal secretomes reported by Zhao et al. [14] revealed that plant pathogens typically have higher numbers of CAZymes than non-pathogens and that biotrophic fungi have a lower diversity of CAZymes than do necrotrophs and hemibiotrophs. The



same work showed that there were similar fractions of each CAZyme class found in pathogens that infect similar hosts, such as monocots or dicots. This evidence confirms what has been revealed by the genome comparison between *C. graminicola* and *C. higginsianum* by O'Connell et al. [17]. Comparative analyses of 18 fungi with diverse pathogenic lifestyles and strategies [56] revealed that the overall distribution of genes encoding CAZyme involved in plant cell wall breakdown reflect the different taxonomic groups, and that it may in fact reflect differences in cellulose, xylan and pectin of the host plants. A more recent work focused on comparative analysis of 79 fungal genomes [79] revealed that more CAZymes were found to be specific to each species compared with the Magnaporthaceae specific clusters suggesting that CAZyme gene sequences are plastic and may contribute to speciation. Furthermore the authors hypothesize that in the Magnaporthaceae CAZymes may vary based on route of infection rather than the type of

host plant [79]. These results are consistent with the idea that different lifestyles, hosts and host tissues present different types of carbohydrate substrates to the pathogen this is reflected by each species' CAZyme repertoire. However to investigate the evolutionary processes associated with gene expansions/losses research needs to focus on closely related species (or isolates belonging to the same species) with different lifestyles, hosts and host specificities.

The CAsc and CGsc genomes have the highest number of fungal CAZymes involved in the degradation of plant cell walls, such as xyloglucan, xylan, pectin and cellulose [80, 81]. CAsc and CGsc are also characterized by the broadest arsenal of secreted peptidases such as metallo and serine peptidases, both families with known roles in plant pathogenesis [82–84]. These proteases may contribute to the degradation of the plant cell wall by targeting structural proteins. Changes in gene families

encoding for degradative enzymes reflect the evolutionary adaptation of species complexes to different hosts and niches and open new opportunities for the identification of novel genes for industrial purposes.

Examination of protein domain and family content using InterPro revealed changes in NLPs with CAsc carrying twice the number of the genes compared to other fungi. NLPs trigger leaf necrosis and immunity associated responses exclusively in dicotyledonous plants [85]. In *C. higginsianum* the cell death inducer ChNLP1 is upregulated during the transition from biotrophy to necrotrophy. Intriguingly, Kleemann et al. [86] have also found some effectors (ChECs) that suppressed cell death induced by ChNLP1 during the initial biotrophic phase, in order to maintain cell viability. The remarkable repertoire of NLPs found in the CAsc might reflect their ability to infect a wide range of hosts. Some of these proteins may function as necrosis inducing proteins, whereas other NLPs may have roles in overcoming early plant defense responses [86].

Many NLPs have been characterized; however, only a few from the genomes used in this study have been shown to induce necrosis in plant. Since most of the previous work has focused on NLPs characterization under laboratory conditions, the hypothesis is that they might have different role in different hosts and/or they might act at different conditions providing “flexibility” to the pathogen. Considering the high number of NLPs encoded by CAsc it could be used as model system to study NLP evolution and their biological role. This study also revealed lineage specific contractions of LSECs and expansions of NLP families within the CAsc, LSECs are important for suppressing the host immune system, enabling pathogens to colonize host tissues [87–89]. Both *C. graminicola* and *C. sublineola* also have reduced numbers of LSECs but the expansion of NLPs is only found in the CAsc indicating that it may be a lineage specific innovation that is unique to members of the CAsc.

Colletotrichum spp. have a large arsenal of putative SM-related backbone genes, clusters and tailoring enzymes compared to other plant pathogenic fungi and thus hold the genetic potential to generate diverse SMs. P450s play central roles in the detoxification of plant-derived antimicrobials [90] and in the biotransformation of industrial-related products [64]. The large diversity of P450s identified in *Colletotrichum* spp. may explain how these pathogens detoxify phytoalexins [91] and why these fungi have been investigated for their capacity to biotransform various products (e.g. terpenoids) [92]. The PKS genes and clusters are most abundant within *Colletotrichum* spp. PKSs are important to host penetration and are commonly responsible for macrolides biosynthesis. This is of great interest, as *Colletotrichum* spp. can synthesize various macrolides (e.g. monocillins I-III,

monoorden, colletodiol, etc.) with phytotoxic and antimicrobial activities including novel macrolides produced by CAsc [93–96]. DMATS are overrepresented in *Colletotrichum* spp. and in those pathogens characterized by a biotrophic stage. Genome-wide expression profiles of *Colletotrichum* spp. reveals low activity of DMATS [17, 67] with few highly active of *C. orbiculare* during plant infection [65]. Secretion of phytotoxic alkaloids by *Colletotrichum* spp. with importance to plant infection has not been found, but identified gene cluster in genome sequencing are potential candidates to different compounds with anticancer and antimicrobial activities among others found within this genus [97–99].

Earlier studies concluded that gene family content in fungi depends more on evolutionary lineage, and suggested that lifestyle is less important in driving changes in gene family size [13]. In this study we focused on one genus and compared several species belonging to species complexes that differ in host range. We compared the gene content of several members of the genus *Colletotrichum* and found that changes in several gene families, particularly CAZymes and proteases are associated with two distantly related lineages of *Colletotrichum* spp. that have broad host range.

Conclusions

In this manuscript, we provide an analysis of adaptations in gene content that are associated with broad host range in *Colletotrichum* spp. This study illustrates the plasticity of fungal genomes, and shows that relatively recent changes in gene content are associated with major changes in host range. It is intriguing to speculate a cause-effect relationship between a decrease in CAZyme and protease diversity and decrease in host range. Future studies of additional taxa both within and outside of the genus *Colletotrichum* will show whether lifestyle changes drive gene duplication and loss in all fungi and whether CAZymes and peptidases are the key families controlling host range. This study also demonstrates the need for higher resolution taxonomic sampling in order to better understand the role of gene duplication and loss in the evolution of fungal genomes and the possible association with biological characters.

Additional files

Additional file 1: Table S1. List of *Colletotrichum* strains used for the phylogeny shown in Additional file 2: Figure S1, indicating their species designation, complex, strain ID and Genbank accession numbers of the sequences retrieved and used for phylogenetic analyses. Strains included in genome sequence projects are highlighted in bold. (XLSX 67 kb)

Additional file 2: Figure S1. Phylogenetic analysis of the 133 *Colletotrichum* spp. listed in Additional file 1: Table S1 based on a multilocus concatenated alignment of the ITS, TUB2, ACT and GAPDH genes. A Markov Chain Monte Carlo (MCMC) algorithm was used to generate phylogenetic trees with Bayesian probabilities using MrBayes 3.2.1. The species complexes

described by Cannon et al. [2] and Liu et al. [49] as well as *C. acutatum* species complex clades described by Damm et al. [5] are shown. *Monilochaetes infuscans* was used as an outgroup. (PDF 593 kb)

Additional file 3: Table S2. InterPro functional domains associated with the proteomes and secretomes of fungal genomes used in this study. (XLSX 645 kb)

Additional file 4: Table S3. Number of secreted CAZyme modules determined according to the CAZy database (<http://www.cazy.org/>) [50, 52]. (XLSX 21 kb)

Additional file 5: Figures S2–S13. Figures S2 and S3. Clustering of secreted carbohydrate-active enzyme (S2) and peptidase (S3) families encoding genes identified in this study. Numbers of genes in each row were normalized using MeV 4.8.1. Hierarchical clustering of genes and species was performed and visualized using the package “pheatmap” 1.0.8 within R. Figures S4–S13. Phylogenetic trees of secreted proteins belonging to specific class of peptidases (Figure S4. A01A; Figure S5. S10; Figure S6. M43B; Figure S7. M35) and carbohydrate-active (Figure S8. AA3; Figure S9. AA7; Figure S10. CE10; Figure S11. CE16; Figure S12. GH5, Figure S13. GH43) enzyme families identified in the genomes analyzed in this study. Proteins were aligned with MAFFT and trees were inferred using the FastTree algorithm implemented in Geneious 8.1.4. Protein names in red are from CAsc species, Protein names in blue are from other *Colletotrichum* spp. (PDF 8509 kb)

Additional file 6: Table S4. Numbers of extracellular secreted protease homologs classified according to the MEROPS database 10.0 [45] in *Colletotrichum* and other fungal genomes used in this study. (XLSX 14 kb)

Additional file 7: Table S5. The secondary metabolite biosynthesis backbone genes and clusters in the four sequenced *Colletotrichum acutatum* species. Clusters were predicted by AntiSMASH version 1.1.2 [48]. BLASTp and RunPrnScan were used to manually identify: nonribosomal peptide synthetases (NRPS; IPR010071, IPR006163, IPR001242), polyketide synthetases (PKS; IPR013968), DMATS-family aromatic prenyltransferases (DMATS: IPR017795, Pfam PF11991), and terpene synthases/cyclases (TS: IPR008949) [47]. (XLSX 766 kb)

Acknowledgements

This research was supported by grants AGL2012-34139 and AGL2015-663 62-R. MINECO, Spain and partially from “The Prograilive” project (grant: RBRE160116CR0530019) funded by the regions of Bretagne and Pays de la Loire and FEADER grants, France. R.B was supported in part by a Senior fellowship from the British Society of Plant Pathology (BSPP). The authors would like to thank Giovanni Cafà and Matthew Ryan from The Centre for Agriculture and Bioscience International (CABI) for depositing the strains in the CABI Genetic Resource Collection (GRC)

Availability of data and materials

The genome sequences and the raw data of the four members of the *Colletotrichum acutatum* species complex are available in GenBank under the following accession numbers: *C. simmondsii*: JFBX00000000; SRP074810, *C. nymphaeae*: JEMN00000000; SRP074816, *C. fiorinia*: JARH00000000; SRP074685, and *C. salicis*: JFFI00000000; SRP074780. All phylogenetic trees presented in this work are available as nexus files in figshare: <https://dx.doi.org/10.6084/m9.figshare.3471782>.

Authors' contributions

Conceived and designed the experiments: MRT, RB, SAS, SSP (Surapreddy Sreenivasaprasad). Performed the experiments: RB, MRT. Analyzed the data: RB, MRT, DBA, AZ, SS, GV, GLF, RH, EH, SAS. Wrote the manuscript: RB, MRT, SAS, SSP. All authors have read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Author details

¹Laboratoire Universitaire de Biodiversité et Ecologie Microbienne (LUBEM), University of Western Brittany, Technopôle Brest-Iroise, 29280 Plouzané,

France. ²Department of Plant and Environmental Sciences and Copenhagen Plant Science Centre, University of Copenhagen, Thorvaldsensvej 40, 1871 Frb. C, Copenhagen, Denmark. ³Department of Agriculture, Food and Environment, University of Pisa, Via del Borghetto 80, 56124 Pisa, Italy. ⁴NIAB-EMR, New Road, East Malling, Kent ME19 6BJ, UK. ⁵School of Life Sciences, Warwick Crop Centre, University of Warwick, Wellesbourne, Warwickshire CV35 9EF, UK. ⁶Instituto Hispano-Luso de Investigaciones Agrarias (CIALE), University of Salamanca, Campus de Villamayor, C/Del Duero, 12, 37185, Villamayor, Salamanca, Spain. ⁷Institute of Biomedical and Environmental Science and Technology (iBEST), University of Bedfordshire, University Square, Luton, Bedfordshire LU1 3JU, UK.

Received: 14 April 2016 Accepted: 7 July 2016

Published online: 05 August 2016

References

- Bailey J, O'Connell R, Pring R, Nash C. Infection strategies of *Colletotrichum* species. In: Bailey JA, Jeger MJ, editors. *Colletotrichum: biology, pathology and control*. Wallingford: CAB International; 1992. p. 88–120.
- Cannon PF, Damm U, Johnston PR, Weir BS. *Colletotrichum* – current status and future directions. *Stud Mycol*. 2012;73:181–213.
- Silva DN, Talhinhas P, Cai L, Manuel L, Gichuru EK, Loureiro A, et al. Host-jump drives rapid and recent ecological speciation of the emergent fungal pathogen *Colletotrichum kahawae*. *Mol Ecol*. 2012;21:2655–70.
- 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: Top 10 fungal pathogens. *Mol Plant Pathol*. 2012;13:414–30.
- Damm U, Cannon PF, Woudenberg JHC, Crous PW. The *Colletotrichum acutatum* species complex. *Stud Mycol*. 2012;73:37–113.
- Nirenberg HI, Feiler U, Hagedorn G. Description of *Colletotrichum lupini* comb. nov. in modern terms. *Mycologia*. 2002;94:307–20.
- Talhinhas P, Baroncelli R, Le Floch G. Anthracnose of lupins caused by *Colletotrichum lupini*: a recent disease and a successful worldwide pathogen. *J Plant Pathol*. 2016;98:5–14.
- von Arx JA. Die Arten der Gattung *Colletotrichum* Cda. *Phytopathol Z*. 1957;29:413–68.
- Weir BS, Johnston PR, Damm U. The *Colletotrichum gloeosporioides* species complex. *Stud Mycol*. 2012;73:115–80.
- Damm U, Cannon PF, Woudenberg JHC, Johnston PR, Weir BS, Tan YP, et al. The *Colletotrichum boninense* species complex. *Stud Mycol*. 2012;73:1–36.
- Damm U, Cannon PF, Liu F, Barreto RW, Guatimosim E, Crous PW. The *Colletotrichum orbiculare* species complex: Important pathogens of field crops and weeds. *Fungal Divers*. 2013;61:29–59.
- Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature*. 2005;434:980–6.
- Krijger J-J, Thon MR, Deising HB, Wirsal SG. Compositions of fungal secretomes indicate a greater impact of phylogenetic history than lifestyle adaptation. *BMC Genomics*. 2014;15:722.
- Zhao Z, Liu H, Wang C, Xu J-R. Comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi. *BMC Genomics*. 2013;14:274.
- Tsai U, Tanaka E, Masuya H, Tanaka R, Hirooka Y, Endoh R, et al. Comparative genomics of *Taphrina* fungi causing varying degrees of tumorous deformity in plants. *Genome Biol Evol*. 2014;6:861–72.
- Morales-Cruz A, Amrine KCH, Blanco-Ulate B, Lawrence DP, Travadon R, Rolshausen PE, et al. Distinctive expansion of gene families associated with plant cell wall degradation, secondary metabolism, and nutrient uptake in the genomes of grapevine trunk pathogens. *BMC Genomics*. 2015;16:469.
- 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.
- Lo Presti L, Lanver D, Schweizer G, Tanaka S, Liang L, Tollot M, et al. Fungal effectors and plant susceptibility. *Annu Rev Plant Biol*. 2015;66:513–45.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80.
- Tamura K, Stecher G, Peterson D, Filipiński A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol*. 2013;30:2725–9.
- Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*. 2003;19:1572–4.

22. Baroncelli R, Sreenivasaprasad S, Sukno SA, Thon MR, Holub E. Draft genome sequence of *Colletotrichum acutatum sensu lato* (*Colletotrichum fioriniae*). Genome Announc. 2014;2:e00112–4.
23. Lardner R, Johnston PR, Plummer KM, Pearson MN. Morphological and molecular analysis of *Colletotrichum acutatum sensu lato*. Mycol Res. 1999;103:275–85.
24. Guerber JC, Liu B, Correll JC, Johnston PR. Characterization of diversity in *Colletotrichum acutatum sensu lato* by sequence analysis of two gene introns, mtDNA and intron RFLPs, and mating compatibility. Mycologia. 2003;95:872–95.
25. Sundelin T, Schiller M, Lübeck M, Jensen DF, Paaske K, Petersen BD. First report of anthracnose fruit rot caused by *Colletotrichum acutatum* on strawberry in Denmark. Plant Dis. 2005;89:432.
26. Sreenivasaprasad S, Talhinhos P. Genotypic and phenotypic diversity in *Colletotrichum acutatum*, a cosmopolitan pathogen causing anthracnose on a wide range of hosts. Mol Plant Pathol. 2005;6:361–78.
27. Baroncelli R, Zapparata A, Sarrocco S, Sukno SA, Lane CR, Thon MR, et al. Molecular diversity of anthracnose pathogen populations associated with UK strawberry production suggests multiple introductions of three different *Colletotrichum* species. PLoS ONE. 2015;10:e0129140.
28. LoBuglio KF, Pfister DH. A *Glomerella* species phylogenetically related to *Colletotrichum acutatum* on Norway maple in Massachusetts. Mycologia. 2008;100:710–5.
29. Baek J-M, Kenerley CM. The *arg2* gene of *Trichoderma virens*: cloning and development of a homologous transformation system. Fungal Genet Biol. 1998;23:34–44.
30. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.
31. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70.
32. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2.
33. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics. 2011;12:491.
34. Consortium TU. The universal protein resource (UniProt) in 2010. Nucleic Acids Res. 2010;38:D142–8.
35. Borodovsky M, Lomsadze A. Eukaryotic gene prediction using GeneMark. hmm-E and GeneMark-ES. Curr. Protoc. Bioinforma. Ed. Board Andreas Baxevanis Al. 2011; CHAPTER 4:Unit – 4.6.1-10.
36. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59.
37. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: *ab initio* prediction of alternative transcripts. Nucleic Acids Res. 2006;34:W435–9.
38. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10.
39. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16:157.
40. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. Bioinformatics. 2005;21:2104–5.
41. Horton P, Park K-J, Obayashi T, Fujita N, Harada H, Adams-Collier CJ, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35:W585–7.
42. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30:1236–40.
43. Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform. 2009;23:205–11.
44. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445–51.
45. Rawlings ND, Barrett AJ, Bateman A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res. 2012;40:D343–50.
46. Pemberton CL, Salmond GPC. The Nep1-like proteins—a growing family of microbial elicitors of plant necrosis. Mol Plant Pathol. 2004;5:353–9.
47. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, et al. Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the Clavicipitaceae reveals dynamics of alkaloid loci. PLoS Genet. 2013;9:e1003323.
48. 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:W339–46.
49. Liu F, Cai L, Crous PW, Damm U. The *Colletotrichum gigasporum* species complex. Persoonia. 2014;33:83–97.
50. Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. 2009;37:D233–8.
51. Have A ten, Tenberge KB, Benen JAE, Tudzynski P, Visser J, Kan JAL van. The contribution of cell wall degrading enzymes to pathogenesis of fungal plant pathogens. F. Kempken, ed. The Mycota, Vol. 11. Springer-Verlag, Berlin, Heidelberg. 2002. p. 341–58
52. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels. 2013;6:41.
53. Monod M, Capoccia S, L  chenne B, Zaugg C, Holdom M, Jousson O. Secreted proteases from pathogenic fungi. Int J Med Microbiol. 2002;292:405–19.
54. Di  guez-Urbeondo J, F  rster H, Adaskaveg JE. Visualization of localized pathogen-induced pH modulation in almond tissues infected by *Colletotrichum acutatum* using confocal scanning laser microscopy. Phytopathology. 2008;98:1171–8.
55. Alkan N, Fluhr R, Sherman A, Prusky D. Role of ammonia secretion and pH modulation on pathogenicity of *Colletotrichum coccodes* on tomato fruit. Mol Plant-Microbe Interact. 2008;21:1058–66.
56. Ohm RA, Feu N, Henrissat B, Schoch CL, Horwitz BA, Barry KW, et al. Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eighteen Dothideomycetes fungi. PLoS Pathog. 2012;8:e1003037.
57. Di Pietro A, Huertas-Gonz  lez MD, Guti  rrez-Corona JF, Mart  nez-Cadena G, M  glez E, Roncero MI. Molecular characterization of a subtilase from the vascular wilt fungus *Fusarium oxysporum*. Mol Plant-Microbe Interact. 2001;14:653–62.
58. Sreedhar L, Kobayashi DY, Bunting TE, Hillman BI, Belanger FC. Fungal proteinase expression in the interaction of the plant pathogen *Magnaporthe poae* with its host. Gene. 1999;235:121–9.
59. Armijos Jaramillo VD, Vargas WA, Sukno SA, Thon MR. New insights into the evolution and structure of *Colletotrichum* plant-like subtilisins (CPLSs). Commun Integr Biol. 2013;6:e25727.
60. Armijos Jaramillo VD, Vargas WA, Sukno SA, Thon MR. Horizontal transfer of a subtilisin gene from plants into an ancestor of the plant pathogenic fungal genus *Colletotrichum*. PLoS ONE. 2013;8:e59078.
61. Ottmann C, Luberacki B, K  fner I, Koch W, Brunner F, Weyand M, et al. A common toxin fold mediates microbial attack and plant defense. Proc Natl Acad Sci. 2009;106:10359–64.
62. Scharf DH, Heinekamp T, Brakhage AA. Human and plant fungal pathogens: the role of secondary metabolites. PLoS Pathog. 2014;10:e1003859.
63. Lim FY, Keller NP. Spatial and temporal control of fungal natural product synthesis. Nat Prod Rep. 2014;31:1277–86.
64. Kelly SL, Kelly DE. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? Philos Trans R Soc Lond B Biol Sci. 2013;368:20120476.
65. 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.
66. Takano Y, Kubo Y, Shimizu K, Mise K, Okuno T, Furusawa I. Structural analysis of PKS1, a polyketide synthase gene involved in melanin biosynthesis in *Colletotrichum lagenarium*. Mol Gen Genet. 1995;249:162–7.
67. Alkan N, Friedlander G, Ment D, Prusky D, Fluhr R. Simultaneous transcriptome analysis of *Colletotrichum gloeosporioides* and tomato fruit pathosystem reveals novel fungal pathogenicity and fruit defense strategies. New Phytol. 2015;205:801–15.
68. Goddard R, Hatton IK, Howard JAK, MacMillan J, Simpson TJ, Gilmore CJ. Fungal products. Part 22. X-Ray and molecular structure of the monoacetate of colletotrichin. J Chem Soc [Perkin 1] 1979; 1494–8.
69. Gohbara M, Kosuge Y, Yamasaki S, Kimura Y, Suzuki A, Tamura S. Isolation, structures and biological activities of Colletotrichins, phytotoxic substances from *Colletotrichum nicotianae*. Agric Biol Chem. 1978;42:1037–43.
70. Lu H, Zou WX, Meng JC, Hu J, Tan RX. New bioactive metabolites produced by *Colletotrichum* sp., an endophytic fungus in *Artemisia annua*. Plant Sci. 2000;151:67–73.
71. Ohra J, Morita K, Tsujino Y, Tazaki H, Fujimori T, Goering M, et al. Production of the phytotoxic metabolite, ferricrocin, by the fungus *Colletotrichum gloeosporioides*. Biosci Biotechnol Biochem. 1995;59:113–4.

72. Trigos A, Reyna S, Gutierrez ML, Sanchez M. Diketopiperazines from cultures of the fungus *Colletotrichum gloeosporioides*. *Nat Prod Lett*. 1997;1:13–6.
73. Guimarães DO, Borges WS, Vieira NJ, de Oliveira LF, da Silva CHTP, Lopes NP, et al. Diketopiperazines produced by endophytic fungi found in association with two Asteraceae species. *Phytochemistry*. 2010;71:1423–9.
74. Miyamoto KT, Komatsu M, Ikeda H. Discovery of gene cluster for mycosporine-like amino acid biosynthesis from Actinomycetales microorganisms and production of a novel mycosporine-like amino acid by heterologous expression. *Appl Environ Microbiol*. 2014;80:5028–36.
75. Leite B, Nicholson RL. Mycosporine-alanine: A self-inhibitor of germination from the conidial mucilage of *Colletotrichum graminicola*. *Exp Mycol*. 1992;16:76–86.
76. García-Pajón CM, Collado IG. Secondary metabolites isolated from *Colletotrichum* species. *Nat Prod Rep*. 2003;20:426–31.
77. Lowe RGT, Howlett BJ. Indifferent, affectionate, or deceitful: lifestyles and secretomes of fungi. *PLoS Pathog*. 2012;8:e1002515.
78. Gan P, Narusaka M, Kumakura N, Tsuchima 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.
79. Okagaki LH, Sailsbery JK, Eyre AW, Dean RA. Comparative genome analysis and genome evolution of members of the magnaporthaceae family of fungi. *BMC Genomics*. 2016;17:135.
80. You B-J, Choquer M, Chung K-R. The *Colletotrichum acutatum* Gene Encoding a putative pH-responsive transcription regulator is a key virulence determinant during fungal pathogenesis on citrus. *Mol Plant Microbe Interact*. 2007;20:1149–60.
81. Vidal-Melgosa S, Pedersen HL, Schückel J, Arnal G, Dumon C, Amby DB, et al. A new versatile microarray-based method for high-throughput screening of carbohydrate-active enzymes. *J Biol Chem*. 2015;290:9020–36. jbc.M114.630673.
82. Redman RS, Rodriguez RJ. Characterization and isolation of an extracellular serine protease from the tomato pathogen *Colletotrichum coccodes*, and its role in pathogenicity. *Mycol Res*. 2002;106:1427–34.
83. Sanz-Martin JM, Pacheco-Arjona JR, Bello-Rico V, Vargas WA, Monod M, Díaz-Minguez JM, et al. A highly conserved metalloprotease effector enhances virulence in the maize anthracnose fungus *Colletotrichum graminicola*. *Mol Plant Pathol* 2015; doi: 10.1111/mpp.12347
84. Jashni MK, Dols IHM, Iida Y, Boeren S, Beenen HG, Mehrabi R, et al. Synergistic action of a metalloprotease and a serine protease from *Fusarium oxysporum* f. sp. *lycopersici* cleaves chitin-binding tomato chitinases, reduces their antifungal activity, and enhances fungal virulence. *Mol Plant Microbe Interact*. 2015;28:996–1008.
85. Gijzen M, Nürnberger T. Nep1-like proteins from plant pathogens: recruitment and diversification of the NPP1 domain across taxa. *Phytochemistry*. 2006;67:1800–7.
86. Kleemann J, Rincon-Rivera LJ, Takahara H, Neumann U, van Themaat EVL, van der Does HC, et al. Sequential delivery of host-induced virulence effectors by appressoria and intracellular hyphae of the phytopathogen *Colletotrichum higginsianum*. *PLoS Pathog*. 2012;8:e1002643. Howlett BJ, editor.
87. Giraldo MC, Valent B. Filamentous plant pathogen effectors in action. *Nat Rev Microbiol*. 2013;11:800–14.
88. Doehlemann G, Hemetsberger C. Apoplastic immunity and its suppression by filamentous plant pathogens. *New Phytol*. 2013;198:1001–16.
89. Vargas WA, Sanz-Martin JM, Rech GE, Armijos-Jaramillo VD, Rivera LP, Echeverria MM, et al. A fungal effector with host nuclear localization and DNA-binding properties is required for maize anthracnose development. *Mol Plant Microbe Interact*. 2015;29:83–95.
90. Crešnar B, Petrič S. Cytochrome P450 enzymes in the fungal kingdom. *Biochim Biophys Acta*. 2011;1814:29–35.
91. Soby S, Caldera S, Bates R, VanEtten H. Detoxification of the phytoalexins maackiain and medicarpin by fungal pathogens of alfalfa. *Phytochemistry*. 1996;41:759–65.
92. García-Pajón CM, Hernández-Galán R, Collado IG. Biotransformations by *Colletotrichum* species. *Tetrahedron Asymmetry*. 2003;14:1229–39.
93. Tianpanich K, Prachya S, Wiyakrutta S, Mahidol C, Ruchirawat S, Kittakoop P. Radical scavenging and antioxidant activities of isocoumarins and a phthalide from the endophytic fungus *Colletotrichum* sp. *J Nat Prod*. 2011;74:79–81.
94. MacMillan J, Simpson TJ. Fungal products. Part V. The absolute stereochemistry of colletodiol and the structures of related metabolites of *Colletotrichum capsici*. *J. Chem. Soc. [Perkin 1]*. 1973; 1487–93.
95. Seebach D, Adam G, Zibuck R, Simona W, Rouilly M, Meyer WL, et al. Gloeosporone – a macrolide fungal germination self-inhibitor total synthesis and activity. *Liebigs Ann Chem*. 1989;1989:1233–40.
96. Mancilla G, Jiménez-Teja D, Femenía-Rios M, Macías-Sánchez AJ, Collado IG, Hernández-Galán R. Novel macrolide from wild strains of the phytopathogen fungus *Colletotrichum acutatum*. *Nat Prod Commun*. 2009;4:395–8.
97. Hu Z, Wang J, Bi X, Zhang J, Xue Y, Yang Y, et al. Colletotrichimine A, a novel indole-pyrazine alkaloid with an unprecedented C16N3-type skeleton from cultures of *Colletotrichum capsici*. *Tetrahedron Lett*. 2014;55:6093–5.
98. Chithra S, Jasim B, Sachidanandan P, Jyothis M, Radhakrishnan EK. Piperine production by endophytic fungus *Colletotrichum gloeosporioides* isolated from *Piper nigrum*. *Phytomedicine*. 2014;21:534–40.
99. Cimmino A, Mathieu V, Masi M, Baroncelli R, Boari A, Pescitelli G, et al. Higginsianins A and B, two diterpenoid α -pyrones produced by *Colletotrichum higginsianum*, with in vitro cytostatic activity. *J Nat Prod*. 2016;79:116–25.
100. Baroncelli R, Sanz-Martin JM, Rech GE, Sukno SA, Thon MR. Draft genome sequence of *Colletotrichum sublineola*, a destructive pathogen of cultivated sorghum. *Genome Announc*. 2014;2:e00540–14.
101. Alkan N, Meng X, 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;26:1345–58.
102. Klosterman SJ, Subbarao KV, Kang S, Veronese P, Gold SE, Thomma BPHJ, et al. Comparative genomics yields insights into niche adaptation of plant vascular wilt pathogens. *PLoS Pathog*. 2011;7:e1002137.
103. Ma L-J, van der Does HC, Borkovich KA, Coleman JJ, Daboussi M-J, Di Pietro A, et al. Comparative genomics reveals mobile pathogenicity chromosomes in *Fusarium*. *Nature*. 2010;464:367–73.
104. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, et al. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature*. 2003;422:859–68.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

