Dissecting the fungal biology of *Bipolaris papendorfii*: from phylogenetic to comparative genomic analysis

Chee Sian Kuan¹, Su Mei Yew¹, Yue Fen Toh¹, Chai Ling Chan¹, Yun Fong Ngeow¹, Kok Wei Lee², Shiang Ling Na¹, Wai-Yan Yee², Chee-Choong Hoh², and Kee Peng Ng¹,*

¹Department of Medical Microbiology, Faculty of Medicine, University of Malaya, Kuala Lumpur 50603, Malaysia, and
²Codon Genomics SB, Jalan Bandar Lapan Belas, Selangor Darul Ehsan 47160, Malaysia

*To whom correspondence should be addressed. Tel. +60 3-79492418. Fax. +60 3-79494881. E-mail: kpng@ummc.edu.my

Edited by Prof. Takashi Ito

Received 17 January 2015; Accepted 28 March 2015

Abstract

*Bipolaris papendorfii* has been reported as a fungal plant pathogen that rarely causes opportunistic infection in humans. Secondary metabolites isolated from this fungus possess medicinal and anticancer properties. However, its genetic fundamental and basic biology are largely unknown. In this study, we report the first draft genome sequence of *B. papendorfii* UM 226 isolated from the skin scraping of a patient. The assembled 33.4 Mb genome encodes 11,015 putative coding DNA sequences, of which, 2.49% are predicted transposable elements. Multilocus phylogenetic and phylogenomic analyses showed *B. papendorfii* UM 226 clustering with *Curvularia* species, apart from other plant pathogenic *Bipolaris* species. Its genomic features suggest that it is a heterothallic fungus with a putative unique gene encoding the LysM-containing protein which might be involved in fungal virulence on host plants, as well as a wide array of enzymes involved in carbohydrate metabolism, degradation of polysaccharides and lignin in the plant cell wall, secondary metabolite biosynthesis (including dimethylallyl tryptophan synthase, non-ribosomal peptide synthetase, polyketide synthase), the terpenoid pathway and the caffeine metabolism. This first genomic characterization of *B. papendorfii* provides the basis for further studies on its biology, pathogenicity and medicinal potential.

Key words: dematiaceous, *Bipolaris papendorfii*, *Cochliobolus*, *Curvularia*, LysM domain

1. Introduction

*Bipolaris* is a large genus of dematiaceous hyphomycetes belonging to the order Pleosporales in the class Dothideomycetes.¹ It is closely related to two other genera, *Cochliobolus* and *Curvularia*. The taxonomy of these three genera has caused confusion for many mycologists as their sexual and asexual states have undergone several nomenclatural changes.² Manamgoda et al.³ showed that *Cochliobolus* species can be separated broadly into two subgroups, one with *Bipolaris* species and another with *Curvularia* species. Although *Bipolaris* and *Curvularia* share many morphological similarities including the same teleomorph of *Cochliobolus*, some species demonstrate intermediate morphology and some *Bipolaris* species cluster phylogenetically with *Curvularia* species instead of with other *Bipolaris* species.³

*Bipolaris* species have been frequently isolated from soil, plant debris, as well as living plants on which they may exist as saprophytes or phytopathogens. The fungal genus *Bipolaris* has an asexual state named *Cochliobolus* are notorious for causing devastating disease...
epidemics of crop plants such as maize, oat, wheat, rice and sorghum, leading to problems of starvation, malnutrition and world economic crisis. Some plant pathogens such as Bipolaris australiensis, Bipolaris hauasiensis and Bipolaris spicifera occasionally cause disease in both immunocompetent and immunocompromised humans, manifesting as allergic or chronic, superficial or deep-seat, localized or disseminated infections.

Bipolaris papendorfii has received little attention, compared with other virulent species that are pathogenic in food crops and humans. It is a rare opportunistic pathogen that has been reported in only three cases of keratitis. Like other plant pathogenic Bipolaris species, it causes corn leaf spot, but the underlying pathogenic mechanism is totally unclear. More interestingly, a recent report described two bioactive compounds, hamgerone and radicinol, from B. papendorfii that exhibited anti-proliferative activity in various cancer cells. Both of these compounds also activated the apoptosis pathway by regulating the expression of BCL-2, caspase 3 and p53.

The development of high-throughput next-generation sequencing technologies (NGS) has provided massive genomic nucleotide sequence data for the study of microbes. Although there are whole genome sequence data for several Cochliobolus species in international databases, the whole genome sequence of B. papendorfii has not yet been determined. The genetic fundamental and basic biology of this fungus are largely unknown. In this study, we report the de novo draft genome sequence of B. papendorfii UM 226 that was isolated from the skin scraping of a patient with dermatomycosis. The draft genome was generated by a hybrid assembly of Illumina and PacBio sequencing technologies to improve genome assembly. To our knowledge, this report is the first detailed description of the B. papendorfii genome. It should provide useful information on the biology, pathogenicity and medicinal potential of this fungus.

2. Materials and methods

2.1. Ethics statement
The isolate used in this study was obtained from an archived fungal collection. All labels on the sample have been erased with the exception of sample type and clinical diagnosis. Thus, there is no information traceable back to the patient from whom the isolate was obtained. As such, this genomic study is exempt from ethical approval in our teaching hospital. (http://www.ummc.edu.my/view/content.php?ID=VGxSWlBRPT0=).

2.2. Fungal isolate
UM 226 was isolated from the skin scraping of a patient with suspected dermatomycosis in the University of Malaya Medical Centre (UMMC), Kuala Lumpur, Malaysia. The isolate was processed according to the laboratory’s standard operating procedures (SOP) with direct wet mount microscopy followed by inoculation on Sabouraud Dextrose Agar (SDA) for incubation at 30°C up to 7 days, with alternate day examination for fungal growth. Macroscopic examination was carried out to observe colonial characteristics, including color, texture and topology. tease mount and slide culture were performed to identify the arrangement of conidia and conidiophores. The isolate was viewed under the scanning electron microscope (SEM) to examine wall ornamentation and fine surface features of fungal conidia.

2.3. DNA sequencing and phylogenetic analysis
The internal transcribed spacer region (ITS), small subunit of the ribosomal RNA gene (SSU) and large subunit of the ribosomal RNA gene (LSU) were used as targets for molecular identification of the fungal isolate. Total DNA extraction, PCR amplification and sequencing were performed as described previously. The primers used are listed in Supplementary Table S1. Annealing temperatures used to amplify ITS, SSU and LSU genes were 58, 52 and 46°C, respectively. The sequenced data were subjected to BLASTn search against the non-redundant (nr) NCBI-nucleotide database for fungal identification. Unique ITS, SSU and LSU sequences from the isolate, together with an additional 145 species of Cochliobolus, Bipolaris and Curvularia and an outgroup strain of Alternaria brassicicola (Supplementary Table S2) were subjected to phylogenetic analysis. Multiple sequence alignments of collected ITS, LSU and SSU sequences were generated using M-Coffee. Individual alignments were concatenated for Bayesian Markov Chain Monte Carlo (MCMC) analysis partitioned by gene. Bayesian tree analyses were performed using MrBayes v3.2.1 with reversible jump MCMC averaging over the entire general time reversible (GTR) rates and gamma-distributed rate heterogeneity for all subsets of partitioned scheme. A total of 5,000,000 generations were run with a sampling frequency of 100, and diagnostics were calculated for every 1,000 generations. The first 12,500 trees were discarded with burn-in setting of 25%. Convergence was assessed with a standard deviation of split frequencies below 0.01, no obvious trend for the plot of the generation versus the log probability of the data, and the potential scale reduction factor (PSRF) close to 1.0 for all parameters.

2.4. Genomic DNA extraction
Genomic DNA was extracted from fresh fungal mycelia in accordance with the protocol of Moslem et al. with slight modification. Briefly, mycelia were scraped off from the agar surface and crushed into fine powder with liquid nitrogen. One gram of crushed fungal powder was suspended in 6 ml of DNA extraction buffer [200 mM Tris–HCl, (pH 8.5), 250 mM NaCl, 25 mM EDTA, 0.6% (w/v) SDS] in a 50-ml centrifuge tube and incubated at 65°C for 10 min. Next, 1.4 ml of 3 M sodium acetate (pH 5.3) was added to the mixture and followed by incubation at 80°C for 20 min. The tube was then centrifuged at 4,000 × g for 5 min and the supernatant was transferred into a new 50-ml centrifuge tube. Genomic DNA was then precipitated with equal volume of 100% (v/v) isopropanol. Genomic DNA was washed from the pellet by adding 500 µl of ice-cold 70% (v/v) ethanol and finally dissolved in nuclease-free distilled water. The purified genomic DNA was treated with RNAase at 37°C for overnight. Then, the mixture was incubated at 65°C for 10 min, followed by centrifugation at 28,433 × g for 5 min. The supernatant was transferred into a new 1.5-ml microcentrifuge tube. The quantity and quality of the extracted genomic DNA were determined using NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific).

2.5. Genome sequencing and de novo hybrid assembly
Illumina library was prepared using TruSeq v3 Reagent Kits (Illumina). UM 226 genomic DNA was sheared into smaller fragments by Covaris S2E210 or Bioruptor. The 500-bp fragments were purified though gel electrophoresis, which then selectively enriched and amplified by PCR. Next, qualified sequencing library were sequenced using Illumina HiSeq 2000 Sequencer (Illumina) with 2 × 90 bp paired-end mode. PacBio RS sequencing was performed using the PacBio RS II system (Pacific Biosciences). UM 226 genomic DNA was used for 10-kb library preparation using PacBio DNA Template Prep Kit 2.0 (3–10 kb). SMRTbell templates were then annealed using PacBio DNA/Polymerase Binding Kit P4. Three SMRT cells were used for sequencing with MagBead standard Seq v2 collection protocol using 180-min movies.
Genome assembly was performed using the hybrid approach using ECTools with Illumina contigs and PacBio long reads. Illumina short reads were first pre-processed trimming bases with a Phred quality below Q20 from the 3' end of the reads, retaining reads ≥ 50 bp and reads with 40% bases having Qr ≤ 20 were filtered out using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The pre-processed reads were corrected using Quake software with 16-mer setting. Corrected reads were then assembled into preliminary contigs using Velvet version 1.2.075 with k-mer setting = 57, scaffolding = no, insert_length = 462 and ins_length_sd = 60.

ECTools then utilize velvet assembled contigs with length ≥ 500 bp to correct errors of the PacBio long reads. Error corrected long reads were then assembled using Celera assembler. Both ECTools and Celera assembly were performed according to the software instruction with default parameters setting. Celera assembled contigs smaller than 10 kb were discarded, and Quiver was used to polish the remaining contigs. The assembly was imported into SMRT analysis software version 2.1.1 through SMRT portal interface as the reference, and three SMRT cells data generated were aligned to the genome using the RS-Resequence protocol iteratively. Through the polishing steps, low-quality contigs with low coverage information from Quiver output were discarded from the final assembly.

2.6. Gene prediction and annotation

Protein coding sequences of B. papendorfii UM 226 were predicted using GeneMark-ES version 2.3e. Annotation of coding sequences for UM 226 was completed using BLAST (Basic Local Alignment Search Tool) searches against the NCBI nr protein and SwissProt databases. Gene Ontology and KEGG metabolic pathways matches were carried out using local BLAST2GO tools. Protein classifications were performed by EuKaryotic Orthologous Group (KOG). The total number of predicted proteins involved in post-translational modification, protein turnover, chaperones (KOG annotation for other Cochliobolus species) was obtained from the Joint Genome Institute (JGI) website, http://genome.jgi.doe.gov/. Protein domain families were matched to Pfam database using InterProScan. Individual rRNA and tRNA sequences were identified using Rfam database using a sampling frequency of 100 for 100,000 generations with a burn-in setting of 25%.

2.8. Phylogenetic analysis

A phylogenetic tree was built using all proteome clusters of the seven Cochliobolus species generated from comparative analysis. The tree is rooted with A. brassicicola ATCC 9636 as outgroup. A total of 6,051 single-copy orthologous genes containing one member in each species was subjected to individual sequence alignments using ClustalW version 2.0. All spurious sequences or poorly aligned regions were discarded with trimAL (with –gt 0.5). The filtered multiple alignments were then concatenated into a superalignment with 2,991,675 characters. Bayesian phylogenetic analysis was run using MrBayes v3.2.1 with mixed amino acid model, gamma-distributed rate variation across sites and a proportion of invariable sites. The MCMC was run using a sampling frequency of 100 for 100,000 generations with a burn-in setting of 25%.

2.9. Prediction of secondary structure and homology modelling

PSIPRED v3.3 was used to predict the secondary structure of the LysM domains. Three-dimensional structural models of the LysM domains were generated using SWISS-MODEL server. Amino acid sequence alignment between UM226-683 domains and Ecp6 was performed using ClustalW version 2.0. The template used in this modelling was selected from template identification function in the SWISS-MODEL program. In this study, a Cladosporium fulvum LysM effect-domain with Ecp6 (Protein Data Bank entry 4b8v) was selected as the template. The predicted models were stored as a PDB output file, and the structures were visualized by Swiss-PdbViewer 4.0.1. The quality of the protein model generated was validated by using the PROCHECK program as described previously.

2.10. Mating type locus identification and phylogenetic analysis

MAT1-2 mating type regions immediately 5 kb upstream and downstream were retrieved from the sequenced genome using Artemis v12.0 sequence viewer. Bayesian tree analysis was carried out by using a procedure similar to ribosomal molecular identification with 500,000 generations run. A phylogenetic tree was constructed using MAT1-2 nucleotide sequences from B. papendorfii UM 226, together with additional nine species of Cochliobolus and Bipolaris. The tree is rooted with Alternaria alternata as outgroup.

3. Results and discussion

3.1. Morphological and molecular identification

The B. papendorfii UM 226 colonies on SDA were fast growing, velvety and compact with abundant short aerial mycelium (Fig. 1a). The surface of the colonies was initially grayish white to olivaceous green and becomes black with a raised grayish periphery after being cultured for 7 days (Fig. 1a). It had a black-green coloration on the reverse side
The diameter of the colony was 60 mm after 7 days of incubation at 30°C. Microscopic morphology demonstrated simple or flexuose branched, pale to medium brown conidiophores (up to 200 µm long, 4–9 µm wide) with zig-zag rachis bearing conidium in a sympodial pattern (Fig. 1c). Conidiogenous nodes were verruculose, distinct, swollen and up to 5 µm wide (Fig. 1d). The conidia (24–40 × 10–20 µm) were curved, kidney shaped or obpyriform (Fig. 1c). The conidia had three pseudoseptates with the widest at the second cell from the base. SEM observation demonstrated typical, roughed, zig-zag conidiophores with several conidia (Fig. 1e). Unlike other...
Bipolaris spp. that have typical smooth walled conidia, the SEM microphotograph of the UM 226 conidia showed that they were verrucose walled with thickening at places of pseudosepta (Fig. 1e). On the basis of these characteristics, the UM 226 clinical isolate was found to be morphologically related to B. papendorfii.

The preliminary morphological identification of the UM 226 isolate was confirmed by PCR amplification of the ITS, SSU and LSU gene regions, which yielded specific amplicons of ~540 (ITS), 1,200 (SSU) and 900 bp (LSU), respectively (Supplementary Fig. S1). The sequenced data were used to construct a phylogram using combined gene analysis with the additional 145 ex-type strains of Cochliobolus, Bipolaris and Curvularia spp. (Fig. 2). As previously noted, the phylogenetic tree analysis separated the Curvularia, Bipolaris and Cochliobolus species into two major, well-supported groups: Groups 1 and 2 (Fig. 2). Phylogenetic analysis of the combined dataset of the three loci revealed that 59 species were clustered in Group 1 and 87 species were clustered in Group 2. Group 1 consisted of Bipolaris and Cochliobolus spp. while Group 2 comprised species of Curvularia, Bipolaris and Cochliobolus. All the plant pathogenic and economically important fungi are identified in genus Bipolaris, which were grouped into Group 1. The Bipolaris species that clustered in Group 2 (Fig. 2). Multilocus phylogenetic analysis revealed that UM 226 was clustered together with B. papendorfii strain CMON22 in Group 2 (Fig. 2). Based on the morphological examination and multi-gene phylogeny, the isolate was identified as B. papendorfii.

3.2. Genome features of B. papendorfii UM 226

In this study, Illumina HiSeq 2000 and PacBio sequencing technologies were used to produce hybrid genome assembly for B. papendorfii UM 226. The total assembly size of B. papendorfii UM 226 genome is 33.4 Mb. The sequencing reads were assembled into 374 contigs with an N50 contig size of 146,099 kb. Hybrid assembly of B. papendorfii UM 226 showed a lower number of contigs and a higher N50 value (Table 1). Summary of the sequence statistics is presented in Table 1. The genome had an average G + C content of 50.65%.

3.3. Transposable elements

Transposable elements (TEs), also known as jumping genes, were important in the variability, gene structure and evolution of the genome. Recently, Santana et al. demonstrated that C. heterostrophus race O genome (~36 Mb) harbours 5.9% of TEs. Here, our bioinformatics prediction revealed that 2.49% (831,552 bp) of the assembled genome of B. papendorfii UM 226 consisted of TEs (Table 2). The transposable elements of Class I, retrotransposons comprised 2.31% of the genome, whereas Class II transposable elements, DNA transposons comprised only 0.19% of the genome. LINE (Long Interspersed Nuclear Element)- and LTR (Long Terminal Repeat)-type retrotransposons are mostly presented in the fungal genomes. Gypsy and Copia elements are the LTR superfamilies that widely distributed in these genomes. As previously noted, Gypsy elements were identified to be located upstream of the PKS1, OXI1 and LAMI1 genes, downstream of PKS2 gene as well as between the RED2 and RED3 genes in the C. heterostrophus race O genome; Copia element was located downstream of the TOX9 gene in the genome. In this study, our analysis showed that 0.24, 0.67 and 0.29% of the sequenced B. papendorfii UM 226 genome, estimated 33.4 Mb, consists of LINE, gypsy and TY1_Copia, respectively. However, the locations of these TEs in the genome remain unknown at this stage of knowledge. The most predominant TE families, both in terms of number of copies and element size, are DDE_1 transposases (Table 2). The DDE_1 transposases contain a functional domain to mediate DNA-binding, protein–protein interaction and ‘cut-and-paste’ activities. The DDE domain has been identified in all eukaryotic cut-and-paste TEs, suggesting their common evolutionary origin.

3.4. Gene prediction

Our pipeline showed that there are 128 tRNAs, 12 5S rRNA, 3 18S rRNA and 3 28S rRNA in the genomic sequence. In addition, the B. papendorfii UM 226 genome revealed a total of 11,015 putative coding DNA sequences (CDS) with an average gene length of 1,425 bp. Gene density of protein coding genes is 3.29 genes/10 kb. The genome size of B. papendorfii UM 226, number of predicted genes and gene density are comparable to the previously sequenced Cochliobolus species genomes (Supplementary Table S3). On average, there are 2.95 exons per gene. A total of 10,708, 6,963 and 8,283 gene-coding sequences are homologous to currently available proteins in the NCBI nr, SwissProt and InterPro databases with 7,759 hypothetical proteins identified based on the top hit of the BLAST result against NCBI nr database. The large amount of hypothetical proteins in the genome indicated the uniqueness of B. papendorfii isolate.

To further characterize the predicted proteins, the genome was mapped to KOG and KEGG databases. All the proteins were ascribed to 26 different functional groups based on KOG classification (Fig. 3a and Supplementary Table S4). Apart from the poorly characterized categories: categories [R] General functions prediction only, [S] Function unknown and [X] Unnamed protein, the top five KOG groups were annotated in were [O] Post-translational modification, protein turnover, chaperones (512 genes), [T] Signal transduction mechanisms (396 genes), [I] Lipid transport and metabolism (382 genes), [C] Energy production and conversion (344 genes) and [Q] Secondary metabolites biosynthesis, transport and catabolism (334 genes) (Fig. 3a). Like other Cochliobolus species [C. heterostrophus C4 (560 copies), C. heterostrophus C5 (608 copies), C. victoriae F13 (597 copies), C. sativus ND90Pr (585 copies), C. carboxum 26-R-13 (597 copies), C. lunatus m118 (587 copies)] that have been previously sequenced by JGI, the majority of the genes in B. papendorfii UM 226 genome was involved in post-translational modifications (Fig. 3a). Post-translational modifications of proteins regulate a wide variety of basic cellular functions in fungi, such as cellular growth, adaptive and developmental processes. Autophagy plays a vital role in cellular growth, degradation as well as cell survival in filamentous fungi, especially in starvation condition. Interestingly, we identified 103 putative serine/threonine-protein kinase involved in autophagy from the KOG group O (Supplementary Table S4). Moreover, various fungal cell surface proteins post-translationally modified in fungi are involved in fungal virulence. To gain insight into the gene functions, we managed to assign 3,074 predicted proteins to their orthologous genes in the KEGG database. At the first glance, there were 1,125 predicted proteins of B. papendorfii UM 226 involved in metabolic pathways in the KEGG database (Supplementary Table S5). The top five categories in KEGG metabolic pathway are carbohydrate metabolism, amino acid metabolism, lipid metabolism, nucleotide metabolism and xenobiotics biodegradation and metabolism (Fig 3b). There are 293 unique reactions corresponding to carbohydrate metabolism. This result suggests that there are enriched carbohydrate mechanisms such as glycolysis, gluconeogenesis and citrate cycle (TCA cycle), that can facilitate the use of diverse substrates of B. papendorfii UM 226 in exposure to various carbohydrate sources in the environment.
Figure 2. Bayesian phylogram generated based on combined genes of ITS, SSU and LSU sequenced data. The tree is rooted with Alternaria brassicicola as outgroup. Numbers on the nodes indicate Bayesian posterior probability based on 100 sampling frequency for a total of 150,000 generations. This figure is available in black and white in print and in colour at DNA Research online.
Table 1. Genome statistics of Bipolaris papendorfii UM 226

| Details | Illumina HiSeq 2000 | PacBio RS II system | Hybrid assembly of Illumina and PacBio |
|---------|----------------------|----------------------|--------------------------------------|
| Sequencing depth | ~87x | ~34x | — |
| Read type | Small insert | Single end | — |
| Library size (bp) | 500 | 20,000 | — |
| Read length (bp) | 90 | 2,742 | — |
| Total number of reads | 33,655,556 | 405,257 | — |
| Total number of bases | 3,029,000,040 | 1,111,215,948 | — |
| Assembly size (Mb) | 33,620,574 | — | 33,379,470 |
| Contigs (N50) (kb) | 127,284 | — | 374 |
| Contigs (N500) (kb) | 127,284 | — | 146,099 |
| Contigs G + C (%) | 50.55 | — | 50.65 |

Table 2. Putative transposable elements in the genome sequence of Bipolaris papendorfii UM 226

| Class | Family name | Total number | Total bases | Percentage of assembled genome |
|-------|-------------|--------------|-------------|-------------------------------|
| I     | DDE_1       | 422          | 369,615     | 1.11                          |
|       | SYNE        | 279          | 223,716     | 0.67                          |
|       | LINE        | 54           | 78,738      | 0.24                          |
|       | Irr_Roo     | 2            | 348         | 0.00001                       |
|       | TY1_Copia   | 87           | 95,805      | 0.29                          |
| II    | helitronORF | 4            | 2,673       | 0.01                          |
|       | hAT         | 21           | 15,879      | 0.05                          |
|       | mariner      | 55           | 36,201      | 0.11                          |
|       | mariner_antI | 6            | 2,793       | 0.01                          |
|       | MuDR_A_B    | 19           | 4,701       | 0.01                          |
|       | cacta       | 10           | 1,083       | 0.00003                       |
| Total |             | 831,552      | 2.49        |

3.5. The CAZymes family

All types of CAZymes, especially plant cell wall degrading enzymes, are commonly produced by fungal plant pathogens to penetrate the rigid barrier of the plant cell wall. Most of the fungal CAZymes were known to degrade plant cell wall materials to obtain carbon source for fungal growth as well as for penetration and infection of their host plants. As B. papendorfii has been reported to cause corn infection, the B. papendorfii UM 226 genome was mapped to the CAZy database to determine the presence of CAZyme-coding homologous genes. A total of 728 putative CAZyme-coding homologous genes were determined in the genome. This comprises 122 with auxiliary activities (AA), 69 carbohydrate-binding module (CBM), 150 carbohydrate esterases (CE), 266 glycoside hydrolases (GH), 104 glycosyl transferases (GT) and 17 polysaccharide lyases (PL) (Fig. 4 and Supplementary Table S6). The number of GH, AA and PL family candidates identified is comparable to the several reported in several studies for necrotrophic and hemibiotrophic Cochliobolus species. The number of GHs in C. heterostrophus C4, C. heterostrophus C5 and C. sativus ND90Pr are 276, 292 and 272, respectively; the number of GTs in C. heterostrophus C4, C. heterostrophus C5 and C. sativus ND90Pr are 96, 103 and 99, respectively; the number of PLs in C. heterostrophus C4, C. heterostrophus C5 and C. sativus ND90Pr are 15, 15 and 15, respectively. Interestingly, the B. papendorfii UM 226 genome contains a higher number of CE-coding genes homologs (150 copies) compared with that of the reported Cochliobolus species, such as C. heterostrophus C4 (46 copies), C. heterostrophus C5 (49 copies) and C. sativus ND90Pr (47 copies). Our findings revealed that the fungus has 12 of the 16 CE families, with family CE2, CE6, CE11 and CE13 missing. The high number of CE1 (42 copies) and CE10 (57 copies) indicated that B. papendorfii UM 226 might produce many putative plant degrading enzymes containing carboxylesterase and endo-1,4-β-xylanase activities, respectively. CE1 and CE10 members were reported to have a wide range of substrate specificities. CE1 and CE10 are hemicellulose-targeting CE1 and CE10, respectively. In this study, we identified seven copies of putative endo-1,4-β-xylanases, which are responsible to degrade the 1,4-glycosidic bond and thus depolymerize the xylan polymer (a major carbohydrate component of hemicellulose) into its monomer constituents (Supplementary Table S6). Our result also demonstrated a high number of GH and AA genes, postulating the ability of our fungal isolate to degrade lignocellulose, which is a main structural constituent of non-woody or woody plants. A total of 16 putative ligninolytic genes: multicopper oxidase (6 copies) and heme peroxidase (10 copies) were identified in this work. Collectively, B. papendorfii UM 226 possesses a wide array of CAZymes that are involved in the degradation of the complex carbohydrates in the plant cell walls.

3.6. Comparative genomics of B. papendorfii UM 226 and other Cochliobolus species

In this work, we selected seven publicly available Cochliobolus species genomes and an A. brassicicola genome for rooting the phylogenomic tree (Fig. 3). A total of 101,438 proteins were clustered into 14,296 orthologous clusters with 6,051 single-copy orthologous genes determined. It should be noted that B. papendorfii UM 226 was tightly clustered with C. lunatus m118. Among the 14,296 gene family clusters, 222 gene families were conserved between B. papendorfii UM 226 and C. lunatus m118. In agreement with the multilocus phylogenetic analysis (Fig. 2), these data showed a close evolutionary relationship between B. papendorfii UM 226 and Curvularia species, which are separated from all plant pathogenic Cochliobolus (anamorph Bipolaris) species (Fig. 5)
and DNA-binding function (Table 3). Among them, particular attention was paid to a hypothetical protein (UM226_638) with multiple lysin motifs (LysMs) domains. Based on the CAZyme annotation, UM226_638 hypothetical protein belongs to the carbohydrate-binding module (CBM50) of CAZyme which is involved in the degradation of chitin or peptidoglycan. Fungal LysM effectors had been shown to have chitin-binding affinity to perturb chitin-induced immunity from their host plant.\textsuperscript{51} A novel virulence factor with LysM domains, known as Ecp6 effector, was well characterized from the tomato pathogenic fungus \textit{C. fulvum}.\textsuperscript{52} The presence of LysM domains in Ecp6 effector was postulated to protect fungal cell walls from plant chitinases hydrolysis by competing with host receptors for chitin.\textsuperscript{52} As previously noted,\textsuperscript{52} the LysMs-containing UM226_638 protein belongs to type A fungal LysM protein (PF01467) and contains three LysM domains. Many of the type A LysM proteins play an important role in the degradation of chitin and peptidoglycan.\textsuperscript{53}
role in infection process. For convenient description hereafter, these three domains were assigned as LysM domain I (75–118 amino acids), LysM domain II (338–384 amino acids) and LysM domain III (454–502 amino acids) (Fig. 6a). Unlike C. fulvum secreted Ecp6,38,39 we found that the UM226_638 protein do not have a signal peptide for secretion. Thus, UM226_638 protein might not be a secreted protein. Moreover, secondary structure prediction showed that only LysM domain II and domain III have a βααβ secondary structure (Fig. 6b). On the contrary, LysM domain I composed of three α-helices and a β-sheet (Fig. 6b).

Tertiary structures of the LysM domain II and LysM domain III were then generated by using the SWISS-MODEL server based on the experimentally solved structural homolog. As previously reported,38,39 the three-dimensional structural prediction showed that the domains comprised two α-helices packed on the same side of the anti-parallel two β-sheets (Fig. 6c, panel 1 and 2). The molecular surface of the third LysM domain of UM226_638 was computed (Fig. 6c, panel 3). In consistent with the structural homology of C. fulvum Ecp6,52 a shallow groove was observed in the LysM domain III that provides a binding site for chitin oligomers. Accuracy of the domain models generated was verified by PROCHECK program, a protein structure validation program (data not shown). The stereochemical parameters were all indicated to be within the acceptable limit, and it can serve as a good hypothetical domain structures for further

Table 3. Specific family clusters in Bipolaris papendorfii UM 226

| Families     | Gene ID     | InterPro ID/domain                                      |
|--------------|-------------|--------------------------------------------------------|
| Bipap11034   | UM226_39    | IPR023213/Chloramphenicol acetyltransferase-like domain |
|              | UM226_347   | IPR023213/Chloramphenicol acetyltransferase-like domain |
|              | UM226_1504  | IPR023213/Chloramphenicol acetyltransferase-like domain |
|              | UM226_638   | IPR018392/LysM domain                                   |
|              | UM226_6391  | IPR018392/LysM domain                                   |
|              | UM226_10258 | IPR000383/Alpha/Beta hydrolase fold                     |
|              | UM226_11642 | IPR000383/Alpha/Beta hydrolase fold                     |
|              | UM226_11099 | IPR021109/Aspartic peptidase domain                     |
|              | UM226_10987 | IPR021109/Aspartic peptidase domain                     |
|              | UM226_6690  | IPR006600/HTH CenpB-type DNA-binding domain             |
|              | UM226_9523  | IPR004875/DDE superfamily endonuclease, CENP-B-like    |
|              | UM226_10253 | IPR029068/Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase |
|              | UM226_11078 | IPR029068/Glyoxalase/Bleomycin resistance protein/Dihydroxybiphenyl dioxygenase |
functional study. Taken together, the putative unique LysM-containing UM226_638 protein might be a fungal avirulence protein that involved in fungal virulence on its host.

3.7. Sexual reproduction

Sexual reproduction is well characterized for most of the known Cochliobolus species. In Cochliobolus species, a single mating type
locus (MAT) was important to regulate sexual development and production of sexual spores. The coexistence of heterothallic and homothallic species are common in the genus of Cochliobolus. Thus, sexual Cochliobolus species can be either heterothallic fungi that require two strains of opposite mating type to participate in the sexual process (MAT1-1 or MAT1-2 gene) or homothallic fungi that carry both MAT1-1 and MAT1-2 genes. In this study, we successfully identified a putative MAT1-2 (UM226_825) gene but no MAT1-1 gene in the UM 226 genome. UM226_825 shared a homology (81% identical) and domain architecture (DNA domain of the high mobility group, HMG domain) with the MAT-2 from Curvularia ellisii. The structural organization of 12 kb at the B. papendorfi UM 226 MAT1-2 region was examined (Fig. 7a). As previously described, three other putative genes have been identified adjacent to the MAT1-2: (i) a β-glucosidase homolog gene (3′ flank of MAT1-2), (ii) a homolog of a yeast gene (ORF1) with unknown function (5′ flank of MAT1-2) and (iii) a GTPase-activating homolog gene (5′ flank of MAT1-2) (Fig. 7a). However, these genes were not involved in sexual reproduction. Turgeon et al. showed that the MAT1-2 gene evolved at a faster rate than other nucleotide sequences in the genome of Cochliobolus species. In this study, a phylogenetic tree was built to reveal evolutionary relationships between B. papendorfi and other related species (Fig. 7b). MAT1-2 genes from three homothallic fungi and six heterothallic fungi were included in the analysis (Supplementary Table S7). Based on the phylogenetic analysis, the MAT1-2 genes can be divided into two groups, which agree well with the tree previously discussed (Fig. 2). B. papendorfi MAT1-2 is tightly clustered together with the MAT1-2 of heterothallic C. ellisii, which are separated from another group that consisted of virulent plant pathogens. The close relationship of asexual B. papendorfi to the sexual C. ellisii suggests that B. papendorfi can also be reproduced sexually in field.

3.8. Secondary metabolism
Fungal secondary metabolites are a double-edged sword. They are the source of medicinal biologically active compounds. They are also

Figure 7. MAT1-2 gene of Bipolaris papendorfi UM 226. (A) Schematic representation of major open reading frames (ORFs) identified in the MAT regions of B. papendorfi UM 226 and Cochliobolus heterostrophus C4. Numbers are in kilobases. (B) Bayesian phylogram generated based on MAT1-2 nucleotide sequences. The tree is rooted with Alternaria alternata as outgroup. Numbers on the nodes indicate Bayesian posterior probability based on 100 sampling frequency for a total of 150,000 generations. This figure is available in black and white in print and in colour at DNA Research online.
notorious mycotoxins that cause diseases on plants, animals and humans. Armed with the knowledge that secondary metabolite biosynthetic genes often occur in clusters with a key backbone gene. In this study, we used SMURF prediction to annotate potential gene cluster backbone enzymes in the *B. papendorfii* UM 226 genome. There are 32 gene clusters encoding core enzymes, including 19 genes encoding polyketide synthase (PKS) or PKS-like, 11 genes encoding non-ribosomal peptide synthetase (NRPS) or NRPS-like, a gene encoding polyketide synthase/non-ribosomal peptide synthase hybrid (PKS-NRPS) and a gene encoding dimethylallyl tryptophan synthase (DMATS), that are important for the biosynthesis of melanin, bioactive compounds and mycotoxins. The only DMATS gene identified in *B. papendorfii* UM 226 is UM226_1500 (48% identical), which is often involved in the first committed step of ergot alkaloids biosynthesis. Ergot alkaloids are mycotoxins but also important natural pharmaceuticals for human use.17

Like most *Cochliobolus* plant pathogens reported, *B. papendorfii* UM 226 is enriched with multidomain NRPS encoding genes (NPS) and PKS genes. NPS2 (UM226_2338, 74% identical), NPS6 (UM226_3398, 87% identical), NPS4 (UM226_5568, 82% identical) and NPS10 (UM226_3258, 93% identical) orthologous genes found in *B. papendorfii* UM 226 were reported to be conserved across all *Cochliobolus* species. NPS2, NPS6, NPS4 and NPS10 harbour multiple domains, including AMP-binding domain, condensation domain, phosphopantetheine-binding domain and amino acid adenylation domain. NPS2 is likely responsible for the synthesis of siderophore and intracellular iron storage. All of the reported *Cochliobolus* pathogens are necrotrophs, which produce Host Selective Toxins (HSTs) to infect cereal crops. Like *C. carbonum* (*Bipolaris zeicola*), *B. papendorfii* has been reported to cause Corn Leaf Spot disease. Thus, it is not surprising that *B. papendorfii* UM 226 contains genes encoding HST. Interestingly, UM226_5568 shared a homology (69% identical) with the HC-toxin synthetase from *Pyrenopeziza tritici-repentis* (strain Pt-1C-BFP). As noted previously, the NPS4 encoded by UM226_5568 might be responsible for producing HC-toxin. In addition, four putative HC-toxin efflux carrier TOXAs were identified in the genome, including UM226_2212 (77% identical), UM226_5722 (74% identical), UM226_9947 (55% identical) and UM226_6817 (50% identical). These putative HC-toxin efflux carrier TOXAs have a potential role in self-protection against secreted HC-toxin.

Melanin production, a hallmark of dematiaceous fungi, is responsible for the protection of the fungi from UV irradiation, high temperatures as well as oxidants. However, this PKS18, encoding a PKS for melanin biosynthesis, is conserved in all *Cochliobolus* plant pathogens. However, this PKS18 gene was not identified in *B. papendorfii* UM 226, implying that there must be other genes, such as PKS1 (UM226_1411, 93% identical), PKS2 (UM226_3388, 84% identical), PKS4 (UM226_8230, 79% identical), PKS7 (UM226_6241, 65% identical), PKS9 (UM226_7130, 77% identical) and PKS11 (UM226_7830, 85% identical) that can be responsible for melanin biosynthesis. Among these, PKS1 (UM226_1411), BRNI (UM226_1408, 98%) and transcription factor CMRI (UM226_1409, 84%) are located in a close group, and their mutual orientation is the same as found in *C. heterostrophus* and *A. brassicicola*. Moreover, two important melanin biosynthetic genes, scytalone dehydratases (SCD1), UM226_59, (98% identical) and UM226_7826 (65% identical), were identified. PKS1, BRNI and SCD1 are the key enzymes involved in the 1,8-dihydroxynaphthalene (DHN) biosynthesis pathway for melanin biosynthesis. Transcription factor CMRI plays an important role to control *BRNI* and *SCD1* genes expression during melanin production.15 Eliahu et al. indicated that the expression of CMRI is regulated by two mitogen-activated protein kinases (MAPKs) known as Mps1 and Chk1 for normal pigmentation; both enzymes were found in our genome. Two Mps1 (UM226_10400, 95% identical and UM226_10401, 93% identical) and two Chk1 (UM226_10509, 87% identical and UM226_10016, 50% identical) were identified. Overall, these findings suggested that like most of the dematiaceous fungi, melanin is biosynthesized via the DHN-melanin pathway, which is mediated by the transcription factor CMRI and two signaling components, Mps1 and Chk1.

Terpenoids, also referred to as isoprenoids, are a diverse and the largest group of natural bioactive compounds. Many terpenoids such as the anticancer drug paclitaxel and the antimarial drug Artemisinin are well recognized for human disease therapy. In this study, we identified 10 core enzymes involved in the mevalonate (MVA) pathway for terpenoid biosynthesis (Supplementary Table S8). The enzymes 3-hydroxy-3-methylglutaryl-coenzyme A reductase and acetyl-CoA C-acetyltransferase are encoded by two and four copies of the genes, respectively. The remaining eight enzymes are encoded by single-copy genes. Taken together, terpenoid backbone biosynthesis in *B. papendorfii* UM 226 might occur mainly through the MVA pathway.

Caffeine (1,3,7-trimethylxanthine), a purine alkaloid, is one of the world’s most popular psychoactive drug in pharmaceuticals, with an estimated consumption of 120,000 tons worldwide per year. Caffeine is a large environmental concern because of the production of millions tons of caffeine-containing waste products every year. There are few publication about microorganisms capable of caffeine degradation have been isolated from different environments. Sauer et al. indicated that caffeine was degraded by cytochrome P450 enzymes in yeast, postulating that the catabolic pathway might be more similar to that in animals. In this work, we searched the *B. papendorfii* UM226 genome for potential genes involved in caffeine metabolism. Two cytochrome P450 enzymes (UM226_867, 78% identical and UM226_7104, 39% identical) were identified in the genome based on the KEGG metabolic pathway analysis. In addition, we also found two cytochrome P450 2E1 (UM226_1044, 81% identical and UM226_4157, 79% identical) that catalyze caffeine metabolism via N-demethylation pathway in human liver cells. The production of methylxanthines during the course of caffeine degradation can be further converted to the methylic acids by xanthine oxidase. In this work, two xanthine oxidases (UM226_7607, 92% identical and UM226_7880, 86% identical) were found in the *B. papendorfii* UM 226 genome. Moreover, we also identified an uricase (UM226_10622, 83% identical) that might responsible to convert free xanthine into simpler compounds. This suggests that *B. papendorfii* UM 226 may be a potential caffeine-resistant strain of fungus.

In this study, we successfully generated a high-quality draft genome sequence of *B. papendorfii* UM 226 using hybrid assembly of Illumina and PacBio sequencing technologies. This is the first genome assembly of *B. papendorfii*, which allows us to carry out phylogenetic and comparative genomic analyses. Phylogenomic analysis revealed that *B. papendorfii* UM 226 tightly clustered with *B. lunatus* m118, which is consistent with the multilocus phylogenetic analysis. However, at this stage of knowledge, the long standing confusion in the taxonomic grouping of *Bipolaris*, *Curvularia* and *Cochliobolus* remains unsolved. In this work, we identified a unique gene encoding LysM-containing protein. The LysM-containing protein has three LysM domains which are predicted to bind chitin to perturb chitin-induced immunity from the host. Furthermore, *B. papendorfii* UM226 are enriched with putative DMAT, PKS and NRPS encoding
genes. The presence of these key enzymes will serve as a landmark for future discovery of bioactive molecules as well as mycotoxins. We hope that the dissection of the \textit{B. papendorfi} UM226 genome sequence and in-depth analysis of the genome content will help to decipher its basic biology and provide a valuable platform for future fungal research.

4. \textbf{Availability}

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession JXCC00000000. The version described in this paper is version JXCC01000000.

\section*{Supplementary data}

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

\section*{Funding}

This study was supported by High Impact Research MoE Grant UM.C/625/1/HIR/MOHE/MED/31 (No. H-20001-00-E000070). Funding to pay the Open Access publication charges for this article was provided by High Impact Research MoE Grant UM.C/625/11/HIR/MOHE/MED/31 (No. H-20001-00-E000070).

\section*{References}

1. da Cunha, K.C., Sutton, D.A., Fothergill, A.W., et al. 2012, Diversity of \textit{B. papendorfi} species in clinical samples in the United States and their antifungal susceptibility profiles, \textit{J. Clin. Microbiol.}, 50, 4061–6.

2. Manamgoda, D.S., Cai, L., McKenzie, E.H.C., et al. 2012, A phylogenetic and taxonomic re-evaluation of the \textit{Bipolaris} - \textit{Cochliobolus} - \textit{Carvudaria} complex, \textit{Fungal Divers.}, 56, 131–44.

3. Manamgoda, D.S., Cai, L., Bahkali, A.H., Chukeatirote, E. and Hyde, K.D. 2011, \textit{Cochliobolus}: an overview and current status of species, \textit{Fungal Divers.}, 51, 3–42.

4. Tatum, L.A. 1971, \textit{The southern corn leaf blight epidemic}, \textit{Science}, 171, 1113–6.

5. Scheffer, R.P. 1997, \textit{The nature of disease in plants}. Cambridge University Press; New York.

6. Condon, B.J., Leng, Y., Wu, D., et al. 2013, Comparative genome structure, secondary metabolite, and effector coding capacity across \textit{Cochliobolus} pathogens, \textit{PLoS Genet.}, 9, e1003233.

7. Brandt, M.E. and Warnock, D.W. 2003, Epidemiology, clinical manifestations, and therapy of infections caused by dematiaceous fungi, \textit{J. Chemother.}, 15(Suppl. 2), 36–47.

8. de Hoog, G.S., Guarro, J., Gené, J. and Figueras, M.J. 2005, \textit{Atlas of clinical fungi}, 2nd edn. Centraalbureau Voor Schimmelcultures: Baarn, the Netherlands.

9. Li, P.P. and Cao, Z.Y. 2013, First report of \textit{Bipolaris papendorfi} causing corn leaf spot in China, \textit{Plant Dis.}, 97, 1506.

10. Giridharan, P., Perekar, S.A., Gohil, A.R., Mishra, P.D., Khanna, A. and Deshmukh, S.K. 2014, Antiproliferative activity of hamagemore and radicnol isolated from \textit{Bipolaris papendorfi}, \textit{Biomed. Res. Int.}, 2014, 890904.

11. Yew, S.M., Chan, C.L., Lee, K.W., et al. 2014, A five-year survey of dematiaceous fungi in a tropical hospital reveals potential opportunistic species, \textit{PLoS ONE}, 9, e104352.

12. Moretti, S., Armougom, F., Wallace, L.M., Higgins, D.G., Jongeneel, C.V. and Notredame, C. 2007, The M-Coffee web server: a meta-method for computing multiple sequence alignments by combining alternative alignment methods, \textit{Nucleic Acids Res.}, 35, W645–8.

13. Ronquist, F. and Huelsenbeck, J.P. 2003, MrBayes 3: Bayesian phylogenetic inference under mixed models, \textit{Bioinformatics}, 19, 1572–4.

14. Moslem, M.A., Bahkali, A.H., Abd-Elsalam, K.A. and Win, P.J. 2010, An efficient method for DNA extraction from \textit{Cladosporium fungi}, \textit{Genet. Mol. Res.}, 9, 2283–91.

15. Lee, H., Gurtowski, J., Yoo, S., Marcus, S., McCombie, W.R. and Schatz, M. 2014, Error correction and assembly complexity of single molecule sequencing reads, Cold Spring Harbor Laboratory, 1–17.

16. Kelley, D.R., Schatz, M.C. and Salzberg, S.L. 2010, Quake: quality-aware detection and correction of sequencing errors, \textit{Genome Biol.}, 11, R116.

17. Myers, E.W., Sutton, G.G., Delcher, A.L., et al. 2000, A whole-genome assembly of \textit{Drosophila}, \textit{Science}, 287, 1926–204.

18. Lomnadeaz, A., Ter-Hovhannisyan, V., Chernoff, Y.O. and Borodovsky, M. 2005, Gene identification in novel eukaryotic genomes by self-training algorithm, \textit{Nucleic Acids Res.}, 33, 6494–506.

19. Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M. and Robles, M. 2005, Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, \textit{Bioinformatics}, 21, 3674–6.

20. Tatusov, R.L., Fedorova, N.D., Jackson, J.D., et al. 2003, The COG database: an updated version includes eukaryotes, \textit{BMC Bioinformatics}, 4, 41.

21. Quevillon, E., Silventoinen, V., Pillai, S., et al. 2005, InterProScan: protein domains identifier, \textit{Nucleic Acids Res.}, 33, W116–20.

22. Lagesen, K., Hallin, P., Rodland, E.A., Staerfeldt, H.H., Rognes, T. and Ussery, D.W. 2007, RNAmmer: consistent and rapid annotation of ribosomal RNA genes, \textit{Nucleic Acids Res.}, 35, 3100–8.

23. Lowe, T.M. and Eddy, S.R. 1997, \textit{fRNAscan-SE}: a program for improved detection of transfer RNA genes in genomic sequence, \textit{Nucleic Acids Res.}, 25, 955–64.

24. Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F. and Xu, Y. 2012, dbCAN: a web resource for automated carbohydrate-active enzyme annotation, \textit{Nucleic Acids Res.}, 40, W445–51.

25. Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. 2011, \textit{SignalP} 4.0: discriminating signal peptides from transmembrane regions, \textit{Nat. Methods}, 8, 785–6.

26. Krogh, A., Larsson, B., von Heijne, G. and Sonnhammer, E.L. 2001, Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes, \textit{J. Mol. Biol.}, 305, 567–80.

27. Khaledi, N., Seffudin, F.T., Turner, G., et al. 2010, \textit{SMURF}: Genomic mapping of fungal secondary metabolite clusters, \textit{Fungal Genet. Biol.}, 47, 736–41.

28. Ohm, B.A., Feau, N., Hemriassat, B., et al. 2012, Diverse lifestyles and strategies of plant pathogenesis encoded in the genomes of eight Dothideomycetes fungi, \textit{PLoS Pathog.}, 8, e1003037.

29. Dean, R.A., Ann, L.P. and Chittaranjan, K. 2014, \textit{Genomics of plant-associated fungi: monoco Gutbacter}, Springer: Berlin, Heidelberg, pp. 41–67.

30. Buchan, D.W., Ward, S.M., Lobley, A.E., Nugent, T.C., Broyson, K. and Jones, D.T. 2010, Protein annotation and modelling servers at University College London, \textit{Nucleic Acids Res.}, 38, W563–8.

31. Basini, M., Bienert, S., Waterhouse, A., et al. 2014, \textit{SWISS-MODEL} : modelling protein tertiary and quaternary structure using evolutionary information, \textit{Nucleic Acids Res.}, 42, W252–8.

32. Li, L., Stoeckert, C.J., Jr and Roos, D.S. 2003, \textit{OrthoMCL}: identification of ortholog groups for eukaryotic genomes, \textit{Genome Res.}, 13, 2178–89.

33. Larkin, M.A., Blackshields, G., Brown, N.P., et al. 2007, \textit{Clustal W} and \textit{Clustal X} version 2.0, \textit{Bioinformatics}, 23, 2947–8.

34. Sanchez-Vallet, A., Saleem-Batcha, R., Kombrink, A., et al. 2013, Fungal effector \textit{Ecp6} outcompetes host immune receptor for chitin binding through intrachain LysM dimerization, \textit{Elife}, 2, e00790.

35. Guox, N. and Peitsch, M.C. 1997, \textit{SWISS-MODEL} and the Swiss-PdbViewer: an environment for comparative protein modeling, \textit{Electrophoresis}, 18, 2714–23.

36. Kuan, C.S., Tien, C.F., See Too, W.C. and Few, L.L. 2012, Sequence analysis and homology modeling of glyceraldehyde-3-phosphate dehydrogenase of an antarctic psychrophilic \textit{Pseudomonas sp}, \textit{Afr. J. Microbiol. Res.}, 6, 881–91.
37. Rutherford, K., Parkhill, J., Crook, J., et al. 2000. Artemis: sequence visualization and annotation, Bioinformatics, 16, 944–5.
38. Santana, M.F., Silva, J.C., Mizabuti, E.S., et al. 2014. Characterization and potential evolutionary impact of transposable elements in the genome of Cochliobolus heterostrophus, BMC Genomics, 15, 536.
39. Novikova, O., Fet, V. and Blinov, A. 2009. Non-LTR retrotransposons in fungi, Funct. Integr. Genomics, 9, 27–42.
40. Muszewska, A., Hoffmann-Sommer, M. and Gryenberg, M. 2011. LTR retrotransposons in fungi, PLoS ONE, 6, e29425.
41. Brillet, B., Bigot, Y. and Auge-Gouillou, C. 2007. Assembly of the Tc1 and mariner transposition initiation complexes depends on the origins of their transposase DNA binding domains, Genetica, 130, 103–20.
42. Yuan, Y.W. and Wessler, S.R. 2011. The catalytic domain of all eukaryotic cut-and-paste transposase superfamily, Proc. Natl Acad. Sci. USA, 108, 7884–9.
43. Leach, M.D. and Brown, A.J. 2012. Posttranslational modifications of proteins in the pathobiology of medically relevant fungi, Eukaryot. Cell, 11, 98–108.
44. Khan, I.A., Lu, J.P., Liu, X.H., Rehman, A. and Lin, F.C. 2012. Multifunction of autophagy-related genes in filamentous fungi, Microbiol. Res., 167, 339–45.
45. Zhao, Z., Liu, H., Wang, C. and Xu, J.R. 2014. Correction: comparative analysis of fungal genomes reveals different plant cell wall degrading capacity in fungi, BMC Genomics, 15, 6.
46. Blackman, L.M., Cullerne, D.P. and Hardham, A.R. 2014. Bioinformatic characterisation of genes encoding cell wall degrading enzymes in the Phytophthora parasitica genome, BMC Genomics, 15, 785.
47. Mitreva-Dautova, M., Rose, E., Overmars, H., et al. 2006. A symbiont-independent endo-1,4-beta-xylanase from the plant-parasitic nematode Meloidogyne incognita, Mol. Plant Microb. Interact., 19, 521–9.
48. Ahmed, S., Riaz, S. and Jamil, A. 2009. Molecular cloning of fungal xylanases: an overview, Appl. Microbiol. Biotechnol., 84, 19–35.
49. Levasseur, A., Salohermo, M., Navarro, D., et al. 2010. Exploring laccase-like multicopper oxidase genes from the ascomycete Trichoderma reesei: a functional, phylogenetic and evolutionary study, BMC Biochem., 11, 32.
50. Piontek, K., Smith, A.T. and Blodiag, W. 2001. Lignin peroxidase structure and function, Biochem. Soc. Trans., 29, 111–6.
51. Kombrik, A. and Thomma, B.P. 2013. LysM effectors: secreted proteins supporting fungal life, PLoS Pathog., 9, e1003769.
52. Bolton, M.D., van Esse, H.P., Vossen, J.H., et al. 2008. The novel Cladosporium fulvum lysis motif effector Ecp6 is a virulence factor with orthologues in other fungal species, Mol. Microbiol., 69, 119–36.
53. de Jonge, R. and Thomma, B.P. 2009. Fungal LysM effectors: extinguishers of host immunity? Trends Microbiol., 17, 151–7.
54. Turgeon, B.G. 1998. Application of mating type gene technology to problems in fungal biology, Annu. Rev. Phytopathol., 36, 115–37.
55. Keller, N.P., Turner, G. and Bennett, J.W. 2005. Fungal secondary metabolism - from biochemistry to genomics, Nat. Rev. Microbiol., 3, 937–47.
56. Metzger, U., Schall, C., Zocher, G., et al. 2009. The structure of dimethylallyl tryptophan synthase reveals a common architecture of aromatic prenyltransferases in fungi and bacteria, Proc. Natl Acad. Sci. USA, 106, 14309–14.
57. Scharl, C.L., Panaccione, D.G. and Tudzynski, P. 2006. Ergot alkaloids–biology and molecular biology, Alkaloids Chem. Biol., 63, 45–86.
58. Turgeon, B.G. and Baker, S.E. 2007. Genetic and genomic dissection of the Cochliobolus heterostrophus Tox1 locus controlling biosynthesis of the polyketide virulence factor T-toxin, Adv. Genet., 57, 219–61.
59. Dadachova, E., Bryan, R.A., Huang, X., et al. 2007. Ionizing radiation changes the electronic properties of melanin and enhances the growth of melanized fungi, PLoS ONE, 2, e457.
60. Elahu, N., Itagiba, A., Rose, M.S., Horwitz, B.A. and Lev, S. 2007. Melanin biosynthesis in the maize pathogen Cochliobolus heterostrophus depends on two mitogen-activated protein kinases, Chk1 and Mps1, and the transcription factor Cmr1, Eukaryot. Cell, 6, 421–9.
61. Tsuji, G., Kenmochi, Y., Takano, Y., et al. 2000. Novel fungal transcriptional activators, Cmr1p of Colletotrichum lagenarium and pgp1p of Magnaporthe grisea, contain Cys2His2 zinc finger and ZnH2Cys6 binuclear cluster DNA-binding motifs and regulate transcription of melanin biosynthesis genes in a developmentally specific manner, Mol. Microbiol., 38, 940–54.
62. Neuman, D.J., Cragg, G.M. and Snader, K.M. 2000. The influence of natural products on drug discovery, Nat. Prod. Rep., 17, 215–34.
63. Gilbert, R.M. 1984. Caffeine consumption, Prog. Clin. Biol. Res., 158, 185–213.
64. Dash, S.S. and Gummaditi, S.N. 2006. Catabolic pathways and biotechnological applications of microbial caffeine degradation, Biotechnol. Lett., 28, 1993–2002.
65. Mazzafera, P. 2004. Catalolism of caffeine in plants and microorganisms, Front Biosci., 9, 1348–59.
66. Sauer, M. 1982. Comparison of the cytochrome P-450 containing monooxygenases originating from two different yeasts, Dev. Biochem., 23, 452–7.