Genome Sequencing and Comparative Genomics Analysis Revealed Pathogenic Potential in *Penicillium capsulatum* as a Novel Fungal Pathogen Belonging to *Eurotiales*

Ying Yang[1,2,3†], Min Chen[4,5,6†], Zongwei Li[7†], Abdullah M. S. Al-Hatmi[5,6,8], Sybren de Hoog[5,6], Weihua Pan[4], Qiang Ye[3,9], Xiaochen Bo[2], Zhen Li[2], Shengqi Wang[2], Junzhi Wang[3]*, Huipeng Chen[1]* and Wanqing Liao[4]*

[1] Beijing Institute of Biotechnology, Beijing, China, [2] Beijing Institute of Radiation Medicine, Beijing, China, [3] National Institutes for Food and Drug Control, Beijing, China, [4] Department of Dermatology, Shanghai Key Laboratory of Molecular Medical Mycology, Shanghai Institute of Medical Mycology, Shanghai Changzheng Hospital, Shanghai, China, [5] CBS-KNAW Fungal Biodiversity Centre, Utrecht, Netherlands, [6] Institute of Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, Netherlands, [7] Center for Hospital Infection Control, Chinese PLA Institute for Disease Control and Prevention, Beijing, China, [8] Directorate General of Health Services, Ibri Hospital, Ministry of Health, Ibri, Oman, [9] Key Laboratory of the Ministry of Health for Research on Quality and Standardization of Biotech Products, Beijing, China

Penicillium capsulatum is a rare *Penicillium* species used in paper manufacturing, but recently it has been reported to cause invasive infection. To research the pathogenicity of the clinical *Penicillium* strain, we sequenced the genomes and transcriptomes of the clinical and environmental strains of *P. capsulatum.* Comparative analyses of these two *P. capsulatum* strains and close related strains belonging to *Eurotiales* were performed. The assembled genome sizes of *P. capsulatum* are approximately 34.4 Mbp in length and encode 11,080 predicted genes. The different isolates of *P. capsulatum* are highly similar, with the exception of several unique genes, INDELs or SNPs in the genes coding for glycosyl hydrolases, amino acid transporters and circumsporozoite protein. A phylogenomic analysis was performed based on the whole genome data of 38 strains belonging to *Eurotiales*. By comparing the whole genome sequences and the virulence-related genes from 20 important related species, including fungal pathogens and non-human pathogens belonging to *Eurotiales*, we found meaningful pathogenicity characteristics between *P. capsulatum* and its closely related species. Our research indicated that *P. capsulatum* may be a neglected opportunistic pathogen. This study is beneficial for mycologists, geneticists and epidemiologists to achieve a deeper understanding of the genetic basis of the role of *P. capsulatum* as a newly reported fungal pathogen.

Keywords: *Penicillium capsulatum*, novel fungal pathogen, genome sequencing, comparative genomics

INTRODUCTION

The term "novel fungal pathogens" has traditionally referred to those non-pathogenic fungi that have been shown to cause human infection, and the number of these pathogens has steadily increased during the last 20 years (Jones et al., 2008). However, the majority of these fungi have been frequently neglected after being reported, although fungal pathogens have recently been recognized as a significant threat to public health (Morens et al., 2004; Garcia-Solache and Casadevall, 2010). Current unprecedented declines in biodiversity may actually increase the risk of fungal infection in mammals, and the potential hazard of novel fungal pathogens cannot be neglected (Keesing et al., 2010; Fisher et al., 2012). Virtually all of these fungal pathogens have been identified using morphological, physiological and multiple locus sequence typing (MLST) analyses (Struelens and Brisse, 2013; Samson et al., 2014). These approaches can identify these fungal pathogens at the species or even the genotype level. However, it is a challenge to investigate the possible causes of variations in the virulence and the evolution, and transmission of novel fungal pathogens using routine identification techniques.

Recently, the application of next-generation sequencing (NGS) technology has allowed us to conduct whole genome scans and to gain insight into the virulence factors and the adaptation of fungal pathogens to human hosts (Pareek et al., 2011; Ferrarini et al., 2013). The second-generation platform Illumina sequencing can provide accurate, high-throughput data with greater sequencing depths, whereas the third generation sequencing platform PacBio RS II can offer increased read lengths and unbiased genome coverage (Ferrarini et al., 2013). Combining the two platforms will produce better sequencing results with increased efficiency. These developments can provide additional insights for obtaining genetic information in novel fungal pathogens to develop new tools for the detection of emergence and the formulation of appropriate responses toward infections, such as infection control. The high levels of variation and unique genomic features of different isolates that are relevant to virulence have been increasingly reported in multiple NGS and comparative studies of human pathogens (Lavezzo et al., 2013). Recent novel phylogenetic findings based on MLST analyses– such as the identification of eurotiomycetous endophytes, which showed close affinities to *Chaetothyriales*, *Eurotiales*, and the characterization of a new order, *Phaeomoniellales*-still required further phylogenetic proof based on additional genomic information (Chen et al., 2013, 2015). However, similar studies in fungal pathogens have been rare to date (Lavezzo et al., 2013).

Species of *Eurotiales* are common and important to both industry and medicine. Some species, such as *Aspergillus fumigatus*, are well known as human pathogens. *Penicillium* is an important genus belonging phylogenetically to *Eurotiales*. Over 350 species are currently defined in the genus, the majority of which are saprobic and commonly occur in soil. Invasive infections caused by *Penicillium* species are currently rare, and such reports will decrease because *Penicillium marneffei* is now known as *Talaromyces marneffei* after its recent transfer to the genus *Talaromyces*, which belongs to the subgenus *Biverticillium* (Samson et al., 2011). *P. capsulatum*, a rare *Penicillium* species

frequently used in the paper-making industry, has never been recognized as a human pathogen (Lyratzopoulos et al., 2002; Houbraken and Samson, 2011). In 2013, we first reported a fungus ball in the left lung of a type 2 diabetic patient caused by *P. capsulatum* (Chen et al., 2013). Routine molecular identification techniques, such as ITS, calmodulin and *RPB2* sequencing (Samson et al., 2014), are insufficient to reveal the hidden genetic basis of the probable cause of novel fungal pathogens such as *P. capsulatum*, and most of these pathogens were reported and neglected.

To further investigate the genomic basis of differences in the pathogenicity of *P. capsulatum* species, we performed genome sequencing and comparative genomics of the clinical (CBS 134186) and environmental (ATCC 48735) strains of *P. capsulatum* using a combination of the second and third generation sequencing platforms. A comparative transcriptome analysis of these two strains was performed using RNA-Seq data. In addition, a new phylogenetic analysis was performed in *Eurotiales* based on the fungal species with available genome data, including *P. capsulatum*. Moreover, we thoroughly researched the potential virulence factors in *P. capsulatum* and its closely related strains, which include both fungal pathogens and non-human pathogens. Our study not only provided additional insights into the pathogenic potential of *P. capsulatum* as an example of novel fungal pathogens belonging to *Eurotiales* but also more accurately explored the phylogenetic placements of *Eurotiales* species.

MATERIALS AND METHODS

Isolates

The clinical strain of *P. capsulatum* (LiaoWQ-2011 = CBS 134186) was isolated from a pulmonary fungus ball in a patient with type 2 diabetes in 2013, and the environmental strain of *P. capsulatum* (ATCC 48735) was isolated from exposed canvas on Gilbert Island in 1945. The representative isolates were included for whole genome sequence analyses in this study, as described in Supplementary Table S1.

Genome DNA and RNA Extraction

The representative strains for *P. capsulatum* (i.e., the clinical strain CBS 134186, and the environmental strain ATCC 48735) were grown on potato dextrose agar (PDA) at 25°C for 14 days. Genomic DNA for whole-genome sequencing was extracted from the pure cultures using the DNeasy Plant Mini kit (QIAGEN Co., Ltd, Hamburg, Germany) according to the instruction manual. RNA for transcriptome sequencing was extracted using the RNeasy Plant Mini kit (QIAGEN Co., Ltd, Hamburg, Germany) following the protocol for the isolation of the total RNA.

Genome Sequencing

Our strategy for whole-genome sequencing involved a combination of the Illumina MiSeq (Illumina Inc., San Diego, CA, USA) and the Pacific Biosciences RS II (Pacific Biosciences, Menlo Park, CA, USA) sequencing platforms, and the sequence data from the Illumina platform were used to proofread the PacBio assembly sequence. Illumina libraries were prepared

using TruSeq DNA sample prep kits (Illumina Inc., San Diego, CA, USA) according to the manufacturer's instructions. Two paired-end sequencing libraries were constructed. One was a 250 Pair End (PE) sequencing library with an insert sizes of approximately 400 bp, and the other was a 150PE library with insert sizes ranging from 200 bp to 1 kb. Pacific Biosciences RS II sequencing technology was used as the sequencing platform. A 10-kb Single Molecule Real Time (SMRT) bell library was prepared from sheared genomic DNA via a 10-kb template library preparation workflow. SMRT sequencing was conducted on the PacBio RS II sequencing platform using C3 sequencing chemistry and P5 polymerase with 16 SMRT cells.

Genome Assembly

The Hierarchical Genome Assembly Process (HGAP) was used to assemble these two sequenced genomes. *De novo* assembly of the PacBio read sequences was carried out using continuous long reads (CLR), followed by the HGAP workflow (PacBioDevNet; Pacific Biosciences) as available in SMRT Analysis v2.1. HGAP consists of preassembly, *de novo* assembly with Celera Assembler (CA), and assembly polishing with Quiver. CA software version 7.0 was utilized in the pre-assembly step, and the PacBioRs_PreAssembler with one module and a default minimum subread length of 500 bp, a minimum read quality of 0.80, and a minimum subread length of 7500 bp was used to perform error correction for the raw data generated by the PacBio RS II platform. To polish the assembled sequence from HGAP, the MiSeq read sequences were mapped using BWA v0.7.5a (Li and Durbin, 2009, 2010), and the SNPs and INDELs were called and corrected by SAMtools v0.1.18 (Li et al., 2009) and an in-house script. The genomes of *P. capsulatum* strains CBS 134186 and ATCC 48735 were deposited in Genbank under the accession numbers JPLR00000000 and JPLQ00000000.

Prediction and Annotation of Gene Structure

InterProScan was used to obtain functional annotations for all predicted genes and to determine motifs and domains. To obtain high confidence gene models for *P. capsulatum*, coding genes were predicted with AUGUSTUS v3.1 (Stanke et al., 2008). All of the predicted gene models were functionally annotated based on their sequence similarity to genes and proteins in the NCBI nucleotide (Nt), non-redundant and UniProt/Swiss-Prot protein databases. The gene models were also annotated based on their protein domains using InterProScan (Jones et al., 2014). All genes were classified according to Gene Ontology (GO), eukaryotic orthologous groups (KOG), and Kyoto Encyclopedia of Genes and Genomes (KEGG) metabolic pathways. Repeat sequences were masked throughout the genome using Repeat Masker v3.2.9 and the RepBase library v16.08 (Jurka et al., 2005). tRNAs were identified using tRNA scan-SE (Lowe and Eddy, 1997; Schattner et al., 2005).

Orthology and Phylogenetic Analysis

For analysis of orthologous and unique genes, we used the Markov clustering program OrthoMCL v2.09 (Li et al., 2003) to define a gene family as a group of genes that descended from a single gene present in the last common ancestor of the considered species. The peptide sequences were subjected to an all-versus-all BLASTp search with a threshold value of E ≤ 1e−5 and then clustered by MCL with an inflation value of 1.5. A pipeline was used to cluster individual genes into gene families and to analyze orthologous and unique genes. The pipeline included four main steps: (1) data preparation, (2) BlastP searches for all of the protein sequences against a database containing a protein dataset from all species with E-values under 1E-5 and a 90% Match Cutoff, (3) clustering by MCL, and (4) extracting gene families.

For the phylogenetic analysis based on all 40 strains, alignments of single-copy orthologs were produced using MAFFT v7.221 (Katoh and Standley, 2013). ProtTestv3.4 (Darriba et al., 2011) was utilized to determine the best-fitting model for the data. A maximum likelihood phylogeny was reconstructed using RAxML v8.1.23 (Stamatakis, 2014). For the phylogenetic analysis based on virulence-related genes, 43 virulence-related genes shared among 22 strains were selected for a similar phylogenetic analysis.

Secondary Metabolite Genes and Gene Cluster Prediction

The gene clusters related to secondary metabolite biosynthesis were predicted using antiSMASH. All protein sequences were also searched against a local polyketide synthase (PKS)/non-ribosomal peptide synthetase (NRPS) database (a subset of the SwissProt database) using BlastP with a threshold E-value ≤ 1E−5 to identify PKS and NRPS. The putative PKS/NRPS protein sequences were further searched against the NCBI Conserved Domain Database (V3.09) to confirm that the three typical domains (PKS, MT, and ACP) were present. The consensus sequences of PKS/NRPS predicted by antiSMASH (Blin et al., 2013) and manually identified proteins were of high confidence.

Protein Families

Whole genome blast searches were conducted against protein sequences in the pathogen–host interaction database[1] (E-value ≤ 1E-10) (Urban et al., 2015). In the cytochrome P450s analysis, the reference CYP sequences were downloaded from http://drnelson.uthsc.edu/P450seqs.dbs.html. All predicted proteins were then used to search the reference CYP data set using the BLASTP program with a cutoff E-value ≤ 1E−10. SignalP (Petersen et al., 2011) was used to predict signal peptide cleavage sites in amino acid sequences, and putative protein sequences were also examined using the TMHMM method (Sonnhammer et al., 1998) for transmembrane domains. The predicted signal peptide sequences without transmembrane domains were identified as exocrine proteins.

Comparative Genomics Analysis

To identify significantly different genes between the genomes of strains CBS 134186 and ATCC 48735, all of the annotated genes were subjected to all-versus-all BLASTP searches with

---

[1]http://www.phi-base.org/

a threshold *E*-value $\leq$ 1E–10. Any gene with a coverage rate below 90% was identified as a specific candidate gene. Specific genes were obtained by filtering out genes without a specific annotated function or with any alignments near the gene boundary. Furthermore, comparisons of nucleotide sequences between strains ATCC 48735 and CBS 134186 were carried out for SNP and INDEL detection using the Mauve program with the default values (minimum match length, 20 bp). An in-house script was used to call SNPs and INDELs between strains ATCC 48735 and CBS 134186.

Synteny Analysis and Visualization

Synteny analysis between ATCC 48735 and CBS 134186 was performed using MUMmer v3.0 (Kurtz et al., 2004) and MAUVE v20150226 (Darling et al., 2010). The two strains of *P. capsulatum* were visualized using the Circos program v0.68 (Krzywinski et al., 2009).

Transcriptome Sequencing and Assembly

Poly-A mRNA was isolated using oligo-dT-coupled beads from 40 μg total RNA of each sample, followed by shearing. The isolated RNA samples were used for first strand cDNA synthesis with random hexamers and Superscript II reverse transcriptase. After end repair and the addition of a 3'dA overhang, the cDNA was ligated to the Illumina paired-end adapter oligomix, and approximately 200 bp fragments were size selected by gel purification. After 16 cycles of PCR, the libraries were sequenced using Illumina HiSeq2500 and the paired-end sequencing module. The RNA expression analysis was based on the predicted genes of CBS 134186 and ATCC 48735. First, Tophatv2.0.10 (Kim et al., 2013) was used to map the mRNA reads to the genome, and Cufflinksv2.1.1 (Trapnell et al., 2010) was then used to calculate the expected fragments per kilobase of transcript per million mapped reads (FPKM) as the expression values for each transcript. RNA-Seq data of CBS 134186 and ATCC 48735 are available at National Center for Biotechnology Information (NCBI) under SRR accession numbers of SRR 4031065 and SRR 4051963.

RESULTS

Genome Sequencing, Assembly, and Assessment

The genomes of both *P. capsulatum* strains were successively sequenced by a combination of the Pacific Biosciences RS II and Illumina MiSeq platforms. A total of 4.9 and 4.4 Gb of raw data were generated for strains CBS 134186 and ATCC 48735, respectively, using the PacBio RSIISMRT platform and a 10-kb library preparation. A total of 3.2 and 3.0 Gb of raw data from strains CBS 134186 and ATCC 48735, respectively, were generated using the Illumina platform. The genome size of *P. capsulatum* was estimated to be approximately 35 Mb. Hence, we sequenced the genomes of the two *P. capsulatum* strains with an average coverage of ~231 × for strain CBS 134186

and ~211× for strain ATCC 48735. For strain CBS 134186, the genome size was approximately 34.34 Mb, with a contigN50 size of 3.30 Mb; for strain ATCC 48735, the genome size was approximately 34.37 Mb, with a contigN50 size of 3.19 Mb. The read data from the Illumina platform were used to correct the assembled sequence based on the PacBio platform because the latter normally results in relatively more mistakes (Ferrarini et al., 2013). Fifty SNPs and 23 INDELs were corrected in the genome of strain CBS 134186, whereas 48 SNPs and 29 INDELs were corrected in the genome of strain ATCC 48735. For the corrected SNPs and INDELs, polyN-type errors were the main error type, which accounted for approximately 75 and 73% of the total corrected sides for CBS 134186 and ATCC 48735, respectively. Two different gene footprint coverage methods were applied to validate the coding region coverage of the genome assemblies. The results obtained using the Core Eukaryotic Genes Mapping Approach (CEGMA) showed that 95.56% (237/248) of the core eukaryotic genes that mapped to the assembly genome of CBS 134186 were identified and that 97.18% (241/248) of the eukaryotic genes that mapped to the genome of ATCC 48735 were identified (**Table 1**).

Functional Annotation and Genome Features

For the genome of strain CBS 134186, a total of 11,080 gene models were identified based on Gene Ontology (GO) annotations, with an average coding sequence length of 1807.8 bp. For the ATCC 48735 genome, 11,076 gene models were obtained, with a total length of 34.37 Mb and an average coding sequence length of 1808.7 bp (**Table 1**; Supplementary Table S2). The average GC content in both strains of *P. capsulatum* was 49.06%. Approximately 1.41 Mb of repeated regions were found in the genome of the clinical strain CBS 134186, which accounted for 4.19% of its genome size, whereas 1.42 Mb of repeated regions were found in the genome of the

TABLE 1 | Genome characteristics of *Penicillium capsulatum* strains CBS 134186 and ATCC 48735.

Nuclear genome	CBS 134186	ATCC 48735
Size (Mb)	34.341	34.374
Coverage (fold)	231x	211x
Contigs	62	65
N50 length (bp)	3304607	3188669
N50 contig	5	5
G+C content (%)	49.058%	49.074%
Number of protein-coding genes (> 30aa)	11,080	11076
Average gene length (bp)	1807.8	1808.7
GC content of protein-coding genes (%)	52%	52%
Mean number of exons/gene	3.3	3.3
Average length of exons (bp)	482.3	482.5
Mean number of introns/gene	2.3	2.3
Average length of introns (bp)	102.8	103.3
tRNA number	185	174
Repetitive DNA (%)	4.19 %	4.22%
Genbank accession number	JPLR00000000	JPLQ00000000

ATCC 48735 strain. For both strains, interspersed repeats were the predominant type of repeat region, accounting for 84.9% and 85.1% of the repeat regions in the CBS 134186 and ATCC 48735 genomes, respectively. In addition, tRNAScan-SE predicted 185 and 174 tRNAs in the genomes of CBS 134186 and ATCC 48735, respectively.

Phylogenetic Analysis

A phylogenetic analysis based on the whole genome sequence was conducted to determine the relationship of *P. capsulatum* with other important species belonging to *Eurotiales*. The included species were pre-divided into four categories: *Eurotiales* species known to be medically important fungal pathogens; opportunistic fungal pathogens; species newly reported to cause invasive infection; and species unreported as human pathogens. Except for the sequencing data of *P. capsulatum*, other publicly available and complete genomic data of *Eurotiales* were downloaded from NCBI, including those data for thirteen *Penicillium* strains, fourteen *Aspergillus* strains, five *Talaromyces* strains, two *Neosartorya* strains, one *Rasamsonia* strain and one *Byssochlamys* strain. Two *zygomycete* pathogens, *Rhizopus oryzae* and *Mucor circinelloides*, were selected as taxonomic out-groups (Supplementary Table S1).

The genomes ranged in size from 26.0 to 35.9 Mb within *Penicillium*, from 24.2 to 39.5 Mb within *Aspergillus*, from 28.6 to 37.6 Mb within *Talaromyces* and from 31.7 to 32.2 Mb within *Neosartorya*. The genome sizes of *Rasamsonia emersonii* and *Byssochlamys spectabilis*, which belong phylogenetically to *Eurotiales*, were 28.2 and 29.8 Mb, respectively. Within *Eurotiales*, *A. rambellii*, and *T. stipitatus* carried significantly higher percentages of repetitive elements (13.85 and 12.43%, respectively). The number of repetitive sequences of other strains in *Eurotiales* was relatively low; only 1.01−8.18% of each genome was classified as repetitive. In contrast to the low GC content in the out-group, the genomic GC content of *Eurotiales* ranged from 46 to 53% with a balanced ratio (**Figure 1**).

Subsequently, all of the single-copy orthologs were selected for the phylogenetic analysis. *Eurotiales* were divided roughly into three groups, in which *Penicillium* had a closer phylogenetic relationship to *Aspergillu* than to *Talaromyces*. Fifteen *Penicillium* strains were grouped together in one clade, suggesting their origin from a single progenitor species (**Figure 1**). The *P. capsulatum* strains were confidently placed in a basal position to *P. chrysogenum* and other species belonging to the subgenus *Penicillium*. The clinical strain of *P. capsulatum* showed a very close phylogenetic relationship with the environmental strain, forming a subclade within the *Penicillum* clade. *Talaromyces marneffei,* which was previously known as *P. marneffei*, was grouped in the genus *Talaromyces* based on its genomic data in this study. In addition, *Neosartorya* was classified with *Aspergillu* into one clade according to their genomic characteristics. These results offered us a new perspective based on whole genome data for future phylogenetic classifications.
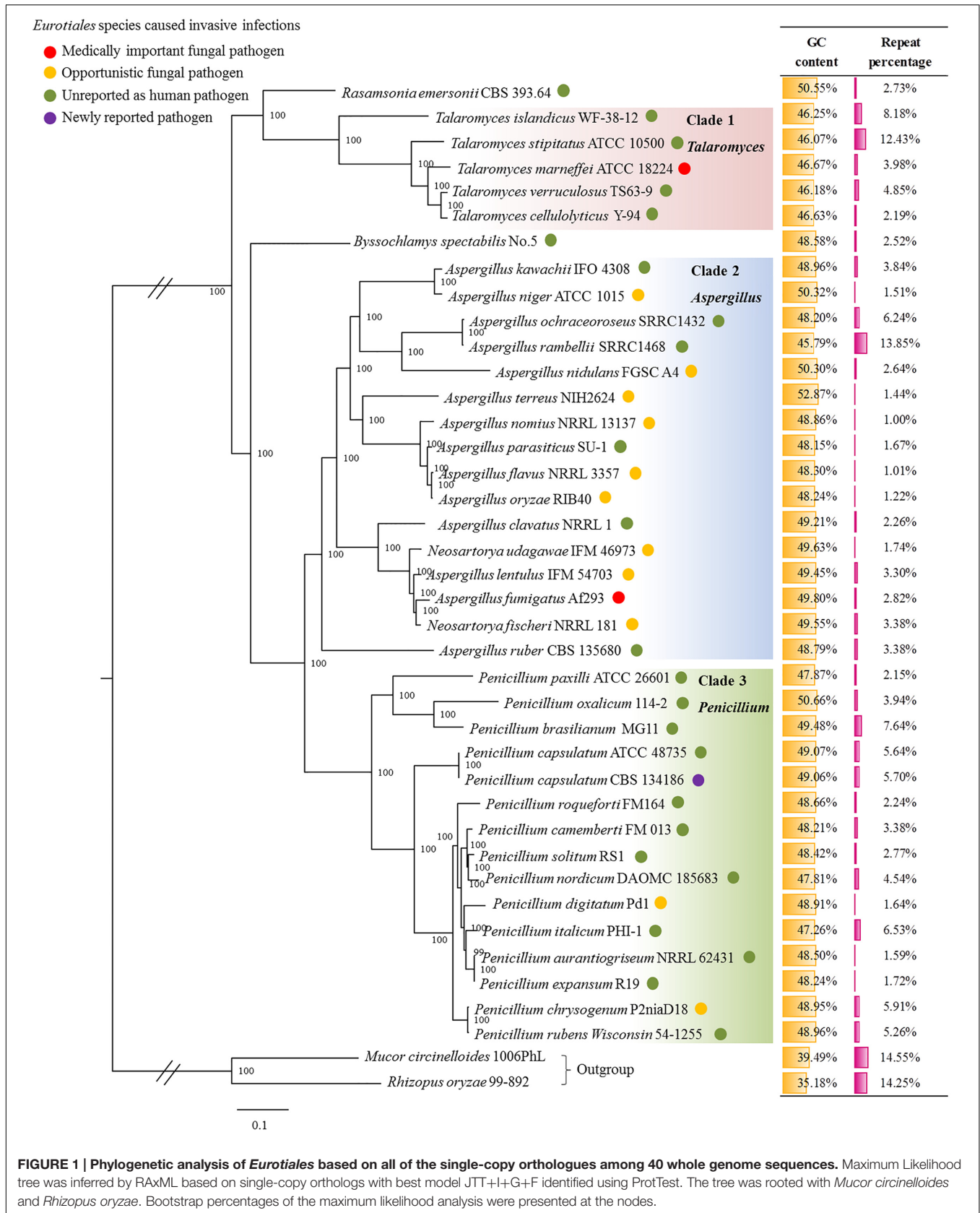
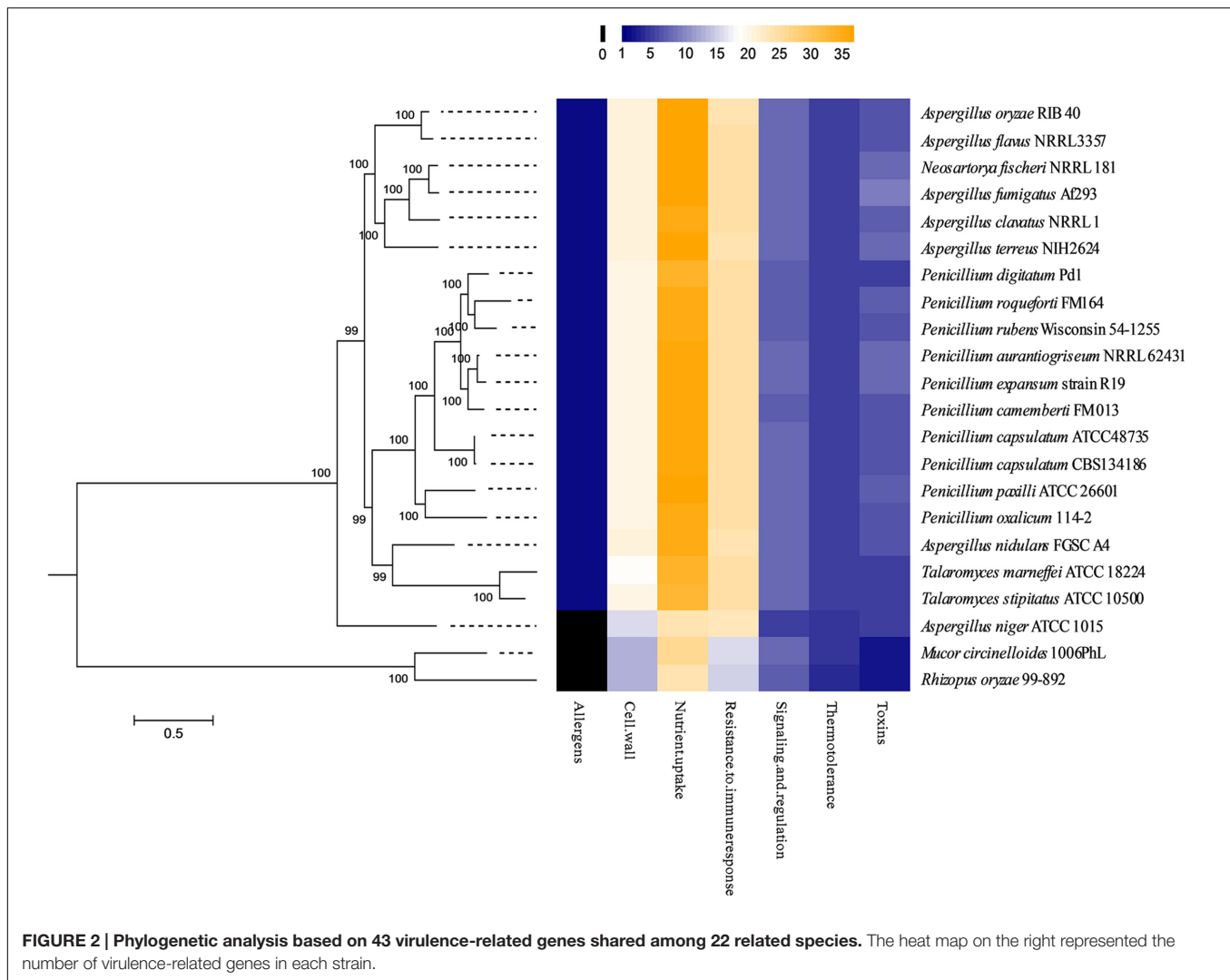## Analysis of Virulence-Related Genes in *P. capsulatum* and Related Species

To further explore the pathogenic factors of *P. capsulatum* and its closely related strains, we performed a phylogenetic analysis based on 106 virulence-related genes identified from medically important fungal pathogens, such as *A. fumigatus*, *T. marneffei et al.* as references for studying the novel fungal pathogens (Supplementary Table S3). Except for two *P. capsulatum* strains, 18 species were selected according to their medical or economic importance and their phylogenetic position, with the intention to obtain a focused representation of *Eurotiales*. The following *Eurotiales* species were included: species known to frequently cause invasive infections as medically important fungal pathogens, including *A. fumigates* and *T. marneffei*; species reported as opportunistic fungal pathogens, including *A. niger*, *A. nidulans*, *A. terreus*, *A. flavus*, *A. oryzae*, *P. chrysogenum*, *P. oxalicum*, *P. digitatum*, and *N. fischeri*; and species unreported as human pathogens, including *T. stipitatus*, *A. clavatus*, *P. camemberti*, *P. paxilli*, *P. aurantiogriseum*, *P. expansum*, and *P. rubens Wisconsin*. *Rhizopus oryzae* and *Mucor circinelloides* were included as taxonomic out-groups.

The presence or absence of virulence-related genes was searched in the 22 studied genomes. We found that a total of 43 virulence-related genes were shared the studied genomes. These genes could be divided into the following six categories: toxins, thermo-tolerance, signaling and regulation, resistance to immune response, cell wall, allergens and nutrient uptake (Supplementary Table S4). Generally, the medically important fungal pathogens have more virulence-related genes than non-human pathogens. In this study, the clinical and environmental isolates of *P. capsulatum* were found to be very similar in terms of their virulence-related genes. Surprisingly, we found in this phylogenic analysis that *P. paxilli*, which is used as a model to study the biochemistry of indol-diterepene biosysnthesis and is seldom reported to be infectious to humans, had a high proportion of virulence-related genes and very close relationship to *P. capsulatum* (**Figure 2**). Thus, we selected *P. paxilli* and two well-known human pathogens, *A. fumigates* and *T. marneffei*, for a more detailed comparison with *P. capsulatum*.

Comparison and Analysis Among the Four Related Strains

Using a Venn diagram of four strains, we compared the *P. capsulatum* genome with a non-human pathogen, *P. paxilli*, which had a close phylogenetic relationship to *P. capsulatum*, as well as *A. fumigates* and *T. marneffei,* two major human pathogens. By analyzing the homologous genes among the four genomes and enriching the virulence genes among these groups, as expected, the majority of homologous gene groups were in the groups of C1&C2&C3&C4. In this shared group among all four strains, most genes were identified as belonging to the cytochrome P450 (CYP) family and the major facilitator superfamily (MFS), which generally play important roles in the biosynthesis and transportation of metabolites. The second largest group of genes shared by these strains was group C2&C3&C4, indicating that the relationships among

FIGURE 1 | Phylogenetic analysis of *Eurotiales* based on all of the single-copy orthologues among 40 whole genome sequences. Maximum Likelihood tree was inferred by RAxML based on single-copy orthologs with best model JTT+I+G+F identified using ProtTest. The tree was rooted with *Mucor circinelloides* and *Rhizopus oryzae*. Bootstrap percentages of the maximum likelihood analysis were presented at the nodes.

FIGURE 2 | Phylogenetic analysis based on 43 virulence-related genes shared among 22 related species. The heat map on the right represented the number of virulence-related genes in each strain.
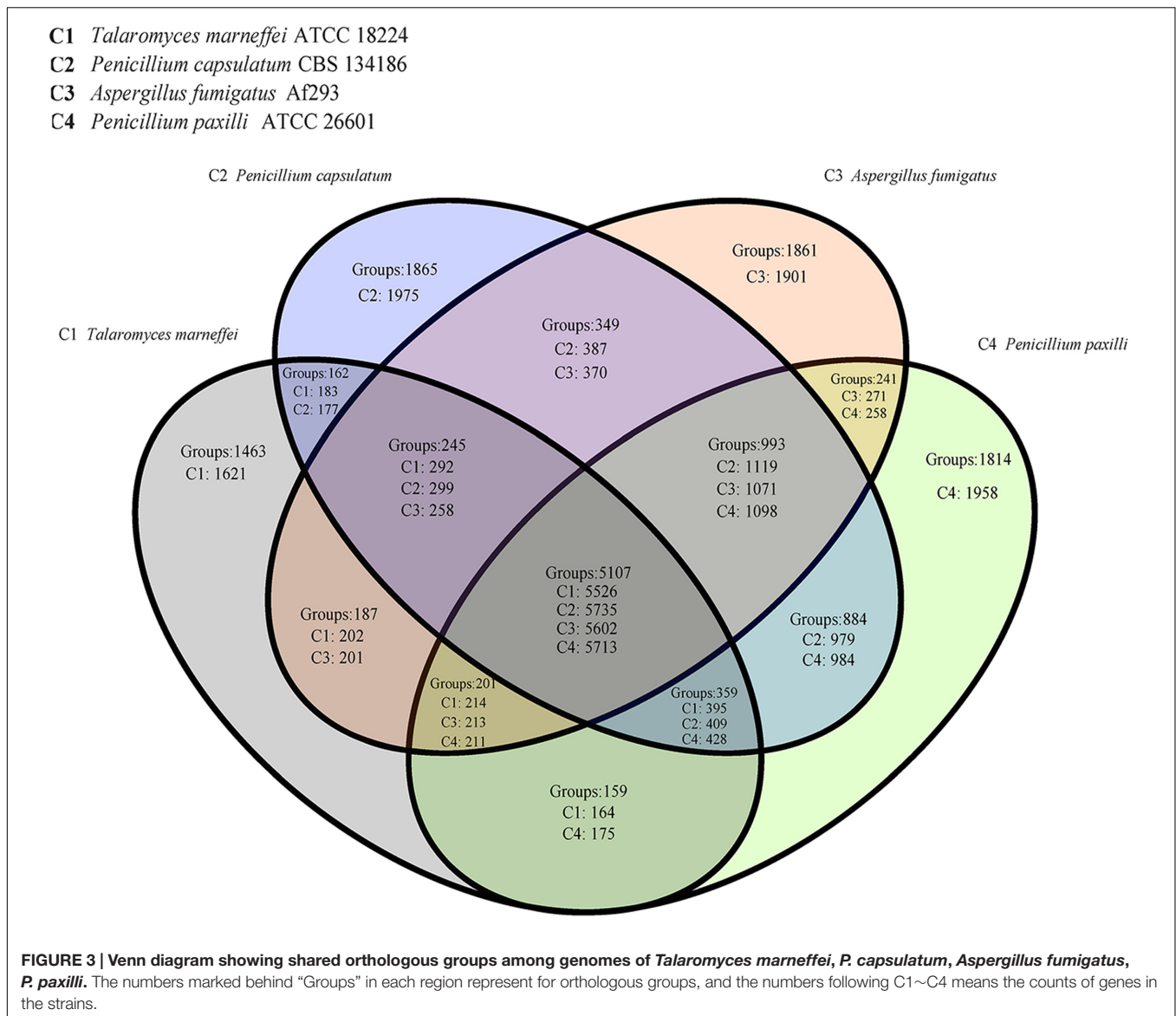
P. capsulatum, *P. paxilli*, and *A. fumigate* were closer than *T. marneffei*. Compared with the other three strains, the unique paralog gene groups of *A. fumigate* in group C3 consisted of typical virulence-related genes, such as genes involved in gliotoxin synthesis (*GliZ*, *Glip*) and fumitremorgin synthesis (*ftmE-H*). Like group C3, group C2 comprised unique paralog gene groups found in *P. capsulatum* CBS 134186. This group had the most unique paralog gene groups, containing 1865 gene groups (1975 predicted genes in total). Some of these genes may increase the pathogenic capacity of these four strains (**Figure 3**, Supplementary Table S5).

By further investigating the virulence genes, we found that only 7 out of 15 groups contained virulence-related genes (Supplementary Table S6). The majority of the virulence genes ($n = 75$) were in the group C1&C2&C3&C4, which had virulence genes among all the four strains. There were 13 virulence factors in group C2&C3&C4, which contained the second greatest number of virulence genes. Interestingly, group C3&C4 contained four virulence genes, which was greater than the number of virulence genes in group C1&C3 and group C2&C3. The enrichment of virulence-related genes in *P. paxilli* indicated the need for paying increased attention to its potential pathogenicity. However, it is possible that having these virulence-related genes is not sufficient to cause disease. Thus, the underlying difference between fungal pathogens and non-pathogens needs further study.

Comparative Genomic Analysis of *P. capsulatum* Strains CBS 134186 and ATCC 48735

The genome of strain CBS 134186 was 0.033 Mb smaller than the genome of ATCC 48735 (**Table 1**). We speculated that the gaps between the large scaffolds in the CBS 134186 genome ($n = 104$) and the ATCC 48735 genome ($n = 101$) may be responsible for the different genome sizes observed between these two strains. Pair-wise genome alignments showed that the two *P. capsulatum* strains were highly syntenic and shared, on average, 99.98% sequence similarity (**Figures 4A,C**). These results suggested that the pathogenic potential of the clinical strain

C1 *Talaromyces marneffei* ATCC 18224
C2 *Penicillium capsulatum* CBS 134186
C3 *Aspergillus fumigatus* Af293
C4 *Penicillium paxilli* ATCC 26601

FIGURE 3 | Venn diagram showing shared orthologous groups among genomes of *Talaromyces marneffei*, *P. capsulatum*, *Aspergillus fumigatus*, *P. paxilli*. The numbers marked behind "Groups" in each region represent for orthologous groups, and the numbers following C1~C4 means the counts of genes in the strains.

CBS 134186 and the environmental strain ATCC 48735 may be extremely similar or that the specific genes, SNPs or INDELs present in the CBS 134186 genome may play a key role in causing invasive infection. Among these probable mutations, a total of seven unique genes, SNPs or INDELs in the CBS134186 genome may be involved in the cause of make itself to be a novel fungal pathogen (**Table 2**, Supplementary Tables S7−S9; **Figure 4B**). Cytochrome P450, secondary metabolism, and exocrine proteins analyses were also performed on these seven genes. One of the unique genes (g512) was predicted as an exocrine protein gene; there were no positive results in the secondary metabolite and cytochrome P450 analyses. The unique genes coding for vegetative incompatibility protein and glycosyl hydrolases (g512 in contig27), the SNP in the gene coding for circumsporozoite protein (g399 in contig17), and the INDEL in the gene containing the F-actin capping protein beta subunit (g8270 in contig44) in the CBS 134186 genome may be related to its adaptation to

stress conditions; a similar phenomenon has been observed in the interaction between *A. fumigates* and humans (Abad et al., 2010).

Comparative Transcriptomic Analysis of Strains CBS 134186 and ATCC 48735

To more accurately annotate the *P. capsulatum* genome, we performed deep transcriptome sequencing of both *P. capsulatum* strains (CBS 134186 and ATCC 48735), which generated 4.0 and 2.9 Gb of RNA sequencing data, respectively. Compared with ATCC 48735, CBS 134186 contained 481 differentially expressed genes and two up-regulated virulence-related genes, which were observed under *in vitro* culture conditions by three replicates to give a general estimate. These virulence-related genes were beta-(1,3) glucanosyltransferase (GEL-3) and mannose-6 phosphate isomerase (MPI), which were up-regulated fivefold and threefold, respectively. Moreover, these two genes are
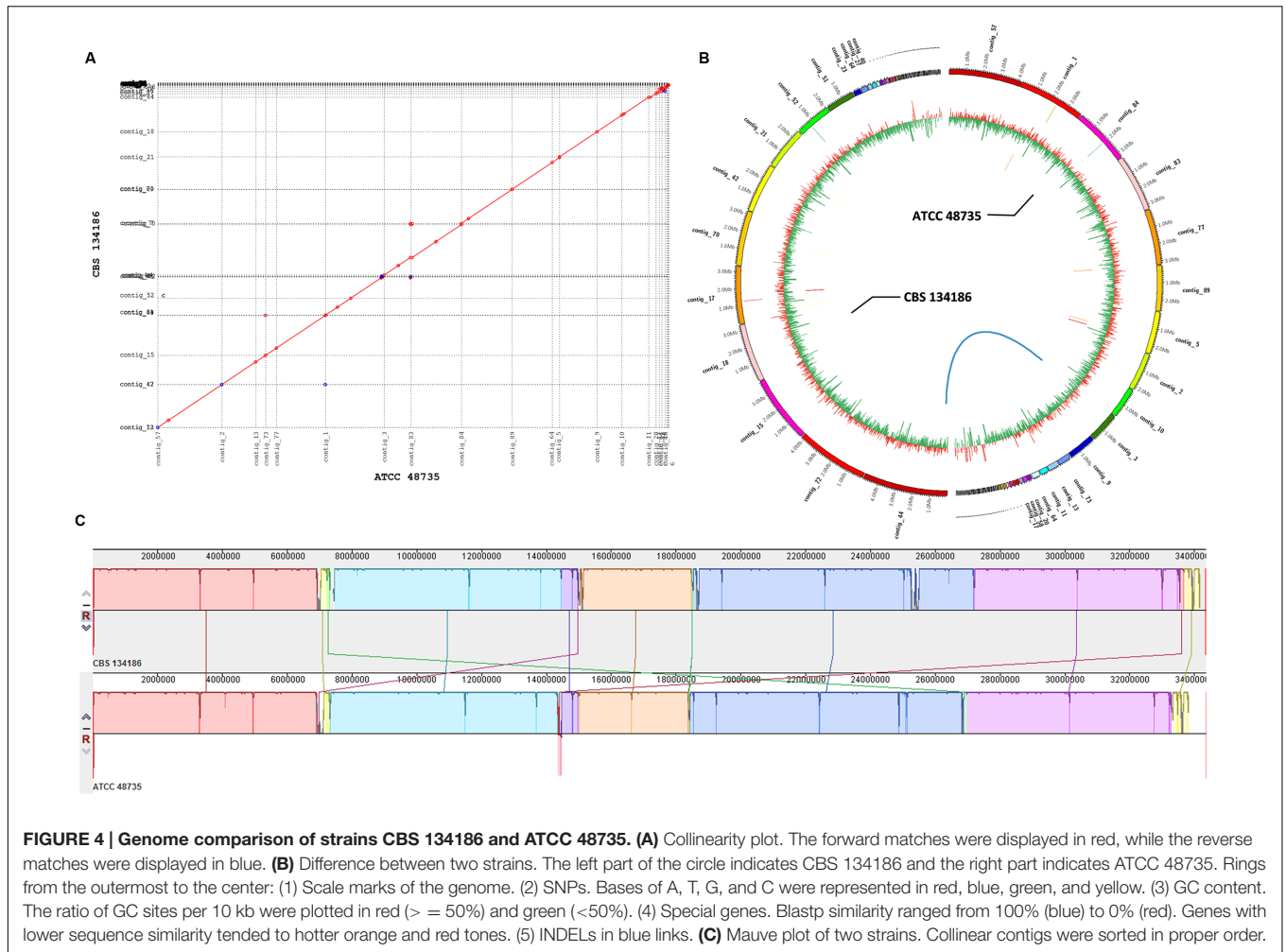
FIGURE 4 | Genome comparison of strains CBS 134186 and ATCC 48735. **(A)** Collinearity plot. The forward matches were displayed in red, while the reverse matches were displayed in blue. **(B)** Difference between two strains. The left part of the circle indicates CBS 134186 and the right part indicates ATCC 48735. Rings from the outermost to the center: (1) Scale marks of the genome. (2) SNPs. Bases of A, T, G, and C were represented in red, blue, green, and yellow. (3) GC content. The ratio of GC sites per 10 kb were plotted in red (> = 50%) and green (<50%). (4) Special genes. Blastp similarity ranged from 100% (blue) to 0% (red). Genes with lower sequence similarity tended to hotter orange and red tones. (5) INDELs in blue links. **(C)** Mauve plot of two strains. Collinear contigs were sorted in proper order.

TABLE 2 | The unique characteristics with pathogenic potential in strain CBS 134186 genome by compared with the environmental strains.

Gene ID	Mutation location	Function	Exocrine protein gene	PHI database	Type	Secondary metabolite	Cytochrome P450s
g97	contig_17:376853-381151	Vegetative incompatibility protein HET-E-1	No	PHI: 211	gene	No	No
g10464	contig_70:1865570-1869484	N amino acid transport system protein	No	PHI: 511	gene	No	No
g512	contig_17:1693336-1696043	1,3-beta-glucanosyltransferase gel3	Yes	PHI: 522	gene	No	No
g399	contig_17:1308904	Circumsporozoite protein	No	—	SNP	No	No
g1265	contig_52:720079	Zn(2)-C6 fungal-type DNA-binding domain	No	PHI: 1163	SNP	No	No
g1265	contig_52:720280	Zn(2)-C6 fungal-type DNA-binding domain	No	PHI: 1163	SNP	No	No
g8270	contig_44:191141	F-actin-capping protein subunit beta	No	PHI: 2568	INDEL	No	No

involved in cell wall composition and maintenance, and the high expression of these genes may increase the pathogenicity of the clinical strain CBS 134186.
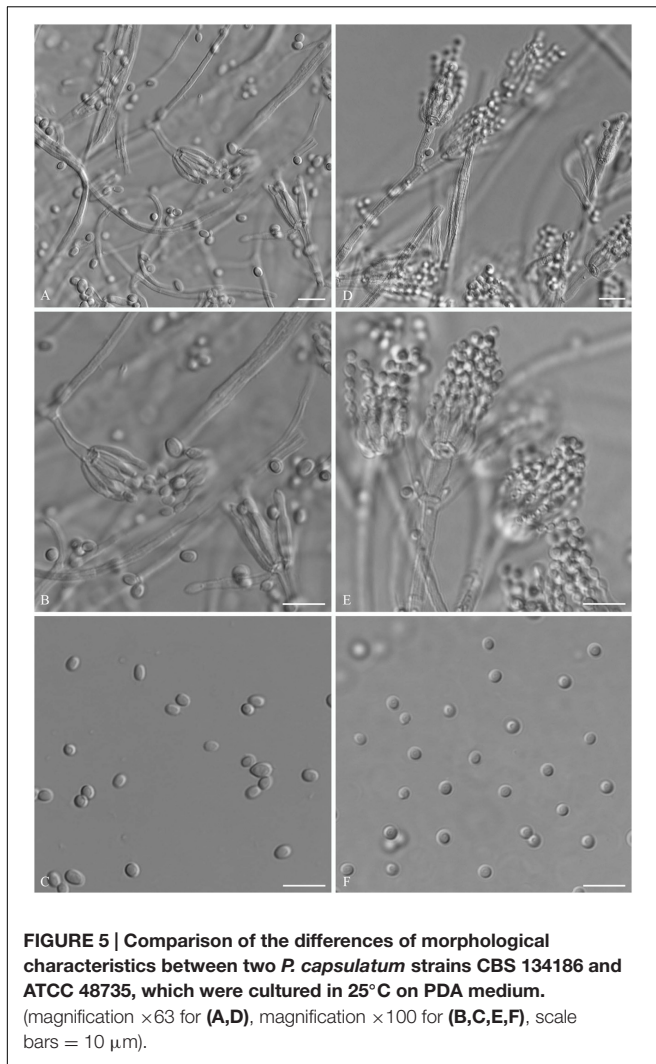
Morphological Characteristics of Strains CBS 134186 and ATCC 48735

We comparatively analyzed the morphological characteristics between strain CBS 134186 and strain ATCC 48735. Both *P. capsulatum* strains had smooth walled conidia that were ellipsoid to slightly cylindrical in shape (**Figure 5**). There were

no significantly different morphological characteristics between the two strains, which was in accordance with their highly similar genomes.

DISCUSSION

In the present study, the genomes of two *P. capsulatum* strains were successively sequenced using the Pacific Biosciences RS II and Illumina MiSeq platforms in combination. There are only 5 contigs length beyond N50 in CBS 134186 and ATCC 48735,

FIGURE 5 | Comparison of the differences of morphological characteristics between two *P. capsulatum* strains CBS 134186 and ATCC 48735, which were cultured in 25°C on PDA medium. (magnification ×63 for **(A,D)**, magnification ×100 for **(B,C,E,F)**, scale bars = 10 μm).

which demonstrated the high integrity of the genome assembly. Our assembly size of the genome is similar to reported *Penicillium* genome (van den Berg et al., 2008). Furthermore, the combined application of these second and third generation sequencing technologies shows good potential for the development of genome sequence assemblies without reference to genomic data, thus resulting in fewer unresolved gaps or ambiguities and a smaller number of contigs.

To explore the pathogenic potential of novel pathogens in *Eurotiales,* such as *P. capsulatum*, we referred to studies of representative fungal pathogens, such as *A. fumigatus, T. marneffei,* and others. As successful fungal pathogens, numerous genes from these species have been systematically researched and verified to be involved in virulence (Pitt, 1994; Abad et al., 2010; Chakrabarti and Slavin, 2011; Sheppard and Filler, 2015). A total of 106 genes were selected after we systematically reviewed the virulence-related genes of these important fungal pathogens (Supplementary Table S3). Then, we conducted a further analysis based on the homologous virulence-related genes between *P. capsulatum* and other related *Eurotiales* species. These

genes can be grouped into seven main functional aspects: thermo-tolerance, cell wall composition and maintenance, resistance to the immune response, toxins, nutrient uptake during invasive growth, signaling, metabolic regulation and response to stress conditions, and allergens. Moreover, these virulence-related genes are considered to result from a combination of biological characteristics of the fungus and the immune status of the patient. Interestingly, a considerable number of virulence-related genes are shared between novel fungal pathogens, such as *P. capsulatum*, and non-pathogenic fungi, such as *P. paxilli*. This suggests the pathogenic potential of these *Penicillium* species, even if some of these species have not yet been reported as human pathogens. Certainly, the bioinformatics analysis has its inherent limitation to predict fungal pathogenicity, further verification and analysis works are needed. Here, we just make a discussion on the pathogenicity evaluation based on the whole genome sequencing data.

A total of 101 possible virulence-related genes were found in the *P. capsulatum* genome (Supplementary Table S10). Of these, we selected some important probable virulence-related genes for further analysis with regard to seven aspects. First, thermo-tolerance is considered to be an essential ability for a fungal pathogen to successfully infect humans. A total of five genes have been found to be essential for thermo-tolerance, including the *thtA*, *afpmt1*, *kre2/afpmt1*, *cgrA*, and *hsp1/asp f12* genes. All of these genes were identified in both strains CBS 134186 and ATCC 48735. Among these genes, *kre2/afpmt1* and *cgrA* genes are required for growth at 48C and 37°C, respectively; notably, the protein encoded by *kre2/afpmt1* has also been considered as a new antifungal target (Bhabhra et al., 2004; Wagener et al., 2008). Notably, contig_42g3968 and contig_21g7395 in CBS 134186 were regarded as the *kre2/afpmt1* and *cgrA* genes by functional annotation and verification using the pathogen-host interactions database. Similar results were obtained in the genome of ATCC 48735, it showed potential thermo-tolerance, which may be crucial for a novel fungal pathogen. A previous study by our group also showed that CBS 134186 grows well in MEA, CYA, and YES cultures at 37°C, and grows faster at 30 and 37°C than at 25°C (Chen et al., 2013). Second, due to its structural integrity and physical protection, the fungal cell-wall is the mainline of defense against the human inner environment and is frequently attacked as a target of the immune system during infection. A total of 20 genes, including *fks1*, *gel1*, and *ecm33*, have been confirmed as necessary for the composition and maintenance of the cell wall (Mouyna et al., 2000a,b, 2004; Chabane et al., 2006; Maubon et al., 2006; Romano et al., 2006; Wagener et al., 2008; Gastebois et al., 2009). These genes were also found in the studied *Penicillium* species, including *P. capsulatum*, suggesting that the composition and maintenance of the cell wall play important roles during infection. Third, a successful fungal pathogen can evade or resist the immune response and can also weaken the host immune response. A total of 25 genes, including *pksP/alb1* and *cat2/katG*, have been found to be associated with resistance to the human immune response in previous studies (Calera et al., 1997; Langfelder et al., 1998; Brakhage and Liebmann, 2005; Tsitsigiannis et al., 2005; Reeves et al., 2006; Abad et al.,

2010; Al-Bader et al., 2010). Interestingly, all 25 of these genes have been found to be similar to genes in the *P. capsulatum*. For example, the *cat2/katG* gene in the *A. fumigates* genome encodes the hyphal catalases, and the deletion of this gene resulted in increased susceptibility of conidia to H_2O_2 in *vitro* (Abad et al., 2010). Our results implied that the profile of resistance to the immune response in *P. capsulatum* may be similar to that of *A. fumigates* during infection, which has a combination of characteristics that helps the fungus to evade or resist to immune response (Abad et al., 2010). Fourth, toxins produced by fungi not only protect the fungi from predators and competitors in their ecological niche (Abad et al., 2010; Steinbach et al., 2006) but may also contribute to the pathogenesis of fungal pathogens, such as *A. fumigates*, because they can directly attack the host. No fewer than nine genes in *A. fumigates* have been shown to attack humans, and most of these genes are associated with secondary metabolites (Abad et al., 2010). Six of these nine genes have homologues in *P. capsulatum* that are also related to secondary metabolites. Hence, *P. capsulatum* may produce similar toxins that can attack humans. Fifth, the uptake of nutrients in the human internal environment is an essential for the success of a fungal pathogen (Abad et al., 2010). Currently, no fewer than 37 genes have been associated with the uptake of nutrients in the human internal environment, including through normal nutrient uptake systems and other activated systems (D'Enfert et al., 1996; Brown et al., 2000; Hissen et al., 2005; Moreno et al., 2007; Schrettl et al., 2007; Al-Bader et al., 2010; Amich et al., 2010; Li et al., 2016). Thirty-six of these 37 homologous genes have been found in *P. capsulatum*. Our results strongly suggested that *P. capsulatum* may have a considerable capacity for nutrient uptake in the human internal environment. Sixth, medically important fungal pathogens can regulate their cellular physiology to adapt to the various, dynamic, and changing conditions relative to the normal environmental niche during infection. These conditions include increased osmolarity, heatshock and nutrient limitation, and adaptation can occur through several defined signaling regulatory mechanisms, such as the mitogen-activated protein kinase (MAPK) pathways, Rasproteins, calcium signaling, and others. To date, no fewer than eight genes have been confirmed to be related to the regulation of signaling during adaptation to the environment in association with infection in the previous reports (May et al., 2005; Abad et al., 2010). Interestingly, all of these eight genes have been found to be similar to genes present in *P. capsulatum* strains. Among these genes, *tcsB*, *Sln1*, *calA*, and *cnaA* have been shown to be related to stress responses through, for instance, the signaling regulation of adaptation by Ras-proteins or MAPK pathways; notably, the deletion of genes involved in this pathway in fungal pathogens results in hypo-virulence (Pott et al., 2000; Clemons et al., 2002; Steinbach et al., 2006, 2007). Seventh, most fungal pathogens can produce various allergenic molecules, but they can also increase the risk associated with aspergillosis because they can redirect the immune response against the fungus by activating *Th2* lymphocytes (Reichard et al., 1990; Kolattukudy et al., 1993; Tang et al., 1993; Jaton-Ogay et al., 1994). We searched the homologous genes of *P. capsulatum*, and not surprisingly, similar genes, such as *asp f* 13, were also

detected in *P. capsulatum* strains. Again, this result suggested the pathogenic potential of *P. capsulatum* as a novel fungal pathogen.

The above virulence-related genes were observed in equal numbers in the CBS 134186 and ATCC 48735 genomes, which not only suggested the pathogenic potential of *P. capsulatum* as a novel fungal pathogen but also implied that the difference in the pathogenic potential of novel fungal pathogens between the clinical and environmental isolates may be very small compared with our expectations. We suppose the absence of virulence-related genes in *P. capsulatum* may play a key role in the divergence between novel fungal pathogens and medically important fungal pathogens.

To better study the differences between the clinical and environmental isolates of *P. capsulatum*, we performed a comparative genomic analysis of strains CBS 134186 and ATCC 48735. The results showed that the pathogenic potential of the clinical strain CBS 134186 and environmental strain ATCC 48735 may be extremely similar and that those specific genes, SNPs or INDELs present in the CBS 134186 genome may play a key role in causing invasive infection. Among these probable mutations, a total of 7 unique genes, SNPs or INDELs in the CBS 134186 genome were as likely candidates for conferring the pathogenic status needed to make this strain a novel fungal pathogen. The unique genes coding for vegetative incompatibility protein and glycosyl hydrolases (g512 in contig27), the SNP in the gene coding for circumsporozoite protein (g399 in contig17), and the INDEL in the gene containing F-actin capping protein beta subunit (g8270 in contig44) in the CBS 134186 genome may be related to its adaptation to stress conditions, as observed in the interaction between *A. fumigates* and humans. Moreover, a transcriptomic analysis showed that the g512 and g9805 genes in CBS 134186 were expressed at significantly higher levels than their respective homologs in ATCC 48735 (g3527 and g10066). Both of these genes have been confirmed to be required for the composition and maintenance of the cell wall in *A. fumigates* (Mouyna et al., 2000a,b, 2004, 2005; Chabane et al., 2006; Maubon et al., 2006; Romano et al., 2006; Gastebois et al., 2009). We also speculate these two genes might help the clinical strain to adapt to the human internal environment, as observed in the interaction between *A. fumigates* and humans (Wagener et al., 2008; Abad et al., 2010). These results are likely applicable beyond the present study of *P. capsulatum* and should be considered for other novel fungal pathogens.

CONCLUSION

We first report the *P. capsulatum* genome by using a combined NGS strategy. A phylogenetic analysis based on whole genome data may offer us comprehensive information of each strain. The analysis of virulence-related genes in *P. capsulatum* and related species revealed that these strains have significant pathogenic potential. In addition, the comparative genomic analysis of the *P. capsulatum* strains showed high similarity between the clinical strain CBS 134186 and the environmental strain ATCC 48735, except for several unique genes, INDELs or SNPs in genes

such as that which codes for circumsporozoite protein; these differences may increase the infectious capacity of this clinical strain. Remarkably, a number of virulence-related genes (such as *gel1-3*, *mirB*, and *afpmt1* in *A. fumigatus*) in CBS 134186, which have also been identified in the *P. capsulatum* genome, may be closely related to the pathogenicity of *P. capsulatum* and the phylogenetically related *Penicillium* species. Moreover, considering the host of CBS 134186 is a type 2 diabetes patient, the inner-host environment, glucose metabolism and host immune status may be different, which could also increase the possibility of *P. capsulatum* infection. We believe that this study deepens our understanding of the genomic features of *P. capsulatum* and reveals a considerable level of pathogenic potential in *P. capsulatum* as an example of a novel but easily neglected *Penicillium* pathogen. Clinicians, mycologists and epidemiologists should be aware of novel fungal pathogens such as *P. capsulatum*.

AUTHOR CONTRIBUTIONS

WL, JW, HC, and SW conceived the project. YY, MC, and ZL prepared the strain samples and did the sequencing test. ZL, XB conducted the bioinformatics analysis. MC, YY, and ZL prepared the manuscript. AA, GdH, WP, QY, and YL participated in discussions and provided suggestions. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2016.01541

REFERENCES

Abad, A., Fernandez-Molina, J. V., Bikandi, J., Ramirez, A., Margareto, J., Sendino, J., et al. (2010). What makes *Aspergillus fumigatus* a successful pathogen? Genes and molecules involved in invasive aspergillosis. *Rev. Iberoam. Micol.* 27, 155–182. doi: 10.1016/j.riam.2010.10.003

Al-Bader, N., Vanier, G., Liu, H., Gravelat, F. N., Urb, M., Hoareau, C. M., et al. (2010). Role of trehalose biosynthesis in *Aspergillus fumigatus* development, stress response, and virulence. *Infect. Immun.* 78, 3007–3018. doi: 10.1128/IAI.00813-09

Amich, J., Vicentefranqueira, R., Leal, F., and Calera, J. A. (2010). *Aspergillus fumigatus* survival in alkaline and extreme zinc-limiting environments relies on the induction of a zinc homeostasis system encoded by the zrfC and aspf2 genes. *Eukaryot. Cell* 9, 424–403. doi: 10.1128/EC.00348-09

Bhabhra, R., Miley, M. D., Mylonakis, E., Boettner, D., Fortwendel, J., Panepinto, J. C., et al. (2004). Disruption of the *Aspergillus fumigatus* gene encoding nucleolar protein CgrA impairs thermotolerant growth and reduces virulence. *Infect. Immun.* 72, 4731–4740. doi: 10.1128/IAI.72.8.4731-4740.2004

Blin, K., Medema, M. H., Kazempour, D., Fischbach, M. A., Breitling, R., Takano, E., et al. (2013). antiSMASH 2.0-a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 41, W204–W212. doi: 10.1093/nar/gkt449

Brakhage, A. A., and Liebmann, B. (2005). *Aspergillus fumigatus* conidial pigment and cAMP signal transduction: significance for virulence. *Med. Mycol.* 43(Suppl. 1), S75–S82. doi: 10.1080/13693780400028967

Brown, J. S., Aufauvre-Brown, A., Brown, J., Jennings, J. M., Arst, H. Jr., and Holden, D. W. (2000). Signature-tagged and directed mutagenesis identify PABA synthetase as essential for *Aspergillus fumigatus* pathogenicity. *Mol. Microbiol.* 36, 1371–1380. doi: 10.1046/j.1365-2958.2000.01953.x

Calera, J. A., Paris, S., Monod, M., Hamilton, A. J., Debeaupuis, J. P., Diaquin, M., et al. (1997). Cloning and disruption of the antigenic catalase gene of *Aspergillus fumigatus*. *Infect. Immun.* 65, 4718–4724.

Chabane, S., Sarfati, J., Ibrahim-Granet, O., Du, C., Schmidt, C., Mouyna, I., et al. (2006). Glycosylphosphatidylinositol-anchored Ecm33p influences conidial cell wall biosynthesis in *Aspergillus fumigatus*. *Appl. Environ. Microbiol.* 72, 3259–3267. doi: 10.1128/AEM.72.5.3259-3267.2006

Chakrabarti, A., and Slavin, M. A. (2011). Endemic fungal infections in the Asia-Pacific region. *Med. Mycol.* 49, 337–344. doi: 10.3109/13693786.2010.551426

Chen, K. H., Miadlikowska, J., Molnar, K., Arnold, A. E., U'Ren, J. M., Gaya, E., et al. (2015). Phylogenetic analyses of eurotiomycetous endophytes reveal their close affinities to Chaetothyriales, Eurotiales, and a new order – Phaeomoniellales. *Mol. Phylogenet. Evol.* 85, 117–130. doi: 10.1016/j.ympev.2015.01.008

Chen, M., Houbraken, J., Pan, W., Zhang, C., Peng, H., Wu, L., et al. (2013). Pulmonary fungus ball caused by *Penicillium capsulatum* in a patient with type 2 diabetes: a case report. *BMC Infect. Dis.* 13:496. doi: 10.1186/1471-2334-13-496

Clemons, K. V., Miller, T. K., Selitrennikoff, C. P., and Stevens, D. A. (2002). fos-1, a putative histidine kinase as a virulence factor for systemic aspergillosis. *Med. Mycol.* 40, 259–262. doi: 10.1080/714031112

Darling, A. E., Mau, B., and Perna, N. T. (2010). progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147. doi: 10.1371/journal.pone.0011147

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* 27, 1164–1165. doi: 10.1093/bioinformatics/btr088

D'Enfert, C., Diaquin, M., Delit, A., Wuscher, N., Debeaupuis, J. P., Huerre, M., et al. (1996). Attenuated virulence of uridine-uracil auxotrophs of *Aspergillus fumigatus*. *Infect. Immun.* 64, 4401–4405.

Ferrarini, M., Moretto, M., Ward, J. A., Surbanovski, N., Stevanovic, V., Giongo, L., et al. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics* 14:670. doi: 10.1186/1471-2164-14-670

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484, 186–194. doi: 10.1038/nature10947

Garcia-Solache, M. A., and Casadevall, A. (2010). Global warming will bring new fungal diseases for mammals. *MBio* 1:e00061-10. doi: 10.1128/mBio.00061-10

Gastebois, A., Clavaud, C., Aimanianda, V., and Latge, J. P. (2009). *Aspergillus fumigatus*: cell wall polysaccharides, their biosynthesis and organization. *Future Microbiol.* 4, 583–595. doi: 10.2217/Fmb.09.29

Hissen, A. H., Wan, A. N., Warwas, M. L., Pinto, L. J., and Moore, M. M. (2005). The *Aspergillus fumigatus* siderophore biosynthetic gene sidA, encoding L-ornithine N5-oxygenase, is required for virulence. *Infect. Immun.* 73, 5493–5503. doi: 10.1128/iai.73.9.5493-5503.2005

Houbraken, J., and Samson, R. A. (2011). Phylogeny of *Penicillium* and the segregation of Trichocomaceae into three families. *Stud. Mycol.* 70, 1–51. doi: 10.3114/sim.2011.70.01

Jaton-Ogay, K., Paris, S., Huerre, M., Quadroni, M., Falchetto, R., Togni, G., et al. (1994). Cloning and disruption of the gene encoding an extracellular metalloprotease of *Aspergillus fumigatus*. *Mol. Microbiol.* 14, 917–928. doi: 10.1111/j.1365-2958.1994.tb01327.x

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., et al. (2008). Global trends in emerging infectious diseases. *Nature* 451, 990–993. doi: 10.1038/nature06536

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031

Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome. Res.* 110, 462–467. doi: 10.1159/000084979

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Keesing, F., Belden, L. K., Daszak, P., Dobson, A., Harvell, C. D., Holt, R. D., et al. (2010). Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468, 647–652. doi: 10.1038/nature09575

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14:R36. doi: 10.1186/gb-2013-14-4-r36

Kolattukudy, P. E., Lee, J. D., Rogers, L. M., Zimmerman, P., Ceselski, S., Fox, B., et al. (1993). Evidence for possible involvement of an elastolytic serine protease in aspergillosis. *Infect. Immun.* 61, 2357–2368.

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645. doi: 10.1101/gr.092759.109

Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., et al. (2004). Versatile and open software for comparing large genomes. *Genome Biol.* 5:R12. doi: 10.1186/gb-2004-5-2-r12

Langfelder, K., Jahn, B., Gehringer, H., Schmidt, A., Wanner, G., and Brakhage, A. A. (1998). Identification of a polyketide synthase gene (pksP) of *Aspergillus fumigatus* involved in conidial pigment biosynthesis and virulence. *Med. Microbiol. Immunol.* 187, 79–89. doi: 10.1007/s004300050077

Lavezzo, E., Toppo, S., Franchin, E., Di Camillo, B., Finotello, F., Falda, M., et al. (2013). Genomic comparative analysis and gene function prediction in infectious diseases: application to the investigation of a meningitis outbreak. *BMC Infect. Dis.* 13:554. doi: 10.1186/1471-2334-13-554

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. doi: 10.1093/bioinformatics/btp698

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, L., Stoeckert, C. J. Jr., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Li, Y., Wang, Z. K., Liu, X., Song, Z. Y., Li, R., Shao, C. W., et al. (2016). Siderophore biosynthesis but not reductive iron assimilation is essential for the dimorphic fungus *Nomuraea rileyi* conidiation, dimorphism transition, resistance to oxidative stress, pigmented microsclerotium formation, and virulence. *Front. Microbiol.* 7:931. doi: 10.3389/fmicb.2016.00931

Lowe, T. M., and Eddy, S. R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25, 955–964. doi: 10.1093/nar/25.5.0955

Lyratzopoulos, G., Ellis, M., Nerringer, R., and Denning, D. W. (2002). Invasive infection due to penicillium species other than *P. marneffei*. *J Infect.* 45, 184–195. doi: 10.1053/jinf.2002.1056

Maubon, D., Park, S., Tanguy, M., Huerre, M., Schmitt, C., Prevost, M. C., et al. (2006). AGS3, an alpha(1-3)glucan synthase gene family member of *Aspergillus fumigatus*, modulates mycelium growth in the lung of experimentally infected mice. *Fungal Genet. Biol.* 43, 366–375. doi: 10.1016/j.fgb.2006.01.006

May, G. S., Xue, T., Kontoyiannis, D. P., and Gustin, M. C. (2005). Mitogen activated protein kinases of *Aspergillus fumigatus*. *Med. Mycol.* 43(Suppl. 1), S83–S86. doi: 10.1080/13693780400024784

Moreno, M. A., Ibrahim-Granet, O., Vicentefranqueira, R., Amich, J., Ave, P., Leal, F., et al. (2007). The regulation of zinc homeostasis by the ZafA transcriptional activator is essential for *Aspergillus fumigatus* virulence. *Mol. Microbiol.* 64, 1182–1197. doi: 10.1111/j.1365-2958.2007.05726.x

Morens, D. M., Folkers, G. K., and Fauci, A. S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature* 430, 242–249. doi: 10.1038/nature02759

Mouyna, I., Fontaine, T., Vai, M., Monod, M., Fonzi, W. A., Diaquin, M., et al. (2000a). Glycosylphosphatidylinositol-anchored glucanosyltransferases play an active role in the biosynthesis of the fungal cell wall. *J. Biol. Chem.* 275, 14882–14889. doi: 10.1074/jbc.275.20.14882

Mouyna, I., Henry, C., Doering, T. L., and Latge, J. P. (2004). Gene silencing with RNA interference in the human pathogenic fungus *Aspergillus fumigatus*. *FEMS Microbiol. Lett.* 237, 317–324. doi: 10.1016/j.femsle.2004.06.048

Mouyna, I., Monod, M., Fontaine, T., Henrissat, B., Lechenne, B., and Latge, J. P. (2000b). Identification of the catalytic residues of the first family of beta(1–3)glucanosyltransferases identified in fungi. *Biochem. J.* 347(Pt 3), 741–747. doi: 10.1042/bj3470741

Mouyna, I., Morelle, W., Vai, M., Monod, M., Lechenne, B., Fontaine, T., et al. (2005). Deletion of GEL2 encoding for a beta(1-3)glucanosyltransferase affects morphogenesis and virulence in *Aspergillus fumigatus*. *Mol. Microbiol.* 56, 1675–1688. doi: 10.1111/j.1365-2958.2005.04654.x

Pareek, C. S., Smoczynski, R., and Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J. Appl. Genet.* 52, 413–435. doi: 10.1007/s13353-011-0057-x

Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods* 8, 785–786. doi: 10.1038/nmeth.1701

Pitt, J. I. (1994). The current role of *Aspergillus* and *Penicillium* in human and animal health. *J. Med. Vet. Mycol.* 32(Suppl. 1), 17–32. doi: 10.1080/02681219480000701

Pott, G. B., Miller, T. K., Bartlett, J. A., Palas, J. S., and Selitrennikoff, C. P. (2000). The isolation of FOS-1, a gene encoding a putative two-component histidine kinase from *Aspergillus fumigatus*. *Fungal Genet. Biol.* 31, 55–67. doi: 10.1006/fgbi.2000.1225

Reeves, E. P., Reiber, K., Neville, C., Scheibner, O., Kavanagh, K., and Doyle, S. (2006). A nonribosomal peptide synthetase (Pes1) confers protection against oxidative stress in *Aspergillus fumigatus*. *FEBS J.* 273, 3038–3053. doi: 10.1111/j.1742-4658.2006.05315.x

Reichard, U., Buttner, S., Eiffert, H., Staib, F., and Ruchel, R. (1990). Purification and characterisation of an extracellular serine proteinase from *Aspergillus fumigatus* and its detection in tissue. *J. Med. Microbiol.* 33, 243–251. doi: 10.1099/00222615-33-4-243

Romano, J., Nimrod, G., Ben-Tal, N., Shadkchan, Y., Baruch, K., Sharon, H., et al. (2006). Disruption of the *Aspergillus fumigatus* ECM33 homologue results in rapid conidial germination, antifungal resistance and hypervirulence. *Microbiology* 152(Pt 7), 1919–1928. doi: 10.1099/mic.0.28936-0

Samson, R. A., Visagie, C. M., Houbraken, J., Hong, S. B., Hubka, V., Klaassen, C. H., et al. (2014). Phylogeny, identification and nomenclature of the genus *Aspergillus*. *Stud. Mycol.* 78, 141–173. doi: 10.1016/j.simyco.2014.07.004

Samson, R. A., Yilmaz, N., Houbraken, J., Spierenburg, H., Seifert, K. A., Peterson, S. W., et al. (2011). Phylogeny and nomenclature of the genus *Talaromyces* and taxa accommodated in *Penicillium* subgenus *Biverticillium*. *Stud. Mycol.* 70, 159–183. doi: 10.3114/sim.2011.70.04

Schattner, P., Brooks, A. N., and Lowe, T. M. (2005). The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33, W686–W689. doi: 10.1093/nar/gki366

Schrettl, M., Bignell, E., Kragl, C., Sabiha, Y., Loss, O., Eisendle, M., et al. (2007). Distinct roles for intra- and extracellular siderophores during *Aspergillus fumigatus* infection. *PLoS Pathog.* 3:e128. doi: 10.1371/journal.ppat.0030128

Sheppard, D. C., and Filler, S. G. (2015). Host cell invasion by medically important fungi. *Cold Spring Harb. Perspect. Med.* 5:a019687. doi: 10.1101/cshperspect.a019687

Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6, 175–182.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033

Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24, 637–644. doi: 10.1093/bioinformatics/btn013

Steinbach, W. J., Cramer, R. A. Jr., Perfect, B. Z., Asfaw, Y. G., Sauer, T. C., Najvar, L. K., et al. (2006). Calcineurin controls growth, morphology, and pathogenicity in *Aspergillus fumigatus*. *Eukaryot. Cell* 5, 1091–1103. doi: 10.1128/EC.00139-06

Steinbach, W. J., Cramer, R. A. Jr., Perfect, B. Z., Henn, C., Nielsen, K., Heitman, J., et al. (2007). Calcineurin inhibition or mutation enhances cell wall inhibitors against *Aspergillus fumigatus*. *Antimicrob. Agents Chemother.* 51, 2979–2981. doi: 10.1128/AAC.01394-06

Struelens, M. J., and Brisse, S. (2013). From molecular to genomic epidemiology: transforming surveillance and control of infectious diseases. *Euro. Surveill.* 18:20386.

Tang, C. M., Cohen, J., Krausz, T., Van Noorden, S., and Holden, D. W. (1993). The alkaline protease of *Aspergillus fumigatus* is not a virulence determinant in two murine models of invasive pulmonary aspergillosis. *Infect. Immun.* 61, 1650–1656.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28, 511–515. doi: 10.1038/nbt.1621

Tsitsigiannis, D. I., Bok, J. W., Andes, D., Nielsen, K. F., Frisvad, J. C., and Keller, N. P. (2005). *Aspergillus* cyclooxygenase-like enzymes are associated with prostaglandin production and virulence. *Infect. Immun.* 73, 4548–4559. doi: 10.1128/IAI.73.8.4548-4559.2005

Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H., and Hammond-Kosack, K. E. (2015). The pathogen-host interactions database (PHI-base): additions and future developments. *Nucleic Acids Res.* 43, D645–D655. doi: 10.1093/nar/gku1165

van den Berg, M. A., Albang, R., Albermann, K., Badger, J. H., Daran, J. M., Driessen, A. J., et al. (2008). Genome sequencing and analysis of the filamentous fungus *Penicillium chrysogenum*. *Nat. Biotechnol.* 26, 1161–1168. doi: 10.1038/nbt.1498

Wagener, J., Echtenacher, B., Rohde, M., Kotz, A., Krappmann, S., Heesemann, J., et al. (2008). The putative alpha-1,2-mannosyltransferase AfMnt1 of the opportunistic fungal pathogen *Aspergillus fumigatus* is required for cell wall stability and full virulence. *Eukaryot. Cell* 7, 1661–1673. doi: 10.1128/Ec.00221-08