REPORT

www.the-innovation.org

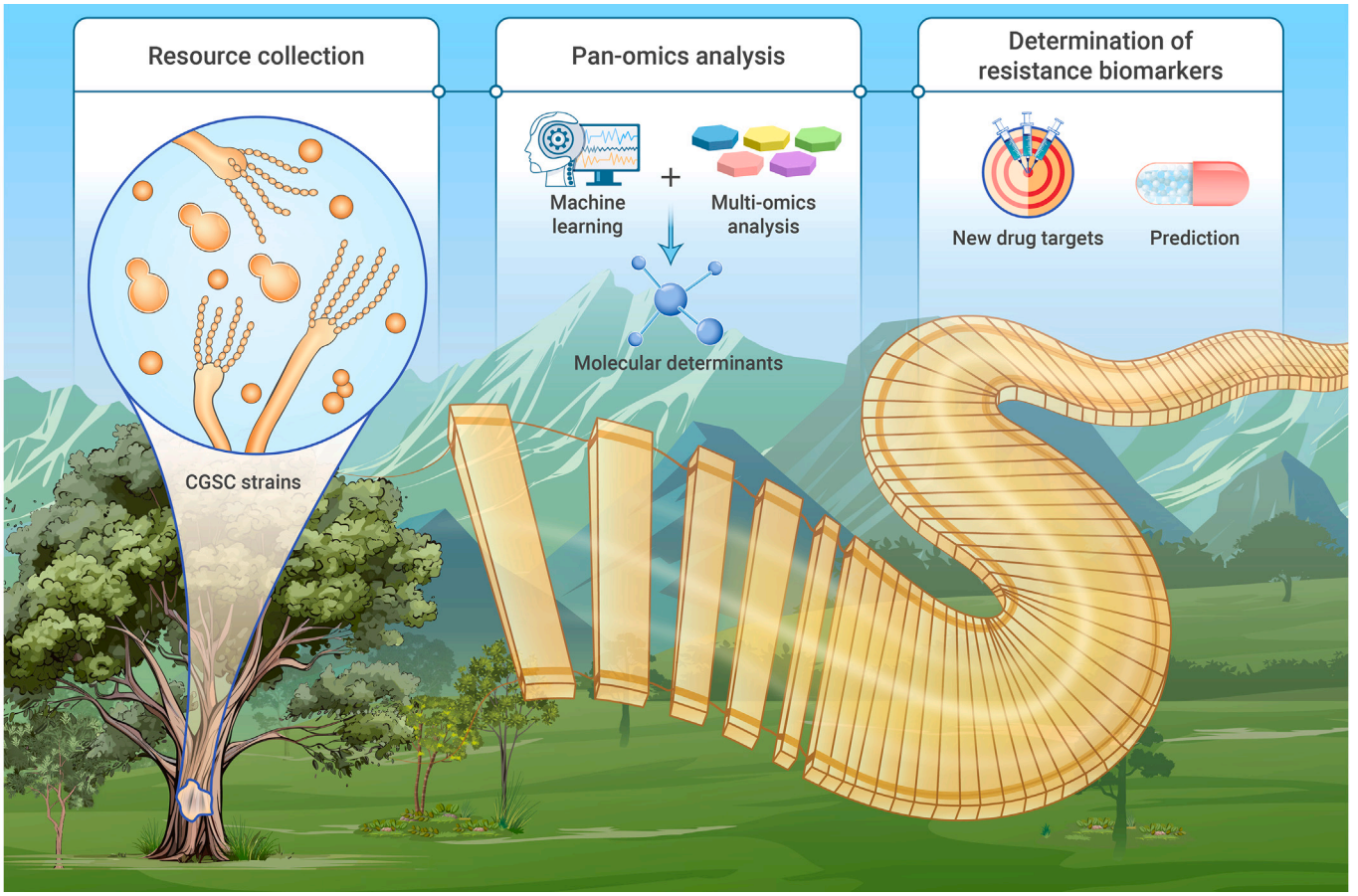
Pan-omics-based characterization and prediction of highly multidrug-adapted strains from an outbreak fungal species complex

Xin Fan,[1,2,3,18] Lei Chen,[3,18] Min Chen,[4,18] Na Zhang,[3,5] Hong Chang,[5,6] Mingjie He,[5,6] Zhenghao Shen,[3,5] Lanyue Zhang,[3,5] Hao Ding,[3,5] Yuyan Xie,[3,5] Yemei Huang,[7] Weixin Ke,[3] Meng Xiao,[8,9] Xuelei Zang,[7] Heping Xu,[10] Wenxia Fang,[11] Shaojie Li,[3,5] Cunwei Cao,[12,13] Yingchun Xu,[8,9] Shiguang Shan,[5,6] Wenjuan Wu,[14] Changbin Chen,[15,16,*] Xinying Xue,[7,17,*] and Linqi Wang[3,5,*]

*Correspondence: cbchen@ips.ac.cn (C.C.); xuexinying2988@bjsjth.cn (X.X.); wanglq@im.ac.cn (L.W.)

Received: January 5, 2024; Accepted: July 28, 2024; Published Online: July 31, 2024; https://doi.org/10.1016/j.xinn.2024.100681

GRAPHICAL ABSTRACT



PUBLIC SUMMARY

- Pan-phenomic assessment revealed 2,821 phenotype-strain associations in the *Cryptococcus gattii* species complex (CGSC).
- Phenotypic analysis revealed a specific set of CGSC strains with high adaptations to three first-line antifungals.
- Integrated pan-omic approaches identified novel multidrug resistance determinants in CGSC strains.
- We identified biomarkers that allow the prediction of highly multidrug-adapted CGSC strains.

REPORT

Pan-omics-based characterization and prediction of highly multidrug-adapted strains from an outbreak fungal species complex

Xin Fan,[1,2,3,18] Lei Chen,[3,18] Min Chen,[4,18] Na Zhang,[3,5] Hong Chang,[5,6] Mingjie He,[5,6] Zhenghao Shen,[3,5] Lanyue Zhang,[3,5] Hao Ding,[3,5] Yuyan Xie,[3,5] Yemei Huang,[7] Weixin Ke,[3] Meng Xiao,[8,9] Xuelei Zang,[7] Heping Xu,[10] Wenxia Fang,[11] Shaojie Li,[3,5] Cunwei Cao,[12,13] Yingchun Xu,[8,9] Shiguang Shan,[5,6] Wenjuan Wu,[14] Changbin Chen,[15,16,*] Xinying Xue,[7,17,*] and Linqi Wang[3,5,*]

[1]Department of Infectious Diseases and Clinical Microbiology, Beijing Institute of Respiratory Medicine and Beijing Chao-Yang Hospital, Capital Medical University, Beijing 100020, China

[2]Beijing Research Center for Respiratory Infectious Diseases, Beijing 100020, China

[3]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[4]Department of Dermatology, Shanghai Key Laboratory of Molecular Medical Mycology, Changzheng Hospital, Shanghai 200003, China

[5]University of Chinese Academy of Sciences, Beijing 100049, China

[6]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

[7]Department of Respiratory and Critical Care, Emergency and Critical Care Medical Center, Beijing Shijitan Hospital, Capital Medical University, Beijing 100038, China

[8]Department of Laboratory Medicine, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, Beijing 100730, China

[9]Beijing Key Laboratory for Mechanisms Research and Precision Diagnosis of Invasive Fungal Diseases, Beijing 100730, China

[10]Department of Clinical Laboratory, First Affiliated Hospital of Xiamen University, Xiamen 361003, China

[11]Institute of Biological Science and Technology, Guangxi Academy of Sciences, Nanning 530007, China

[12]Department of Dermatology, The First Affiliated Hospital of Guangxi Medical University, Nanning 530021, China

[13]Guangxi Key Laboratory of Mycosis Prevention and Treatment, Nanning 530021, China

[14]Department of Laboratory Medicine, Shanghai East Hospital, Tongji University School of Medicine, Shanghai 200120, China

[15]The Unit of Pathogenic Fungal Infection & Host Immunity, Shanghai Institute of Immunity and Infection, Chinese Academy of Sciences, Shanghai 200031, China

[16]Nanjing Advanced Academy of Life and Health, Nanjing 211135, China

[17]Department of Respiratory and Critical Care, Shandong Second Medical University, Weifang 261035, China

[18]These authors contributed equally

*Correspondence: cbchen@ips.ac.cn (C.C.); xuexinying2988@bjsjth.cn (X.X.); wanglq@im.ac.cn (L.W.)

Received: January 5, 2024; Accepted: July 28, 2024; Published Online: July 31, 2024; https://doi.org/10.1016/j.xinn.2024.100681

Citation: Fan X., Chen L., Chen M., et al., (2024). Pan-omics-based characterization and prediction of highly multidrug-adapted strains from an outbreak fungal species complex. The Innovation **5(5)**, 100681.

Strains from the *Cryptococcus gattii* species complex (CGSC) have caused the Pacific Northwest cryptococcosis outbreak, the largest cluster of life-threatening fungal infections in otherwise healthy human hosts known to date. In this study, we utilized a pan-phenome-based method to assess the fitness outcomes of CGSC strains under 31 stress conditions, providing a comprehensive overview of 2,821 phenotype-strain associations within this pathogenic clade. Phenotypic clustering analysis revealed a strong correlation between distinct types of stress phenotypes in a subset of CGSC strains, suggesting that shared determinants coordinate their adaptations to various stresses. Notably, a specific group of strains, including the outbreak isolates, exhibited a remarkable ability to adapt to all three of the most commonly used antifungal drugs for treating cryptococcosis (amphotericin B, 5-fluorocytosine, and fluconazole). By integrating pan-genomic and pan-transcriptomic analyses, we identified previously unrecognized genes that play crucial roles in conferring multidrug resistance in an outbreak strain with high multidrug adaptation. From these genes, we identified biomarkers that enable the accurate prediction of highly multidrug-adapted CGSC strains, achieving maximum accuracy and area under the curve (AUC) of 0.79 and 0.86, respectively, using machine learning algorithms. Overall, we developed a pan-omic approach to identify cryptococcal multidrug resistance determinants and predict highly multidrug-adapted CGSC strains that may pose significant clinical concern.

INTRODUCTION

Cryptococcosis is a severe fungal disease that affects people worldwide. This life-threatening illness is caused by pathogenic yeasts from the *Cryptococcus neoformans* species complex (CNSC) and the *Cryptococcus gattii* species complex (CGSC), which are among the most significant human fungal pathogens in the Basidiomycota phylum—a major fungal group with over 40,000 described species.[1–3] Currently, there are only three classes of drugs available to treat CNSC and CGSC infections: polyenes (mainly amphotericin B [AmB] and its lipid formulations), azoles (primarily fluconazole), and the pyrimidine analog 5-fluorocytosine (5-FC).[4–8] The limited number of antifungal options is further complicated by the ongoing emergence of drug-resistant variants, heightening the threat posed by CNSC and CGSC infections.[9–12]

Previous studies have identified several virulence traits common to both CNSC and CGSC pathogens.[5,13] For instance, both can thrive at typical mammalian body temperatures, a characteristic that sets them apart from non-pathogenic *Cryptococcus* species. Despite these similarities, the CNSC and CGSC exhibit different clinical features.[14–16] CNSC pathogens (*C. neoformans* and *C. deneoformans*) primarily infect immunocompromised individuals, such as patients with AIDS, often leading to central nervous system infections. In contrast, CGSC strains are more commonly associated with infections in immunocompetent individuals and can cause severe lung diseases without spreading to other organs.[13,17]

CGSC pathogens are classified into five recognized genotypes: VGI, VGII, VGIII, VGIV, and VGV. These genotypes are currently proposed as five distinct species: *C. gattii sensu stricto*, *C. deuterogattii*, *C. bacillisporus*, *C. tetragattii*, and *C. decagattii*, respectively.[18] To date, *C. gattii sensu stricto* and *C. deuterogattii* account for the majority of CGSC infections worldwide.[19–22] Infections caused by the other three species are relatively rare and mainly reported in patients from the Americas, Sub-Saharan Africa, and the Indian subcontinent.[20–22] Among the CGSC pathogens, *C. deuterogattii* (VGII) has garnered significant attention because it was responsible for a major cryptococcosis outbreak in the Pacific Northwest of North America—the largest recorded outbreak of invasive fungal disease in otherwise healthy hosts.[23,24] Over the past two decades, this pathogen, along with *C. gattii* (VGI), has been detected in various regions across temperate, subtropical, and tropical zones.[19–22] This suggests that they can adapt to diverse environmental niches with varying stressors, which may have contributed to their emergence as global pathogens capable of infecting healthy individuals.

Several studies have examined the differences between CGSC strains in their adaptation to specific stressors, such as antifungal drugs, which can lead to varying clinical treatment outcomes among strains.[25–29] Despite these valuable studies, the absence of a systematic pan-phenomic evaluation has limited our understanding of the survival advantages between and within CGSC species in response to antifungal agents or host-related and environmental stresses. These factors are strongly associated with the ability of CGSC strains to tolerate treatment, cause disease, and spread.

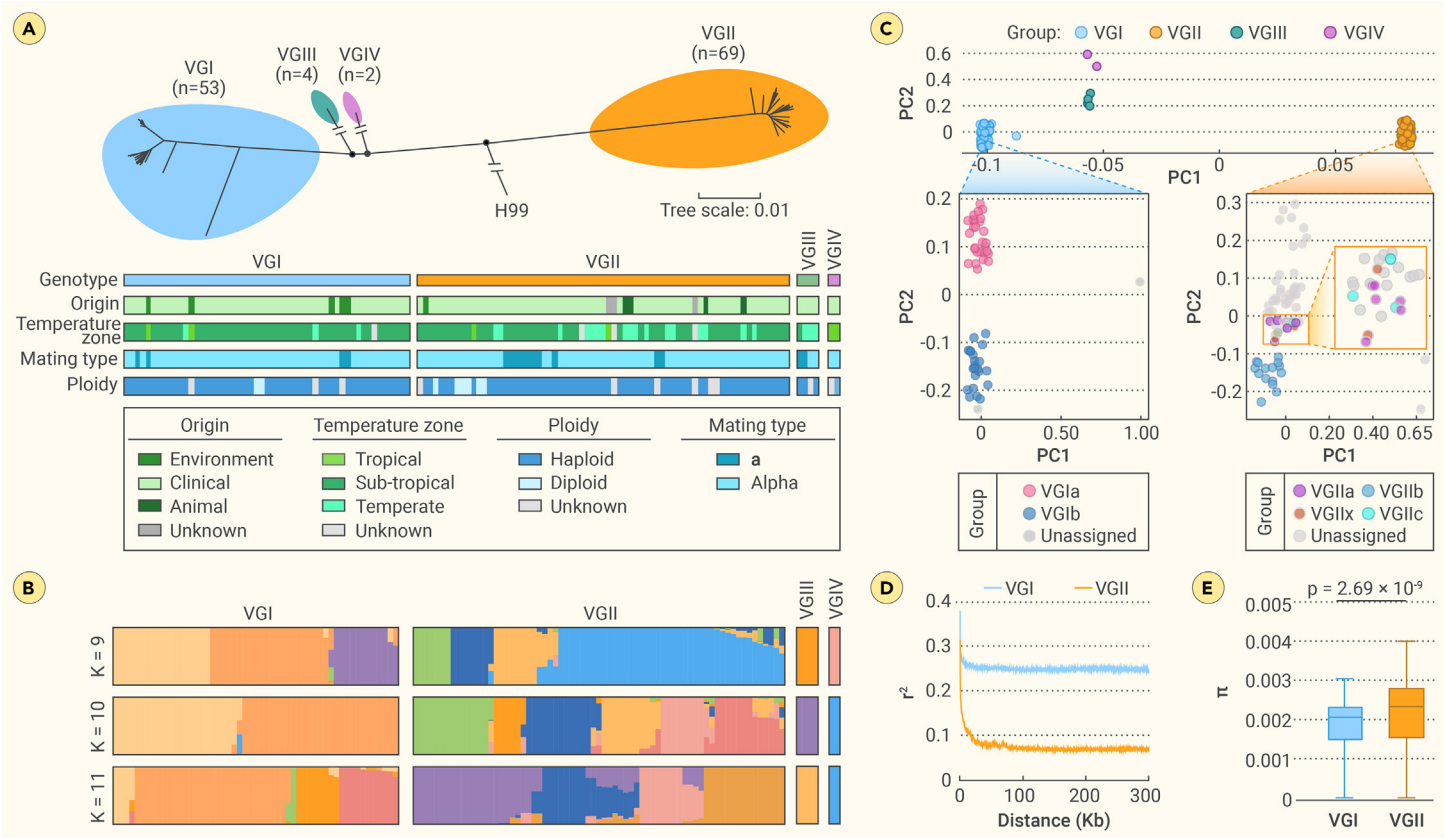


Figure 1. Phylogenetic diversity and population structure of CGSC strains (A) Maximum likelihood trees of 128 CGSC isolates based on SNPs from whole-genome sequences (top). The origins, temperature zones, ploidy levels, and mating types of the isolates are indicated by different bars (bottom). (B) Admixture population structure analysis of CGSC clades at K = 9, 10, and 11 using ADMIXTURE. (C) Principal-component analysis (PCA) showing sequence similarity of CGSC strains. The positions of the clades and subclades of CGSC strains are indicated, each distinguished by a unique color. (D) Decay of linkage disequilibrium (expressed in terms of correlation coefficient, r^2) as a function of distance averaged over 250 kb for VGI and VGII isolates. (E) Diversity metric (π) for VGI and VGII isolates. Boxplots display the 25th, 50th (median), and 75th quantiles as well as the minimum and maximum values. Statistical analyses were performed using the Mann-Whitney U test.

In this study, we conducted a comprehensive phenotypic evaluation of 91 strains of *C. deuterogattii* (VGII, 48 strains) and the closely related species *C. gattii* (VGI, 43 strains), the most common and widespread CGSC pathogens. Phenotypic assessments of fungal growth fitness under a wide range of stress conditions generated 2,821 phenotype-strain associations, providing unprecedented insights into inter- and intraspecific differences in stress adaptation. This phenome-based assessment allowed us to identify a specific group of CGSC strains with high adaptation to all three commonly used first-line antifungal drugs. By integrating pan-genomic and pan-transcriptomic approaches, we identified a set of genes involved in multidrug resistance in an outbreak CGSC strain. Furthermore, we discovered biomarkers that enable the accurate prediction of CGSC strains with high multidrug adaptation, which may be of potential clinical concern.

RESULTS

Population genomic analysis of 115 CGSC strains

We collected 115 CGSC strains from ten countries across Asia, Africa, Europe, North America, South America, and Oceania. These included 107 clinical isolates, 3 environmental isolates, 4 animal isolates, and 1 strain of unknown origin (Figure 1A; Table S1). The hospitals from which the clinical strains were sourced were located in temperate, subtropical, and tropical zones (Figure 1A). Consistent with most *Cryptococcus* species, the majority (87.8%, 101/115) of the CGSC strains were of the α mating type, while only 12.2% (14/115) were of the **a** mating type (Figure 1A; Table S1). Ploidy assessment using flow cytometry revealed that 93.0% (107/115) of the strains were haploid, with the remainder being diploid (Figure 1A; Table S1).

We isolated single colonies of all 115 CGSC strains and sequenced their genomes. To facilitate an unbiased analysis of the genomic diversity of the CGSC, we generated *de novo* genome assemblies of these isolates. The mean number of contigs in our assemblies was 187, the mean N50 was 283,565 bp, and the mean genome size was 17.58 Mb (Figure S1; Table S1).

To provide context for our isolates, we included the publicly available genome sequences of 11 CGSC strains reported by Farrer et al.[30] These sequences were of high quality and included the reference strains for VGI (WM276) and VGII (R265), bringing the total number of genomes for subsequent phylogenetic analyses to 128 isolates (Figure S1).

Phylogenetic analysis of the CGSC genomes revealed a clear separation into four distinct clades (Figures 1A and S2). This separation was further supported by population ancestry analyses using STRUCTURE (k = 10) and sequence similarity assessed via principal-component analysis (PCA) (Figures 1B and 1C). Among the 128 isolates, 53 were VGI (41.4%), 69 were VGII (53.9%), and 4 were VGIII. Only two isolates were VGIV, and no VGV isolates were included in this study (Figures 1A and S2). As VGIII, VGIV, and VGV are rare, particularly outside of the Indian subcontinent,[20–22] their low representation in our sample—comprising mainly East Asian strains (86/115, 74.8%)—was expected.

Our phylogenetic analysis, in line with previous studies, indicated that *C. deuterogattii*/VGII is the basal lineage of the CGSC, further subdivided into four subclades (Figures 1A, 1C, and S2). In contrast, VGI strains primarily clustered into two subclades (Figures 1C and S2). These results suggest greater genetic diversity within VGII strains compared to VGI strains.

Supporting this notion, the decay of linkage disequilibrium analysis for the CGSC strains showed that VGII exhibited evidence of <0.1 r^2 within a 250 kb region (Figure 1D). Conversely, approximately 0.26 r^2 was observed in the VGI population (Figure 1D). Additionally, genome-wide calculations of the diversity metric π revealed a statistically significant difference between the values for VGI and VGII strains, further corroborating the higher genetic diversity in VGII compared to VGI strains (Figure 1E).

Systematic phenotypic profiling of CGSC strains

The remarkable genetic diversity of CGSC strains revealed by population genomic analysis likely reflects variations in their ability to adapt to stress. To comprehensively assess the stress adaptation of CGSC strains, we used a

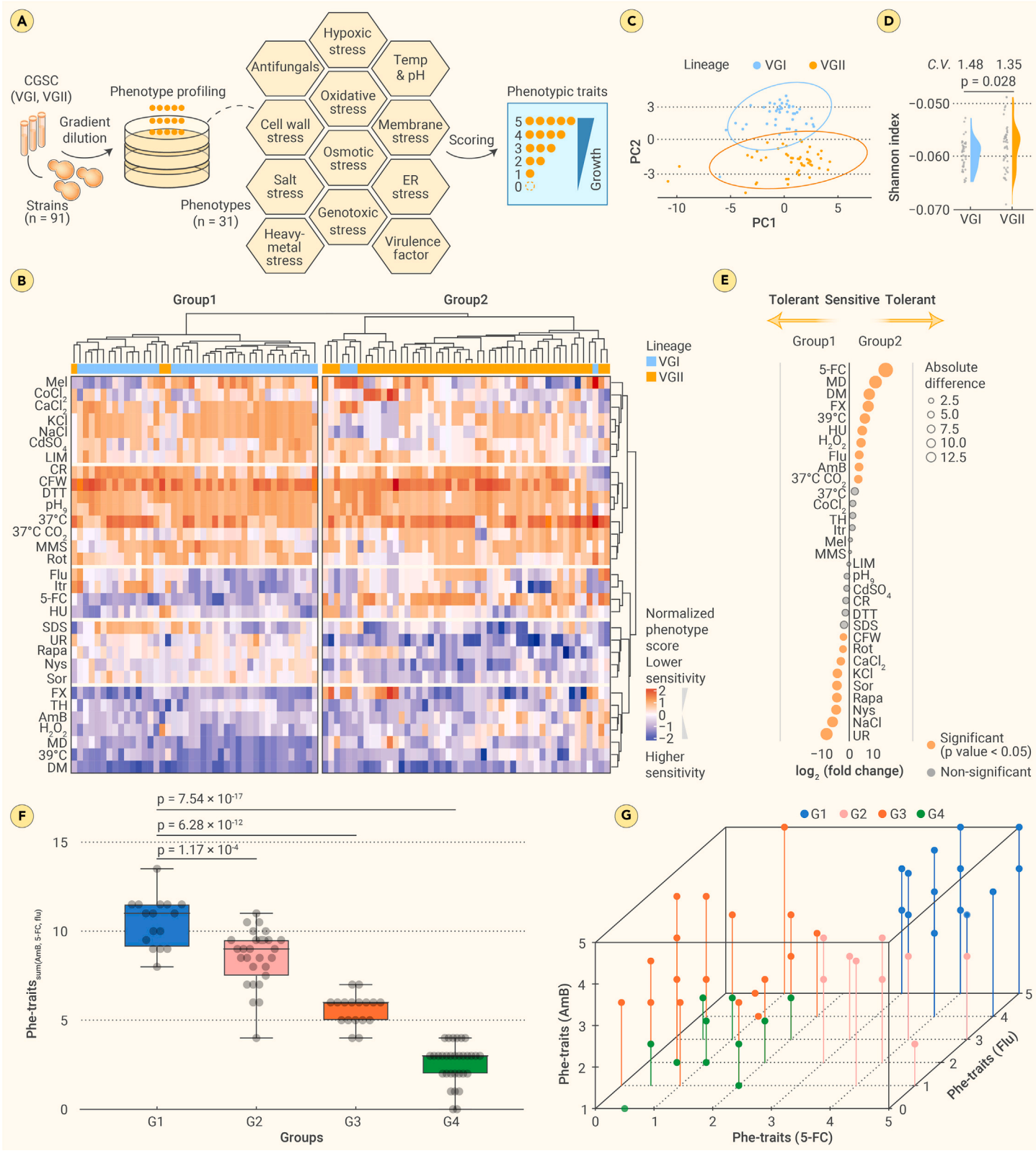


Figure 2. Systematic phenotypic profiling reveals a set of CGSC isolates with strong adaptation to three antifungals (A) Phenotypic traits of each CGSC strain were evaluated under different stress conditions. Phenotypic traits were scored using five grades based on growth fitness (0, no growth; 1, strong growth inhibition; 2, moderate growth inhibition; 3, weak growth inhibition; 4, slight growth inhibition; and 5, no growth inhibition). (B) Hierarchical clustering of different phenotypic traits in CGSC strains. The heatmap was organized into columns for phenotypes and rows for strains. Blue and red indicate high and low sensitivities to stress, respectively. The data are presented as individual values from two biological replicates. Mel, melanin production; LIM, iron limitation; CR, Congo red; CFW, calcofluor white; DTT, dithiothreitol; MMS, methyl methanesulfonate; Rot, rotenone; Flu, fluconazole; Itr, itraconazole; 5-FC, 5-flucytosine; HU, hydroxyurea; SDS, sodium dodecyl sulfate; UR, urease production; Rapa, rapamycin; Nys, nystatin; Sor, sorbitol; FX, fludioxonil; TH, *tert*-butyl hydroperoxide; AmB, amphotericin B; MD, menadione; DM, diamide. (C) PCA showing phenotypic trait similarities between the CGSC strains. Each point represents the mean of two biological replicates. (D) Violin plot of Shannon index differences based on phenotypic traits between the VGI and VGII strains. Coefficients of variation (CVs) are shown. Statistical differences were determined using a two-tailed unpaired Student's t test. (E) The differences in phenotypic traits between group 1 and group 2 in (A) are shown in a bubble chart.

(legend continued on next page)

semi-quantitative phenome assessment strategy as previously reported by Jung et al.[31] The phenotypic assessment included 31 different growth condition tests for virulence factor production and adaptation to various stresses (Figure 2A). Due to the extreme rarity of VGIII, VGIV, and VGV strains, only the VGI and VGII strains (43 VGI isolates and 48 VGII isolates) were subjected to these stress assessment tests (Figures 2A and S1).

For phenotypic traits, we evaluated the relative growth fitness of each CGSC strain under different stress conditions, generating 2,821 strain-phenotype associations (Figures 2A and 2B). Pearson's correlation analyses of the phenotypic data indicated high correlations between phenotypes related to similar stresses, such as cell wall stress (calcofluor white and Congo red), osmotic stress (NaCl and KCl), and azole stress (fluconazole and itraconazole) (Figure S3). These results supported the reliability of the phenome method employed.

PCA-based evaluation of the phenotypic profiles indicated that the VGI and VGII strains were largely represented as two separate groups, suggesting a strong interlineage distinction in their ability to tolerate different stresses (Figure 2C). However, the subclades of VGI and VGII did not cluster independently, revealing the complexity of genetic-phenotypic connections between strains of the same species (Figure S4). Using the Shannon index, we found a higher variation in stress adaptation among the VGII strains than among the VGI strains (Figure 2D). This difference could be explained by the higher within-species genetic diversity of the VGII strains than that of the VGI strains, as illustrated in Figure 1.

We also used hierarchical clustering to assess the similarity of the phenotypic profiles between the different strains (Figure 2B). These results are consistent with the PCA data and show that the CGSC strains can be divided into two groups: group 1 (VGI 92.7%) and group 2 (VGII 91.8%), dominated by VGI and VGII, respectively (Figure 2C).

We further evaluated the contribution of each phenotype to the differences in phenotypic profiles between groups 1 and 2. As shown in Figure 2E, the VGI group exhibited an increased ability to produce urease and significantly greater resistance to eight stresses, including those related to salt (NaCl, KCl, and $CaCl_2$), antifungal agents (nystatin and rapamycin), sorbitol, rotenone, and calcofluor white. Conversely, the VGII group demonstrated stronger adaptation to ten stresses, growing more robustly at 37°C in a 5% CO_2 atmosphere, at 39°C, and in the presence of four antifungal agents (5-FC, fluconazole, AmB, and fludioxonil), three oxidative stressors (menadione, diamide, and H_2O_2), and a genotoxic stressor (hydroxyurea).

Among the stress phenotypes that differed between the groups, we focused on AmB, fluconazole, and 5-FC, the most commonly used antifungal drugs to treat CGSC infections.[7] Using a Gaussian mixture modeling (GMM) approach based on the phenotypic traits reflecting overall tolerance to these three antifungal drugs, we identified a set of strains—approximately 17.6% (16/91) of the phenotypically tested isolates—that appeared to have better adaptation to all three drugs (Figures 2F and S1). A 3D plot of the GMM-based clustering results indicated that these strains could be grouped independently of the others (Figure 2G). We referred to these strains as multidrug hyperresistant (MHR) strains.

A functional pan-genomic strategy identifies VGII-specific genes involved in drug resistance in an MHR strain

As all MHR strains were VGII strains, we hypothesized that VGII-specific genes might contribute to multidrug resistance. To test this hypothesis, we utilized *de novo* genome assemblies to define and characterize the pan-genome (Figure S1). We identified 9,443 gene clusters in the CGSC strains, condensed into 7,983 non-redundant orthogroups. The CGSC pan-genome comprised a core genome of 4,688 orthogroups present in all 91 isolates (58.7% of the pan-genome), 843 soft core orthogroups found in >95% of the isolates (10.6% of the pan-genome), 1,643 shell genes in 5%–95% of the isolates (20.6% of the pan-genome), and a cloud genome of 809 genes present in less than 5% of the isolates (10.1% of the pan-genome; Figure 3A). On average, each isolate contained 6,275 orthogroups and 1,587 orthologous accessory gene clusters, corresponding to 21.4%–27.0% of the total genome of the strain. The pan-genome was considered closed, as the number of pan-genes did not increase after the addition of approximately 80 genomes (Figure 3B).

Based on the pan-genomic data, we identified 12 genes present in over 90% of the VGII genomes but absent in all VGI genomes (Figure 3C). Among these, eight genes were found in all MHR strains. Through transcriptomic assays on the MHR strain R265 cultured under host-mimicking conditions (DMEM medium at 37°C in a 5% CO_2 atmosphere), we observed that four out of these eight VGII-specific genes showed either no detectable transcription or very low levels (transcripts per million [TPM] < 5) and were excluded from further functional assays. The remaining four genes underwent gene knockout experiments, and the resulting mutants were evaluated for their ability to adapt to various antifungal agents.

As depicted in Figure 3D, all of these VGII-specific genes were implicated in cellular adaptation to at least one antifungal drug. Mutants lacking CNBG_9317 or CNBG_0992 exhibited increased sensitivity to fluconazole and 5-FC, respectively, while CNBG_3001Δ and CNBG_9316Δ mutants showed heightened susceptibility to fluconazole and 5-FC. However, none of the VGII-specific genes appeared to affect the adaptation of R265 to AmB. Collectively, these findings suggest that while VGII-specific genes contribute to cellular adaptation to one or two antifungal drugs, they do not fully confer resistance to all three antifungal drugs.

Pan-transcriptomic assessment enables identification of key multidrug resistance genes in MHR strains

To further identify the determinants of multidrug resistance, we used a pan-transcriptome method based on the fact that variations in gene expression in different strains can be related to phenotypic traits, as demonstrated in different fungal species.[26,32,33]

We performed high-coverage RNA sequencing (RNA-seq) analyses on 91 CGSC strains, and the RNA-seq reads from each strain were aligned to the corresponding genome to maximize the coverage of transcripts from the CGSC strains (Figure S1). We detected a total of 6,082 transcripts (average TPM > 10) with an expression coverage of 98.4% for core orthogroups, 95.5% for softcore orthogroups, and 41% for shell and cloud orthogroups. Compared with using only the genome of strain R265 as a reference genome, the pan-genome-based approach yielded 970 additional transcripts, the majority of which (586 transcripts) could be aligned to the publicly available genomes of other Cryptococcus isolates.

To identify the determinants underlying CGSC adaptation to different stresses, we conducted weighted gene co-expression network analysis based on pan-transcriptomic data to explore potential gene modules significantly correlated with phenotypic traits related to each stress condition (Figures 4A and S1). This approach identified 16 gene modules (Figure 4A), with 11 out of 16 modules associated with at least two phenotypes, indicating the pleiotropic roles of genes in orchestrating fungal adaptation to various stresses.

One specific gene module, highlighted in turquoise color (Figure 4B), was significantly and positively correlated with MHR levels. Gene Ontology (GO) analysis of genes within this module revealed enrichment in metabolic processes (such as fatty acid and lipid metabolism) and oxidation and redox processes (Figure 4C). Previous studies have implicated these biological processes in fungal cellular responses to antifungal drugs,[34–37] underscoring the potential importance of genes within the turquoise-colored module in conferring antifungal resistance.

To investigate the potential functional role of turquoise color block genes in multidrug resistance, we selected the top 10 candidate genes with the highest correlation with the MHR phenotype (Table S2). Each of these genes was deleted in the R265 strain. Gene knockouts were achieved for nine genes; however, we failed to obtain CNBG_5527 mutants after multiple attempts, suggesting that it may be necessary for R265 growth. Indeed, the homolog of this gene has been shown to be essential in *Saccharomyces cerevisiae*.[38]

All nine identified genes affected the adaptation of R265 cells to at least two antifungal drugs (Table S2). Notably, four genes (CNBG_1863, CNBG_3326, CNBG_6193, and CNBG_4483) were found to significantly reduce resistance to all three antifungal agents upon their deletion (Figures 4D and 4E). The roles of

Statistical significance was determined using a two-tailed unpaired Student's t test. (F) Boxplot showing the clustering of the phenotypic traits reflecting tolerance to the three antifungal drugs using the Gaussian mixture modeling (GMM)-based method. Group 1 (G1) represents MHR strains. Boxplots display the 25th, 50th (median), and 75th quantiles as well as the minimum and maximum values. Two-tailed unpaired Student's t test. (G) 3D plot of the GMM-based clustering results. Each point represents the sum of the phenotypic traits (phe-traits) of the three antifungal agents.

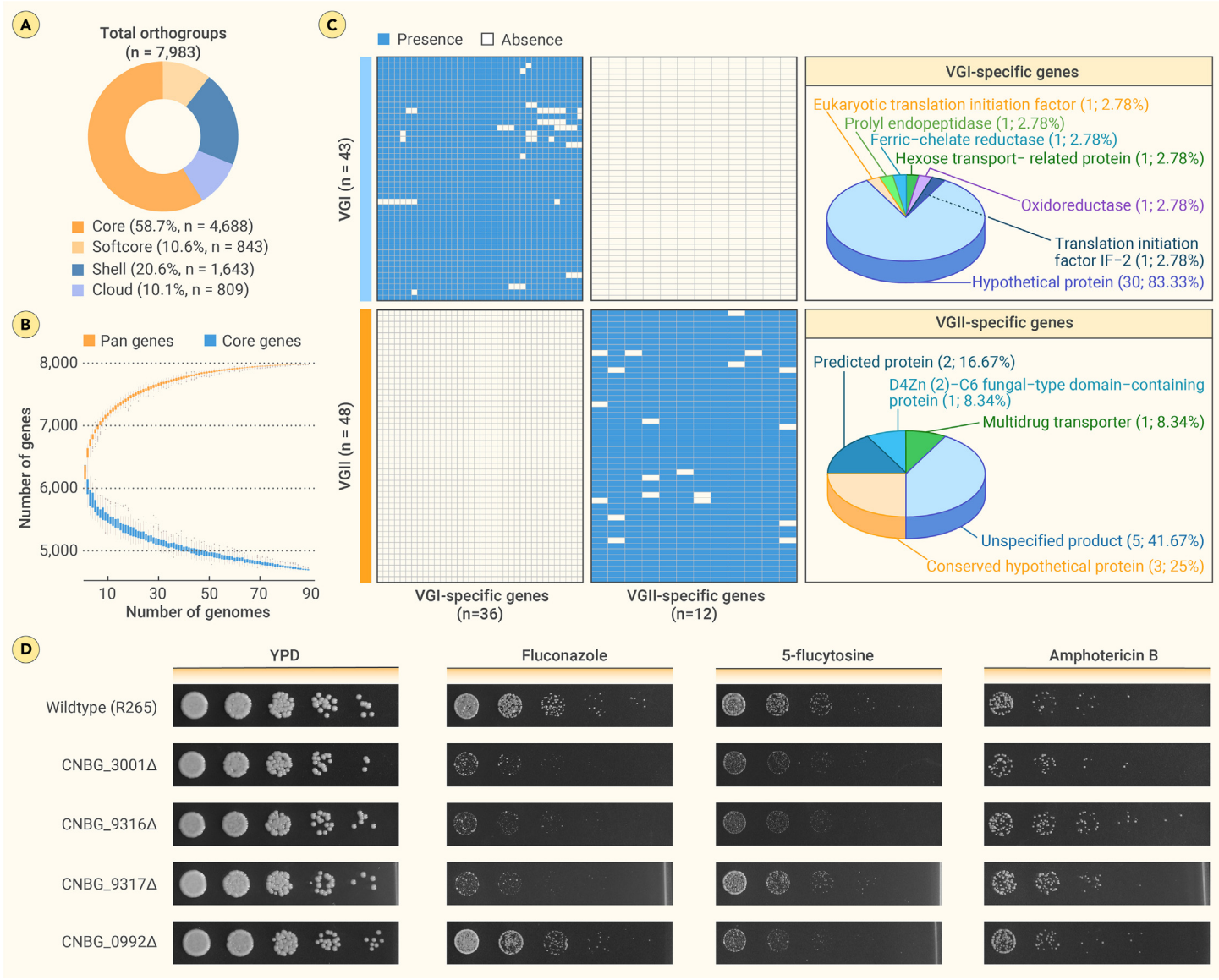


Figure 3. Functional pan-genomics analysis reveals the impact of VGII-specific genes on drug resistance (A) Percentage of core, softcore, shell, and cloud orthogroups in pan-genome analysis. (B) Pan and core gene size as the number of included genomes increases; boxplots display the 25th and 75th percentiles as well as the minimum and maximum values. (C) Presence/absence matrix and functional annotation of VGI- and VGII-specific genes identified by pan-genome analysis. (D) Wild-type (R265) and four VGII-specific genes mutant cells were spotted onto YPD plates with amphotericin B (0.4 μg/mL), fluconazole (15 μg/mL), or 5-flucytosine (300 μg/mL). Images are representative of more than 3 trials.

these genes in multidrug resistance have not been previously characterized in CGSC pathogens or other fungal species (Table S4).

To further evaluate the impact of these genes on multidrug resistance, their respective mutants underwent transcriptomic analyses to identify differentially expressed genes (DEGs) influenced by their absence (Table S3). We observed significant enrichment of DEGs within the turquoise-colored gene module associated with MHR, indicating their crucial role in this phenotype (Figure 5A). As depicted in Figure 5B, a total of 211 overlapping genes were identified across all sets of DEGs. GO analysis revealed the significant enrichment of genes involved in processes related to cellular adaptation to oxidative stress among these overlapping DEGs (Figure 5C).

Antifungals induce the generation of endogenous reactive oxygen species (ROS).[37,39–42] In this context, the ability to reduce the production of intracellular ROS or enhance antioxidant defenses is closely related to fungal adaptation to antifungals. We performed a flow-cytometry-based evaluation to compare the levels of intracellular ROS in strains with mutations in MHR-related genes to those of the parental strain R265 and found that the levels of ROS in these strains did not appear to be remarkably different, regardless of drug treatment (Figure S5). These results suggested that MHR-related genes are not involved in the generation of intracellular ROS upon antifungal challenge, implying that MHR-related genes may play a role in the inhibition of oxidative damage.

In support of this notion, mutants lacking the four identified multidrug resistance genes exhibited increased sensitivity to H_2O_2 compared to the wild-type R265 strain (Figure 5D). Additionally, phenotypic profiling revealed a significant correlation between the MHR phenotype and enhanced cell growth under oxidative stress conditions, including diamide, H_2O_2, menadione, and *tert*-butyl hydroperoxide (Figure 5E). Detailed analyses further demonstrated that MHR strains displayed greater resistance to H_2O_2 compared to other CGSC strains (Figure 5F). These findings suggest that robust tolerance to oxidative stress is closely associated with adaptation to various antifungal agents, consistent with previous studies on fungal species highlighting the interplay between oxidative stress defense mechanisms and antifungal adaptation.

A machine-learning-integrated multiomic feature analysis to predict MHR CGSC isolates

The remarkable adaptation of MHR strains to three first-line anticryptococcal drugs suggests their potential clinical significance. Phylogenetic analyses showed that the MHR strains did not belong to a specific subclade but were discretely distributed across the various subclades of VGII, suggesting that they could not be characterized by phylogeny because of the complexity of genetic variation among their genomes (Figure S6).

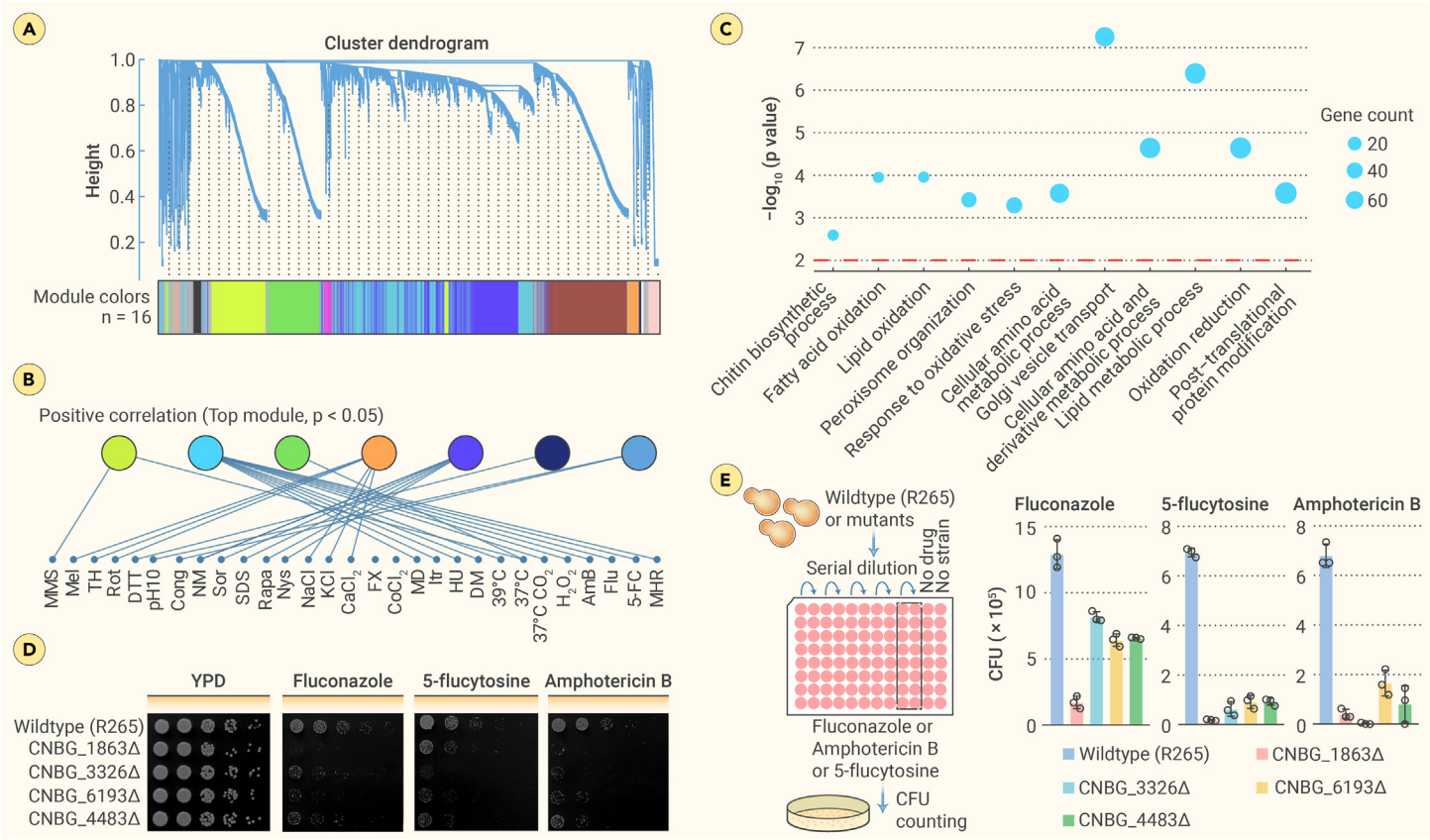


Figure 4. Pan-transcriptome-based WGCNA enables the identification of MHR-related genes (A) Hierarchical cluster tree showing the gene co-expression modules identified using weighted gene co-expression network analysis (WGCNA). Each gene is represented by a tree leaf, and each module is represented by a major tree branch. The heatmap shows the gene co-expression module assignments. (B) Degree of positive correlation between phenotypes and module eigengenes; yellow line represents the presence of a positive correlation ($p < 0.05$). (C) GO analysis of the genes included in the turquoise module. (D) Cells of wild-type (R265) and 4 MHR-related gene mutant strains were spotted onto the noted YPD plates with amphotericin B (0.45 μg/mL), fluconazole (15 μg/mL), and 5-flucytosine (300 μg/mL). Images are representative of more than three trials. (E) Quantitative antifungal susceptibility phenotypic schematic diagram (left) and histograms of colony-forming units (CFUs) of wild-type (R265) and four MHR-related gene mutant strains cultured in RPMI medium supplemented with amphotericin B (0.0625 μg/mL), fluconazole (1 μg/mL), or 5-flucytosine (1 μg/mL). Data are presented as mean ± SD.

Recently, multiomic data integrating expression and genetic variation have proven effective in identifying predictors that can accurately forecast complex disease-associated traits in humans across diverse genetic backgrounds using machine learning.[43–45] However, this approach has yet to be applied extensively to human fungal pathogens. In this study, we focused on genes identified through our multiomic approach to discover biomarkers (features) capable of predicting MHR strains (Figure S1). These genes encompassed four VGII-specific genes and ten genes from the turquoise module that exhibited the strongest correlation with the MHR phenotype. We employed model training to develop independent predictive models for MHR strains, utilizing input data that included biomarkers, the presence or absence of VGII-specific genes, and variations in the expression levels of turquoise module genes (Figure 6A).

These models are based on a multistep predictor pipeline. Inside the pipeline, each ensemble consisted of three algorithms acting in parallel: logistic regression with elastic net regularization, random forest, and XGBoost. The scores of the three algorithms are averaged to form a predictor. Model hyperparameters were optimized using a 10-fold cross-validation scheme (Figures 6A and S1).

The fully trained model was validated on independent outer layers including 22 CGSC strains. The percentages of MHR strains were similar between the training and validation datasets (Figure 6A). The optimized model achieved area under the curve (AUC) values of 0.91. In the outer validation group, the AUC predicted for MHR reached 0.78 (Figure 6B).

We investigated the significance of features used in the integrated training model and found notable variation in their contributions (coefficient of variation = 1.13) (Figure 6C). This variability suggests the potential for reducing the number of features used in model training. To test this hypothesis, we ranked the features by importance and observed that model accuracy in validation could be enhanced by excluding less influential features (Figure 6D). Notably, focusing on key features, such as the presence or absence of CNBG_9316 and variations in expression levels of CNBG_1863, CNBG_3686, CNBG_5527, and CNBG_9373, maximized the model's AUC, achieving a peak value of 0.86 (Figure 6D). Specifically, by using three features—the presence or absence of CNBG_9316 and variations in expression of CNBG_1863 and CNBG_5527—we achieved robust predictive performance (AUC = 0.85 for the training set, 0.82 for the validation set) (Figure 6D). These findings highlight the utility of pan-omic data in identifying a concise set of three biomarkers capable of accurately predicting MHR CGSC isolates.

DISCUSSION

Among CGSC pathogens, *C. gattii*/VGI and *C. deuterogattii*/VGII are widely distributed and capable of infecting healthy individuals. Using a systematic phenotypic evaluation of VGI and VGII strains, we provide a comprehensive understanding of lineage- and strain-specific survival advantages in response to various stresses relevant to CGSC tolerance and pathogenicity.[16,25–27] The data revealed significant differences between VGI and VGII populations in their ability to adapt to diverse stress conditions. PCA of their phenotypic profiles grouped VGI and VGII strains into distinct clusters, supporting the proposition that VGI and VGII, despite their evolutionary proximity, represent distinct fungal species.[18] Our findings also indicated greater heterogeneity in stress adaptation among VGII strains compared to VGI strains, likely due to VGII's greater genetic diversity as the basal lineage of CGSC, resulting in varied phenotypic traits among strains.

Phenotypic clustering highlighted significant correlations among traits associated with different stresses, indicating that certain strains exhibit robust resistance to multiple stresses. Particularly noteworthy was a subset of strains highly adaptive to AmB, fluconazole, and 5-FC—commonly used anticryptococcal drugs—which exclusively belonged to the VGII lineage. This finding aligns with previous studies emphasizing VGII's heightened clinical significance compared to other CGSC pathogens.[17,23,26,27,46]

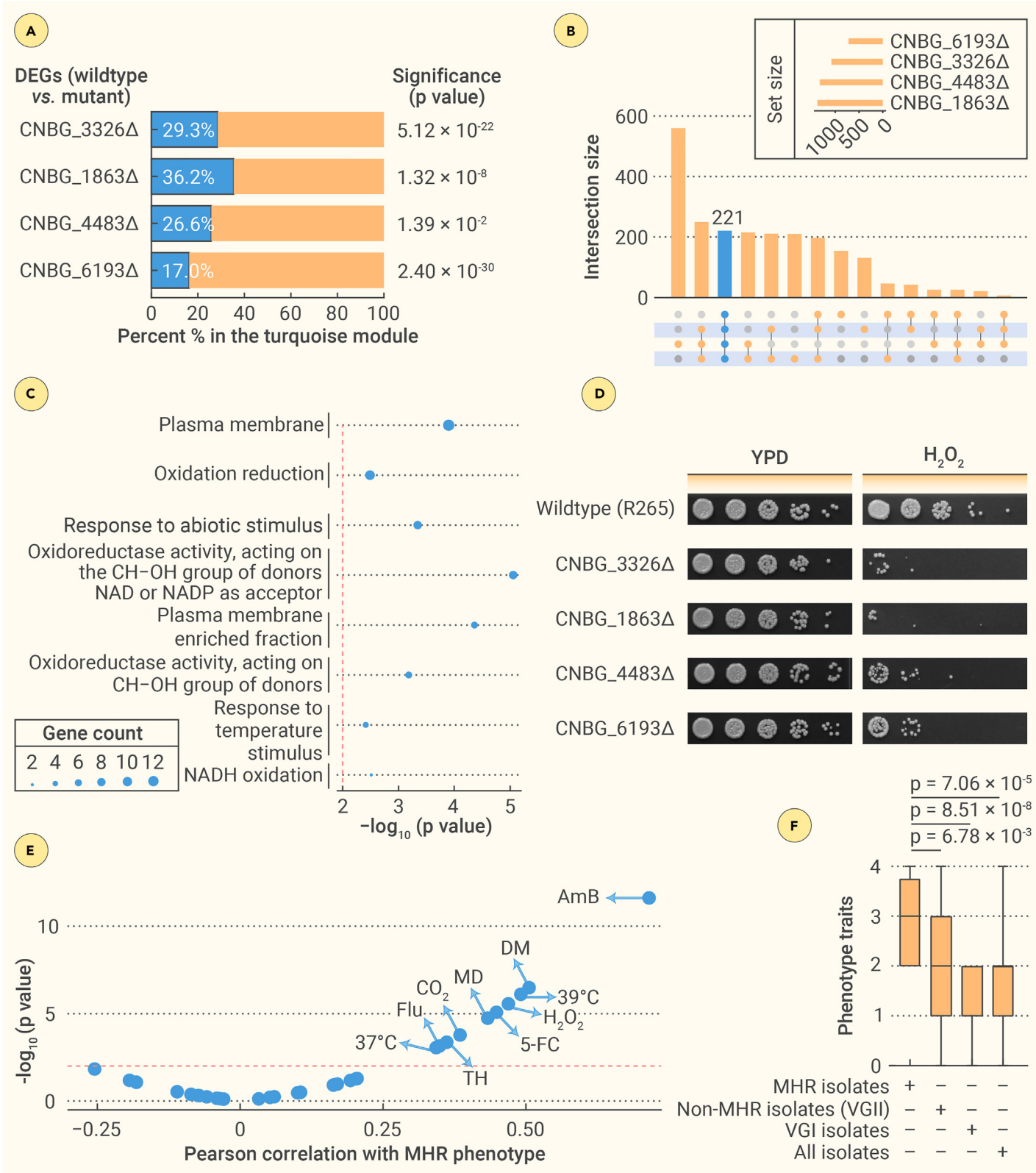


Figure 5. Functional role of the MHR-related genes in multidrug resistance (A) Percentage of turquoise module genes from Figure 4A regulated by each of the MHR-related genes; significance was evaluated using Fisher's exact test. (B) UpSet plot showing the combination of the four MHR-related genes. Yellow represents differentially expressed genes (DEGs) co-regulated by all MHR-related genes. (C) GO analysis of DEGs co-regulated by all four MHR-related genes. (D) Wild-type (R265) and the four MHR-related gene mutant strains were spotted onto YPD plates containing 0.0175% H_2O_2. Images are representative of more than 3 trials. (E) Pearson's correlation between the MHR phenotype and various stresses. The dashed red line indicates p = 0.01. DM, diamide; MD, menadione; CO_2, 37°C 5% CO_2; TH, *tert*-butyl hydroperoxide; Flu, fluconazole; 5-FC, 5-flucytosine; AmB, amphotericin B. (F) Comparison of tolerance to H_2O_2 stress among the different isolates. Boxplots display the 25th, 50th (median), and 75th quantiles as well as the minimum and maximum values. Two-tailed unpaired Student's t test.

We employed a pan-omic approach to predict genes associated with the MHR phenotype and identify novel multidrug resistance determinants in the CGSC. Among the nine candidate genes predicted and amenable to deletion, knockout mutants of all were found to be resistant to at least two of three antifungal drugs, with four genes (44.4%) showing reduced resistance to all three drugs, validating our predictive approach for multidrug resistance genes.

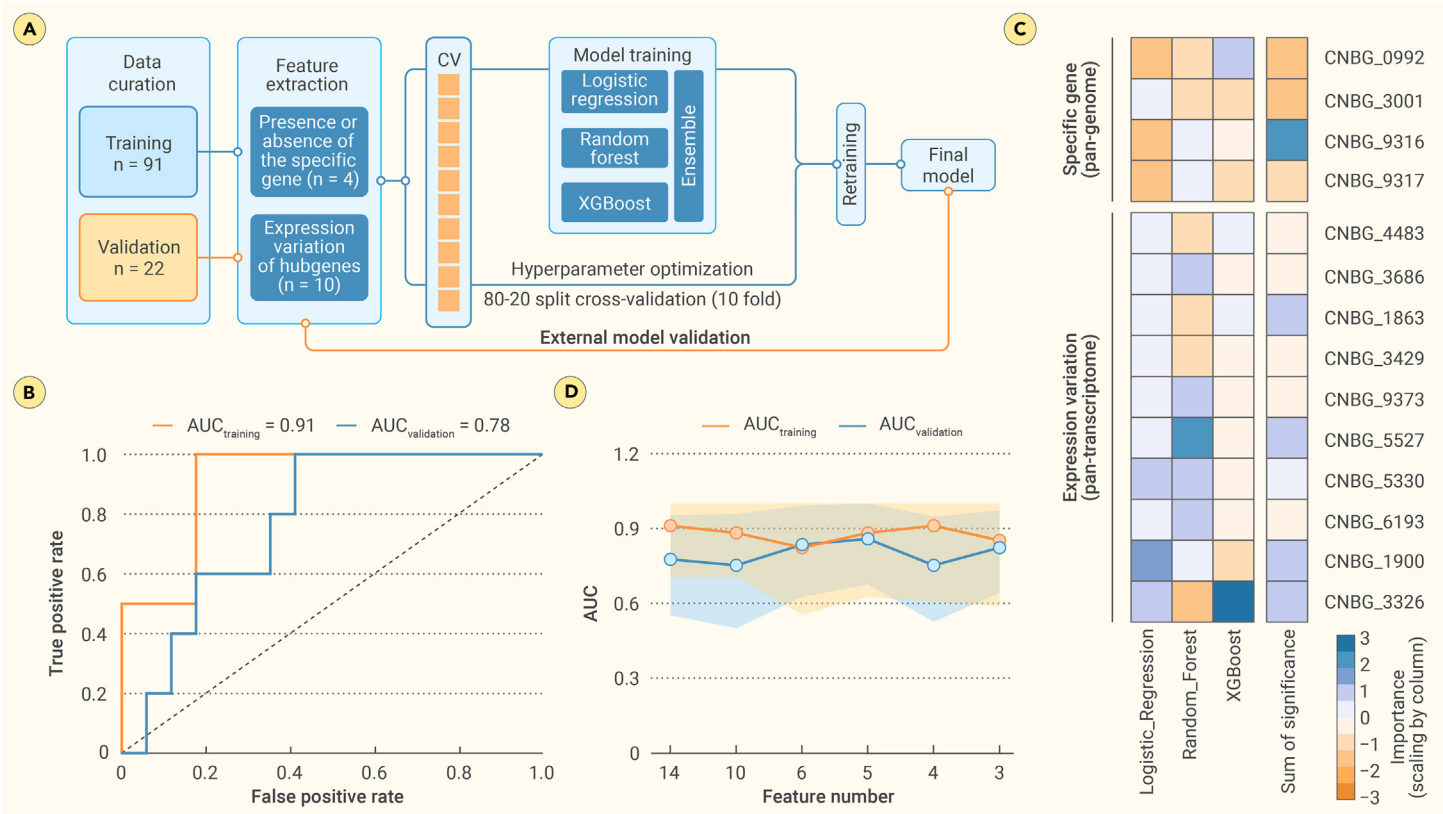


Figure 6. Prediction of MHR CGSC isolates using machine learning-integrated multiomic features (A) Schematic of the machine learning framework. CV, cross-validation. (B) Receiver operating characteristic (ROC) curves for the prediction of MHR CGSC isolates in the training and validation sets. (C) Feature importance based on the three model algorithms As indicated by the colored bars, higher importance values are shown in blue, whereas lower importance values are shown in orange. (D) Area under the curve (AUC) in training and validation sets after eliminating features of low importance. The 95% confidence intervals were calculated using bootstrap sampling, with shaded colors representing the lower and upper bounds of the training and validation sets, respectively.

Our phenome data revealed significant correlations between phenotypic traits linked to MHR and those related to oxidative stress. Prior studies have implicated antioxidant defenses in resistance against AmB, fluconazole, and 5-FC across various fungi.[47–49] These findings, coupled with our phenotypic correlations, suggest that the MHR phenotype in these strains may stem from robust antioxidant capacities. Supporting this notion, our data demonstrated enhanced tolerance to oxidative agents in MHR strains compared to others. Nevertheless, elucidating the mechanisms underlying the heightened oxidative stress tolerance in MHR strains requires further investigation.

Using a machine learning approach, we identified features that enabled accurate prediction of MHR strains (AUC > 0.85). Importantly, our method is applicable with small sample sizes, making it versatile for predicting clinically relevant phenotypic traits in other pathogens. Our study represents a significant advance in utilizing pan-omics methods to explore novel multidrug resistance genes and predict clinically relevant MHR strains within the CGSC. While our focus was on CGSC pathogens, we anticipate that these findings and methodologies will inspire further research and provide crucial experimental frameworks for identifying resistance determinants and predicting resistant variants across different fungal pathogenic species or species complexes.

MATERIALS AND METHODS

See the supplemental information for details.

REFERENCES

1. He, M.Q., Zhao, R.L., Liu, D.M., et al. (2022). Species diversity of basidiomycota. Fungal Divers. ***114***(1): 281–325.
2. Byrnes, E.J., 3rd, Bartlett, K.H., Perfect, J.R., et al. (2011). *Cryptococcus gattii*: an emerging fungal pathogen infecting humans and animals. Microbes Infect. ***13***(11): 895–907.
3. Loftus, B.J., Fung, E., Roncaglia, P., et al. (2005). The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science ***307***(5713): 1321–1324.
4. Zhao, Y., Ye, L., Zhao, F., et al. (2023). *Cryptococcus neoformans*, a global threat to human health. Infect. Dis. Poverty ***12***(1): 20.
5. Iyer, K.R., Revie, N.M., Fu, C., et al. (2021). Treatment strategies for cryptococcal infection: challenges, advances and future outlook. Nat. Rev. Microbiol. ***19***(7): 454–466.
6. May, R.C., Stone, N.R.H., Wiesner, D.L., et al. (2016). *Cryptococcus*: from environmental saprophyte to global pathogen. Nat. Rev. Microbiol. ***14***(2): 106–117.
7. Perfect, J.R., Dismukes, W.E., Dromer, F., et al. (2010). Clinical practice guidelines for the management of cryptococcal disease: 2010 update by the infectious diseases society of America. Clin. Infect. Dis. ***50***(3): 291–322.
8. WHO (2022). In Guidelines for Diagnosing, Preventing and Managing Cryptococcal Disease Among Adults, Adolescents and Children Living with HIV (World Health Organization).
9. Gullo, F.P., Rossi, S.A., Sardi, J.d.C.O., et al. (2013). Cryptococcosis: epidemiology, fungal resistance, and new alternatives for treatment. Eur. J. Clin. Microbiol. Infect. Dis. ***32***(11): 1377–1391.
10. Priest, S.J., Yadav, V., Roth, C., et al. (2022). Uncontrolled transposition following RNAi loss causes hypermutation and antifungal drug resistance in clinical isolates of *Cryptococcus neoformans*. Nat. Microbiol. ***7***(8): 1239–1251.
11. Firacative, C., Meyer, W., and Castañeda, E. (2021). *Cryptococcus neoformans* and *Cryptococcus gattii* species complexes in Latin America: a map of molecular types, genotypic diversity, and antifungal susceptibility as reported by the Latin American cryptococcal study group. J. Fungi (Basel) ***7***(4): 282.
12. Bermas, A., and Geddes-McAlister, J. (2020). Combatting the evolution of antifungal resistance in *Cryptococcus neoformans*. Mol. Microbiol. ***114***(5): 721–734.
13. Kwon-Chung, K.J., Fraser, J.A., Doering, T.L., et al. (2014). *Cryptococcus neoformans* and *Cryptococcus gattii*, the etiologic agents of cryptococcosis. Cold Spring Harbor Perspect. Med. ***4***(7): a019760. https://doi.org/10.1101/cshperspect.a019760.
14. Chan, M., Lye, D., Win, M.K., et al. (2014). Clinical and microbiological characteristics of cryptococcosis in Singapore: predominance of *Cryptococcus neoformans* compared with *Cryptococcus gattii*. Int. J. Infect. Dis. ***26***: 110–115.
15. Jin, L., Cao, J.R., Xue, X.Y., et al. (2020). Clinical and microbiological characteristics of *Cryptococcus gattii* isolated from 7 hospitals in China. BMC Microbiol. ***20***(1): 73.
16. Ke, W., Xie, Y., Hu, Y., et al. (2022). A forkhead transcription factor contributes to the regulatory differences of pathogenicity in closely related fungal pathogens. mLife ***1***(1): 79–91.
17. Chen, S.C.A., Meyer, W., and Sorrell, T.C. (2014). *Cryptococcus gattii* infections. Clin. Microbiol. Rev. ***27***(4): 980–1024.
18. Hagen, F., Khayhan, K., Theelen, B., et al. (2015). Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet. Biol. ***78***: 16–48.
19. Hitchcock, M., and Xu, J. (2023). Global analyses of multi-locus sequence typing data reveal geographic differentiation, hybridization, and recombination in the *Cryptococcus gattii* species complex. J. Fungi (Basel) ***9***(2): 276.

20. Francisco, E.C., de Jong, A.W., and Hagen, F. (2021). *Cryptococcosis* and *Cryptococcus*. Mycopathologia ***186***(5): 729–731.
21. Herkert, P.F., Hagen, F., Pinheiro, R.L., et al. (2017). Ecoepidemiology of *Cryptococcus gattii* in developing countries. J. Fungi (Basel) ***3***(4): 62.
22. Bellet, V., Roger, F., Krasteva, D., et al. (2022). Multilocus sequence typing of strains from the *Cryptococcus gattii* species complex from different continents. Mycoses ***65***(1): 88–96.
23. Byrnes, E.J., 3rd, Bildfell, R.J., Frank, S.A., et al. (2009). Molecular evidence that the range of the Vancouver Island outbreak of *Cryptococcus gattii* infection has expanded into the Pacific Northwest in the United States. J. Infect. Dis. ***199***(7): 1081–1086.
24. Kidd, S.E., Hagen, F., Tscharke, R.L., et al. (2004). A rare genotype of *Cryptococcus gattii* caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). Proc. Natl. Acad. Sci. USA ***101***(49): 17258–17263.
25. Fernandes, K.E., Dwyer, C., Campbell, L.T., et al. (2016). Species in the *Cryptococcus gattii* complex differ in capsule and cell size following growth under capsule-inducing conditions. mSphere ***1***(6). e00350–16.
26. Xue, X., Zang, X., Xiao, M., et al. (2022). Significance of differential expression profiles of ABC transporters in azole susceptibility between *Cryptococcus gattii* VGI and VGII strains. Med. Mycol. ***60***(7): myac035. https://doi.org/10.1093/mmy/myac035.
27. Farrer, R.A., Ford, C.B., Rhodes, J., et al. (2018). Transcriptional heterogeneity of *Cryptococcus gattii* VGII compared with non-VGII lineages underpins key pathogenicity pathways. mSphere ***3***(5). e00445–18.
28. Montoya, M.C., Magwene, P.M., and Perfect, J.R. (2021). Associations between *Cryptococcus* genotypes, phenotypes, and clinical parameters of human disease: a review. J. Fungi (Basel) ***7***(4): 260.
29. Beale, M.A., Sabiiti, W., Robertson, E.J., et al. (2015). Genotypic diversity is associated with clinical outcome and phenotype in cryptococcal meningitis across southern Africa. PLoS Negl. Trop. Dis. ***9***(6): e0003847. https://doi.org/10.1371/journal.pntd.0003847.
30. Farrer, R.A., Desjardins, C.A., Sakthikumar, S., et al. (2015). Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio ***6***(5): e00868-15. https://doi.org/10.1128/mbio.00868-15.
31. Jung, K.W., Yang, D.H., Maeng, S., et al. (2015). Systematic functional profiling of transcription factor networks in *Cryptococcus neoformans*. Nat. Commun. ***6***: 6757.
32. Kvitek, D.J., Will, J.L., and Gasch, A.P. (2008). Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. PLoS Genet. ***4***(10): e1000223. https://doi.org/10.1371/journal.pgen.1000223.
33. Angiolella, L., Stringaro, A.R., De Bernardis, F., et al. (2008). Increase of virulence and its phenotypic traits in drug-resistant strains of *Candida albicans*. Antimicrob. Agents Chemother. ***52***(3): 927–936.
34. Chayakulkeeree, M., Rude, T.H., Toffaletti, D.L., et al. (2007). Fatty acid synthesis is essential for survival of *Cryptococcus neoformans* and a potential fungicidal target. Antimicrob. Agents Chemother. ***51***(10): 3537–3545.
35. Pan, J., Hu, C., and Yu, J.H. (2018). Lipid biosynthesis as an antifungal target. J. Fungi (Basel) ***4***(2): 50.
36. Perez-de Los Santos, F.J., Garcia-Ortega, L.F., Robledo-Marquez, K., et al. (2021). Transcriptome analysis unveils Gln3 role in amino acids assimilation and fluconazole resistance in *Candida glabrata*. J. Microbiol. Biotechnol. ***31***(5): 659–666.
37. Gonzalez-Jimenez, I., Perlin, D.S., and Shor, E. (2023). Reactive oxidant species induced by antifungal drugs: identity, origins, functions, and connection to stress-induced cell death. Front. Cell. Infect. Microbiol. ***13***: 1276406. https://doi.org/10.3389/fcimb.2023.1276406.
38. Giaever, G., Chu, A.M., Ni, L., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. Nature ***418***(6896): 387–391.
39. Belenky, P., Camacho, D., and Collins, J.J. (2013). Fungicidal drugs induce a common oxidative-damage cellular death pathway. Cell Rep. ***3***(2): 350–358.
40. Sokol-Anderson, M.L., Brajtburg, J., and Medoff, G. (1986). Amphotericin B-induced oxidative damage and killing of *Candida albicans*. J. Infect. Dis. ***154***(1): 76–83.
41. Guirao-Abad, J.P., Sánchez-Fresneda, R., Alburquerque, B., et al. (2017). ROS formation is a differential contributory factor to the fungicidal action of amphotericin B and micafungin in *Candida albicans*. Int. J. Med. Microbiol. ***307***(4–5): 241–248.
42. Peng, C.A., Gaertner, A.A.E., Henriquez, S.A., et al. (2018). Fluconazole induces ROS in *Cryptococcus neoformans* and contributes to DNA damage *in vitro*. PLoS One ***13***(12): e0208471. https://doi.org/10.1371/journal.pone.0208471.
43. Liu, Z., Zhao, Y., Kong, P., et al. (2023). Integrated multi-omics profiling yields a clinically relevant molecular classification for esophageal squamous cell carcinoma. Cancer Cell ***41***(1): 181–195.e9.
44. Sammut, S.J., Crispin-Ortuzar, M., Chin, S.F., et al. (2022). Multi-omic machine learning predictor of breast cancer therapy response. Nature ***601***(7894): 623–629.
45. Reel, P.S., Reel, S., Pearson, E., et al. (2021). Using machine learning approaches for multi-omics data analysis: a review. Biotechnol. Adv. ***49***: 107739. https://doi.org/10.1016/j.biotechadv.2021.107739.
46. Applen Clancey, S., Ciccone, E.J., Coelho, M.A., et al. (2019). *Cryptococcus deuterogattii* VGIIa infection associated with travel to the Pacific Northwest outbreak region in an anti-granulocyte-macrophage colony-stimulating factor autoantibody-positive patient in the United States. mBio ***10***(1). e02733–18.
47. Sanglard, D., Coste, A., and Ferrari, S. (2009). Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. FEMS Yeast Res. ***9***(7): 1029–1050.
48. Kanafani, Z.A., and Perfect, J.R. (2008). Antimicrobial resistance: resistance to antifungal agents: mechanisms and clinical impact. Clin. Infect. Dis. ***46***(1): 120–128.
49. Yaakoub, H., Mina, S., Calenda, A., et al. (2022). Oxidative stress response pathways in fungi. Cell. Mol. Life Sci. ***79***(6): 333.

ACKNOWLEDGMENTS

We thank members from the Wang lab and Dr. Ence Yang for sharing the constructs, protocols, and scientific discussions and Dr. Tong Zhao for ploidy measurements. This work was financially supported by the National Key R&D Program of China (2021YFC2302100); the National Natural Science Foundation of China (82370005 and 82172291); the National Key R&D Program of China (2022YFC2303000 and 2021YFC230000); the CAS Interdisciplinary Innovation Team, the Beijing Research Center for Respiratory Infectious Diseases Project (BJRID2024-008 and BJRID2024-011); the R&D Program of Beijing Municipal Education Commission (KM202410025012); the Reform and Development Program of Beijing Institute of Respiratory Medicine (Ggyfz202328 and Ggyfz202418); the National Key R&D Program of China (2020YFA0907200); Shanghai Science and Technology Innovation Action Plan 2023 "Basic Research Project" (23JC1404201); the Shanghai "Belt and Road" Joint Laboratory Project (22490750200); the National Natural Science Foundation of China (82370005) and National High Level Hospital Clinical Research Funding (2022-PUMCH-C-052).

AUTHOR CONTRIBUTIONS

All authors contributed to data analysis. X.F., L.C., X.X., and L.W. designed this study; X.F., L.C., M.C., L.Z., Y.Xie and W.K. performed the experiments; L.C., N.Z., H.C., M.H., Z.S., and H.D. conducted the bioinformatics analysis; M.X., X.Z., Y.H., H.X., W.F., S.L., CC., Y.Xu., W.W., C.C., X.X., and S.S. contributed strains/materials/analysis tools; and L.C., N.Z., and L.W. wrote the manuscript with contributions from other authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.1016/j.xinn.2024.100681.

LEAD CONTACT WEBSITE

http://mycolab.im.ac.cn/wanglinqi/zzjj/201907/t20190731_507787.html.