

Cryptococcus gattii VGI Subtypes: Geographical Distribution, Molecular Traits, and Virulence Difference

Xinying Xue

Capital Medical University Affiliated Beijing Shijitan Hospital

Wei Tang

ShanghaiTech University

Xuelel Zang

Medical School of Chinese PLA: Chinese PLA General Hospital

Weixin Ke

Chinese Academy of Sciences

Yuxia Liu

Peking Union Medical College Hospital

Zhiqiang Li

Shanghai Jiaotong University: Shanghai Jiao Tong University

Mei Xie

Medical School of Chinese PLA: Chinese PLA General Hospital

Lifeng Wang

Chinese PLA General Hospital

Yiting Shao

ShanghaiTech University

Chongchong Wu

Chinese PLA General Hospital

Jie Gao

Chinese PLA General Hospital

Xidong Ma

Medical School of Chinese PLA: Chinese PLA General Hospital

Sheng Zhao

Peking Union Medical College Hospital

Pengfei Liu

Capital Medical University

Qing Zhang

Shanghai Jiaotong University: Shanghai Jiao Tong University

Hui Deng

Capital Medical University Affiliated Beijing Shijitan Hospital

Jialin Song

Weifang Medical University

Hengyu Deng

Peking University

Shikun Guo

Peking University

Xue Wang

China Agricultural University

Yemei Huang

Capital Medical University

Chengwen Gao

Qingdao University

Meng Zhou

Binzhou Medical University

Sanhong Liu (✉ liush@shutcm.edu.cn)

Shanghai University of Traditional Chinese Medicine <https://orcid.org/0000-0002-8695-0343>

Meng Xiao

Peking Union Medical College Hospital

Linqi Wang

Chinese Academy of Sciences

Dingxia Shen

Chinese PLA General Hospital

Yongyong Shi

Shanghai Jiaotong University: Shanghai Jiao Tong University

Liye Zhang

ShanghaiTech University

Research

Keywords: Cryptococcus gattii, whole-genome sequencing, multi-locus sequence typing, Single Nucleotide Polymorphism

Posted Date: October 5th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-942229/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: As a life-threatening fungus, *Cryptococcus gattii* (*C gattii*) species complex is emerging worldwide. However, the geographical distribution, molecular traits, and virulence difference are poorly characterized in China.

Results: From 2011 to 2017, we collected 32 strains of *C gattii* from 18 hospitals across China, of which 27 [84.4%] strains molecular traits were profiled by whole-genome sequencing (WGS) and multi-locus sequence typing (MLST) and compared with strains previously described in China from 2006 to 2020. Totally 119 clinical cases caused by *C gattii* strains (87 in previous reports and 32 in our study) distributed widely in 20 provincial-level administrative regions of China, of which 114 strains molecular types were obtained. The majority molecular type was VGI (81/114, 71.1%) and the other was VGII (33/114, 28.9%). Four major subtypes of VGI (VGIa, VGIb, VGIc, and VGI d) were revealed from global *C gattii* VGI (n=308), respectively accounting for 52.9% (163/308), 36.0% (111/308), 3.9% (12/308), and 4.2% (13/308). The other nine strains could not be assigned to these four subtypes clearly. Our clinical data suggested that VGIb cases had a worse clinical outcome than VGIa, which was consistent with *in vitro* and *in vivo* experiments. In addition, a candidate virulence SNP on SOD2 in VGIa was initially identified by comparing high-quality de novo reference genome.

Conclusions: The geographical distribution of *C gattii* species complex was first described in China. *C gattii* VGI could be clearly segregated into four major subtypes based on genomics profiles and VGIb was more virulent than VGIa in China. Our study suggests the molecular type of *C gattii* is necessary for personalized treatment in clinic.

Introduction

C gattii species complex is a special fungus with a high fatality, which mainly infects the lung and central nervous system (CNS) of patients.¹⁻³ In terms of global geographical distribution, *C gattii* has occurred in 44 countries on six continents and the number is still increasing steadily.^{4,5} However, there has been no systematic description on the geographical distribution of *C gattii* species complex in China.

Genome sequencing on *C gattii* in recent years has become a powerful tool to reveal novel lineages as well as genetic variations that were associated with phenotypes and virulence in different lineages and/or subtypes within the same lineage.⁶⁻⁹ *C gattii* can be divided into six unique molecular types (VGI-VGVI).^{5,6} VGII is the main molecular type in clinical strains across the world and the subtype VGIIa is more virulent than others.¹⁰ However, VGI is an emerging and more abundant molecular type in Asia, Australia, and Europe¹¹, and some studies showed obvious differences in virulence among individual VGI strains.^{8,12} Therefore, we suspected that some subtypes of *C gattii* VGI might be more virulent, but no subtype classification system was available for *C gattii* VGI up to now.

In this study, we aimed to answer whether *C gattii* VGI could be genetically categorized into subgroups, hoping to explain difference in virulence among individual VGI strains. First of all, we collected the largest WGS data of VGI strains, from home (n=21 in China) and abroad (n=24). We used these data (n=45) to construct a phylogenetic tree and identify four VGI subtypes. Considering very limited WGS data was available for VGI strains, we developed a novel computational pipeline to predict subtypes based on MLST to generalize our conclusion on subtype prevalence and geographical distribution patterns. In addition to multi-omics (WGS and RNA-Seq) experiments, we also performed *in vitro* and *in vivo* evaluation to characterize the virulence trait of divergent molecular subtypes of *C gattii* VGI strains. Finally, candidate subtype specific SNP was also predicted to explain the difference in virulence.

Materials And Methods

Ethical statement

The study protocol was approved by the Medical Ethics Committee of Beijing Shijitan Hospital: sjtky11-1x-2020 (20).

Strains of *C gattii* in previous reports and their molecular types

We used "*Cryptococcus gattii*" and "China (or specific location)" to search all related publications in PubMed and China National Knowledge Infrastructure (CNKI). After screening abstracts and reviewing full texts, we found 22 studies containing 86 strains between 2006 and 2020 with molecular types.

Strain isolation and molecular types of *C gattii* in our study

32 clinical *C gattii* strains from 18 hospitals between 2011 and 2017 in China were isolated and were stored at -80°C in glycerol. After subculture, the isolates were re-identified by canavanine glycine bromothymol blue (CGB) agar or matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS, Bruker), but five [15.6%] of them could not be recovered due to improper storage. The molecular type (VGI and VGII) was determined by traditional PCR based MLST.¹³

Clinical data of patients with *C gattii* in our study

We retrospectively collected clinical data of 32 patients infected with *C gattii*, including gender, age, onset address, immune status, contact history, symptoms, lung CT, brain MRI presentation, intracranial pressure (ICP), biochemistry of cerebrospinal fluid (CSF), initial diagnosis, strain source, strain isolation time, admission to diagnosis time, treatment, survival prognosis, final diagnosis, and genotyping (table 1).

DNA/RNA Sequencing for *C gattii* strains

DNA extraction was performed similarly as previous report.¹⁴ All genomes were sequenced by Illumina HiSeq X Ten platform by 150 bp paired-end sequencing. Nanopore Sequencing was performed on four representative isolates (three VGIIb and one VGIIa, appendix p 6). RNA was extracted by following the

protocol from commercial Kit (CW BIO, CW0581, China) and was sequenced by Illumina HiSeq X Ten platform by 150 bp paired-end sequencing.

DNA-Seq data of *C gattii* strains analyzed from publicly available data

Additional public *C gattii* sequencing data was downloaded from NCBI (appendix p 7-8), which was then converted to fastq format by fastq-dump, version 2.8.2.¹⁵ Then the adapters were removed by fastp, version 0.20.0.¹⁶

Variant calling

Each Illumina data was aligned to the reference genome (VGIa reference WM276 in phylogenetic analysis and VGIb assembled reference from our G8 strain in SNP calling analysis) by bwa-mem2.¹⁷ To call variants, the Picard tools (<http://picard.sourceforge.net>) were used for pre-processing, followed by GATK HaplotypeCaller, version v4.1.8.1,¹⁸ with recommended filtering parameters. The final vcf file was annotated with a customized database by SnpEff, version 4.3t.¹⁹ In addition, ploidies were estimated based on the ratio of heterozygous sites given by the visualized output of ploidy NGS, version 3.1.2,²⁰ and copy number variation (CNV) was called by CNV kit, version 0.9.6.²¹

Phylogenetic analysis

The single nucleotide polymorphisms (SNPs) were retained if all samples were homozygous, and then were converted to phylip format, which was used to construct a phylogenetic tree by RAxML, version 8.2.12, with 1000 bootstraps.²² VGI strain R265 was set as the outgroup, and the final phylogenetic tree was visualized by iTOL, version 6.²³

Multiple plate-based virulence assays

The strains were incubated in YPD liquid medium at 30°C for 16 h, centrifuged, and then washed twice with phosphate-buffered saline (PBS). The strains were uniformly adjusted to 5×10^7 cells/ml, and then gradient diluted to 10^7 cells/ml, 2×10^6 cells/ml, 4×10^5 cells/ml, 8×10^4 cells/ml, and 1.6×10^4 cells/ml. The cells (3 ml from each concentration) were respectively plated onto yeast extract-peptone-dextrose (YPD) agar medium to observe the growth at different temperatures (30°C, 37°C, and 39°C), plated onto YPD agar medium containing hydrogen peroxide (0.03% H₂O₂) and diamide (3 mM) to observe the growth at 30°C plated onto YPD agar medium containing different concentrations of antifungal drugs to observe the growth at 30°C (Flucytosine: 300 µg/ml; Fluconazole: 20 µg/ml; Amphotericin B: 0.75 µg/ml).²⁴

Induction of melanin using minimal medium containing L-3,4-dihydroxyphenylalanine (L-DOPA): Culture as described above, the strains were also uniformly adjusted to 5×10^7 cells/ml, and then gradient diluted to 10^7 cells/ml, 2×10^6 cells/ml, 4×10^5 cells/ml, 8×10^4 cells/ml, and 1.6×10^4 cells/ml. The cells (3 ml from each concentration) were respectively plated onto minimal medium containing L-3,4-

dihydroxyphenylalanine (L-DOPA) and incubated at 30°C for three days. Incubated at 30°C for three days and photographed for melanin production.²⁵

Induction of capsule using Dulbecco's minimal essential media (DMEM) medium: The culture conditions are as described above. In brief, the strains were washed three times with PBS and then uniformly adjusted to 2×10^5 cells/ml. Cells were resuspended in DMEM and incubated at 37°C and 5% CO₂ for 24 hours. Staining with India ink, acquire images using Carl Zeiss Microscopy, and use the cell measurement software (Zen 2011) to manually measure capsule and cell diameter.²⁶

***In vivo* virulence study in mouse**

Survival test: In order to verify the difference of virulence between VGla and VGlb, representative strains were selected for mouse experiment, a VGII strain was used as the reference strain. The cells were adjusted to 2×10^6 cells/ml. Female C57BL/6 mice aged 8-10 weeks were used for survival test, and each group of VGla, VGlb, and VGII contained ten mice. After the mice were anesthetized with 10% chloral hydrate, 50 μ l (2×10^6 cells/ml) was slowly dropped into the nasal cavity of each mouse. The state of the mice was observed every day and the death was recorded.

Lung and brain fungal burden test: The process of mouse infection was the same as survival test, five mice in each testing group. After 14 days of infection, the mice were killed, the lung and brain tissues were dissected and placed in a 15 ml centrifuge tube containing 4 ml PBS, the tissues were broken up with a homogenizer. The tissue suspension was serially diluted (ten times), 100 μ l of tissue suspension were plated onto YPD agar medium, incubated at 30°C and counted after two days.

Hematoxylin and eosin (H&E) staining of lung tissues: H&E staining was performed as previously described.²⁷ Briefly, the entire lung of Cryptococcus infected mice was dissected, fixed with 4% PFA and embedded with paraffin. Maximum cross section was selected and stained. The entire sections were scanned for analysis by Panoramic midi: 3D/histech. CaseViewer, version 2.4, was utilized to visualize the stained section.

Assembly and annotation of *C gattii* genomes

Initial assemblies of four isolates with Nanopore sequencing reads were generated by Canu, version 2.0, with parameter genome size = 18,000,000.²⁸ Illumina reads were aligned to genome drafts by bwa-mem2¹⁷, which was used to polish for two rounds with Pilon, version 1.23.²⁹ Assembly contigs were discarded if they represented the mitochondrial genome, the rest of fully assembled contigs were aligned against WM276 with NUCmer, version 3.1,³⁰ to evaluate whether a chromosome-level assembly was obtained correctly and examine potential recombination events. QUILT, version 5.0.2, was also used to evaluate the assembly quality.³¹ After we identified soft-masked repeat regions by RepeatMasker, version 4.0.9.³² BRAKER2, version 2.1.4, was performed to predict genes in *C gattii*.³³ RNA-seq reads were aligned with STAR, version 2.5.3a,³⁴ which was incorporated to improve gene callings. tRNA and rRNA was

separately predicted by tRNAscan-SE, version 2.0.5³⁵ and rRNAmmer, version 1.2.³⁶ EggNog-mapper version2 was then used to predict gene functions,³⁷ which were filtered if over 30% coding sequences overlapped with noncoding regions.

Statistical analysis

We applied Kruskal-Wallis with Dunn's post-hoc test when comparing *in vitro* experiments and time from onset of symptoms to final diagnosis of VGIa, VGIb, and VGII. We used one-way ANOVA test and a Tukey's honestly significant difference (Tukey's HSD) post-hoc test in organ fungal burden test between three groups. In addition, Wilcoxon rank sum test was used to compare behaviors between two groups. All statistical analyses were done with R, version 3.6.3.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Geographical distribution and clinical characteristics of *C gattii* isolated in China

In our study, a multi-center cohort of *C gattii* from 15 provincial-level administrative regions has been built (table 1). 32 patients infected by *C gattii* with relatively complete clinical data were collected, of which 27 (84.4%) were males and five (16.6%) were females. Mean age was 44 (range 20-81) years. Based on genome sequences profiled by WGS, 20 (62.5%) strains were genotyped as VGI while seven (21.9%) strains were genotyped as VGII. 26 [96.3%] strains were isolated from the cerebrospinal fluid (CSF), only one from the lung.

To gain a more comprehensive and accurate picture of geographical distribution, we also collected *C gattii* strains information from published studies from 2006 to 2020. Totally, 119 *C gattii* appeared in at least 20 provincial-level administrative regions of China. Most of strains were concentrated in the southeastern coastal areas with subtropical climates, but a few strains had also begun to appear in the northeastern areas with temperate climates. Interestingly, only VGI strains were found in the northeastern areas with temperate climates (figure 1). The number of cases decreases as the latitude elevates. Out of 114 strains with molecular types, VGI and VGII accounted for 71.1% (81/114) and 28.9% (33/114) respectively.

Four major subtypes of *C gattii* VGI based on WGS

Given the higher prevalence of *C gattii* VGI in our study and previous reports in China³⁸ and Asia³⁹, we focused on the VGI lineage based on WGS. In addition to 20 VGI strains in our study, we retrieved all publicly available VGI strains as well as each representative VGII, VGIII, and VGIV strains (appendix p 7-8).

By aligning to VGI reference genome WM276, we constructed a phylogenetic tree, which clearly segregated 45 VGI strains into four major subgroups (figure 2a). Based on the descending order of the strain number in WGS data, we named them as VGla, VGlb, VGlc, and VGld, they were 35.6% (16/45), 26.7% (12/45), 22.2% (10/45), and 15.6% (7/45) respectively. The largest number of strains was VGla subtype. Strains of VGla mostly consisted of clinical strains from China (our study), Australia, and US. Clinical Strains of VGlb were from China and US, and environmental strains of VGlb were all from Africa. VGlc exclusively consisted of clinical strains of Southeastern states of US.⁴⁰ Most of VGld strains were environment-originated from Africa, but one clinical strain from our study. To better illustrate the relationship between VGI strains, an unrooted tree for VGI were constructed (figure 2b). Interestingly, there was no clear segregation between environmental and clinical strains as to both VGla and VGlb strains, which suggested *C gattii* VGI environmental strains could be pathogenic without addition of genetic alterations. *C gattii* VGII possessed similar phenomenon that VGII strains of environmental, veterinary, and clinical source shared the same ST genotypes.⁴¹

Global distribution of VGI subtype based on MLST signatures defined by WGS

In order to infer MLST sequences from WGS data, especially for public WGS data without MLST information, a computational pipeline was established to predict MLST subtypes based on WGS data (appendix p 2). The performance of the pipeline was validated by traditional PCR based MLST on the same VGI strains. *In silico* MLST prediction yielded 100% consistent results with standard PCR based MLST. The sequences extracted from PCR based and WGS based data were used for construction of a phylogenetic tree. The phylogenetic tree assigned most of strains into the four subtypes as we previously defined (figure 3a). Nine strains out of 308 (2.9%) did not show enough similarity to the four subtypes, thus were assigned to “others”. Several information can be used to deduce subtypes (figure 3b). By assigning MLST genotyped strains to four subtypes we defined, we found that the overall pattern was largely consistent with that based on the limited sample sized WGS dataset. VGla and VGlb, accounting for 52.9% (163/308) and 36.0% (111/308), were the most prevalent subtypes of VGI globally (figure 3c and appendix p 9). VGlc (12/308, 3.9%) was exclusively found in the US, while VGld (13/308, 4.2%) were sparsely present in Asia and Africa.

Higher virulence of VGlb supported by clinical outcomes, in vitro and in vivo experiments

As VGla and VGlb were the most prevalent subtypes both in China and all over the world, these two major subtypes were further studied to explore whether the subtypes were associated with different clinical outcomes. VGII strains were included as a reference. The patients infected by VGlb strains showed a worse overall survival compared with patients infected by VGla and VGII (log-rank test p value=0.0554, figure 4a). Noticeably, patients infected by VGlb showed a much shorter period of time between the initial appearance of symptoms and confirmative diagnosis than those by VGla (figure 4b).

Host temperature adaptation, capsular production, and melanin are the main virulence factors of species belonging to *Cryptococcus* pathogenic species complex, including *C gattii*. Our in vitro semi-quantitatively

phenotypic assessment revealed that VGIb isolates showed significant higher melanin production compared with VGIa, but similar levels compared with VGII strains, while VGII exhibited the significant higher resistance to high temperature (figure 4c). There was no clear difference in term of capsule size and growth defect in body temperature (appendix p 3). Next, we evaluated the drug resistance of the most commonly used drug for cryptococcal meningitis. VGIb strains showed significantly lower susceptibility to fluconazole and flucytosine than VGIa, and higher susceptibility to these two drugs compared with VGII strains (figure 4d). There was no significant difference among VGIa, VGIb, and VGII in terms of amphotericin B sensitivity (figure 4d).

In vitro assay partially explained the higher virulence of VGIb versus VGIa, however, VGII strains showed similar or even higher virulence compared with VGIb strains. Therefore, we performed *in vivo* assay to examine the virulence of these strains in a murine inhale model. Consistently, VGIb showed not only a significant worse survival (figure 4e) but also a significant higher CFU in both lung and brain compared with VGIa and VGII strains (figure 4f). In addition, H&E staining showed that significantly fewer tertiary lymphoid structures (as shown in the right panels of figure 4g) were formed on unit area in the lung of mice infected with VGIb strains ($0.71/\text{mm}^2$) in comparison to that of mice infected with either VGIa strains ($1.78/\text{mm}^2$) or VGII strains ($1.31/\text{mm}^2$), suggesting that stronger immune responses were induced in mice infected by VGIa and VGII strains (figure 4g). Taken together, these results suggested that high virulence of *C. gattii* VGIb might be due to its higher CFU in both lung and brain, possibly through escaping the surveillance of the immune system. These features make VGIb more pathogenic than VGIa and VGII.

Comparison of VGIa and VGIb genomes

There was no high-quality reference genome assembly available for VGIb strains, the best available reference on VGIb is RU294, with 53 scaffolds.⁹ So, additional Oxford nanopore sequencing (134-282X coverage) was performed for representative VGIa and VGIb strains, a complete and high-quality assembly of 14 chromosomes was achieved by combining illumina short reads with Oxford nanopore long reads (figure 5a, appendix p 10). The paired RNA-Seq data for representative VGIa and VGIb strains was generated to obtain high quality gene annotation (appendix p 6, appendix p 11).

We next tried to identify whether there were VGIa and VGIb specific genes. However, only one specifically deleted gene was found in VGIa, while this gene was present in other VGI subtypes (figure 5b) and was not highly expressed in the VGIb (data not shown). In addition, a focal deletion in VGIa also supported that a VGIa lineage specific deletion might lead to the loss of this gene, which contained a tetratricopeptide repeat (TPR) domain, its function was still unclear. Therefore, the deletion of this gene in VGIa was unlikely to explain its reduced virulence.

Given the aneuploidy was known to associated with clinical outcome,⁴² we first examined whether there was clear aneuploidy in VGIb strains. However, none of VGIb strains showed aneuploidy (appendix p 4). We also computationally inferred the ploidy of each strain, and still no clear differences were observed between VGIa and VGIb on the ploidies (appendix p 6).

Then we tried to identify VGla and VGlb specific SNPs or Indels and annotated these sites based on the nucleotides observed in VGlc and VGld strains (figure 5c). SNPs of VGlb were more consistent with other VGI strains. However, two SNPs that differed between VGla and VGlb were found on the known virulence genes. One of them was located on Phospholipase C-1 (PLC1), but this SNP was not confirmed on the RNA-Seq data (data not shown). The other one was located on manganese-dependent SOD (SOD2) gene, which was previously considered to be related to virulence in *C gattii*.⁴³ This mutation caused an amino acid change from glutamic acid (Glu) to lysine (Lys) in VGla (figure 5d), and was supported by RNA-Seq data as well (appendix p 5). Given its role in oxygen metabolism, we suspected that VGla and VGlb exhibited differential levels in handling oxidative stress. Indeed, VGlb strains showed a significant growth advantage compared with VGla strains with oxidative stress introduced by diamide (figure 5e).

Discussion

C gattii VGII is the most prevalent molecular type in America, while VGI is mostly prevalent in Asia, Europe, and Australia.¹¹ There was no systematic review of *C gattii* with different molecular types in China. In this study, we carried out by far the most comprehensive study of *C gattii* in China and found the overall classification of *C gattii* in China was consistent with that in Asia.

Based on WGS of all available VGI strains around the world, we categorized the VGI into four major subtypes (VGla, VGlb, VGlc, and VGld). By integrative analysis of WGS and MLST, we obtained statistics for the prevalence of each subtype of VGI and results that VGla and VGlb were the major clinical VGI strains, both of which were present in all continents except Antarctica. While VGlc and VGld were less often observed with a clear geographical preference. Our integrative analysis of WGS and MLST also identified SNP sites that could distinguish VGI subtypes, which will facilitate the VGI subtype classification on the clinical setting using commonly available tool of MLST genotyping based on seven genes.

To understand potential clinical significance of the classification, we focused on the commonly observed VGla and VGlb subtypes. Among them, VGlb subtype showed a significant worse clinical outcome compared with VGla and VGII strains. It turned out the VGI strain exhibited higher virulence than VGII strains⁴⁴ belonged to the VGlb subtype. Due to the small sample size, such results needed to be validated in a larger cohort in the future. Consistently, the results of *in vitro* assays showed higher virulence, higher resistance to antifungal drug of VGlb subtype compared with the VGla subtype, weaker immune response.

With the high-quality reference, we were able to eliminate the possibility of extensive genome rearrangement within VGI subtypes, which was observed between genomes of *C gattii* VGI and VGII.⁸ Aneuploidy was found in resistant *C gattii* VGI and VGII strains.⁸ No aneuploidy was found in neither VGla nor VGlb treatment naïve clinical strains of this study. We suspected that *C gattii* aneuploidy might occur under the selection pressure of antifungal antibiotics. Finally, we identified the subtype specific SNPs and one such SNP on SOD2 (related to oxidation-reduction process) might contribute to differences in

virulence. SOD2 protects the fungi from superoxide anion in the mitochondrial, thus allowing them to survive Reactive Oxidative Species (ROS) released by host innate immune cells. SOD2 SNP on VG1a strain is likely to be neutral in the fitness during the growth in the environment without ROS stress, and thus is not deselected in the evolution. However, additional functional studies were needed to validate these predictions.

To summarize, we believed our classification system, as well as the high-quality data generated in this study, will facilitate future studies on *C gattii*. It will be interesting to characterize the differences between clinical features and molecular subtypes of *C gattii* VG1 strains in larger cohorts as well as in the other countries. In the future, subtype specific drug screening and treatment option might be applied.

Declarations

Ethical Approval and Consent to participate

The study was approved by the Medical Ethics Committee of Beijing Shijitan Hospital: sjtky11-1x-2020 (20) and informed consent forms were obtained from all patients prior to their participation in this study. All methods were carried out in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable.

Availability of data and materials

All the data generated and analyzed during this study are included in this published article. The raw data of *C gattii* isolates produced by Illumina sequencing and Oxford Nanopore have been deposited on Sequence Read Archive under the BioProject accession number PRJNA721774 and PRJNA721775.

Competing interests

We declare no competing interests.

Funding

This work was supported by National Natural Science Fund Youth Project (81700007); Beijing Natural Science Foundation (2019A10); Excellent talents in Beijing “Youth top individual” (2019YXBJ1); Capital’s Funds for Health Improvement and Research (2020-2Z-2086); “Sailing” plan for clinical technology innovation project in Beijing (XMLX202115); National Key Research and Development Program of China Research on the Precision Diagnosis, Treatment, and Integrated Prevention, Control for the elderly with common infectious disease (2020YFC2005404); National Natural Science Foundation of China (31871332, 81772798).

Author contributions

XX, LZ, LW, and DS designed, initiated, and coordinated the study. CW, JG, MX, YL, LW, SZ, PL, QZ, XM, HD, YH, JS, HD, SG, SG, and MZ contributed to data collection. XZ, WK, and XW performed the basic experiments. WT, MX, CG, and YS analyzed the data with the help from SL, LZ, XX, WT, CW, DS, LW, and XZ wrote and revised the paper. All authors read, commented on, and approved the final manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Acknowledgments

The authors wish to thank Qiqi Wang and Jie Wang for the help of this manuscript. This work was supported by HPC platform of ShanghaiTech University.

References

1. Wanderley de Melo S, Jr., Turner KR. Diagnosis of *Cryptococcus gattii* Infection Based on Endoscopic Ultrasound and Fine Needle Aspiration of a Lung Mass. *Clin Gastroenterol Hepatol* 2020; **18**(2): A25.
2. Huynh J, Saddi V, Cooper P, et al. Unusual Presentation of Severe Endobronchial Obstruction Caused by *Cryptococcus gattii* in a Child. *J Pediatric Infect Dis Soc* 2020; **9**(1): 67-70.
3. Franco-Paredes C, Womack T, Bohlmeyer T, et al. Management of *Cryptococcus gattii* meningoencephalitis. *Lancet Infect Dis* 2015; **15**(3): 348-55.
4. Acheson ES, Galanis E, Bartlett K, Mak S, Klinkenberg B. Searching for clues for eighteen years: Deciphering the ecological determinants of *Cryptococcus gattii* on Vancouver Island, British Columbia. *Med Mycol* 2018; **56**(2): 129-44.
5. Harris J, Lockhart S, Chiller T. *Cryptococcus gattii*: where do we go from here? *Medical Mycology* 2012; **50**(2): 113-29.
6. Farrer RA, Chang M, Davis MJ, et al. A New Lineage of *Cryptococcus gattii* (VGV) Discovered in the Central Zambezi Miombo Woodlands. *mBio* 2019; **10**(6).
7. Billmyre RB, Croll D, Li W, et al. Highly recombinant VGII *Cryptococcus gattii* population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. *mBio* 2014; **5**(4): e01494-14.
8. D'Souza CA, Kronstad JW, Taylor G, et al. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *mBio* 2011; **2**(1): e00342-10.
9. Farrer RA, Desjardins CA, Sakthikumar S, et al. Genome Evolution and Innovation across the Four Major Lineages of *Cryptococcus gattii*. *mBio* 2015; **6**(5): e00868-15.
10. Ngamskulrungraj P, Serena C, Gilgado F, Malik R, Meyer W. Global VGIIa isolates are of comparable virulence to the major fatal *Cryptococcus gattii* Vancouver Island outbreak genotype. *Clin Microbiol Infect* 2011; **17**(2): 251-8.
11. Cogliati M. Global Molecular Epidemiology of *Cryptococcus neoformans* and *Cryptococcus gattii*: An Atlas of the Molecular Types. *Scientifica (Cairo)* 2013; **2013**: 675213.

12. Okubo Y, Tochigi N, Wakayama M, et al. How Histopathology Can Contribute to an Understanding of Defense Mechanisms against Cryptococci. *Mediators of Inflammation* 2013; **2013**: 1-11.
13. Meyer W, Aanensen DM, Boekhout T, et al. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Med Mycol* 2009; **47**(6): 561-70.
14. Pitkin JW, Panaccione DG, Walton JD. A putative cyclic peptide efflux pump encoded by the TOXA gene of the plant-pathogenic fungus *Cochliobolus carbonum*. *Microbiology (Reading)* 1996; **142 (Pt 6)**: 1557-65.
15. Edwards R. fastq-dump. 2015 <https://edwards.sdsu.edu/research/fastq-dump>.
16. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics (Oxford, England)* 2018; **34**(17): i884-i90.
17. Li H, Durbin RJB. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics (Oxford, England)* 2009; **25**(14): 1754-60.
18. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010; **20**(9): 1297-303.
19. Cingolani P, Platts A, Wang le L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012; **6**(2): 80-92.
20. Riaño-Pachón DM. ploidyNGS: Visually exploring ploidy with Next Generation Sequencing data. 2017. <https://github.com/diriano/ploidyNGS>.
21. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* 2016; **12**(4): e1004873.
22. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)* 2014; **30**(9): 1312-3.
23. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* 2019; **47**(W1): W256-W9.
24. Jung KW, Yang DH, Maeng S, et al. Systematic functional profiling of transcription factor networks in *Cryptococcus neoformans*. *Nat Commun* 2015; **6**: 6757.
25. Lee D, Jang EH, Lee M, et al. Unraveling Melanin Biosynthesis and Signaling Networks in *Cryptococcus neoformans*. *mBio* 2019; **10**(5).
26. Guess T, Lai H, Smith SE, et al. Size Matters: Measurement of Capsule Diameter in *Cryptococcus neoformans*. *J Vis Exp* 2018; (132).
27. Chen X, Cao X, Sun X, et al. Bcl-3 regulates TGFbeta signaling by stabilizing Smad3 during breast cancer pulmonary metastasis. *Cell Death Dis* 2016; **7**(12): e2508.
28. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k/mer weighting and repeat separation. *bioRxiv* 2017: 071282.

29. Walker BJ, Abeel T, Shea T, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* 2014; **9**(11).
30. Kurtz S, Phillippy A, Delcher AL, et al. Versatile and open software for comparing large genomes. *Genome Biol* 2004; **5**(2): R12-R.
31. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)* 2013; **29**(8): 1072-5.
32. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* 2009; **Chapter 4**: Unit 4.10.
33. Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. Whole-Genome Annotation with BRAKER. *Methods Mol Biol* 2019; **1962**: 65-95.
34. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* 2013; **29**(1): 15-21.
35. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research* 1997; **25**(5): 955-64.
36. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**(9): 3100-8.
37. Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol Biol Evol* 2017; **34**(8): 2115-22.
38. Chen M, Xu Y, Hong N, et al. Epidemiology of fungal infections in China. *Front Med* 2018; **12**(1): 58-75.
39. Chen SC, Meyer W, Sorrell TC. Cryptococcus gattii infections. *Clin Microbiol Rev* 2014; **27**(4): 980-1024.
40. Lockhart SR, Roe CC, Engelthaler DM. Whole-Genome Analysis of Cryptococcus gattii, Southeastern United States. *Emerg Infect Dis* 2016; **22**(6): 1098-101.
41. Engelthaler DM, Hicks ND, Gillece JD, et al. Cryptococcus gattii in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. *mBio* 2014; **5**(4): e01464-14.
42. Stone NR, Rhodes J, Fisher MC, et al. Dynamic ploidy changes drive fluconazole resistance in human cryptococcal meningitis. *J Clin Invest* 2019; **129**(3): 999-1014.
43. Narasipura SD, Chaturvedi V, Chaturvedi S. Characterization of Cryptococcus neoformans variety gattii SOD2 reveals distinct roles of the two superoxide dismutases in fungal biology and virulence. *Mol Microbiol* 2005; **55**(6): 1782-800.
44. Xue X, Wu H, Wang K, Cao J, Shen D. Cryptococcosis by Cryptococcus gattii in China. *Lancet Infect Dis* 2015; **15**(10): 1135-6.

Figures

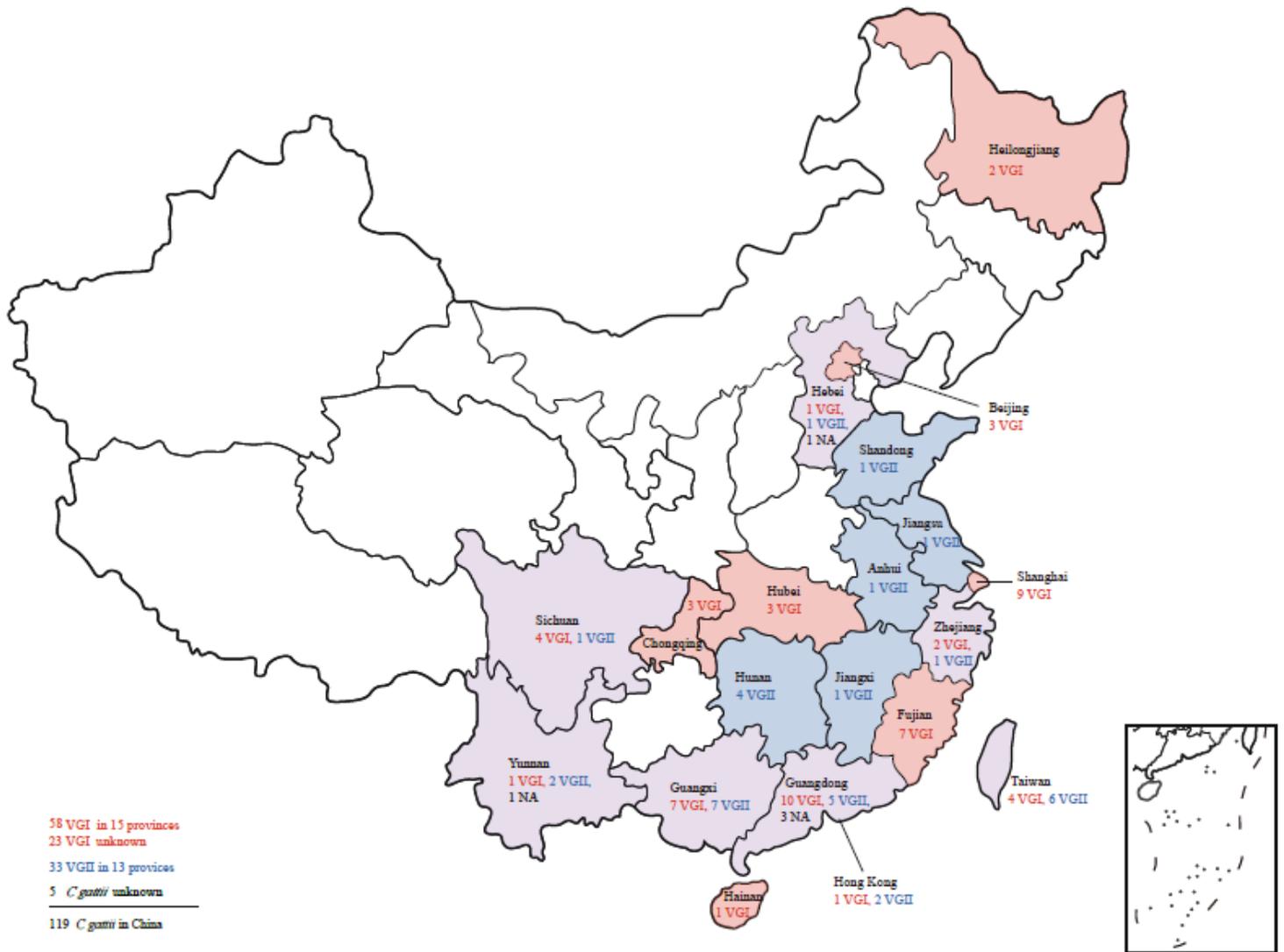


Figure 1

Geographical distribution of *C. gattii* strains isolated in China. Red, blue, and purple represented provincial-level administrative region with VGI, VGII, and both VGI and VGII respectively.

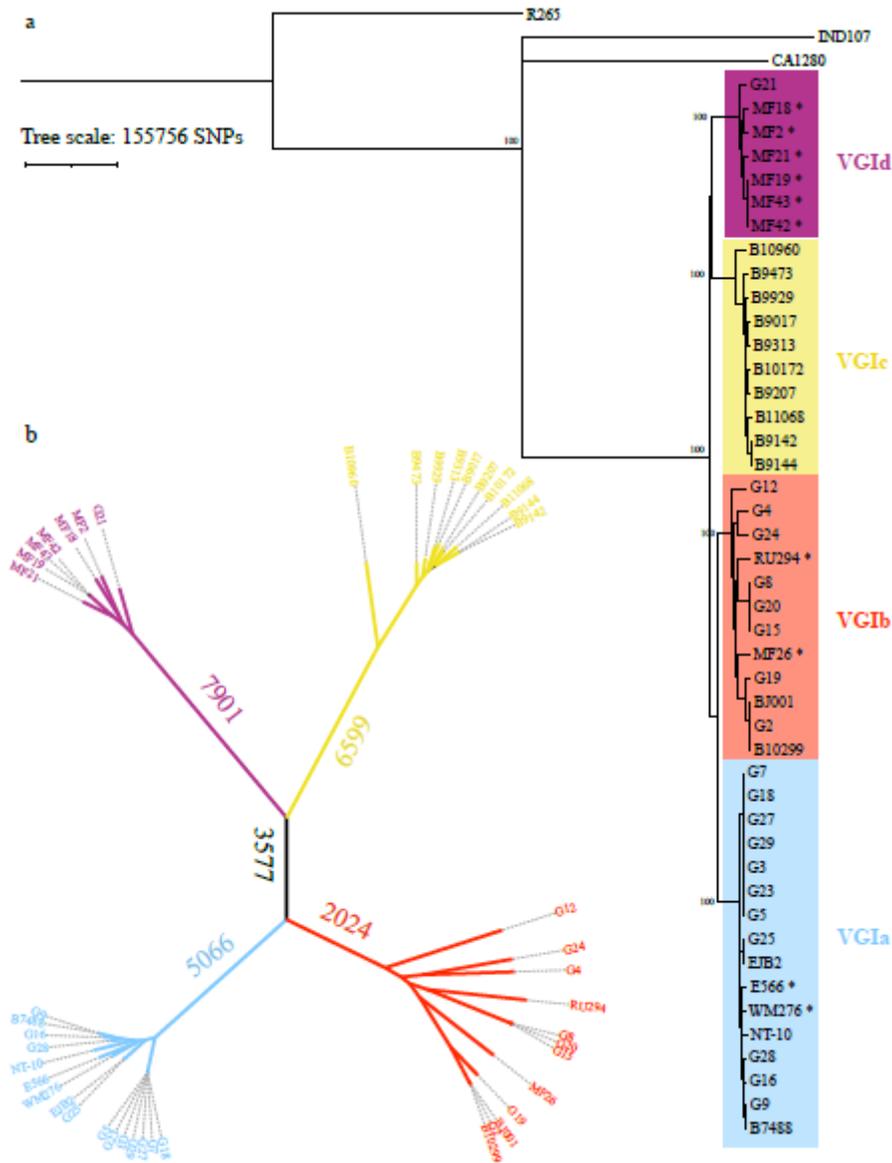


Figure 2

Whole genome sequencing phylogeny tree of *C. gattii* VGI genomes. Previously published genome sequences were included. Environmental strains were indicated with an asterisk. (A) Branches that determine subgroups were labeled with bootstrap values, and four different colors were separately assigned to VGIa, VGIb, VGIc, and VGI d. (B) An unrooted phylogeny of VGI strains is labeled with unique and shared SNPs.

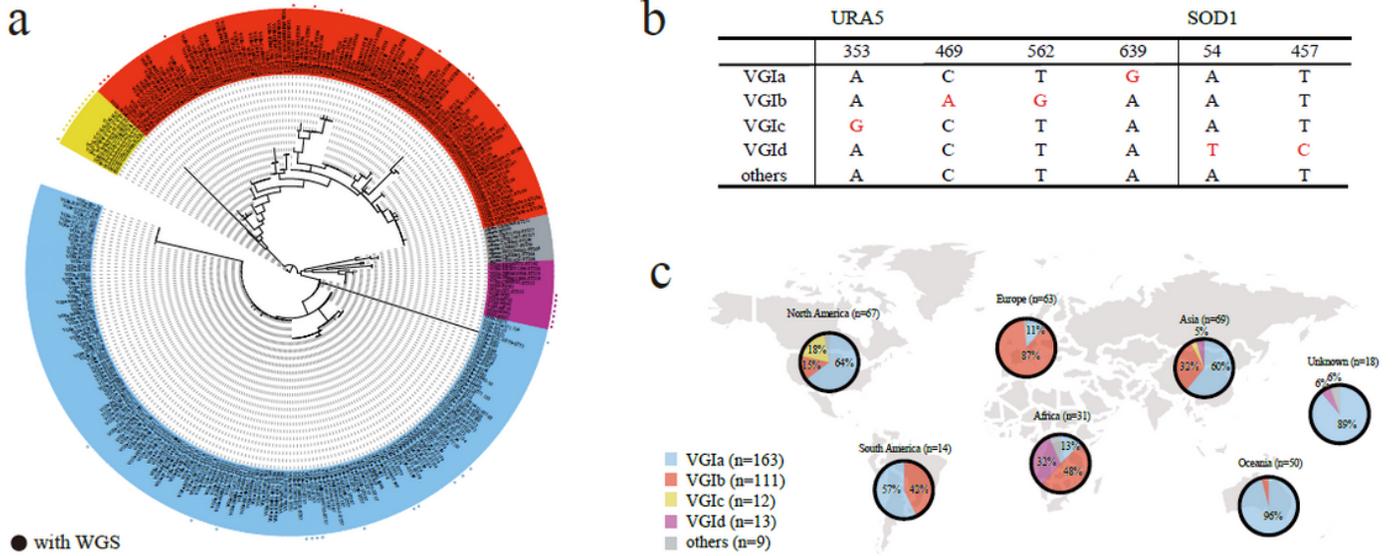


Figure 3

Global distribution of VGI subgroups based on published MLST and in silico MLST. (A) 311 samples were included for phylogeny tree. Samples with WGS are labeled with dot, which were used as anchors for annotation. (B) SNP signatures from two MLST genes associated with VGI subgroups. (C) The world distribution of VGIa, VGIb, VGIc, and VGId based on MLST.

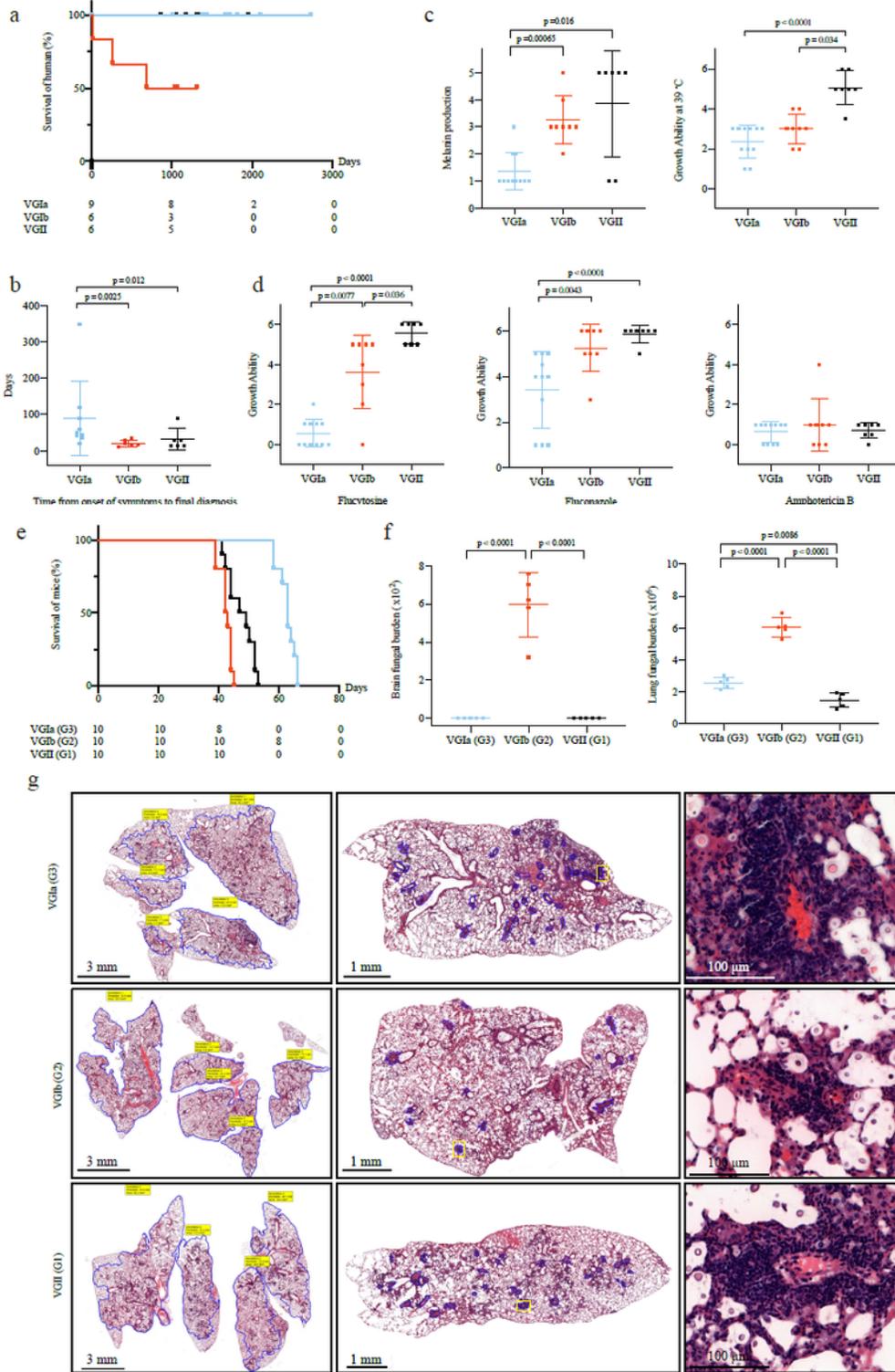


Figure 4

Different clinical outcomes and virulence between VGla and VGlb (A) Kaplan-Meier survival curve highlights differences in survival of patients infected by VGla, VGlb, and VGII. (B) Time of reaching a confirmative diagnosis in patients infected with VGla, VGlb, and VGII. (C) Plate-based Melanin production and heat resistance to 39oC. (Mean with SD) (D) Drug resistance to flucytosine, fluconazole and Amphotericin B. (Mean with SD) (E) Kaplan-Meier survival curve of mice (log-rank test $p < 0.001$).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix.pdf](#)