

# *Cryptococcus gattii* in the Age of Whole-Genome Sequencing

Wieland Meyer<sup>a,b</sup>

Molecular Mycology Research Laboratory, Centre for Infectious Diseases and Microbiology, Sydney Medical School-Westmead Hospital, Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Westmead Millennium Institute, Sydney, Australia<sup>a</sup>; Mycology Laboratory, National Institute of Infectious Diseases Evandro Chagas, FIOCRUZ, Rio de Janeiro, Brazil<sup>b</sup>

**ABSTRACT** *Cryptococcus gattii*, the sister species of *Cryptococcus neoformans*, is an emerging pathogen which gained importance in connection with the ongoing cryptococcosis outbreak on Vancouver Island. Many molecular studies have divided this species into four major lineages: VGI, VGII, VGIII, and VGIV. This commentary summarizes the whole-genome sequencing (WGS) studies that have been carried out with this species, re-emphasizing the phylogenetic relationships, showing chromosomal rearrangements between those four groups, and identifying VGII as ancestral population within *C. gattii*. In addition, WGS specific to VGII, containing the Vancouver Island outbreak genotypes and those from the Pacific Northwest region of the United States, has placed the origin of this lineage within South America and identified specific genes responsible for either brain or lung infection. It also showed, that many genotypes are spread across a number of different continents, as has been previously shown by multi-locus sequence typing (MLST). In addition, it showed that recombination occurs more frequently between mitochondrial than nuclear genomes.

*Cryptococcus gattii* is an emerging human pathogen that causes infections mainly in immunocompetent patients, in contrast to its sister species, *Cryptococcus neoformans*, which causes most of the cryptococcosis cases in immunocompromised individuals worldwide (1). *C. gattii* gained importance in connection with the outbreak, ongoing since 1999, on Vancouver Island, British Columbia, Canada, and its subsequent spread to the Pacific Northwest of the United States (1). This resulted in a large global effort to identify the origin of the outbreak-related strains and to search for specific genetic determinants of the different phenotypes identified. Molecular epidemiology using a large set of different genotyping techniques, ranging from simple restriction fragment length polymorphism (RFLP) analysis to DNA and PCR fingerprinting, amplified fragment length polymorphism (AFLP), multilocus microsatellite typing (MLMT), multilocus sequence typing (MLST), and whole-genome sequence typing (WGST), has recognized four major lineages among *C. gattii* isolates (VGI, VGII, VGIII, and VGIV) (1, 2). These four lineages are so genetically, epidemiologically, and geographically different that they most likely constitute different species, as recently proposed (3).

Open questions concern (i) the evolutionary origin of the species, (ii) its global distribution, (iii) the emergence of virulence traits, and (iv) how they are manifested genetically. With the advance of whole-genome sequencing (WGS), great progress has been made to start to answer some of those questions.

The Pacific Northwest outbreak caused by three distinct nearly clonal sublineages of VGII—VGIIa (minor), VGIIb (major), and VGIIc (novel)—was surprising, as *C. gattii* was until that point thought to be found mainly in tropical and subtropical regions. Besides the new ecological niche, the dominant emerging population was characterized by an increased virulence, causing mainly pulmonary infections, which differed from the mostly neurological diseases until then associated with *C. gattii* (1). MLST analysis originally pointed to Australian origins, based on the similarity of MLST data obtained for VGIIb (4). The first comparative whole-genome analysis between a single VGI isolate (WM276), representing the molecular type causing the majority of the global *C. gattii* infections, and a VGII isolate (R265), representing the highly virulent VGIIa Vancouver Island outbreak genotype, re-

vealed numerous differences in the chromosome copy numbers and genomic rearrangements and a large number of sequence polymorphisms specific to either VGI or VGIIa (5). In a second study, WGS resulted in the reconstruction of the phylogenetic relationships among the different outbreak subgenotypes, with a greater resolution among the strains than was obtained with the previous MLST analysis. While MLST had not revealed single-nucleotide polymorphisms (SNPs) within either of the VGIIa and VGIIb isolates and revealed only 1 SNP among the VGIIc isolates, WGST recognized 1,512 SNPs within VGIIa, 132 SNPs in VGIIb, and 137 SNPs in VGIIc isolates (6).

After these initial reports, two subsequent studies applied WGS. The first one investigated 118 VGII genomes (7), and the second one investigated 53 VGII genomes (8). The first study showed clearly that all three nearly clonal Pacific Northwest outbreak lineages likely originated in South America. The phylogenetic SNP analysis of those 118 genomes indicated multiple dispersal events from South America to North America and other parts of the world (7). Among the Pacific Northwest outbreak lineages, 717, 246, and 528 SNPs were identified within VGIIa, VGIIb, and VGIIc, respectively. VGIIa isolates were found in North and South America; VGIIb isolates were more broadly dispersed, being present in North and South America, Australia, Asia, and Europe. VGIIc isolates were found only in the Pacific Northwest region of the United States, but they were most closely related to South American isolates. Gene content comparisons between the Pacific Northwest outbreak lineages and also with the global VGII population revealed genes potentially related to habitat adaptation, virulence, and pulmonary versus neurological disease presentation. Different gene content, which was either specific to the Pacific Northwest outbreak lineages or shared between

Published 17 November 2015

Citation: Meyer W. 2015. *Cryptococcus gattii* in the age of whole-genome sequencing. mBio 6(6):e01761-15. doi:10.1128/mBio.01761-15.

Copyright © 2015 Meyer. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported license, which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to wieland.meyer@sydney.edu.au.

one of those lineages and the global population, was found, suggesting recombination. In addition, a gene introgression from *C. neoformans* var. *grubii* was found in the Pacific Northwest populations which was seldom found in the global VGII population (7). The second study, which also used whole-genome SNP phylogeny, confirmed the clonal expansion of the Pacific Northwest outbreak lineages and placed the origin of VGIIa most likely in South America. However, the origin of VGIIb was still placed in Australia, and VGIIc was found only in the Pacific Northwest region of the United States (8), which reflects the small sample size analyzed. The authors suggest that the highly virulent VGIIa population arose from a less virulent lineage which had a mutation in the *MSH2* ortholog but then reverted in the VGIIa strains, implementing a transient mutator phenotype, which may have contributed to the adaptation and increased virulence in the Pacific Northwest outbreak strains. They also found gene content shared between the Pacific Northwest outbreak lineages or the global VGII population, indicating sexual macroevolution and asexual microevolution, with novel virulent genotypes resulting either from bisexual or unisexual recombination. In addition, an introgression of 3 genes in the basal VGII population was shown, which most likely originated from the VGI lineage (8).

Besides the emergence of the VGII lineage in the Pacific Northwest region of the United States, isolates of a second major *C. gattii* lineage, VGIII, have been increasingly causing infections especially in HIV/AIDS patients in southern California (9). Using mainly MLST analysis of 19 environmental isolates and 11 clinical samples as well as WGS of a small subset of those isolates (9 strains), 3 environmental populations were established, and those populations were demonstrated to be the most likely environmental reservoirs for recent clinical infections (9).

The recent study by Farrer et al. (10) is the first one to use 15 *de novo* assembled and annotated genomes representing clinical and environmental isolates of all four major *C. gattii* lineages to investigate genome variation among these lineages. In addition, it reannotated the genome of the representative VGIIa major Vancouver Island outbreak strain (R265), and in particular, it resolved 124 kb of previously ambiguous sites, replaced 4,166 single bases, and introduced 2,382 insertions and deletions (9). Using 5,319 single-copy orthologous protein-coding genes, it reconstructed the phylogenetic relationships between and among the four lineages, reconfirming these four major lineages as distinct groups, as previously established by MLST, and further indicating that they deserve species recognition.

The authors also identified the VGII lineage as the ancestral population within *C. gattii*, diverging before VGIV, followed by VGIII and VGI, which are the most closely related lineages (10), as had previously been established by multigene analysis based on the *ACT1*, *IDE*, *PLB1*, and *URA5* genes (2). Within *C. gattii*, the chromosome structure was highly conserved, with the most conservation in the VGII lineage. Syntenic variation was mainly confined within the three closely related lineages VGI, VGIII, and VGIV (10), with 2.6% of each of the 16 genomes being rearranged with respect to each other. The majority of the large chromosomal rearrangements were located close to the centromeres, indicating that they are primarily whole-chromosome arm rearrangements, which may impact the capability of interlineage genetic exchange. In addition, isolates of the VGII and VGIII lineages contained *MATa* locus-based aneuploidies, which have been suggested to influence the virulence potential of the sister species *C. neoformans*.

Upon investigation of the gene content, all key virulence genes, including 32 of the 35 genes involved in cryptococcal capsule formation, were also found as single-copy genes, indicating that the polysaccharide capsule synthesis and virulence genes are conserved between the two sibling species (10).

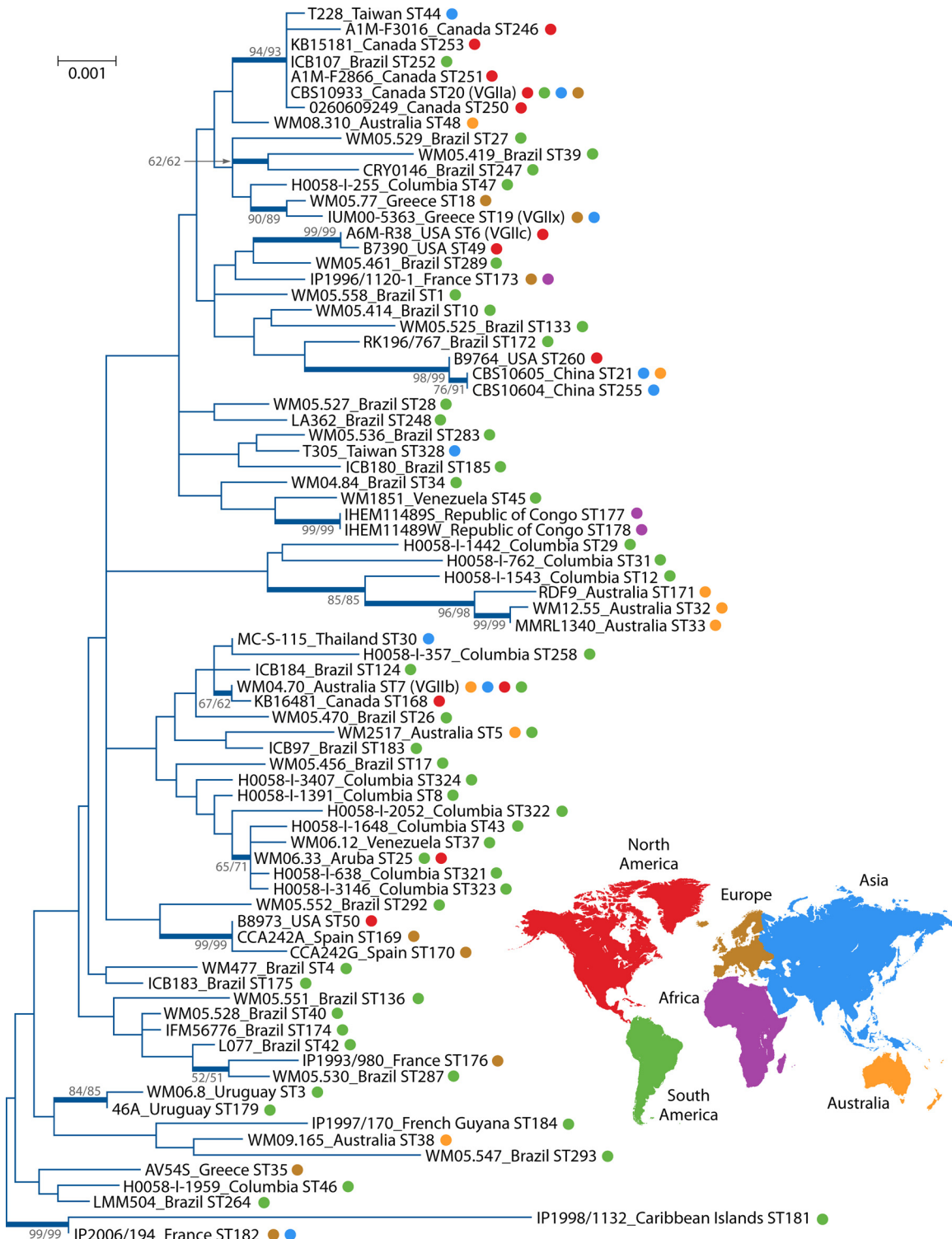
Farrer et al. suggest that the VGII lineage has lost 146 genes compared to the VGI, VGIII, and VGIV lineages (10). This could be one scenario, if the genome was compared to a common ancestral population of all four lineages. Considering that VGII was identified as the ancestral population within *C. gattii* (2, 10), it is also possible that the VGI, VGIII, and VGIV lineages have gained those genes. The VGII lineage carries an expanded array of secretory carrier membrane proteins, unique genes enriched for HSP70, COX6B, and iron-binding domains, which all could contribute to its greater virulence. In addition, it is missing crucial genes for nuclear and mitochondrial maintenance, which could explain the different tubular mitochondrial morphology. The VGIII lineage was characterized by 79 lineage-specific genes, and VGIV has the greatest number (170) of lineage-specific genes. The loss of enolases and Ctr copper transporter family domains in the VGIII and VGIV lineages may be responsible for the lower infection rates in immunocompetent patients. The most widely distributed lineage, VGI, has the fewest gene losses, which may explain its ability to survive in such a broad range of habitats (10).

Little genetic exchange was found between the four *C. gattii* lineages, suggesting that they have remained mainly isolated since their divergence, despite overlapping geographic distribution and shared ecological niches (10).

WGS analysis has also identified a new transcontinental spread of an additional genotype, named VGIIx, represented by isolates from Greece (CBS10090) and Brazil (LA55), within the VGII lineages (10). This is not an unusual finding, as MLST had already identified, within 529 global isolates of the VGII lineage, 8 out of 78 sequence types with an intercontinental spread (Fig. 1). This includes the genotype VGIIx (10), newly described by Farrer et al., equivalent to ST19 (11). Similar findings have been observed within the VGIII population (C. Firacative and W. Meyer, unpublished data) and should be expected if further isolates within the VGI and VGIV lineages are investigated. These analyses, either by MLST or WGS, identified intercontinental spreads, which indicates that *C. gattii* is actively spreading and is able to adapt continuously to new environmental niches through ongoing genetic recombination via sexual macroevolution and asexual microevolution.

An unexpected finding of the whole-genome analysis by Farrer et al. (10) is that recombination occurs more commonly between mitochondrial than nuclear genomes in *C. gattii*. This once again emphasizes the increasing role of mitochondria in disease progression and outcome (12). The authors have identified a number of nucleus-specific genes that respond to oxidative stress, import into the mitochondrial inner membrane, and mitochondrial maintenance and morphology, suggesting that differences in those genes may partially explain differences in the mitochondrial phenotypes (10).

Overall, whole-genome analysis, and specifically the study by Farrer et al. (10), has identified highly valuable genome structure differences between the different *C. gattii* lineages. These will form the basis for substantial further studies into differences in virulence and disease phenotype, progression, and outcome.



**FIG 1** Phylogenetic analysis of *Cryptococcus gattii* VGII isolates inferred by maximum-likelihood (ML) and neighbor-joining (NJ) methods, based on the Kimura 2-parameter model plus discrete gamma distribution (5 categories [parameter = 0.0500]) and invariable sites (0.2512% sites), using the concatenated data set of the seven loci from the International Society for Human and Animal Mycology (ISHAM) consensus MLST scheme. A total of 79 *C. gattii* VGII strains representing each previously published sequence type (<http://mlst.mycologylab.org>) were included in the analysis. The ML tree with the highest log likelihood (−8,607.0150) is shown and drawn to scale, with branch lengths measuring the number of substitutions per site. There were a total of 4,172 positions in the final data set. Numbers at each branch indicate bootstrap values of >50% based on 1,000 replicates by each of the two (ML and NJ) algorithms, which presented similar topologies. The taxa nomenclature includes the sequence type number (ST), country of isolation, and sublineages VGIIa, VGIIb, VGIIc, and VGIIx (10) (=ST19 [11]).

## ACKNOWLEDGMENTS

I am thankful to E. Castañeda, K. Ferreira-Paim, C. Firacative, and L. Trilles for critical reading of the manuscript.

This work was supported by NH&MRC grant no. APP1031943 from Australia, NIH grant no. R21AI098059 from the United States, and CAPES Science without Frontiers grant no. 098/2012 from Brazil.

## REFERENCES

1. Chen SC-, Meyer W, Sorrell TC. 2014. *Cryptococcus gattii* infections. Clin Microbiol Rev 27:980–1024. <http://dx.doi.org/10.1128/CMR.00126-13>.
2. Ngamskulrunroj P, Gilgado F, Faganello J, Litvintseva AP, Leal AL, Tsui KM, Mitchell TG, Vainstein MH, Meyer W. 2009. Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. PLoS One 4:e5862. <http://dx.doi.org/10.1371/journal.pone.0005862>.
3. Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, Falk R, Parmen S, Lumsch HT, Boekhout T. 2015. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. Fungal Genet Biol 78:16–48. <http://dx.doi.org/10.1016/j.fgb.2015.02.009>.
4. Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, Diezmann S, Allen A, Stajich JE, Dietrich FS, Perfect JR, Heitman J. 2005. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature 437:1360–1364. <http://dx.doi.org/10.1038/nature04220>.
5. Kronstad JW, D'Souza CA, Taylor G, Warren R, Yuen M, Hu G, Jung WH, Sham A, Kidd SE, Tangen K, Lee N, Zellmaker T, Sawkins J, McVicker G, Shah S, Gnerre S, Griggs A, Zeng Q, Bartlett K, Li W, Wang X, Heitman J, Stajich JE, Fraser J, Meyer W, Carter D, Schein J, Krzywinski M, Kwon-Chung KJ, Varma A, Wang J, Brunham R, Fyfe M, Ouellette BFF, Siddiqui A, Marra M, Jones S, Holt R, Birren BW, Galagan JE, Cuomo CA. 2011. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. mBio 2:e00342-10. <http://dx.doi.org/10.1128/mBio.00342-10>
6. Gillice JD, Schupp JM, Balajee SA, Harris J, Pearson T, Yan Y, Keim P, DeBess E, Marsden-Haug N, Wohrle R, Engelthaler DM, Lockhart SR. 2011. Whole genome sequence analysis of *Cryptococcus gattii* from the Pacific Northwest reveals unexpected diversity. PLoS One 6:e28550. <http://dx.doi.org/10.1371/journal.pone.0028550>.
7. Engelthaler DM, Hicks ND, Gillice JD, Roe CC, Schupp JM, Driebe EM, Gilgado F, Carriconde F, Trilles L, Firacative C, Nagamskulrunroj P, Castañeda E, Lazera MDS, Melhem MSC, Pérez-Bercoff A, Huttley G, Sorrell TC, Voelz K, May RC, Fisher MC, Thompson GR, Lockhart SR, Keim P, Meyer W. 2014. *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. mBio 5:e01464-14. <http://dx.doi.org/10.1128/mBio.01464-14>.
8. Billmyre RB, Croll D, Li W, Mieczkowski P, Carter DA, Cuomo CA, Kronstad JW, Heitman J. 2014. Highly recombinant VGII *Cryptococcus gattii* population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. mBio 5:e01494-14. <http://dx.doi.org/10.1128/mBio.01494-14>.
9. Springer DJ, Billmyre RB, Filler EE, Voelz K, Pursall R, Mieczkowski PA, Larsen RA, Dietrich FS, May RC, Filler SG, Heitman J. 2014. *Cryptococcus gattii* VGIII isolates causing infections in HIV/AIDS patients in Southern California: identification of the local environmental source as arboreal. PLoS Pathog 10:e1004285. <http://dx.doi.org/10.1371/journal.ppat.1004285>.
10. Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, Chen Y, Voelz K, Heitman J, May RC, Fisher MC, Cuomo CA. 2015. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio 6:e00868-15. <http://dx.doi.org/10.1128/mBio.00868-15>.
11. Firacative C, Ferreira-Paim K, Trilles L, Engelthaler DM, Meyer W. 2015. Australia in the global picture of the molecular epidemiology of *Cryptococcus gattii* molecular type VGII. Microbiol Aust 36:67–70. <http://dx.doi.org/10.1071/MA15023>.
12. Ma H, May RC. 2010. Mitochondria and the regulation of hypervirulence in the fatal fungal outbreak on Vancouver Island. Virulence 1:197–201. <http://dx.doi.org/10.4161/viru.1.3.11053>.

The views expressed in this Commentary do not necessarily reflect the views of this journal or of ASM.