RESEARCH ARTICLE

Cryptococcus gattii in North American Pacific Northwest: Whole-Population Genome Analysis Provides Insights into Species Evolution and Dispersal

David M. Engelthaler,[a] Nathan D. Hicks,[a] John D. Gillece,[a] Chandler C. Roe,[a] James M. Schupp,[a] Elizabeth M. Driebe,[a] Felix Gilgado,[b] Fabian Carriconde,[b,c] Luciana Trilles,[b,d] Carolina Firacative,[b,e] Popchai Ngamskulrungroj,[b,f] Elizabeth Castañeda,[e] Marcia dos Santos Lazera,[d] Marcia S. C. Melhem,[g] Åsa Pérez-Bercoff,[b,h] Gavin Huttley,[h] Tania C. Sorrell,[b] Kerstin Voelz,[i,j] Robin C. May,[i,j] Matthew C. Fisher,[k] George R. Thompson III,[l] Shawn R. Lockhart,[m] Paul Keim,[a,n] Wieland Meyer[b]

Translational Genomics Research Institute, Flagstaff, Arizona, USA[a]; Molecular Mycology Research Laboratory, Center 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[b]; Institut Agronomique Néo-Calédonien (IAC), Noumea, New Caledonia[c]; Laboratório de Micologia, Instituto de Pesquisa Clínica Evandro Chagas, FIOCRUZ, Rio de Janeiro, Brazil[d]; Microbiology Group, Instituto Nacional de Salud, Bogotá, Colombia[e]; Department of Microbiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand[f]; Public Health Reference Laboratory, Institute Adolfo Lutz, São Paulo, SP, Brazil[g]; John Curtin School of Medical Research, Australian National University, Canberra, Australia[h]; Institute of Microbiology and Infection and School of Biosciences, University of Birmingham, Birmingham, United Kingdom[i]; National Institute for Health Research Surgical Reconstruction and Microbiology Research Centre, Queen Elizabeth Hospital, Birmingham, Birmingham, United Kingdom[j]; Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom[k]; University of California, Davis, Davis, California, USA[l]; Centers for Disease Control and Prevention, Atlanta, Georgia, USA[m]; Microbial Genetics and Genomics Center, Northern Arizona University, Flagstaff, Arizona, USA[n]

D.M.E. and N.D.H. contributed equally to this article.

ABSTRACT The emergence of distinct populations of *Cryptococcus gattii* in the temperate North American Pacific Northwest (PNW) was surprising, as this species was previously thought to be confined to tropical and semitropical regions. Beyond a new habitat niche, the dominant emergent population displayed increased virulence and caused primary pulmonary disease, as opposed to the predominantly neurologic disease seen previously elsewhere. Whole-genome sequencing was performed on 118 *C. gattii* isolates, including the PNW subtypes and the global diversity of molecular type VGII, to better ascertain the natural source and genomic adaptations leading to the emergence of infection in the PNW. Overall, the VGII population was highly diverse, demonstrating large numbers of mutational and recombinational events; however, the three dominant subtypes from the PNW were of low diversity and were completely clonal. Although strains of VGII were found on at least five continents, all genetic subpopulations were represented or were most closely related to strains from South America. The phylogenetic data are consistent with multiple dispersal events from South America to North America and elsewhere. Numerous gene content differences were identified between the emergent clones and other VGII lineages, including genes potentially related to habitat adaptation, virulence, and pathology. Evidence was also found for possible gene introgression from *Cryptococcus neoformans* var. *grubii* that is rarely seen in global *C. gattii* but that was present in all PNW populations. These findings provide greater understanding of *C. gattii* evolution in North America and support extensive evolution in, and dispersal from, South America.

IMPORTANCE *Cryptococcus gattii* emerged in the temperate North American Pacific Northwest (PNW) in the late 1990s. Beyond a new environmental niche, these emergent populations displayed increased virulence and resulted in a different pattern of clinical disease. In particular, severe pulmonary infections predominated in contrast to presentation with neurologic disease as seen previously elsewhere. We employed population-level whole-genome sequencing and analysis to explore the genetic relationships and gene content of the PNW *C. gattii* populations. We provide evidence that the PNW strains originated from South America and identified numerous genes potentially related to habitat adaptation, virulence expression, and clinical presentation. Characterization of these genetic features may lead to improved diagnostics and therapies for such fungal infections. The data indicate that there were multiple recent introductions of *C. gattii* into the PNW. Public health vigilance is warranted for emergence in regions where *C. gattii* is not thought to be endemic.

Cryptococcus gattii, an environmental basidiomycetous yeast and the primary cause of cryptococcosis in immunocompetent patients, has been previously primarily associated with central nervous system disease in tropical and subtropical regions of the world (1). This distribution and disease phenotype are in contrast with those of another causative agent of cryptococcosis, *Cryptococcus neoformans*, which is primarily an opportunistic pathogen of immunosuppressed patients and is distributed worldwide. In addition, *C. neoformans* is classically found in bird excreta and decaying wood from a wide variety of trees (2–4), whereas *C. gattii* is primarily found growing in tree bark, tree detritus, and decaying wood. *C. gattii* has been closely associated with a number of trees, notably *Eucalyptus* species, which were thought to be the dominant environmental source allowing for transportation of *C. gattii* out of Australia where it is endemic to locales, such as California, Greece, and elsewhere (5, 6). It has since been shown to be endemic in South America where it is found in close association with a large number of tropical trees (2, 7–9). *C. gattii* is currently classified into four distinct populations, referred to previously as "major molecular types": VGI (AFLP4), VGII (AFLP6), VGIII (AFLP5), and VGIV (AFLP7) (10), based on multilocus sequence typing (MLST), amplified fragment length polymorphism (AFLP), and other subgenomic genotyping methodologies. Each of these populations or molecular types can be divided further into distinct lineages or subtypes, as defined by AFLP profiling or MLST sequence typing, which identify individual clonal lineages within each molecular type.

In the late 1990s, infections caused by *C. gattii* molecular type VGII were reported on Vancouver Island in British Columbia, Canada (11–13); until this time, reports of *C. gattii* infection were uncommon in North America (1). Two clonal subtypes within the VGII molecular type known as VGIIa (AFLP6A) and VGIIb (AFLP6B) were identified as the cause of the outbreak, which subsequently spread to the Canadian mainland (14) and other regions of the Pacific Northwest (PNW) (15, 16). Within a decade, an additional novel VGII subtype, VGIIc, appeared in the PNW, primarily in the states of Oregon and Washington (16). Given the genomic similarity of isolates within each subtype and finding only the α mating type locus, all three subtypes were thought to be distinct clones (17). While all three PNW subtypes of *C. gattii* display a shared novel biome preference (presence in a temperate versus tropical climate), only VGIIb maintains the classical clinical features of neurologic dominance (18). The VGIIa and VGIIc subtypes, on the other hand, share distinct novel phenotypes (dominantly pulmonary disease and increased virulence, both *in vitro* and *in vivo*) (18).

Given their geographic proximity, time frame of emergence, and genotypic and phenotypic similarity, it was proposed that VGIIa and VGIIc originated relatively recently from a common ancestor (15, 19). Several studies were conducted to better understand the genotypic and phylogenetic nature of the *C. gattii* strains associated with these outbreaks (11, 13, 15, 17). VGIIa and VGIIb strains have now been found in other non-North American locales, and genotyping studies have provided evidence that these strains emerged independently from a South American source (9). The VGIIc subtype has not been identified elsewhere thus far.

A number of studies have described *in vitro* and *in vivo* clinical phenotypic variances (i.e., virulence in animal model, clinical presentation, etc.) between the emergent North American strains (15, 18, 20). In animal models of infection, VGIIa and VGIIc strains have been shown to be more virulent than VGIIb and other *C. gattii* molecular types (9, 21). It is also clear that the virulence profile within subtype VGIIa does not remain consistent globally with isolates of lesser virulence being described (13, 21; C. Firacative and W. Meyer, unpublished data).

Niche adaptation and phenotypic changes in pathogenic fungi are known to happen through a series of evolutionary mechanisms, including gene duplication, sexual recombination, base mutation, and to some degree, horizontal gene transfer (22). The potential genetic causes for the phenotypic and niche differences in the PNW *C. gattii* strains are not well understood, although potential contributing factors have been deduced from targeted expression analyses (20, 23). Whole-genome analyses provide a more comprehensive view of phylogenetic relationships and genomic evolution. Here we apply population-level whole-genome approach with the following aims: (i) to more thoroughly understand the population genomics of the *C. gattii* VGII molecular type; (ii) to explore the emergence of its subtypes in North America; and (iii) to identify possible genetic causes for its changing phenotype.

RESULTS

Whole-genome sequencing and SNP analysis. Whole genomes from 118 *Cryptococcus gattii* isolates, representing both the Pacific Northwest (PNW) subtypes and the known global VGII molecular genotype diversity, along with isolates of each of the other major *C. gattii* molecular types (VGI, VGIII, and VGIV), were sequenced and assembled *de novo*, using data from multiple next-generation sequencers (all Illumina technology). We obtained at least 10× coverage across an average of 98.27% of the genome (mode, 98.64%; range, 90.63% to 99.43%) compared to the assembled R265 reference (VGIIa), with an average depth of 111× (range, 24× to 532×). The average assembly size was ~17.55 Mb (range, 17.36 to 17.83 Mb).

We determined the presence of single nucleotide polymorphisms (SNPs) across and between subtypes. A total of 1,282,876 SNPs were identified from all genomes, including the VGI, VGII, VGIII, and VGIV molecular types; 544,881 SNPs were parsimony informative, that is they distinguished between more than one strain and were not autapomorphic (Fig. 1). Within the 115 VGII genomes, 310,969 SNPs were found, with 221,248 being parsimony informative (Fig. 2). Consistent with recently emerged and low-diversity clones, only 717, 246, and 528 SNPs were identified within the VGIIa, VGIIc, and VGIIb populations, respectively (Fig. 3, Fig. 4, and Fig. 5, respectively). A combined analysis of only the three PNW clones identified little to no homoplasy (data not shown), as previously described (17).

Genome sequence data analysis. Population-level genome analyses were initiated with an SNP-based assessment of phylogenetics and population structure, followed by gene presence/absence comparisons to understand the genetic relationships between and among the VGII subtypes in the PNW and other geographically diverse subtypes. A detailed description of the data analysis flowchart is shown in Fig. S1 in the supplemental material.

Phylogenetics. Isolates chosen for phylogenetic analysis were representative of the global diversity of *C. gattii* VGII based on their multilocus sequence type (MLST), including a number of isolates from each of the PNW lineages (VGIIa, VGIIb, and VGIIc). Additionally, isolates of the other three *C. gattii* popula-

70,000 SNPs
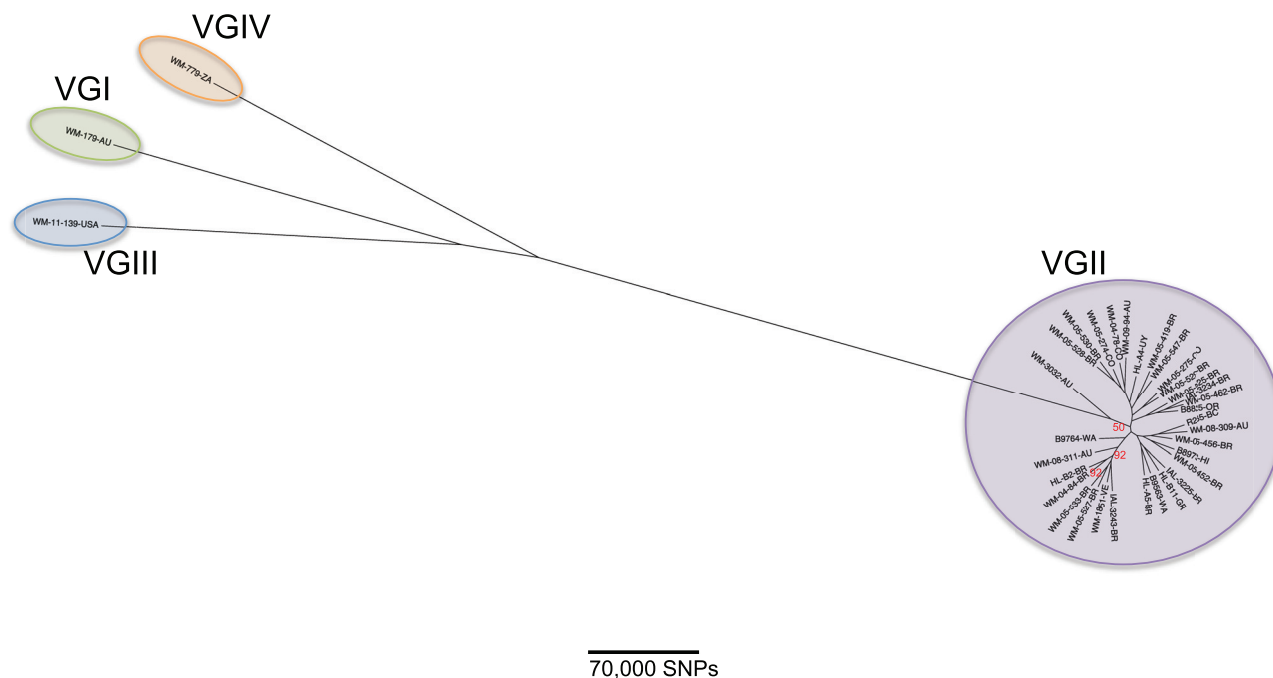
FIG 1 Phylogenetic analysis of *C. gattii* isolates representing the major molecular types VGI to VGIV. Maximum-parsimony phylogenetic analysis was performed on WGST SNP data from 35 *C. gattii* genomes (one each of VGI, VGIII, and VGIV molecular types and 32 genomes representing the diversity of the VGII subtypes). The analysis found 1,282,876 total SNPs with 544,881 of them being parsimonious SNPs, with a consistency index (CI) of 0.729. The tree shown is not rooted. Branch lengths represent numbers of SNPs between taxa, with the unit bar in the figure. Unless otherwise shown in red, all branches had bootstrap values of 100 with 500 generations. The taxon nomenclature includes a unique strain identifier and a two-letter country or state abbreviation for the geographic source of the original isolate (AR, Argentina; AU, Australia; AW, Aruba; BR, Brazil; CO, Colombia; DK, Denmark; GR, Greece; JP, Japan; MY, Malaysia; TH, Thailand; UY, Uruguay; VE, Venezuela; BC, British Columbia, Canada; ZA, South Africa; and for the United States, CA, California; FL, Florida; HI, Hawai'i; ID, Idaho; OR, Oregon; WA, Washington).

tions were analyzed. Based on whole-genome SNP typing (WGST), the VGII strains were clearly separated from other molecular types (i.e., VGI, VGIII, and VGIV), with greater than 500,000 SNPs between VGII and any other major molecular type (Fig. 1).

Using maximum-parsimony analysis for WGST, the VGII isolates separated mostly along the lines of their MLST (Table 1; see Fig. S2 and S3 in the supplemental material). A number of isolates with distinct sequence types (STs) were clearly more related to each other and likely represent members of the same clone with single or limited nucleotide variation within the MLST markers (e.g., VGIIc includes ST6 and ST49). Multiple clonal lineages are seen in the whole VGII phylogeny, including the PNW lineages (Fig. 2 and Fig. S2), and the long terminal branches indicate substantial genetic distance between most lineages. The maximum-parsimony tree (Fig. S2) had a relatively low consistency index of 0.4, indicating either a high degree of homoplasy from multiple mutations at the same sites or that VGII is not evolving in a strictly tree-like manner. To better visualize the relationships between lineages, we developed a neighbor-joining phylogenetic split tree network, or "neighbor net" (Fig. 2). The network contained many phylogenetic splits, which could not be condensed into a single tree phylogeny. This finding may indicate that sexual recombination events between distinct subtypes are contributing to the ge-

netic diversity of VGII, as these events will cause additional splits due to different regions of the genome being best described by different trees.

When the phi test for recombination was performed using the SNP data, the test indicated that recombination was present with a very high level of significance ($P = 0.0$). We then examined the phi value across the genome to see whether recombination could be localized to a particular subset of the genome. Using nonoverlapping windows of 500 concatenated SNP positions, we tested for recombination separately in 673 "windows" across the genome (encompassing variable genome window sizes due to the nonuniform density of SNPs in the genomes). Recombination was detected at a significant level in 631 of these windows when the P value cutoff was adjusted for making 673 comparisons ($P < 7.2^{-5}$), indicating that the recombination signal was common across most of the genome and not localized to a single region.

Due to the evidence for recombination among isolates, a model-based Bayesian approach (24) for identifying population structure and individual assignment to populations was added to our phylogenetic analyses. fineStructure analysis placed the 32 distinct VGII subtypes, as identified by WGS and MLST (see Fig. S2 and S3 in the supplemental material), into 11 "populations" or groups related by shared genomic regions, with some of these containing only a single representative (Fig. 6). For several of
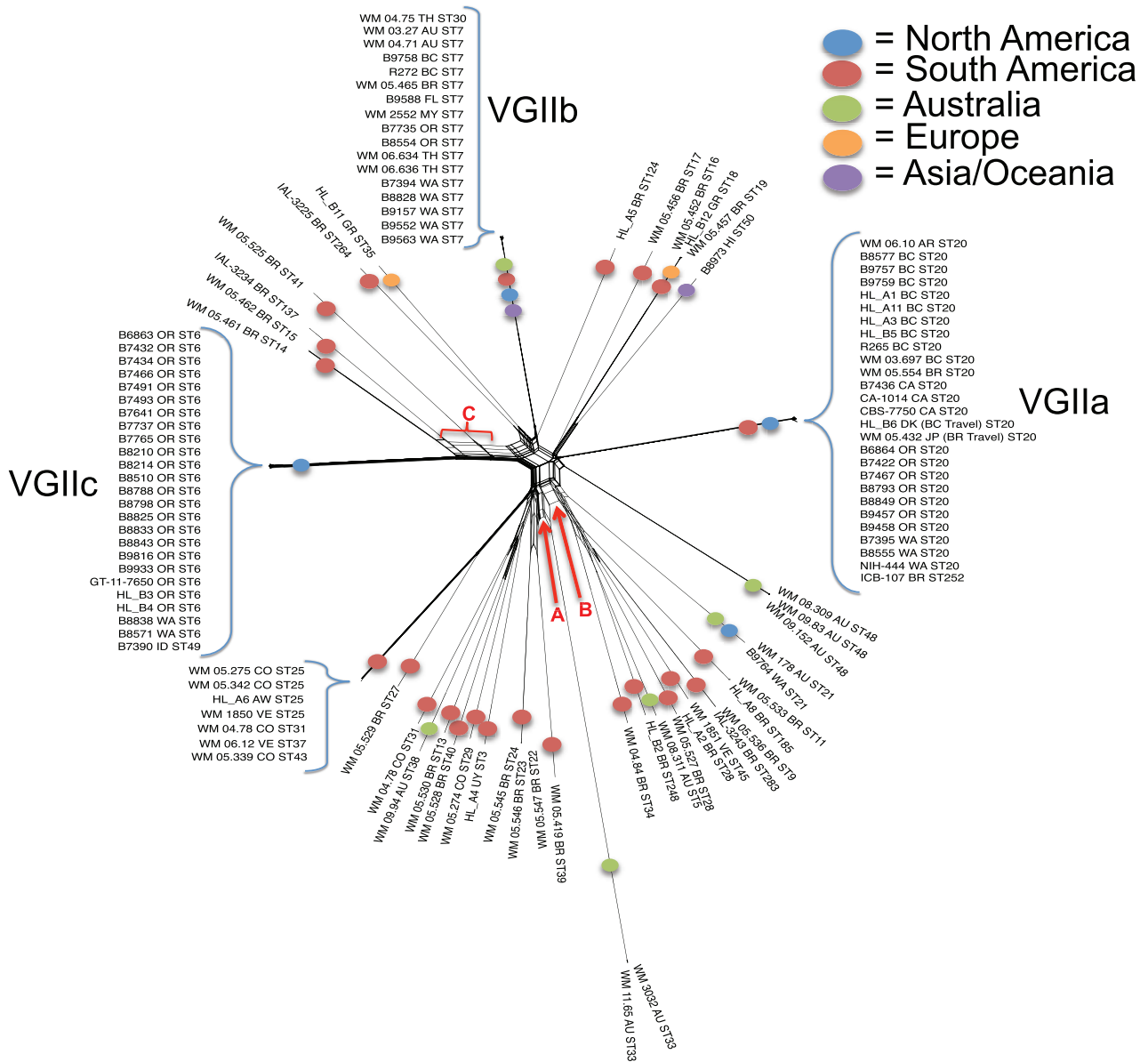
FIG 2 Phylogenetic network of *C. gattii* VGII. A neighbor net representation of the relationships among all VGII subtypes based on WGST SNP data, using the uncorrected *p*-distance transformation. Each band of parallel edges indicates a split. For example, the two splits that group ST33 with two distinct groups of other VGII subtypes are labeled with red A and B letters. The split joining VGIIc with three additional STs is labeled with a red C letter.

these groups, members shared many genome regions with members from other groups, indicating incomplete or recent separation between groups. This finding is also supported by the presence of isolates, such as B9764 and WM 05.525, which apparently share genetic material with several populations and may represent admixed individuals. Importantly, all three of the PNW subtypes were placed into separate fineStructure "populations," all of which include members from South America. The neighbor net network (Fig. 2), maximum parsimony-based WGST (Fig. S2), and MLST (Fig. S3) analyses all demonstrate broad dispersal of subtypes from South America throughout the VGII population.

The members of the PNW subtypes (i.e., VGIIa, VGIIc, and VGIIb) showed strict clonality within their respective subtype, with consistency indices (CI) of 1.0 for each subtype tree (Fig. 3 to

5), and significant separation from each other, as described previously (17). Bioinformatic attempts to determine a date for the most recent common ancestor of each subtype did not yield useful estimates (analysis not shown), which may be due to the small amount of SNP data available or potentially the nonclocklike evolution of *C. gattii*.

The VGIIa subtype contained two distinct clades, a North American-dominated group and a South American group (Fig. 3). The North American VGIIa clade showed little topological structure with a 15-branch polytomy consistent with a radiation emanating from a recent common ancestor. One subclade, joined by 13 shared SNPs, contained a set of three isolates previously described as having lower virulence (13, 20), one of which was isolated in Brazil in 1981 and was genomically separated from the
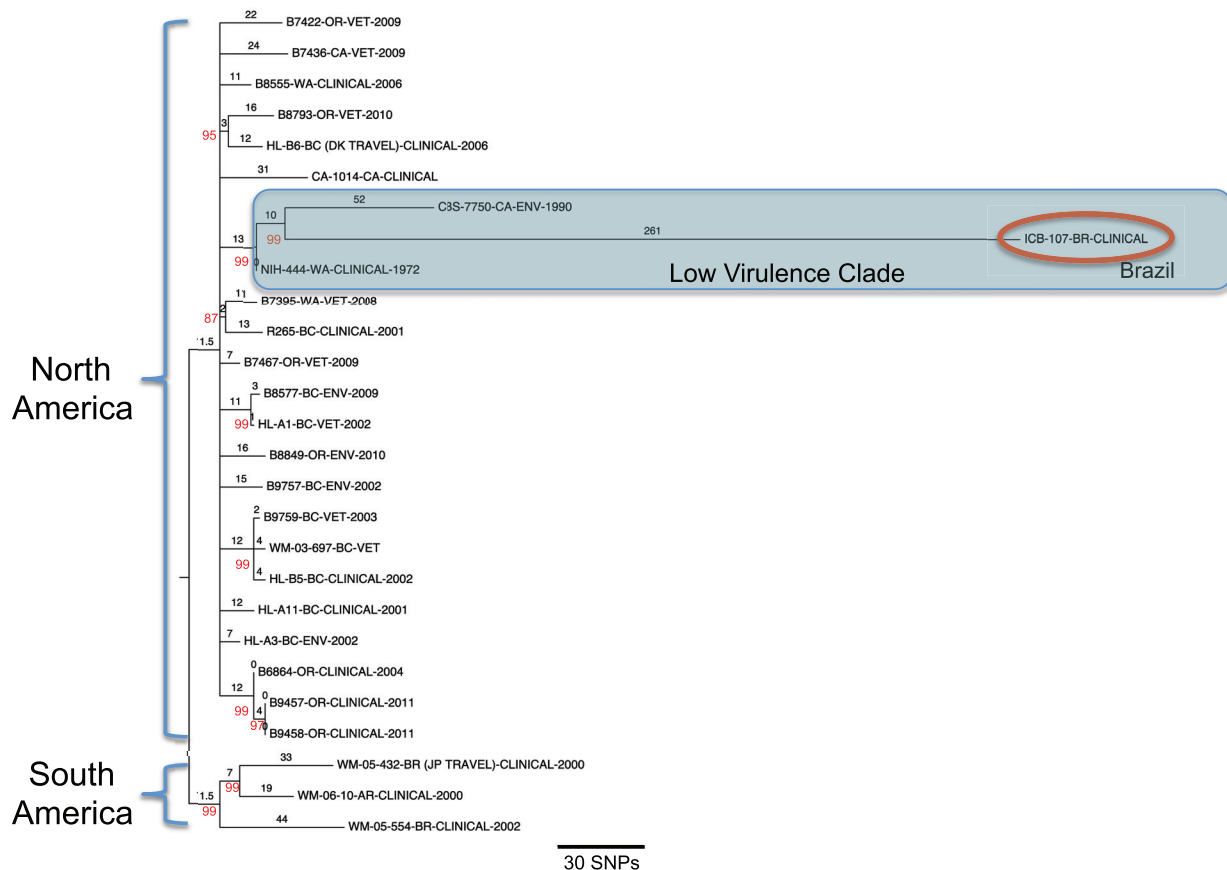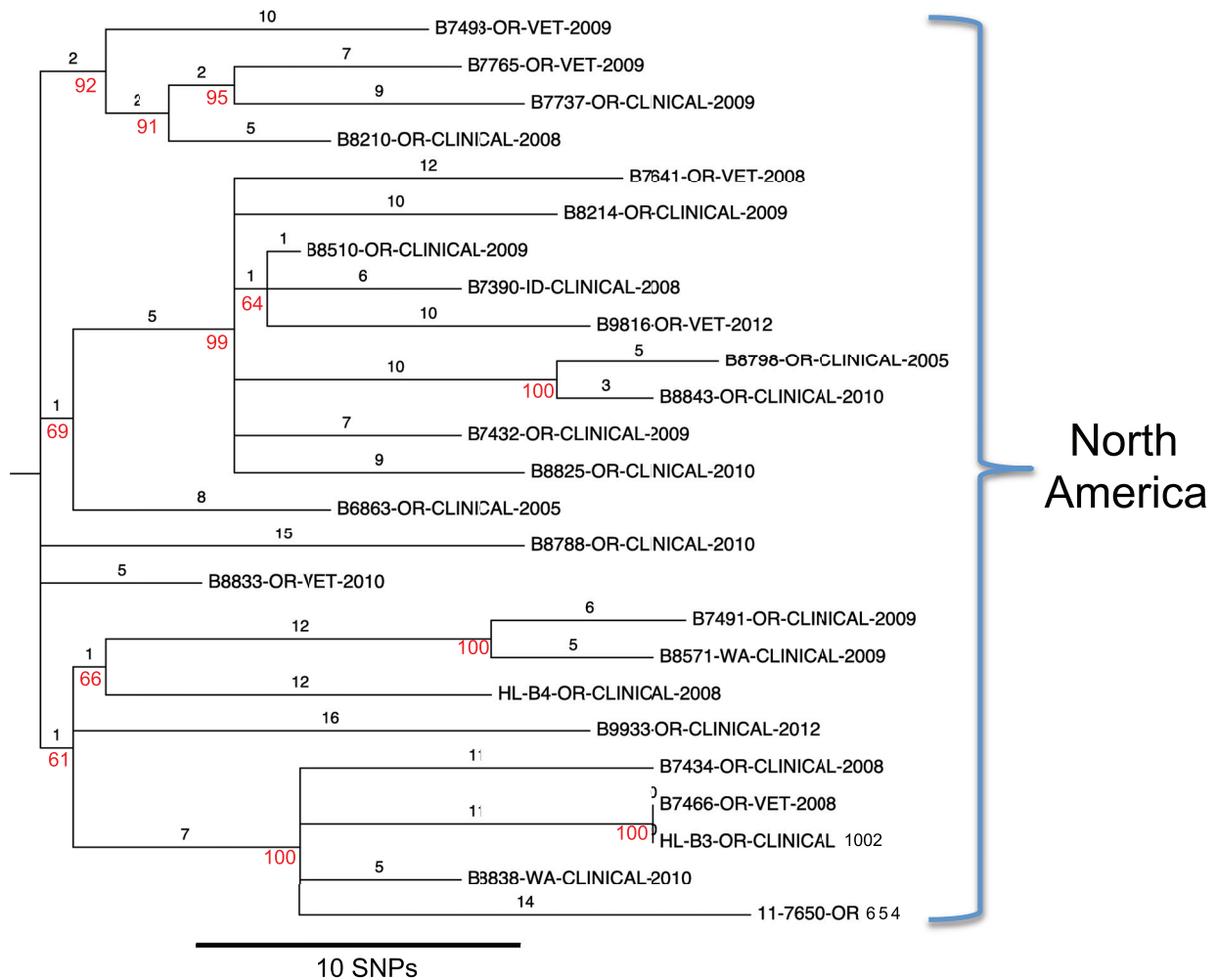
FIG 3 SNP phylogeny from whole genomes from VGIIa isolates. Maximum-parsimony phylogenetic analysis was performed on all isolate genomes within VGIIa. The VGIIa tree includes 717 SNPs (97 parsimony informative) from 17 isolates (CI = 1.0). For country and state abbreviations, see the legend to Fig. 1. Numbers in black on top of branches are SNP length, and numbers in red next to branches are the bootstrap values generated from 500 generations.

others by 261 additional SNPs. The other two members of this lower-virulence subclade were isolated in Washington State in 1972 and California in 1990 and had SNP branch lengths of 0 and 52, respectively. The VGIIc population was also strictly clonal (CI = 1.0) with a total of 55 parsimony informative SNPs (Fig. 4). This subtype phylogeny contained a 5-branched polytomy but with several distinct subclades. The branch lengths were short (<16 SNPs); notably, all of these isolates were collected between 2008 and 2012.

Unlike VGIIa and VGIIc, the VGIIb subtype is dispersed beyond the Western Hemisphere (6, 13, 15, 25). This subtype contained 528 SNPs (147 parsimony informative) and multiple distinct clades (Fig. 5). North American isolates were found in multiple locations across the tree, including in one large derived clade where they shared membership with isolates that were collected in Asia and Australia. The Asian-Australian/North American subclade shows polytomy, which is consistent with the presence of a genetic bottleneck and subsequent radiation, or may be a reflection of limited sampling. A single isolate from Florida shared membership in a separate basal clade with an isolate from Australia.

Gene content. *Ab initio* gene prediction using the pretrained Augustus software identified a median of 6,276 genes per genome (range, 6,213 to 6,359). The gene content analysis initially focused on differences and similarities between the PNW subtypes. The use of BLAT score ratio (BSR) for this analysis allowed us to make gene predictions from the genome sequences, where exon-coding DNA sequences were concatenated into single contiguous sequences to create the *in silico* coding DNA (is-cDNA), the presence or absence of which could then be scored for the PNW genomes. The gene sequences from the three PNW subtypes clustered into 6,302 is-cDNAs. The gene sequences from the remaining 29 non-PNW subtypes clustered into 6,541 is-cDNAs. The individual and shared gene differences in the PNW subtypes were scored against representative genomes from the other known global VGII subtypes. Additionally, we surveyed for the presence or absence of specific genes from any identified PNW clone (e.g., PNW-VGIIa) within the representative genomes from the other *C. gattii* major molecular types (VGI, VGIII, and VGIV), as well as within the *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans* reference genomes (Table 2). A heat map was developed (Fig. 7) to visualize the gene presence, absence, partial presence, and/or presence of genes with low sequence similarity. We identified several gene content similarities and differences in all of the comparisons (Table 2). Many of the identified putative genes, or is-cDNAs, were located in contiguous blocks of assembled genome sequence, labeled clusters 1 to 7 in Fig. 7. When the PNW clones were compared with each other, VGIIa had the largest number of PNW clone-specific gene content differences. Additionally, the PNW VGIIa-specific genes were typically absent in

FIG 4 SNP phylogeny from whole genomes from VGIIc isolates. Maximum-parsimony phylogenetic analysis was performed on all isolate genomes within VGIIc. The VGIIc tree includes 246 SNPs (55 parsimony informative) from 25 isolates (CI = 1.0). For state abbreviations, see the legend to >Fig. 1. Numbers in black on top of branches are SNP length, and numbers in red next to branches are the bootstrap values generated from 500 generations.

the other VGII subtypes and non-VGII genomes. Conversely, the PNW VGIIb-specific and VGIIc-specific genes (i.e., those genes found only in their respective subtype in the PNW) were commonly found in the other VGII subtypes and non-VGII *C. gattii* genomes. A large number of genes that were present in both VGIIa and VGIIc were absent from VGIIb. These genes, most of which were located adjacent to each other in clusters 1 and 2, were found in nearly all other *C. gattii* genomes, indicating a likely gene loss event in the PNW VGIIb population. An analysis of genes common to all three PNW populations but absent in other VGII subtypes identified a three-gene cluster ("cluster 6") that was found to be absent in most *C. gattii* genomes but was present, with a high degree of similarity, in the *C. neoformans* var. *grubii* genome (Fig. 7).

A comparison of phylogenies based on linked SNPs and gene content showed dissimilar relationships among VGII subtypes (data not shown). This suggests that differential gene content has not become fixed within the VGII lineages. As such, placement of an isolate in a phylogeny or within a subtype using SNP data is not necessarily a good indicator of its gene content, and conversely, gene content is not a good indicator of phylogenetic placement.

DISCUSSION

The emergence of *C. gattii* in temperate British Columbia in Canada and the U.S. PNW was unexpected in that *C. gattii* was previously thought to be prevalent only in tropical and semitropical regions. There has since been much speculation on the origins of the outbreak strains (9, 12, 15, 16). In this study, whole-genome sequencing was performed on 118 *C. gattii* isolates, primarily of the molecular type VGII, to better ascertain the source and genomic adaptations behind its emergence in the PNW. Our primary strategy was to analyze SNP mutations and gene content and to compare the PNW subtype genomes with a genomically diverse global collection of isolates. This study shows the following. (i) There were most likely multiple distinct North American introductions of at least two of the three PNW VGII subtypes. (ii) Gene content varies between the VGII subtypes. (iii) *C. gattii* has developed a number of evolutionary strategies that allow for continual niche adaptation. This study also explores the relationship of the North American VGII subtypes to a global *C. gattii* collection and the potential role of unique genes in virulence and niche adaption.
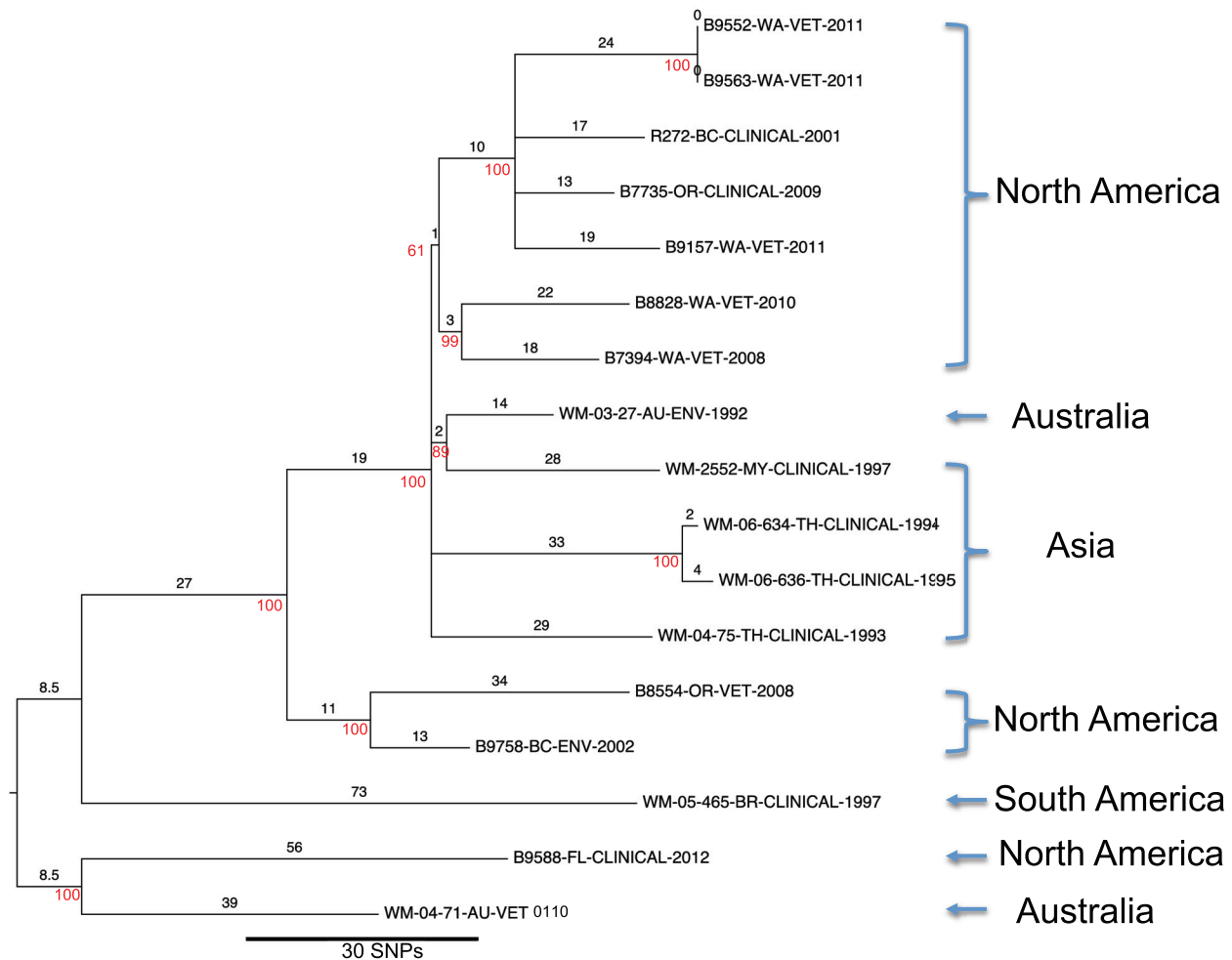
FIG 5 SNP phylogeny from whole genomes from VGIIb isolates. Maximum-parsimony phylogenetic analysis was performed on all isolate genomes within VGIIb. The VGIIb tree includes 528 SNPs (147 parsimony informative) from 17 isolates (CI = 1.0). For country or state abbreviations, see the legend to >Fig. 1. Numbers in black on top of branches are SNP length, and numbers in red next to branches are the bootstrap values generated from 500 generations.

Out of South America? (phylogenetic analyses). Analysis of the population genome structure of the PNW VGII clones provides additional significant evidence of multiple and continued emergence of apparently regionally fit lineages from within VGII originating in South America, as has been proposed previously (9, 26). As the apparent level of recombination across VGII negates the use of a single SNP-based tree to understand the true phylogenetic relationships among all VGII members, we conducted a neighbor net analysis, providing a better understanding of relationships within the phylogenetic network. For example, the Australian ST33 subtype, which appears most genetically distant on the WGST SNP tree (see Fig. S2 in the supplemental material), still appears most genetically distant in the network; however, there are several distinct splits uniting this subtype with other subtypes (see splits A and B on Fig. 2). The ST33 lineage, therefore, may have continued to exchange genetic material with these taxa after its derivation from the bulk of currently sampled VGII taxa.

From within the genomic diversity of VGII, we can see the emergence of the three low-diversity and highly fit clones (i.e., VGIIa, VGIIb, and VGIIc) that have become prevalent in North America. The emergence of these three clonal subtypes clearly represents at least three distinct events. Our phylogenetic analyses suggest that the phenotypically differentiated VGIIa and VGIIc lineages evolved in South America and spread to the PNW. Additionally, the widely disseminated VGIIb lineage and the VGIIa subtype were likely exported, possibly out of South America, on multiple occasions, and the VGIIb subtype has emerged in multiple geographic locations.

(i) VGIIa. The likely most recent translocation of the VGIIa lineage, as inferred from the very short branch lengths within its respective clade (Fig. 3), resulted in its distinct emergence on Vancouver Island, spreading to the Canadian mainland and into Northwest United States. There is a distinct separation of the PNW dominant VGIIa clade and a clade of South American origin (i.e., Brazil, Argentina, and Japan [in a Brazilian emigrant]) strains. The two VGIIa clades likely have a recent common ancestor, and the lack of its presence elsewhere suggests that this subtype originated in South America. Additionally, WGST allowed us to identify genetic separation between the strains identified during the recent emergence on Vancouver Island and the subclade composed of the older, but less virulent, North American VGIIa strains: NIH-444 (1972; Washington) and CBS7750 (1990; Cali-

TABLE 1 *Cryptococcus gattii* and *C. neoformans* isolates used in this study[a]

Genotype	Strain	ST	Country	Source	Yr of isolation	Figure inclusion
VGI	WM 179	51	Australia	Clinical	1993	1, 6, 7
VGIIa	B6864	20	USA	Clinical	2004	2, 3, S2, S3
	B7395	20	USA	Veterinary	2008	2, 3, S2, S3
	B7422	20	USA	Veterinary	2009	2, 3, S2, S3
	B7436	20	USA	Veterinary	2009	2, 3, S2, S3
	B7467	20	USA	Veterinary	2009	2, 3, S2, S3
	B8555	20	USA	Clinical	2006	2, 3, S2, S3
	B8577	20	Canada	Environmental	2009	2, 3, S2, S3
	B8793	20	USA	Veterinary	2010	2, 3, S2, S3
	B8849	20	USA	Environmental	2010	2, 3, S2, S3
	B9457	20	USA	Clinical	2011	2, 3, S2, S3
	B9458	20	USA	Clinical	2011	2, 3, S2, S3
	B9757	20	USA	Environmental	2002	2, 3, S2, S3
	B9759	20	USA	Veterinary	2003	2, 3, S2, S3
	CA-1014	20	USA	Clinical	?	2, 3, S2, S3
	CBS 7750	20	USA	Environmental	1990	2, 3, S2, S3
	R265	20	Canada	Clinical	2001	1, 2, 3, 6, 7, S2, S3
	HL_A1	20	Canada	Veterinary	2002	2, 3, S2, S3
	HL_A3	20	Canada	Environmental	2002	2, 3, S2, S3
	HL_A11	20	Canada	Clinical	2001	2, 3, S2, S3
	HL_B5	20	Canada	Clinical	2002	2, 3, S2, S3
	HL_B6	20	Denmark	Clinical	2006	2, 3, S2, S3
	NIH 444	20	USA	Clinical	1975	2, 3, S2, S3
	WM 03.697	20	Canada	Veterinary	2001	2, 3, S2, S3
	WM 05.432	20	Japan-Brazil	Clinical	2000	2, 3, S2, S3
	WM 05.554	20	Brazil	Clinical	2002	2, 3, S2, S3
	WM 06.10	20	Argentina	Clinical	2000	2, 3, S2, S3
	ICB-107	252	Brazil	Clinical	1981	2, 3, S2, S3
VGIIb	B7394	7	USA	Veterinary	2008	2, 5, S2, S3, S4
	B7735	7	USA	Clinical	2009	2, 5, S2, S3, S4
	B8554	7	USA	Veterinary	2008	2, 5, S2, S3, S4
	B8828	7	USA	Veterinary	2010	2, 5, S2, S3, S4
	B9157	7	USA	Veterinary	2011	2, 5, S2, S3, S4
	B9552	7	USA	Veterinary	2011	2, 5, S2, S3, S4
	B9563	7	USA	Veterinary	2011	1, 2, 5, 6, 7, S2, S3, S4
	B9588	7	USA	Clinical	2012	2, 5, S2, S3, S4
	B9758	7	USA	Environmental	2002	2, 5, S2, S3, S4
	CDC R272	7	Canada	Clinical	2001	2, 5, S2, S3, S4
	WM 2552	7	Malaysia	Clinical	1997	2, 5, S2, S3, S4
	WM 03.27	7	Australia	Environmental	1992	2, 5, S2, S3, S4
	WM 04.71	7	Australia	Veterinary	1991	2, 5, S2, S3, S4
	WM 05.465	7	Brazil	Clinical	1997	2, 5, S2, S3, S4
	WM 06.634	7	Thailand	Clinical	1994	2, 5, S2, S3, S4
	WM 06.636	7	Thailand	Clinical	1995	2, 5, S2, S3, S4
	WM 04.75	30	Thailand	Clinical	1993	2, 5, S2, S3, S4
VGIIc	B6863	6	USA	Clinical	2005	2, 4, S2, S3
	B7432	6	USA	Clinical	2009	2, 4, S2, S3
	B7434	6	USA	Clinical	2008	2, 4, S2, S3
	B7466	6	USA	Veterinary	2008	2, 4, S2, S3
	B7491	6	USA	Clinical	2009	2, 4, S2, S3
	B7493	6	USA	Veterinary	2009	2, 4, S2, S3
	B7641	6	USA	Veterinary	2008	2, 4, S2, S3
	B7737	6	USA	Clinical	2009	2, 4, S2, S3
	B7765	6	USA	Veterinary	2009	2, 4, S2, S3
	B8210	6	USA	Clinical	2008	2, 4, S2, S3
	B8214	6	USA	Clinical	2009	2, 4, S2, S3
	B8510	6	USA	Clinical	2009	2, 4, S2, S3
	B8571	6	USA	Clinical	2009	2, 4, S2, S3
	B8788	6	USA	Clinical	2010	2, 4, S2, S3
	B8798	6	USA	Clinical	2005	2, 4, S2, S3

(Continued on following page)

TABLE 1 (Continued)

Genotype	Strain	ST	Country	Source	Yr of isolation	Figure inclusion
	B8825	6	USA	Clinical	2010	1, 2, 4, S2, S3
	B8833	6	USA	Clinical	2010	2, 4, S2, S3
	B8838	6	USA	Clinical	2010	2, 4, S2, S3
	B8843	6	USA	Clinical	2010	2, 4, S2, S3
	B9816	6	USA	Veterinary	2012	2, 4, S2, S3
	B9933	6	USA	Clinical	2012	2, 4, S2, S3
	GT-11-7650	6	USA	Veterinary	?	2, 4, S2, S3
	HL_B3	6	USA	Clinical	2005	2, 4, S2, S3
	HL_B4	6	USA	Clinical	2008	2, 4, S2, S3
	B7390	49	USA	Clinical	2008	2, 4, S2, S3
VGII	B8973	50	USA	Clinical	2010	1, 2, 6, 7, S2, S3
	B9764	21	USA	Veterinary	2012	1, 2, 6, 7, S2, S3
	IAL-3225	264	Brazil	Veterinary	1994	1, 2, 6, 7, S2, S3
	IAL-3234	137	Brazil	Veterinary	1998	1, 2, 6, 7, S2, S3
	IAL-3243	283	Brazil	Veterinary	2000	1, 2, 6, 7, S2, S3
	HL_A2	28	Brazil	Clinical	Pre-2008	2, S2, S3
	HL_A4	3	Uruguay	Environmental	1996	1, 2, 6, 7, S2, S3
	HL_A5	124	Brazil	Environmental	Pre-2008	1, 2, 6, 7, S2, S3
	HL_A6	25	Aruba	Veterinary	1953	2, S2, S3
	HL_A8	185	Brazil	Environmental	Pre-2008	2, S2, S3
	HL_B2	248	Brazil	Veterinary	2000	1, 2, 6, 7, S2, S3
	HL_B11	35	Greece	Clinical	1996	1, 2, 6, 7, S2, S3
	HL_B12	18	Greece	Clinical	1998	2, S2, S3
	WM 178	21	Australia	Clinical	1991	2, S2, S3
	WM 1850	25	Venezuela	Clinical	1999	2, S2, S3
	WM 1851	45	Venezuela	Clinical	1999	1, 2, 6, 7, S2, S3
	WM 3032	33	Australia	Clinical	1983	1, 2, 6, 7, S2, S3
	WM 04.78	31	Colombia	Clinical	1998	1, 2, 6, 7, S2, S3
	WM 04.84	34	Brazil	Clinical	1986	1, 2, 6, 7, S2, S3
	WM 05.274	29	Colombia	Clinical	2002	1, 2, 6, 7, S2, S3
	WM 05.275	25	Colombia	Clinical	2001	1, 2, 6, 7, S2, S3
	WM 05.339	43	Colombia	Environmental	2005	2, S2, S3
	WM 05.342	25	Colombia	Environmental	2005	2, S2, S3
	WM 05.419	39	Brazil	Clinical	1988	1, 2, 6, 7, S2, S3
	WM 05.452	16	Brazil	Clinical	1995	1, 2, 6, 7, S2, S3
	WM 05.456	17	Brazil	Environmental	1994	1, 2, 6, 7, S2, S3
	WM 05.457	19	Brazil	Clinical	1995	2, S2, S3
	WM 05.461	14	Brazil	Clinical	1997	2, S2, S3
	WM 05.462	15	Brazil	Clinical	1997	1, 2, 6, 7, S2, S3
	WM 05.525	41	Brazil	Clinical	1997	1, 2, 6, 7, S2, S3
	WM 05.527	28	Brazil	Clinical	1997	1, 2, 6, 7, S2, S3
	WM 05.528	40	Brazil	Clinical	2001	1, 2, 6, 7, S2, S3
	WM 05.529	27	Brazil	Clinical	1997	1, 2, 6, 7, S2, S3
	WM 05.530	13	Brazil	Clinical	1999	1, 2, 6, 7, S2, S3
	WM 05.533	11	Brazil	Clinical	1997	1, 2, 6, 7, S2, S3
	WM 05.536	9	Brazil	Clinical	1997	2, S2, S3
	WM 05.545	24	Brazil	Clinical	2001	2, S2, S3
	WM 05.546	23	Brazil	Clinical	2001	2, S2, S3
	WM 05.547	22	Brazil	Clinical	2001	1, 2, 6, 7, S2, S3
	WM 06.12	37	Venezuela	Clinical	1996	2, S2, S3
	WM 08.309	48	Australia	Veterinary	1997	1, 2, 6, 7, S2, S3
	WM 08.311	5	Australia	Veterinary	1996	1, 2, 6, 7, S2, S3
	WM 09.94	38	Australia	Veterinary	2001	1, 2, 6, 7, S2, S3
	WM 09.152	48	Australia	Environmental	2009	2, S2, S3
	WM 11.65	33	Australia	Veterinary	2011	2, S2, S3
VGIII	WM 11.139	143	USA	Veterinary	2011	1, 7
VGIV	WM 779	70	South Africa	Veterinary	1994	1, 7
Cryptococcus neoformans						
VNI	H99	2	USA	Clinical	1978	7
VNIV	JEC21	126	USA	Clinical	1997	7

[a] The strain designation, major molecular type, MLST sequence type, and figures that the strains are shown on are indicated.
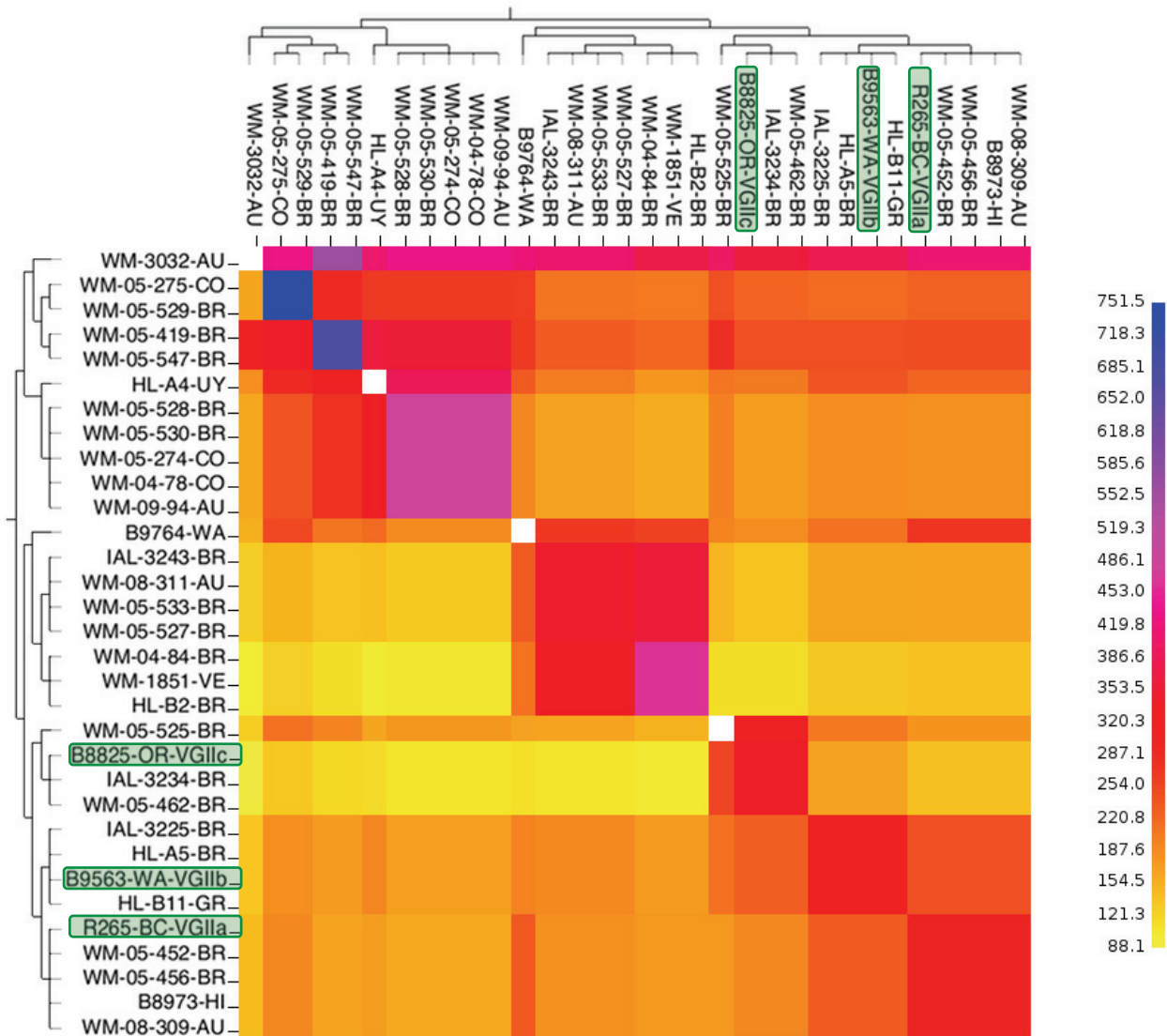
FIG 6 fineStructure analysis of dominant VGII subtypes. To identify the population structure and possible recombination events within VGII, fineStructure analysis was performed using the SNP matrix developed for >Fig. 2. Using one representative from each of the 32 VGII subtypes, whole-genome SNP data were reduced to a pairwise similarity matrix. The population structure is identified using this similarity matrix, where the underlying model assumes individuals within populations will share more regions of their genome with each other and have a similar amount of admixture with individuals from different populations. Identified populations are merged one at a time, at each step using the most likely population merge, to relate populations to each other through a tree. The x-axis analysis represents the strain as a "recipient," and the y axis represents the strain as a "donor" of genomic regions. The scale bar represents the number of shared genome regions with blue being the greatest amount of sharing and yellow being the least. The PNW-associated subtype representatives are outlined in green.

fornia) (16). This lower-virulence subclade also contained a 1981 Brazilian isolate (ICB-107). It is possible that the latter isolate represents the originating lineage in Brazil and that the two North American isolates represent two distinct translocation events of that same lineage to North America prior to the eventual PNW emergence. Strain ICB-107 was isolated from a patient in Recife, a major port city in Brazil, representing a possible point of export. A lack of epidemiological information associated with that particular isolate prevents further understanding at this time.

(ii) VGIIc. The VGIIc subtype has been found only in the PNW. Its most recent splits (i.e., closest genetic relatives) in the phylogenetic network were found exclusively in South America

(see split A in Fig. 2). The very short branch lengths and lack of distinct separate clades within this subtype (Fig. 4) are indicative of an individual emergence of this subtype, likely from South America, where all of the most closely related subtypes were identified.

(iii) VGIIb. VGIIb is the most diverse of the VGII lineages found in the PNW and likely has been translocated to North America and other parts of the world on multiple occasions. The origin of PNW VGIIb strains is not clear. A VGIIb subclade defined by the putative loss of a 12-gene 10-kb contiguous genomic region (see Fig. S4 in the supplemental material; cluster 1 in Fig. 7) contains most, but not all, PNW VGIIb strains along with strains

TABLE 2 BLAT score ratio gene content comparison

Gene content similarity and is-cDNA	Family hit(s)	Pfam	E value(s)
Present in PNW VGIIa/c			
is-cDNA 1	Transmembrane amino acid transporter protein family	PF01490.13	1.90E−33
is-cDNA 2	Endoribonuclease L-PSP	PF01042.16	8.30E−39
is-cDNA 3	Aminotransferase class I and II domain	PF00155.16	5.20E−47
is-cDNA 4	No significant matches		
is-cDNA 5	Zinc finger double domain, fungus-specific transcription factor domain	PF13465.1, PF04082.13	4.4E−7, 2.3E−16
is-cDNA 6	Major facilitator superfamily	PF07690.11	5.80E−11
is-cDNA 7	Peptidase family M20/M25/M40, peptidase dimerization domain	PF01546.23, PF07687.9	3E−6, 4.2E−9
is-cDNA 8	Major facilitator superfamily, fungus-specific transcription factor domain	PF07690.11, PF04082.13	5.5E−29, 9.6E−12
is-cDNA 9	No significant matches		
is-cDNA 10	Arylsulfotransferase (ASST) family	PF14269.1	9.90E−14
is-cDNA 11	Major facilitator superfamily	PF07690.11	7.20E−23
is-cDNA 12	Phosphatidylethanolamine-binding protein domain	PF01161.15	5.70E−18
is-cDNA 13	Fungus-specific transcription factor domain	PF04082.13	2.90E−05
is-cDNA 14	Glycosyl hydrolase family 32 N-terminal domain, glycosyl hydrolase family 32 C-terminal domain	PF00251.15, PF08244.7	1.8E−51 4.8E−10
is-cDNA 15	No significant matches		
is-cDNA 16	Alcohol dehydrogenase GroES-like domain, zinc-binding dehydrogenase	PF08240.7, PF13602.1	8.5E−6, 2.8E−19
is-cDNA 17	No significant matches		
Present in PNW VGIIb			
is-cDNA 18	E1-E2 ATPase, haloacid dehalogenase-like hydrolase		3.6E−47, 3.1E−21
is-cDNA 19	No significant matches		
is-cDNA 20	No significant matches		
Present in PNW VGIIa			
is-cDNA 21	Sugar (and other) transporter	PF00083.19	1.20E−94
is-cDNA 22	Alcohol dehydrogenase GroES-like domain, zinc-binding dehydrogenase	PF08240.7, PF00107.21	1.6E−28, 1.6E−16
is-cDNA 23	Major facilitator superfamily	PF07690.11	1.20E−17
is-cDNA 24	No significant matches		
is-cDNA 25	Aldo/keto reductase family	PF00248.16	1.30E−52
is-cDNA 26	Protein of unknown function, putative collagen-binding domain of a collagenase	PF13204.1, PF12904.2	2.1E−54, 7.4E−19
is-cDNA 27	Sugar (and other) transporter	PF00083.19	3.70E−93
is-cDNA 28	Alcohol dehydrogenase GroES-like domain, zinc-binding dehydrogenase	PF08240.7, PF00107.21	5.3E−28, 1.2E−15
is-cDNA 29	Fungus-specific transcription factor domain, fungal Zn(2)-Cys(6) binuclear cluster domain	PF04082.13, PF00172.13	2E−16, 1.8E−4
is-cDNA 30	No significant matches		
is-cDNA 31	No significant matches		
is-cDNA 32	Aspartyl protease	PF09668.5	3.30E−05
is-cDNA 33	Helicase conserved C-terminal domain, DEAD/DEAH box helicase	PF00271.26, PF00270.24	4.6E−7, 1E−4
Present in PNW VGIIc			
is-cDNA 34	Fungus-specific transcription factor domain	PF04082.13	2.50E−11
is-cDNA 35	Lyase, adenylosuccinate lyase C terminus	PF00206.15, PF10397.4	1.5E−47, 2.6E−18
is-cDNA 36	Sugar (and other) transporter	PF00083.19	5.40E−80
is-cDNA 37	No significant matches		
is-cDNA 38	No significant matches		
is-cDNA 39	Protein of unknown function	PF11175.3	1.40E−80
is-cDNA 40	Sugar (and other) transporter	PF00083.19	8.30E−63
is-cDNA 41	Protein of unknown function	PF11175.3	6.70E−83

(Continued on following page)

TABLE 2 (Continued)

Gene content similarity and is-cDNA	Family hit(s)	Pfam	E value(s)
is-cDNA 42	No significant matches		
is-cDNA 43	No significant matches		
Present in PNW VGIIa/b/c and missing in half of VGII STs			
is-cDNA 44	Glycosyl hydrolase family 3 N-terminal domain, fibronectin type III-like domain	PF00933, PF14310	1.9E−70,7.5E−20
is-cDNA 45	Alpha-L-rhamnosidase N-terminal domain, bacterial alpha-L-rhamnosidase	PF08531, PF05592	04.7E−36, 9.5E−166
is-cDNA 46	Sugar (and other) transporter	PF00083	2.00E−57
is-cDNA 47	No significant matches		
Missing in PNW and present in half of VGII STs			
is-cDNA 48	Fungus-specific transcription factor domain	PF04082.13	1.90E−23
is-cDNA 49	Proline dehydrogenase	PF01619.13	4.60E−31
is-cDNA 50	Transmembrane amino acid transporter protein	PF01490.13	2.8E−17, 1E−9

collected in the previous decade in Australia and Asia (6, 25). These strains are the most commonly found VGIIb strains in the PNW and they reportedly are less virulent than the other VGIIb strains found in the PNW (16). This 10-kb contig is absent from the derived clade yet occurs in all other, more-basal VGIIb strains, including two additional isolates found in PNW, which strongly suggests a single gene loss event in a parental strain. This event, again, provides evidence for multiple independent introductions of VGIIb to North America, which are discernible by both SNP states and genetic content. Further studies are necessary to assess the phenotypic impact of the loss of this genomic component.

The maximum possible consistency indices (CI = 1.0) for the VGIIa, VGIIb, and VGIIc phylogenetic trees demonstrate a lack of observable recombination occurring within or among the North America VGII isolates. These data, taken with the fact that all PNW strains contain only the mating type α locus, indicate that there is limited to no sexual reproduction occurring in these strains. Same-sex mating was previously hypothesized as an evolutionary driver of these subtypes (13). As no phylogenetic inconsistencies were identified among strains within each PNW subtype, there is no obvious evidence of same-sex mating in the region as well, although the progeny of such events may not be readily identified due to the genomic similarities of the parental strains. As sexual reproduction is typically necessary for the development of infectious propagules, it is possible that individual isolates undergo unisexual reproduction (selfing) or haploid fruiting, which have been previously described in *C. neoformans* (27, 28), to induce a yeast-to-hyphal transition.

Searching for North American adaptations (comparative genomics). In order to identify genetic content associated with niche adaptation to the PNW, we explored the population genome content for genes associated with each of the North American VGII subtypes compared to each other and to other *C. gattii* molecular types. We were able to identify a subtype-specific gene

present (compared to other PNW subtypes) in each of the three subtypes, as well as shared genes in VGIIa and VGIIc compared to the VGIIb subtype. These genomic differences could be the basis for specific niche adaptations and possible differences in virulence and tissue tropism (18). None of the PNW VGIIa, VGIIb, and/or VGIIc clone-defining genes or gene regions was completely unique to North American populations; all were identified in at least once in other global VGII genotypes analyzed.

(i) Notable gene differences. The gene regions specific to VGIIa in the PNW were rarely seen in other global subtypes and may represent greater likelihood of novel genes. The majority of these VGIIa-specific regions were in a contiguous gene cluster (gene cluster 3, which contains is-cDNA 21 to 29, in Fig. 7) that was seen only in three other global subtypes, originating from Brazil, Colombia, and Australia. This gene cluster contains several genes associated with sugar transport and alcohol dehydrogenase and is possibly important for regional plant sugar metabolism (Table 2). This cluster also includes a collagen-binding domain, likely from a collagenase enzyme, which has been identified in other pathogenic fungi (29). Collagen is a component of all major lung cells (30), and collagen binding in the lung has been associated with colonization of *Aspergillus* in the asthmatic lung (31) and increased virulence in other pulmonary infections (32). The presence of a collagen-binding domain, therefore, may be relevant to binding of cryptococci to host lung cells and contribute to increased virulence or lung tissue tropism or both in the PNW VGIIa infections (18).

Interestingly, the VGIIa cluster 3 gene cluster is also absent from the other major *C. gattii* types (VGI, VGIII, and VGIV), except for variants of two of the sugar transport/metabolic domains that were present in the VGI genome, possibly an indication of a historical gene transfer event. Additionally, this gene cluster was absent from the *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans* reference genomes (Fig. 7).
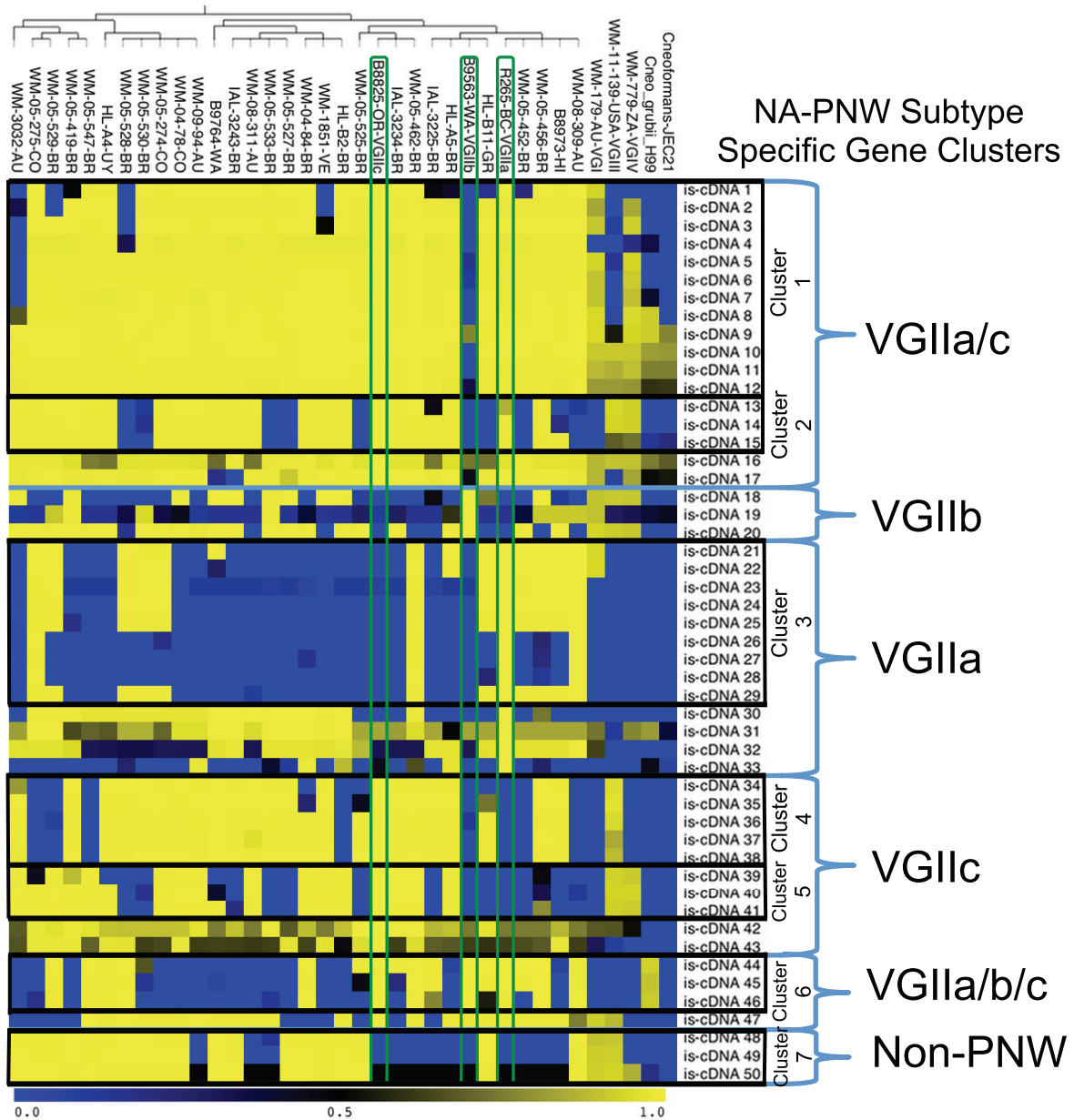
FIG 7 Population-level gene content analysis. To identify shared and differing gene content between and among the PNW subtypes compared to other VGII subtypes, other *C. gattii* molecular types, and other *Cryptococcus* species, BLAT score ratio analysis was performed, which is described in detail in the Materials and Methods. The genomes scored are the same as in >Fig. 6 plus one representative each from VGI, VGIII, VGIV, *C. neoformans* var. *neoformans*, and *C. neoformans* var. *grubii*. VGII isolates are presented in the populations identified by fineStructure analysis where the tree on top is reproduced from >Fig. 6. The rows on the left are divided by comparison group (VGIIa only, VGIIa and VGIIc only, etc.); these rows are further grouped by identified contiguous gene clusters (clusters 1 to 7). Genes of high homology between groups are yellow (complete homology = 1.0 on the heat map bar), and no homology (complete absence of the scored gene = 0.0) is represented by blue, with intermediate colors representing a gradient of similarity. The PNW population representatives and their scores are outlined in green.

An additional gene complex found in the VGIIa group, a DEAD box RNA helicase (is-cDNA 33), was rarely seen in VGII subtypes outside the PNW. Semihomologous genes or gene fragments were identified in VGIIc and four additional VGII subtypes and in the VGIII genome, as well as in the *C. neoformans* var. *grubii* genome. The DEAD box RNA helicase has been found previously in multiple eukaryotic and prokaryotic genomes and was recently established as an important virulence factor in *C. neoformans*, with reported impacts on environmental sensing, pathogenesis, and RNA metabolism (33). Additionally, RNA helicases have been associated with fungal cold response (34). As such, its presence/partial presence in VGIIa and VGIIc and relatively few other VGII lineages may suggest a linkage to differential virulence and/or temperate climate phenotype. Its absence from most other *C. gat-*

tii genomes suggests one or more intertype or interspecies gene transfer events.

An analysis of the genome content of the three North American VGII subtypes against all other VGII genomes in the current study revealed an ~10-kb region, containing three genes, that was consistently found in all VGIIa, VGIIb, and VGIIc genomes. This cluster (cluster 6, which contains is-cDNA 44 to 46, in Fig. 7) was present in only a limited number of other VGII subtypes and was absent from the representative VGI, VGIII, and VGIV strains. BSR analysis revealed that the region had high similarity with a homologous region of the same size in the *C. neoformans* var. *grubii* reference genome but was missing from *C. neoformans* var. *neoformans*. The three genes in this cluster were identified as encoding rhamnosidase, avenacinase, and an unspecified sugar transporter. Rhamnosidase has been previously identified in *Cryptococcus albidus* (35), *Aspergillus* spp. (36), and other basidiomycetes and ascomycetes (37). The presence of rhamnosidase may enable more efficient use of rhamnose as a carbohydrate source. Of particular note, rhamnose is a free sugar produced by *Rhamnus* (buckthorn), a plant with a large distribution in the PNW. Rhamnose is also found in members of the *Uncaria* family, a genus found throughout the tropical region. It is a component of many plant and bacterial polysaccharides, as well as plant glycosides. Additionally, rhamnosidase in soft-rot fungi has been shown to have lignocellulolytic activity, which is involved in the degradation of hardwoods (38). Its presence in the PNW population may also provide an adaptive advantage to old-growth forests of the region, as *C. gattii* been recovered from rotting and decaying wood in other geographic areas (39, 40).

The genes found in VGIIc, but not VGIIa or VGIIb (is-cDNA 34 to 41), relate primarily to sugar transport, metabolism, and gene expression and are commonly seen in other VGII subtypes and non-VGII cryptococcal genomes, suggesting that the lack of homologous genes in VGIIa and VGIIb represents independent gene loss events.

Among the three gene regions found only in the North American VGIIb strains is a haloacid dehalogenase-like hydrolase (is-cDNA 18). This gene was rather infrequent in other VGII subtypes, although a highly homologous region was seen in VGI, VGIII, and VGIV. A haloacid dehalogenase was found to be a virulence factor in *Paracoccidioides brasiliensis*, necessary for lung epithelial cell adhesion (41). No function for this gene region has been established in *Cryptococcus*.

The majority of all other between-PNW subtype genetic differences include genes or domains that are seen relatively frequently in other VGII subtypes. The presence or absence of these genes will need to be further explored to understand any functional role associated with the PNW populations.

The 12-gene cluster that was absent in the derived clade of VGIIb (cluster 1, which contains is-cDNA 1 to 12, in Fig. 7; see Fig. S4 in the supplemental material) was present in all other VGIIb isolates, including the separate clade found in North America, and was found, wholly or partially, in all other *Cryptococcus* isolates. This gene complement is composed of structural, regulatory, transport, and repair genes. The significance of this deletion remains unknown, but it is a marker of the derived lower-virulence clade of VGIIb. Phenotypic studies are necessary to determine whether there are other differences between the derived clade and the rest of the VGIIb subtype.

(ii) Evidence of recombination. Beyond the clear differences of gene content between and among VGII subtypes, multiple lines of evidence demonstrate the likelihood of "genome sharing" throughout the VGII population. In addition to the "all VGII" tree consistency index of ~0.4 (meaning a high degree of homoplasy) (see Fig. S2 in the supplemental material), and a significant phi test for recombination ($P = 0.0$), it is evident that gene content does not dictate phylogeny. We assume that recombination throughout the VGII population accounts for the high degree of homoplasy, which may be accomplished through larger genome content sharing events (i.e., bisexual or same-sex mating [19]). While the neighbor net phylogenetic network recovers a mostly star-like shape, there are apparent relationships between lineages (Fig. 2). The fineStructure analysis conducted here assigned lineages to specific "populations" and assessed levels of relatedness between genomes using adjacent linked SNPs to identify the nearest neighbor relationships for portions of the genome bounded by recombination events. Evidence of recombination was identified in South American isolates. For example, isolates WM 05.275 (originating in Colombia) and WM 05.529 (originating in Brazil) share many of their linked loci, although they are distinct subtypes and are separated by WGST, and they have distinct gene content by BSR. The same pattern is seen with isolates WM 05.419 and WM 05.547, both Brazilian clinical isolates. The VGIIa (R265 [British Columbia, Canada]) and VGIIb (B9563 [Washington]) representatives showed clear evidence of limited genome sharing with most isolates in the analysis. Conversely, the VGIIc genome showed almost no sharing with the vast majority of other VGII subtypes, except its nearest phylogenetic relatives from Brazil (samples WM 05.462 and IAL-3234). Taken together, these results indicate a relatively weak population structure within VGII or a recent subdivision of populations that have not yet allowed gene absence and presence to become fixed within recombining groups. Again, the high consistency indices for whole-genome SNP phylogenies of each PNW subtype (Fig. 3 to 5) indicate a lack of observable recombination occurring within these clones.

The presence of the three-gene cluster with a high degree of homology to *C. neoformans* var. *grubii* (cluster 6) provides evidence of a possible previous hybridization event between *C. gattii* and *C. neoformans* var. *grubii*, resulting in the possible introgression of this cluster in either a limited number of *C. gattii* subtypes or in the *C. neoformans* var. *grubii* genome. More whole-genome sequencing of the latter will be needed to ascertain the direction of introgression. Natural *C. gattii* and *C. neoformans* var. *grubii* hybrids have been recently identified from clinical samples in Colombia and Brazil (42). Additionally, genetic remnants of previous hybridization events, also referred to as "identity islands," have been reported in the case of a nonreciprocal exchange of a multigene cluster from *C. neoformans* var. *grubii* to *C. neoformans* var. *neoformans* (43). The existence of hybrids and subsequent introgression are an additional mechanism of genome evolution (beyond recombination, SNP mutation, and gene gain, loss, and duplication) and may allow continued successful emergence of novel phenotypes.

(iii) Conclusions. No clear progenitor strains have been identified for VGIIa, VGIIb, or VGIIc, although the first two have been found in South America. The phylogenetic evidence shown here is consistent with multiple translocations of VGIIb to the PNW and elsewhere, as well as possible multiple introductions of VGIIa into North America. The VGIIc subtype lacks representation outside

the PNW; however, its consistent SNP and gene content relationship with subtypes from Brazil provides compelling evidence of a likely South American origin for VGIIc as well. The results of the analysis also confirm that South America is a major source of VGII diversity and, with numerous subtypes and extensive recombination, it appears to represent an evolutionary accelerator for the VGII subtype and clones emerging from it.

Our extensive whole-population genome analysis enabled an *in silico* search for genetic factors associated with differing phenotypes in the PNW populations; similar *in silico* genome-wide association studies (GWAS) have been useful for studying adaptations in other widely dispersed fungi (34). The initial SNP and gene analyses based on this population-level study provide an important perspective on *C. gattii* diversity and will serve as a baseline for future studies. The identification of variant sugar transport and metabolism genes, as well as previously unexplored virulence genes (e.g., collagen adhesion gene in VGIIa), provides excellent candidates to pursue in functional studies that will help explain the PNW outbreaks and other *C. gattii* emergence events. The comprehensive SNP analyses will also be useful for future mutation-based functional analyses. Given the obvious gene content differences seen within VGIIb, even demonstrably clonal subtypes are not uniform and may require intrasubtype comparative analyses.

We recognize that limitations of this study include sampling bias and hence possible overrepresentation of isolates from some geographic regions and underrepresentation from others. There could be further diversity of VGII in other regions that were not sampled or that were sparsely sampled. Additionally, given the finding of *C. gattii* VGII strains in undisturbed Amazon rainforest and the amount of diversity already seen in South America (7), there is doubtless much more unsampled diversity on that continent. Additionally, this study was designed to identify important genomic features from whole-genome sequence data and was not designed to identify important mutations impacting gene regulation or expression. Nevertheless, our findings provide a significant advance in our understanding of the emergence of *C. gattii* in the PNW though the importance of the gene differences is unknown and requires further functional genomic analyses. As has been seen in recent years, *C. gattii* continues to spread to new locales, and a comprehensive understanding of its population genomics may help researchers, public health officials, and clinicians to better respond to the impacts of its continued evolution and emergence.

MATERIALS AND METHODS

***Cryptococcus gattii* isolates and Illumina sequencing.** One hundred eighteen *C. gattii* isolates were included in the analyses (Table 1). Information for clinical strains was collected with the approval and oversight of the institutional ethic committee. Additionally, the two near neighboring taxa, *C. neoformans* var. *neoformans* strain JEC21 and *C. neoformans* var. *grubii* strain H99, were obtained as whole-genome assemblies from NCBI (http://www.ncbi.nlm.nih.gov/Genbank/index.html) and the Broad Institute of Harvard and MIT (http://www.broadinstitute.org/annotation/genome/cryptococcus_neoformans_b), respectively. The 118 *C. gattii* isolates were selected based on previously established MLST sequence types (Table 1) to ensure global genomic diversity. The previously published VGIIa reference genome (R265) (see the Broad Institute link above) was resequenced for this analysis, due to the relatively high number of likely errors in the existing Sanger-based reference genome (17). The genomes of all 118 *C. gattii* isolates were sequenced using Illumina HiSeq, MiSeq, or

GAIIx sequencing technology as previously described (17, 44). High-molecular-weight DNA was extracted using the ZR Fungal/Bacterial DNA MiniPrep kit (catalog no. D6005; Zymo Research, Irvine, CA, USA). DNA samples were prepared for paired-end sequencing using the Kapa Biosystems library preparation kit (catalog no. KK8201; Woburn, MA, USA) protocol with an 8-bp index modification. Briefly, 2 μg of double-stranded DNA (dsDNA) sheared to an average size of 600 bp was used for the Kapa Illumina paired-end library preparation as described by the manufacturer. Modified oligonucleotides (Integrated DNA Technologies, Coralville, IA, USA) with 8-bp indexing capability (45), were substituted at the appropriate step. Prior to sequencing, the libraries were quantified with quantitative PCR (qPCR) on the ABI 7900HT (Life Technologies Corporation, Carlsbad, CA, USA) using the Kapa library quantification kit (catalog no. KK4835; Woburn, MA, USA). Libraries were sequenced to a read length of 100 bp or 150 bp on the Illumina HiSeq system, to 75 bp or 100 bp on the GAIIx system, or to 250 bp on the MiSeq system. All WGS data files have been deposited in the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/bioproject; project PRJNA244927).

SNP detection and phylogenetic comparison. (i) Genome assembly. All sequenced genomes were *de novo* assembled using the SPAdes assembler v2.5.0 (46). Raw contigs were filtered by length and coverage, as SPAdes tends to produce many short and low-coverage contigs. For isolates assembled from 100-bp or 150-bp paired-end data, contigs shorter than 200 bp were filtered and removed, while contigs shorter than 500 bp were filtered/removed in genomes assembled from 250-bp paired-end MiSeq data. Additionally, for every genome, contigs with coverage below 20% of the average coverage of the 20 largest contigs were filtered out.

(ii) MLST typing. The seven MLST loci, *CAP59*, *GPD1*, *LAC1*, *SOD1*, *URA5*, *PLB1*, and intergene sequence (IGS), were aligned to the contigs according to the consensus MLST scheme for the *C. neoformans*/*C. gattii* species complex (10). Allele types (ATs) and sequence types (STs) were assigned according to the MLST database at http://mlst.mycologylab.org, and new ATs and STs were added to this database. A maximum likelihood tree was constructed based on ST locus scores using the program MEGA v. 5.05 (47).

(iii) SNP variant detection. Illumina read data were aligned against the herein newly obtained R265 (high virulent VGIIa Vancouver Island outbreak strain) assembly using Novoalign 3.00.03 (Novocraft Technologies, Selangor, Malaysia), and then SNP variants were identified using the GATK Unified Genotyper v2.4 (48). SNP calls were then filtered (http://tgennorth.github.io/NASP/) to remove positions with less than 10× coverage, with less than 80% variant allele calls, or identified by Nucmer as being within duplicated regions in the reference. In the cases of the VGIIb and VGIIc analyses, the read data were aligned by the same methods using the genome assemblies for B9563 and B8825, respectively. In total, five SNP matrices with different taxa were produced: (i) *C. gattii* molecular types VGI to VGIV; (ii) *C. gattii* molecular type VGII only with all previously established MLST sequence types; (iii) *C. gattii* VGIIa; (iv) *C. gattii* VGIIb; and (v) *C. gattii* VGIIc.

(iv) Phylogenetic analysis. To understand relationships between isolates, we conducted whole-genome SNP typing (WGST) as previously described (17, 39, 49). Maximum-parsimony SNP trees, based on each of the above five SNP matrices, were constructed, using PAUP* v.4.0b10 (50) and visualized in FigTree v.1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). The *C. gattii* VGII tree was rooted using the large *C. gattii* tree, while the VGIIa, VGIIb, and VGIIc trees were each rooted using the other VGII subtypes as outgroups. Roots are shown simply as the midpoint on the branch determined by outgroup rooting, because the exact number of SNPs assigned to each branch varies slightly with the inclusion of different taxa, due to the SNP filtering strategies employed. The phi test for recombination was conducted on the whole VGII SNP data set, as well as the same data set broken into 500-bp nonoverlapping windows, using PhiPack software (51). Due to recombination signal strongly present in the VGII SNP data, a neighbor-joining split tree network or "neighbor net" (52) was drawn to visualize the relationship between isolates in the

VGII tree. The neighbor net was drawn using SplitsTree4 (53) with the uncorrected P-distance transformation. The GTR+G model of nucleotide evolution, selected by jModelTest2 (54), was also used to calculate distances for input to the neighbor net algorithm, which produced a network showing the same divisions among samples (data not shown).

(v) ChromoPainter and fineStructure population analysis. In addition to whole-genome SNP phylogenies, fineStructure analysis (25) was performed using SNP matrix 2 (see "SNP variant detection" above) to help identify the population structure within VGII. Using one representative from each of the 32 distinct VGII subtypes (i.e., MLST sequence types) (Table 1), visualized by placement in the SNP phylogeny, whole-genome SNP data were reduced to a pairwise similarity matrix using ChromoPainter, which estimates the number of regions in two genomes that coalesce prior to coalescing with another genome. ChromoPainter uses the linkage disequilibrium present between closely positioned SNP loci to help assign a given region of a genome to its best near neighbor, increasing the sensitivity of population detection over unlinked methods. ChromoPainter was run using the linkage model, assuming uniform rates of recombination per base pair of sequence. Next, the population structure is identified by fineStructure using this similarity matrix, where the underlying model assumes individuals within populations will share more regions of their genome with each other, and have a similar amount of admixture with individuals from different populations. After identifying populations, fineStructure then merges populations one at a time, at each step using the most likely population merge, to relate populations to each other through a tree.

Gene content comparison. (i) BSR analysis of gene content. Potential gene content comparison of taxa was performed using BLAT score ratio (BSR) analysis, as previously described for prokaryotic genomes (55), with modifications made to accommodate eukaryotic genes containing introns (described below). Gene sequences were predicted for each assembled genome using the *ab initio* eukaryotic gene prediction software Augustus v2.6.1 (56). For each gene prediction, the coding DNA region sequences were concatenated into a single contig, removing introns to create the putative *in silico* coding DNA (is-cDNA). is-cDNA sequences from all isolates were then clustered and dereplicated using USEARCH v5.0.144 (57) with a 90% identity cutoff, creating a nonredundant set of consensus is-cDNA predictions encompassing the entire diversity of samples. Putative is-cDNAs were then aligned back to each genome assembly using BLAT v35x1 (58), which can align the cDNA sequence against a reference allowing for intervening intron sequences, using an 80% identity cutoff. BLAT alignments were converted to scores replicating the calculation used by the web-based BLAT server (https://genome.ucsc.edu/). For each putative is-cDNA, the raw score for an isolate was converted into a ratio by dividing the score of a particular isolate by the top score achieved by any isolate, as previously described (55). These BSRs were then imported into the MultiExperiment Viewer (MEV) to generate a heat map which could be manually inspected to identify genes present and absent across different sets of taxa (Fig. 7). Thus, a value of 1.0 (yellow on the heat map) represents the best hit, and a value of 0 (blue) represents no alignment by BLAT.

Using the BSR method described above, the potential gene content between VGIIa, VGIIb, and VGIIc sequence types was compared using representative isolates from each subtype present in the PNW (Table 2). Representative strains were selected based on assembly quality and being part of the PNW outbreak subtypes (i.e., isolates typed as VGIIa and VGIIb but not from the PNW were not included in these sets; all VGIIc isolates were from the PNW). This set included 9 VGIIa isolates, 6 VGIIb isolates, and 12 VGIIc isolates (Table 1). Putative consensus is-cDNAs showing differential content across the three PNW sequence types were identified using a cutoff BSR score of above 0.9 in the taxa of interest and below 0.9 in the excluded taxa. Multiple assemblies from each subtype were included to help reduce the number of genes that may appear to be absent due to variability of draft genome quality with multiple contigs per chromosome. Genes with a low BSR value in a single member of a subtype

almost invariably were due to contig breaks in assembly. Unique putative genes were identified for each of the VGIIa, VGIIb, and VGIIc groups as well as for genes contained in both VGIIa and VGIIc but missing in VGIIb, as this comparison may help explain the previously described phenotype differences seen in the PNW.

Expanding the analysis beyond differences within the PNW sequence types, the putative is-cDNA sequences that define the PNW subtypes were scored in each of the additional 29 global VGII subtypes (Table 1 and Fig. 7), a representative sequence from each of the VGI, VGIII, and VGIV major molecular types, as well as *C. neoformans* var. *grubii* (H99) and *C. neoformans* var. *neoformans* (JEC21). This allowed for the gene differences found in the PNW subtypes to be examined across the diversity of the *Cryptococcus neoformans/C. gattii* species complex, as well as for the identification of genes present in all of the PNW isolates but not universally present in other VGII subtypes. In addition, the full complement of is-cDNAs was scored in all 29 global VGII subtypes.

To characterize genes absent in the PNW populations but present in many of the other global VGII subtypes, a second set of putative genes (is-cDNAs) was generated using the gene predictions in non-PNW VGII subtypes. While this would presumably contain many of the same genes found in the PNW gene set, it would also contain any genes that may be absent in all of the PNW subtypes. Using the same cutoff scheme as described above, putative genes were identified that were missing or truncated in all of the PNW subtypes but present in at least half of the other VGII subtypes. is-cDNA sequences identified as having a differential content pattern of interest were then located in the genome assemblies of strains R265 (VGIIa), B9563 (VGIIb), B8825 (VGIIc), depending on which taxa contained the is-cDNAs, to look for blocks of adjacent genes showing similar patterns of presence and absence.

(ii) BSR confirmation with sequence read data alignment. Because a low BSR score may indicate gene loss or truncation as well as problems caused by assembly issues in draft genomes, putative gene sequences identified as differential across taxa of interest were subsequently assessed by raw read data alignment. This provides evidence for confirmation of the differential gene content and, in the case of scores between 0.0 and 0.9, helps to identify the type of gene structure differences present between isolates with different scores. To avoid potential problems in aligning genomic read data against putative is-cDNA sequences, which do not contain introns, full-length gene predictions containing introns were dereplicated with USEARCH to serve as read alignment references. Putative is-cDNAs of interest were aligned against the full-length gene sequences using BLAT to identify the corresponding full-length gene. Read data from each isolate in the PNW analysis was then aligned to the full-length gene. These alignments were inspected for each gene to ensure read coverage across the entire gene sequence for isolates with high BSR scores and to ensure a lack of read coverage for isolates with BSR scores indicating gene absence. Moderate BSR scores occurred in a few cases when limited read data successfully aligned to a putatively missing gene (e.g., is-cDNA 9 in VGIIb), or when read data aligned to truncated genes. In an effort to be conservative in identifying differential gene content using short-read data, gene differences that could not be confirmed with read data are not reported. This most often occurred in the case of putative genes that contained highly repetitive sequences, which cause assembly and definitive alignment issues in short-read data.

(iii) Putative gene characterization. is-cDNAs, which were confirmed as variable with read data (see above), were translated into amino acid sequences and then searched against the Pfam database (http://pfam.sanger.ac.uk/) to help identify potential protein functions. Protein families and domains achieving a significant score are reported in Table 2. Additional characterization for selected is-cDNAs was performed using a blastp search of the NCBI nonredundant protein database (http://www.ncbi.nlm.nih.gov/RefSeq/).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.01464-14/-/DCSupplemental.

Figure S1, EPS file, 0.9 MB.
Figure S2, PDF file, 1.7 MB.
Figure S3, EPS file, 0.3 MB.
Figure S4, EPS file, 1 MB.

ACKNOWLEDGMENTS

REFERENCES

1. **Kwon-Chung KJ, Bennett JE.** 1984. High prevalence of *Cryptococcus neoformans* var. *gattii* in tropical and subtropical regions. Zentralbl. Bakteriol. Mikrobiol. Hyg. A **257**:213–218.

2. **Lazera MS, Salmito Cavalcanti MA, Londero AT, Trilles L, Nishikawa MM, Wanke B.** 2000. Possible primary ecological niche of *Cryptococcus neoformans*. Med. Mycol. **38**:379–383. http://dx.doi.org/10.1080/mmy.38.5.379.383.

3. **Escandón P, Sánchez A, Martínez M, Meyer W, Castañeda E.** 2006. Molecular epidemiology of clinical and environmental isolates of the *Cryptococcus neoformans* species complex reveals a high genetic diversity and the presence of the molecular type VGII mating type a in Colombia. FEMS Yeast Res **6**:625–635. http://dx.doi.org/10.1111/j.1567-1364.2006.00055.x.

4. **Firacative C, Torres G, Rodríguez MC, Escandón P.** 2011. First environmental isolation of *Cryptococcus gattii* serotype B, from Cúcuta, Colombia. Biomedica **31**:118–123. http://dx.doi.org/10.1590/S0120-41572011000100014.

5. **Ellis DH, Pfeiffer TJ.** 1990. Natural habitat of *Cryptococcus neoformans* var. *gattii*. J. Clin. Microbiol **28**:1642–1644.

6. **Carriconde F, Gilgado F, Arthur I, Ellis D, Malik R, van de Wiele N, Robert V, Currie BJ, Meyer W.** 2011. Clonality and alpha-a recombination in the Australian *Cryptococcus gattii* VGII population—an emerging outbreak in Australia. PLoS One **6**:. http://dx.doi.org/10.1371/journal.pone.0016936e16936.

7. **Fortes ST, Lazéra MS, Nishikawa MM, Macedo RC, Wanke B.** 2001. First isolation of *Cryptococcus neoformans* var. *gattii* from a native jungle tree in the Brazilian Amazon rainforest. Mycoses **44**:137–140. http://dx.doi.org/10.1046/j.1439-0507.2001.00651.x.

8. **Baltazar LDM, Ribeiro MA.** 2008. First isolation of *Cryptococcus gattii* from the environment in the State of Espírito Santo. Rev. Soc. Bras. Med. Trop. **41**:449–453. (In Portuguese.) http://dx.doi.org/10.1590/S0037-86822008000500003.

9. **Hagen F, Ceresini PC, Polacheck I, Ma H, van Nieuwerburgh F, Gabaldón T, Kagan S, Pursall ER, Hoogveld HL, van Iersel LJ, Klau GW, Kelk SM, Stougie L, Bartlett KH, Voelz K, Pryszcz LP, Castañeda E, Lazera M, Meyer W, Deforce D, Meis JF, May RC, Klaassen CH, Boekhout T.** 2013. Ancient dispersal of the human fungal pathogen *Cryptococcus gattii* from the Amazon rainforest. PLoS One **8**:e71148. http://dx.doi.org/10.1371/journal.pone.0071148.

10. **Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR, Esposto MC, Fisher M, Gilgado F, Hagen F, Kaocharoen S, Litvintseva AP, Mitchell TG, Simwami SP, Trilles L, Viviani MA, Kwon-Chung J.** 2009. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med. Mycol. **47**:561–570. http://dx.doi.org/10.1080/13693780902953886.

11. **Kidd SE, Hagen F, Tscharke RL, Huynh M, Bartlett KH, Fyfe M, Macdougall L, Boekhout T, Kwon-Chung KJ, Meyer W.** 2004. A rare genotype of *Cryptococcus gattii* caused the cryptococcosis outbreak on Vancouver Island (British Columbia, Canada). Proc. Natl. Acad. Sci. U. S. A. **101**:17258–17263. http://dx.doi.org/10.1073/pnas.0402981101.

12. **Fraser JA, Subaran RL, Nichols CB, Heitman J.** 2003. Recapitulation of the sexual cycle of the primary fungal pathogen *Cryptococcus neoformans* var. *gattii*: implications for an outbreak on Vancouver Island, Canada. Eukaryot. Cell **2**:1036–1045. http://dx.doi.org/10.1128/EC.2.5.1036-1045.2003.

13. **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.

14. **MacDougall L, Fyfe M.** 2006. Emergence of *Cryptococcus gattii* in a novel environment provides clues to its incubation period. J. Clin. Microbiol. **44**:1851–1852. http://dx.doi.org/10.1128/JCM.44.5.1851-1852.2006.

15. **Byrnes EJ, III, Bildfell RJ, Frank SA, Mitchell TG, Marr KA, Heitman J.** 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**:1081–1086. http://dx.doi.org/10.1086/597306.

16. **Byrnes EJ, III, Li W, Lewit Y, Ma H, Voelz K, Ren P, Carter DA, Chaturvedi V, Bildfell RJ, May RC, Heitman J.** 2010. Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the northwest United States. PLoS Pathog. **6**:e1000850. http://dx.doi.org/10.1371/journal.ppat.1000850.

17. **Gillece 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.

18. **Harris JR, Lockhart SR, Debess E, Marsden-Haug N, Goldoft M, Wohrle R, Lee S, Smelser C, Park B, Chiller T.** 2011. *Cryptococcus gattii* in the United States: clinical aspects of infection with an emerging pathogen. Clin. Infect. Dis. **53**:1188–1195. http://dx.doi.org/10.1093/cid/cir723.

19. **Springer DJ, Phadke S, Billmyre B, Heitman J.** 2012. *Cryptococcus gattii*, no longer an accidental pathogen? Curr. Fungal Infect. Rep. **6**:245–256. http://dx.doi.org/10.1007/s12281-012-0111-0.

20. **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.

21. **Ngamskulrungroj P, Serena C, Gilgado F, Malik R, Meyer W.** 2011. Global VGIIa isolates are of comparable virulence to the major fatal *Cryptococcus gattii* Vancouver Island outbreak genotype. Clin. Microbiol. Infect. **17**:251–258. http://dx.doi.org/10.1111/j.1469-0691.2010.03222.x.

22. **Moran GP, Coleman DC, Sullivan DJ.** 2011. Comparative genomics and the evolution of pathogenicity in human pathogenic fungi. Eukaryot. Cell **10**:34–42. http://dx.doi.org/10.1128/EC.00242-10.

23. **Ngamskulrungroj P, Price J, Sorrell T, Perfect JR, Meyer W.** 2011. *Cryptococcus gattii* virulence composite: candidate genes revealed by microarray analysis of high and less virulent Vancouver Island outbreak strains. PLoS One **6**:e16076. http://dx.doi.org/10.1371/journal.pone.0016076.

24. **Kaocharoen S, Ngamskulrungroj P, Firacative C, Trilles L, Piyabongkarn D, Banlunara W, Poonwan N, Chaiprasert A, Meyer W, Chindamporn A.** 2013. Molecular epidemiology reveals genetic diversity amongst isolates of the *Cryptococcus neoformans*/*C. gattii* species complex in Thailand. PLoS Negl. Trop. Dis. **7**:e2297. http://dx.doi.org/10.1371/journal.pntd.0002297.

25. **Lawson DJ, Hellenthal G, Myers S, Falush D.** 2012. Inference of population structure using dense haplotype data. PLoS Genet. **8**:e1002453. http://dx.doi.org/10.1371/journal.pgen.1002453.

26. **Ngamskulrungroj 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.

27. **Phadke SS, Feretzaki M, Heitman J.** 2013. Unisexual reproduction enhances fungal competitiveness by promoting habitat exploration by hyphal growth and sporulation. Eukaryot. Cell **12**:1155–1159. http://dx.doi.org/10.1128/EC.00147-13.

28. **Tscharke RL, Lazera M, Chang YC, Wickes BL, Kwon-Chung KJ.** 2003. Haploid fruiting in *Cryptococcus neoformans* is not mating type alpha-specific. Fungal Genet. Biol. **39**:230–237. http://dx.doi.org/10.1016/S1087-1845(03)00046-X.

29. **Nishikaku AS, Ribeiro LC, Molina RF, Albe BP, Cunha CDS, Burger E.** 2009. Matrix metalloproteinases with gelatinolytic activity induced by *Paracoccidioides brasiliensis* infection. Int. J. Exp. Pathol. **90**:527–537. http://dx.doi.org/10.1111/j.1365-2613.2009.00663.x.

30. **Laurent GJ.** 1986. Lung collagen: more than scaffolding. Thorax **41:** 418–428. http://dx.doi.org/10.1136/thx.41.6.418.

31. **Bromley IM, Donaldson K.** 1996. Binding of *Aspergillus fumigatus* spores to lung epithelial cells and basement membrane proteins: relevance to the asthmatic lung. Thorax **51:**1203–1209. http://dx.doi.org/10.1136/thx.51.12.1203.

32. **Wagner C, Khan AS, Kamphausen T, Schmausser B, Unal C, Lorenz U, Fischer G, Hacker J, Steinert M.** 2007. Collagen binding protein Mip enables *Legionella pneumophila* to transmigrate through a barrier of NCI-H292 lung epithelial cells and extracellular matrix. Cell. Microbiol. **9:**450–462. http://dx.doi.org/10.1111/j.1462-5822.2006.00802.x.

33. **Panepinto J, Liu L, Ramos J, Zhu X, Valyi-Nagy T, Eksi S, Fu J, Jaffe HA, Wickes B, Williamson PR.** 2005. The DEAD-box RNA helicase Vad1 regulates multiple virulence-associated genes in *Cryptococcus neoformans*. J. Clin. Invest. **115:**632–641. http://dx.doi.org/10.1172/JCI23048.

34. **Ellison CE, Hall C, Kowbel D, Welch J, Brem RB, Glass NL, Taylor JW.** 2011. Population genomics and local adaptation in wild isolates of a model microbial eukaryote. Proc. Natl. Acad. Sci. U. S. A. **108:**2831–2836. http://dx.doi.org/10.1073/pnas.1014971108.

35. **Rzaieva OM,** Varbanets' LD, Hudzenko OV. 2011. Optimization of cultivation conditions of *Cryptococcus albidus*—producers of alpha-l-rhamnosidase. Mikrobiol. Z. 73:10–16. (In Ukrainian.) PubMed. PubMed.

36. **Liu T, Yu H, Zhang C, Lu M, Piao Y, Ohba M, Tang M, Yuan X, Wei S, Wang K, Ma A, Feng X, Qin S, Mukai C, Tsuji A, Jin F.** 2012. *Aspergillus niger* DLFCC-90 rhamnoside hydrolase, a new type of flavonoid glycoside hydrolase. Appl. Environ. Microbiol. **78:**4752–4754. http://dx.doi.org/10.1128/AEM.00054-12.

37. **Yadav V, Yadav PK, Yadav S, Yadav KDS.** 2010. α-L-Rhamnosidase: a review. Proc. Biochem. **45:**1226–1235. http://dx.doi.org/10.1016/j.procbio.2010.05.025.

38. **Nghi DH, Bittner B, Kellner H, Jehmlich N, Ullrich R, Pecyna MJ, Nousiainen P, Sipilä J, Houng LM, Hofrichter M, Liers C.** 2012. The wood rot ascomycete *Xylaria polymorpha* produces a novel GH78 glycoside hydrolase that exhibits α-L-rhamnosidase and feruloyl esterase activities and releases hydroxycinnamic acids from lignocelluloses. Appl. Environ. Microbiol. **78:**4893–4901. http://dx.doi.org/10.1128/AEM.07588-11.

39. **Lazéra MS, Cavalcanti MA, Trilles L, Nishikawa MM, Wanke B.** 1998. *Cryptococcus neoformans* var. *gattii*— evidence for a natural habitat related to decaying wood in a pottery tree hollow. Med. Mycol. **36:**119–122. http://dx.doi.org/10.1046/j.1365-280X.1998.00120.x.

40. **Grover N, Nawange SR, Naidu J, Singh SM, Sharma A.** 2007. Ecological niche of *Cryptococcus neoformans* var. *grubii* and *Cryptococcus gattii* in decaying wood of trunk hollows of living trees in Jabalpur City of Central India. Mycopathologia **164:**159–170. http://dx.doi.org/10.1007/s11046-007-9039-2.

41. **Hernández O, Almeida AJ, Gonzalez A, Garcia AM, Tamayo D, Cano LE, Restrepo A, McEwen JG.** 2010. A 32-kilodalton hydrolase plays an important role in *Paracoccidioides brasiliensis* adherence to host cells and influences pathogenicity. Infect. Immun. **78:**5280–5286. http://dx.doi.org/10.1128/IAI.00692-10.

42. **Aminnejad M, Diaz M, Arabatzis M, Castañeda E, Lazera M, Velegraki A, Marriott D, Sorrell TC, Meyer W.** 2012. Identification of novel hybrids between *Cryptococcus neoformans* var. *grubii* VNI and *Cryptococcus gattii* VGII. Mycopathologia **173:**337–346. http://dx.doi.org/10.1007/s11046-011-9491-x.

43. **Kavanaugh LA, Fraser JA, Dietrich FS.** 2006. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. Mol. Biol. Evol. **23:**1879–1890. http://dx.doi.org/10.1093/molbev/msl070.

44. **Etienne KA, Gillece J, Hilsabeck R, Schupp JM, Colman R, Lockhart SR, Gade L, Thompson EH, Sutton DA, Neblett-Fanfair R, Park BJ, Turabelidze G, Keim P, Brandt ME, Deak E, Engelthaler DM.** 2012. Whole genome sequence typing to investigate the *Apophysomyces* outbreak following a tornado in Joplin, Missouri, 2011. PLoS One 7:e49989. http://dx.doi.org/10.1371/journal.pone.0049989.

45. **Kozarewa I, Turner DJ.** 2011. 96-plex molecular bar coding for the Illumina Genome Analyzer. Methods Mol. Biol. **733:**279–298. http://dx.doi.org/10.1007/978-1-61779-089-8_20.

46. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. **19:**455–477. http://dx.doi.org/10.1089/cmb.2012.0021.

47. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28:**2731–2739. http://dx.doi.org/10.1093/molbev/msr121.

48. **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ.** 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. **43:**491–498. http://dx.doi.org/10.1038/ng.806.

49. **Engelthaler DM, Kelley E, Driebe EM, Bowers J, Eberhard CF, Trujillo J, Decruyenaere F, Schupp JM, Mossong J, Keim P, Even J.** 2013. Rapid and robust phylotyping of spa t003, a dominant MRSA clone in Luxembourg and other European countries. BMC Infect. Dis. **13:**339. http://dx.doi.org/10.1186/1471-2334-13-339.

50. **Swofford DL.** 2003. PAUP*: Phylogenetic Analysis Using Parsimony (and other methods), version 4. Sinauer Associates, Sunderland, MA.

51. **Bruen TC, Philippe H, Bryant D.** 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics **172:**2665–2681. http://dx.doi.org/10.1534/genetics.105.048975.

52. **Bryant D, Moulton V.** 2004. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol. Biol. Evol. **21:**255–265. http://dx.doi.org/10.1093/molbev/msh018.

53. **Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. **23:**254–267.

54. **Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods **9:**772. http://dx.doi.org/10.1038/nmeth.2111.

55. **Sahl JW, Gillece JD, Schupp JM, Waddell V, Driebe EM, Engelthaler DM, Keim P.** 2013. Evolution of a pathogen: a comparative genomics analysis identifies a genetic pathway to pathogenesis in *Acinetobacter*. PLoS One 8:e54287. http://dx.doi.org/10.1371/journal.pone.0054287.

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

57. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:**2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461.

58. **Kent WJ.** 2002. BLAT—the BLAST-like alignment tool. Genome Res. **12:**656–664. http://dx.doi.org/10.1101/gr.229202.