

Whole Genome Sequence Analysis of *Cryptococcus gattii* from the Pacific Northwest Reveals Unexpected Diversity

John D. Gillece¹*, James M. Schupp^{1*}, S. Arunmozhi Balajee², Julie Harris², Talima Pearson³, Yongpan Yan⁴, Paul Keim³, Emilio DeBess⁵, Nicola Marsden-Haug⁶, Ron Wohrle⁷, David M. Engelthaler¹, Shawn R. Lockhart²

1 Pathogen Genomics Division, Translational Genomics Research Institute, Flagstaff, Arizona, United States of America, **2** Mycotic Diseases Branch, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America, **3** Center for Microbial Genetics and Genomics, Northern Arizona University, Flagstaff, Arizona, United States of America, **4** Center for Genetic Epidemiology and Prevention, Van Andel Institute, Grand Rapids, Michigan, United States of America, **5** Oregon State Department of Human Services, Portland, Oregon, United States of America, **6** Washington State Department of Health, Shoreline, Washington, United States of America, **7** Washington State Department of Health, Tumwater, Washington, United States of America

Abstract

A recent emergence of *Cryptococcus gattii* in the Pacific Northwest involves strains that fall into three primarily clonal molecular subtypes: VGIIa, VGIIb and VGIIc. Multilocus sequence typing (MLST) and variable number tandem repeat analysis appear to identify little diversity within these molecular subtypes. Given the apparent expansion of these subtypes into new geographic areas and their ability to cause disease in immunocompetent individuals, differentiation of isolates belonging to these subtypes could be very important from a public health perspective. We used whole genome sequence typing (WGST) to perform fine-scale phylogenetic analysis on 20 *C. gattii* isolates, 18 of which are from the VGII molecular type largely responsible for the Pacific Northwest emergence. Analysis both including and excluding (289,586 SNPs and 56,845 SNPs, respectively) molecular types VGI and VGIII isolates resulted in phylogenetic reconstructions consistent, for the most part, with MLST analysis but with far greater resolution among isolates. The WGST analysis presented here resulted in identification of over 100 SNPs among eight VGIIc isolates as well as unique genotypes for each of the VGIIa, VGIIb and VGIIc isolates. Similar levels of genetic diversity were found within each of the molecular subtype isolates, despite the fact that the VGIIb clade is thought to have emerged much earlier. The analysis presented here is the first multi-genome WGST study to focus on the *C. gattii* molecular subtypes involved in the Pacific Northwest emergence and describes the tools that will further our understanding of this emerging pathogen.

Citation: Gillece JD, Schupp JM, Balajee SA, Harris J, Pearson T, et al. (2011) Whole Genome Sequence Analysis of *Cryptococcus gattii* from the Pacific Northwest Reveals Unexpected Diversity. PLoS ONE 6(12): e28550. doi:10.1371/journal.pone.0028550

Editor: Dee A. Carter, University of Sydney, Australia

Received: April 19, 2011; **Accepted:** November 10, 2011; **Published:** December 7, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: All work was funded by the Translational Genomics Research Institute and the Centers for Disease Control and Prevention. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: jschupp@tgen.org

† These authors contributed equally to this work.

Introduction

Cryptococcus gattii is an emerging cause of community-acquired fungal pneumonia, meningitis, and disseminated disease. In the past decade, infections caused by *C. gattii* have been increasingly identified outside of the previously recognized tropical and subtropical endemic regions and have emerged in temperate regions, including British Columbia (BC), the US Pacific Northwest (PNW), and Japan [1–4].

Emergence of *C. gattii* in BC and the PNW is remarkable for a number of reasons, including the adaptation of the pathogen to a climate previously unrecognized as potentially hospitable, an apparent increased virulence of certain molecular subtypes in mouse models [5], and the identification of novel clonal molecular subtypes [6–8]. Veterinary cases were first identified in 2000 on Vancouver Island, BC [9], and subsequent surveillance identified human cases beginning in 1999 on Vancouver Island and in 2004 in the US PNW [2,3]. By 2011, cases from this emergence had been identified on Vancouver

Island, the BC mainland, and in the states of Washington, Oregon, Idaho and California in the US.

Strains of *C. gattii* have been differentiated through the use of restriction fragment length polymorphism (RFLP), amplified fragment length polymorphism (AFLP), and multi-locus sequence typing (MLST) into four main molecular types: VGI, VGII, VGIII and VGIV. Further work using MLST has determined that these groups are monophyletic and genotypically distinct enough to deserve varietal status [10,11]. While there is genetic diversity within each of the *C. gattii* molecular types, a unique feature of the *C. gattii* emergence in BC and the PNW is the distinct clonality of these isolates. The predominant *C. gattii* molecular type VGII sorts into three closely related clonal subtypes designated VGIIa, VGIIb, and VGIIc [6,7,12,13]. The major subtype in BC was designated VGIIa because it was the predominant subtype and it had not been previously recognized from global collections [6,7]. The minor subtype in BC was designated VGIIb, and has been previously identified from other parts of the world, which led to speculation that it may be the progenitor of subtype VGIIa [7,14]. The strains identified in Oregon and Washington include VGIIa,

VGIIb and a unique VGII subtype not identified in BC and has since been designated VGIIc [12]. VGIIc shares a number of MLST alleles with both VGIIa and VGIIb and has novel MLST alleles as well [12,13]. These novel VGIIc strains were highly virulent in a murine pulmonary model of infection and had similar virulence features to VGIIa strains studied from Vancouver Island [5]. Isolates belonging to the VGIIc subtype group have not been identified outside of the PNW.

While MLST has been able to distinguish adequately between the VGIIa, VGIIb and VGIIc subtypes within the BC and PNW emergence, neither MLST nor variable number tandem repeat analysis (VNTR) have been successful at distinguishing among isolates within each clonal group. To date, only one single nucleotide polymorphism (SNP) has been identified within the VGIIc clonal group, and one VNTR difference has been identified within the VGIIa clonal group using these methods [5,13]. Like most typing systems, MLST and VNTR are limited by the small portion of the genome that is surveyed. Whole Genome Sequence Typing (WGST) is a methodology that maximizes the data available for inference of genetic diversity and has been successfully used to distinguish among highly related isolates of the fungus *Coccidioides immitis* [15]. Here we describe the WGST analysis of *C. gattii* using isolates from the PNW emergence. This next generation typing methodology allowed us to generate a unique genetic fingerprint for each study isolate, effectively subtyping a “clonal” population.

Materials and Methods

Isolates and DNA extraction

C. gattii isolates were received at the Centers for Disease Control and Prevention from state and local health departments, clinicians, veterinarians and ongoing environmental studies (Table 1) in the

PNW US and BC. Isolates were grown on YPD plus 0.5% NaCl and DNA was prepared using an UltraClean DNA Isolation Kit as described by the manufacturer (MO BIO Laboratories, Carlsbad, CA). Twenty *C. gattii* isolates representing three of the four previously described molecular types (VGI, VGII, and VGIII) and two published *C. gattii* genomes [16] were used in the analysis. The majority of the isolates (n=18) were known to belong to the VGIIa (6 isolates), VGIIb (4 isolates) or VGIIc (8 isolates) molecular subtypes, with an additional one each from the VGI and VGIII types. All isolates were from the PNW and BC emergence.

MLST analysis

MLST gene fragments were amplified using the seven consensus primers of Meyer et al. [17] as previously described [13]. All loci were sequenced using BigDye terminator technology (ABI, Foster City, CA) with an ABI Prism 3730 DNA sequencer in both forward and reverse directions with the same primers as those used for the PCR reactions.

Whole Genome Sequencing

The DNA samples were prepared for multiplexed, paired-end sequencing on the Illumina GAIIX Genome Analyzer (Illumina, Inc, San Diego, CA). For each isolate, 1–5 µg dsDNA in 200 µl was sheared in a 96-well plate with the SonicMAN™ (Part # SCM1000-3, Matrical Bioscience, Spokane, WA) to a size range of 200–1000 base pairs with the majority of material at ca. 600 base pairs using the following parameters: Pre Chill - 0°C for 75 sec; Cycles - 20; Sonication - 10 sec; Power - 100%; Lid Chill - 0°C for 75 sec; Plate Chill - 0°C for 10 sec; Post Chill - 0°C for 75 sec. The sheared DNA was purified using the QIAquick PCR Purification kit (Cat #28106, Qiagen, Valencia, CA). The

Table 1.

MLST Type	Strain ID	Geographic Source	Clinical Source	Collection Date	Reference Genome covered at $\geq 10\times$ (%)	Average Read Depth
IIa	B7395	Washington	Dog	2008	96.79	39
	B7467	Oregon	Porpoise	2009	96.69	39
	B8849	Oregon	Environmental	2010	96.93	57
	B8577	British Columbia	Environmental	2009	96.71	42
IIb	B7422	Oregon	Cat	2009	97.41	94
	B7436	N. California	Alpaca	2009	97.47	108
	B7394	Washington	Cat	2008	86.65	18
	B7735	Oregon	Human	2009	87.00	19
	B8554	Oregon	Dog	2008	95.82	70
	B8828	Washington	Porpoise	2010	93.72	34
IIc	B8571	Washington	Human	2009	95.06	46
	B8843	Oregon	Human	2010	95.26	58
	B8838	Washington	Human	2010	95.19	55
	B7466	Oregon	Cat	2008	94.76	42
	B7737	Oregon	Human	2009	94.74	39
	B6863	Oregon	Human	2005	94.09	32
	B7390	Idaho	Human	2008	93.69	30
	B7432	Oregon	Human	2009	92.90	25
I	B7488	Oregon	Human	2009	72.71	33
III	B8212	Oregon	Human	2007	64.25	16

doi:10.1371/journal.pone.0028550.t001

enzymatic processing (end-repair, phosphorylation, A-tailing, and adaptor ligation) of the DNA followed the guidelines as described in the Illumina protocol (“Preparing Samples for Multiplexed Paired-End Sequencing”, Catalog #PE-930-1002, Part #1005361). The enzymes for processing were obtained from New England Biolabs (Cat #E6000L, New England Biolabs, Ipswich, MA) and the oligonucleotides and adaptors were obtained from Illumina (Cat #PE-400-1001). After ligation of the adaptors, the DNA was run on a 2% agarose gel for 2 hours, after which a gel slice containing 500–600 bp fragments of each DNA sample was isolated and purified using the QIAquick Gel Extraction kit (Cat #28706, Qiagen, Valencia, CA). Individual libraries were quantified with qPCR on the ABI 7900HT (Part #4329001, Life Technologies Corporation, Carlsbad, CA) in triplicate at two concentrations, 1:1000 and 1:2000, using the Kapa Library Quantification Kit (part # KK4832 or KK4835, Kapa Biosystems, Woburn, MA). Based on the individual library concentrations, equimolar pools of no more than 5 indexed *Cryptococcus* libraries were prepared at a concentration of at least 1 nM using 10 mM Tris-HCl, pH 8.0+0.05% Tween 20 as the diluent. To ensure accurate loading onto the flowcell, the same quantification method was used to quantify the final pools. The pooled libraries were sequenced on the Illumina GAIIX. A 75 bp read paired end run was used for the first 2 isolates sequenced; the remaining isolates were done with a 101 bp read paired end run.

MLST and WGST Sequence Analysis

The seven MLST gene fragment sequences were concatenated, aligned, and subjected to maximum parsimony analysis using MEGA4 [18].

Illumina WGS data sets were aligned against the *Cryptococcus neoformans* var *gattii* R265 genome [19] (molecular subtype VGIIa) using the short-read alignment component of the Burrows-Wheeler Aligner [20].

Reads containing insertions or deletions, and those mapping to multiple locations in the reference were removed from the final alignments. Each alignment was analyzed for SNPs using SolSNP (<http://sourceforge.net/projects/solsnp/>). SNPs were excluded if they did not meet a minimum coverage of 10× and if the variant was present in less than 90% of the base calls for that position. In parallel, publically available genomes were aligned against R265 using MUMmer 3.22. SNPs were extracted from the alignments using a custom script. Subsequently, regions found to be duplicated with the R265 reference genome were identified using MUMmer version 3.22 [21]. SNPs residing within these repetitive regions were then removed. Loci that lacked reference sequence coverage data for one or more isolates were removed from the final analysis. The remaining loci were compiled in a matrix of orthologous SNPs loci shared across all genomes. Phylogenetic analyses were performed using the maximum parsimony algorithm, default parameters, in MEGA5 [18]. Where shown, 1000 generations were run for bootstrap analysis. Reference genome

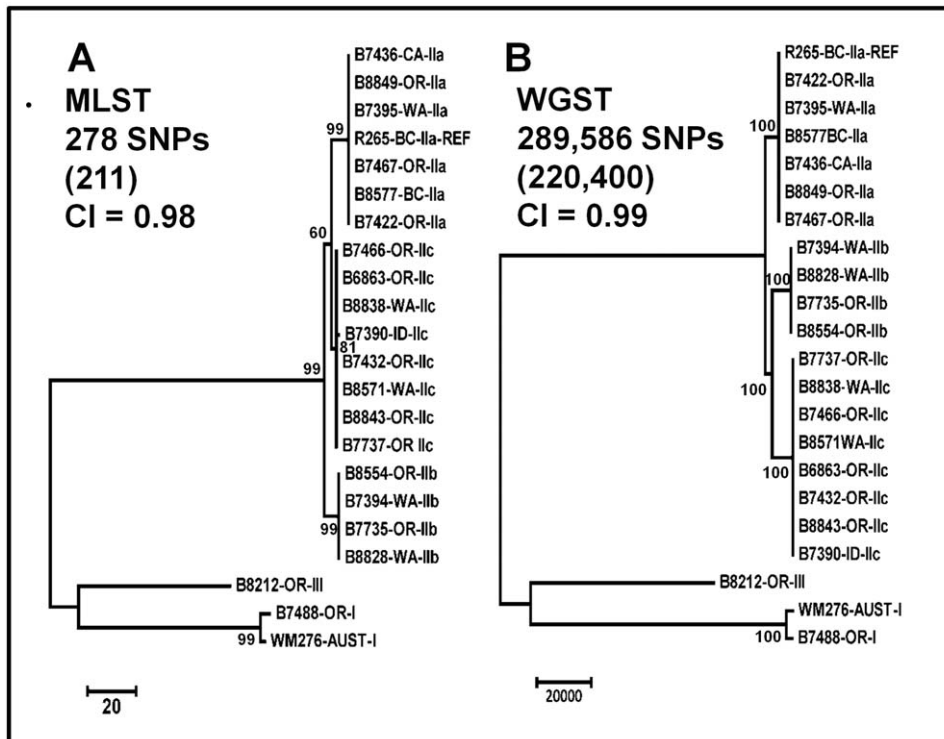


Figure 1. Phylogenetic analysis comparison of MLST and WGST data from 22 *C. gattii* isolates. Maximum parsimony phylogenetic analysis was performed in MEGA4 on 22 *C. gattii* genomes with MLST (panel A, one of 270 most parsimonious trees shown) and WGST (panel B, one of 190 most parsimonious trees shown) data. The R265 whole genome sequence was used as the reference for SNP discovery. Both trees are rooted on the VGI/VGIII branch. Bootstrap values less than 50% and for intra-VGIIa, VGIIb and VGIII branches are not shown. The taxa nomenclature includes a four digit unique identifier, the location of origin (BC = British Columbia, CA = California, ID = Idaho, OR = Oregon and WA = Washington state), and the molecular subtype as determined by MLST analysis (I, IIa, IIb, IIc and III). The number of SNPs included in each data set is indicated (number of parsimony-informative SNPs in parentheses), as is the consistency index (CI) as calculated by MEGA4. While WGST analysis found unique genotypes for all isolates, they are not visible on this tree due to the large numbers of SNPs separating the VGII from VGI and VGIII isolates. doi:10.1371/journal.pone.0028550.g001

mapping and read depth statistics were determined using Lasergene's Seqman NGEN™ V. 2.2 software (Lasergene, Madison, WI). The Illumina WGS data files have been deposited in the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra>, accession # SRP006436).

Results

Twenty two *C. gattii* genomes representing three of the four previously described molecular types (VGI, VGII, and VGIII) were subjected to MLST analysis using the consensus seven loci [17]. MLST analysis delineated 278 single nucleotide polymorphisms (SNPs) among the 22 *C. gattii* genomes, 211 of which were parsimonious, with the majority being among the different subtypes (Figure 1). Among the three VGII subtypes, MLST was able to distinguish 18 SNPs, 11 of which were parsimonious. Within each of the VGII subtypes all genomes were clonal with the exception of a single non-parsimonious SNP among the VGIIc isolates.

The same set of 22 genomes was analyzed by WGST analysis (Figure 1). The average read depth of the VGII isolates, when aligned to the R265 reference genome (VGIIa, 17.5 Mbp), ranged from 18–108×, with reads from all VGII isolates mapping to >86% of the reference genome. The VGI and VGIII genomes had average read depths of 33× and 16× with reference genome coverage of 72.7% and 64.2%, respectively. The average distance between paired reads was 450 bp. Among all 22 genomes, 289,586 SNPs were identified, 220,400 (76.1%) of which were parsimony-informative. There were a small number of insertion and deletion differences between VGI/VGIII and VGII, but none among the VGII genomes. These insertions and deletions were not incorporated into the phylogenetic analysis. The trees shown were rooted with VGI and VGIII genomes as we established that they formed an outgroup to the VGII group (Figure S1). When the VGI and VGIII genomes were excluded, WGST analysis found 56,845 SNPs, 56,095 (98.7%) of which were parsimony-informative, among all three VGII subtypes (Figure 2). Excluding more distantly related genomes from the WGS SNP discovery analysis increases the portion of the reference genome shared and, therefore, the number of shared SNP loci found among the remaining genomes. Branch lengths for the VGIIa, VGIIb, and VGIIc subclades were 18,701, 17,239, and 19,666 SNPs, respectively. All other branch lengths on this tree were less than 500 SNPs. The Consistency Index (CI) for the VGII isolate Maximum Parsimony analysis was 0.99. Separate analyses on each of the VGII subtypes found a single most parsimonious tree for each subtype and 1,512 SNPs (9 (0.6%) parsimony-informative), 132 SNPs (4 (3%) parsimony-informative) and 137 SNPs (15 (10.9%) parsimony-informative) within the VGIIa, VGIIb and VGIIc subtypes, respectively (Figure 3). Within-subtype analyses resulted in similarly high CI values, indicating minimal homoplasy within these data sets.

The phylogenetic relationships among the various subtypes and general topology of the trees were consistent, for the most part, between MLST and WGST analysis, both providing clear separation of each subtype (Figure 1). The major difference between the two types of analysis was the improved resolving power of WGST over MLST within each of the VGII clonal subtypes. Only a single SNP difference was found within any of the subtypes (VGIIc) using MLST, whereas unique genotypes were generated for each isolate by WGST.

Discussion

Employing next generation sequencing to analyze a large number of *C. gattii* genomes (1) demonstrate that WGST has

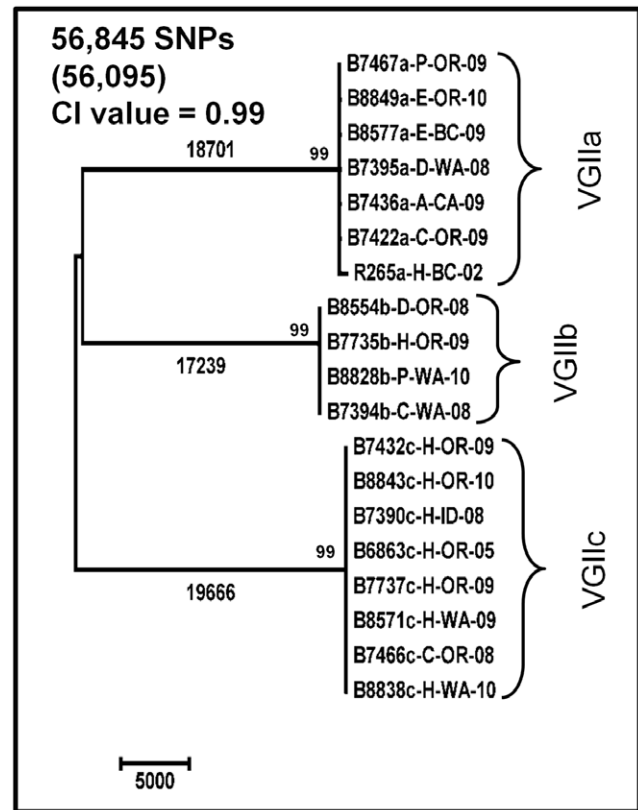


Figure 2. Phylogenetic analysis of WGST data from *C. gattii* VGII molecular type isolates. Maximum parsimony phylogenetic analysis was performed in MEGA4 on WGST SNP data from 18 VGII *C. gattii* genomes using the R265 whole genome sequence as the reference for SNP discovery (R265a-H-BC-02). The analysis found 76 most parsimonious trees, one of which is shown. The total number of SNPs included in the analysis is indicated (number of parsimonious SNPs in parentheses) as is the consistency index (CI) as calculated by MEGA4. Bootstrap values less than 50% are not shown. The tree shown is rooted on the mid-point. Branch lengths as calculated by MEGA4 are indicated above the three major branches. All other branch lengths are less than 500. The taxa nomenclature include a unique four digit identifier, the molecular subtype (a, b or c), the source of the isolate (A = alpaca, C = cat, D = dog, E = environmental, H = human and P = porpoise), the location of origin (BC = British Columbia, CA = California, ID = Idaho, OR = Oregon and WA = Washington state), and the year of collection. For example, B7395a-D-WA-08 indicates that isolate B7395 is a VGIIa subtype collected from a dog in Washington state in 2008. doi:10.1371/journal.pone.0028550.g002

superior resolving power as compared to available sub-typing tools; (2) reveals the presence of unique genotypes within VGII “clonal” subtypes; and (3) yields evidence that while the VGIIb sublineage may have radiated earlier than the other two sublineages, it may not have given rise to them.

In the present study, *C. gattii* WGST analysis identified a substantial number of genomic differences that were undetectable by the currently available typing methodologies that include MLST and VNTR analysis [5]. Analyses of twenty two genomes by both MLST and WGST revealed that while the former method generated only 18 SNPs, the latter method revealed nearly 57,000 SNPs that discriminated among and within the three VGII molecular subtypes. Further, WGST analysis revealed over one hundred SNPs within each VGII clonal subtype, more than MLST revealed across all VGII subtypes. More importantly, MLST analysis revealed no SNPs among the VGIIa and VGIIb

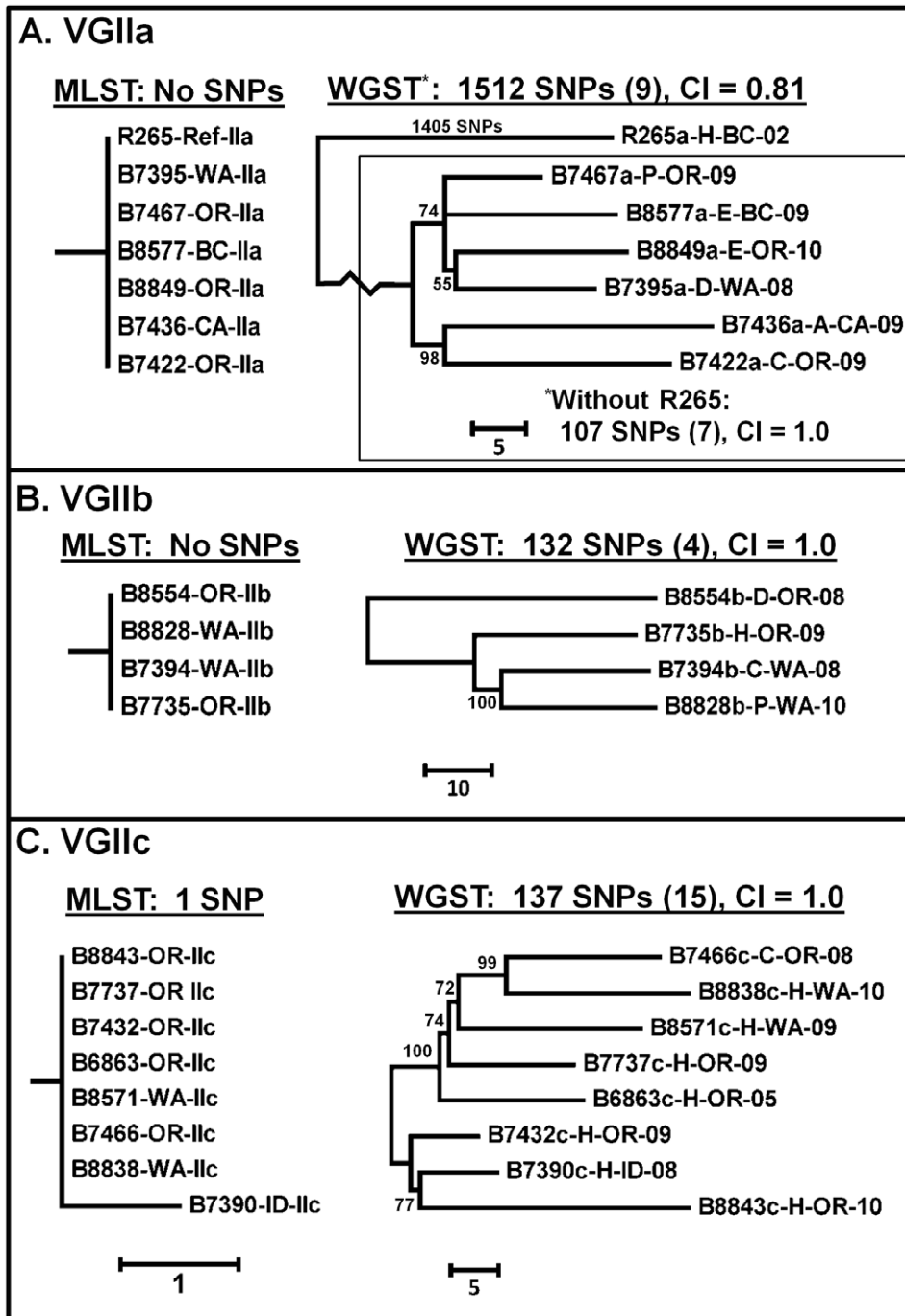


Figure 3. Comparison of MLST and WGST analysis of VGIIa, VGIIb and VGIIc *C. gattii* isolates. Maximum parsimony phylogenetic analysis was performed in MEGA4 on MLST and WGST SNP data from VGIIa (panel A), VGIIb (panel B) and VGIIc (panel C) genomes. Scale bars indicate relative branch lengths for each analysis. The R265 whole genome sequence was used as the reference for SNP discovery (R265a-H-BC-02). VGIIa (panel A) analysis is shown both with and without R265, VGIIb and VGIIc shown only without R265. The numbers of SNPs included in each analysis is indicated. The consistency index value (CI) was calculated in MEGA4. Meaning of taxa nomenclature is described in legends for Figures 1 (MLST) and 2 (WGST). Trees shown are mid-point rooted, bootstrap values less than 50% are not shown.
doi:10.1371/journal.pone.0028550.g003

subtypes and only one SNP among the VGIIc isolates, leaving individual isolates undifferentiated, while WGST analysis was able to generate 22 unique genotypes among the 22 genomes analyzed in addition to establishing the phylogenetic structure among all isolates and defining fine scale evolutionary relationships between the individual isolates.

Of note is the observation that most of the SNPs in VGIIa are found on the branch leading to the single R265 genome; when this genome is removed from the analysis, only 107 SNPs are identified among the remaining VGIIa isolates, seven of which are parsimony-informative (Figure 3). One reason for this could be that isolate R265 was collected in BC while the majority of other isolates were

from the PNW. However, B8577 is an environmental isolate from BC so it might be expected to be more closely related to R265, but this was not the case in this analysis. Another explanation could be that isolate R265 is also temporally separated from the other isolates, however this only represents a six year difference. An alternative, and we believe more plausible, explanation for this divergence could be the presence of sequencing errors in the reference strain, R265, which was sequenced with an average coverage of $6.5\times$, using Sanger sequencing. [16]. Sequencing errors in the reference would have limited impact on the diversity among the VGIIb and VGIIc clades in this analysis. As only shared orthologous SNPs were used in the analyses, a lower average read depth in a given WGST data set would decrease the shared portion of the reference genome interrogated for SNPs in each analysis, potentially decreasing the number of SNPs found among any of the groups. The lowest percent reference genome coverage in the VGIIa data set was greater (96.7%) than the lowest coverage found in the VGIIc data set (92.9%; Table 1).

While VGII subtypes had been identified in several collections globally distributed in Australia, Europe and South America [22–24], neither the VGIIa nor the VGIIc subtypes had been previously identified from any global collections. The novelty of the subtypes and their recent identification in the PNW are indications that these are recently emerged subtypes. The VGIIb subtype, which has been previously identified in other parts of the world, could have appeared in the PNW much earlier than VGIIa and VGIIc [7,12,14]. While the number of WGST SNPs identified among the VGIIa (after removing the R265 sequence from the analysis) and VGIIc isolates is similar to the number identified among the VGIIb isolates (Figure 3), this may simply be due to fewer VGIIb isolates being included in the analysis. There appears to be as much, if not more, genetic diversity among the four VGIIb isolates than any four of the VGIIa or VGIIc isolates as indicated by the relative branch lengths of the isolates. While further analysis with additional VGIIb isolates is necessary for confirmation, the data presented here is consistent with the current belief that the VGIIb clone radiated earlier than VGIIa or VGIIc clones, thus having more time to establish distinct sublineages.

Our study provides further evidence of the distinct nature of the *C. gattii* clonal subtypes causing disease in the PNW. WGST revealed definitively that the VGIIc subtype of *C. gattii*, the predominant causative agent of recent cryptococcal illness in Oregon [12,13], is genetically distinct from both the VGIIa and VGIIb subtypes. Additionally, while MLST analysis shows that VGIIa and VGIIc are slightly more closely related to each other than to VGIIb, WGST analysis indicates that VGIIb and VGIIc are slightly more closely related to each other than to VGIIa. All three subtypes do appear to be equally genetically distinct from one another, as indicated by the nearly identical branch lengths for each of the subtype clades. As such, it seems unlikely that the VGIIb clone gave rise to the VGIIa and/or VGIIc clones, as has been previously suggested. WGST analysis of an expanded collection of VGII isolates, including analyzing patterns of shared diversity, is planned and may help to better characterize the relationships among these three important subtypes.

The VGI and VGIII isolates were included in this study as outgroups to anchor the analysis, but the genome data provide important information. The diversity ($\sim 290,000$ SNPs) seen among the VGI, VGII and VGIII molecular types is consistent with previous reports that assigned varietal status to these types [11,16]. While this was not a goal of the current analysis, further analysis of the WGST data will help to delineate the genomic differences between the *C. gattii* molecular types and subtypes as well as differences in pathogenesis [5,7,11].

The comprehensive SNP data sets provided by WGS analysis will help facilitate the development of an accurate molecular clock, which will facilitate a better understanding of the evolutionary history of this organism. One of the most important aspects of this analysis is that we will now be able to individually delineate every *C. gattii* isolate collected. Large numbers of SNPs enable robust phylogenetic analyses, leading to more accurate phylogenetic reconstructions, especially in recombining organisms [15,25]. The delineated phylogenetic relationships may allow us to link travel to exposure, map the spread of clonal lineages, and with the future development of a library of environmental isolates we may be able to delineate both the geographic and clonal origin of this emergence. We may eventually be able to link molecular subtypes as well as specific genotypes with pattern or severity of illness which would be of significant public health importance. SNP-based WGST phylogeny may not capture all small genetic differences due to sexual recombination that could lead to phenotypic changes. Detection of these types of changes would require more comprehensive genome sequence comparison to find structural variation such as rearrangements or insertion/deletion events. However, the fine-scale resolution of WGST has been shown to be a powerful and discriminatory molecular tool that can link epidemiologically related isolates in epidemiologic investigations [15,26,27].

Data generated from this study have established a framework from which computational analyses can be designed and conducted to better characterize the *C. gattii* and to provide robust tools for rapidly detecting emerging strains and efficiently responding to outbreaks.

Supporting Information

Figure S1 Phylogenetic analysis of WGST data from *C. gattii*, *C. grubii* and *C. neoformans* isolates. Maximum parsimony phylogenetic analysis was performed in MEGA4 on *C. gattii*, *C. neoformans* and *C. grubii* whole genome sequence data [19]. The R265 whole genome sequence was used as the reference for SNP discovery. The tree is rooted on the *C. grubii*/*C. neoformans* branch; one of 190 most parsimonious trees is shown. Bootstrap values less than 50% and for intra-VGIIa, VGIIb and VGIIc nodes are not shown. The numbers of SNPs included in the analysis is indicated (number of parsimony-informative SNPs in parentheses), as is the consistency index (CI) as calculated by MEGA4. While WGST analysis found unique genotypes for all isolates, they are not visible on this tree due to the large numbers of SNPs separating the VGII isolates from the other isolates. (TIF)

Acknowledgments

This work would not be possible without the efforts of BC and US state public health officials who have contributed all isolates for this study. The authors acknowledge the support of Cyndi Free, Dawn Daly, Ben Sun, Randall Nett and Tom Chiller. The authors thank Jim Kronstad for generously sharing the completed *C. gattii* WM276 genome sequence in advance of its public release.

The findings and conclusions of this article are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Author Contributions

Conceived and designed the experiments: JDG JMS DME SAB SRL. Performed the experiments: JDG SRL. Analyzed the data: JDG SRL SAB DME JMS PK TP YY. Contributed reagents/materials/analysis tools: DME SRL SAB. Wrote the paper: JDG JMS SAB JH TP YY PK ED NM RW DME SRL.

References

- 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.
- Hoang LM, Maguire JA, Doyle P, Fyfe M, Roscoe DL (2004) *Cryptococcus neoformans* infections at Vancouver Hospital and Health Sciences Centre (1997–2002): epidemiology, microbiology and histopathology. *J Med Microbiol* 53: 935–940.
- MacDougall L, Kidd SE, Galanis E, Mak S, Leslie MJ, et al. (2007) Spread of *Cryptococcus gattii* in British Columbia, Canada, and detection in the Pacific Northwest, USA. *Emerg Infect Dis* 13: 42–50.
- Okamoto K, Hatakeyama S, Itoyama S, Nukui Y, Yoshino Y, et al. (2010) *Cryptococcus gattii* genotype VGIIa infection in man, Japan, 2007. *Emerg Infect Dis* 16(7): 1155–1157.
- Byrnes EJ, 3rd, Li W, Lewit Y, Ma H, Voelz K, et al. (2010) Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the northwest United States. *PLoS Pathog* 6(4): e1000850.
- Kidd SE, Hagen F, Tschärke RL, Huynh M, Bartlett KH, et al. (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.
- Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, et al. (2005) Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. *Nature* 437: 1360–1364.
- Byrnes EJ, 3rd, Heitman J (2009) *Cryptococcus gattii* outbreak expands into the Northwestern United States with fatal consequences. *F1000 Biol Rep* 1.
- Stephen C, Lester S, Black W, Fyfe M, Raverty S (2002) Multispecies outbreak of cryptococcosis on southern Vancouver Island, British Columbia. *Can Vet J* 43: 792–794.
- Bovers M, Hagen F, Kuramae EE, Boekhout T (2008) Six monophyletic lineages identified within *Cryptococcus neoformans* and *Cryptococcus gattii* by multi-locus sequence typing. *Fungal Genet Biol* 45: 400–421.
- Ngamskulrungron P, Serena C, Gilgado F, Malik R, Meyer W (2010) Global VGIIa isolates are of comparable virulence to the major fatal *Cryptococcus gattii* Vancouver Island outbreak genotype. *Clin Microbiol Infect* 17(2): 251–258.
- Byrnes EJ, 3rd, Bildfell RJ, Frank SA, Mitchell TG, Marr KA, et al. (2009) Molecular evidence that the range of the Vancouver Island outbreak of *Cryptococcus gattii* infection has expanded into the Pacific Northwest in the United States. *J Infect Dis* 199(7): 1081–1086.
- Iqbal N, DeBess EE, Wohrle R, Sun B, Nett RJ, et al. (2010) Correlation of genotype and in vitro susceptibilities of *Cryptococcus gattii* strains from the Pacific Northwest of the United States. *J Clin Microbiol* 48(2): 539–544.
- Kidd SE, Guo H, Bartlett KH, Xu J, Kronstad JW (2005) Comparative gene genealogies indicate that two clonal lineages of *Cryptococcus gattii* in British Columbia resemble strains from other geographical areas. *Eukaryot Cell* 4: 1629–1638.
- Engelthaler DM, Chiller T, Schupp JA, Colvin J, Beckstrom-Sternberg SM, et al. (2011) Next-Generation Sequencing of *Coccidioides immitis* isolated during cluster investigation. *Emerg Infect Dis* 17(2): 227–232.
- D'Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, et al. (2011) Genome Variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *MBio* 2: e00342-10.
- Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR, et al. (2009) Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. *Med Mycol* 47(6): 561–570.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–1599.
- Cryptococcus gattii* Sequencing Project. Broad Institute of Harvard and MIT. Available: http://www.broadinstitute.org/annotation/genome/cryptococcus_neoformans_b/Home.html.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* 25: 1754–60.
- Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, et al. (2004) Versatile and open software for comparing large genomes. *Genome Biol* 5: R12.
- Baró T, Torres-Rodríguez JM, De Mendoza MH, Morera Y, Alía C (1998) First identification of autochthonous *Cryptococcus neoformans* var. *gattii* isolated from goats with predominantly severe pulmonary disease in Spain. *J Clin Microbiol* 36: 458–61.
- Campbell LT, Currie BJ, Krockenberger M, Malik R, Meyer W, et al. (2005) Clonality and recombination in genetically differentiated subgroups of *Cryptococcus gattii*. *Eukaryot Cell* 4: 1403–1409.
- Santos WR, Meyer W, Wanke B, Costa SP, Trilles L, et al. (2008) Primary endemic *Cryptococcosis gattii* by molecular type VGII in the state of Pará, Brazil. *Mem Inst Oswaldo Cruz* 103(8): 813–818.
- Pearson T, Giffard P, Beckstrom-Sternberg S, Auerbach R, Hornstra H, et al. (2009) Phylogeographic reconstruction of a bacterial species with high levels of lateral gene transfer. *BMC Biol* 7: 78.
- Hendriksen RS, Price LB, Schupp JM, Gillette JD, Kaas RS, et al. (2011) Population genetics of *Vibrio cholerae* from Nepal in 2010: Evidence on the origin of the Haitian outbreak. *MBio* 2(4): e00157-11.
- Engelthaler DM, Bowers J, Schupp JA, Pearson T, Ginther J, et al. (2011) Molecular investigations of an autochthonous case of melioidosis in Southern AZ, USA. *PLoS Negl Trop Dis* 5(10): e1347.