Advances in *Cryptococcus* genomics: insights into the evolution of pathogenesis

Christina A Cuomo[1/+], Johanna Rhodes[2], Christopher A Desjardins[1]

[1]Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA
[2]Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

Cryptococcus species are the causative agents of cryptococcal meningitis, a significant source of mortality in immunocompromised individuals. Initial work on the molecular epidemiology of this fungal pathogen utilized genotyping approaches to describe the genetic diversity and biogeography of two species, *Cryptococcus neoformans* and *Cryptococcus gattii*. Whole genome sequencing of representatives of both species resulted in reference assemblies enabling a wide array of downstream studies and genomic resources. With the increasing availability of whole genome sequencing, both species have now had hundreds of individual isolates sequenced, providing fine-scale insight into the evolution and diversification of *Cryptococcus* and allowing for the first genome-wide association studies to identify genetic variants associated with human virulence. Sequencing has also begun to examine the microevolution of isolates during prolonged infection and to identify variants specific to outbreak lineages, highlighting the potential role of hyper-mutation in evolving within short time scales. We can anticipate that further advances in sequencing technology and sequencing microbial genomes at scale, including metagenomics approaches, will continue to refine our view of how the evolution of *Cryptococcus* drives its success as a pathogen.

Key words: *Cryptococcus neoformans* - *Cryptococcus gattii* - MLST - genome - sequencing - evolution - recombination

The genome sequencing of *Cryptococcus*, as for many fungi, initially focused on generating assemblies of a few reference isolates (Cuomo & Birren 2010, Cuomo 2017). These assemblies provided high quality representations of the genome, including some that consist of chromosomal size scaffolds; however, due to difficulties in sequencing and assembling some genomic regions, gaps remained typically at centromeres or chromosome ends and occasionally at other internal regions. With the early release of this genomic data, the availability of the nearly complete catalog of genes for these species were rapidly utilized in studies of individual genes, gene families and biological pathways; the sequence was also leveraged to build functional genomic resources.

Recent years have seen dramatic expansion in the scale of fungal genome sequencing, with the falling cost and increased availability of sequencing technologies. This includes annotated genome assemblies representing each of the varieties of *C. neoformans*, var. *grubii* and var. *neoformans*, and each of the four major monophyletic lineages of *C. gattii* (VGI, VGII, VGIII, and VGIV) (Table) (Loftus et al. 2005, D'Souza et al. 2011, Janbon et al. 2014, Farrer et al. 2015, Rhodes et al. 2017b). Each of these six subdivisions and one additional novel group were proposed to be separate species and assigned new names (Table), though this revision of the taxonomy is debated (Hagen et al. 2015, Kwon-Chung et al. 2017). While *Cryptococcus* isolates are primarily haploid, diploid hybrid isolates of *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans* have also been reported (Table) (Franzot et al. 1999, Desnos-Ollivier et al. 2015). An increasing number of large scale studies take a re-sequencing approach utilizing high quality reference genome assemblies to detect variants across hundreds of genomes, in many cases leveraging prior multilocus sequence typing (MLST) data to select isolates. Comparing genomes at this scale has provided a fine scale view of the genetic diversity and level of exchange between lineages and across geographic regions, characterized how gene gain and loss have impacted these species and lineages, and enabled genome-wide association studies of phenotypes related to virulence. Here, we review the progress in genome sequencing of *Cryptococcus* and highlight how increasing scale has been applied to examine both macro and microevolution of both *C. neoformans* and *C. gattii*.

Population studies by MLST - While a number of genotyping methods have been used to study *C. neoformans* var. *grubii* and *C. gattii*, in 2009 the International Society for Human and Animal Mycology (ISHAM) *Cryptococcus* Working Group agreed upon MLST as the standardized genotyping approach. MLST can have high discriminatory power with appropriately selected loci, give reproducible results between laboratories, and has been applied to differentiate species and to study the epidemiology of fungal pathogens (Taylor & Fisher 2003). Seven loci, comprised of six protein coding genes (*CAP59*, *GPD1*, *LAC1*, *PLB1*, *SOD1* and *URA5*) and the intergenic spacer IGS1 (Meyer et al. 2009), were selected to identify all eight major molecular types of both species. The resulting allele types are assigned according to either Litvintseva et al. (2006) for *C. neoformans* var. *grubii*, or Fraser et al. (2005) for *C. gattii*.

TABLE

Major subdivisions within the *Cryptococcus neoformans* and *Cryptococcus gattii* species

Species	Variety	Molecular Type	AFLP	Proposed species name
C. neoformans	var. *grubii*	VNI	AFLP1	*C. neoformans*
		VNII		
		VNB		
	var. *neoformans*	VNIV	AFLP2	*C. deneoformans*
	AD hybrid	VNII	AFLP3	*C. neoformans* x *C. deneoformans* hybrid
C. gattii		VGI	AFLP4	*C. gattii*
		VGII (a, b and c sub-genotypes)	AFLP6	*C. deuterogattii*
		VGIII (a, b and c sub-genotypes)	AFLP5	*C. bacillisporus*
		VGIV	AFLP7	*C. tetragattii*
		VGIV/VGIIIc	AFLP10	*C. decagattii*

MLST analysis revealed subpopulations within *C. neoformans* var. *grubii* and *C. gattii* (Litvintseva et al. 2006, Bovers et al. 2008, Ngamskulrungroj et al. 2009) and subsequently identified fine-grain geographical associations with particular sequence types (Chowdhary et al. 2011, Beale et al. 2015, Lam et al. 2017). In providing initial estimates of variation within the population, higher estimates of diversity in South African *C. neoformans* var. *grubii* isolates led to the hypothesis that this population was ancestral and isolates spread around the globe 'out of Africa' (Litvintseva et al. 2011, Simwami et al. 2011, Litvintseva & Mitchell 2012). By contrast, high similarity of MLST profiles can suggest more recent transitions, such as the apparent migration of the VNI lineage from African populations to Asia, infecting predominantly to HIV-positive patients (Simwami et al. 2011, Khayhan et al. 2013). Similar MLST analyses identified high diversity of the *C. gattii* VGII lineage in northern Brazil, suggesting this location could be the source for the global expansion of VGII including the outbreak in the Pacific Northwest (Hagen et al. 2013, Souto et al. 2016). These comparisons also revealed the introduction of Australian VGII into Vancouver Island, Canada (Fraser et al. 2005, Byrnes 3rd et al. 2009, 2010). Thus, these MLST studies were fundamental in describing the population structure of these species and describing associations with geography.

A limitation of MLST is that only a small fraction of the sequence diversity is captured (Beale et al. 2015) and this may provide incomplete or in some cases inaccurate measures of the species relationships. More specifically, comparing inadequate data and relying on congruence can lead to incorrect assertions of recombination (Tibayrenc & Ayala 2014) and of the relationships of isolates, as most phylogenetic methods rely on an assumption of vertical descent. Some of the *Cryptococcus* MLST loci have been shown to have undergone recombination (Litvintseva et al. 2006) and therefore including them among a small number of loci in a multi-gene phylogeny may result in an inaccurate view of the species-level relationships of some isolates.

Reference genome sequencing projects - The earliest completed fungal genomes, sequenced using Sanger technology, included several *Cryptococcus* isolates (Ga-lagan et al. 2005, Cuomo & Birren 2010). The first report described the genomes of two related isolates of *C. neoformans* var. *neoformans*, JEC21 and B-3501 (Loftus et al. 2005). Several years later, the genomes of two isolates of *C. gattii* were compared; WM276 was selected to represent the predominant VGI lineage and R265 to represent the VGII lineage responsible for an outbreak in the Pacific Northwest (Fraser et al. 2005, D'Souza et al. 2011). The genome for the H99 VNI isolate of *C. neoformans* var. *grubii* was the first to incorporate RNA-Seq, generated from yeast grown in a variety of conditions, to predict a deep catalog of alternatively spliced transcripts and non-coding genes (Janbon et al. 2014). More recently, high quality reference assemblies for larger numbers of isolates have been generated from Illumina sequencing. To date, 6 *C. gattii*, 3 *C. neoformans* var. *neoformans*, and 49 *C. neoformans* var. *grubii* assemblies have been deposited onto NCBI. These annotated genomes include representatives of all the major molecular types of *C. gattii* (including VGIII and VGIV isolates) and *C. neoformans* var. *grubii* (including VNII and VNB isolates) (Farrer et al. 2015, Rhodes et al. 2017a).

Overall genome structure was found to be highly conserved across *Cryptococcus* with a small number of rearrangements detected between species or molecular types. The chromosomal sequence of isolates of the different molecular types of *C. gattii* were found to be highly colinear, with a small number of rearrangements detected; sequence divergence of ~7% on average (D'Souza et al. 2011, Farrer et al. 2015) supported a suggestion that these could represent different species (D'Souza et al. 2011). The genomes of *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans* also vary in structure, with three translocations and two large inversions; more extensive rearrangements are observed for these genomes in comparison to *C. gattii* (Janbon et al. 2014).

Characterization of gene content and structure revealed differences in *Cryptococcus* compared to other fungi. *Cryptococcus* genes are intron-rich (Loftus et al. 2005) with extensive alternative splicing predicted from RNA-Seq (Janbon et al. 2014). While gene content is highly conserved in *C. gattii*, variable genes are enriched for specific functions including response to oxidative stress, mitochondrial import, and metal binding

and transport (Farrer et al. 2015). One notable difference between the *C. gattii* molecular types involves the loss of genes involved in RNA interference in VGII isolates (D'Souza et al. 2011, Farrer et al. 2015). Variation in gene content has also been examined *in C. neoformans* var. *grubii*; a recent study assembled and annotated 39 genomes representing each of the major lineages (VNI, VNII, and VNB) (Rhodes et al. 2017b). A small number of genes specific to each lineage were identified, including transporters and transcription factors; these functional categories were also found to include the most rapidly evolving genes (Rhodes et al. 2017b).

The sequenced reference genomes have also enabled the construction of large-scale functional genomic resources. A deletion collection of all *C. neoformans* var. *grubii* genes in the H99 strain background is underway (Liu et al. 2008), which helps support routine analysis of the requirement for specific genes under different conditions. Smaller functionally-focused deletion collections of transcription factors (Jung et al. 2015) and kinases (Lee et al. 2016) have already been completed and analyzed, making a host of new connections between specific genes and virulence or virulence-related phenotypes.

Population sequencing of C. gattii - Building on the analysis of reference genomes and assemblies, several studies have examined genetic diversity in *C. gattii* using whole genome sequencing. Available genome sequence represents four major molecular types of *C. gattii* (VGI, VGII, VGIII, and VGIV) and some more rarely observed clades related to these lineages. Isolates of VGII have been well sampled by sequencing to understand the recent outbreak in the Pacific Northwest. In addition, VGIII has been deeply sequenced in an effort to characterize an increasing number of infections in the USA, South and Central America. By contrast, the VGI and VGIV groups are not as deeply sequenced to date; VGI isolates are common to Asia, Australia, and Europe, whereas VGIV is most prevalent in India and Africa.

Initial studies of VGII diversity examined the relationship of isolates in the outbreak lineages in the Pacific Northwest to those found in other geographic regions. In 1999, an outbreak of *C. gattii* occurred on Vancouver Island in Canada and expanded to the Pacific Northwest, a region outside the typically tropical and subtropical range of this pathogen (Hoang et al. 2004, Fraser et al. 2005). These studies confirmed that there are three major subgroups of VGII, including the VGIIa and VGIIc lineages that are largely restricted to the Pacific northwest and the more widely distributed VGIIb (Billmyre et al. 2014, Engelthaler et al. 2014). Using phylogenetic methods, South American isolates were placed as outgroups to both VGIIa and VGIIb (Billmyre et al. 2014, Engelthaler et al. 2014); in addition, novel VGII genotypes outside these major subgroups most frequently include isolates from South America (Engelthaler et al. 2014), supporting an origin of South America for this lineage as also suggested by aforementioned MLST studies (Hagen et al. 2013, Souto et al. 2016). While the major subgroups each appear highly clonal, as they have one mating type and no evidence of recombination, there is evidence of

ancestral recombination between the subgroups (Billmyre et al. 2014, Engelthaler et al. 2014). Examining mutations specific to the VGIIa outbreak lineage revealed a frameshift mutation in an ortholog of the *MSH2* DNA mismatch repair gene; this was found only in the VGIIa-like isolates which are lower in virulence than other VGIIa isolates (Billmyre et al. 2014). The rate of mutation was higher in homopolymer runs in the VGIIa-like isolates; however, these were not detected in other VGIIa outbreak lineages or ancestral branches, suggesting that loss of *MSH2* and higher mutation rate and may not have played a role in adaption in the virulent isolates within the outbreak lineage (Billmyre et al. 2017).

Recent studies have characterized the subdivisions and geographical associations of isolates in the VGIII group by comparing genome sequences. In Southern California, *C. gattii* infections of HIV/AIDS patients are predominantly caused by VGIII isolates (Byrnes 3rd et al. 2011); isolation of local environmental isolates revealed a very close relationship, with three pairs of clinical and environmental isolates each separated by an average of only 106 SNPs, suggesting a local environmental reservoir of *C. gattii* VGIII (Springer et al. 2014). MLST study of VGIII isolates revealed two separated subgroups, VGIIIa and VGIIIb (Byrnes 3rd et al. 2011), corresponding to two different serotypes (Firacative et al. 2016). A wider study of 60 VGIII genomes supported these two clades and notably identified two small related clades of VGIII-like isolates, one of which shares most MLST alleles with the AFLP10/*C. decagattii* isolate (Hagen et al. 2015, Firacative et al. 2016). Further analyses of the sequenced genomes of VGIIIa, VGIIIb, and AFLP10/*C. decagattii* isolates are needed to characterize the level of genetic exchange between these groups.

Farrer et al. (2015) combined population-level sequencing of 37 isolates with 16 reference assemblies of *C. gattii*, and identified a new transcontinental sublineage of VGII, named VGIIx. Comparison of the mitochondrial and nuclear genomes revealed evidence of recombination between the VGII and VGIII mitochondrial but not nuclear genomes, providing a glimpse of genetic exchange processes within *C. gattii*. Furthermore, tests of selection between all major lineages and sub-lineages of *C. gattii* identified multidrug transporters, aconitases (iron-sulfur proteins), capsule genes, heat-shock proteins and protein kinases under positive selection in multiple sub-lineages (Farrer et al. 2015, 2016).

Population sequencing of C. neoformans - A number of recently published population-level studies have greatly expanded the knowledge of genomic diversity of *C. neoformans* var. *grubii* available (Desjardins et al. 2017, Rhodes et al. 2017b, Vanhove et al. 2017). These studies recapitulated the three traditional lineages (VNI, VNII, and VNB) but also supported splitting VNB into two lineages, VNBI and VNBII, a pattern that had been noted previously in AFLP- and MLST-based studies (Litvintseva et al. 2003, 2006, Chen et al. 2015). Recombination was found to be occurring within all tested lineages including the largely unisexual VNI, although at a slightly slower rate than VNBI and VNBII (Desjardins

et al. 2017). Recombination between the three lineages appeared limited, suggesting some degree of reproductive isolation. A number of inter-lineage hybrids were identified, however, including some that likely resulted from recent meiotic crossing-over, as evidenced by large blocks of single-lineage ancestry across each chromosome (Rhodes et al. 2017b). Aneuploidy was found across all chromosomes in both single-lineage and hybrid strains, and a number of diploid hybrids between *C. neoformans* var. *grubii* and *C. neoformans* var. *neoformans* showed aneuploidy leading to loss-of-heterozygosity across various chromosomes. In sum, these analyses point to limited but definitive genetic exchange between the *C. neoformans* var. *grubii* lineages.

The division of VNB into VNBI and VNBII enabled characterization of notable phenotypic differences between these two groups. In Botswana, VNBII was enriched for clinical isolates relative to VNBI (Desjardins et al. 2017) and the same trend was seen in Zambian isolates (Vanhove et al. 2017). Surprisingly, high-throughput phenotyping showed that VNBI environmental isolates were more resistant to oxidative stress and more heavily melanized that VNBI clinical isolates, perhaps reflecting a greater breadth of selective pressures in the environment than in the human host (Desjardins et al. 2017).

The characterization of population structure and recombination levels also guided selection of isolates for genome-wide association studies. Within VNB, mutations in virulence factors and stress response genes were found to be associated with clinical isolates versus environmental, and loss-of-function mutations in the transcription *BZP4* linked to melanization capacity (Desjardins et al. 2017). Thus, analyses of these large-scale genomic data have begun to identify a host of potential virulence factors and mutations to be directly tested in the laboratory setting.

Sequencing isolates from poorly represented geographic regions has provided invaluable insight into the biogeography of *Cryptococcus*. Vanhove et al. (2017) found that *C. neoformans* var. *grubii* and *C. gattii* occupied different ecological niches, with the former primarily recovered from Zambezi Mopane woodlands and the latter primarily recovered from Central Miombo woodlands. Sequencing of a number of *C. neoformans* var. *grubii* clinical isolates from Brazil revealed that both the VNBI and VNBII lineages are naturally occurring there (Rhodes et al. 2017b). Phylogenomic analysis revealed that deep branches separated South American and African isolates in both VNBI and VNBII, suggesting that migration of these lineages between continents occurred prior to their diversification (Rhodes et al. 2017b). This suggests that the transcontinental spread of *C. neoformans* var. *grubii* probably occurred much earlier than previously thought.

Serial isolates from patients and microevolution - In contrast to macro-evolutionary patterns described with genomic studies of global populations, genome sequencing has also been leveraged to study micro-evolution of *Cryptococcus* within patients by comparing isolates at initial presentation and weeks later during relapse of infection (Ormerod et al. 2013, Chen et al. 2017, Rhodes

et al. 2017a). Ormerod et al. (2013) sequenced a pair of initial and relapse isolates with marked virulence-related phenotypic differences (including growth rate at high temperature, capsule size, and melanization); a small number of genetic changes were identified by comparing variants, including deletion of a transcriptional regulator and changes in copy number on the arms of chromosome 12. Chen et al. (2017) sequenced 38 initial and relapse isolates from 18 patients and found mutations occurring in genes involved in a number of virulence-related phenotypes, including growth at 39ºC, stress response, and capsule production. They also identified a 13-fold amplification of a small region of chromosome 1 including *ERG11* and linked this to fluconazole resistance, pointing to the importance of aneuploidy to within-host evolution.

Beyond aneuploidy, a number of studies have identified clinical isolates capable of rapid adaption through hypermutation, largely linked to mutations in mismatch repair gene *MSH2* in *C. gattii* VGII (*C. deuterogattii*) (Billmyre et al. 2014, 2017) and *C. neoformans* var. *grubii* (Boyce et al. 2017, Rhodes et al. 2017a). Experimental measurements by Boyce *et al.* and Billmyre *et al.* (Billmyre et al. 2017, Boyce et al. 2017) demonstrated that deletion of *MSH2*, *MLH1*, and *PMS1* all produce hypermutator phenotypes, and deletion of *MSH2* specifically increases the rate of insertions and deletions in homopolymer tracts. This can increase the rate of drug resistance acquisition, particularly involving genes with long homopolymer tracts. Mutation of *PMS1* resulted in reduced virulence whereas mutation of *MLH1* and *MSH2* maintained fully virulent phenotypes (Boyce et al. 2017), supporting the hypothesis that hypermutator phenotypes can occur without a significant fitness tradeoff.

Cryptococcus microevolution has also been studied in the context of the laboratory strain *C. neoformans* var. *grubii* H99 (Janbon et al. 2014). This strain decreased in virulence after laboratory passaging and increased in virulence after passage in a rabbit model, comparing isolates from multiple laboratories (Janbon et al. 2014). Sequencing of three isolates exhibiting phenotypic differences revealed a frameshift in *LMP1* resulted in strain H99C being largely avirulent and infertile (Janbon et al. 2014). This illustrates the variation present in common laboratory isolates; understanding these differences and updating strain names to reflect this is critical for comparison of the genetic background of strains used in experiments across laboratories, as isolates may have diverged from wild type genotypes.

Future directions leveraging technological advances - As the above studies illustrate, genome sequencing of *Cryptococcus* is taking place at an increasing scale encompassing hundreds of isolates. As much of this data is being generated on a single platform, Illumina, combining and comparing data from different studies can be simplified. However, it is important to consider how differences in coverage level, library preparation chemistry, and sequencing bias could impact results from combined data sets (Rhodes et al. 2014); for example, sequencing bias could lead to under-representation of some regions of the genome, affecting analyses of copy number varia-

tion. Heterozygous isolates are also occasionally detected and require different protocols for variant identification and analysis (Rhodes et al. 2017b). In addition, such data re-use is potentiated by clearly documented metadata that typically includes the source of isolation and occasionally additional phenotypes measured experimentally.

These genomic studies rely on high quality reference assemblies and annotations. Current analysis of sequence data is based on comparisons to a small number of reference assemblies, which have benefitted from collaborative work to generate and curate their content. However, a single genome does not completely represent a species; gene gain and loss has been observed between isolates of *C. neoformans* var. *grubii* (Day et al. 2017, Rhodes et al. 2017b). In addition, most assemblies are not yet complete telomere to telomere representations of the genome, even for these reference genomes. Illumina data has been utilized to generate highly complete assemblies for additional isolates of *Cryptococcus*, with highest continuity where paired-end data from larger insert libraries is included; yet these assemblies do not represent complete genomes as they have many more scaffolds and contigs than chromosomes (Day et al. 2017, Rhodes et al. 2017b). Assembly of long reads generated by Pacific Biosciences or Oxford Nanopore Technologies may provide more contiguous assemblies, as they should help resolve repetitive regions including the centromeres and subtelomeres. However, targeted finishing effort may also be required to resolve some regions and achieve an end to end representation of these genomes.

Incorporation of additional genome-wide data sets have also provided another layer of annotation to reference genomes. The gene sets of the *C. neoformans* var. *grubii* (H99) and *C. neoformans* var. *neoformans* (JEC21) reference genomes have benefitted from RNA-Seq guided gene structure prediction (Janbon et al. 2014, González-Hilarion et al. 2016); this has improved the accuracy of splice site annotation in gene models and has also suggested the presence of large numbers of noncoding transcripts. RNA-Seq has also been used to update gene sets for each of the three main lineages of *C. neoformans* var. *grubii* and supported the prediction of lineage-specific genes (Rhodes et al. 2017a). Recent studies have begun to map modified bases across the *Cryptococcus* genome (Catania et al. 2017), and three dimensional genome mapping techniques utilizing sequencing to map higher order adjacency of genomic regions may also be applied in the near future. The integration of such diverse data sets into a more comprehensive model of the genome, and how it changes across time and varies within the population, will provide a more complete framework to study the evolution of these pathogens.

Increased population-level sequencing will also be required to further identify genetic mechanisms of virulence, drug resistance, and within-host adaptation. Genome-wide association studies require extremely large sample sizes to link variants to complex phenotypes, such as infectivity or virulence; while studies have begun to build this dataset (see Desjardins et al. 2017), further increases in sample size will enable further genetic insight. However small sets of isolates have the potential to link genotype to phenotype; for example, a comparing genomes for closely related isolates differing in the production of Titan cells identified stop codon mutations in the protein kinase A regulatory gene *PKR1*, for which prior work had demonstrated a role in Titan cell production (Choi et al. 2012, Hommel et al. 2017). Furthermore, while these genomic studies have examined individual isolates that are colony purified from clinical or environmental samples, metagenomic approaches could assist uncovering diversity within an individual host or site. For example, metagenomic sequencing of the population from serial isolates within a single patient over time would enable analysis of how mutations occur and become fixed. In sum, large-scale sequencing of both global populations and within single patients will continue to play critical roles in understanding the genetic basic of human virulence in *Cryptococcus*.

REFERENCES

Beale MA, Sabiiti W, Robertson EJ, Fuentes-Cabrejo KM, O'Hanlon SJ, Jarvis JN, et al. Genotypic diversity is associated with clinical outcome and phenotype in cryptococcal meningitis across southern Africa. PLOS Negl Trop Dis. 2015; 9(6): e0003847.

Billmyre RB, Clancey SA, Heitman J. Natural mismatch repair mutations mediate phenotypic diversity and drug resistance in *Cryptococcus deuterogattii*. eLife. 2017; 6: e28802.

Billmyre RB, Croll D, Li W, Mieczkowski P, Carter DA, Cuomo CA, et al. Highly recombinant VGII *Cryptococcus gattii* population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. mBio. 2014; 5(4): e01494-01414.

Bovers M, Hagen F, Kuramae EE, Boekhout T. Six monophyletic lineages identified within *Cryptococcus neoformans* and *Cryptococcus gattii* by multi-locus sequence typing. Fungal Genet Biol. 2008; 45(4): 400-21.

Boyce KJ, Wang Y, Verma S, Shakya VPS, Xue C, Idnurm A. Mismatch repair of DNA replication errors contributes to microevolution in the pathogenic fungus *Cryptococcus neoformans*. mBio. 2017; 8(3): e00595-17.

Byrnes 3rd EJ, Bildfell RJ, Frank SA, Mitchell TG, Marr KA, Heitman J. 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. 2009; 199(7): 1081-6.

Byrnes 3rd EJ, Li W, Lewit Y, Ma H, Voelz K, Ren P, et al. Emergence and pathogenicity of highly virulent *Cryptococcus gattii* genotypes in the Northwest United States. PLoS Pathog. 2010; 6(4): e1000850.

Byrnes 3rd EJ, Li W, Ren P, Lewit Y, Voelz K, Fraser JA, et al. A diverse population of *Cryptococcus gattii* molecular type VGIII in southern Californian HIV/AIDS patients. PLoS Pathog. 2011; 7(9): e1002205.

Catania S, Dumesic PA, Stoddard C, Cooke S, Burke J, Cuomo CA, et al. Epigenetic maintenance of DNA methylation after evolutionary loss of the de novo methyltransferase. bioRxiv. 2017; 149385.

Chen Y, Farrer RA, Giamberardino C, Sakthikumar S, Jones A, Yang T, et al. Microevolution of serial clinical isolates of *Cryptococcus neoformans* var. *grubii* and *C. gattii*. mBio. 2017; 8(2): e00166-17.

Chen Y, Litvintseva AP, Frazzitta AE, Haverkamp MR, Wang L, Fang C, et al. Comparative analyses of clinical and environmental populations of *Cryptococcus neoformans* in Botswana. Mol Ecol. 2015; 24(14): 3559-71.

Choi J, Vogl AW, Kronstad JW. Regulated expression of cyclic AMP-dependent protein kinase A reveals an influence on cell size and the secretion of virulence factors in *Cryptococcus neoformans*. Mol Microbiol. 2012; 85(4): 700-15.

Chowdhary A, Hiremath SS, Sun S, Kowshik T, Randhawa HS, Xu J. Genetic differentiation, recombination and clonal expansion in environmental populations of *Cryptococcus gattii* in India. Environ Microbiol. 2011; 13(7): 1875-88.

Cuomo CA, Birren BW. The fungal genome initiative and lessons learned from genome sequencing. Methods Enzymol. 2010; 470: Chapter 34.

Cuomo CA. Harnessing whole genome sequencing in medical mycology. Curr Fungal Infect Rep. 2017; 11(2): 52-9.

D'Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, Hu G, et al. Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. mBio. 2011; 2(1): e00342-10.

Day JN, Qihui S, Thanh LT, Trieu PH, Van AD, Thu NH, et al. Comparative genomics of *Cryptococcus neoformans* var. *grubii* associated with meningitis in HIV infected and uninfected patients in Vietnam. PLoS Negl Trop Dis. 2017; 11(6): e0005628.

Desjardins CA, Giamberardino C, Sykes SM, Yu C-H, Tenor JL, Chen Y, et al. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. 2017; 27(7): 1207-19.

Desnos-Ollivier M, Patel S, Raoux-Barbot D, Heitman J, Dromer F, French Cryptococcosis Study Group. Cryptococcosis serotypes impact outcome and provide evidence of *Cryptococcus neoformans* speciation. mBio. 2015; 6(3): e00311.

Engelthaler DM, Hicks ND, Gillece JD, Roe CC, Schupp JM, Driebe EM, et al. *Cryptococcus gattii* in North American Pacific Northwest: whole-population genome analysis provides insights into species evolution and dispersal. mBio. 2014; 5(4): e01464-14.

Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, et al. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio. 2015; 6(5): e00868-15.

Farrer RA, Voelz K, Henk DA, Johnston SA, Fisher MC, May RC, et al. Microevolutionary traits and comparative population genomics of the emerging pathogenic fungus *Cryptococcus gattii*. Philos Trans R Soc Lond B Biol Sci. 2016; 371(1709): pii: 0160021.

Firacative C, Roe CC, Malik R, Ferreira-Paim K, Escandón P, Sykes JE, et al. MLST and whole-genome-based population analysis of *Cryptococcus gattii* VGIII links clinical, veterinary and environmental strains, and reveals divergent serotype specific sub-populations and distant ancestors. PLoS Negl Trop Dis. 2016; 10(8): e0004861.

Franzot SP, Salkin IF, Casadevall A. *Cryptococcus neoformans* var. *grubii*: separate varietal status for *Cryptococcus neoformans* serotype A isolates. J Clin Microbiol. 1999; 37(3): 838-40.

Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, Diezmann S, et al. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature. 2005; 437(7063): 1360-64.

Galagan JE, Henn MR, Ma LJ, Cuomo CA, Birren B. Genomics of the fungal kingdom: insights into eukaryotic biology. Genome Res. 2005; 15(12): 1620-31.

González-Hilarion S, Paulet D, Lee K-T, Hon C-C, Lechat P, Mogensen E, et al. Intron retention-dependent gene regulation in *Cryptococcus neoformans*. Sci Rep. 2016; 6: 32252.

Hagen F, Ceresini PC, Polacheck I, Ma H, van Nieuwerburgh F, Gabaldón T, et al. Ancient dispersal of the human fungal pathogen *Cryptococcus gattii* from the Amazon rainforest. PLoS ONE. 2013; 8(8): e71148.

Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, et al. Recognition of seven species in the *Cryptococcus gattii/Cryptococcus neoformans* species complex. Fungal Genet Biol. 2015; 78: 16-48.

Hoang LMN, Maguire JA, Doyle P, Fyfe M, Roscoe DL. *Cryptococcus neoformans* infections at Vancouver Hospital and Health Sciences Centre (1997-2002): epidemiology, microbiology and histopathology. J Med Microbiol. 2004; 53(Pt 9): 935-40.

Hommel B, Mukaremera L, Cordero RJB, Desjardins CA, Sturny-Leclere A, Perfect J, et al. Identification of environmental and genetic factors important for *Cryptococcus neoformans* titan cell formation using new in vitro inducing conditions. bioRxiv. 2017; doi: https://doi.org/10.1101/191668.

Janbon G, Ormerod KL, Paulet D, Byrnes EJ, Yadav V, Chatterjee G, et al. Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. 2014; 10(4): e1004261.

Jung K-W, Yang D-H, Maeng S, Lee K-T, So Y-S, Hong J, et al. Systematic functional profiling of transcription factor networks in *Cryptococcus neoformans*. Nat Commun. 2015; 6: 6757.

Khayhan K, Hagen F, Pan W, Simwami S, Fisher MC, Wahyuningsih R, et al. Geographically structured populations of *Cryptococcus neoformans* variety *grubii* in Asia correlate with HIV status and show a clonal population structure. PloS ONE. 2013; 8(9): e72222.

Kwon-Chung KJ, Bennett JE, Wickes BL, Meyer W, Cuomo CA, Wollenburg KR, et al. The case for adopting the "Species Complex" nomenclature for the etiologic agents of Cryptococcosis. mSphere. 2017; 2(1): pii: e00357-16.

Lam TT, Phan TH, Rattanavong S, Nguyen TM, Duong AV, Dacon C, et al. Multilocus sequence typing reveals a unique co-dominant population structure of *Cryptococcus neoformans* var. *grubii* in Vietnam. bioRxiv. 2017; 190785.

Lee K-T, So Y-S, Yang D-H, Jung K-W, Choi J, Lee D-G, et al. Systematic functional analysis of kinases in the fungal pathogen *Cryptococcus neoformans*. Nat Commun. 2016; 7: 12766.

Litvintseva AP, Carbone I, Rossouw J, Thakur R, Govender NP, Mitchell TG. Evidence that the human pathogenic fungus *Cryptococcus neoformans* var. *grubii* may have evolved in Africa. PLoS ONE. 2011; 6(5): e19688.

Litvintseva AP, Marra RE, Nielsen K, Heitman J, Vilgalys R, Mitchell TG. Evidence of sexual recombination among *Cryptococcus neoformans* serotype A isolates in sub-Saharan Africa. Eukaryotic Cell. 2003; 2(6): 1162-8.

Litvintseva AP, Mitchell TG. Population genetic analyses reveal the African origin and strain variation of *Cryptococcus neoformans* var. *grubii*. PLoS Pathog. 2012; 8(2): e1002495.

Litvintseva AP, Thakur R, Vilgalys R, Mitchell TG. Multilocus sequence typing reveals three genetic subpopulations of *Cryptococcus neoformans* var. *grubii* (serotype A), including a unique population in Botswana. Genetics. 2006; 172(4): 2223-38.

Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. Cell. 2008; 135(1): 174-88.

Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science (New York). 2005; 307(5713): 1321-4.

Meyer W, Aanensen DM, Boekhout T, Cogliati M, Diaz MR, Esposto MC, et al. Consensus multi-locus sequence typing scheme for *Cryptococcus neoformans* and *Cryptococcus gattii*. Med Mycol. 2009; 47(6): 561-70.

Ngamskulrungroj P, Gilgado F, Faganello J, Litvintseva AP, Leal AL, Tsui KM, et al. Genetic diversity of the *Cryptococcus* species complex suggests that *Cryptococcus gattii* deserves to have varieties. PLoS ONE. 2009; 4(6): e5862.

Ormerod KL, Morrow CA, Chow EWL, Lee IR, Arras SDM, Schirra HJ, et al. Comparative genomics of serial isolates of *Cryptococcus neoformans* reveals gene associated with carbon utilisation and virulence. G3 (Bethesda). 2013; 3(4): 675-86.

Rhodes J, Beale MA, Fisher MC. Illuminating choices for library prep: a comparison of library preparation methods for whole genome sequencing of *Cryptococcus neoformans* using Illumina HiSeq. PLoS ONE. 2014; 9(11): e113501.

Rhodes J, Beale MA, Vanhove M, Jarvis JN, Kannambath S, Simpson JA, et al. A population genomics approach to assessing the genetic basis of within-host microevolution underlying recurrent cryptococcal meningitis infection. G3 (Bethesda). 2017a; 7(4): 1165-76.

Rhodes J, Desjardins CA, Sykes SM, Beale MA, Vanhove M, Sakthikumar S, et al. Tracing genetic exchange and biogeography of *Cryptococcus neoformans* var. *grubii* at the global population level. Genetics. 2017b; 207(1): 327-46.

Simwami SP, Khayhan K, Henk DA, Aanensen DM, Boekhout T, Hagen F, et al. Low diversity *Cryptococcus neoformans* variety *grubii* multilocus sequence types from Thailand are consistent with an ancestral African origin. PLoS Pathog. 2011; 7(4): e1001343.

Souto ACP, Bonfietti LX, Ferreira-Paim K, Trilles L, Martins M, Ribeiro-Alves M, et al. Population genetic analysis reveals a high genetic diversity in the Brazilian *Cryptococcus gattii* VGII population and shifts the global origin from the Amazon Rainforest to the semi-arid desert in the Northeast of Brazil. PLOS Negl Trop Dis. 2016; 10(8): e0004885.

Springer DJ, Billmyre RB, Filler EE, Voelz K, Pursall R, Mieczkowski PA, et al. *Cryptococcus gattii* VGIII isolates causing infections in HIV/AIDS patients in southern California: identification of the local environmental source as arboreal. PLoS Pathog. 2014; 10(8): e1004285.

Taylor JW, Fisher MC. Fungal multilocus sequence typing - it's not just for bacteria. Curr Opin Microbiol. 2003; 6(4): 351-6.

Tibayrenc M, Ayala FJ. *Cryptosporidium*, *Giardia*, *Cryptococcus*, pneumocystis genetic variability: cryptic biological species or clonal near-clades? PLoS Pathog. 2014; 10(4): e1003908.

Vanhove M, Beale MA, Rhodes J, Chanda D, Lakhi S, Kwenda G, et al. Genomic epidemiology of *Cryptococcus* yeasts identifies adaptation to environmental niches underpinning infection across an African HIV/AIDS cohort. Mol Ecol. 2017; 26(7): 1991-2005.