

Harnessing Whole Genome Sequencing in Medical Mycology

Christina A. Cuomo¹

Published online: 9 June 2017
© Springer Science+Business Media New York 2017

Abstract

Purpose of Review Comparative genome sequencing studies of human fungal pathogens enable identification of genes and variants associated with virulence and drug resistance. This review describes current approaches, resources, and advances in applying whole genome sequencing to study clinically important fungal pathogens.

Recent Findings Genomes for some important fungal pathogens were only recently assembled, revealing gene family expansions in many species and extreme gene loss in one obligate species. The scale and scope of species sequenced is rapidly expanding, leveraging technological advances to assemble and annotate genomes with higher precision. By using iteratively improved reference assemblies or those generated de novo for new species, recent studies have compared the sequence of isolates representing populations or clinical cohorts. Whole genome approaches provide the resolution necessary for comparison of closely related isolates, for example, in the analysis of outbreaks or sampled across time within a single host.

Summary Genomic analysis of fungal pathogens has enabled both basic research and diagnostic studies. The increased scale of sequencing can be applied across populations, and new metagenomic methods allow direct analysis of complex samples.

Keywords Whole genome sequencing · Human fungal pathogens · Genome sequencing · Medical mycology · Fungal infections · Review

Introduction

Whole genome sequencing provides a powerful lens for the investigation of fungal pathogens. In providing a comprehensive snapshot of the gene content and thereby the functional potential, the genomic studies of human pathogenic fungal species have revealed the repertoire of proteins that contribute to host interactions [1•, 2, 3•, 4], predicted metabolic capabilities and requirements [1•, 5•], uncovered the potential for sexual reproduction in some species [6, 7], and have been used as the platform to study specific genes as well as for systematic functional genomic approaches [8–10]. For diagnostic purposes, the complete sequence of clinical isolates can type the infecting species and the subgroup or lineage with high accuracy. In addition, the identification of specific mutations that are clinically actionable, including those that confer or promote resistance to antifungal drugs, has the potential to guide treatment decisions for individual patients.

The ease and falling cost of generating whole genome sequence has dramatically expanded the use of this data. The ability to sequence genomes on demand has dramatically expanded the scope of sequenced species, and importantly allows rapid response to new or emerging pathogens. Further, comparing the genome-wide variation of clinical isolates from an outbreak with isolates from environmental or other associated samples can precisely determine how clonal an outbreak is across patients, the identity to isolates from other sources, and establish transmission chains. For recurrent infections or in the context of prolonged outbreaks, genome sequence can trace how pathogens evolve over time, tracking the emergence

This article is part of the Topical Collection on *Advances in Diagnosis of Invasive Fungal Infections*

✉ Christina A. Cuomo
cuomo@broadinstitute.org

¹ Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA

of new genotypes and the spread of particularly virulent or drug-resistant groups.

Whole Genome Sequencing Approaches

There are two general approaches for genomic analysis of fungal pathogens. One involves generation of a genome assembly de novo, such as for a species that has not been previously sequenced and assembled. In the other approach, commonly termed re-sequencing, variants are identified between an existing reference assembly and a sequenced isolate via alignment of sequence reads to the reference. Both methods rely on generating high depth whole genome shotgun sequence, necessary to achieve a high-quality consensus across the genome. However, the choice of technology selected to generate the sequence is influenced both by the approach selected and by the goals of the study. In the initial years of fungal genome sequencing [11], a small number of prioritized species were sequenced using long, paired-end reads produced with Sanger technology from multiple libraries. By contrast, current approaches both for assembly and for variant detection leverage short read sequence such as Illumina and often only a single small insert library. While this approach is sufficient for many goals, incorporating reads from multiple libraries including those with larger inserts and longer reads from technologies such as Pacific Biosciences or Oxford Nanopore are particularly important for accurate assembly of repetitive genomes and for examining structural variation (Fig. 1).

Generation of a new genome assembly for a species, also known as de novo sequencing and assembly (Fig. 1), has been applied to generate reference genome assemblies for all of the major human fungal pathogens (Table 1). Many of these species have now achieved chromosome-scale assemblies, although there may still be gaps at subtelomeres or within each chromosomal sequence often in repetitive or low complexity regions. The choice of a genome sequencing strategy depends both on the properties of the genome and on the goal of the analysis. While the assembly of a short read from a single library can provide a good overview of gene content, repetitive sequences sharing high identity may not be resolved, resulting in gaps in the genome assembly [12]. This issue can be overcome either by the generation of larger insert mate pair libraries for short read sequencing or by the incorporation of longer reads to provide linkage information; highly repetitive genomes may benefit from a strategy of exclusively long read sequencing and assembly. Similar to the prior use of physical or genetic maps in validating and anchoring draft assemblies to chromosomes, technologies such as Hi-C that map the three-dimensional space of a genome also can be used for higher order scaffolding of assemblies [13, 14]. In diploid genomes, heterozygosity may also impact genome assembly;

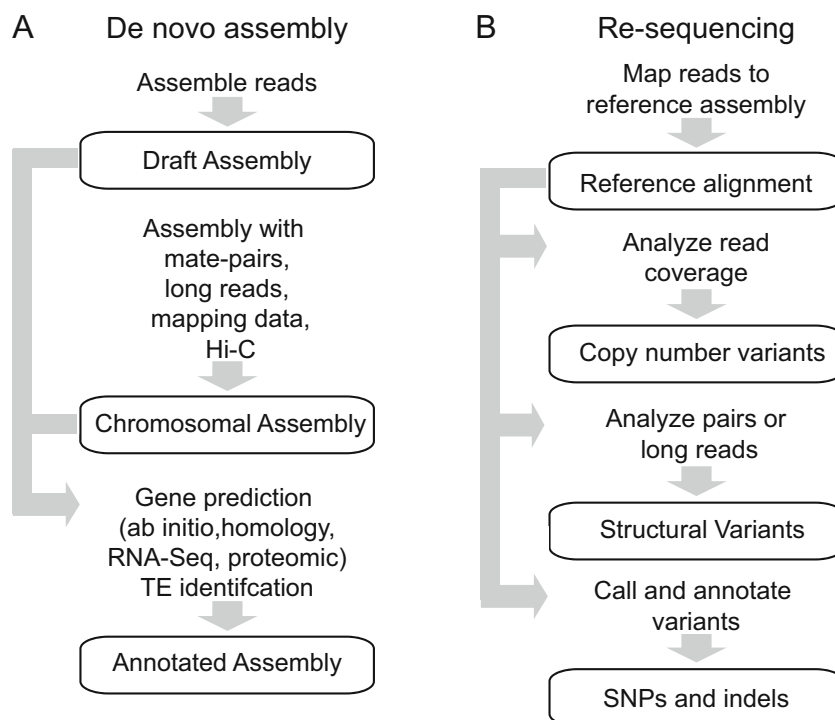
most methods seek to generate a haploid version of the reference, and the generation of a consensus sequence from heterozygous regions can incorrectly merge haplotypes. A phased diploid assembly for *Candida albicans* was constructed using a panel of strains homozygous for specific chromosomes [15•]; a more general approach using long reads was recently reported [16]. De novo assemblies of either haploid or diploid genomes can then be annotated by predicting the structure and function of protein coding genes, using de novo, homology-based, and evidence-based prediction algorithms [17]. The ability to generate deep coverage of RNA-Seq from multiple conditions enables a higher level of accuracy and validation of gene structures, and has been applied to systematically improve gene structures and predict alternatively spliced transcripts.

With the decreased cost of whole genome sequencing, de novo assemblies have been generated on demand to examine new species. This includes representing both rarely observed pathogens and many nonpathogenic species related to common pathogens, with the goal of using comparative genomic approaches to identify differences that could contribute to pathogenesis. The ability to rapidly generate genomes for new species is also important for the response to emerging pathogens and in the context of recent fungal outbreaks. Recent studies have also expanded our view within a single species by examining the genome of more than one “reference” isolate for some species, which can characterize differences in gene content between isolates of the same species. From a larger perspective, the increasing number and diversity of sequenced genomes enable a wide range of studies focused on comparisons of specific genes, as well as a set of references for alignment-based approaches including both metagenomics and sequencing of single isolates.

For species for which a high-quality reference assembly is available, re-sequencing is an alternative approach to identify genome-wide variants. Typically short read sequence is generated from one or more isolates of the same species, reads are aligned to a reference assembly, and high-quality variants identified from the alignments (Fig. 1). These methods have been applied to both haploid and diploid fungal genomes. The full-genome resolution and scalability of this approach make it ideal for examining transmission links and in the context of an outbreak and pathogen evolution during the course of an infection, during which few variants may be expected. Variants can also be mapped to genes, on the reference genome, to infer changes in important genes involved in drug resistance (see below).

In addition to these approaches that rely on sequencing of individual isolates, the increase of metagenomic sequencing has driven the development of methods to look directly at populations within a single sample. Also using a shotgun sequencing approach, the sequence of a pool of samples can be used to determine the species within a single sample, to categorize the gene content to suggest functional capacity, and recently to examine species level variation.

Fig. 1 Overview of whole genome sequencing approaches. **a** De novo assembly approach; while draft assemblies are more fragmented than finished or chromosomal assemblies, both can be annotated for gene structures and repetitive elements. **b** Re-sequencing approach; this approach starts with initial alignment of sequence reads to a reference assembly, with separate processes required to identify copy number variants, structural variants, and SNPs and indels



Recent Genome Sequencing Findings

With the advent of highly multi-parallel sequencing, the increased ease and low cost of generating whole genome sequence led to a dramatic expansion of the number of fungal genomes available. Notably, the 1000 fungal genomes project at the US Department of Energy Joint Genome Institute (<http://1000.fungalgenomes.org>) aims to provide a comprehensive representation of the fungal kingdom, where each family level division would be represented by at least two genomes. The pace of sequencing is already eclipsing the scale of this project, with over 2100 fungal genome assemblies available in NCBI (<https://www.ncbi.nlm.nih.gov/genome/browse/#>). Of these, only 812 have gene annotations deposited in NCBI, highlighting the more limited scope of easily available gene sets for comparative analysis.

Recent years have seen advances in the generation or improvement of reference genomes for the major human fungal pathogens (Table 1). The assembly of *Cryptococcus neoformans* var. *grubii* (serotype A) into 14 chromosomes incorporated deep RNA-Seq for annotation, allowing assessment of the structures of nearly all genes to provide a comprehensive view of coding and noncoding transcripts [24]. Building on the initial genome sequence of two lineages of *Cryptococcus gattii* [25], genome comparison across 16 genome assemblies representing all four predominant lineages of *C. gattii* identified variation in gene content including RNAi, iron-binding, and stress-related genes and selection

pressure of transporters [25, 26]. Sequencing of the obligate pathogenic *Pneumocystis* species required optimization of strategies to purify fungal from host pulmonary tissue and iteration of sequencing strategy to improve assemblies [1•, 5•]. The high copy major surface glycoprotein gene family, which encodes the most abundant cell surface protein in *Pneumocystis*, was not well represented in assemblies of short read data [5•] but improved using longer PacBio reads [1•]. Building on single genome studies of the species causing mucormycosis [32], a comparative genomic study of 30 species of *Rhizopus* and *Mucor* revealed that the CotH invasin gene family is a unique feature of all invasive Mucorales and that copy number appears to correlate with species prevalence [3•]. These and other recent studies collectively provide reference genomes for the study of all the major human fungal pathogens (Table 1).

In addition to studies focusing on a single genome as representative of a species, multiple studies have used re-sequencing to examine variation across multiple isolates of a single species. Large studies of *C. gattii* have characterized the relationship of global isolates [34, 35] and identified a loss of function mutation in the mismatch repair gene *MSH2* in one sublineage, VGIIa [35]. One of the largest studies to date in *C. neoformans* var. *grubii* compared the sequence of 387 isolates from clinical and environmental origin; genome-wide association study (GWAS) variants associated with the isolation source identified virulence factors and stress response genes [36•]. A parallel GWAS of melanization in these isolates identified loss of function mutation in clinical isolates in

Table 1 Genome assemblies for human fungal pathogens

Species	Genome size (Mb)	Assembly state	References
<i>Aspergillus fumigatus</i>	29	Chromosomal, draft	[18, 19]
<i>Blastomyces dermatitidis</i>	67–75	Draft	[20]
<i>Candida albicans</i> , <i>Candida glabrata</i>	12–15	Chromosomal, diploid	[4, 15, 21, 22]
<i>Coccidioides immitis</i>	29	Chromosomal	[23]
<i>Cryptococcus neoformans</i> , <i>Cryptococcus gattii</i>	17–19	Chromosomal, draft	[24–27]
<i>Histoplasma capsulatum</i>	30–41	Draft	[28]
<i>Paracoccidioides brasiliensis</i>	29–30	Draft	[2, 29]
<i>Penicillium marneffeii</i>	29	Draft	[30, 31]
<i>Pneumocystis jirovecii</i>	8	Chromosomal	[1, 5]
<i>Rhizopus oryzae</i> , <i>Rhizopus delemar</i>	39–46	Draft	[3, 32]

Species representing the ten most significant invasive fungal infections, adapted from Brown et al. [33]

the *BZP4* transcription factor required for melanin production [36•]; while melanin is a virulence factor in *Cryptococcus*, the presence of multiple loss of function mutations in clinical isolates suggests that loss of melanin production is observed clinically. A study of a panel of 20 clinical isolates of *C. albicans* characterized frequent loss of heterozygosity and pinpointed a loss of function of *EFG1*, a gene required for filamentous growth; this isolate further showed a competitive advantage during gastrointestinal growth over isolates that were isogenic except for the addition of a wild type copy of this gene, suggesting that this change could have provided an advantage during commensal growth [37]. Additional studies in dimorphic fungi have defined new population subdivisions and the level of genetic exchange between these groups [2, 38].

Refining Phylogenetic Relationships

Genome sequence is also utilized to assess phylogenetic relationships between species and has helped resolve inconsistencies in species naming. The Phylogenetic Species Concept requires consistency across multiple gene trees [39], as single genes could be subject to recombination or introgression and not reflect the true species relationships. This can highlight conflicts in the naming of genera or species grouped by morphological and phenotypic information and suggest how to refine species boundaries. However, where to set species boundaries and the decision of what evidence justifies changes in species naming is debated [40, 41]. More straightforward cases are those where phylogenetic analysis highlighted inconsistencies in the current genus naming; a re-assessment of the *Emmonsia* genus including many newly reported clinical cases [42] led to a proposed re-organization of the taxonomy of this group including a new genus name [43]. Such assessments incorporate analysis of the support for phylogenetic subdivisions and the genetic distance between groups.

While phylogenies based on whole genome data may capture the same major phylogenetic relationships and subdivisions as those observed in phylogenies based on small numbers of loci, analysis of whole genome data provides a more comprehensive view of genetic exchange between subdivisions. For example, in *C. neoformans*, four well-separated lineages (VNI, VNII, VNB-I, and VNB-II) in whole genome phylogenies appear similar to those identified in multi-locus phylogenies, and at a finer scale, subdivision of VNI into three subgroups is also strongly supported from phylogenetic analysis of whole genome data. However, while recombination is limited between the four lineages, the level of recombination appears similar across VNI as within each subgroup, suggesting that the phylogenetic subdivisions within VNI do not reflect genetic isolation [36•]. Such analyses of level of genetic exchange and separation can help support or question subdivisions made based only on multi-gene phylogenies. In addition, these studies can highlight unusual phylogeographic patterns, which can motivate further population sampling to evaluate rare or unexpected subgroups.

Outbreaks and Emerging Species

Multiple species of fungi have been responsible for major outbreaks of infections in the USA within the last 10 years. In contrast to the predominant species that cause of human fungal infections, many outbreaks have resulted from organisms that are not a common cause of infection, and consequently some of these species are not well studied or previously sequenced. For such cases, a primary goal for whole genome sequencing has been to generate a reference genome that could be used for identification of genome-wide variants across outbreak samples as well as for further genomic and transcriptomic studies of pathogenesis. Comparing the genomes of patient and environmental isolates from populations of these pathogens can help trace the origin and transmission patterns in an outbreak; if

isolates from an outbreak and potential source show very few genome-wide differences, this supports a clonal outbreak mechanism with strong link to the potential source. In addition, the gene set predicted from the genome of outbreak isolates can help develop biomarkers and new diagnostics, and can potentially guide our understanding of what enabled a strain to cause a suddenly high rate of severe infections.

One major treatment-acquired fungal outbreak in the USA resulted from injection of methylprednisolone, as a treatment for pain management, contaminated with the phaeoid fungus *Exserohilum rostratum*. As of January 2013, *E. rostratum* had caused more than 750 cases of phaeohyphomycotic meningitis and at least 61 deaths in 19 US states [44]. A very similar but smaller fungal outbreak occurred 10 years previously, caused by a steroid contamination with *Wangiella (Exophiala) dermatitidis* [45]. Both species of phaeoid fungi (black or dark brown pigmented) are infrequently the cause of superficial infections, however in rare cases they result in systemic neurotropic infections. Whole genome sequencing of *E. rostratum* purified from clinical samples from patients injected with contaminated steroids and from steroids lots established that a clonal fungus was present in both patients and steroid lots [46]. This analysis incorporated both de novo and re-sequencing approaches (Fig. 1): a reference assembly was generated from for one of the outbreak strains, and SNPs were identified by aligning reads from all other samples to this assembly. Analysis of variants revealed that genomes from outbreak isolates were nearly identical, both from 19 patients and 6 from steroid lots; only two SNPs were found between any isolate from a patient and compounding vial. By contrast, over 136,000 SNPs differentiated the outbreak isolates from other environmental isolates, though these were collected in years prior to the outbreak in different geographic regions. This genomic analysis provided strong evidence that the fungal strains found in all patients and in the suspected steroid vials were identical.

One recent report demonstrated how whole genome sequencing can pinpoint isolate relationships and suggest new transmission patterns. To determine whether clinical cases of Coccidioidomycosis in Washington State were the first reports of local exposure or resulted from transmission during travel to the southwestern USA, patient isolates were compared with environmental isolates from the local area in Washington and from the southwestern USA. Remarkably, whole genome sequencing revealed that *Coccidioides immitis* isolates from these patient cases were nearly identical to local soil isolates, differing by only three SNPs across the entire 28 Mb genome [47], suggestive of local transmission and a potentially an expanded endemic area for this pathogen.

More recently, genomic analysis of drug resistance *Candida auris* established that patient isolates from specific geographic regions are highly identical [48, 49]. In one study, a de novo genome assembly was generated for one isolate of *C. auris*, and SNPs identified in other isolates using the re-sequencing

approach [48]. While isolates from a given geographic area appear closely related, there is more variation between regions; in addition, drug-resistant isolates show candidate-resistant mutations in *ERG11*, based on mapping such sites from *C. albicans*.

Outbreaks may also occur when there is a change in a pathogen that increases resistance to stress conditions or enables survival in a new environment. Mutations in the mismatch repair gene *MSH2* initially identified in the outbreak lineage of *C. gattii* may enable more rapid adaptation to stressful conditions or new environments by allowing a higher rate of mutation [35]. Loss of *MSH2* has also been detected in *Candida glabrata*, where this appears to accelerate the acquisition of drug resistance [50]; however another recent study failed to find a correlation between MSH mutation and azole resistance [Dellière S et al. Fluconazole and Echinocandin Resistance of *Candida glabrata* Correlates Better with Antifungal Drug Exposure Rather than with *MSH2* Mutator Genotype in a French Cohort of Patients Harboring Low Rates of Resistance. Front Microbiol. 2016 Dec 23;7:2038.], suggesting that other factors in the genetic background could play a role in drug resistance. These studies suggest that loss of DNA repair genes in fungi could result in a higher mutation rate may provide an advantage under adverse conditions.

Evolution Within Patients

Where fungal infections persist in patients, studies of how the genome changes during chronic infection can highlight mechanisms of adaptation. Recent studies of *C. albicans* and *Cryptococcus* have used whole genome re-sequencing to identify how isolates of these species change during infection. One recent study examined serial isolates of *C. albicans* from 11 patients with oral candidiasis; this revealed that during clinical passage, isolates acquired new mutations, including some linked to host adaptation [51]. In addition, genome regions showing loss of heterozygosity during passage include genes implicated in drug resistance. Another study of serial isolates of *C. neoformans* and *gattii* compared isolates during initial presentation of disease and after 120 days or more during a relapse [52]. These cases were also highly clonal, demonstrating that a second independently infecting strain was not the origin of the relapse. The lower rate is consistent with a prior report; however, higher rates of change in some isolates in a separate study were suggested to result from changes in mismatch repair proteins [53]. Analysis of wider sets of isolates is needed to validate whether such mutations are common in *Cryptococcus*.

Antifungal Resistance

Genome sequencing can type known drug resistance mutations, in some cases suggesting whether particular drugs will

fail to control an infection. Whole genome variants could be screened for point mutations in specific drug targets that are highly correlated with resistance. For example, specific mutations in the target of azole drugs [54] or in the transcription factors that control the expression of drug efflux transporters [55, 56] can be identified from whole genome sequence data only in isolates that display drug resistance [37]. In addition, copy number variation of both drug targets and transporters can also lead to drug resistance; genomic regions showing higher sequencing read depth for such genes are also associated with drug resistance [57]. Where this method could be applied to metagenomic population sequencing, it may be possible to detect early arising drug-resistant mutants that have not yet swept through the population, and precede treatment failure.

Diagnostics

Direct genome sequencing of microbial samples can provide precise diagnostic information. With approaches that meet the need of rapid turnaround for clinical samples, similar methods can be applied to complex samples that contain multiple microbes in addition to human DNA. While analyzing the sequence of a set of organisms requires different approaches than single isolate sequencing, specialized methods can identify the potential pathogen sequence in metagenomic data from a mixed population [58, 59]. Metagenomic data can also be used to examine how much the sequence of a specific species varies within a single sample, separating out the contribution of different strains involved in mixed infections [60, 61]. With sufficient coverage or target enrichment, metagenomic sequence can also provide precise typing of specific mutations of clinical impact, such as those that promote or provide drug resistance. In this way, such approaches can tailor treatments to patients, particularly those at high risk of invasive infections, by using genomic data to predict how a patient will respond to a specific treatment.

Conclusions

Genome sequencing is becoming an increasingly common approach to study human fungal pathogens. Current studies seek to compare the genomes of hundreds of isolates of a given species and to examine pathogen populations at unprecedented scale. Sequencing is no longer a bottleneck; however, analysis approaches also need to continue to scale. Genome data offers the resolution needed to examine microevolution of isolates during the course of infection and to pinpoint the source and transmission networks involved in outbreaks. The use of genomic approaches in diagnosis may become more routine and offers the potential to provide additional clinically

actionable information such predictions of the level or potential for drug resistance. Even prior to some treatments, metagenomic information may identify fungi and other microbes that are a cause for concern. While real challenges exist to the wide implementation of microbial sequencing methods in the clinic, including turnaround time, ease of use and interpretation, and cost, these approaches enable a high-resolution view of the specific fungal isolates causing disease.

Compliance with Ethical Standards

Conflict of Interest Christina A. Cuomo was supported by Federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Grant Number U19AI110818 to the Broad Institute.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by the author.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance

1. Ma L, Chen Z, Huang DW, Kutty G, Ishihara M, Wang H, et al. Genome analysis of three *Pneumocystis* species reveals adaptation mechanisms to life exclusively in mammalian hosts. *Nat Commun*. 2016;7:10740. **Chromosomal assemblies of *Pneumocystis jirovecii*, *P. murina*, and *P. carinii*. Comparative analysis revealed loss of enzymes responsible for chitin and higher order mannose synthesis, suggesting a more flexible cell wall structure that may also impact detection by the host immune system**
2. Muñoz JF, Farrer RA, Desjardins CA, Gallo JE, Sykes S, Sakthikumar S, et al. Genome diversity, recombination, and virulence across the major lineages of *Paracoccidioides*. *mSphere*. 2016;1.
3. Chibucos MC, Soliman S, Gebremariam T, Lee H, Daugherty S, Orvis J, et al. An integrated genomic and transcriptomic survey of mucormycosis-causing fungi. *Nat Commun*. 2016;7:12218. **Large comparative study of mucorales pathogens**
4. Braun BR, van Het Hoog M, d'Enfert C, Martchenko M, Dungan J, Kuo A, et al. A human-curated annotation of the *Candida albicans* genome. *PLoS Genet*. 2005;1:36–57.
5. Cissé OH, Pagni M, Hauser PM. De novo assembly of the *Pneumocystis jirovecii* genome from a single bronchoalveolar lavage fluid specimen from a patient. *MBio*. 2012;4:e00428–12. **First genome assembly for *Pneumocystis jirovecii*, utilizing purification and host sequence filtering to assemble this obligate pathogen**
6. Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*. 2005;438:1105–15.
7. O’Gorman CM, Fuller HT, Dyer PS. Discovery of a sexual cycle in the opportunistic fungal pathogen *Aspergillus fumigatus*. *Nature*. 2009;457:471–4.
8. 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;135:174–88.

9. Roemer T, Jiang B, Davison J, Ketela T, Veillette K, Breton A, et al. Large-scale essential gene identification in *Candida albicans* and applications to antifungal drug discovery. *Mol Microbiol.* 2003;50:167–81.
10. Vyas VK, Barrasa MI, Fink GR. A *Candida albicans* CRISPR system permits genetic engineering of essential genes and gene families. *Sci Adv.* 2015;1:e1500248.
11. Cuomo CA, Birren BW. The fungal genome initiative and lessons learned from genome sequencing. *Methods Enzymol.* 2010;470:Chapter 34.
12. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet.* 2011;13:36–46.
13. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat Biotechnol.* 2013;31:1119–25.
14. Lieberman-Aiden, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* 2009;326:289–293.
15. Muzzey D, Schwartz K, Weissman JS, Sherlock G. Assembly of a phased diploid *Candida albicans* genome facilitates allele-specific measurements and provides a simple model for repeat and indel structure. *Genome Biol.* 2013;14:R97. **Production of a phased diploid assembly for *Candida albicans*, enabling allele-specific analysis**
16. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* 2016;13:1050–4.
17. Haas BJ, Zeng Q, Pearson MD, Cuomo CA, Wortman JR. Approaches to fungal genome annotation. *Mycology.* 2011;2:118–41.
18. Fedorova ND, Khaldi N, Joardar VS, Maiti R, Amedeo P, Anderson MJ, et al. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. *PLoS Genet.* 2008;4:e1000046.
19. Nierman WC, Pain A, Anderson MJ, Wortman JR, Kim HS, Arroyo J, et al. Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*. *Nature.* 2005;438:1151–6.
20. Muñoz JF, Gauthier GM, Desjardins CA, Gallo JE, Holder J, Sullivan TD, et al. The dynamic genome and transcriptome of the human fungal pathogen *Blastomyces* and close relative *Emmonsia*. *PLoS Genet.* 2015;11:e1005493.
21. van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H, et al. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. *Genome Biol.* 2007;8:R52.
22. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, et al. Genome evolution in yeasts. *Nature.* 2004;430:35–44.
23. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Joardar VS, et al. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. *Genome Res.* 2009;19:1722–31.
24. 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:e1004261.
25. 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.
26. 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:e00868–15.
27. 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.* 2005;307:1321–4.
28. Histoplasma Genome Project | Broad Institute [Internet]. [cited 2017 Mar 21]. Available from: <https://www.broadinstitute.org/fungal-genome-initiative/histoplasma-genome-project>
29. Desjardins CA, Champion MD, Holder JW, Muszewska A, Goldberg J, Bailao AM, et al. Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. *PLoS Genet.* 2011;7:e1002345.
30. Nierman WC, Fedorova-Abrams ND, Andrianopoulos A. Genome sequence of the AIDS-associated pathogen *Penicillium marneffei* (ATCC18224) and its near taxonomic relative *Talaromyces stipitatus* (ATCC10500). *Genome Announc.* 2015;3.
31. Woo PCY, Lau SKP, Liu B, Cai JJ, Chong KTK, Tse H, et al. Draft genome sequence of *Penicillium marneffei* strain PM1. *Eukaryot Cell.* 2011;10:1740–1.
32. Ma LJ, Ibrahim AS, Skory C, Grabherr MG, Burger G, Butler M, et al. Genomic analysis of the basal lineage fungus *Rhizopus oryzae* reveals a whole-genome duplication. *PLoS Genet.* 2009;5:e1000549.
33. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. *Sci Transl Med.* 2012;4:165rv13.
34. 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:e01464–14.
35. 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:e01494–14.
36. Desjardins C, Giamberardino C, Sykes S, Yu C-H, Tenor J, Chen Y, et al. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. *bioRxiv.* 2017;118323. **Large population genomic study in *Cryptococcus* establishing the major subdivisions in the populations and associated phenotypes and application of GWAS to identify variants associated with clinical origin and phenotypes.**
37. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome Res.* 2015;25:413–25.
38. Engelthaler DM, Roe CC, Hepp CM, Teixeira M, Driebe EM, Schupp JM, et al. Local population structure and patterns of western hemisphere dispersal for *Coccidioides* spp., the fungal cause of valley fever. *MBio.* 2016;7:e00550–16.
39. Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, et al. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol.* 2000;31:21–32.
40. 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.
41. 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.
42. Schwartz IS, Kenyon C, Feng P, Govender NP, Dukik K, Jiang Y, et al. 50 years of *Emmonsia* disease in humans: the dramatic emergence of a cluster of novel fungal pathogens. *PLoS Pathog.* 2015;11:e1005198.
43. Dukik K, Muñoz JF, Jiang Y, Feng P, Sigler L, Stielow JB, et al. Novel taxa of thermally dimorphic systemic pathogens in the Ajellomycetaceae (Onygenales). *Mycoses.* 2017.

44. Andes D, Casadevall A. Insights into fungal pathogenesis from the iatrogenic epidemic of *Exserohilum rostratum* fungal meningitis. *Fungal Genet Biol* FG B. 2013;61:143–5.
45. Centers for Disease Control and Prevention. *Exophiala* infection from contaminated injectable steroids prepared by a compounding pharmacy—United States, July–November 2002. *Morb Mortal Week Rep*. 2002;51:1109–12.
46. Litvintseva AP, Hurst S, Gade L, Frace MA, Hilsbeck R, Schupp JM, et al. Whole-genome analysis of *Exserohilum rostratum* from an outbreak of fungal meningitis and other infections. *J Clin Microbiol*. 2014;52:3216–22.
47. Litvintseva AP, Marsden-Haug N, Hurst S, Hill H, Gade L, Driebe EM, et al. Valley fever: finding new places for an old disease: *Coccidioides immitis* found in Washington State soil associated with recent human infection. *Clin Infect Dis Off Publ Infect Dis Soc Am*. 2015;60:e1–3.
48. Lockhart SR, Etienne KA, Vallabhaneni S, Farooqi J, Chowdhary A, Govender NP, et al. Simultaneous emergence of multidrug-resistant *Candida auris* on 3 continents confirmed by whole-genome sequencing and epidemiological analyses. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am*. 2017;64:134–40. **Population genomic study of *Candida auris* outbreak isolates and characterization of geographic substructure and the basis of drug resistance; example of application of a de novo assembly and re-sequencing approach**
49. Sharma C, Kumar N, Pandey R, Meis JF, Chowdhary A. Whole genome sequencing of emerging multidrug resistant *Candida auris* isolates in India demonstrates low genetic variation. *New Microbes New Infect*. 2016;13:77–82.
50. Healey KR, Zhao Y, Perez WB, Lockhart SR, Sobel JD, Farmakiotis D, et al. Prevalent mutator genotype identified in fungal pathogen *Candida glabrata* promotes multi-drug resistance. *Nat Commun*. 2016;7:11128.
51. Ford CB, Funt JM, Abbey D, Issi L, Guiducci C, Martinez DA, et al. The evolution of drug resistance in clinical isolates of *Candida albicans*. *elife*. 2015;4:e00662. **Initial study of microevolution and the acquisition of new phenotypes in serial isolates of *Candida albicans***
52. 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:e00166–17.
53. 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 Md. 2017.
54. White TC. The presence of an R467K amino acid substitution and loss of allelic variation correlate with an azole-resistant lanosterol 14alpha demethylase in *Candida albicans*. *Antimicrob Agents Chemother*. 1997;41:1488–94.
55. Dunkel N, Blass J, Rogers PD, Morschhäuser J. Mutations in the multi-drug resistance regulator MRR1, followed by loss of heterozygosity, are the main cause of MDR1 overexpression in fluconazole-resistant *Candida albicans* strains. *Mol Microbiol*. 2008;69:827–40.
56. Coste A, Turner V, Ischer F, Morschhäuser J, Forche A, Selmecki A, et al. A mutation in *Tac1p*, a transcription factor regulating CDR1 and CDR2, is coupled with loss of heterozygosity at chromosome 5 to mediate antifungal resistance in *Candida albicans*. *Genetics*. 2006;172:2139–56.
57. Sionov E, Lee H, Chang YC, Kwon-Chung KJ. *Cryptococcus neoformans* overcomes stress of azole drugs by formation of disomy in specific multiple chromosomes. *PLoS Pathog*. 2010;6:e1000848.
58. Kostic AD, Ojesina AI, Pedamallu CS, Jung J, Verhaak RGW, Getz G, et al. PathSeq: software to identify or discover microbes by deep sequencing of human tissue. *Nat Biotechnol*. 2011;29:393–6.
59. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, et al. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res*. 2014;24:1180–92.
60. Morowitz MJ, Denef VJ, Costello EK, Thomas BC, Poroyko V, Relman DA, et al. Strain-resolved community genomic analysis of gut microbial colonization in a premature infant. *Proc Natl Acad Sci U S A*. 2011;108:1128–33.
61. Truong DT, Tett A, Pasolli E, Huttenhower C, Segata N. Microbial strain-level population structure and genetic diversity from metagenomes. *Genome Res*. [Internet]. 2017 [cited 2017 Mar 14]; Available from: <http://genome.cshlp.org/content/early/2017/03/10/gr.216242.116>