Evolutionary Perspectives on Human Fungal Pathogens

John W. Taylor

University of California, Berkeley, California 94720-3102
Correspondence: jtaylor@berkeley.edu

Owing to their small size and paucity of phenotypic characters, progress in the evolutionary biology of microbes in general, and human pathogenic fungi in particular, has been linked to a series of advances in DNA sequencing over the past quarter century. Phylogenetics was the first area to benefit, with the achievement of a basic understanding of fungal phylogeny. Population genetics was the next advance, finding cryptic species everywhere, and recombination in species previously thought to be asexual. Comparative genomics saw the next advance, in which variation in gene content and changes in gene family size were found to be important sources of variation. Fungal population genomics is showing that gene flow among closely related populations and species provides yet another source of adaptive, genetic variation. Now, two means to associate genetic variation with phenotypic variation, “reverse ecology” for adaptive phenotypes, and genome-wide association of any phenotype, are letting evolutionary biology make a profound contribution to molecular developmental biology of pathogenic fungi.

The modern study of fungal evolution began with the access to nucleic acid variation enabled by the invention of the polymerase chain reaction (PCR) in 1988 (Saiki et al. 1988) because medically important fungi, apart from poisonous mushrooms, are small and provide too few morphological or physiological phenotypic characters for effective comparative evolutionary biology (Taylor et al. 2006). Their microscopic nature makes all the more remarkable Agustino Bassi’s discovery in 1835 of the first microbe causing an animal disease, the Ascomycota, *Beauveria bassiana*, on silkworms (Porter 1973), a discovery in advance of Schwann’s discovery of yeast cells in beer in 1837 (Schwann 1837), Pasteur’s research on the microbiology of wine in 1866 (Pasteur 1866), and Koch’s postulates on proving the microbial origin of disease in 1890 (Grimes 2006).

Perusal of Rippon’s 1988 text on medical mycology demonstrates the challenges faced by evolutionary mycologists of that era (Rippon 1988). Rippon mentions many phylogenetic mysteries, including *Coccidioides immitis*, which today is known to be an Ascomycota despite never having revealed its sexual reproduction to mycologists. Rippon wrote, “Thus, we have a fungus (*C. immitis*) that has characteristics of a Zygomycete in its parasitic stage and is gymnoasaceous in its saprophytic stage. The problem will be fully settled only when the [sexual] state of *C. immitis* is found.” There also are
complete surprises, for example, the Ascomycota, *Pneumocystis*, is not even mentioned by Rippon, because it was considered to be a protist until nucleic acid sequence was obtained for its ribosomal RNA (Stringer et al. 1989).

The evolutionary biology of medically important fungi, along with fungi in general, has advanced on four fronts, each launched by an advance in access to DNA sequence variation: comparative phylogenetics launched by the aforementioned PCR, population genetics launched by the transition from manual to automated, Sanger, DNA sequencing, comparative genomics of fungi spurred by the human genome project, and population genomics launched by next-generation sequencing (NGS). This article will be organized to follow these four fronts to note the basic evolutionary biology that has been learned from each effort, with an emphasis on medically important fungi. Although the phylogenetic and population fields launched at different times, each continues to thrive, now relying on NGS for most applications. Recent reviews of medical mycology also contain material on the epidemiology (Pfaller and Diekema 2010), extent and origin (Heitman 2011), and evolution (Taylor 2006) of fungal pathogens.

**MOLECULAR PHYLOGENETICS**

Mycology followed the lead of bacteriologists in using ribosomal RNA sequences for molecular phylogenetics because kingdom-wide primers could be designed to match ultraconserved ribosomal regions that flanked the phylogenetically informative, variable regions (White et al. 1990). Protein coding genes were added to the studies, principally proteins associated with DNA replication, ribosomal proteins, and elongation factors, as well as an ATP subunit (James et al. 2006). The rDNA regions continue to be useful as fungal, molecular barcodes (Schoch et al. 2012) and for ecological studies in which fungi are identified from DNA isolated from environmental samples (Adams et al. 2013).

Phylogenetics of medically important fungi had to be studied in the context of all fungi, because fungi capable of parasitizing humans come from throughout the fungal kingdom, especially those pathogens seen in patients whose immune system had been compromised through disease or suppressed to avoid rejection of transplanted bone marrow or solid organs. In fact, there probably has been no more profound effect on the field of medical mycology, and the study of fungal molecular phylogenetics in general, than the recognition in 1987 that infections with certain fungi were AIDS-defining diseases (CDC 1987). These opportunistic fungal pathogens included species of *Candida*, *Cryptococcus*, and *Pneumocystis*, in addition to the long-recognized, systemic pathogens, *Coccidioides* and *Histoplasma*, and were joined shortly by *Penicillium marneffii*. In 2000, the *Atlas of Clinical Fungi* (de Hoog et al. 2000) contained 314 additional species, almost all of them opportunistic pathogens and none more important to patients in transplant or cancer therapy than *Aspergillus* species, principally *Aspergillus fumigatus*.

The most complete molecular phylogeny of fungi has come from the AFTOL project with publications on the entire kingdom (James et al. 2006) and individual phyla (see Mycologia, 2006, Vol. 98, No. 6). Most animal parasitic fungi are Ascomycota and two clades of filamentous Ascomycota (Pezizomycotina) are particularly rich in pathogens, the Hypocreales and the Onygenales (Fig. 1). The Hypocreales is home to parasites of insects, including Bassi’s silkworm pathogen, the Zombie Ant fungus, *Ophiocordyceps unilateralis*, expert manipulator of host behavior (Evans et al. 2011; Hughes et al. 2011), and the human pathogen, *Pseudallescheria*. The Onygenales embrace the human and animal dermatophytes, including *Trichophyton* species, as well as fungi causing systemic disease of mammals, that is, the systemic pathogens, *Coccidioides*, *Histoplasma*, *Blastomyces*, and *Paracoccidioides*. These dangerous, BSL3 containment pathogens and their fellow Eurotiomycetes, *Aspergillus* and *Penicillium*, are joined by fungi from other clades of filamentous Ascomycota, including *Exophiala*, *Sporothrix*, and *Neotestudina* species. The most important genus of fungal pathogens in North America, the yeast *Candida*, represents the Saccharo-
mycotina and the third Ascomycota subphylum, Taphrinomycotina, harbors the aforementioned *Pneumocystis*. 

Apart from Ascomycota, the Basidiomycota harbors the next most important agents of human disease, *Cryptococcus* species that can be opportunistic pathogens of immense importance to AIDS patients (i.e., *Cryptococcus neoformans*), or emerging pathogens of otherwise healthy hosts (i.e., *Cryptococcus gattii*). In the same subphylum of Basidiomycota, *Trichosporon*, agent of dermatomycoses, and in the subphylum, Ustilaginomycotina, is found *Malassezia*, skin commensal associated with dandruff. The human, opportunistic pathogens *Mucor, Rhizopus*, and *Basidiobolus* are found in clades of zygomycetes, and at the very base of the Kingdom Fungi emerge the Microsporidia, a species that also causes human disease (Didier and Weiss 2011).

Other basal, fungal phyla harbor animal pathogens (i.e., Blastocladiales and Chytridiomycota) (home to *Batrachochytrium dendrobatidis*, scourge of amphibians), but no human pathogens. Outside the fungal kingdom, several groups traditionally studied by mycologists harbor species capable of causing severe human disease (e.g., the phylogenetically close Mesomyctozoa [Ichthyosporae] contains *Rhinosporidium* (Mendoza et al. 2002) and the distant Oomycota harbors *Pythium insidiosum* (Mendoza and Newton 2005).

Examination of phylogenies of these fungi, which can be fit to a geological timescale (albeit with significant uncertainty) (Berbee and Taylor 2010), has led to formulation of general hypotheses about the evolution of fungal pathogens. Among them is the observation that adaptation to parasitize amoebae may have prepared fungi to live in amoebae-like host cells devoted to

---

**Figure 1.** Phylogeny of fungi. Most human pathogenic fungi are Ascomycota, with *Candida* species in Saccharomycotina, *Pneumocystis* in Taphrinomycotina, and the Onygenales clade of systemic pathogens and dermatophytes in Pezizomycotina. Basidiomycotina also harbors important pathogens, *Cryptococcus* and *Trichosporon* species in Agaricomycotina and *Malassezia* species in Ustilaginomycotina. (From Stajich et al. 2009; reprinted, with permission, from the authors. Based on approaches and phylogenies in James et al. 2006 and Taylor and Berbee 2006.)
immunological response (Steenbergen et al. 2001) and that the kingdom-wide scarcity of thermotolerant fungi may have been a selective force in the evolution of endothermy in mammals and birds (Casadevall 2005; Robert and Casadevall 2009; Taylor and Gurr 2014).

POPULATION GENETICS

Just as PCR combined with manual DNA sequencing made comparative phylogenomics possible, PCR combined with semiautomated DNA sequencing, whether by gels or gel-filled capillaries, made population genetics possible. The ability to study tens to hundreds of isolates from one morphological species discovered hidden genetic variation that led to fundamental changes in how we recognize fungal species and how we view fungal reproduction.

Biological species recognition, the gold standard for species recognition in plants and animals, had little impact on fungal systematics in general and medical mycology in particular because it required experimental mating. To be sure, biological species recognition was practiced with model fungal organisms, among them, the Ascomycota Saccharomyces cerevisiae and Neurospora crassa, and the Basidiomycota, Schizophyllum commune, Heterobasidion annosum and Armellaria mellea; however, explicitly excluded were the ca. 20% of fungi for which sexual reproduction had not been observed. Included in this group are many medically important fungi (e.g., Candida albicans, A. fumigatus, Coccioidoides immitis, and Sporothrix schenckii), as well as human pathogens for which sexual reproduction is difficult to achieve in the laboratory (e.g., Histoplasma or Blastomyces). Like countries that have leapt from inadequate landline telephone service directly to cell phones, mycology went directly from morphological species recognition to phylogenetic species recognition, which led to the discovery that almost all morphological species of medically important fungi harbored more than one phylogenetic species, typically termed “cryptic species,” because they had not been recognized morphologically. Thus, C. albicans also had Candida dubliniensis (Sullivan et al. 1995), Coccioidoides immitis also had C. posadasii (Fisher et al. 2002), C. neoformans also had C. grubii, albeit described as a variety (Franzot et al. 1999), Paracoccidioides brasiliensis also had P. lutzii (Matute et al. 2006; Teixeira et al. 2014), Blastomyces dermatidis also had B. gilchristii (Brown et al. 2013), and Histoplasma capsulatum had at least six additional species (Kasuga et al. 2003), as well as a new group recently found to be associated with cats (Balajee et al. 2013). Even model, biological species (e.g., S. cerevisiae and N. crassa), were found to have additional, phylogenetic species, among which mating was still possible (e.g., Saccharomyces paradoxus [Martini 1989] or Neurospora perkinsii [Villalta et al. 2009]). It is unusual to find a medically important, morphological species that has just one phylogenetic species. A. fumigatus is one example of a globally distributed phylogenetic species, despite being studied by two different groups, who used different collections of isolates and different variable DNA regions (Paolletti et al. 2005; Pringle et al. 2005; Rydholm et al. 2006). This opportunistic pathogen is ubiquitous in soils and, as a thermophile, in compost. Perhaps it has been able to maintain a global population by associating with the spread of humans and their compost.

A fair concern about phylogenetic species, as compared to morphological species, is there any clinical relevance (Latge 2013)? The well-differentiated, cryptic species, C. albicans and C. dubliniensis showed differences in virulence (Gilfillan et al. 1998). Within C. albicans in China, it has been reported that different genotypes associate with different phenotypes of disease presentation and antifungal resistance (Bai 2014), and that his association can involve clades of pathogenic strains that are geographically distinct (Ge et al. 2012). European C. albicans has also shown association between genotype and antifungal resistance, as well as cell-surface repeats, but not pathogenicity (Odds 2010). Very recently, phylogeny was combined with antifungal susceptibility testing for nearly 10,000 yeast strains in nearly 20 species, again showing a correlation between susceptibility and phylogeny (Schmalreck et al. 2014). With filamentous Ascomycota, even the more weak-
ly differentiated cryptic species (e.g., *C. immitis* and *C. posadasii*), showed variation in salt tolerance (Fisher et al. 2002). Differences in virulence are hypothesized for cryptic species in *Paracoccidioides* (Teixeira et al. 2013) and *Blastomyces* (Meece et al. 2013) and morphological differences and differences in cultivation are known for unnamed species-level clades of *H. capsulatum* (Edwards and Rappleye 2011). In most of these cases, morphological correlates of phylogenetic species were not observed until it was possible to sort strains into the various cryptic species, only then were the phenotypic differences recognized. For example, with *Sporothrix* species, after phylogenetic species and populations had been identified, tests of antifungal drug susceptibility found associations between genotype and phenotype that could be used to predict therapeutic outcomes (Rodrigues et al. 2014).

These studies of fungal species recognition, revealing multiple, genetically isolated taxa, often occupying geographically distinct ranges, provided evidence to refute the hypothesis that microbes, because of unlimited dispersal, show no endemism and have global, panmictic populations (Finlay 2002). With the exception of *A. fumigatus* and the dermatophytes, studies of medically important fungi show that “everything is not everywhere” (Taylor 2006).

Recombination is instrumental in promoting adaptation, as has been experimentally demonstrated with *S. cerevisiae* (Goddard et al. 2005), and recombining and clonal reproduction can be distinguished using the same data that enable phylogenetic species recognition. However, population tests for reproductive mode can be made only after the phylogenetic species have been recognized because they are confounded if isolates from different, genetically isolated groups are mixed. Beginning with *C. immitis* (Burt et al. 1996), evidence for recombination was found in medically important species in which clonal conidia were obvious and sexual reproduction had not been observed. This unexpected observation challenged the notion that *Coccidioides* and, by inference, the other 20% of fungi that lacked the morphology of sexual reproduction, were limited to clonal reproduction. Subsequent studies of *C. posadasii* (Fisher et al. 2000), a related Onygenaceae, *Paracoccidioides brasiliensis* (Matute et al. 2006), and Eurotiales, *A. fumigatus* (Paoletti et al. 2005; Pringle et al. 2005) and *P. marneffei* (Henk et al. 2012) have also found evidence for recombination in collections of wild isolates from single phylogenetic species. Evidence that the mechanism of recombination is sexual comes from observations of a balance of mating types in wild populations of *Coccidioides* (Fraser et al. 2007; Mandel et al. 2007), *A. fumigatus* (Rydholm et al. 2006) and *P. marneffei* (Henk et al. 2012), the presence of intact mating loci, and, in the case of *A. fumigatus*, confirmation of sexual reproduction by successful laboratory matings (O’Gorman et al. 2009). These fungi, *Coccidioides*, *Aspergillus*, and *Penicillium*, are haploid and outbreeding (heterothallic, to use the mycological jargon), making interpretation of their reproductive behavior fairly straightforward. Although, in the case of *P. marneffei*, very low variation among interbreeding groups of wild isolates made detection of recombination more difficult (Henk et al. 2012). However, two of the most important pathogens, *C. albicans* and *C. neoformans*, are typically diploid and unraveling their reproductive modes has been more challenging, as discussed below.

The potential for reproduction in fungi goes beyond simple sexual recombination and clonal spread. Most fungi capable of sexual recombination have at least two mating alleles (termed idiomorphs because of their extreme divergence) in the population, which can promote outbreeding and restrict inbreeding. Some of the most important medically important fungi are exceptional in that natural isolates are either diploids with two copies of the same allele, as in *C. albicans*, or haploids with a population of isolates having just one mating allele, as in *C. neoformans*. Although it is possible that each species has unsampled populations of haploid individuals with a balance of mating alleles from which the animal- or human-associated strains emerged, skewed by some environmental factor in their distribution in ploidy and mating type, but it does not seem likely given the attention paid to these pathogens. Recent studies...
show that fungi with identical mating alleles, or haploids derived from diploids by aneuploidy (Hickman et al. 2013), have the potential to mate and recombine in both C. albicans (Alby et al. 2009) and C. neoformans var. grubii (Roach et al. 2014). A key question is whether or not these fungi, which clearly can spread clonally in nature and can be mated in the lab, also recombine in nature. Population genetic tests were used to report recombination in C. albicans in early studies (Gräser et al. 1996), and have been supported recently in a study of nearly 1400 isolates (Odds et al. 2007). Other Candida species, including C. parapsilosis and its close relatives, C. metapsilosis and C. orthopsilosis, have also been investigated, with the finding that C. parapsilosis and C. metapsilosis isolates have just one mating idiomorph, whereas C. orthopsilosis has a balance of both (Sai et al. 2011). By inference, one might expect that only C. orthopsilosis would be recombining in nature, as demonstrated by the discovery of natural hybridization among strains from two of its populations (Pryszcz et al. 2014). However, comparison of genomes of just a few C. parapsilosis strains also showed genetic variation consistent with recombination and independent adaptation to an environmental stress, arsenite (Pryszcz et al. 2013).

The mechanism of mating between diploid C. albicans is closely tied to pathogenesis through a developmental change necessary for both processes, the switch from white to opaque cells. A thorough review of the topic (Pujol and Soll 2011) concludes that the C. albicans population structure supports recombination as well as clonality, but notes three possible sources of the recombined genotypes. The first source would be sexual reproduction involving diploids, tetraploids, and meiosis; in support of this option, they note that a closely related species, C. lusitaniae, is sexual (Reedy et al. 2009; Butler 2010). The second source would be parasexuality involving diploids, tetraploids, and aneuploidy reductions, as has been demonstrated in the laboratory. Here, a close relative of C. albicans, C. tropicalis, which also requires a phenotypic switch to initiate mating (Porman et al. 2011), has a parasexual cycle involving diploid cells mating to form tetraploids, which then reduce to the diploid state by aneuploidy (Seerai et al. 2013). The third possible explanation would be recombination that persists from an ancestral, sexual population. The difficulty of tracking the source by mimicking the natural environment in the laboratory was recently demonstrated when wild C. tropicalis strains, few of which would switch from white to opaque in the laboratory, were stimulated to do so by providing N-acetylglucosamine as the carbon source (Xie et al. 2012). This study was followed by a similar discovery in C. albicans, in which an increase in the carbon dioxide concentration in addition to providing N-acetylglucosamine stimulated switching in wild isolates (Xie et al. 2013). These studies show that the environment is critical to producing natural behavior, and caution against using laboratory observation to limit the range of possible natural behaviors.

As noted above, a similar reproductive situation exists with the important Basidiomycota yeast pathogen, C. neoformans. Here, population genetic analyses detected recombination in wild isolates collected from populations having both mating idiomorphs, or just a single one (Saul et al. 2008). This phenomenon can be explained from laboratory research showing that mating and meiosis can occur between nonisogenic strains of the same mating idiomorph (Lin et al. 2005). More recently, aneuploidy has been shown to result from matings between partners having the same idiomorph, as well as those with the two, different idiomorphs, with the interesting twist that these aneuploidy progeny show more phenotypic variation than those arising from meiosis alone (Ni et al. 2013).

To make matters even more interesting, two phylogenetic species of C. neoformans var. grubii coexist in Botswana, one has a global distribution with almost all individuals having just one mating allele, the other is restricted to Botswana and has an even distribution of mating alleles (Litvintseva et al. 2003). Analyses showed recombination, but not between the two phylogenetic species. One interpretation is that the global phylogenetic species emerged from Africa (Litvintseva et al. 2003) but another...
is that the global phylogenetic species, strongly biased toward one mating idiomorph, emerged from some other location and dispersed to Africa. Studies of natural isolates of another Cryptococcus species, C. gattii have also detected evidence of recombination in nature (Ngamskulrungroj et al. 2009).

As plastic as reproduction seems in the medically important yeasts, the model fungus, S. cerevisiae, is even more flexible in its reproductive cycle. Most commonly, it inbreeds with siblings of the opposite mating type in the ascus, and, less commonly, it switches mating alleles to mate with itself when there is no partner of the opposite mating type and, least commonly, it outbreeds. In spite of all the inbreeding and selfing, it has a recombining population structure (Tsai et al. 2008). There is a genetic effect of all of this inbreeding and selfing in that linkage blocks are at least three to nine times larger in S. cerevisiae than in what is thought to be an obligate outbreeder, N. crassa (Liti et al. 2009; Ellison et al. 2011a).

In only a few cases have population genetic studies failed to find recombination in fungi, and even in medically important fungi. Included in this select group are fungi that are transmitted from host to host, rather than being acquired from the environment. For example, no evidence of recombination and only evidence of clonal spread has been found in the athlete’s foot fungus, Trichophyton rubrum (Gräser et al. 1999, 2006) or the histoplasmosis of horses and donkeys (Kasuga et al. 2003). Both of these cases demonstrate that clonal lineages of pathogenic fungi can emerge from recombinating populations, as has been noted for both of these cases. Detecting cases in which clonal lineages have emerged from recombinating populations is difficult and requires thorough collecting from different geographic areas in both nature and the clinic. A good example is provided by the recent C. gattii outbreak in the Pacific Coast of British Columbia, Canada and points further south in the United States. The initial interpretation of the source of the outbreak strains as recombinants from clonal lineages in Australia and North America was based on the outbreak strains and a small sample of environmental isolates (Fraser et al. 2005). The interpretation changed to clonal genotypes emerging from a recombinating population in South America when sampling enlarged to include that continent (Hagen et al. 2013).

The need to sample has risen in prominence as the cost of DNA sequencing has plummeted; in fact, the limiting resource in population genetics has become the collection of isolates. The studies mentioned above on C. neoformans (Litzintseva et al. 2003), A. fumigatus (Pringle et al. 2005), C. immitis (Fisher et al. 2002), N. crassa (Ellison et al. 2011a), or S. cerevisiae (Liti et al. 2009) all had tens to hundreds of isolates collected from throughout their ranges and, with the exception of C. immitis, at least as many environmental as clinical isolates. Population genetic studies are essential for understanding reproductive mode in nature, but do not specify the mechanism of recombination and can be difficult to apply where collections of wild isolates are limiting. Here, laboratory studies are needed, sometimes finding intact, meiotic genes in fungi that have yet to reveal their sexual mechanisms, as recently reviewed (Butler 2010).

Idiomorphs, the mating loci that are too strongly diverged to deserve the name “allele,” can have interesting evolutionary histories. Idiomorphs and mating regions diverge because they are prevented from recombining and they enlarge when additional regions of somatic chromosomes become trapped in the zones blocked from recombining. C. neoformans displays this trait (Heitman et al. 2013; Heitman 2014) and larger nonrecombining regions, approaching entire chromosomes, are seen in some Neurospora species (Menkis et al. 2008, 2010; Ellison et al. 2011a). Although the divergence is similar to animal sex chromosomes, the degeneration of one ideomorph relative to the other does not occur, presumably because the fungi must exist for at least a part of their life cycles as haploids. In Ascomycota, the mating locus itself can be introgressed from one species to another, as seen in some plant pathogenic fungi (Inderbitzin et al. 2005) and Neurospora (Strandberg et al. 2010; Sun et al. 2012b). It also has been recently shown that even large idio-
morphic regions of suppressed recombination, as seen in Cryptococcus, can be subject to recombination via gene conversion (Sun et al. 2012a).

Even the naming of fungi has been affected by the mixture of clonal and recombining reproductive systems. For nearly 150 years, Ascomycota and some Basidiomycota could have more than one name, one for the sexual state and one for each clonal state. The same access to nucleic acid variation that promoted evolutionary biology of fungi made the system of dual nomenclature unnecessary because sexual and clonal states of the same species have the same, distinctive genomes (Taylor 2011). As of January 1, 2013, each fungus must have only one name (McNeill and Turland 2011). The challenge now is to choose among the names attached to each fungus. For example, the genus Candida was applied to all morphologically undistinguished, clonal Ascomycota yeasts. Now the name can be retained for one species, logically, *C. albicans*, and the phylogenetically distant species should gain new names that emphasize evolutionary differences. Similar reasoning can be applied to species of Cryptococcus, the name that was applied to all morphologically undistinguished Basidiomycota yeasts, and even Aspergillus, where half-a-dozen clades with distinct sexual morphology as are distant from each other as they are from Penicillium (Pitt and Taylor 2014).

COMPARATIVE GENOMICS

Nowhere is comparative genomics better practiced than with fungi because of the plethora of sequenced, fungal genomes (cf., F1000, genome.jgi.doe.gov/pages/fungi-1000-projects.jsf; Fungidb, fungidb.org/fungidb; GOLD, www.genomesonline.org). Beginning with *S. cerevisiae* and relatives (Kellis et al. 2003), followed shortly by filamentous Aspergillus species (Galagan et al. 2005), many fungi, including the important fungal pathogens, have been compared to their relatives, sometimes with just a few other strains or species but increasingly in larger projects. The “low hanging fruit,” to use a dated genomics phrase, has proved to be discoveries of gene gain, gene loss, and gene duplication, processes that can lead to gene family expansion or contraction. These sources of genetic variation often have been correlated with important phenotypes in pathogenic fungi.

Experiments designed to detect chromosome duplication were conducted with yeast using strains in which one of a pair of duplicated genes had been deleted. Robust growth was restored when the region containing the remaining gene spontaneously duplicated and rapid sequencing of these strains showed that the duplications were large (Koszul et al. 2004). A reasonable inference is that all regions of fungal genomes are being duplicated all the time, but that these duplications rarely confer any selective advantage and cells or nuclei harboring them are lost. In those rare cases in which the duplication or loss does confer a fitness advantage, it is retained. The many cases in which gene duplications have been discovered through comparative genomics in fungi, to be discussed below, support this inference.

Much has been written about the fate of duplicated genes and how they can evolve to take on different functions. However, a comparison of genomes throughout the Ascomycota, including Saccharomycotina, Pezizomycotina, and Taphrinomycotina, found that duplicated genes evolved to different modes of regulation, but not different functions (Wapinski et al. 2007). In the same study, it was noted that genes associated with signal reception, transport, cell walls, and stress responses were more often duplicated than those essential for viability and growth, and whose gene products are found in nuclei and nucleoli, endoplasmic reticulum, and the Golgi apparatus, or mitochondria. Given that expansion to new environments, whether physical in the case of geographical range extensions, or biological in the case of host range expansion, involves adaptation to new stresses, this result seems logical.

The largest concentration of fungi that cause disease of otherwise healthy humans is in the Onygenales, and this group has been a focus of comparative genomic studies concerned with gene duplications. In fact, the first report of gene duplication in this group pre-dates genomics, and came from studies of en-
zymes that digest animal protein, the metalloprotease M36 family of fungalysins in the genera *Trichophyton* and *Microsporum* (Jousson et al. 2004).

When genomes of *C. immitis* and *C. posadasii* became available, they were compared to genomes from a close, nonpathogenic relative, *Uncinocarpus reesii*, a more distantly related pathogen (*Histoplasma*), and several even more distant, nonpathogenic Ascomycota in the Eurotiomycetes (*Aspergillus*) and Sordariomycetes (*Fusarium*, *Neurospora*). The most striking findings were that genes coding for enzymes involved in the deconstruction of plant cell walls were absent from all the pathogens and *U. reesii*, whereas they were common in the non-pathogens, and that genes coding for proteases (i.e., subtilisin N, peptidase S8, and peptidase M35), had significantly higher copy numbers in the two *Coccidioides* species and *U. reesii* than in the nonpathogens or *H. capsulatum* (Sharpton et al. 2009). The authors inferred that the ancestral substrate of filamentous Ascomycota was plant carbon and that the ancestor of the Onygenales adapted to use other sources of nutrition. In the case of the ancestor of the *Coccidioides* clade, including *U. reesii*, animal protein became the substrate of choice.

As genomes of dermatophyte Onygenales became available, comparative genomics continued, focusing on secondary metabolite gene clusters. Here, the dermatophytes have large numbers of these biosynthetic clusters compared to the systemic pathogens *C. immitis* and *H. capsulatum*, leading the authors to suggest that the systemic parasite had lost the clusters as they became more adapted to the full force of the host immune response (Burmester et al. 2011).

Continued study of the fungal proteases confirmed the disparity in gene number between pathogens and saprobics (Muszewska et al. 2011). These studies then turned to a phylogenetic, likelihood approach, leading reports of positive natural selection acting on genes and on specific amino acid positions for both M35 and M36 families in the dermatophytes and for the M35 family in the *Coccidioides* species (Li et al. 2012; Li and Zhang 2014). A similar approach had previously been used to detect selection on *Coccidioides* antigens and the ribosomal repeat region (Johannesson et al. 2004, 2005).

When genomes of three *Paracoccidioides* strains became available, they were added to the analyses, again finding a loss of genes involved in metabolizing carbohydrates in these species (Desjardins et al. 2011). These authors also experimented with the growth of *U. reesii* on 190 carbon substrates, including carbohydrates and proteins. They found that this close relative of *Coccidioides* has a preference for growth on protein over carbohydrate, as predicted from the comparative genomic comparisons described above.

Although distantly related to Onygenales, comparative genomics of the zygomycete genus, *Rhizopus*, also revealed expansion of protease gene families, in this case those for aspartic proteases and subtilase (Ma et al. 2009). This study also detected an expansion of genes in the ergosterol synthesis pathway, including the genes whose product, ERG11, is the target of azole antifungal drugs.

Comparative genomics has also been an active pursuit with *Candida* species. An early comparison among six *Candida* species and ten other yeasts found significant expansions in genes associated with cell walls, secretion, and molecular transport, all of them suggesting a role in virulence (Butler et al. 2009). Another comparative genomic study between *C. albicans* and its more weakly pathogenic relative, *C. dubliniensis*, found elaboration of gene families of transcription factors and transmembrane proteins in *C. albicans* and a loss of genes regulating filamentous growth in *C. dubliniensis* (Jackson et al. 2009). More recently, comparison of genomes of *Candida glabrata* and relatives, which are closer relatives of *S. cerevisiae* than *C. albicans*, found that adhesin genes, considered important for pathogenicity, had expanded in the ancestor of both pathogenic and nonpathogenic species (Gabalón et al. 2013). This result indicates that these fungi were preadapted to pathogenicity, echoing aforementioned thinking about thermotolerance and life within amoebae being necessary preadaptations for the pathogen life-
style. Comparative genomics can also have a practical side. A search for 55 genes essential to \textit{C. albicans} or \textit{A. fumigatus} found that ten were also found in other pathogens and that six of these were absent from humans; four of these are now the subject of pharmaceutical research (Abadio et al. 2011).

Switching to Basidiomycota, comparative genomics has been applied to very closely related genotypes of the important pathogen, \textit{C. neoformans}. Variants of \textit{C. neoformans} strain H99 were assembled that differ in melanization, mating capacity, and virulence in animal models as a result of their having been passaged in vivo and in vitro. These strains also showed variation in environmental and pharmaceutical stress response as well as urease production. Comparisons of their genomes found indels in the passaged strains, and these could be associated, in a mating population study, with the phenotypes of reduced mating and melanization (Janbon et al. 2014). The authors also found minor chromosome length variants, which were more common near chromosome ends. Interestingly, these genetic variants may not have phenotypes because the genes involved were either of unknown function or also found elsewhere in the genome, and these genetic differences did not associate with any phenotype. This study shows the power of comparative genomics to detect effects of evolution over short time periods, and also provides a clear means to ensure that different labs, or even different workers in the same lab, are truly working on the same strains.

Also in Basidiomycota, genomes of the agent of atopic eczema, \textit{Malassezia sympodialis} and that associated with dandruff, \textit{Malassezia globosa}, have been compared with the expected finding in these lipophilic yeasts of no genes for fatty acid synthesis and few genes for lipid biosynthesis, and the unexpected result that there were few genes known to mediate adhesion or biofilm production in organisms that seemingly must adhere to their hosts (Gioti et al. 2013). Given the role that fungi must play in allergies, an exciting aspect of the study was a focus on antigenic proteins, one of which, Mal S 7, has history of gene expansion.

**POPULATION GENOMICS**

It was the latest advance in genomic sequencing, NGS, which moved genomics from comparisons of one individual per species to comparisons of many individuals within species, or population genomics. Population genomics will obviously advance the field of fungal evolution, but will also advance fungal molecular and developmental biology. This latter advance comes in two varieties, one to find adaptive genes and another to associate genetic variation with phenotypic variation, that is, to find the genes behind any phenotype, virulence included. Comparison of individuals in a population is also revealing new sources of genetic variation that were impossible to detect by comparing single representatives of more distantly related taxa. Examples of this nascent field, which come from studies of plant and animal pathogens as well as some model fungi, show that genome regions move between species by hybridization and introgression more commonly than previously thought and that the retention of specific regions is because of positive natural selection for adaptive genes that lie in the regions. A recent review describes some of these studies (Gladieux et al. 2014) and another discusses the genetic mechanisms behind hybridization and introgression (Morales and Dujon 2012).

\textit{Coccidioides} reaped the benefits of prior population genetic research by being chosen, in the era of Sanger sequencing, for population genomics of more than a dozen individuals from the two species, \textit{C. immitis} and \textit{C. posadasii} (Neafsey et al. 2010). Evolutionarily, the most exciting finding was introgression of a surprisingly large fraction (7%) of the genome between the two species. One region moving into \textit{C. immitis} from \textit{C. posadasii} contained a metalloprotease gene (\textit{mep 4}) located near one edge of one of the introgressed regions. This gene was hypothesized to be the one responsible for sweeping the region between species. However, subsequent transcription studies found that \textit{mep 4} was not up-regulated in the shift from the environmental, hyphal form to the parasitic, endosporulating spherule form. Instead, a small,
In addition to animal pathogenic fungi, population genomics is beginning to be applied to plant pathogens. A recent, thorough study of the wheat pathogen *Mycosphaerella graminicola* and closely related strains from wild grasses, between which much ancestral variation has been preserved, found that almost no neutral, nonsynonymous mutations between the two populations, implying very strong selection on the pathogen's genes, which included those coding for secreted proteins and effectors that fungi use to parasitize plants (Stukenbrock et al. 2011).

Fungal population genomics is also practiced with model fungi, exemplified in a recent population genomic study that compared genomes chosen to span the natural variation of *S. cerevisiae* and *S. paradoxus*, but biased toward European isolates of both species- and human-associated isolates of *S. cerevisiae* (Bergström et al. 2014). The most startling result is that the ratio of gene gains, losses, and duplications as compared to SNPs is far higher in *S. cerevisiae* than *S. paradoxus*. Phenotypically, there is also more variation in *S. cerevisiae* than *S. paradoxus*. The genes showing copy number variation in *S. cerevisiae* produce proteins whose functions indicate an interaction with the environment (e.g., sugar transport and metabolism, cell adhesion and cell-to-cell flocculation, and metal transport and metabolism). For genes involved in resistance to arsenic (a source of stress previously noted for Candida species [Pryszcz et al. 2013]), the European clade had higher copy numbers, consistent with adaptation to geographically different environments. Saccharomyces species are among the few in which population genomic studies could include hundreds to thousands of individuals combined with high-throughput phenotype analysis and hypothesis testing by gene deletion or allele swaps. The medically important yeasts, *C. albicans* and *C. neoformans* are two others. We can only hope.

The filamentous, model Ascomycota, *N. crassa*, also has been the subject of population genomic analysis. As with *Coccidioides* and *Saccharomyces*, previous research on population genetics and phylogenetic species recognition

unnamed gene that is located right between *mep4* and the edge of the introgressed region, CIMG_00509, is strongly up-regulated in the shift from environmental to pathogen phenotype (Whiston et al. 2012). The gene product of this gene is predicted to be secreted and the hypothesis about adaptive genes has shifted from *mep 4* to this gene, a hypothesis that awaits validation experiments.

Human pathogens excepted, the most important fungal pathogen of animals must be the Chytridiomycota, *B. dendrobatidis*, which threatens an astounding 30% of the world’s amphibian species. Two recent population genomic studies have reached very similar conclusions about the evolution of this pathogen (Farrer et al. 2013; Rosenblum et al. 2013). They find that one clade in the species complex, termed the Global Panzooctic Lineage, is responsible for the amphibian decline. This lineage is not simply a clonal expansion from an as yet undiscovered ancestral population, but a complex clade harboring significant genetic variation. Contributing to the diversity is chromosome copy number variation, cryptic recombination within and among isolates, large, unrecombined regions that contain genes known to contribute to pathogenicity in other fungi, and genes under directional selection. Among the genes under directional selection are proteases, as has been discussed for Onygenales.

Chromosome duplication in pathogenetic lineages is not restricted to *B. dendrobatidis*. Chromosome duplication also has been implicated in azole antifungal resistance in *C. albicans*, in which it has been most carefully studied. Recently, it has been shown that aneuploids develop in *C. albicans* via a tetraploid intermediate and subsequent unequal DNA segregation (Harrison et al. 2014). These aneuploids were seen to develop in response to azole antifungals, indicating that aneuploids arise because of antifungal treatment during the infection process and are retained through selection for the increased azole resistance that they confer. Chromosome polymorphisms and copy number variation have also been researched in other pathogenic yeasts, including *C. glabrata* (Muller et al. 2009) and *C. neoformans* (Morrow and Fraser 2013).
paved the way for genomic studies. Analysis of the genomes of 50 individuals from what was thought to be one population based on concordance of four gene genealogies found seven populations, fortunately, two of them, from Louisiana and the Caribbean, happened to be well sampled (Ellison et al. 2011a). Comparison of the genomes from these two populations, between which only 10% of ancestral polymorphism had been lost, revealed two regions of exceptional divergence (Fig. 2). Although the regions were strongly diverged between the populations, they were quite similar within populations, consistent with their having entered the populations by hybridization and introgression from an as-yet-unsampled source, and then having been swept through the populations. If this scenario were true, there should be genes in the regions that confer an adaptive advantage in response to some parameter differing between the environments experienced by the two populations. Given the presence of a cold-stress RNA helicase in one region, and a cold-stress prefoldin in the other, it was hypothesized that these genes provided adaptation to the 10°C lower temperature experienced by the Louisiana strains as compared to their Caribbean neighbors. Fitness experiments with wild-type and gene-deletion mutants at high and low temperatures failed to disprove the hypothesis (or validated it, to use current parlance) that these regions swept through the Louisiana population, driven by selection for adaptation to what, for N. crassa, is an unusually cold environment (Ellison et al. 2011a). This approach is the reverse of the ecology practiced with large organisms (i.e., plants or animals), in which environmental factors and adaptive phenotypes are easier to identify and in which one can then predict which genes are likely to be responsible for the phenotype; for that reason, the microbial approach has been termed “reverse ecology.”

Just as studies of population genetics and phylogenetic species recognition set the stage for population genomic research, population genomics sets the stage for studies that have the potential to revolutionize the genetics of human pathogenic fungi by enabling genome-wide association studies (GWAS). Examples from two fungi, one a plant pathogen and one a saprobe, show the potential of this approach.

One attribute that makes both of these examples of interest to medical mycology is that relatively few strains need to be phenotyped. In each study, fewer than 100, even as few as two dozen, were phenotyped as compared to the thousands of strains that must be phenotyped in a mutational screen or gene-deletion library or the 500 commonly used in a mating population quantitative trait locus (QTL) study. A second attribute is that the association is to a single gene, or even part of a single gene, owing

Figure 2. Aligned N. crassa genomes showing a region with extreme divergence between Louisiana and Caribbean populations and relative uniformity within populations, consistent with its origin by hybridization and introgression from a more diverged source. Strains for which MRH4-like (a dead box RNA helicase) had been deleted showed reduced growth at low temperature, consistent with the hypothesis that this gene drove a selective sweep in the northern, Louisiana population (Ellison et al. 2011a). Rows are individual genomes from Louisiana (LA), Caribbean (Carib), and related (Out) populations. Columns are polymorphic nucleotide positions. Each color signifies one of the four nucleotides. Deletion of genes coding for phospholipase C (plc-1) or an unidentified protein found in mitochondria (NCU06247*) had no significant effect on growth at low temperature. (From Ellison et al. 2011a; reprinted, with permission, from the author.)
to the effect of recombination from a multitude of mating events in a wild population as compared to the one mating in a QTL crossing study (Fig. 3).

The first GWAS of a fungus was for the fungal pathogen of forest trees, *H. annosum* (Dalman et al. 2013). Here, just 23 genomes were used to associate genetic variation with the phenotypic variation of virulence on spruce and pine hosts. The analysis found seven genes that associated with virulence, four of which were known to be virulence factors in other plant pathogenic fungi and three new ones. Hypotheses about these genes await validation.

At nearly the same time, GWAS was applied to a complex morphological and physiological trait in *N. crassa* (Palma-Guerrero et al. 2013). Having first used population genomics to recognize a truly panmictic population with no detectable population structure (Ellison et al. 2011a), the researchers assessed variation among ca. 100 strains for the trait of germling fusion. In this phenomenon, conidia form pairs that then communicate so that their germination tubes grow toward each other to fuse before they initiate a large colony. Using a sample of 24 strains that spanned the range of variation found in all 100, the authors found strong association between the trait and one gene (*cse-1*) that codes for a calcium-sensitive protein involved in signaling in mammals and yeast (Fig. 3). Knowledge of two proteins that interact with this gene in mammals and yeast increased to three the number of putative genes involved in the trait. By relaxing the false discovery rate for genes associating with the trait in the GWAS, another 18 putative genes were added to the total. Tests of the hypotheses that the genes associated with the phenotype were truly involved ensued, using wild-type and gene-deletion mutants. These tests validated the association of *cse-1*, the two interacting proteins and six of the 18 genes identified by relaxing the false-positive threshold. In addition to finding nine new genes involved in the trait, three of them were the first to be found whose deletion positively affected the trait, rather than decreasing it. In addition to *Candida* species, other pathogenic fungi with tremendous potential for GWAS include the Botswana population of *C. neoformans* var. *grubii*, the South American population of *C. gattii*, and populations of both *Coccidioides* species.

**THE FUTURE**

Advances in DNA sequencing, beginning with PCR, and moving through improvements in Sanger sequencing and now NGS of various types have brought increasingly sophisticated evolutionary studies to fungi. It seems inevitable that sequencing costs will continue to fall and that all evolutionary studies will then employ complete sequences for all isolates. Phylogenomics is already replacing phylogenetics, and the challenge is to select the set of genes that best address questions about particular divergences (Lopez-Giraldez and Townsend 2011). Barcoding, presently focused on the hypervariable ITS region, will be supplanted by genome sequencing to allow researchers to pick regions of genomic variation appropriate for their evolutionary or ecological questions. Species and populations
J.W. Taylor

will be recognized by population genomics, which will replace phylogenetic species recognition or multilocus sequence typing. Once populations have been recognized, the most exciting application of NGS will become routine (i.e., finding the genes responsible for specific phenotypes). Where researchers want to let evolution identify the adaptive phenotypes, reverse ecology will be used to find the adaptive genes. More commonly, researchers will want to choose the phenotype of interest, and, here, GWAS will be the approach. The challenge to GWAS will be to increase the number of individuals from 100 to 1000 or 10,000, as now is common in GWAS for humans. The challenge for both GWAS and reverse ecology will be testing the hypotheses about genes responsible for phenotypes in fungi outside the few, well-studied model organisms.

ACKNOWLEDGMENTS

J.W.T. acknowledges the support of National Science Foundation (NSF) Grant DBI-10-46115 and NSF Grant DEB-12-57528.

REFERENCES


CDC. 1987. Revision of the CDC surveillance case definition for acquired immunodeficiency syndrome. MMWR 36: S1–S15.


Evans HC, Elliot SL, Hughes DP. 2011. Hidden diversity behind the zombie-ant fungus Ophiocordyceps unilateralis: Four new species described from carpenter ants in Minas Gerais, Brazil. Plos ONE 6: e17024.
Evolution of Human Fungal Pathogens


J.W. Taylor


Evolution of Human Fungal Pathogens

Strandberg R, Nygren K, Menkis A, James TY, Wik L, Stajich JE, Johannesson H. 2010. Conflict between reproductive gene trees and species phylogeny among heterothallic and pseudohomothallic members of the filamentous as-
J.W. Taylor


Evolutionary Perspectives on Human Fungal Pathogens

John W. Taylor

Cold Spring Harb Perspect Med 2015; doi: 10.1101/cshperspect.a019588 originally published online November 10, 2014

Subject Collection  Human Fungal Pathogens

Evolutionary Perspectives on Human Fungal Pathogens
  John W. Taylor

Black Molds and Melanized Yeasts Pathogenic to Humans
  Anuradha Chowdhary, John Perfect and G. Sybren de Hoog

Fungal Pathogens: Survival and Replication within Macrophages
  Andrew S. Gilbert, Robert T. Wheeler and Robin C. May

Innate Defense against Fungal Pathogens
  Rebecca A. Drummond, Sarah L. Gaffen, Amy G. Hise, et al.

Antifungal Pharmacokinetics and Pharmacodynamics
  Alexander J. Lepak and David R. Andes

Human Fungal Pathogens of Mucorales and Entomophthorales
  Leonel Mendoza, Raquel Vilela, Kerstin Voelz, et al.

Functional Profiling of Human Fungal Pathogen Genomes
  Alexi I. Goranov and Hiten D. Madhani

Aspergillus fumigatus and Related Species

Thermally Dimorphic Human Fungal Pathogens—Polyphyletic Pathogens with a Convergent Pathogenicity Trait
  Anita Sil and Alex Andrianopoulos

Mechanisms of Antifungal Drug Resistance
  Leah E. Cowen, Dominique Sanglard, Susan J. Howard, et al.

Treatment Principles for Candida and Cryptococcus
  Laura C. Whitney and Tihana Bicanic

The Human Mycobiome
  Patrick C. Seed

Treatment Principles for the Management of Mold Infections
  Dimitrios P. Kontoyiannis and Russell E. Lewis

Adaptive Immunity to Fungi
  Akash Verma, Marcel Wüthrich, George Deepe, et al.

The Candida Pathogenic Species Complex
  Siobhán A. Turner and Geraldine Butler

Fungal Morphogenesis
  Xiaorong Lin, J. Andrew Alspaugh, Haoping Liu, et al.

For additional articles in this collection, see http://perspectivesinmedicine.cshlp.org/cgi/collection/