Comparative genomics and the evolution of pathogenicity in human pathogenic fungi

Gary P. Moran, David C. Coleman and Derek J. Sullivan*

Microbiology Research Unit, Division of Oral Biosciences, Dublin Dental University Hospital, Trinity College Dublin, University of Dublin, Dublin 2, Republic of Ireland

Key words: Genome sequencing, comparative genomic analysis, fungi, pathogenesis, evolution.

* For correspondence. E-mail Derek.Sullivan@dental.tcd.ie; Tel. +353 1 612 7275; Fax +353 1 612 7295

ABSTRACT

Because most fungi have evolved to be free-living in the environment and because the infections they cause are usually opportunistic in nature, it is often difficult to identify specific traits that contribute to fungal pathogenesis. In recent years there has been a surge in the number of sequenced genomes of human fungal pathogens, and comparison of these sequences has proved to be an excellent resource for exploring commonalities and differences in how these species interact with their hosts. In order to survive in the human body fungi must be able to adapt to new nutrient sources and environmental stresses. Therefore, genes involved in carbohydrate and amino acid metabolism and transport and genes encoding secondary metabolites tend to be overrepresented in pathogenic species (e.g. *Aspergillus fumigatus*). However, it is clear that human commensal yeast species, such as *Candida albicans* have also evolved a range of specific factors that facilitate direct interaction with host tissues. The evolution of virulence across the human pathogenic fungi has occurred largely through very similar mechanisms. One of the most important mechanisms is gene duplication and the expansion of gene families, particularly in subtelomeric regions. Unlike prokaryotic pathogens, horizontal transfer of genes between species and other genera does not seem to have played a significant role in the evolution of fungal virulence. New sequencing technologies promise the prospect of even greater numbers of genome sequences, facilitating the sequencing of multiple genomes and transcriptomes within individual species, and will undoubtedly contribute to a deeper insight into fungal pathogenesis.

Despite the myriad of yeasts and fungal spores in the environment, relatively few of these impact directly on human health. However, those fungi that do cause disease in humans continue to represent a significant challenge for infectious diseases physicians and clinical microbiologists. Of the 100 or so fungal species that have been implicated in human disease, species belonging to the genus *Candida* and *Aspergillus* are widely recognized as the most important human fungal pathogens. Aside from the human costs associated with these infections, in 2002, it was estimated that the direct costs associated with systemic fungal infections in the United States alone amounted to $2.6 billion (60). For the most part these infections are opportunistic in nature, usually occurring in patients who are hospitalized and immunocompromised, although cases of community-acquired candidemia have recently been reported (49). The incidence of these infections continues to increase (43) and mortality rates remain high due to difficulties associated with the early diagnosis of fungal infections resulting in failure to provide appropriate antifungal therapy in time to treat fungal infections effectively.

As most fungal infections are opportunistic in nature, it is predominantly host factors that lead to the establishment of infection. However, not all fungi have the capacity to cause disease in even the most immunocompromised patient, and human pathogenic fungi must be able to express specific traits that allow them to grow in humans. While it is sometimes relatively easy to identify virulence factors directly associated with human disease in many bacterial and viral pathogens, the identification of virulence factors or, more correctly, virulence-associated factors in fungal pathogens is usually less clear-cut. Because of this complexity, our understanding of the pathogenesis of fungal infections lags behind that of other classes of microbial infections and we still have much to learn regarding how fungi

cause disease. Fortunately, new genome sequencing technologies offer future hope for improving our understanding of fungal pathogenesis. Over the past decade the genomes of many of the most important human fungal pathogens have been sequenced and the world of the "-omics" has been opened up to medical mycologists. In this Minireview we discuss how analysis of these genome sequences and, in particular, how comparative genomic analysis (CGA) of related species, has contributed to our understanding of the pathogenesis of some of the most common human fungal pathogens, i.e. *Aspergillus* spp., *Candida* spp. *Coccidioides* spp., and *Cryptococcus* spp. and how virulence has evolved in these organisms.

Identification of lineage-specific genes in human pathogenic fungi

The overriding goal of CGA in human pathogenic fungi is the identification of pathogen-specific genes required to efficiently colonize and infect the human host. CGA can produce large volumes of data, however, implicating specific genetic differences between pathogenic and non-pathogenic fungi in the disease process can be problematic. Initial CGA studies on pathogenic fungi mainly focused on comparisons with previously sequenced and distantly related non-pathogenic model organisms. Such studies have provided broad insights into genome evolution within the relevant genera, but rarely pinpointed specific genetic differences associated with virulence. In order to maximize the potential power of CGA to provide clues to identify virulence-associated genes, it is essential to choose species that are taxonomically close, yet distinct in the phenotype of interest, in this case the capacity of an organism to cause human infection. CGA within the genus *Aspergillus* demonstrates how analyses of organisms of varying relatedness can provide different

perspectives into the evolution of the genus and in particular the evolution of virulence-traits in the opportunistic pathogen *Aspergillus fumigatus*. Aspergilli are free-living filamentous saprophytic fungi that are ubiquitous in the environment, growing mainly in soil and decaying vegetation (21). These include *Aspergillus nidulans*, a model species long used in genetic studies, the industrially important *Aspergillus niger* and *Aspergillus oryzae*, and *A. fumigatus*, one of the most important human fungal pathogens. Of all environmental filamentous fungi, *A. fumigatus* is by far the most common cause of invasive human disease (11). Although this saprophyte has evolved specifically to grow in compost, some of the traits that permit growth in this environmental niche, by happenstance also allow it to grow in humans with compromised immune systems. In an attempt to identify *A. fumigatus* genes specifically associated with the capacity to cause human infection, CGA was used to compare its genome sequence with the genome sequences of *A. oryzae* and *A. nidulans.* This analysis revealed the presence of nine *A. fumigatus*-specific allergens (including ribotoxin, an enzyme that cleaves ribosomal 28s RNA in two) and more than a dozen clusters of genes encoding secondary metabolites (e.g. fumagillin) that have been suggested to play a role in virulence (21). In total, more than 500 genes were found to be *A. fumigatus*-specific, although, most of these have no known function. It is important to note, however, that these three species are in fact distantly related, (orthologous proteins share approx. 70% identity on average, which is similar to the relationship between mammals and fish (15)) and that *A. fumigatus* is the only one of the three species that can be described as pathogenic in humans. So while comparison of these genomes has been informative in terms of general physiology and genome evolution (e.g. the detection of a set of genes involved in a sexual mating cycle in *A. fumigatus*), in relation to pathogenicity, there are other species that are

more appropriate for comparative genomic analysis. Subsequent analyses applied CGA to two of the species most closely related to *A. fumigatus,* namely *Neosartorya fischeri* and *Aspergillus clavatus* (18). The former is a species that has only very rarely been associated with infection, while *A. clavatus,* which is known to produce allergens and mycotoxins, is a rare cause of alveolar inflammation, particularly in maltworkers (6, 10). At a gross level, significant differences in genome size were observed which may be explained by the greater number of transposable elements in the *N. fischeri* genome. Closer analysis confirmed the taxonomic relatedness of the three species and revealed a high level of synteny, with more than 7,500 orthologous core genes present in the three genomes. However, despite the relatively close relatedness of these species, a significant number of species-specific genes were identified. Relative to *A. nidulans* and *A. oryzae,* the genomes of these three organisms were enriched for genes involved in ion transport, choline transport and carbohydrate metabolism, as well as for genes involved in the production of secondary metabolites. Specific analysis of *A. fumigatus* also revealed the presence of species-specific genes involved in the catabolism of carbohydrates, polysaccharides and amino-sugars and in amine transport (18). Although the production of specific secondary metabolites may play a role in virulence in these organisms, it is clear that the capacity to grow in vivo is associated with the presence of genes with roles in nutrient transport and catabolism, suggesting that growth in humans may require a high degree of catabolic flexibility. Interestingly, many of these *A. fumigatus*-specific genes were on average smaller than core shared *Aspergillus* genes, had fewer introns and displayed a significant telomeric bias, the significance of which is described in greater detail below.

The genus *Candida* which is comprised of more than 150 disparate yeast species has also been the subject of several noteworthy CGA studies. The majority of *Candida* spp. are environmental saprophytes and only a dozen or so of these have been associated with human colonization or infection. *Candida albicans* is a common human commensal and is often referred to as the most pathogenic fungal species as it is responsible for most cases of candidiasis, although other species, such as *Candida glabrata, Candida tropicalis* and *Candida parapsilosis* can also cause infection (40). The most common form of candidiasis, referred to as superficial candidiasis or thrush, occurs on the mucosal surfaces of the mouth or vagina as a result of immunosupression or disturbances in the normal microbial flora. In severe cases of immunosuppression (e.g. following chemotherapy or neutropenia) especially if the integrity of the gut wall is compromised, endogenous *Candida* cells can invade and penetrate into the bloodstream, resulting in candidemia and systemic infection (42). Unlike the vast majority of *Candida* species and the *Aspergillus* species described earlier, these pathogenic yeast species are members of the human normal flora and have evolved to colonize the oral cavity, the gastrointestinal tract and the vagina. Initial comparative analysis of the *C. albicans* genome with that of the model yeast *Saccharomyces cerevisiae* revealed striking differences in metabolic capabilities and the size and expansion of a range of gene families in *C. albicans* (7, 58). Many of these gene families have roles in nutrient acquisition (e.g. secreted aspartyl proteinases and lipases) and nutrient uptake (e.g. ferric reductases, iron transporters, oligopeptide and amino acid permeases) and may play a role in host colonization and infection. *Candida albicans* also differs from *S. cerevisiae* in having additional genes required for respiratory catabolism, including acyl-CoA oxidases, fatty acid-CoA synthases and numerous oxidoreductases, all of which could be expected to play a

role in nutrient acquisition and energy production in vivo. Indeed, it has already been demonstrated that genes involved in the metabolism of alternative carbohydrate sources are required for virulence *C. albicans* (46). Further catabolic diversity is provided by the presence of several genes encoding amino acid oxidases (7). As in the case of the pathogenic filamentous fungus *A. fumigatus,* the enhanced nutrient uptake and catabolic capacity of *C. albicans* suggests that growth in vivo requires greater metabolic flexibility, and the genes involved in these processes have either been acquired by *C. albicans* and related species or lost by *S. cerevisiae* since the two species diverged.

Subsequently, a broader comparative analysis of the candidal genomes has provided a deeper insight into the evolution of pathogenic mechanisms (8). This study included several members of the so-called 'CTG clade', a group of yeasts that translate the CUG codon as serine and includes yeasts that are pathogenic (*Candida albicans, Candida tropicalis* and *Candida parapsilosis),* moderately pathogenic (*C. lusitaniae, C. guilliermondii*) and non-pathogenic (*Lodderomyces elongisporus Debaryomyces hansenii*). Despite the taxonomic distance separating these species, CGA could identify *Candida*-specific gene families with some of these families specifically enriched in the genomes of the most pathogenic species (see below) (8). These include the *ALS* gene family, which encodes a group of well-characterized proteins, proposed to act primarily as adhesins. Other gene families encode the rapidly evolving Hyr/Iff protein family, most members of which possess GPI-anchor sites, suggesting that they are cell wall-attached and possibly involved in host-pathogen interaction. A third gene family that is over-represented in the *Candida* clade encodes Pga30-like proteins, however, the role of these proteins in virulence has yet to be assessed. Although these families are enriched in the pathogenic species

they are found in most of the *Candida* genomes so far examined. For instance, while there are 8 members of the *ALS* family in *C. albicans* there are 16 *ALS* genes in the less pathogenic *C. tropicalis* and 4 in the exceedingly rare pathogen *L. elongisporus* (8). Consequently while these gene family expansions may explain why the *Candida* clade is generally more pathogenic than *Saccharomyces* species they don't fully explain why *C. albicans* is by far the most pathogenic member of the clade.

More recently, CGA has been applied to the closely related species *C. albicans* and *Candida dubliniensis*. First identified in 1995, *C. dubliniensis* is, by far, the species most closely related to *C. albicans* (22, 53, 54) and even though the two species diverged from a common ancestor approx. 20 million years ago (39) they are sufficiently closely related to mate in vitro (45). Consequently, *C. dubliniensis* shares many phenotypic characteristics with *C. albicans*, to the point where the two species are very difficult to distinguish using phenotypic methods. Despite the very close phenotypic and phylogenetic relatedness of *C. albicans* and *C. dubliniensis*, epidemiological data suggest that the former is a far more effective pathogen. The significantly lower virulence of *C. dubliniensis* has also been confirmed using a range of experimental infection models, including the murine systemic (2, 22, 59) and oro-gastric (52) models of systemic disease and in the ex vivo oral Reconstituted Human Epithelial (RHE) infection model of superficial oral infection (50, 52). On this basis *C. albicans* might be expected to exhibit a more extensive array of virulence factors than *C. dubliniensis*. However, many of the virulence factors proposed to play a role in candidal pathogenicity, such as adherence, dimorphism, phenotypic switching and the ability to produce secreted aspartyl proteinases (SAPs), are shared by both species (13, 22, 24, 25, 37). The ability to switch between yeast and true hyphal forms has long been recognized as an important *C. albicans* virulence trait. However, although

C. dubliniensis is capable of producing true hyphae it does so less efficiently than *C. albicans,* both in vivo and under a wide range of in vitro conditions (2, 41, 52, 59). As has been observed in all of the other comparative analyses of fungal genomes described earlier, the main disparity between *C. albicans* and *C. dubliniensis* is primarily due to differences in specific gene families in *C. albicans,* however, these differences are often quite subtle. Differences include genes with a known role in virulence, including a number of genes that are only expressed by *C. albicans* hyphae. Some of the most notable genes absent from the *C. dubliniensis* genome are two hypha-specific *SAP* genes and *ALS3* which is also hypha-specific and appears to have arisen in *C. albicans* by a unique transposition event. Als3 is believed to play an important role in *C. albicans* virulence as it has been demonstrated to have invasin-like properties (44) and it has been proposed to play a role in iron acquisition from ferritin in vivo (1). In addition to the subtle differences in the size of gene families known to be involved in virulence, the main disparities in gene family size occur in families with no known function such as the *IFA* genes that are proposed to encode a family of transmembrane proteins and the *TLO* gene family. While there are 14 TeLOmere-associated (*TLO*) genes in *C. albicans* (58) there are only two in *C. dubliniensis*, representing the most notable difference in gene content between the two species. The function of the *TLO* genes is unknown, however, the presence of a conserved Med2 domain suggests they may encode a novel family of transcriptional regulators. Disruption of the *C. dubliniensis TLO* genes results in defective morphogenesis, suggesting that they may play a role in candidal pathogenesis (27).

Studies to compare the genomes of the human pathogens *Coccidiodes immitis* and *C. posadasii* with the closely related non-pathogenic species, *Uncinocarpus reesii,* and the more distantly related pathogenic species *Histoplasma capsulatum*

have also revealed pathogen-specific adaptations (48). *Coccidiodes* are fungi endemic in the south west of the United States and arid areas of northern Mexico and are the causative agents of "Valley fever", a pulmonary disease that affects 100,000-300,000 people in the US each year. The fungus grows in the soil as filaments that produce arthroconidia (asexual spores), which when inhaled into the mammalian lung develop into multinucleate spherical structures, called spherules, which are filled with endospores. (43). CGA analysis identified 93 genes specific to the *Coccidiodes* lineage since divergence from *U. reesii* that exhibited spherule-specific expression, including genes involved in energy metabolism, and a gene required for the use of allantoin as a nitrogen source. This gene set also included a range of integral membrane and cell surface proteins. Additionally, *Coccidiodes* spp. have retained or acquired a set of heme-binding proteins, suggesting that as with other pathogenic fungi, acquisition of iron in vivo is crucial for virulence. As with *Aspergillus* species, many of the species-specific genes in *Coccidioides* spp. are found in "genomic islands" in subtelomeric regions, although the role of these genes in virulence has yet to be established (48). Comparison of the Onygenales genome sequences with those of members of the sister order, Eurotiales, (fungi that are primarily associated with plants, including the aspergilli), has revealed that Onygenales species either lack or have reduced numbers of genes associated with growth on plant matter, e.g. plant cell wall degrading enzymes, such as cellulases, cutinase, pectate lyase, pectin esterases as well as genes required for carbohydrate metabolism. In contrast, the *Coccidioides* spp. and *U. reesii* have larger families of protease genes encoding extracellular serine-proteases such as keratinase, and the deuterolysin metalloprotease family, some of which are only found in the *Coccidioides* spp.. These data suggest that the

Coccidioides spp. might not be soil saprophytes, instead it has been proposed that they have evolved to associate specifically with animals hosts, probably rodents (48).

Unlike the pathogenic fungi described thus far, which are all members of the phylum Ascomycota, the pathogenic yeast *Cryptococcus neoformans* belongs to the phylum Basidiomycota. *Cryptococcus neoformans* is the causative agent of cryptococcal meningitis, an infection of the immunocompromised patient, particularly those with HIV-infection. Two varieties of *C. neoformans* are known to cause infection in humans, serotype A (*var. grubbi*) and serotype D (var. *neoformans*), with serotype A accounting for the majority of infections in AIDS patients (34). Comparison of the *C. neoformans* genome with that of ascomycetous yeasts reveals considerable differences in genomic architecture, with the cryptococcal genome being rich in introns and antisense messages (35). Comparison of the whole genomes of var. *grubii* and var. *neoformans* strains revealed that an ancient, non-reciprocal transfer of 40 kb had occurred from var. *grubii* to var. *neoformans* approximately 2 million years ago and is widespread in the natural population. This event was likely the result of the creation of a serotype AD hybrid intermediate. Such hybrids can be created in the laboratory and are found in nature, but are genetically unstable and often aneuploid. This region, referred to as an 'identity island' is widespread in naturally occurring strains of var. *neoformans* and may have imparted a selective advantage which has allowed it to become fixed in the population (29).

Evolution of virulence-associated genes in pathogenic fungi

The application of comparative genomics to human pathogenic fungi has revealed that several different genetic processes have played important roles in the

acquisition of virulence-associated genes. These processes are largely driven by gene duplication to provide the raw material for new ORFs, although horizontal gene transfer may also play a minor role. Specialization for life as a commensal may also lead to gene loss, as costly genetic material associated with previous lifestyles is lost to increase reproductive fitness. Although these processes are not unique to pathogenic fungi, their importance in virulence gene evolution has been repeatedly highlighted in CGA studies of pathogenic fungi. The transfer of fungi from the environment to a living mammalian host introduces them to a range of challenging new environments. As the subsequent section outlines, fungi have used a variety of evolutionary mechanisms to facilitate growth in these potentially stressful environments.

Gene family expansion

Gene duplication is an important force in evolution. Duplicated genes can increase fitness in an environment where gene dosage is important or duplicated genes may diversify to take on new functions. Gene duplication and the formation of gene clusters or families may therefore lead to specialization for specific environmental conditions. Expansion of tandem gene arrays often occurs under selective pressure when increased gene dosage is required. Comparative analysis of the sequenced *Saccharomyces* and *Candida* genomes revealed that three cell wall-associated gene families are particularly enriched in the pathogenic species. These families, which encode the Als, Iff and Pga30 proteins, show evidence for tandem duplications and subsequent divergence and have been proposed to play important roles in host interactions (8). In *C. glabrata*, several large gene clusters have been described that exhibit evidence of functional diversification. One is a cluster of six YPS genes, encoding the yapsins, which are extracellular GPI-linked aspartyl proteinases and

another is a cluster of 8 alpha-1,3-mannosyl transferases (14, 15). One of the most important *C. glabrata* virulence factors is the Epa family, which is a family of GPI-anchored cell wall proteins that facilitate host recognition and adhesion by *C. glabrata*. The number of *EPA* genes differs from strain to strain (up to 23 paralogs have been found in one strain) but they are all located in clusters adjacent to telomeres, where they are subject to transcriptional silencing (12, 47). Evidence of tandem gene duplication also occurs in *C. albicans*, however most gene families in this species appear to be dispersed, with subgroups of related genes on the same chromosome. This model of duplication and dispersion is well exemplified by the *LIP* family of lipases. Two related clusters of *LIP* genes can be identified on chromosome 1 (*LIP1, 2, 3, 6* and *10*) and chromosome 7 (*LIP5, 8* and *9*) with an orphan gene, *LIP4* on chromosome 6, indicating that at least two translocations between chromosomes have occurred followed by expansion on chromosomes 1 and 7 (26, 58). Interestingly, the less pathogenic species *C. parapsilosis* only has two *LIP* genes, only one of which appears to be functional (20). Although the functional *LIP* gene in *C. parapsilosis* has been shown to be required for virulence, the large size of the *LIP* gene family in *C. albicans* may in some way contribute to enhancing the ability of this species to colonize and infect humans. Comparison of the *SAP* gene cluster in *C. albicans* and *C. dubliniensis* on chromosome 6 provides insight into the mechanisms of gene family expansion along a single chromosome. *Candida albicans* possesses three closely related genes termed *SAP4,5* and *6* on chromosome 6, whereas *C. dubliniensis* possesses one gene (*CdSAP456*) orthologous to this subfamily. It appears that in *C. albicans*, the ancestral '*SAP456*' gene underwent two separate duplications and segmental inversions in the region that dispersed the cluster along chromosome 6 (Fig. 1). However, the origin of *SAP1* in this region is more difficult to explain, as

SAP1 is more closely related to *SAP2* on chromosome R, indicating that proximity is not always a reliable indicator of relatedness. *SAP1* was likely derived from a separate translocation event, perhaps involving homologous flanking sequences (27, 58).

Several gene family expansions have also been identified in *Coccidiodes*. Both *Coccidiodes* spp. and the non-pathogenic *U. reesii* harbor an expanded family of extracellular serine proteases, perhaps required to acquire nutrients from mammalian tissues. This expansion most likely occurred before divergence of *U. reesii* from *Coccidiodes* and may have evolved to allow growth on decaying animal matter rather than virulence within a living host. However, another family of metalloproteases homologous to the known virulence factor *MEP1* was identified in these organisms, and in this case three additional members could be identified in the pathogenic *Coccicodes* spp. relative to *U. reesii (48).*

Telomeric "gene factories"

In many pathogenic fungi, the telomere proximal regions appear to be locations where pathogenic species have acquired novel genes that are often absent from closely related, non-pathogenic relatives. This phenomenon has been particularly well described in *A. fumigatus*. In contrast to *C. albicans*, where novel, duplicated genes are usually dispersed along a chromosome, almost 50% of the *A. fumigatus*-specific genes can be clustered together in blocks or 'genomic islands' of 10 or more genes (18). Many of these clusters of genes are involved in the production of secondary metabolites such as mycotoxins and fumigaclavine. These genomic islands show a strong telomeric bias, with the majority located within 300 kb of telomore ends. Although initial studies suggested that these clusters may have arisen by horizontal gene transfer, subsequent studies have identified paralogous genes in

other aspergilli, allowing the evolution of these genes to be traced within the genus and suggesting that they were recently duplicated and translocated to telomere proximal locations (18). Unexpectedly, a tendency was found for these *A. fumigatus* lineage-specific genes to be shorter than core *Aspergillus* genes and to contain fewer introns. This may be due to the reduced selective pressures on these duplicated genes that may have allowed loss of introns and the accumulation of premature stop codons. The tendency of these genes to cluster at telomere proximal regions may be linked to their rapid evolution due to the relaxed selective constraints at these regions. Telomeric loci have been associated with accelerated evolution in protozoa (4). Fedorova et al. (18) speculated that the telomeric proximal regions of *Aspergillus* spp. may act as 'gene factories', where duplicated genes may undergo significant divergence (or pseudogenization) in the absence of pressure to maintain function and are subsequently translocated to other areas of the genome. This theory has been termed the duplication, differentiation and differential gene loss (DDL) hypothesis. Alternatively, clustering of these genes at subtelomeric loci may be a strategy for coordinate regulation of virulence gene expression. Movement of these clusters into subtelomeric regions would place them under the regulatory control of *LaeA*, a factor that regulates gene expression through chromatin remodeling (9). Microarray data support the idea that clustering facilitates the coordinated epigenetic regulation of virulence gene expression, as ~30% of clustered genes are induced during initiation of invasive aspergillosis in a mouse infection model (38).

Genes belonging to expanded gene families have also been identified at telomere proximal regions in *Candida spp.* The *EPA* genes, mentioned above, occur in tandem arrays at the telomeres of *C. glabrata* (12), a location that may have promoted their functional divergence and facilitated their regulation by transcriptional

silencing. As also described above, a family of TeLOmere (*TLO*) specific genes has also been identified in *C. albicans* (58). Comparative genomics with the less pathogenic species *C. dubliniensis* reveals that while 14 copies are present in *C. albicans* SC5314, only two copies of this gene are present in the sequenced type strain of *C. dubliniensis* (i.e. Cd36, (27). The specific expansion of this family in *C. albicans* suggests that this family may play a specific role in commensalism or virulence of *C. albicans*. The telomere-specific location of these genes suggests a method of duplication and dispersal distinct from the gene families of *C. albicans* described previously. Dispersal of this family in *C. albicans* may have occurred by telomeric recombination. In addition, each *C. albicans TLO* gene is flanked at the 5' end by the LTR kappa, possibly implicating the movement of retrotransposons in their evolution. Furthermore, comparison of the sequences of the encoded *C. albicans* Tlo proteins with the *C. dubliniensis* and the single *C. tropicalis* ortholog suggests that the *C. albicans* family has diverged significantly from the ancestral state (27). The sequences of the *C. albicans* TLOs are highly similar, suggesting a recent expansion or one that has occurred to increase gene dosage rather than functional diversity.

Horizontal gene transfer

Horizontal gene transfer (HGT), whereby genes are exchanged between individual strains and species, is a significant force in prokaryotic evolution and plays a major role in the spread of genes involved in bacterial virulence and antibiotic resistance (28). In fungi, the role of horizontal gene transfer in evolution, either from bacteria or other fungi, has been considered to be marginal. However, the relative dearth of evidence for HGT in fungi may be due to the relatively small number of fungal genomes available for analysis and because these events may be very ancient and difficult to identify. Indeed, numerous good candidates for HGT events in fungi

have recently been described. For example, the *S. cerevisiae URA1* gene appears to have been acquired from a *Lactobacillus* species, and may have facilitated adaptation to anaerobic growth (23). Evidence also exists for the transfer of a whole cluster of genes involved in secondary metabolism from the plant pathogen *Magnaporthe grisea* to *A. clavatus* (31). There is also evidence that suggests the reason why the *A. oryzae* genome contains more than a thousand extra genes in comparison with *A. fumigatus* and *A. nidulans* is because of HGT from a range of fungi, including Sodariomycetes. This suggests that A. oryzae may be more competent to take up foreign DNA than other fungal species (32). Recently, genome sequence analyses provided evidence for the interkingdom transfer of a gene encoding a proline racemase and a gene involved in the metabolism of phenazine from *Burkholderia spp.* to *C. parapsilosis* (19). The paucity of examples of recent HGT in the *Candida* CTG clade may be due to their unique codon usage which may act to restrict HGT. Therefore, although there are examples of inter-kingdom and inter-genus HGT in fungi, there is little evidence thus far to suggest that this phenomenon has played a significant role in the evolution of virulence in human pathogenic fungi.

Adaptive evolution of gene sequences

Positive selection can lead to rapid amino acid substitution rates in particular genes, and this can often be detected by determining the ratio of the rate of non-synonymous substitutions (K_a) to the rate of synonymous substitutions (K_s) between homologous genes. Genes with high K_a/K_s ratios are usually said to be evolving rapidly under positive selection. Four virulence-associated genes of *A. fumigatus* were shown to exhibit evidence of accelerated evolution within the *Aspergillus* clade (18). This may be due to a relaxation of selection or more likely to a positive selection process leading to rapid functional diversification. The four genes are involved in

nutrient acquisition and oxidative stress response (*PabaA, fos-1, pes1* and *pksP*). Comparison of *Coccidiodes spp.* with the closely related *U. reesii* identified 67 *Coccidiodes* genes that exhibit rapid evolution. This group included the well-characterized immunization antigen 1 protein that can confer protective immunity in mice (48).

Several gene familes were identified in *C. albicans* that exhibit positive selection, including the *IFF-HYR1*, *ALS* and *PGA30*-like families of cell surface proteins that are enriched in the most pathogenic *Candida* species (8). Recombination may also play a role in the functional diversification of cell surface proteins. Phylogenetic comparison of the *C. albicans* and *C. dubliniensis ALS* genes reveals that syntenic genes at the corresponding genomic locations often lack the expected sequence similarities. Evidence exists for *ALS* mosaicism, indicating a high degree of intergenic recombination, which may play a major role in generating functional diversity in this gene family (27).

Gene loss

Although a member of the genus *Candida*, *C. glabrata*, is more closely related to *S. cerevisiae* than to *C. albicans* (15). Comparison of the *S. cerevisae* and *C. glabrata* genomes confirms the shared presence of sister chromosomal regions or 'blocks', thought to be the result of an ancestral whole genome duplication event. However, the number and size of duplicated blocks in *C. glabrata* is lower than that of *S. cerevisiae*, indicative of significant gene loss in the former species since divergence. When gene content was compared with *S. cerevisiae* and *Kluyveromyces lactis*, *C. glabrata* was found to have lost genes involved in galactose metabolism, phosphate metabolism and nitrogen and sulfur metabolism (14, 15). This level of

difference is not unexpected considering the evolutionary distance between these species and may reflect metabolic specialization of *C. glabrata* for life in the human gastrointestinal tract.

The impact of gene loss or pseudogenization on genome content is also clearly evident when the genomes of *C. albicans* and *C. dubliniensis* are compared. It is apparent that *C. dubliniensis* is undergoing reductive evolution, with many genes potentially involved in morphogenesis and virulence (e.g. *HYR1,* a member of the virulence associated *IFF* gene family (3)*,* which confers neutrophil resistance on *C. albicans* (36)) having been lost entirely or in the process of being lost through pseudogenization. *Candida dubliniensis* was found to possess 78 pseudogenes which possessed intact orthologues in *C. albicans*, including 16 genes designated as filamentous growth regulators (FGRs) (27, 57). Another gene family, the *IFA* family encoding putatative transmembrane proteins are unique to *C. albicans*, *C. dubliniensis* and *C. tropicalis*. In *C. albicans* and *C. dubliniensis*, there is evidence of widespread expansion of this gene family, with 31 loci in *C. albicans* and 21 loci in *C. dubliniensis*. However, a substantial component of the *C. dubliniensis IFA* gene repertoire appears to be in a state of mutational decay. Many of the *IFA* ORFs are heavily decayed gene relics, while others contain a few point mutations or frame-shifts, suggesting that gene loss is an ongoing process (27). Clearly, the evolutionary pressure on *C. albicans* and *C. dubliniensis* is to become successful commensals in the human host. However, given the opportunity (e.g. damage to the gut wall, neutropenia due to immunosuppression therapy, etc.) both species have the capacity to become pathogens when they overgrow and invade tissue, ultimately leading to systemic candidiasis. Taken together, the CGA data suggest that *C. dubliniensis* may be undergoing reductive evolution for specialized growth in an as yet unidentified

anatomic niche, while the greater repertoire of *C. albicans* genes may allow this species to adapt to and thrive under a more diverse range of environmental conditions, including those found in a wider range of sites in the body. Therefore, the presence of colonizing *C. albicans* in greater numbers in more anatomic sites results in this species being more pathogenic than *C. dubliniensis* and therefore responsible for a greater number of cases of invasive infection. Further evidence from the pathogenic fungi that genome reduction is an important evolutionary mechanism that has lead to niche specialization is found in the genomes of *Coccidioides* spp., which lack many of the genes necessary for growth on plant matter that are present in species such as the aspergilli (48).

Functional genomics.

While CGA has proved to be a very helpful tool in investigating fungal virulence, it is clear that analysis of gene content alone is not sufficient to explain differences between, or indeed, within fungal species. For example, comparison of two strains of *C. albicans* that differ in their ability to invade tissue and cause infection *in vivo* using CGH suggested that the gene content of the two species was identical. However, global transcription comparisons revealed significant differences in gene expression profiles, suggestive of transcriptional rewiring, which could contribute to differences in growth rate between strains in vivo (55). These data indicate that functional genomic approaches will be very helpful in identifying differences in gene expression and function between strains and closely related species. For example, cross-species forward genetic screens have been used successfully for whole genome functional comparisons between *C. albicans* and *C. dubliniensis* (17, 51).

Analysis of transcriptional networks in *C. albicans* and other ascomycete species using transcript profiling and ChIP-CHIP analysis shows that some transcriptional pathways have been specifically modified in *Candida* species. In *C. albicans*, these changes are associated with high turnover of transcription factor binding sites, as has been shown for *MCM1*, and changes in the regulatory networks controlling glycolysis and ribosome biogenesis (56). These types of changes are likely to have played a significant role in the evolution of *C. albicans* from an ancestral, environmental yeast, and have recently been reviewed by Lavoie et al. (33). In terms of host adaptation, *C. albicans* has been shown to possess numerous transcriptional programmes activated in response to conditions commonly encountered in vivo, including alkaline pH, oxidative stresses and morphogenetic signals (5). Recent transcript profiling analysis of the less virulent *C. dubliniensis* suggests that many of these transcriptional programmes are conserved (41). However, the nature of the stimulus required to trigger these transcriptional responses appears to differ in the two species and may help account for the differences in pathogenicity of these two organisms (41, 50).

Conclusions

Fungi have evolved through several specific evolutionary mechanisms. CGA has revealed that virulence and virulence-associated factors have evolved in primary and opportunistic human pathogens through the same mechanisms (summarized in Fig. 2). One of the most important of these mechanisms involves gene duplication and subsequent expansion of specific gene families and clusters. Many of the genes expressed by pathogenic fungi confer flexibility in nutrient acquisition and metabolic

diversity (e.g. *A. fumigatus*) as well as in host recognition and adhesion (e.g. *C. albicans*). Expanded gene clusters and gene islands implicated in virulence are often present in subtelomeric regions (e.g. the *EPA* genes of *C. glabrata*), with families undergoing expansion adjacent to telomeres and being exchanged between the telomeres of individual chromosomes (e.g the *TLO* genes of *C. albicans*). Lineage specific genes have also been identified in the subtelomeric regions of non-pathogenic fungi belonging to the genus *Saccharomyces*, suggesting that the chromatin structure adjacent to telomeres facilitates accelerated evolution of genes located there and that these evolutionary hotspots are a common feature of fungi in general (30). As well as evidence for gene "gain" through expansion, there are also clear cases of gene "loss" occurring in pathogenic fungi, particularly in species that are specifically adapted for life as mammalian commensals, suggestive of a process of reductive evolution (e.g. *C. glabrata*). Genomic analysis also indicates that evolution of virulence in fungi has taken a different path from that of bacterial pathogens. In bacteria, one of the most important means of acquiring virulence is through HGT of virulence factor genes (16). However, although there is evidence for HGT between different fungal species (23, 31) and between bacteria and fungi (19), it appears to be relatively rare and there have been no examples detected of the transfer of genes that might contribute directly to virulence in humans.

Although the genomes of the most important fungal pathogens have only become available recently, they have already provided a very useful insight into how fungi cause disease in humans and into how fungal virulence has evolved. Up to now comparative genomics has been hampered by the lack of publicly available completed fungal genome sequences. However, new rapid and high through put DNA and RNA sequencing technologies offer the potential to generate genomic and transcriptomic

data for multiple strains and species, allowing the comparison of any number of strains that differ in their capacity to cause disease or their host or geographic range. Comparison of the genome sequences of multiple strains will facilitate the functional analysis of the many genes of unknown function in all of the fungal pathogens and will facilitate the identification of rapidly evolving genes, which are most likely to play a role in host-pathogen interaction. Similar comparative studies using whole transcriptome shotgun sequencing (also referred to as RNA-seq) have the potential to revolutionize fungal pathogenomics and identify new virulence-associated genes. We are now on the brink of an explosion of genomic information, and while a major bottleneck in the past was the lack of sequence data, the coming tsunami of genomic data will provide a new challenge to bioinformaticians and evolutionary biologists.

Acknowledgements

Work in the authors' laboratory is supported by the Board of the Dublin Dental Hospital, Science Foundation Ireland (04/IN3/B463) and the Health Research Board (RP/2004/235 & HRA/2009/3).

References

1. **Almeida, R. S., S. Brunke, A. Albrecht, S. Thewes, M. Laue, J. E. Edwards, S. G. Filler, and B. Hube.** 2008. The hyphal-associated adhesin and invasin Als3 of *Candida albicans* mediates iron acquisition from host ferritin. PLoS Pathog. **4:**e1000217.
2. **Asmundsdottir, L. R., H. ErlendsdÛttir, B. A. Agnarsson, and M. Gottfredsson.** 2009. The importance of strain variation in virulence of *Candida dubliniensis* and *Candida albicans*: results of a blinded histopathological study of invasive candidiasis. Clin. Microbiol. Infect. **15:**576-585.
3. **Bates, S., J. M. de la Rosa, D. M. MacCallum, A. J. P. Brown, N. A. R. Gow, and F. C. Odds.** 2007. *Candida albicans* Iff11, a secreted protein required for cell wall structure and virulence. Infect. Immun. **75:**2922-2928.
4. **Berriman, M., N. Hall, K. Sheader, F. Bringaud, B. Tiwari, T. Isobe, S. Bowman, C. Corton, L. Clark, G. A. M. Cross, M. Hoek, T. Zanders, M. Berberof, P. Borst, and G. Rudenko.** 2002. The architecture of variant surface glycoprotein gene expression sites in *Trypanosoma brucei*. Mol. Biochem. Parasitol. **122:**131-140.
5. **Biswas, S., P. Van Dijck, and A. Datta.** 2007. Environmental sensing and signal transduction pathways regulating morphopathogenic determinants of *Candida albicans*. Microbiol. Mol. Biol. Rev. **71:**348-376.
6. **Blyth, W., I. W. Grant, E. S. Blackadder, and M. Greenberg.** 1977. Fungal antigens as a source of sensitization and respiratory disease in Scottish maltworkers. Clin. Allergy **7:**549-562.
7. **Braun, B. R., M. van Het Hoog, C. d'Enfert, M. Martchenko, J. Dungan, A. Kuo, D. O. Inglis, M. A. Uhl, H. Hogues, M. Berriman, M. Lorenz, A. Levitin, U. Oberholzer, C. Bachewich, D. Harcus, A. Marcil, D. Dignard, T. Iouk, R. Zito, L. Frangeul, F. Tekaia, K. Rutherford, E. Wang, C. A. Munro, S. Bates, N. A. Gow, L. L. Hoyer, G. Kohler, J. Morschhauser, G. Newport, S. Znaidi, M. Raymond, B. Turcotte, G. Sherlock, M. Costanzo, J. Ihmels, J. Berman, D. Sanglard, N. Agabian, A. P. Mitchell, A. D. Johnson, M. Whiteway, and A. Nantel.** 2005. A human-curated annotation of the *Candida albicans* genome. PLoS Genet. **1:**36-57.
8. **Butler, G., M. D. Rasmussen, M. F. Lin, M. A. S. Santos, S. Sakthikumar, C. A. Munro, E. Rheinbay, M. Grabherr, A. Forche, J. L. Reedy, I. Agrafioti, M. B. Arnaud, S. Bates, A. J. P. Brown, S. Brunke, M. C. Costanzo, D. A. Fitzpatrick, P. W. J. de Groot, D. Harris, L. L. Hoyer, B. Hube, F. M. Klis, C. Kodira, N. Lennard, M. E. Logue, R. Martin, A. M. Neiman, E. Nikolaou, M. A. Quail, J. Quinn, M. C. Santos, F. F. Schmitzberger, G. Sherlock, P. Shah, K. A. T. Silverstein, M. S. Skrzypek, D. Soll, R. Staggs, I. Stansfield, M. P. H. Stumpf, P. E. Sudbery, T. Srikantha, Q. Zeng, J. Berman, M. Berriman, J. Heitman, N. A. R. Gow, M. C. Lorenz, B. W. Birren, M. Kellis, and C. A. Cuomo.** 2009. Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature **459:**657-662.
9. **Cairns, T., F. Minuzzi, and E. Bignell.** 2010. The host-infecting fungal transcriptome. FEMS Microbiol. Lett. **307:**1-11.

10. **Chim, C. S., P. L. Ho, and K. Y. Yuen.** 1998. Simultaneous *Aspergillus fischeri* and Herpes simplex pneumonia in a patient with multiple myeloma. Scand. J. Infect. Dis. **30:**190-191.
11. **Dagenais, T. R., and N. P. Keller.** 2009. Pathogenesis of *Aspergillus fumigatus* in invasive aspergillosis. Clin. Microbiol. Rev. **22:**447-465.
12. **De Las Peñas, A., S.-J. Pan, I. Castaño, J. Alder, R. Cregg, and B. P. Cormack.** 2003. Virulence-related surface glycoproteins in the yeast pathogen *Candida glabrata* are encoded in subtelomeric clusters and subject to *RAP1*- and *SIR*-dependent transcriptional silencing. Genes Dev. **17:**2245-2258.
13. **de Repentigny, L., F. Aumont, K. Bernard, and P. Belhumeur.** 2000. Characterization of binding of *Candida albicans* to small intestinal mucin and its role in adherence to mucosal epithelial cells. Infect. Immun. **68:**3172-3179.
14. **Dujon, B.** 2010. Yeast evolutionary genomics. Nat. Rev. Genet. **11:**512-524.
15. **Dujon, B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. Marck, C. Neuveglise, E. Talla, N. Goffard, L. Frangeul, M. Aigle, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J. M. Beckerich, E. Beyne, C. Bleykasten, A. Boisrame, J. Boyer, L. Cattolico, F. Confanioleri, A. De Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J. M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G. F. Richard, M. L. Straub, A. Suleau, D. Swennen, F. Tekaia, M. Wesolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, and J. L. Souciet.** 2004. Genome evolution in yeasts. Nature **430:**35-44.
16. **Ehrlich, G., N. L. Hiller, and F. Hu.** 2008. What makes pathogens pathogenic. Genome Biol. **9:**225.
17. **Enjalbert, B., G. P. Moran, C. Vaughan, T. Yeomans, D. M. MacCallum, J. Quinn, D. C. Coleman, A. J. P. Brown, and D. J. Sullivan.** 2009. Genome-wide gene expression profiling and a forward genetic screen show that differential expression of the sodium ion transporter Ena21 contributes to the differential tolerance of *Candida albicans* and *Candida dubliniensis* to osmotic stress. Mol. Microbiol. **72:**216-228.
18. **Fedorova, N. D., N. Khaldi, V. S. Joardar, R. Maiti, P. Amedeo, M. J. Anderson, J. Crabtree, J. C. Silva, J. H. Badger, A. Albarraq, S. Angiuoli, H. Bussey, P. Bowyer, P. J. Cotty, P. S. Dyer, A. Egan, K. Galens, C. M. Fraser-Liggett, B. J. Haas, J. M. Inman, R. Kent, S. Lemieux, I. Malavazi, J. Orvis, T. Roemer, C. M. Ronning, J. P. Sundaram, G. Sutton, G. Turner, J. C. Venter, O. R. White, B. R. Whitty, P. Youngman, K. H. Wolfe, G. H. Goldman, J. R. Wortman, B. Jiang, D. W. Denning, and W. C. Nierman.** 2008. Genomic islands in the pathogenic filamentous fungus *Aspergillus fumigatus*. PLoS Genet. **4:**e1000046.
19. **Fitzpatrick, D., M. Logue, and G. Butler.** 2008. Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. BMC Evol. Biol. **8:**181.
20. **Gácser, A., D. Trofa, W. Schäfer, and J. D. Nosanchuk.** 2007. Targeted gene deletion in *Candida parapsilosis* demonstrates the role of secreted lipase in virulence. J. Clin. Invest. **117:**3049-3058.

21. **Galagan, J. E., S. E. Calvo, C. Cuomo, L. J. Ma, J. R. Wortman, S. Batzoglou, S. I. Lee, M. Basturkmen, C. C. Spevak, J. Clutterbuck, V. Kapitonov, J. Jurka, C. Scazzocchio, M. Farman, J. Butler, S. Purcell, S. Harris, G. H. Braus, O. Draht, S. Busch, C. D'Enfert, C. Bouchier, G. H. Goldman, D. Bell-Pedersen, S. Griffiths-Jones, J. H. Doonan, J. Yu, K. Vienken, A. Pain, M. Freitag, E. U. Selker, D. B. Archer, M. A. Penalva, B. R. Oakley, M. Momany, T. Tanaka, T. Kumagai, K. Asai, M. Machida, W. C. Nierman, D. W. Denning, M. Caddick, M. Hynes, M. Paoletti, R. Fischer, B. Miller, P. Dyer, M. S. Sachs, S. A. Osmani, and B. W. Birren.** 2005. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. Nature **438:**1105-1115.

22. **Gilfillan, G. D., D. J. Sullivan, K. Haynes, T. Parkinson, D. C. Coleman, and N. A. R. Gow.** 1998. *Candida dubliniensis*: phylogeny and putative virulence factors. Microbiology **144:**829-838.

23. **Hall, C., S. Brachat, and F. S. Dietrich.** 2005. Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. Eukaryot. Cell **4:**1102-1115.

24. **Hannula, J., M. Saarela, B. Dogan, J. Paatsama, P. Koukila-Kahkola, S. Pirinen, H. L. Alakomi, J. Perheentupa, and S. Asikainen.** 2000. Comparison of virulence factors of oral *Candida dubliniensis* and *Candida albicans* isolates in healthy people and patients with chronic candidosis. Oral Microbiol. Immunol. **15:**238-244.

25. **Hoyer, L. L., R. Fundyga, J. E. Hecht, J. C. Kapteyn, F. M. Klis, and J. Arnold.** 2001. Characterization of agglutinin-like sequence genes from non-albicans *Candida* and phylogenetic analysis of the *ALS* family. Genetics **157:**1555-1567.

26. **Hube, B., F. Stehr, M. Bossenz, A. Mazur, M. Kretschmar, and W. Schäfer.** 2000. Secreted lipases of *Candida albicans*: cloning, characterisation and expression analysis of a new gene family with at least ten members. Arch. Microbiol. **174:**362-374.

27. **Jackson, A. P., J. A. Gamble, T. Yeomans, G. P. Moran, D. Saunders, D. Harris, M. Aslett, J. F. Barrell, G. Butler, F. Citiulo, D. C. Coleman, P. W. de Groot, T. J. Goodwin, M. A. Quail, J. McQuillan, C. A. Munro, A. Pain, R. T. Poulter, M. A. Rajandream, H. Renauld, M. J. Spiering, A. Tivey, N. A. Gow, B. Barrell, D. J. Sullivan, and M. Berriman.** 2009. Comparative genomics of the fungal pathogens *Candida dubliniensis* and *C. albicans*. Genome Res. **19:**2231-2244.

28. **Jain, R., M. C. Rivera, J. E. Moore, and J. A. Lake.** 2003. Horizontal gene transfer accelerates genome innovation and evolution. Mol. Biol. Evol. **20:**1598-1602.

29. **Kavanaugh, L. A., J. A. Fraser, and F. S. Dietrich.** 2006. Recent evolution of the human pathogen C*ryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. Mol. Biol. Evol. **23:**1879-1890.

30. **Kellis, M., N. Patterson, M. Endrizzi, B. Birren, and E. S. Lander.** 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:**241-254.

31. **Khaldi, N., J. Collemare, M. H. Lebrun, and K. H. Wolfe.** 2008. Evidence for horizontal transfer of a secondary metabolite gene cluster between fungi. Genome Biol. **9:**R18.

32. **Khaldi, N., and K. H. Wolfe.** 2008. Elusive origins of the extra genes in *Aspergillus oryzae.* PLoS One **3:**e3036.
33. **Lavoie, H., H. Hogues, and M. Whiteway.** 2009. Rearrangements of the transcriptional regulatory networks of metabolic pathways in fungi. Curr. Opin. Microbiol. **12:**655-663.
34. **Lin, X.** 2009. *Cryptococcus neoformans*: morphogenesis, infection, and evolution. Infect. Genet. Evol. **9:**401-416.
35. **Loftus, B. J., E. Fung, P. Roncaglia, D. Rowley, P. Amedeo, D. Bruno, J. Vamathevan, M. Miranda, I. J. Anderson, J. A. Fraser, J. E. Allen, I. E. Bosdet, M. R. Brent, R. Chiu, T. L. Doering, M. J. Donlin, C. A. D'Souza, D. S. Fox, V. Grinberg, J. Fu, M. Fukushima, B. J. Haas, J. C. Huang, G. Janbon, S. J. Jones, H. L. Koo, M. I. Krzywinski, J. K. Kwon-Chung, K. B. Lengeler, R. Maiti, M. A. Marra, R. E. Marra, C. A. Mathewson, T. G. Mitchell, M. Pertea, F. R. Riggs, S. L. Salzberg, J. E. Schein, A. Shvartsbeyn, H. Shin, M. Shumway, C. A. Specht, B. B. Suh, A. Tenney, T. R. Utterback, B. L. Wickes, J. R. Wortman, N. H. Wye, J. W. Kronstad, J. K. Lodge, J. Heitman, R. W. Davis, C. M. Fraser, and R. W. Hyman.** 2005. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science **307:**1321-1324.
36. **Luo, G., Ashraf S. Ibrahim, B. Spellberg, Clarissa J. Nobile, Aaron P. Mitchell, and Y. Fu.** 2010. *Candida albicans* Hyr1p confers resistance to neutrophil killing and is a potential vaccine target. J. Infect. Dis. **201:**1718-1728.
37. **McCullough, M., B. Ross, and P. Reade.** 1995. Characterization of genetically distinct subgroup of *Candida albicans* strains from oral cavities of patients infected with human immunodeficiency virus. J. Clin. Microbiol. **33:**696-700.
38. **McDonagh, A., N. D. Fedorova, J. Crabtree, Y. Yu, S. Kim, D. Chen, O. Loss, T. Cairns, G. Goldman, D. Armstrong-James, K. Haynes, H. Haas, M. Schrettl, G. May, W. C. Nierman, and E. Bignell.** 2008. Sub-telomere directed gene expression during initiation of invasive aspergillosis. PLoS Pathog. **4:**e1000154.
39. **Mishra, P. K., M. Baum, and J. Carbon.** 2007. Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. Mol. Genet. Genomics **278:**455-65.
40. **Moran, G. P., D. J. Sullivan, and D. C. Coleman.** 2002. Emergence of non-*Candida albicans Candida* species as pathogens, p. 37 - 53. *In* R. A. Calderone (ed.), *Candida* and candidiasis. ASM Press, Washington D.C.
41. **O'Connor, L., N. Caplice, D. C. Coleman, D. J. Sullivan, and G. P. Moran.** 2010. Differential filamentation of *Candida albicans* and *C. dubliniensis* is governed by nutrient regulation of *UME6* expression. Eukaryot. Cell. **9:**1383-1397
42. **Perlroth, J., B. Choi, and B. Spellberg.** 2007. Nosocomial fungal infections: epidemiolohy, diagnosis, and treatment. Med. Mycol. **45:**321-346.
43. **Pfaller, M. A., and D. J. Diekema.** 2010. Epidemiology of invasive mycoses in North America. Crit. Rev. Microbiol. **36:**1-53.
44. **Phan, Q. T., C. L. Myers, Y. Fu, D. C. Sheppard, M. R. Yeaman, W. H. Welch, A. S. Ibrahim, J. E. Edwards, Jr., and S. G. Filler.** 2007. Als3 is a *Candida albicans* invasin that binds to cadherins and induces endocytosis by host cells. PLoS Biol. **5:**64.

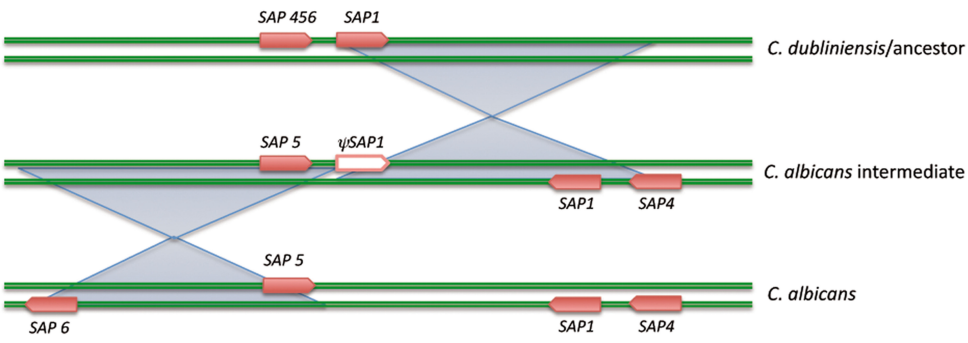
45. **Pujol, C., K. J. Daniels, S. R. Lockhart, T. Srikantha, J. B. Radke, J. Geiger, and D. R. Soll.** 2004. The closely related species *Candida albicans* and *Candida dubliniensis* can mate. Eukaryot. Cell **3:**1015-1027.
46. **Ramirez, M. A., and M. C. Lorenz.** 2007. Mutations in alternative carbon utilization pathways in *Candida albicans* attenuate virulence and confer pleiotropic phenotypes. Eukaryot. Cell **6:**280-290.
47. **Rosas-Hernandez, L. L., A. Juarez-Reyes, O. E. Arroyo-Helguera, A. De Las Penas, S.-J. Pan, B. P. Cormack, and I. Castano.** 2008. yKu70/yKu80 and Rif1 regulate silencing differentially at telomeres in *Candida glabrata*. Eukaryot. Cell **7:**2168-2178.
48. **Sharpton, T. J., J. E. Stajich, S. D. Rounsley, M. J. Gardner, J. R. Wortman, V. S. Jordar, R. Maiti, C. D. Kodira, D. E. Neafsey, Q. Zeng, C. Y. Hung, C. McMahan, A. Muszewska, M. Grynberg, M. A. Mandel, E. M. Kellner, B. M. Barker, J. N. Galgiani, M. J. Orbach, T. N. Kirkland, G. T. Cole, M. R. Henn, B. W. Birren, and J. W. Taylor.** 2009. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. Genome Res. **19:**1722-1731.
49. **Sofair, A. N., G. M. Lyon, S. Huie-White, E. Reiss, L. H. Harrison, L. T. Sanza, B. A. Arthington-Skaggs, and S. K. Fridkin.** 2006. Epidemiology of community-onset candidemia in Connecticut and Maryland. Clin. Infect. Dis. **43:**32-39.
50. **Spiering, M. J., G. P. Moran, M. Chauvel, D. M. Maccallum, J. Higgins, K. Hokamp, T. Yeomans, C. d'Enfert, D. C. Coleman, and D. J. Sullivan.** 2010. Comparative transcript profiling of *Candida albicans* and *Candida dubliniensis* identifies *SFL2*, a *C. albicans* gene required for virulence in a reconstituted epithelial infection model. Eukaryot. Cell **9:**251-265.
51. **Staib, P., and J. Morschhauser.** 2005. Differential expression of the *NRG1* repressor controls species-specific regulation of chlamydospore development in *Candida albicans* and *Candida dubliniensis*. Mol. Microbiol. **55:**637-652.
52. **Stokes, C., Moran, G.P., M. J. Spiering, G. T. Cole, D. C. Coleman, and D. J. Sullivan.** 2007. Lower filamentation rates of *Candida dubliniensis* contribute to its lower virulence in comparison with *Candida albicans.* Fungal Genet. Biol. **44:**920-931.
53. **Sullivan, D. J., G. P. Moran, and D. C. Coleman.** 2005. *Candida dubliniensis*: Ten years on. FEMS Microbiol. Lett. **253:**9-17.
54. **Sullivan, D. J., T. J. Westerneng, K. A. Haynes, D. E. Bennett, and D. C. Coleman.** 1995. *Candida dubliniensis* sp. nov.: phenotypic and molecular characterization of a novel species associated with oral candidosis in HIV-infected individuals. Microbiology **141:**1507-1521.
55. **Thewes, S., G. P. Moran, B. B. Magee, M. Schaller, D. J. Sullivan, and B. Hube.** 2008. Phenotypic screening, transcriptional profiling, and comparative genomic analysis of an invasive and non-invasive strain of *Candida albicans*. BMC Microbiol. **8:**187.
56. **Tuch, B. B., D. J. Galgoczy, A. D. Hernday, H. Li, and A. D. Johnson.** 2008. The Evolution of combinatorial gene regulation in fungi. PLoS Biol. **6:**e38.
57. **Uhl, M. A., M. Biery, N. Craig, and A. D. Johnson.** 2003. Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen *C. albicans*. EMBO J **22:**2668-2678.

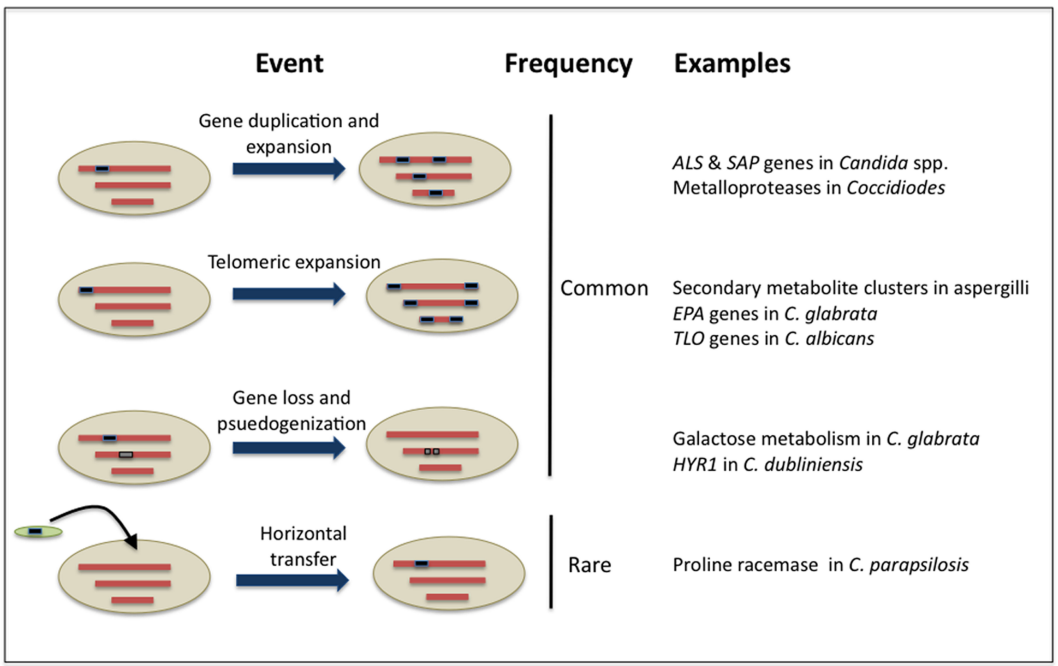
58. **van het Hoog, M., T. Rast, M. Martchenko, S. Grindle, D. Dignard, H. Hogues, C. Cuomo, M. Berriman, S. Scherer, B. B. Magee, M. Whiteway, H. Chibana, A. Nantel, and P. T. Magee.** 2007. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol. **8:**R52.
59. **Vilela, M. M., K. Kamei, A. Sano, R. Tanaka, J. Uno, I. Takahashi, J. Ito, K. Yarita, and M. M.** 2002. Pathogenicity and virulence of *Candida dubliniensis*: comparison with *C. albicans*. Med. Mycol. **40:**249 - 257.
60. **Wilson, L. S., C. M. Reyes, M. Stolpman, J. Speckman, K. Allen, and J. Beney.** 2002. The direct cost and incidence of systemic fungal infections. Value Health **5:**26-34.

Figure Legends

FIG. 1. Cartoon depicting how the *SAP* gene family may have expanded on chromosome 6 in *C. albicans*. It is proposed that *C. dubliniensis* represents the ancestral state (top) and that a single inversion event led to the duplication of the tandem pair of gene *SAP456* and *SAP1* to create an intermediate strain (middle) with *SAP5, SAP1* and *SAP4*, but which has lost one of the copies of *SAP1* (*ψSAP1*). It is proposed that another inversion resulted in the duplication of *SAP5* to create *SAP6*, representing the current complement of genes in *C. albicans* (bottom) Adapted from Jackson et al. (27).

FIG. 2. Summary of the genomic processes responsible for the evolution of human fungal pathogens.


SAP 456
SAP1
C. dubliniensis/ancestor
SAP 5
ψSAP1
C. albicans intermediate
SAP1
SAP4
SAP 5
C. albicans
SAP 6
SAP1
SAP4


Event
Frequency
Examples
Gene duplication and expansion
ALS & SAP genes in Candida spp.
Metalloproteases in Coccidiodes
Telomeric expansion
Common
Secondary metabolite clusters in aspergilli
EPA genes in C. glabrata
TLO genes in C. albicans
Gene loss and psuedogenization
Galactose metabolism in C. glabrata
HYR1 in C. dubliniensis
Horizontal transfer
Rare
Proline racemase in C. parapsilosis