Analysis of Cyp51 protein sequences shows 4 major Cyp51 gene family groups across fungi

Brandi N. Celia-Sanchez [ID] ,[1] Brandon Mangum,[1] Marin Brewer [ID] ,[2] Michelle Momany [ID] [1,*]

[1]Department of Plant Biology, University of Georgia, Athens, GA 30606, USA,
[2]Department of Plant Pathology, University of Georgia, Athens, GA 30606, USA

*Corresponding author: Department of Plant Biology, University of Georgia, Athens, GA 30606, USA. Email: mmomany@uga.edu

Abstract

Azole drugs target fungal sterol biosynthesis and are used to treat millions of human fungal infections each year. Resistance to azole drugs has emerged in multiple fungal pathogens including *Candida albicans*, *Cryptococcus neoformans*, *Histoplasma capsulatum*, and *Aspergillus fumigatus*. The most well-studied resistance mechanism in *A. fumigatus* arises from missense mutations in the coding sequence combined with a tandem repeat in the promoter of *cyp51A*, which encodes a cytochrome P450 enzyme in the fungal sterol biosynthesis pathway. Filamentous members of Ascomycota such as *A. fumigatus* have either 1 or 2 of 3 Cyp51 paralogs (Cyp51A, Cyp51B, and Cyp51C). Most previous research in *A. fumigatus* has focused on Cyp51A due to its role in azole resistance. We used the *A. fumigatus* Cyp51A protein sequence as the query in database searches to identify Cyp51 proteins across fungi. We found 435 Cyp51 proteins in 295 species spanning from early-diverging fungi (Blastocladiomycota, Chytridiomycota, Zoopagomycota, and Mucormycota) to late-diverging fungi (Ascomycota and Basidiomycota). We found these sequences formed 4 major Cyp51 groups: Cyp51, Cyp51A, Cyp51B, and Cyp51C. Surprisingly, we found all filamentous Ascomycota had a Cyp51B paralog, while only 50% had a Cyp51A paralog. We created maximum likelihood trees to investigate the evolution of Cyp51 in fungi. Our results suggest Cyp51 is present in all fungi with 3 paralogs emerging in Pezizomycotina, including Cyp51C which appears to have diverged from the progenitor of the Cyp51A and Cyp51B groups.

Keywords: Cyp51; azole resistance; Cyp51A; Cyp51B; Cyp51C

Introduction

Fungal pathogens caused over 9 million diagnosed infections in 2017 in the United States, but the true fungal burden is hard to estimate since many cases are likely undiagnosed (Vallabhaneni *et al.* 2016; Benedict *et al.* 2019). The infections caused by fungal pathogens include severe chronic conditions, complex chronic respiratory conditions, recurrent infections, and many life-threatening invasive diseases (Perlin *et al.* 2017). Invasive fungal infections generally occur in individuals with suppressed or compromised immune systems (Brown *et al.* 2012). These infections have a high mortality rate if not treated early with appropriate antifungal drugs (Brown *et al.* 2012). Major drugs used to treat invasive fungal infections are echinocandins, polyenes, flucytosine, and azole drugs (Pound *et al.* 2011). Azoles, which target synthesis of the fungal-specific membrane component ergosterol, are among the most highly used antifungal drugs.

Cyp51 proteins, also known as Erg11 in Ascomycota yeast in the Saccharomycotina and Taphrinomycotina subphyla, are in all biological kingdoms and are highly conserved (Lepesheva and Waterman 2011). Cyp51 proteins have 6 substrate recognition sites (SRS), an oxygen-binding motif (AGXDTT), PER, and EXXR motifs that create an E–R–R triad within the heme pocket, and a conserved heme-binding motif (FXXGXXXCXG) (Fig. 1) (Lepesheva and Waterman 2007; Córdova *et al.* 2017). Azole drugs

competitively bind to sterol 14 alpha-demethylase (Cyp51, Erg11), a cytochrome P450 in the ergosterol biosynthesis pathway in fungi. Azole drugs consist of a heterocyclic ring with either 2 (imidazoles) or 3 (triazoles) nitrogens and a sidechain. The side chain of azoles interacts with the Cyp51 polypeptide while the nitrogen in the azole heterocyclic ring interacts directly with the sixth ligand of the heme ferric ion, a cofactor of the Cyp51 protein (Jefcoate *et al.* 1969). Cytochrome P450 proteins conduct a 3-step reaction within the sterol biosynthesis pathway leading to the production of cholesterol in animals, sitosterol in plants, and ergosterol in fungi (Schaller 2003; Dufourc 2008). Sterols are integrated into the cell membrane where they aid in membrane fluidity and permeability (Dufourc 2008). The binding of azoles to Cyp51 depletes intracellular ergosterol and causes accumulation of methylated sterols and toxic intermediate sterols within the fungal cell membrane causing arrested growth and cell membrane stress (Cowen and Steinbach 2008).

Many fungi have acquired mutations in *cyp51* that alter the ability of azoles to bind and inhibit Cyp51 (Denning and Perlin 2011; Leroux and Walker 2011; Flowers *et al.* 2015; Kano *et al.* 2017). Many missense mutations that have been shown to decrease sensitivity to azoles as determined by minimum inhibitory concentration in the human fungal pathogens *Candida albicans*, *Cryptococcus neoformans*, *Histoplasma capsulatum*, and *A. fumigatus* (Rodero *et al.* 2003; Wheat *et al.* 2006; Lockhart *et al.* 2011; Sionov

Fig. 1. Typical organization of Cyp51 domains. Cyp51 proteins contain 6 substrate recognition sites (SRS1–6), an oxygen-binding motif (AGXDTT), PER and EXXR motifs, and a conserved heme-binding motif (FXXGXXXCXG). Black boxes represent SRS domains. Gray shading represents other motifs. Diagram is based on *A. fumigatus* Cyp51A (XP_752137.1) and is shown to scale.

et al. 2012; Flowers *et al.* 2015) occur in SRS causing azoles to interact and bind differently within Cyp51. Increased expression levels of *cyp51A* due to 34-, 46-, 53-, and 120-bp tandem repeats in the promoter have occurred in *A. fumigatus* leading to high levels of pan-azole resistance, which is resistance to more than 1 azole drug (Snelders *et al.* 2008; Hodiamont *et al.* 2009; Snelders *et al.* 2010; Hare *et al.* 2019). Tandem repeats in the *cyp51A* promoter reduce the affinity of the promoter and the CGAAT-binding complex, which binds to CGAAT in the promoter and downregulates *cyp51A* expression (Gsaller *et al.* 2016). Although Cyp51A has been the focus of most studies in *A. fumigatus*, a second paralog (Cyp51B) has also been documented to cause resistance through upregulation and missense mutations (Buied *et al.* 2013; Gonzalez-Jimenez *et al.* 2020). Like human pathogens, plant pathogens (such as *Erysiphe necator*, *Mycosphaerella* spp., *Penicillium digitatum*, and *Venturia inaequalis*) undergo changes in the *cyp51A* promoter (substitutions, insertions, and duplications) and/or mutations in SRS to alter expression and binding of Cyp51A (Délye *et al.* 1997; Schnabel and Jones 2001; Ghosoph *et al.* 2007; Cañas-Gutiérrez *et al.* 2009; Stammler *et al.* 2009; Cools *et al.* 2012; Gadoury *et al.* 2012; Pfeufer and Ngugi 2012; Cools and Fraaije 2013; Sun *et al.* 2013).

Previous studies were limited to fewer than 86 fungal sequences (Hawkins *et al.* 2014; Song *et al.* 2018) or only to *Aspergillus* spp. (Dos Santos *et al.* 2020; Perez-Cantero *et al.* 2020). The goal of our study was to better understand the relationships among all fungal Cyp51 proteins.

Materials and methods
NCBI protein Blast
Aspergillus fumigatus Cyp51A protein sequence (XP_752137.1) was used in an NCBI Protein BLAST to search the reference sequence database for other Cyp51 proteins in Fungi. The following setting were used for the Protein BLAST: Database: Reference Proteins, Exclude: uncultured/environmental sample sequences, Algorithm: blastp (protein-protein BLAST), Max Target Sequences: 1000, Expect Threshold: 0.001, Word size: 6, Max matches in a query range: 0, Matrix: BLOSUM62, Gap Costs: Existence: 11 Extension: 1, and Compositional adjustments: Conditional compositional score matrix adjustment. Sequences with <50% coverage and <30% percent identity were eliminated. To ensure all clades were represented, the top hit from each clade was used to conduct another protein BLAST search (Supplementary Table 1). Some clades were not represented in the reference sequence database, so identical settings were used to search and filter the nonredundant protein sequences database. The unfiltered searches resulted in a total of 4,404 sequence hits. Sequences were filtered to have >50% coverage and more than 70% percent identity. FASTA files of the 480 sequences resulting from filtering were downloaded and opened in Geneious Prime 2019.1.1. To confirm the sequences were a Cyp51, sequences were checked for the presence of SRS1–6, the

oxygen-binding motif AGXDTT, PER, and EXXR motifs, and the conserved FXXGXXXCXG heme-binding domain.

The following sequences were eliminated for missing amino acids in SRS1–6 or missing amino acids in conserved motifs: EPZ30787.1, EPZ31936.1, KXN68292.1, RKP07181.1, RKP16598.1, RKP17874.1, RKP18653.1, RKP18926.1, XP_001218650.1, XP_0025 63403.1, XP_002583031.1, XP_002842283.1, XP_003005233.1, XP_00 3325369.2, XP_007375289.1, XP_007756389.1, XP_007802603.1, XP_008039623.1, XP_009649122.1, XP_013258864.1, XP_01540401 5.1, XP_018230821.1, XP_018249826.1, XP_018270027.1, XP_018712 692.1, XP_020066776.1, XP_022578172.1, XP_025599710.1, XP_027 619241.1, XP_027619242.1, XP_031034290.1, XP_031059536.1, ORZ 32486.1, XP_017991977.1, XP_003017020.1, XP_003019064.1, and XP_033461214.1. The following 19 fusion proteins were found in various members in Ascomycota: XP_022511803.1, XP_0132789 94.1, XP_024670717.1, XP_031899359.1, XP_031927262.1, XP_02662 1438.1, XP_025554838.1, XP_018192447.1, XP_031935516.1, XP_02 5433272.1, XP_015404994.1, XP_024709601.1, XP_007688940.1, XP_014073145.1, XP_025394250.1, XP_014555703.1, XP_0077128 64.1, XP_033384229.1, XP_018700143.1. *Aspergillus flavus* was reported to contain 3 Cyp51 proteins (Liu *et al.* 2012; Paul *et al.* 2015), but these were later removed by NCBI (https://www.ncbi.nlm.nih.gov/protein/XP_002383931.1/) (XP_002375123.1, XP_0023 79130.1, XP_002383931.1). Some *Aspergillus* spp. have a duplication of Cyp51A or Cyp51B which has previously been called Cyp51C (Perez-Cantero *et al.* 2020). NCBI refSeq is not exclusively based on whole genome sequences so it should be noted that it is possible that the full complement of Cyp51 orthologs in each of the 295 species was not identified.

Phylogenetic analyses
Protein sequences were aligned once with MAFFT version 7.407 then once with PASTA version 1.8.5 (Katoh and Standley 2013; Mirarab *et al.* 2015). Maximum likelihood trees were constructed with RAxML version 8.2.11 with a PROTGAMMAAUTO or GTRGAMMA substitution model and 1000 bootstraps (Stamatakis 2014). Interactive Tree of Life (iTOL) was used for visualization and annotation of the trees (Letunic and Bork 2019). Species were represented by a single Cyp51 protein sequence. Subphylum Pezizomycotina was represented by group Cyp51B where duplicate Cyp51B sequences were removed.

Amino acid analyses
Geneious Prime (version 2021.2.2) was used to generate pairwise identities and consensus sequences. Similarity tables for the whole protein and motifs are based on pairwise identity in each group, number of similar amino acids divided by the total number of amino acids in the protein or motif. "Weblogo-like" diagrams were created manually for visualization of conservation across groups. The height of 1 letter amino acid designation was based on frequency across all 4 consensus sequences. Colors and symbols were used as described in Fig. 5 to denote conservation within groups.

Results and discussion

Four hundred thirty-five fungal Cyp51 proteins were analyzed

Cyp51 proteins were previously defined as having 6 SRS, an oxygen-binding motif (AGXDTT), PER and EXXR motifs that create an E–R–R triad within the heme pocket, and a conserved heme-binding motif (FXXGXXXCXG) (Lepesheva and Waterman 2007; Córdova *et al.* 2017) (Fig. 1). To understand Cyp51 genes across Fungi, the Cyp51A protein from *A. fumigatus* strain Af293—a well-studied clinical strain used by multiple laboratories (Garcia-Rubio *et al.* 2018; Keizer *et al.* 2021)—was used as a reference in a protein Basic Local Alignment Search Tool (BLAST) (Altschul *et al.* 1990). A total of 4,404 protein sequences resulted and were filtered to retain those with >50% coverage and >30% percent identity to the reference sequence, resulting in 480 sequences (Supplementary Table 1). The resulting protein sequences were analyzed for the presence of full length SRS1–6 domains and the 4 Cyp51 motifs. Of these, 435 proteins had SRS1–6 domains, the oxygen-binding motif AGXDTT, the PER and EXXR motifs, and the conserved heme-binding motif FXXGXXXCXG and were considered to be functional Cyp51 proteins (Supplementary Table 1). Among the 435 Cyp51 sequences that met our criteria for inclusion, JGI and NCBI listed 15 apparent fusion proteins; 13 between Cyp51B and a kinase immediately upstream and 2 between Cyp51C and an acetyltransferase immediately upstream. To determine whether the 15 apparent fusion proteins were bona fide fusion proteins or 2 separate proteins merged by in silico errors, we searched syntenic regions in related species in FungiDB and MycoCosm (Grigoriev *et al.* 2012; Amos *et al.* 2022). Syntenic regions showed some separate proteins with intergenic regions between them, some overlapping proteins, and some fused proteins (Grigoriev *et al.* 2012; Amos *et al.* 2022). The Cyp51 portions of the fusion proteins were extracted and kept in the analyses resulting in a total of 435 Cyp51 proteins from 295 species (Supplementary Tables 1 and 2). It should be noted that our search parameters ensure that all 435 sequences we analyzed are true Cyp51 proteins; however, it is possible that the full complement of Cyp51 paralogs in each of the 295 species was not identified because refSeq is not exclusively based on whole genome sequences.

Fungal Cyp51 proteins fall into 4 groups

To investigate how fungal Cyp51 proteins are related to each other, we used RAxML to create a maximum likelihood tree with 435 fungal Cyp51 proteins and 2 human Cyp51 proteins to root the tree (Supplementary Fig. 1). To aid in visualization, we collapsed branches based on phyla, subphyla, or classes (Fig. 2). We found fungal Cyp51 proteins fell into 4 major groups which we designated as Cyp51, Cyp51A, Cyp51B, and Cyp51C based on naming in previous literature (Fig. 2; Supplementary Fig. 1) (Hawkins *et al.* 2014; Perez-Cantero *et al.* 2020). Cyp51 in members of Saccharomycotina and Taphrinomycotina are also known as "Erg11" in the literature. The topology of our Cyp51 protein tree largely followed the topology of the fungal tree of life [Fig. 2; Supplementary Fig. 1 (James, Stajich, *et al.* 2020)]. Proteins from early-diverging fungi (Blastocladiomycota, Chytridiomycota, Monoblepharidomycota, Zoopagomycota, and Mucormycota), Basidiomycota, Saccharomycotina, and Taphrinomycotina fell into group Cyp51. Our phylogenetic analyses show a divergence of 3 Cyp51 paralogs in filamentous Ascomycota (Fig. 2; Supplementary Fig. 1). Those from Pezizomycotina fell into Cyp51A, Cyp51B, and Cyp51C. Divergence of Cyp51 from the common ancestor of paralogs Cyp51A, Cyp51B, and Cyp51C has strong support (100% bootstrap support), but divergence of paralogs Cyp51A, Cyp51B, and Cyp51C from each other does not have strong support (41%, 36%, and 41% bootstrap support, respectively).

Most fungal species in our study had 1 or 2 Cyp51 paralogs (180/295 and 100/295, respectively) (Fig. 3; Supplementary Table 2). Fewer had 3 (14/295) and only 1 fungus, *Basidiobolus meristosporus*, had 4 copies of Cyp51 (Fig. 3). Fungi with 1 or 2 Cyp51 proteins were found across all taxonomic groups (Fig. 3). Fungi with 3 Cyp51 proteins were found in Basidiomycota and Ascomycota. Members in Pezizomycotina had different combinations of Cyp51 paralogs (Fig. 3). Most species had Cyp51A and Cyp51B paralogs or only Cyp51B (69/171 and 63/171, respectively) (Fig. 3). All members of Pezizomycotina contained a Cyp51B paralog (Fig. 3). As shown in Supplementary Table 2, we named fungal Cyp51 proteins based on the group assignment in Supplementary Fig. 1.

We postulate 2 possible evolutionary paths for Cyp51 paralogs as shown in Fig. 4. In the first possible evolutionary path, shown in Fig. 4a, after an initial Cyp51 duplication paralog C diverged followed by another duplication event and divergence of paralogs A and B. The divergence of paralogs A (41%), B (36%), and C (41%) has low support. In the second possible evolutionary path, shown in Fig. 4b, the poorly supported nodes are removed so that the 3 paralogs diverged after 2 unresolved duplication events or a triplication event placing them on the same branch. In either scenario it is possible that subsequent gene loss(es) or duplication(s) led to species with the different combinations of paralogs shown in Fig. 3.

Distinguishing between these possible evolutionary paths is complicated by the small number of Cyp51C sequences and the relatively low number of characters (414–624 amino acids) in Cyp51 genes resulting in low bootstrap support for some nodes. Our analysis only had 29 Cyp51C sequences compared with Cyp51A and B with 87 and 171 sequences, respectively. To see if we could better resolve the relationships among Cyp51A, Cyp51B, and Cyp51C, we analyzed pairwise conservation of all Cyp51 protein sequences using Geneious Prime (Supplementary Table 3). We found that individual members of the Cyp51 group varied the most from each other, with only 46% similarity. Similarity within the Cyp51A, Cyp51B, and Cyp51C groups was much higher (64.7–68.8%). Comparing between groups, members of the Cyp51 group were 45–50% similar to members of Cyp51A, Cyp51B, or Cyp51C groups while members of Cyp51A, Cyp51B, and Cyp51C groups were roughly 60% similar to each other (Supplementary Table 3). We then examined conservation within the highly conserved motifs (SRS1–6, AGXDTT, PER, EXXR, and FXXGXXXCXG) (Supplementary Table 4). Once more the general trend was that motifs within the Cyp51 group were more variable that those in other groups.

Consensus sequences from the 4 Cyp51 groups were compared with each other to create a visual representation of differences in motifs across groups (Fig. 5; Supplementary Table 4).

The most conserved regions (>95% across all Cyp51 proteins) are in the EXXR, PER, and FXXGXXXCIG motifs (Fig. 5; Supplementary Table 4). These 3 motifs are presumably highly conserved due to their roles in Cyp51 structure and function. The **E**XX**R** and PE**R** motif form the **E**–**R**–**R** triad that stabilizes the core structure of Cyp51, while the FXXGXXXCIG motif is a heme-binding domain that is essential for Cyp51 function (Deng *et al.* 2007; Sezutsu *et al.* 2013). Out of the 17 amino acids within these 3 motifs, 13 are more than 95% conserved (Fig. 5; Supplementary Table 4). Looking at all motifs, there are 59 amino acids shared
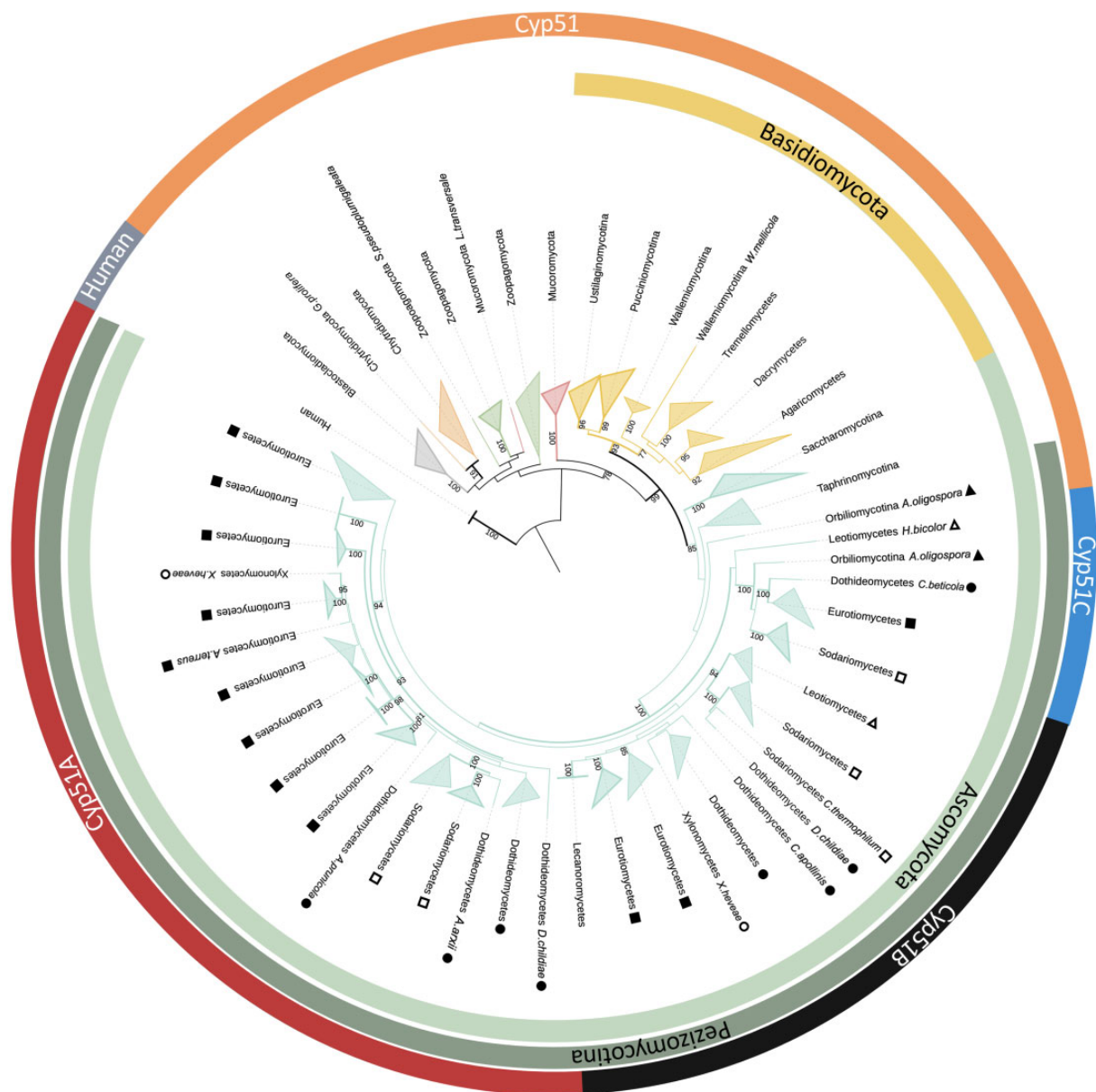
Fig. 2. Cyp51 protein tree for fungi. Maximum likelihood tree of 435 Cyp51 proteins with collapsed branches based on phyla, subphyla, or classes. Branch colors match the colors used for taxonomic clades in the Fungal Tree of Life (James, Stajich, *et al.* 2020). Branches for Blastocladiomycota, Chytridiomycota, Monoblepharidomycota, Zoopagomycota, Mucoromycota, Basidiomycota, and Ascomycota are represented by gray, orange, blue, green, red, yellow, and teal, respectively. Branches with bootstrap support of at least 90 are in bold. Collapsed branches represent phyla, subphyla and classes and are named accordingly. Shapes represent subphyla and classes in Ascomycota. Filled triangles represent subphylum Orbiliomycotina. Empty triangles, filled circles, empty circles, filled squares, and empty squares represent classes Leotiomycetes, Dothideomycetes, Xylonomycetes, Eurotiomycetes, and Sodariomycetes, respectively. The innermost ring shows phyla Basidiomycota (yellow) and Ascomycota (teal). The second ring shows subphylum Pezizomycotina (dark teal). The outer ring shows human Cyp51 proteins used as the outgroup (grey) and 4 groups of fungal Cyp51 proteins—Cyp51, Cyp51C, Cyp51B, and Cyp51A represented by orange, blue, black and red, respectively.

between the 4 consensus sequences with 39 having >95% conservation (Fig. 5; Supplementary Table 4). Cyp51A and Cyp51B have the highest amount of shared amino acids in motifs (88%) (Fig. 5; Supplementary Table 4). Cyp51C has 23 amino acids across motifs that are unique compared with 15 in Cyp51, 8 in Cyp51A, and 5 in Cyp51B, suggesting Cyp51C may be a specialized group, though the low number of Cyp51C sequences included in the analysis might also explain the pattern (Fig. 5).

Research on Cyp51 in filamentous Ascomycota tends to be focused on the Cyp51A paralog because of its role in azole resistance in fungal pathogens of plants and animals. To our surprise,

we found Cyp51A in only half of the species of filamentous Ascomycota we analyzed (86/171) and Cyp51B in all species (Fig. 3; Supplementary Table 2). The low bootstrap support we saw for divergence of Cyp51A from Cyp51B suggests they may play very similar roles with Cyp51B being the essential paralog and the possible Cyp51 ortholog (Fig. 4a). Indeed, in *A. fumigatus*, Cyp51A and Cyp51B act in a compensatory manner; when one paralog is knocked out expression of the other paralog increases (Roundtree *et al.* 2020). Our results confirm and expand previous work by Hawkins *et al.* (2014) that compared 86 fungal Cyp51 proteins in 54 different species and found all species of filamentous
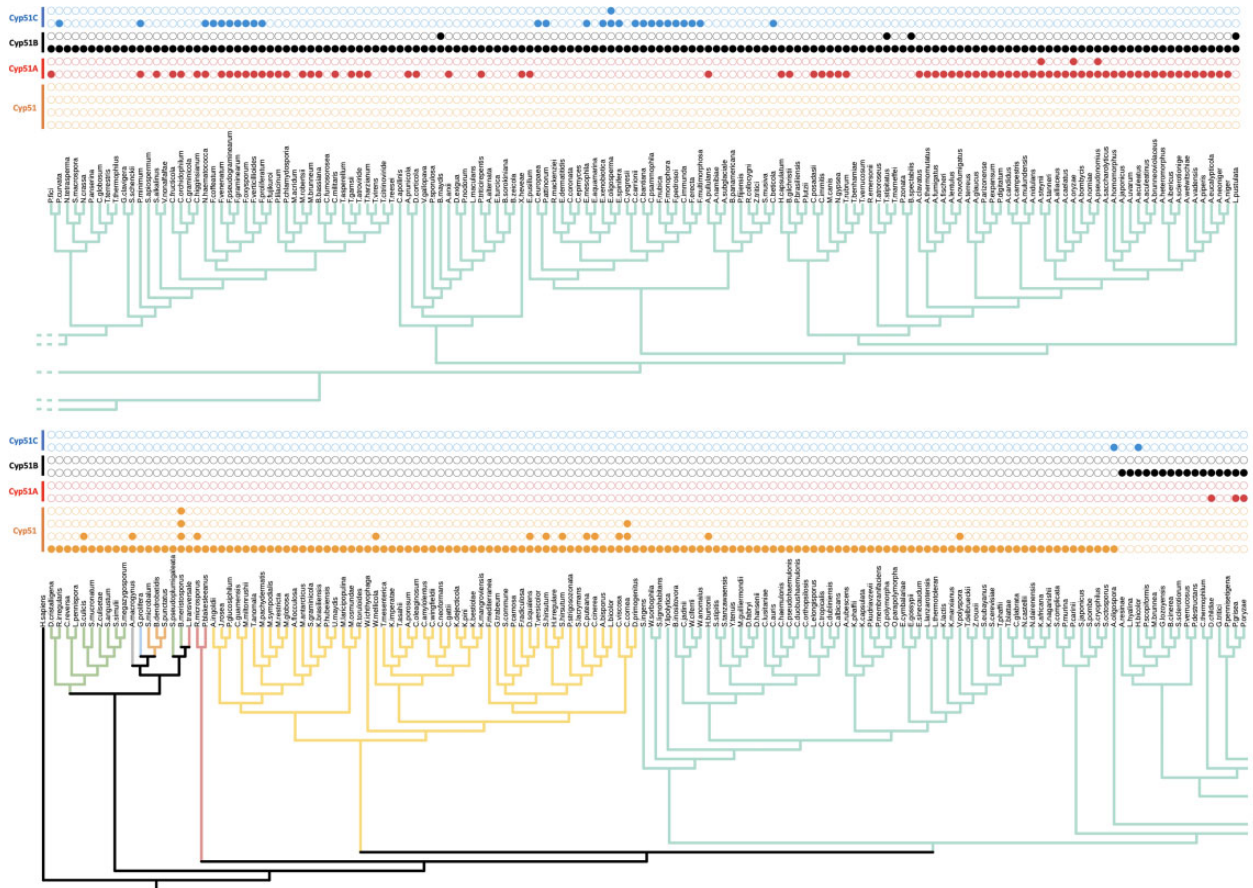
Fig. 3. Fungal species have different numbers of Cyp51 paralogs. Overview of 435 Cyp51 paralogs in 295 fungal species based on Supplementary Table 2 and Supplementary Fig. 1 maximum likelihood tree. Orange, red, black, and blue circles show number of Cyp51, Cyp51A, Cyp51B, and Cyp51C paralogs, respectively. For ease of viewing, where multiple strains represent a species in the maximum likelihood tree, all paralogs were combined in a single entry for that species. Branch colors match the colors used for taxonomic clades in the Fungal Tree of Life (James, Stajich, *et al.* 2020). Branches for Blastocladiomycota, Chytridiomycota, Monoblepharidomycota, Zoopagomycota, Mucoromycota, Basidiomycota, and Ascomycota are represented by gray, orange, blue, green, red, yellow, and teal, respectively.
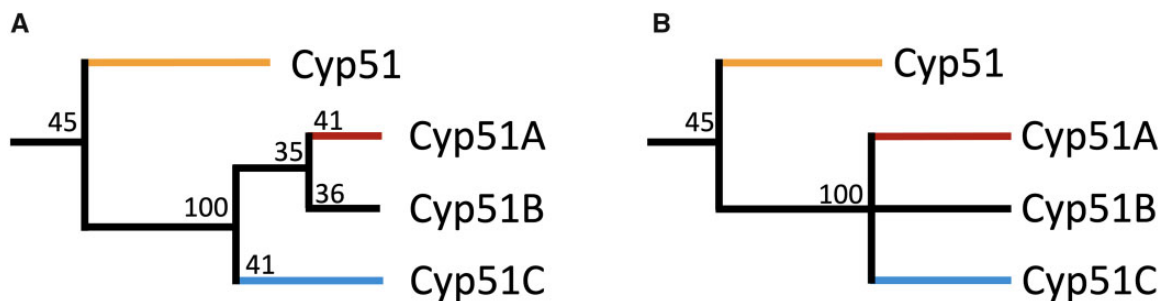

Fig. 4. Possible Cyp51 homolog evolutionary paths. Simplified diagrams of possible Cyp51 evolutionary paths based on Supplementary Fig. 1 (A) and Fig. 2 (B). The Cyp51 branch in early-diverging fungi, Basidiomycota, Saccharomycotina, and Taphrinomycotina. Filamentous Cyp51A, Cyp51B, and Cyp51C branches represent Ascomycota (Pezizomycotina). Numbers represent bootstrap support.

Ascomycota retained a Cyp51B paralog, but Cyp51A had been lost in multiple lineages, and Cyp51C was only found in *Fusarium* spp. (Becher *et al.* 2011; Hawkins *et al.* 2014; James, Lamping, *et al.* 2020). Previous publications have identified Cyp51C in *A. flavus* (Liu *et al.* 2012; Paul *et al.* 2015); however, these sequences were removed from NCBI (https://www.ncbi.nlm.nih.gov/protein/XP_002383931.1/). Other *Aspergillus* spp. have been reported to have a Cyp51C, but these were a duplication of Cyp51A or Cyp51B

(Perez-Cantero *et al.* 2020). Cyp51C was subsequently reported in *Fusarium* spp., *Gibberella zeae*, and *Nectria haematoccoa* (Becher *et al.* 2011; Hawkins *et al.* 2014). We found Cyp51C in 9 other genera (Supplementary Table 2). Interestingly, all are pathogens of plants or animals (Supplementary Table 2). Cyp51C has been shown to be necessary for invasion of plant host tissues in *Fusarium* (Fan *et al.* 2013). This raises the interesting possibility that it could play a similar role in other genera, though further

Fig. 5. Conservation of motifs in Cyp51 proteins across Fungi. Consensus sequences for each of the 4 Cyp51A groups are shown arranged from the N- to C-terminus. Amino acids conserved across all 435 Cyp51 protein sequences are denoted by an *. Amino acids found in more than 95% of all Cyp51s are denoted by >. Lower than 95% conservation in all Cyp51s are denoted by #. Numbers represent amino acid position within motifs.

functional studies are needed to test the role of Cyp51C in invasion and virulence.

Maximum likelihood analysis suggested 2 possible evolutionary paths for Cyp51 paralogs: (1) Cyp51C diverged before Cyp51A and Cyp51B (Fig. 4a); (2) Cyp51A, Cyp51B, and Cyp51C diverged from each other at the same time (Fig. 4b). Based on the higher amino acid conservation between Cyp51A and Cyp51B, the evolutionary path shown in Fig. 4a seems more likely; that is to say Cyp51C is a specialized group of proteins that diverged first (Figs. 4a and 5; Supplementary Table 4). More functional studies

are needed to understand shared and unique roles Cyp51A, Cyp51B, and Cyp51C paralogs play in fungi.

Data availability

All Cyp51 sequences used are listed in Supplementary Table 2 and are publicly available through NCBI (https://www.ncbi.nlm.nih.gov/).

Supplemental material is available at G3 online.

Funding

Conflicts of interest

None declared.

Literature cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–410.

Amos B, Aurrecoechea C, Barba M, Barreto A, Basenko EY, Bażant W, Belnap R, Blevins AS, Böhme U, Brestelli J, *et al.* VEuPathDB: the eukaryotic pathogen, vector and host bioinformatics resource center. Nucleic Acids Res. 2022;50(D1):D898–D911.

Becher R, Weihmann F, Deising HB, Wirsel SG. Development of a novel multiplex DNA microarray for *Fusarium graminearum* and analysis of azole fungicide responses. BMC Genomics. 2011;12:52.

Benedict K, Jackson BR, Chiller T, Beer KD. Estimation of direct healthcare costs of fungal diseases in the United States. Clin Infect Dis. 2019;68(11):1791–1797.

Brown GD, Denning DW, Gow NA, Levitz SM, Netea MG, White TC. Hidden killers: human fungal infections. Sci Transl Med. 2012; 4(165):165rv113.

Buied A, Moore CB, Denning DW, Bowyer P. High-level expression of cyp51B in azole-resistant clinical *Aspergillus fumigatus* isolates. J Antimicrob Chemother. 2013;68(3):512–514.

Cañas-Gutiérrez GP, Angarita-Velásquez MJ, Restrepo-Flórez JM, Rodríguez P, Moreno CX, Arango R. Analysis of the CYP51 gene and encoded protein in propiconazole-resistant isolates of *Mycosphaerella fijiensis*. Pest Manage Sci. 2009;65(8):892–899.

Cools HJ, Bayon C, Atkins S, Lucas JA, Fraaije BA. Overexpression of the sterol 14alpha-demethylase gene (MgCYP51) in *Mycosphaerella graminicola* isolates confers a novel azole fungicide sensitivity phenotype. Pest Manage Sci. 2012;68(7):1034–1040.

Cools HJ, Fraaije BA. Update on mechanisms of azole resistance in *Mycosphaerella graminicola* and implications for future control. Pest Manage Sci. 2013;69(2):150–155.

Córdova P, Gonzalez A-M, Nelson DR, Gutiérrez M-S, Baeza M, Cifuentes V, Alcaíno J. Characterization of the cytochrome P450 monooxygenase genes (P450ome) from the carotenogenic yeast *Xanthophyllomyces dendrorhous*. BMC Genomics. 2017;18(1):540.

Cowen LE, Steinbach WJ. Stress, drugs, and evolution: the role of cellular signaling in fungal drug resistance. Eukaryot Cell. 2008;7(5): 747–764.

Délye C, Laigret F, Corio-Costet MF. A mutation in the 14 alpha-demethylase gene of *Uncinula necator* that correlates with resistance to a sterol biosynthesis inhibitor. Appl Environ Microbiol. 1997;63(8):2966–2970.

Deng JX, Carbone I, Dean RA. The evolutionary history of Cytochrome P450 genes in four filamentous Ascomycetes. BMC Evol Biol. 2007;7(1):1–22.

Denning DW, Perlin DS. Azole resistance in Aspergillus: a growing public health menace. Future Microbiol. 2011;6(11):1229–1232.

Dos Santos RAC, Steenwyk JL, Rivero-Menendez O, Mead ME, Silva LP, Bastos RW, Alastruey-Izquierdo A, Goldman GH, Rokas A. Genomic and phenotypic heterogeneity of clinical isolates of the human pathogens *Aspergillus fumigatus*, *Aspergillus lentulus*, and *Aspergillus fumigatiaffinis*. Front Genet. 2020;11:459.

Dufourc EJ. Sterols and membrane dynamics. J Chem Biol. 2008; 1(1–4):63–77.

Fan J, Urban M, Parker JE, Brewer HC, Kelly SL, Hammond-Kosack KE, Fraaije BA, Liu X, Cools HJ. Characterization of the sterol 14alpha-demethylases of *Fusarium graminearum* identifies a novel genus-specific CYP51 function. New Phytol. 2013;198(3):821–835.

Flowers SA, Colon B, Whaley SG, Schuler MA, Rogers PD. Contribution of clinically derived mutations in ERG11 to azole resistance in *Candida albicans*. Antimicrob Agents Chemother. 2015; 59(1):450–460.

Gadoury DM, Cadle-Davidson L, Wilcox WF, Dry IB, Seem RC, Milgroom MG. Grapevine powdery mildew (*Erysiphe necator*): a fascinating system for the study of the biology, ecology and epidemiology of an obligate biotroph. Mol Plant Pathol. 2012;13(1):1–16.

Garcia-Rubio R, Monzon S, Alcazar-Fuoli L, Cuesta I, Mellado E. Genome-wide comparative analysis of *Aspergillus fumigatus* strains: the reference genome as a matter of concern. Genes (Basel). 2018;9(7):363.

Ghosoph JM, Schmidt LS, Margosan DA, Smilanick JL. Imazalil resistance linked to a unique insertion sequence in the PdCYP51 promoter region of *Penicillium digitatum*. Postharv Biol Technol. 2007; 44(1):9–18.

Gonzalez-Jimenez I, Lucio J, Amich J, Cuesta I, Sanchez Arroyo R. A Cyp51B mutation contributes to azole resistance in *Aspergillus fumigatus*. J Fungi (Basel). 2020;6(4):1–16.

Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, Goodstein D, Kuo A, Minovitsky S, Nikitin R, Ohm RA, *et al.* The genome portal of the Department of Energy Joint Genome Institute. Nucleic Acids Res. 2012;40(Database Issue):D26–D32.

Gsaller F, Hortschansky P, Furukawa T, Carr PD, Rash B, Capilla J, Müller C, Bracher F, Bowyer P, Haas H, *et al.* Sterol biosynthesis and azole tolerance is governed by the opposing actions of SrbA and the CCAAT binding complex. PLoS Pathog. 2016;12(7): e1005775.

Hare RK, Gertsen JB, Astvad KMT, Degn KB, Løkke A, Stegger M, Andersen PS, Kristensen L, Arendrup MC. In vivo selection of a unique tandem repeat mediated azole resistance mechanism (TR120) in *Aspergillus fumigatus* cyp51A, Denmark. Emerg Infect Dis. 2019;25(3):577–580.

Hawkins NJ, Cools HJ, Sierotzki H, Shaw MW, Knogge W, Kelly SL, Kelly DE, Fraaije BA. Paralog re-emergence: a novel, historically contingent mechanism in the evolution of antimicrobial resistance. Mol Biol Evol. 2014;31(7):1793–1802.

Hodiamont CJ, Dolman KM, Ten Berge IJM, Melchers WJG, Verweij PE, Pajkrt D. Multiple-azole-resistant *Aspergillus fumigatus* osteomyelitis in a patient with chronic granulomatous disease successfully treated with long-term oral posaconazole and surgery. Med Mycol. 2009;47(2):217–220.

James JE, Lamping E, Santhanam J, Milne TJ, Abd Razak MF, Zakaria L, Cannon RD. A 23 bp cyp51A promoter deletion associated with voriconazole resistance in clinical and environmental isolates of *Neocosmospora keratoplastica*. Front Microbiol. 2020;11:272.

James TY, Stajich JE, Hittinger CT, Rokas A. Toward a fully resolved fungal tree of life. Annu Rev Microbiol. 2020;74:291–313.

Jefcoate CR, Gaylor JL, Calabrese RL. Ligand interactions with cytochrome P-450. I. Binding of primary amines. Biochemistry. 1969; 8(8):3455–3463.

Kano R, Okubo M, Hasegawa A, Kamata H. Multi-azole-resistant strains of *Cryptococcus neoformans* var. *grubii* isolated from a FLZ-resistant strain by culturing in medium containing voriconazole. Med Mycol. 2017;55(8):877–882.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772–780.

Keizer EM, Valdes ID, Forn-Cuni G, Klijn E, Meijer AH, Hillman F, Wösten HAB, de Cock H. Variation of virulence of five *Aspergillus fumigatus* isolates in four different infection models. PLoS ONE. 2021;16(7):e0252948.

Lepesheva GI, Waterman MR. Sterol 14alpha-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. Biochim Biophys Acta. 2007;1770(3):467–477.

Lepesheva GI, Waterman MR. Structural basis for conservation in the CYP51 family. Biochim Biophys Acta. 2011;1814(1):88–93.

Leroux P, Walker AS. Multiple mechanisms account for resistance to sterol 14alpha-demethylation inhibitors in field isolates of *Mycosphaerella graminicola*. Pest Manage Sci. 2011;67(1):44–59.

Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256–W259.

Liu W, Sun Y, Chen W, Liu W, Wan Z, Bu D, Li R. The T788G mutation in the cyp51C gene confers voriconazole resistance in *Aspergillus flavus* causing aspergillosis. Antimicrob Agents Chemother. 2012;56(5):2598–2603.

Lockhart SR, Frade JP, Etienne KA, Pfaller MA, Diekema DJ, Balajee SA. Azole resistance in *Aspergillus fumigatus* isolates from the ARTEMIS global surveillance study is primarily due to the TR/L98H mutation in the cyp51A gene. Antimicrob Agents Chemother. 2011;55(9):4465–4468.

Mirarab S, Nguyen N, Guo S, Wang L-S, Kim J, Warnow T. PASTA: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences. J Comput Biol. 2015;22(5):377–386.

Paul RA, Rudramurthy SM, Meis JF, Mouton JW, Chakrabarti A. A novel Y319H substitution in CYP51C associated with azole resistance in *Aspergillus flavus*. Antimicrob Agents Chemother. 2015;59(10):6615–6619.

Perez-Cantero A, Lopez-Fernandez L, Guarro J, Capilla J. Azole resistance mechanisms in *Aspergillus*: update and recent advances. Int J Antimicrob Agents. 2020;55(1):105807.

Perlin DS, Rautemaa-Richardson R, Alastruey-Izquierdo A. The global problem of antifungal resistance: prevalence, mechanisms, and management. Lancet Infect Dis. 2017;17(12):e383–e392.

Pfeufer EE, Ngugi HK. Orchard factors associated with resistance and cross resistance to sterol demethylation inhibitor fungicides in populations of *Venturia inaequalis* from Pennsylvania. Phytopathology. 2012;102(3):272–282.

Pound MW, Townsend ML, Dimondi V, Wilson D, Drew RH. Overview of treatment options for invasive fungal infections. Med Mycol. 2011;49(6):561–580.

Rodero L, Mellado E, Rodriguez AC, Salve A, Guelfand L, Cahn P, Cuenca-Estrella M, Davel G, Rodriguez-Tudela JL. G484S amino acid substitution in lanosterol 14-alpha demethylase (ERG11) is related to fluconazole resistance in a recurrent *Cryptococcus neoformans* clinical isolate. Antimicrob Agents Chemother. 2003;47(11):3653–3656.

Roundtree MT, Juvvadi PR, Shwab EK, Cole DC, Steinbach WJ. *Aspergillus fumigatus* Cyp51A and Cyp51B proteins are compensatory in function and localize differentially in response to antifungals and cell wall inhibitors. Antimicrob Agents Chemother. 2020;64(10):1–12.

Schaller H. The role of sterols in plant growth and development. Prog Lipid Res. 2003;42(3):163–175.

Schnabel G, Jones AL. The 14 alpha-demethylase (CYP51A1) gene is overexpressed in *Venturia inaequalis* strains resistant to myclobutanil. Phytopathology. 2001;91(1):102–110.

Sezutsu H, Goff GL, Feyereisen R. Origins of P450 diversity. Philos Trans Roy Soc B Biol Sci. 2013;368(1612):1–11.

Sionov E, Chang YC, Garraffo HM, Dolan MA, Ghannoum MA, Kwon-Chung KJ. Identification of a *Cryptococcus neoformans* cytochrome P450 lanosterol 14alpha-demethylase (Erg11) residue critical for differential susceptibility between fluconazole/voriconazole and itraconazole/posaconazole. Antimicrob Agents Chemother. 2012;56(3):1162–1169.

Snelders E, Karawajczyk A, Schaftenaar G, Verweij PE, Melchers WJ. Azole resistance profile of amino acid changes in *Aspergillus fumigatus* CYP51A based on protein homology modeling. Antimicrob Agents Chemother. 2010;54(6):2425–2430.

Snelders E, van der Lee HAL, Kuijpers J, Rijs AJMM, Varga J, Samson RA, Mellado E, Donders ART, Melchers WJG, Verweij PE, *et al*. Emergence of azole resistance in *Aspergillus fumigatus* and spread of a single resistance mechanism. PLoS Med. 2008;5(11):e219.

Song JX, Zhang SZ, Lu L. Fungal cytochrome P450 protein Cyp51: what we can learn from its evolution, regulons and Cyp51-based azole resistance. Fungal Biol Rev. 2018;32(3):131–142.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313.

Stammler G, Cordero J, Koch A, Semar M, Schlehuber S. Role of the Y134F mutation in cyp51 and overexpression of cyp51 in the sensitivity response of *Puccinia triticina* to epoxiconazole. Crop Protection. 2009;28(10):891–897.

Sun X, Xu Q, Ruan R, Zhang T, Zhu C, Li H. PdMLE1, a specific and active transposon acts as a promoter and confers *Penicillium digitatum* with DMI resistance. Environ Microbiol Rep. 2013;5(1):135–142.

Vallabhaneni S, Mody RK, Walker T, Chiller T. The global burden of fungal diseases. Infect Dis Clin North Am. 2016;30(1):1–11.

Wheat LJ, Connolly P, Smedema M, Durkin M, Brizendine E, Mann P, Patel R, McNicholas PM, Goldman M. Activity of newer triazoles against *Histoplasma capsulatum* from patients with AIDS who failed fluconazole. J Antimicrob Chemother. 2006;57(6):1235–1239.