Fungal Cytochrome P450 Monooxygenases: Their Distribution, Structure, Functions, Family Expansion, and Evolutionary Origin

Wanping Chen[1,2], Mi-Kyung Lee[2], Colin Jefcoate[3], Sun-Chang Kim[4], Fusheng Chen[1], and Jae-Hyuk Yu[2,]*

[1]Department of Food Microbiology, College of Food Science and Technology, Huazhong Agricultural University, Wuhan, Hubei Province, China

[2]Department of Bacteriology and Genetics, University of Wisconsin-Madison

[3]Department of Cell and Regenerative Biology, University of Wisconsin-Madison

[4]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Dae-Jon, Republic of Korea

*Corresponding author: E-mail: jyu1@wisc.edu.

Abstract

Cytochrome P450 (CYP) monooxygenase superfamily contributes a broad array of biological functions in living organisms. In fungi, CYPs play diverse and pivotal roles in versatile metabolism and fungal adaptation to specific ecological niches. In this report, CYPomes in the 47 genomes of fungi belong to the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota have been studied. The comparison of fungal CYPomes suggests that generally fungi possess abundant CYPs belonging to a variety of families with the two global families CYP51 and CYP61, indicating individuation of CYPomes during the evolution of fungi. Fungal CYPs show highly conserved characteristic motifs, but very low overall sequence similarities. The characteristic motifs of fungal CYPs are distinguishable from those of CYPs in animals, plants, and especially archaea and bacteria. The four representative motifs contribute to the general function of CYPs. Fungal CYP51s and CYP61s can be used as the models for the substrate recognition sites analysis. The CYP proteins are clustered into 15 clades and the phylogenetic analyses suggest that the wide variety of fungal CYPs has mainly arisen from gene duplication. Two large duplication events might have been associated with the booming of Ascomycota and Basidiomycota. In addition, horizontal gene transfer also contributes to the diversification of fungal CYPs. Finally, a possible evolutionary scenario for fungal CYPs along with fungal divergences is proposed. Our results provide the fundamental information for a better understanding of CYP distribution, structure and function, and new insights into the evolutionary events of fungal CYPs along with the evolution of fungi.

Key words: cytochrome P450, characteristic motif, fungi, evolution, duplication.

Introduction

Cytochrome P450s (CYPs), constituting a superfamily of heme-containing monooxygenases found in all three domains of life, are involved in the metabolism of a diverse array of endogenous and xenobiotic compounds (Doddapaneni et al. 2005; Bernhardt 2006; Park et al. 2008; Kelly et al. 2009; Moktali et al. 2012). Especially, CYPs extensively participate in a wide variety of physiological reactions in fungi that contribute to the fitness and fecundity of fungi in various ecological niches. Fungi, especially filamentous fungi, produce a vast array of secondary metabolites of biomedical, agricultural, and industrial significance, many of which are biosynthesized with the involvement of various CYPs (Hoffmeister and Keller 2007;

Kelly et al. 2009). For example, some renowned compounds of fungi, such as aflatoxins and lovastatin, are modified by the action of CYPs in their biosynthetic pathways (Kelkar et al. 1997; Barriuso et al. 2011). CYPs are also associated with the physiological traits of fungi, for example, pathogenicity of the fungi (Siewers et al. 2005; Karlsson et al. 2008; Leal et al. 2010). Expansions and functional diversifications of the fungal CYP families have been associated with the evolution of fungal pathogenicity (Soanes et al. 2008). CYPs also contribute to the ecological roles of fungi as saprobes or decomposers. For example, the CYP system in the model white rot fungus *Phanerochaete chrysosporium* is involved in the biodegradation of a vast array of xenobiotic compounds such as

the natural aromatic polymer lignin and a broad range of environmental toxic chemicals (Syed and Yadav 2012). In addition to highly specialized functions, CYPs also play a housekeeping role in fungi. For example, CYP51 involved in sterol biosynthesis is recognized as the housekeeping CYP in fungi, and has been a popular antifungal target for the control of fungal diseases in humans and crop plants (Kelly et al. 2009; Becher and Wirsel 2012).

Nomenclature of CYPs is based on their amino acid sequence similarity. In general, any two CYPs with amino acid sequence identity greater than 40% belong to a single CYP family, and with sequence identity greater than 55% belong to a subfamily (Nelson 2006a). Currently, fungal CYP families are grouped to CYP51–CYP69, CYP501–CYP699, and CYP5001–CYP6999 (Kelly et al. 2009). However, the classification of fungal CYPs has two challenges: The extraordinary diversity of functions and evolution of fungal CYPs and the rapidly increasing number of sequenced fungal genomes (Deng et al. 2007; Hoffmeister and Keller 2007). Accordingly, there are many fungal CYPs remain to be newly assigned. Clans have been proposed as a higher order for grouping CYP families that consistently cluster together on phylogenetic trees (Nelson 2006a). CYP families within a single clan have likely been diverged from a common ancestor gene (Nelson 1999a). However, clan membership parameters have not been clearly defined (Nelson 2006a; Deng et al. 2007). CYP clan arrangements may be slightly different according to the different identity cutoffs. For example, 168 CYP families in four filamentous Ascomycetes were classified into 115 clans, whereas in a recent classification of CYPs from 213 fungal and Oomycete genomes also led to 115 clustered clans (Deng et al. 2007; Moktali et al. 2012).

Generally, fungal CYPs share little sequence similarity, except for a few conserved domains for key characteristics of CYPs, corresponding to the preserved tertiary structure and enzyme functions (Deng et al. 2007; Moktali et al. 2012). The most conserved region FXXGXXXCXG is the heme-binding domain containing the axial Cys ligand to the heme; the motifs EXXR and PER form the E–R–R triad is important for locking the heme pocket into position and to assure stabilization of the core structure; and the motif AGXDTT contributes to oxygen binding and activation (Werck-Reichhart and Feyereisen 2000; Deng et al. 2007; Kelly et al. 2009; Sezutsu et al. 2013). Although CYPs all preserve the basic structural fold, in response to the enormously wide range of substrate specificities, their substrate-binding regions are much more variable, yet may possess a signature motif (Moktali et al. 2012). In addition, most CYPs display significant substrate promiscuity, and therefore, their substrate-binding pockets are well known for the high structural plasticity and the ability to change shape and volume depending on the chemical structure they accommodate (Hargrove et al. 2012). Six putative substrate recognition sites (SRSs) for CYPs have been proposed based on the analysis of the CYP2

family and CYPs structure (Gotoh 1992). Since then, various studies have tried to reveal the interaction between CYPs and substrates by means of X-ray crystallographic analysis or site-directed mutagenesis (Hasler et al. 1994; Hasemann et al. 1995; Harlow and Halpert 1997; Graham and Peterson 1999; Lepesheva et al. 2003; Lepesheva and Waterman 2004). Particularly, the CYP51 family, firstly identified from *Saccharomyces cerevisiae*, has been extensively studied for fundamental CYP structure–function due to its wide presence in all biological kingdoms and its importance as a drug target for the pathogenic fungi (Yoshida and Aoyama 1984; Podust et al. 2001; Lepesheva et al. 2003; Lepesheva and Waterman 2004, 2007; Chen et al. 2010; Becher and Wirsel 2012; Hargrove et al. 2012).

The complete CYP complement of one species, called CYPome, is a collection of CYP genes in the genome of that species (Lamb et al. 2002). Detailed investigation of CYPome evolution could be of great help for better understanding the evolutionary processes of fungi. On the one hand, despite the unclear origin of the CYP family, the ancestral CYP must have emerged early in the evolution of life forms, possibly before atmospheric molecular oxygen appeared on the Earth about 2.4 billion years ago (2.4 Ga), far earlier than the emergence of fungi (Kelly and Kelly 2013; Sezutsu et al. 2013). The phylogenetic analysis suggested that CYP51, CYP61, and CYP530 were present in the last common ancestor of all fungi (Moktali et al. 2012). Except CYP530, which is specific to fungi, CYP51 is widely distributed in all the biological kingdoms whereas CYP61 is frequently found in Plantae and fungi (Morikawa et al. 2006; Nelson 2006b; Moktali et al. 2012; Kelly and Kelly 2013). Interestingly, CYP61 was thought to be evolved from the duplication of CYP51 (Nelson 1999a, 1999b). The phylogenetic relationship of CYPs among remote taxon could provide the information on the origin of fungi. On the other hand, the large biodiversity of fungal CYPs, mainly arose from gene duplication (Feyereisen 2011), is closely associated with fungal living habits in the environments. For example, the large number of CYPs in white- and brown-rot fungi contributes to the breakdown of plant material in the environment (Eastwood et al. 2011; Syed and Yadav 2012). Thus, the expansion and diversification of CYPomes could also provide the information on fungal evolutionary adaptation to ecological niches.

The availability of whole-genome sequences for a number of fungi opens new research avenues to reach a global understanding of the CYPomes. Currently, a number of studies on fungal CYPomes have extensively performed in model fungi, such as *Aspergillus nidulans* and *Penicillium chrysosporium*, and some important fungi such as plant pathogens *Fusarium graminearum* and *Magnaporthe grisea* (Doddapaneni et al. 2005; Deng et al. 2007; Kelly et al. 2009). There are two large and systemic public databases for fungal CYPs: CYP Database (http://drnelson.uthsc.edu/CytochromeP450.html, last accessed June 1, 2014) and

Fungal CYP Database (http://p450.riceblast.snu.ac.kr, last accessed June 1, 2014) (Park et al. 2008; Nelson 2009). In this study, the 47 genomes of fungi from the four traditionally recognized phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota have been surveyed to identify all possible CYP members with hidden Markov models, and these CYPs have been annotated following the International P450 Nomenclature Committee. Then, the characteristic motifs, phylogeny, specific protein features, and SRSs of these CYP proteins are analyzed. Based on the phylogenetic analyses of CYPomes, we propose possible evolutionary events and hypothetical scenarios for the evolution of CYPs along with fungal divergences.

Materials and Methods

Sequence Data

Overall protein sequences of 47 species/strains of fungi from the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota were used in this study. The related information is presented in table 1.

Annotation of CYP Genes

The annotation pipeline of the CYP genes in the selected fungi was in a two-step procedure of identification and annotation. The identification step of CYP family was performed by using HMMER 3.0 (http://hmmer.janelia.org/, last accessed June 1, 2014) with hmmsearch of profile hidden Markov models derived from the Pfam seed alignment flatfile of PF00067 (downloaded from the Pfam protein families database, http://pfam.xfam.org/, last accessed June 1, 2014) against selected fungal proteomes. The cutoff of positive hits was set at E value of 10^{-3}. Then, the positive hits were subject to the annotation procedure involving BLASTP comparisons against the database of all named fungal CYPs (http://blast.uthsc.edu/, last accessed June 1, 2014) (Nelson 2009). These predicted CYPs were assigned to corresponding family types based on their highest sequence similarity (at least 40%) against all named fungal CYPs as followed by the International P450 Nomenclature Committee.

Construction of Phylogenetic Trees

Alignment of annotated CYPs was performed by HMMER package with hmmalign of the corresponding profile hidden Markov models. Then, the phylogenetic trees from alignments of protein sequences were constructed by FastTree version 2.1.4 with maximum-likelihood method (http://www.microbesonline.org/fasttree/, last accessed June 1, 2014) (Price et al. 2009). The tree data were submitted to iTOL (http://itol.embl.de/upload.cgi, last accessed June 1, 2014) for viewing phylogenetic trees and making figures (Letunic and Bork 2007).

Structural Feature Analysis of Protein Sequences

Structural features were explored on homologous protein groups based on their phylogenetic relationships to reveal a clade-specific conservation pattern, essentially conserved within each clade but differing across clade. Multiple protein sequence alignments built by HMMER package were edited by Jalview version 2.7 (Waterhouse et al. 2009). The residues assigned to match states that conserved against the Pfam annotations were reserved for the profile analysis. Consensus logos of the alignments automatically generated by WebLogo were used for visualization of the conservation of primary structure by plotting a stack of amino acids for each position (Schneider and Stephens 1990; Crooks et al. 2004).

Results and Discussion

Distribution of CYPome in the Sequenced Fungal Genomes

The distribution information of the CYP-encoding genes on their frequency, family diversity, and proportion in genomes is summarized in the tested fungi (table 1). All tested fungi from phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota contained CYP genes. It suggested that CYP gene is conserved and plays a critical role in fungi. However, its total in the tested fungi vary greatly, ranging from single to hundreds. Generally, filamentous fungi such as those from the group Eurotiales possess high numbers of CYP genes, but yeasts such as those from the group Saccharomycotina contain a very few CYP genes. The Basidiomycota species show considerable numbers of CYP genes; even some species like *Postia placenta* also possess a large number of CYP genes. The Zygomycota species have plenty of CYP genes whereas the Chytridiomycota species possess a limited number of CYP genes.

Compared with species with high CYP counts from other kingdoms (http://drnelson.uthsc.edu/CytochromeP450.html, last accessed June 1, 2014), generally, the CYP number in filamentous fungi is less than that in plant but more than that in animal. However, when genome size is considered, filamentous fungi have the highest density of CYPs.

There are more than 337 CYP gene families in the tested fungi belonging to CYP51–CYP68, CYP501–CYP698, CYP5025–CYP5307, and CYP6001–CYP6004, and some unassigned families due to their low sequence similarities with currently identified fungal CYPs (supplementary table S1, Supplementary Material online). Based on the frequency distribution, 46.80% of CYPs are classified into CYP501–CYP698, followed by 27.63% into CYP5025–CYP5307, and 22.15% into CYP51–CYP68. As for the individual family, CYP52 shows the highest frequency, accounting for 4.79%, followed by CYP65 (3.73%), CYP51 (3.48%), and CYP61 (2.99%). However, based on the occurrence of CYP families in these fungal species, only CYP51 and CYP61 seem

GBE

Table 1

Distribution of Putative CYPs in 47 Fungal Proteomes

Phylum	Taxonomic Group	Species	Strains	Source	Number	Family Type	Genomic Percentage
Ascomycota	Dothideomycetes	*Leptosphaeria maculans*	JN3	NCBI	66	53	0.25
		Zymoseptoria tritici	IPO323	NCBI	79	60	0.32
	Eurotiales	*Aspergillus flavus*	NRRL 3357	AspGD	153	93	0.65
		As. fumigatus	Af293	AspGD	75	57	0.41
		As. nidulans	FGSC A4	AspGD	119	90	0.64
		Monascus ruber	M7	F.Chen	40	34	0.26
		Penicillium chrysogenum	Wisconsin54-1255	NCBI	98	63	0.48
	Onygenales	*Ajellomyces capsulatus*	G186AR	NCBI	41	35	0.23
		Paracoccidioides brasiliensis	Pb01	NCBI	37	31	0.21
	Leotiomycetes	*Botryotinia fuckeliana*	B05.10	NCBI	121	67	0.42
	Orbiliomycetes	*Arthrobotrys oligospora*	ATCC 24927	NCBI	37	28	0.16
	Pezizomycetes	*Tuber melanosporum*	Mel28	NCBI	28	21	0.03
	Saccharomycotina	*Candida albicans*	WO-1	NCBI	9	6	0.10
		C. dubliniensis	CD36	NCBI	10	6	0.11
		C. glabrata	CBS 138	NCBI	3	3	0.04
		C. tropicalis	MYA-3404	NCBI	12	6	0.13
		Clavispora lusitaniae	ATCC 42720	NCBI	8	6	0.11
		Debaryomyces hansenii	CBS767	NCBI	9	5	0.12
		Eremothecium cymbalariae	DBVPG#7215	NCBI	1	1	0.02
		E. gossypii	ATCC 10895	NCBI	3	3	0.05
		Kluyveromyces lactis	NRRL Y-1140	NCBI	5	5	0.07
		Komagataella pastoris	CBS 7435	NCBI	4	4	0.07
		Lachancea thermotolerans	CBS 6340	NCBI	3	3	0.05
		Lodderomyces elongisporus	NRRL YB-4239	NCBI	10	5	0.10
		Meyerozyma guilliermondii	ATCC 6260	NCBI	9	6	0.13
		Naumovozyma castellii	CBS 4309	NCBI	3	3	0.04
		Ogataea parapolymorpha	DL-1	NCBI	5	5	0.09
		Saccharomyces cerevisiae	YJM789	NCBI	3	3	0.04
		Scheffersomyces stipitis	CBS 6054	NCBI	10	6	0.10
		Tetrapisispora phaffii	CBS 4417	NCBI	3	3	0.04
		Torulaspora delbrueckii	CBS 1146	NCBI	3	3	0.05
		Yarrowia lipolytica	CLIB122	NCBI	17	6	0.13
		Zygosaccharomyces rouxii	CBS 732	NCBI	3	3	0.05
	Sordariomycetes	*Hypocrea jecorina*	QM6a	NCBI	73	51	0.35
		Magnaporthe oryzae	70-15	NCBI	135	78	0.52
		Neurospora crassa	OR74A	NCBI	41	39	0.17
	Taphrinomycotina	*Schizosaccharomyces japonicus*	yFS275	NCBI	2	2	0.03
		S. pombe	972h-	NCBI	2	2	0.02
Basidiomycota	Agaricomycotina	*Cryptococcus gattii*	WM276	NCBI	5	5	0.04
		Laccaria bicolor	S238N-H82	NCBI	76	22	0.19
		Postia placenta	Mad-698-R	NCBI	106	39	0.20
	Pucciniomycotina	*Melampsora larici-populina*	98AG31	NCBI	29	14	0.04
		Puccinia graminis f. sp. *tritici*	CRL 75-36-700-3	NCBI	18	9	0.03
	Ustilaginomycotina	*Sporisorium reilianum*	SRZ2	NCBI	15	14	0.14
		Ustilago maydis	521	BI	20	17	0.17
Chytridiomycota	Chytridiomycetes	*Batrachochytrium dendrobatidis*	JAM81	NCBI	9	7	0.06
Zygomycota	Mucoromycotina	*Rhizopus oryzae*	RA 99-880	BI	49	14	0.15

NOTE.—Taxonomy information of above fungi is extracted from Taxonomy Browser in NCBI (http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi, last accessed June 1, 2014). The overall protein sequences were downloaded from the AspGD (http://www.aspgd.org/, last accessed June 1, 2014), the Broad Institute (BI, http://www.broadinstitute.org/scientific-community/data, last accessed June 1, 2014), the JGI (http://genome.jgi.doe.gov/programs/fungi/index.jsf, last accessed June 1, 2014), and NCBI (http://www.ncbi.nlm.nih.gov/genome/browse/, last accessed June 1, 2014). Putative CYP proteins were identified by HMMER searches against overall protein sequences of each species with the corresponding profile hidden Markov model from Pfam (http://pfam.xfam.org/, last accessed June 1, 2014) and their positive hits were annotated following by BLASTP comparisons against the database of all named fungal P450s (http://blast.uthsc.edu/, last accessed June 1, 2014). Genomic percentage was based on the proportion of overall CYP gene sequences in genomes.

widespread in fungi. CYP51 is present in 46 out of 47 species (absent in *P. placenta*), covering species from the phyla Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota whereas CYP61 is found in 42 fungal species (absent in *Batrachochytrium dendrobatidis*, *Eremothecium cymbalariae*, *Melampsora larici-populina*, *P. placenta*, and *Puccinia graminis* f. sp. *tritici*). Moreover, the CYP51 and CYP61 genes are conserved in their number. Generally, most fungi have one each of the CYP51 and CYP61 gene, but some species have two CYP51 genes (even three CYP51 genes in *As. flavus* and three CYP61 genes in *Rhizopus oryzae*). The conserved distribution of CYP51 and CYP61 implies their important roles played in fungi. Previous studies suggested that only CYP51 and CYP61 play housekeeping functions in sterol biosynthesis at least in filamentous fungi (Kelly et al. 2009). It is worth mentioning that CYP51 is found in all kingdoms: plants, animals, lower eukaryotes, and bacteria, and is a common target of antifungal drugs (e.g., miconazole and ketoconazole) that inhibit CYP51 activity and formation of ergosterol (Nelson 2009; Nebert et al. 2013). Probably, the absence of CYP51 and CYP61 genes in the above-mentioned species could be due to their obligate lifestyle, wherein they may utilize essential sterols from their plant/animal hosts (Moktali et al. 2012).

There are some locally frequent CYP families. CYP52 and CYP56 are frequently present in the phyla Ascomycota. CYP65, CYP68, CYP505, CYP532, CYP537, CYP539, CYP540, CYP548, CYP578, CYP584, CYP617, and CYP682 are widely distributed in filamentous fungi from the Ascomycota, whereas CYP501 and CYP5217 are common in the yeasts (Saccharomycotina). CYP53, CYP504, CYP530, CYP505, and CYP6001 are found both in the Ascomycota and Basidiomycota. The related information could contribute to our understanding of the relationship between fungal taxonomy and CYP families. For example, the frequent presence of CYP52 in the Ascomycota might suggest the emergence of a progenitor CYP52 in the ancestral Ascomycota. Meanwhile, it also indicates that CYP52—oxidation of *n*-alkyl chains— might play an important role in the Ascomycota (Sanglard and Loper 1989).

In general, despite relative species showed some similarities in CYPome distribution, family diversity of CYP genes differs considerably between species. It is not only reflected in their family number, but also their family type. Take close relatives in *Aspergillus* as an example, *As. flavus*, *As. fumigatus*, and *As. nidulans* possess 93, 57, and 90 family types, respectively, but only 45 types being shared. Gene duplications are common especially in fungal species with numerous CYPs. For example, there are seven CYP620 genes in *As. flavus*, and 12 CYP52 genes in *Yarrowia lipolytica*. The expansion of CYP genes may be related to the potential demand of some new physiological processes.

We also addressed whether CYP gene expansion is associated with the genome size by investigating the percentage

of overall CYP gene sequence size taken up in each fungal genome in the 47 species (table 1). The results show large variations of the percentage among species. Generally, filamentous fungi especially from the genus *Aspergillus* have high proportions of CYP genes in their genomes whereas yeasts possess pretty low proportions. For the phylum of Basidiomycota, fungi from Agaricomycotina and Ustilaginomycotina have obvious higher proportions than those from Pucciniomycotina. *Rhizopus oryzae*, as the representative filamentous fungus from the phylum Zygomycota, shows a moderate proportion compared with the tested fungi, whereas the chytrid fungus *B. dendrobatidis* has a low proportion of CYP genes in its genome. Therefore, the great difference in proportions between species makes it clear that CYP gene expansion is not necessarily correlated to the genome size.

Characteristic Motifs of the Fungal CYPs

In agreement with the HMM logo from CYP family on Pfam (http://pfam.xfam.org/family/PF00067, last accessed June 1, 2014), the primary structure analysis showed a few very well-conserved sequence regions despite a considerable variation in sequence. These identifiable sequence motifs correspond to the conserved tertiary structure and enzyme functions in spite of the wide sequence diversity and functions of CYPs. There are four widely recognized consensus regions and they greatly facilitate the detection of CYPs from genomes. The most characteristic motif FXXGXRXCXG (located at position d in fig. 1) is designated as the heme-binding domain (Kelly et al. 2009). It was worth mentioning that Cys herein was previously recognized as the invariant residue across all CYP genes and the indispensable role in ligand to the heme (Werck-Reichhart and Feyereisen 2000; Deng et al. 2007; Kelly et al. 2009; Sezutsu et al. 2013). We found six potential exceptions in the annotated CYPs, where the conserved Cys is replaced with Arg (gi|242210285| from *P. placenta*), Asn (gb|EGR45174.1| from *Trichoderma reesei*), Phe (gb|EHA56235.1| from *M. oryzae*), Pro (RO3T_07773 from *R. oryzae*), Tyr (gb|EHA51695.1| from *M. oryzae*), or Val (gb|EGF81099.1| from *B. dendrobatidis*). However, modifications in the heme-binding domain are more frequently found in CYPs with catalytic activity, often not requiring oxygen, and may indicate novel catalytic activities in these exceptions (Song et al. 1993; Li et al. 2008; Kelly et al. 2009). The second conserved motif EXXR (located at position b in fig. 1) and the third consensus PER (located at position c in fig. 1) form E–R–R triad that is important for locking the heme pocket into position and to assure stabilization of the core structure (Deng et al. 2007; Kelly et al. 2009). The forth relatively conserved motif AGXDTT (located at position a in fig. 1) contributes to oxygen binding and activation (Kelly et al. 2009).

We then compared the sequence logos of the conserved motifs from the tested fungi against those of taxonomic group
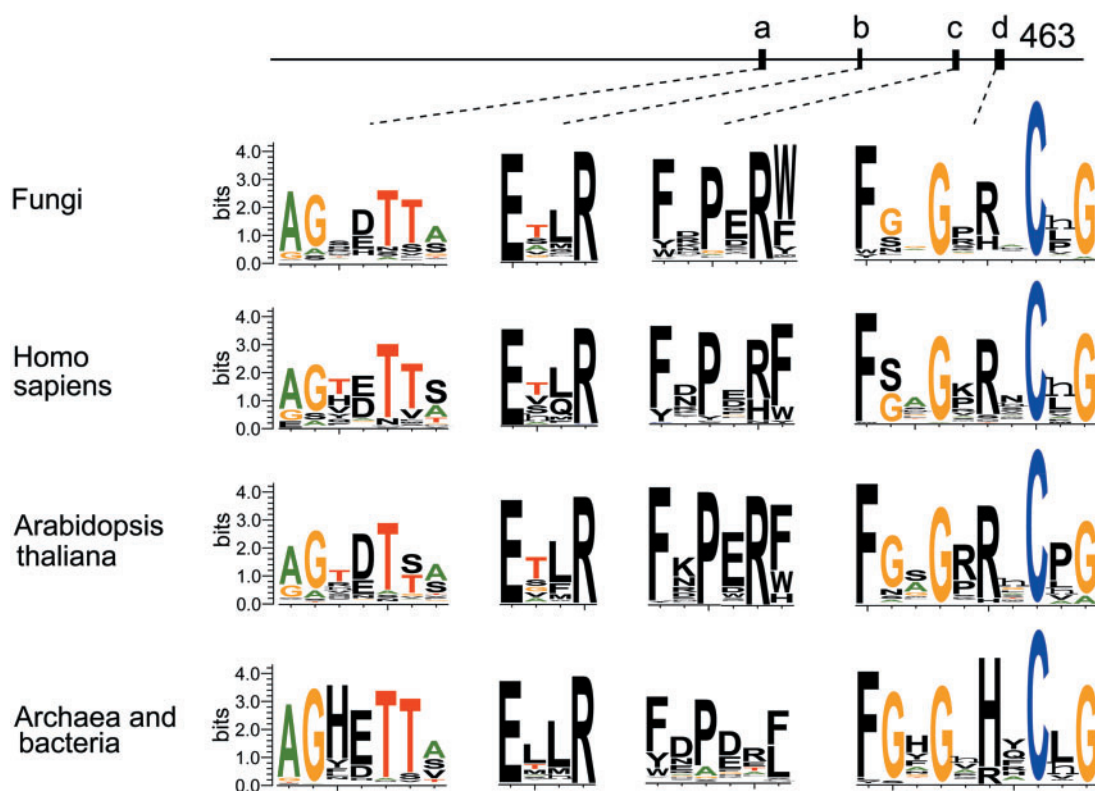
FIG. 1.—Sequence logos of the conserved CYP motifs from the tested fungi and their comparison against human, plant, and prokaryotes. The CYP proteins from *Homo sapiens* (60 CYPs), *Arabidopsis thaliana* (288 CYPs), archaea (27 CYPs), and bacteria (1,105 CYPs) were download from the CYP Database (http://drnelson.uthsc.edu/CytochromeP450.html, last accessed June 1, 2014). Multiple alignments of CYP proteins were performed by aligning them to the profile hidden Markov model of PF00067 with HMMER package. Residues assigned to match states were reserved for the profile analysis and their consensus logos were generated by WebLogo (http://weblogo.threeplusone.com/create.cgi, last accessed June 1, 2014) (Schneider and Stephens 1990; Crooks et al. 2004). The four regions a, b, c, and d correspond to the positions 273–279, 330–333, 383–388, and 405–414, respectively.

animal, plant, archaea, and bacteria to identify some potential signatures assigned to the fungal CYPs (fig. 1). Interestingly, these taxonomic groups showed some noticeable differences among the motifs in spite of their widely recognized conservativeness. In general, the most obvious distinction of the conserved motifs among these taxonomic groups is reflected in those from prokaryotes (represented by archaea and bacteria) against eukaryotes (represented by the tested fungi, *Homo sapiens* and *Arabidopsis thaliana*). It may suggest the early divergence of CYP evolution between prokaryotes and eukaryotes. Meanwhile, it also may improve our understanding of the relationship between CYP structure and function. It is surprising that the motifs in prokaryotes are obviously different from the widely recognized CYP motifs (http://pfam.xfam.org/family/PF00067#tabview=tab4, last accessed June 1, 2014). For example, prokaryotes CYPs have the predominant His to replace Arg in the heme-binding motif FXXGXRXCXG. In addition, the motif PER is not well conserved in prokaryotes CYPs. Relatively, comparison of CYP conserved motifs from the tested fungi, *Ar. thaliana* and *H. sapiens*, showed that

these taxonomic groups shared high similarities of the CYP motifs, likely suggesting the conservative evolution of CYP motifs in eukaryotes. The most distinguishable feature of the fungal CYPs against CYPs of *Ar. thaliana* and *H. sapiens* in the conserved motifs is that fungal CYPs have the predominant Trp over Phe in the motif PERF/W. The difference of CYP conserved motifs among taxonomic groups may provide some information on CYP evolution, structure, and function, even species evolution.

Phylogenetic Tree of Fungal CYPs and Their Clade Features

The phylogeny of all annotated CYPs was constructed based on their consensus sequences and their clade features were analyzed (fig. 2). Results show numerous branches of CYPs in the phylogenetic tree, indicating their highly evolved divergence. However, distribution of CYPs in phylogeny is varying between taxonomic groups. Particularly, CYPs from the group Eurotiales are highly evolved and widespread in many
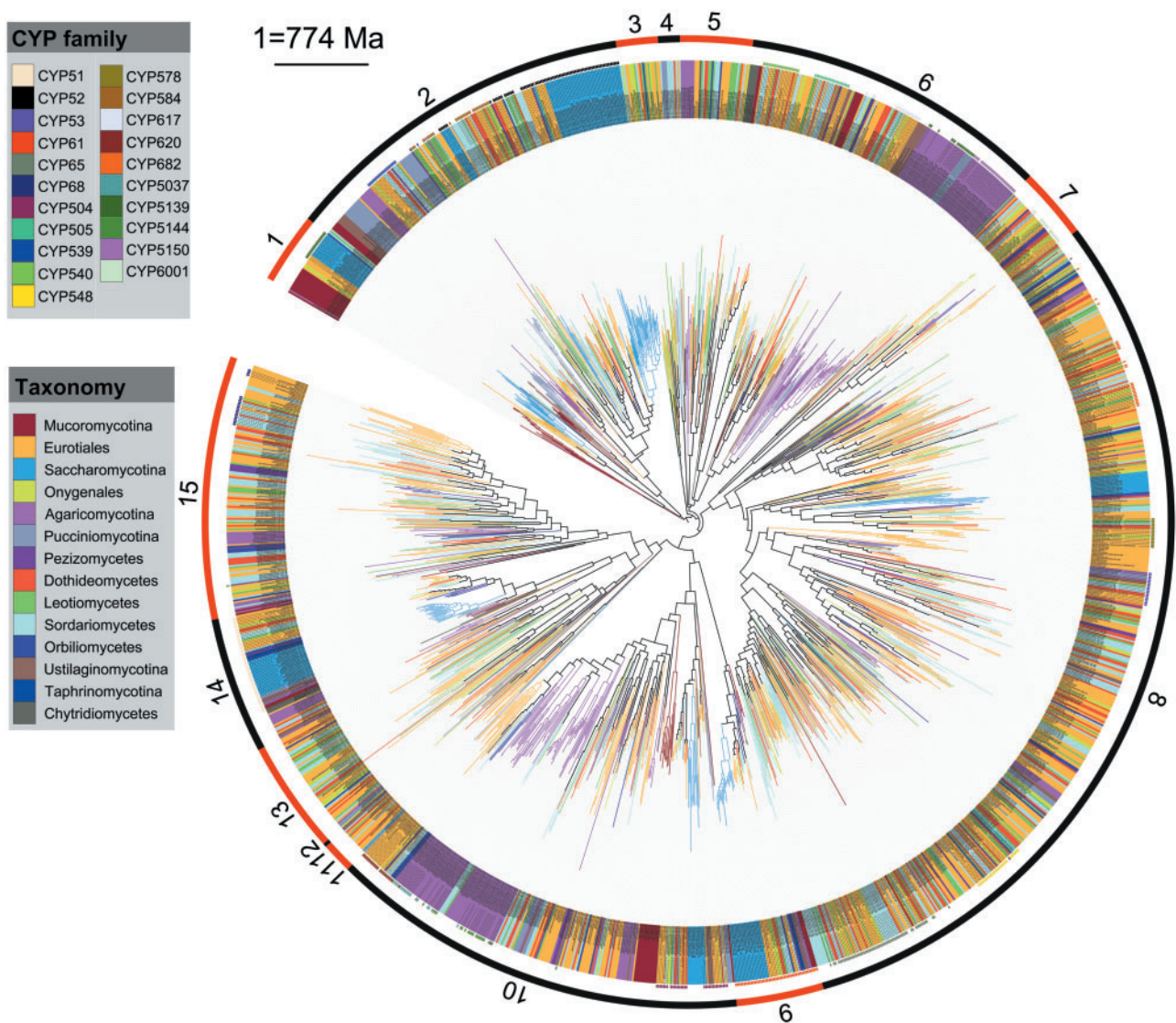
Fig. 2.—Phylogenetic tree of the annotated fungal CYPs. The inner circle is the phylogenetic tree based on the consensus sequences of fungal CYPs. The branches with different colors show their taxonomic groups, as indicated in the legend. The middle circle is the corresponding CYPs, which are covered by different colors to show their taxonomic groups (please refer to supplementary fig. S1, Supplementary Material online, for high resolution one). Each taxon links the branch with a dotted line. Distribution of CYP families is indicated by the scattered colored blocks outside the corresponding taxons, only presenting CYP families with frequencies over 1% among the annotated CYPs. The outer numbers indicate the 15 clades derived in this study, and their ranges are marked by alternating red and black. The calibration of evolutionary rate in CYPs was based on CYP51 and CYP61 (table 3).

branches. But CYPs from some taxonomic groups seem conserved. It is worth mentioning that CYPs from the group Saccharomycotina are gathered in few branches, suggesting their conservatism in evolution. Distribution of CYP families with frequencies over 1% in the tested fungi shows that CYPs in the same family are generally clustered together in the phylogenetic tree, which suggests that the consensus sequences extracted from complete CYP proteins by adjusting to the profile hidden Markov model (PF00067) could well reflect the core domain of CYPs.

Clans have been proposed as a higher order for grouping CYPs, defining as groups of CYP families that consistently cluster together on phylogenetic trees (Nelson 2006a). CYPs within a clan likely diverged from a common ancestor gene (Nelson 1999a). However, clan membership parameters have not been clearly defined (Nelson 2006a; Deng et al. 2007). Thus, we classified CYP families in the same distinctive clade of phylogenetic tree into clans. The fungal CYPs are gathered into 15 clades based on their phylogenetic relationships (fig. 2). Moreover, the distribution of CYP families and fungi

SMBE

Table 2

Distribution of CYP Families in the 15 Fungal CYPs Clades

Clade	CYP Family	Phyla
1	CYP56, CYP661, CYP509, CYP5099, CYP5211, and CYP5212	Ascomycota, Basidiomycota, and Zygomycota
2	CYP52, CYP63, CYP66, CYP509, CYP538, CYP539, CYP544, CYP584, CYP585, CYP655, CYP656, CYP5025, CYP5026, CYP5087, CYP5113, CYP5202, CYP5203, CYP5216, CYP5221, CYP5233, and CYP5288	Ascomycota, Basidiomycota, and Zygomycota
3	CYP59, CYP586, CYP587, CYP662, CYP5192, and CYP5247	Ascomycota
4	CYP526, CYP591, CYP5173, and CYP5230	Ascomycota and Basidiomycota
5	CYP534, CYP589, CYP590, CYP666, CYP667, CYP5075, CYP5106, CYP5141, CYP5154, CYP5171, CYP5181, CYP5228, CYP5243, and CYP5305	Ascomycota, Basidiomycota, and Chytridiomycota
6	CYP505, CYP540, CYP541, CYP547, CYP581, CYP582, CYP617, CYP618, CYP5031-5034, CYP5070, CYP5137, CYP5139, CYP5150, CYP5151, CYP5155, CYP5179, CYP5198, CYP5205, CYP5210, CYP5215, CYP5224, CYP5226, CYP5227, CYP5250, and CYP5287	Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota
7	CYP55, CYP549, CYP687, CYP5116, CYP5190, and CYP6001-6004	Ascomycota and Basidiomycota
8	CYP53, CYP57, CYP58, CYP60, CYP62, CYP65, CYP67, CYP507, CYP511, CYP527, CYP528, CYP531, CYP532, CYP535-537, CYP542, CYP548, CYP551, CYP552, CYP561-568, CYP570, CYP572-578, CYP583, CYP627-632, CYP643, CYP663, CYP669-684, CYP5028-5030, CYP5035, CYP5043, CYP5044, CYP5062, CYP5064, CYP5076-5078, CYP5080, CYP5081, CYP5083, CYP5089, CYP5092, CYP5095, CYP5096, CYP5102, CYP5104, CYP5105, CYP5109, CYP5114, CYP5121, CYP5128, CYP5132, CYP5140-5142, CYP5168, CYP5178, CYP5187, CYP5188, CYP5194, CYP5196, CYP5197, CYP5199, CYP5208, CYP5217, CYP5223, CYP5234, CYP5246, CYP5252, CYP5257, and CYP5307	Ascomycota, Basidiomycota, and Zygomycota
9	CYP61	Ascomycota, Basidiomycota, and Zygomycota
10	CYP64, CYP501, CYP502, CYP504, CYP529, CYP530, CYP533, CYP543, CYP545, CYP546, CYP592, CYP593, CYP619-621, CYP664, CYP665, CYP5027, CYP5037, CYP5042, CYP5046, CYP5047, CYP5050, CYP5052, CYP5053, CYP5056, CYP5058, CYP5063, CYP5065, CYP5066, CYP5068, CYP5069, CYP5097, CYP5108, CYP5146, CYP5148, CYP5152, CYP5158, CYP5206, CYP5207, CYP5209, CYP5220, and CYP5231	Ascomycota, Basidiomycota, and Zygomycota
11	CYP613, CYP685, CYP686, CYP5082, CYP5251, and CYP5286	Ascomycota and Basidiomycota
12	Unassigned	Basidiomycota
13	CYP550, CYP553, CYP610-612, CYP633, CYP635, CYP637-639, CYP641, CYP642, CYP657-660, CYP5090, CYP5100, CYP5101, CYP5111, CYP5189, CYP5201, CYP5222, CYP5232, CYP5240, CYP5248, CYP5249, CYP5263, CYP5274, and CYP5278	Ascomycota and Basidiomycota
14	CYP51, CYP609, CYP5060, CYP5156, CYP5193, CYP5225, CYP5229, CYP5282, and CYP5301	Ascomycota, Basidiomycota, Chytridiomycota, and Zygomycota
15	CYP54, CYP68, CYP503, CYP512, CYP559, CYP560, CYP595-599, CYP601-608, CYP622, CYP623, CYP646-654, CYP698, CYP5048, CYP5061, CYP5067, CYP5073, CYP5074, CYP5085, CYP5086, CYP5091, CYP5093, CYP5103, CYP5107, CYP5110, CYP5125, CYP5144, CYP5157, CYP5191, CYP5195, CYP5200, CYP5204, CYP5213, CYP5245, CYP5281, CYP5284, CYP5285, and CYP5289	Ascomycota, Basidiomycota, and Zygomycota

taxonomy is investigated in the 15 clades (table 2). Clade 8 named CYP53 clan has the most family members, more than 106 CYP families, whereas Clade 9 named CYP61 clan is constituted by only CYP61. It seems that CYP61 is unique compared with other families in the phylogenetic tree, which might imply that CYP61 is evolutionarily conserved and the progenitor CYP61 has not evolved into other families. Clade 15 (CYP54 clan) is also a large branch with more than 56 CYP families. With respect to fungal taxonomy, Clades 6 (CYP505 clan) and 14 (CYP51 clan) consist of members from four

phyla, suggesting the presence of their progenitors in the early fungi. It has been widely considered that CYP51 was present in the primitive fungi, and even in the ancestral eukaryotes (Moktali et al. 2012). Clades 1 (CYP56 clan), 2 (CYP52 clan), 5 (CYP534 clan), 8 (CYP53 clan), 9 (CYP61 clan), 10 (CYP64 clan), and 15 (CYP54 clan) cover members from three phyla. The wide taxonomy of above clades indicates their long evolutionary histories. Meanwhile, members of Clades 7 (CYP55 clan), 11 (CYP613 clan), and 13 (CYP550 clan) are all from the phyla Ascomycota and Basidiomycota,

which cues the CYP family expansion in the common ancestor of the Ascomycota and Basidiomycota. Particularly, members in Clade 3 (CYP59 clan) are solely of Ascomycota, whereas those in Clade 12 are solely of Basidiomycota. It is worth mentioning that Clade 12 consists of few members with unassigned families due to their low sequence similarities against currently identified CYP families.

However, it needs to be mentioned that CYP clan arrangements may be slightly different due to the different identity cutoffs. For example, in the studies of CYPome for four filamentous Ascomycetes, 168 CYP families were classified into 115 clans (Deng et al. 2007). In a recent classification of CYP proteins from 213 fungal and Oomycete genomes, 115 clans were clustered, too (Moktali et al. 2012). Certainly, classifications of clans are not of conflict, all based on the phylogenetic relationship of CYP families, just arose from different cutoffs. In our opinion, due to large numbers of currently identified and new emerging fungal CYPs, the cutoff should be broaden to cluster more CYP families into a clan, avoiding too numerous clans to handle. In this study, 1,607 fungal CYPs are clustered into 15 clans, which would be helpful for the evolutionary studies.

Structure and Function of Fungal CYPs

Structural features among the 15 clades were compared (supplementary fig. S2, Supplementary Material online). There are obvious differences among the 15 clades in their primary structures even in the characteristic domains. Especially, the conserved motif (correspond to signature AGXDTT at position a in fig. 1 and the consensus sequences at position 273–278 in supplementary fig. S1, Supplementary Material online) contributing to oxygen binding and activation is varying greatly. For example, this motif in Clades 5 (AGHETTA) and 6 (AGHETTS/A) is similar to that in archaea and bacteria (AGHETTA/S in fig. 1). However, we tend to maintain that the motif similarity of Clades 5 and 6, and CYPs in archaea and bacteria more likely reflect their functional, rather than phylogenetic, relationships. The motif in Clade 9 (ASQDAS/T) consisting of sole CYP61 family is also very unique. It is noteworthy that Clade 9 shows a high degree of consensus on the whole sequence, which is distinguishable among the 15 clades. It is supposed to be one of the oldest CYP forms, which may help to understand the early structure of CYPs. Relatively, the other three characteristic motifs are conserved in spite of minor differences among clades. In addition, deletions of CYPs are reflected in some clades. Notably, Clade 12 shows extensive deletions with the residues from 259 to 456, basically keeping the range of four characteristic motifs. Even more, some CYPs in Clade 7 have lost the characteristic motif at position 273–278. It may be inferred that maintaining the basic functions of CYPs needs at least three core motifs (domains b, c, and d in fig. 1).

Are there any conserved domains or residues except the four characteristic motifs? After comparison of consensuses among clades, seven regions were of considerable consistency at least within clades, most of which adjacent to the four characteristic motifs. These seven conserved domains in the tested fungi are $W/HX_3RK/RX_5F$ (position 95–106), DX_6FG (position 151–159), LX_3PX_5LRXE (position 289–302), LPYLXAV (position 321–327), $RX_{13}PXG$ (position 344–360), H/NR/HD/NP/EXXF/W/YPD/NP/A (position 371–380), and $F/LAXXEX_{11}F/Y$ (position 417–433). Probably, these domains also play an important role in functions of CYPs.

CYP enzymes participate in a large number of metabolic reactions, collectively involving thousands of substrates (Guengerich 2007). It is of fundamental importance to investigate the interaction between CYPs and their substrates. The six predicted SRSs were firstly proposed for the largest and most catalytically diverse CYP family (CYP2) (Gotoh 1992). In addition, the location of substrate-binding residues in the same secondary structural elements has been found in other CYPs (Nebert et al. 2013). Thus, in this study, fungal CYP51 and CYP61 were used for SRSs analysis due to their housekeeping functions for fungi, and their wide presence and strict substrate specificity. A number of studies have been performed on structure/function relation in CYP51 family and six SRSs have been identified (Aoyama et al. 1996; Podust et al. 2001; Eastwood et al. 2011). The structural feature of fungal CYP51 SRSs was compared with their animal, bacteria, and plant counterparts (fig. 3A). However, some SRSs were not consistent among different taxonomic groups. For example, SRS2 and SRS6 seem to be changeable, which may suggest their noncritical role in substrate recognition. Meanwhile, SRS1 and SRS4 are the most conserved regions and the two corresponding motifs, $YXXF/LX_5PXFGXXVXF/YD$ (position 72–90, SRS1) and GQHT/SS (position 274–278, SRS4), have been proposed as CYP51 signature that can be used to identify a CYP as a CYP51 family member (Lepesheva and Waterman 2007). However, this signature, especially for the motif in SRS4, may not do the same for the CYP51s from bacteria. For example, obviously, for the motif in SRS4, bacteria CYP51s had His replaced with Gln[275]. It is worth mentioning that, for the motif in SRS1, F/L[75] is recognized as the phyla-specific residue (F in plant and L in animal/fungal CYP51s), which leads to their different substrate preferences (Lepesheva and Waterman 2007). At this phyla-specific position, bacteria CYP51s stand together with plant CYP51s, which may suggest their similar substrate preferences. Generally, SRSs characteristics of bacteria CYP51s are more similar to those of plant CYP51s, whereas fungal CYP51s are closer to animal CYP51s, which may also be reflected in their phylogenetic relationships (fig. 3C). Regardless of low sequence identity in the family, the conserved amino acid residues in SRSs ensure a common configuration of CYP51 substrate-binding pockets. Comparison of CYP51 SRSs across the biological kingdoms is
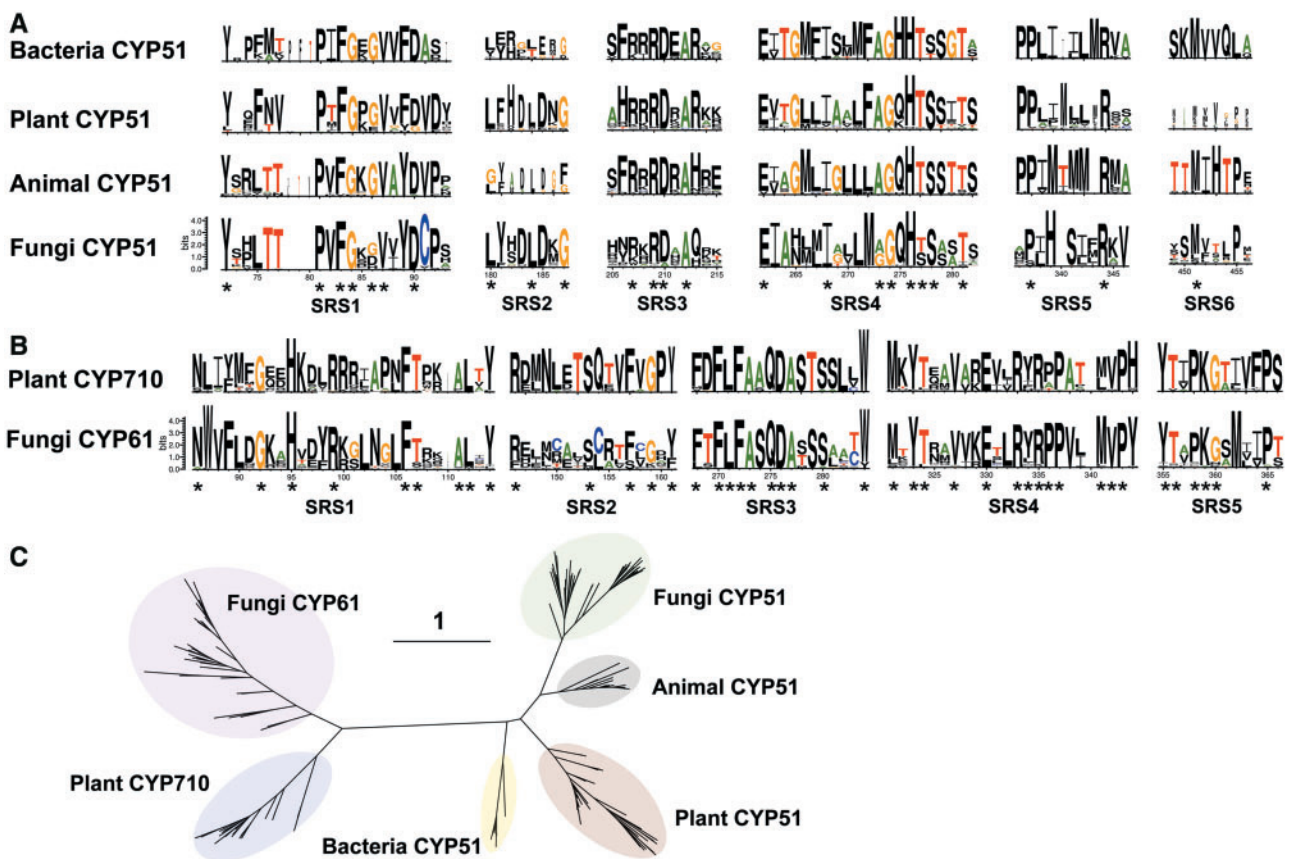
Fig. 3.—Predicted SRSs of fungal CYP51 and CYP61 and their comparisons to animal, plant, and bacterial counterparts. The consensus logos were generated by WebLogo (http://weblogo.threeplusone.com/create.cgi, last accessed June 1, 2014) based on 19 animal CYP51s, 10 bacteria CYP51s, 42 plant CYP51s, and 30 plant CYP710s (extracted from the CYP Database, http://drnelson.uthsc.edu/CytochromeP450.html, last accessed April 19, 2014). The conserved residues are indicated by asterisks. (A) Comparison of fungal CYP51 SRSs with animal, bacteria, and plant counterparts. CYP51 SRSs are based on the analysis of Lepesheva and Waterman (2004). (B) Comparison of fungal CYP61 SRSs with plant CYP710. The conserved motifs from the alignment of fungal CYP61s and plant CYP710s are predicted as their SRSs. (C) The phylogeny of fungal CYP51s and CYP61s, and their counterparts from animal, plant, and bacteria. The phylogenetic tree was inferred by FastTree from the alignments of CYPs constructed by adjusting them to the profile hidden Markov model of PF00067 with HMMER package.

informative for our understanding of the structure/function interaction.

CYP710 family, equivalent of CYP61 as the sterol C22-desaturase, was widely distributed in plants (Morikawa et al. 2006; Nelson 2006b; Kelly and Kelly 2013). Five SRSs of fungal CYP61s are predicted by their comparisons to plant CYP710s (fig. 3B). The sequence comparison reveals striking sequence conservation between plant CYP710 proteins and fungal CYP61 proteins. The two motifs, $NX_5GX_2HX_3RX_6FTX_3ALXY$ (position 86–114, SRS1) and FD/TFLFAA/SQDAS/TT/SS (position 268–280, SRS3) can be considered as the signature of CYP61 or CYP710. For the motif in SRS3, D/T^{269}, S/A^{274}, and T/S^{279} can be used as the phyla-specific residues (D^{269}, A^{274}, and T^{279} in plant CYP710s and T^{269}, S^{274}, and S^{279} in fungal CYP61s). The residue difference at these three positions might be related with substrate preferences of CYP61 and CYP710. The residues D^{276} and A^{277} are specific and absolutely conserved in the CYP710 and CYP61 family proteins. It inferred that these two residues are essential for their common configuration of substrate-binding pockets. The information on the conserved residues will be useful for substrate recognition study of CYP61 and CYP710.

However, it should be noted that SRSs study is difficult for most other CYPs. On the one hand, most CYPs display significant substrate promiscuity. Although they all preserve the basic CYP structural fold, their substrate-binding pockets are well known for high structural plasticity, being able to change shape and volume significantly depending on the chemical structure of the substrates (Hargrove et al. 2012). And even at some extremes, a single amino acid change is sufficient to change the regiospecificity and catalytic efficiency (Schalk and Croteau 2000). On the other hand, CYPs are of high-evolutionary diversity, not only for their considerable variations in sequence, but also for their tremendous functional

diversification (Sezutsu et al. 2013). For example, the tested fungi reflected their individuation of CYPomes. Only CYP51 and CYP61 are widely distributed in the tested fungi. Even for the very close species in *Aspergillus*, they had a large number of species-specific CYPs. Likely, the substrate promiscuity of CYPs was an important driving force for their evolutionary biodiversification aimed to accommodate the increasing number of organic compounds appearing in nature.

Evolutionary Events of Fungal CYPs

Our analyses show that the majority of fungal CYP families have close phylogenetic relationships. It suggests that the diversity of fungal CYPs has mainly arisen from gene duplications. On the one hand, the presence of multiple CYPs that are identical or nearly so in their amino acid sequence is common in individual species, which are considered to be from recent duplications (Sezutsu et al. 2013). On the other hand, in the long term, the duplications could provide redundant genes, which might diverge into new CYP families. The close phylogenetic relationships among CYP families may cue the possible duplication events occurred in the evolutionary history of fungal CYPs (table 2). Generally, it has been recognized that gene duplication could well explain the blooms and great diversity of CYPs (Feyereisen 2011). Meanwhile, it is worth mentioning that a CYP gene from *R. oryzae* (RO3T_09818|RO3G_09819|, CYP5211A1) seems unique in fungal CYPs, not classified into the above-mentioned 15 clades. Based on BLASTP against the NCBI database, it has been shown that this CYP has a high sequence similarity to those from Cyanobacteria species such as *Coleofasciculus chthonoplastes*, *Oscillatoria* sp., *Microcoleus vaginatus*, *Calothrix* sp., and *Nostoc punctiforme*. These Cyanobacteria species often occur in symbiotic associations with fungi to form lichens (Meeks 1998; Redecker 2002). As a possibility, the ancestors of *R. oryzae* and Cyanobacteria-formed symbionts and the gene in *R. oryzae* might have been horizontally transferred from a remote Cyanobacteria species. Likewise, Clade 12 has only two members, one from *Sporisorium reilianum* SRZ2 (emb|CBQ69179.1|) and the other from *Ustilago maydis* 521 (UM04362), with low sequence similarities against currently assigned fungal CYP families. Surprisingly, BLASTP analysis shows their close phylogeny to those from diverse animals such as *Pimephales promelas*, *Ochotona princeps*, and *Jaculus jaculus*. Thus, it can be speculated that the gene was transferred to the ancestor of *S. reilianum* and *U. maydis* from the early animals. More interestingly, Clade 7 (CYP55 clan) shows a high sequence similarity to those from Actinomycetes according to the BLASTP analysis of the consensus. It can be speculated that a gene transfer occurred between the early Actinomycetes and ancestor of Ascomycota and Basidiomycota, but it seemed difficult to infer the transfer direction since both contained numerous CYPs from a wide taxonomy. It is noted that Clade 7 shows unique sequence

features compared with other fungal clades (supplementary fig. S2, Supplementary Material online). For example, the third characteristic motif (located at position 383–388) is not conservative and its signature is very different from others. Besides, the history of Clade 7 is much longer than that of Ascomycota and Basidiomycota. Therefore, this clade is likely to be evolved from the gene transferred from the early Actinomycetes. However, horizontal gene transfer is seldom reported in CYP genes. The gene transfer scenario could be supplemented to the understanding of CYP evolution.

The CYP51 and CYP61 families are highly conserved. Only CYP51 and CYP61 are widely spread in fungi, which are identified in almost all the tested fungi. Moreover, CYP51 and CYP61 show a relatively independent evolution. From the phylogenetic tree, they reflect a relatively distant phylogenetic relationship to other CYPs (fig. 2). More importantly, the phylogenetic trees of CYP51 and CYP61 are consistent with their taxonomic relationships. It suggests that CYP51 and CYP61 are evolutionarily conserved, which could be used to trace the evolution of fungi. Accordingly, the most parsimonious explanation is that CYP51 and CYP61 were present in the last common ancestor of all fungi (Moktali et al. 2012). Indeed, CYP51 is thought to be present even in the ancestral eukaryotes, and it is probable that CYP61 has evolved from a duplication and divergence of the CYP51 gene (Kelly et al. 1997; Nelson 1999b). However, it seems that CYP51 and CYP61 show a remote phylogenetic relationship in the tree (fig. 2). So, even if CYP61 has evolved from CYP51, after a prolonged and separated evolution (longer than fungi history), they show much difference in their sequence characteristics. The conservativeness of CYP51 and CYP61 is probably attributed to their essential roles in fungi-housekeeping functions in sterol biosynthesis (Kelly et al. 2009).

Evolutionary characteristics of CYP51 and CYP61 could provide the important information on evolution of fungi and CYPomes. Accordingly, the recognized divergence times of fungal lineage were applied to calibrate the evolutionary rates of CYP51 and CYP61 (table 3). Surprisingly, CYP51 and CYP61 show very consistent evolutionary rates, which suggest their applicability to be used as fungal molecular clock trees and the stable and consistent evolutionary rates of CYPs. Based on the phylogenetic relationship of CYP51 and CYP61, their divergence time is estimated at around 1.5 Ga. Meaningfully, it could be inferred that the history of fungi is less than 1.5 Gyr since the divergence of CYP51 and CYP61 was prior to the last common ancestor of all fungi (Moktali et al. 2012). It is pretty useful to understand the origin of fungi. Dating estimates on the origin of fungi are very inconsistent, with large time span from 660 Myr to 2.5 Gyr (Taylor and Berbee 2006). Our analysis tends to support the estimate at between 760 Ma and 1.06 Ga (Lucking et al. 2009). It suggests that, before the presence of primitive fungi, CYP61 had endured the separated evolution from the duplication of CYP51 for about 500–740 Myr. Thus, CYP61 was already in more primitive species, later evolved into primitive

Table 3

Evolutionary Rates Calibration of CYP51 and CYP61

Calibration Points (Ma)	Evolutionary Rates (Ma per Unit Distance)	
	CYP51	CYP61
633 (Zygomycota)	791	952
516 (Ascomycota and Basidiomycota)	765	706
414 (Saccharomycotina)	640	788

NOTE.—The distance between CYPs or nodes in the phylogenetic tree was calculated as the average of distances to the divergence node. The length of clade was calculated as the average of distances of two children branches. CYP51 and CYP61 were used as molecular clocks to date fungal divergences and their evolutionary rates were calibrated by the recognized divergence times of fungal lineage (Lucking et al. 2009). It was assumed a globally constant evolutionary rate of CYPs and the rates were estimated to the average rate of CYP51 and CYP61 (774 Ma per unit distance).

fungi. This information could provide the cues on the evolutionary history of primitive fungi as CYP61 homologs are present in other taxonomic groups.

Inspiringly, CYP61 has the same function—sterol C22-desaturase—and a very close phylogeny with plant CYP710 (Morikawa et al. 2006; Morikawa et al. 2009). Some even suggest that these two CYP families should be unified (Kelly and Kelly 2013). CYP710 is thought to be conserved in all plant taxa from unicellular green algae *Chlamydomonas reinhardtii* to higher plants *Populus* (Nelson 2006b). It suggests that a progenitor CYP61 was probably presented in the common ancestor of fungi and plants. Meanwhile, CYP61 is also present in the choanoflagellates, ancestors of fungi and animals (Kodner et al. 2008). However, to date, CYP61 has not been found in animals, even not found in the genome of the sponge *Amphimedon queenslandica*, a model for studying animal evolution (Srivastava et al. 2010; Kelly and Kelly 2013). Likely, CYP61 had been lost in the ancestor of animals due to its nonessential role for animals. Perhaps it can be speculated that Animalia were descended from the ancestor prior to the occurrence of CYP51 duplication.

The dating for the divergence of early eukaryotic groups was estimated based on the phylogeny of CYP51, CYP61, and CYP710 (fig. 3C). The separation of fungi and plant ancestors was estimated at around 1,100 Ma based on their evolutionary distance of CYP61 and CYP710. Later, the fungi and animal ancestors diverged at around 850 Ma estimated from their phylogeny of the CYP51 family. Bacteria CYP51s are thought to be transferred horizontally from plant (Režen et al. 2004), at around 900 Ma. These time points for early eukaryotic divergence are reasonable for current understanding on life evolution (Knoll et al. 2006; Parfrey et al. 2011). The high-evolutionary conservation and phylogenetic relationships among CYP51, CYP61, and CYP710 are useful for understanding evolution of early eukaryotes. A possible scenario is proposed in figure 4.

Fungi possess a wide variety of CYP families, more than 338 CYP gene families in the annotation, but few widespread

CYP families. However, most of fungal CYPs show a close relationship in phylogeny, which reflects a common origin. The fungal CYPs are divided into 15 main clades based on their phylogenetic relationships (fig. 2), which could provide the information on evolutionary events of fungal CYPs such as family expansion. Likely, there were at least nine CYP clans in the primitive fungi: CYP51, CYP52, CYP53, CYP54, CYP56, CYP61, CYP64, CYP505, and CYP534 based on their wide taxonomy (table 2). And then, clans CYP55, CYP550, and CYP613 originated in the ancestor of the Ascomycota and Basidiomycota. The most recent clan CYP59 took a shape in the early Ascomycota. There might be a big duplication event of CYPs in the ancestor of the Ascomycota and Basidiomycota. The redundant CYPs radiating to different CYP families with diverse functions are likely to improve the physiological fitness of Ascomycota and Basidiomycota and promote their prosperity. Another duplication event might have occurred in the early Ascomycota, which may lead to prolific metabolism and booming of filamentous fungi in the Ascomycota. Generally, CYPs show a strong radiation capacity and a progenitor CYP could differentiate into a wide variety of CYP families. For example, the branch of Clade 8, which likely arose from the common ancestor, has evolved into more than 106 CYP families. It might indicate that the motifs of CYPs related with their functions are dynamic in evolution to accommodate diverse functional requirements. CYP gene loss is also a common event in CYP evolution. On the one hand, some globally or locally conserved CYP families are absent in certain species. For example, even if the most conserved CYP51, it was absent in several fungi species. On the other hand, some taxonomic groups might endure obvious CYP gene loss. For example, the yeasts of Saccharomycotina contain few CYP families compared with filamentous fungi in the Ascomycota, dispersed in clans CYP51, CYP52, CYP53, CYP56, CYP61, CYP64, and CYP613. However, based on the phylogenetic analysis, clans CYP54, CYP55, CYP505, CYP526, CYP534, and CYP550 should be in the yeasts of Saccharomycotina. Likely, these clans were lost in the early Saccharomycotina. Moreover, even for the locally conserved families such as CYP52, CYP56, and CYP501 in the Saccharomycotina, their absence in some yeasts might also be attributed to gene loss. Probably, the extensive gene loss in the yeasts might be derived from their limited metabolic demand for CYPs as the yeasts show low abilities in metabolic synthesis compared with filamentous fungi. Accordingly, maintaining numerous CYPs seemed not necessary for the yeasts.

Conclusions

Our investigations of CYPome in 47 fungal genomes from four phyla have led to the fundamental understanding of CYP distribution, structure, function, family expansion, and evolutionary events. The distribution of CYPome differs greatly
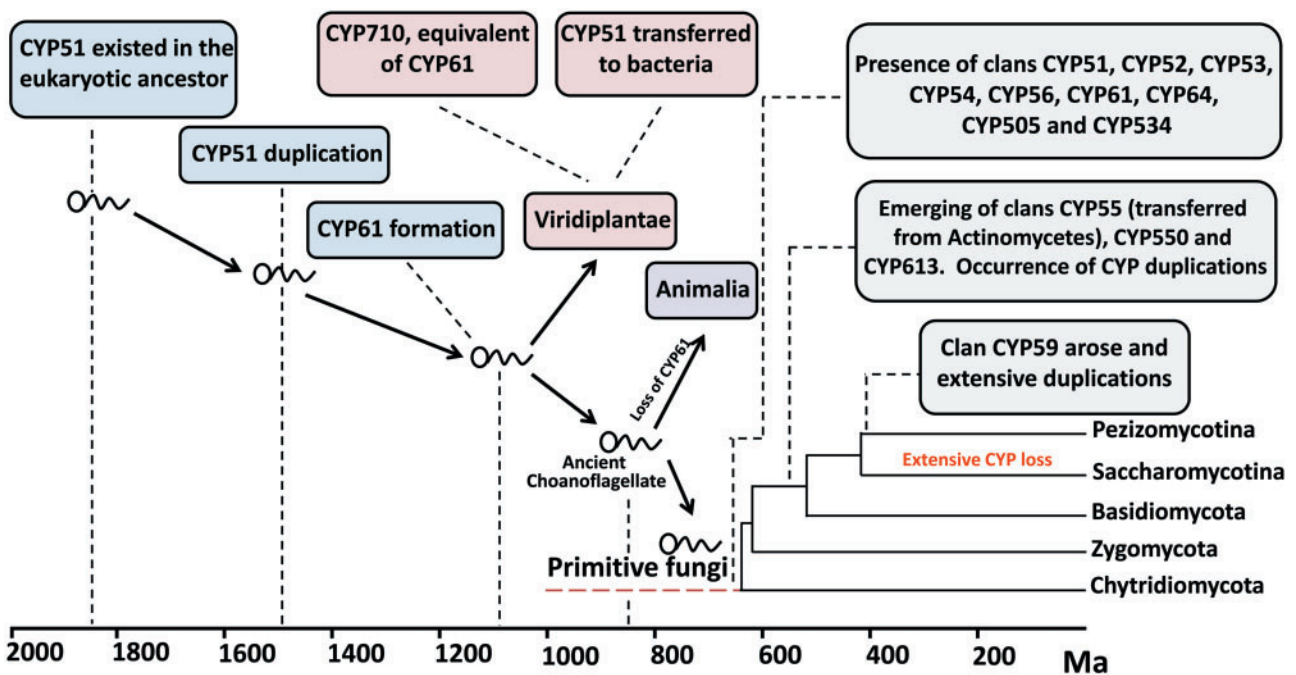
FIG. 4.—A possible evolutionary scenarios of CYPs in fungi. The earliest eukaryotes date from 1,850 Ma in unicellular, flagellated, and aquatic forms (Fedonkin 2003; James et al. 2006; Knoll et al. 2006). CYP51 is thought to be the first eukaryotic CYP (Nelson 1999a). Around 1,500 Ma, CYP51 duplication was occurred in the ancestral eukaryote. The CYP51 duplicate had evolved into the progenitor CYP61 before the separation of Viridiplantae ancestor. CYP710, an equivalent of CYP61, is widespread in Viridiplantae (Nelson 2006b). CYP61 existed in the choanoflagellates, ancestors of fungi, and animals (Kodner et al. 2008), but later likely lost in the early Animalia. The timeline in the fungal tree of life is referred from studies of Lucking et al. (2009). Evolutionary events were inferred based on the distribution and phylogenetic relationship of fungal CYPs.

between taxonomic groups, with CYP number from single to over a hundred. Generally, filamentous fungi such as from the group Eurotiales have high numbers of CYP genes, but yeasts such as from the group Saccharomycotina contain very few CYP genes, and CYP gene expansion is not clearly correlated to genome size. However, the fungi share only two global families, CYP51 and CYP61, as housekeeping functions. The individuation of CYPomes in fungi suggests their highly specialized functions for evolutionary adaptation to ecological niches.

Fungal CYPs showed highly conserved characteristic motifs, but very low overall sequence similarities. The characteristic motifs of fungal CYPs are also highly similar to those of animal, plant, and even archaea and bacteria. The high consistency of characteristic motifs across three domains of life suggests their core roles, probably in maintaining general function of CYP proteins, withstanding long-term evolutionary pressure. However, it should be stressed that the characteristic motifs of fungal CYPs are distinguishable from those of animal, plant, and especially archaea and bacteria. The differences of characteristic motifs between these taxonomic groups could further our understanding on the interaction between CYP structure and function, and CYP evolution. Fungal CYP51s and CYP61s are the good models for

fundamental CYP structure/function studies. The comparison of their SRSs to animal, bacteria, and plant counterparts is useful for their substrate recognition study.

Despite the wide variety and high divergence of fungal CYP families, they can be clustered into 15 clades based on their phylogenetic relationships. The close phylogeny of CYP families suggests that gene duplication was the main force contributing to the large number and variety of CYPs. Moreover, radiation of two possible large duplications in the early Ascomycota and Basidiomycota led to their CYP family expansion and thus may have promoted the later blooming of Ascomycota and Basidiomycota. Meanwhile, some fungal CYPs were arisen from horizontal gene transfer, indicating its important role, far more than hitherto thought, in the development of the diversified CYP superfamily. Conversely, the scarcity of CYPs in yeasts was likely arisen from extensive gene loss coupled with reduced metabolic demands. The phylogeny of CYP51 and CYP61 is highly conserved and consistent with fungal divergences, showing their potential as molecular clocks for tracking fungal evolution. Meanwhile, the phylogenetic relationship between CYP51 and CYP61 could provide some cues on the timeline of early fungi and other early eukaryotic groups. An inferred evolutionary scenario for fungal CYPs along with fungal divergences is generated based on the

phylogenetic and taxonomic relationships among fungal CYP families (fig. 4), which helps understanding the current distribution of CYPomes in fungi and their evolutionary adaptation to ecological niches.

Supplementary Material

Supplementary figures S1 and S2 and table S1 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

Literature Cited

Aoyama Y, et al. 1996. Sterol 14-demethylase P450 (P45014DM*) is one of the most ancient and conserved P450 species. J Biochem. 119:926–933.

Barriuso J, et al. 2011. Double oxidation of the cyclic nonaketide dihydromonacolin L to monacolin J by a single cytochrome P450 monooxygenase, LovA. J Am Chem Soc. 133:8078–8081.

Becher R, Wirsel SG. 2012. Fungal cytochrome P450 sterol 14 alpha-demethylase (CYP51) and azole resistance in plant and human pathogens. Appl Microbiol Biotechnol. 95:825–840.

Bernhardt R. 2006. Cytochromes P450 as versatile biocatalysts. J Biotechnol. 124:128–145.

Chen CK, et al. 2010. Structural characterization of CYP51 from *Trypanosoma cruzi* and *Trypanosoma brucei* bound to the antifungal drugs posaconazole and fluconazole. PLoS Negl Trop Dis. 4:e651.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14:1188–1190.

Deng JX, Carbone I, Dean RA. 2007. The evolutionary history of cytochrome P450 genes in four filamentous Ascomycetes. BMC Evol Biol. 7:30.

Doddapaneni H, Chakraborty R, Yadav JS. 2005. Genome-wide structural and evolutionary analysis of the P450 monooxygenase genes (P450ome) in the white rot fungus *Phanerochaete chrysosporium*: evidence for gene duplications and extensive gene clustering. BMC Genomics 6:92.

Eastwood DC, et al. 2011. The plant cell wall-decomposing machinery underlies the functional diversity of forest fungi. Science 333:762–765.

Fedonkin MA. 2003. The origin of the Metazoa in the light of the Proterozoic fossil record. Paleontol Res. 7:9–41.

Feyereisen R. 2011. Arthropod CYPomes illustrate the tempo and mode in P450 evolution. Biochim Biophys Acta. 1814:19–28.

Gotoh O. 1992. Substrate recognition sites in cytochrome P450 family 2 (CYP2) proteins inferred from comparative analyses of amino acid and coding nucleotide sequences. J Biol Chem. 267:83–90.

Graham SE, Peterson JA. 1999. How similar are P450s and what can their differences teach us? Arch Biochem Biophys. 369:24–29.

Guengerich FP. 2007. Mechanisms of cytochrome P450 substrate oxidation: MiniReview. J Biochem Mol Toxicol. 21:163–168.

Hargrove TY, et al. 2012. Structural complex of sterol 14α-demethylase (CYP51) with 14α-methylenecyclopropyl-Δ7-24, 25-dihydrolanosterol. J Lipid Res. 53:311–320.

Harlow GR, Halpert JR. 1997. Alanine-scanning mutagenesis of a putative substrate recognition site in human cytochrome P450 3A4: role of residues 210 and 211 in flavonoid activation and substrate specificity. J Biol Chem. 272:5396–5402.

Hasemann CA, Kurumbail RG, Boddupalli SS, Peterson JA, Deisenhofer J. 1995. Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure 3:41–62.

Hasler JA, et al. 1994. Site-directed mutagenesis of putative substrate recognition sites in cytochrome P450 2B11: importance of amino acid residues 114, 290, and 363 for substrate specificity. Mol Pharmacol. 46:338–345.

Hoffmeister D, Keller NP. 2007. Natural products of filamentous fungi: enzymes, genes, and their regulation. Nat Prod Rep. 24:393–416.

James TY, et al. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. Nature 443:818–822.

Karlsson M, Elfstrand M, Stenlid J, Olson A. 2008. A fungal cytochrome P450 is expressed during the interaction between the fungal pathogen *Heterobasidion annosum* sensu lato and conifer trees. DNA Seq. 19:115–120.

Kelkar HS, Skloss TW, Haw JF, Keller NP, Adams TH. 1997. *Aspergillus nidulans* stcL encodes a putative cytochrome P450 monooxygenase required for bisfuran desaturation during aflatoxin/sterigmatocystin biosynthesis. J Biol Chem. 272:1589–1594.

Kelly DE, Krasevec N, Mullins J, Nelson DR. 2009. The CYPome (cytochrome P450 complement) of *Aspergillus nidulans*. Fungal Genet Biol. 46:S53–S61.

Kelly SL, Kelly DE. 2013. Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? Philos Trans R Soc B. 368:20120476.

Kelly SL, Lamb DC, Baldwin BC, Corran AJ, Kelly DE. 1997. Characterization of *Saccharomyces cerevisiae* CYP61, sterol Delta(22)-desaturase, and inhibition by azole antifungal agents. J Biol Chem. 272:9986–9988.

Knoll AH, Javaux EJ, Hewitt D, Cohen P. 2006. Eukaryotic organisms in Proterozoic oceans. Philos Trans R Soc B. 361:1023–1038.

Kodner RB, Summons RE, Pearson A, King N, Knoll AH. 2008. Sterols in a unicellular relative of the metazoans. Proc Natl Acad Sci U S A. 105:9897–9902.

Lamb DC, et al. 2002. The cytochrome P450 complement (CYPome) of *Streptomyces coelicolor* A3(2). J Biol Chem. 277:24000–24005.

Leal GA, Gomes LH, Albuquerque PSB, Tavares FCA, Figueira A. 2010. Searching for *Moniliophthora perniciosa* pathogenicity genes. Fungal Biol. 114:842–854.

Lepesheva GI, Virus C, Waterman MR. 2003. Conservation in the CYP51 family. Role of the B′ helix/BC loop and helices F and G in enzymatic function. Biochemistry 42:9091–9101.

Lepesheva GI, Waterman MR. 2004. CYP51—the omnipotent P450. Mol Cell Endocrinol. 215:165–170.

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

Letunic I, Bork P. 2007. Interactive Tree of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics 23:127–128.

Li YC, et al. 2008. Structures of prostacyclin synthase and its complexes with substrate analog and inhibitor reveal a ligand-specific heme conformation change. J Biol Chem. 283:2917–2926.

Lucking R, Huhndorf S, Pfister DH, Plata ER, Lumbsch HT. 2009. Fungi evolved right on track. Mycologia 101:810–822.

Meeks JC. 1998. Symbiosis between nitrogen-fixing cyanobacteria and plants—the establishment of symbiosis causes dramatic morphological and physiological changes in the cyanobacterium. Bioscience 48: 266–276.

Moktali V, et al. 2012. Systematic and searchable classification of cytochrome P450 proteins encoded by fungal and oomycete genomes. BMC Genomics 13:525.

Morikawa T, Saga H, Hashizume H, Ohta D. 2009. CYP710A genes encoding sterol C22-desaturase in Physcomitrella patens as molecular evidence for the evolutionary conservation of a sterol biosynthetic pathway in plants. Planta 229:1311–1322.

Morikawa T, et al. 2006. Cytochrome P450 CYP710A encodes the sterol C-22 desaturase in Arabidopsis and tomato. Plant Cell 18:1008–1022.

Nebert DW, Wikvall K, Miller WL. 2013. Human cytochromes P450 in health and disease. Philos Trans R Soc B. 368:20120431.

Nelson DR. 1999a. Cytochrome P450 and the individuality of species. Arch Biochem Biophys. 369:1–10.

Nelson DR. 1999b. Note on P450 evolution in yeasts and early eukaryotes. [cited 2013 Aug 19]. Available from: http://drnelson.uthsc.edu/yeastP450.evol.html. Memphis (TN): The University of Tennessee Health Science Center.

Nelson DR. 2006a. Cytochrome P450 nomenclature, 2004. In: Phillips IR, Shephard EA, editors. Cytochrome P450 protocols. Totowa (NJ): Springer. p. 1–10.

Nelson DR. 2006b. Plant cytochrome P450s from moss to poplar. Phytochem Rev. 5:193–204.

Nelson DR. 2009. The cytochrome P450 homepage. Hum Genomics. 4: 59–65.

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A. 108:13624–13629.

Park J, et al. 2008. Fungal cytochrome P450 database. BMC Genomics 9: 402.

Podust LM, Stojan J, Poulos TL, Waterman MR. 2001. Substrate recognition sites in 14 α-sterol demethylase from comparative analysis of amino acid sequences and X-ray structure of Mycobacterium tuberculosis CYP51. J Inorg Biochem. 87:227–235.

Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol. 26:1641–1650.

Redecker D. 2002. New views on fungal evolution based on DNA markers and the fossil record. Res Microbiol. 153:125–130.

Režen T, Debeljak N, Kordiš D, Rozman D. 2004. New aspects on lanosterol 14 α-demethylase and cytochrome P450 evolution: lanosterol/cycloartenol diversification and lateral transfer. J Mol Evol. 59:51–58.

Sanglard D, Loper JC. 1989. Characterization of the alkane inducible cytochrome P450 (P450alk) gene from the yeast Candida tropicalis—identification of a new P450 gene family. Gene 76:121–136.

Schalk M, Croteau R. 2000. A single amino acid substitution (F363I) converts the regiochemistry of the spearmint (−)-limonene hydroxylase from a C6-to a C3-hydroxylase. Proc Natl Acad Sci U S A. 97: 11948–11953.

Schneider TD, Stephens RM. 1990. Sequence logos—a new way to display consensus sequences. Nucleic Acids Res. 18:6097–6100.

Sezutsu H, Le Goff G, Feyereisen R. 2013. Origins of P450 diversity. Philos Trans R Soc B. 368:20120428.

Siewers V, et al. 2005. Functional analysis of the cytochrome P450 monooxygenase gene bcbot1 of Botrytis cinerea indicates that botrydial is a strain-specific virulence factor. Mol Plant Microbe Ineract. 18: 602–612.

Soanes DM, et al. 2008. Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS One 3:e2300.

Song WC, Funk CD, Brash AR. 1993. Molecular cloning of an allene oxide synthase—a cytochrome P450 specialized for the metabolism of fatty acid hydroperoxides. Proc Natl Acad Sci U S A. 90:8519–8523.

Srivastava M, et al. 2010. The Amphimedon queenslandica genome and the evolution of animal complexity. Nature 466:720–723.

Syed K, Yadav JS. 2012. P450 monooxygenases (P450ome) of the model white rot fungus Phanerochaete chrysosporium. Crit Rev Microbiol. 38:339–363.

Taylor JW, Berbee ML. 2006. Dating divergences in the fungal tree of life: review and new analyses. Mycologia 98:838–849.

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25:1189–1191.

Werck-Reichhart D, Feyereisen R. 2000. Cytochromes P450: a success story. Genome Biol. 1:REVIEWS3003.

Yoshida Y, Aoyama Y. 1984. Yeast cytochrome P450 catalyzing lanosterol 14-alpha-demethylation. I. Purification and spectral properties. J Biol Chem. 259:1655–1660.

Associate editor: Greg Elgar