Accepted Manuscript

CYPome of the conifer pathogen *Heterobasidion irregulare*: Inventory, phylogeny and transcriptional analysis of the response to biocontrol

Anthony C. Mgbeahuruike, Andriy Kovalchuk, Wimal Ubhayasekera, David R. Nelson, Jagjit.S. Yadav

PII: S1878-6146(16)30173-8

DOI: 10.1016/j.funbio.2016.11.006

Reference: FUNBIO 782

To appear in: *Fungal Biology*

Received Date: 22 March 2016

Revised Date: 25 October 2016

Accepted Date: 26 November 2016

Please cite this article as: Mgbeahuruike, A.C., Kovalchuk, A., Ubhayasekera, W., Nelson, D.R., Yadav, J.S., CYPome of the conifer pathogen *Heterobasidion irregulare*: Inventory, phylogeny and transcriptional analysis of the response to biocontrol, *Fungal Biology* (2017), doi: 10.1016/j.funbio.2016.11.006.

CYPome of the conifer pathogen *Heterobasidion irregulare*: Inventory, phylogeny and transcriptional analysis of the response to biocontrol.

Anthony C. Mgbeahuruike[1,2*], Andriy Kovalchuk[3], Wimal Ubhayasekera[4], David R. Nelson[5], Jagjit. S. Yadav[6].

[1]Department of Microbiology, Faculty of Biological Sciences, University of Nigeria Nsukka, PMB 420001, Enugu State, Nigeria

[2]Department of Veterinary Pathology and Microbiology, Faculty of Veterinary Medicine, University of Nigeria, Nsukka, PMB, 420001, Enugu State, Nigeria.

[3]Department of Forest Sciences, University of Helsinki, Finland.

[4]Structure and Molecular Biology Program, Department of Cell and Molecular Biology, Uppsala University, Box 596, Biomedical Center, SE-751 24, Uppsala, Sweden.

[5]Department of Microbiology, Immunology and Biochemistry, University of Tennessee, Memphis, TN 38163, USA.

[6]Environmental Genetics and Molecular Toxicology Division, Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, Ohio 45267-0056, USA .

*Corresponding author: **Anthony.mgbeahuruike@unn.edu.ng**

Abstract

Heterobasidion annosum s.l is the causative agent of root and butt rot disease of conifer trees. The mechanisms by which the fungus colonizes living conifers are not fully understood. Also, the molecular mechanisms underlying the interaction between the pathogen and the biocontrol fungus, *Phlebiopsis gigantea* have not been fully elucidated. Members of the cytochrome P450 (CYP) protein family may contribute to the detoxification of components of chemical defense of conifer trees by *H. annosum* during infection. At the same time, they may be involved in the interaction between *H. annosum* and *P. gigantea.* A genome-wide analysis of CYPs in *H. irregulare,* a member of *H. annosum* species complex for which the complete genome sequence is available, was carried out alongside gene expression studies.

According to the Standardized CYP Nomenclature criteria, the *H. irregulare* genome has 121 P450 genes and 17 CYP pseudogenes classified into 11 clans, 35 families and 64 subfamilies. Our analysis of CYP distribution showed the presence of tandem gene arrays. The identified arrays consist of closely related genes belonging to the same family and subfamily, an indication that these arrays originated from gene duplications. Some of the arrays included putative pseudogenes with in frame stop codons, frame shifts or deletions. The phylogenetic analysis showed that all the families of *H. irregulare* CYPs were monophyletic groups except for the family CYP5144. Microarray analysis revealed the transcriptional pattern for 130 transcripts of CYP-encoding genes during growth on culture filtrate produced by *P. gigantea*.

The high level of P450 gene diversity identified in this study could result from extensive gene duplications presumably caused by the high metabolic demands of *H. irregulare* in its ecological niches.

Key words: CYPome, genome, P450, plant pathogen, *Heterobasidion annosum* s. l., phylogeny, homology modelling.

1 Introduction

The fungal pathogen *Heterobasidion annosum* (Fr.) Bref. *sensu lato* (s.l.) is the major causative fungal agent of root and butt rot disease of conifers (Stenlid 1987; Asiegbu et al., 2005). *H. annosum* s.l. causes a huge economic loss in the timber industry in the Northern hemisphere (Asiegbu et al., 2005; Woodward et al., 1998). The species complex includes five closely related species, Eurasian *H. annosum*, *H. parviporum* and *H. abietinum*, and North American *H. irregulare* and *H. occidentale*. *H. irregulare* has recently been recognized as an independent species belonging to the *H. annosum* complex (Otrosina and Garbelotto, 2010). It is the only species within *H. annosum* species complex, for which the complete genome sequence is available (Olson et al., 2012). *H. irregulare* was introduced into Lazio region in Italy, most probably during World War II by US troops via infected wood material (Garbelotto et al., 2013). Research studies have demonstrated that *H. irregulare* has spread from its initial introduction site (Castelporziano) and is currently causing extensive mortality in several *Pinus pinea* stands (Garbelotto et al., 2013; Otrosina and Garbelotto, 2010).

The biology and ecology of *Heterobasidion* have been extensively studied, but the molecular mechanisms by which the fungus colonizes living conifers have not been fully understood. Insights from *H. irregulare* genome sequence data shows that the fungus employs a wide arsenal of genes that are involved in different biological processes such as toxin production, protection against plant defenses, processing of low oxygen pressure and other abiotic stresses during pathogenic colonization of living wood (Piri et al., 1990; Olson et al., 2012). One class of genes

that may play crucial roles in detoxification of plant toxins during *H. annosum* s.l. colonization of living wood is cytochrome P450. Cytochrome P450 monooxygenases (CYPs) are a large superfamily of monooxygenases found in diverse living organisms (Ortiz de Montellano, 2005). P450s are heme-thiolate proteins, which perform a wide variety of reactions such as hydroxylation, epoxidation, dealkylation, sulfoxydation, deamination, desulphuration, dehalogenation, and nitro reduction (Ortiz de Montellano, 2005). Diversification of CYPs in different organisms have been suggested to be as a result of extensive gene duplication events and other evolutionary processes such as gene amplification, conversion, genome duplication, gene loss and lateral transfer (DiGuistini et al., 2011).

Among the main fungal phyla, the basidiomycota and zygomycota species show considerable numbers of CYP genes while the chytridiomycota and the ascomycota species possess a limited number of CYPs (Chen et al., 2014; Martinez et al., 2004; Martinez et al., 2009; Eastwood et al., 2009; Fernandez-Fueyo et al., 2012; Floudas et al., 2012). CYPs contribute to significant secondary metabolic processes such as the biosynthesis of fungal toxins, virulence factors and detoxification of xenobiotics (Denison and Whitlock, 1995). Understanding the role of CYPs in a basidiomycete plant pathogen in contrast to saprophytes offers a unique opportunity to detect CYPs specifically evolved as the result of an adaptation to the pathogenic life style. The function, regulation and expression of CYPs have been well characterized in some model organisms because of their important roles in plant, animal and bacterial metabolism and physiology (Gonzalez and Lee, 1996; Black and Coon, 1986). However, not as much is known about CYPs in non-model fungal species including *H.annosum* s.l., despite the fact that these proteins have some properties that may be crucial in bioconversion of lignin and a variety of aromatic compounds, including environmental pollutants (Bezalel et al., 1996; Bezalel et al.,

1996b; Masaphy et al., 1996). Recent genome-wide comparison of CYP gene numbers has shown that *Postia placenta* has a higher number of CYP genes than *Phanerochaete chrysosporiuim* (Martinez et al., 2009; Hirosue et al., 2011). In addition, several studies involving phylogenetic analysis of fungal CYPs have been reported (Syed et al., 2014). A recent study involving the analysis of the CYPs in *P. chrysosporium* identified 12 CYP families which were classified into 11 clans based on a phylogeny (Yadav et al., 2006). Other studies have investigated the divergence of CYP proteins in different fungi using different bioinformatics approaches (Chen et al., 2014; Deng et al., 2007; Park et al., 2008). For example, in *P. chrysosporium*, 154 CYPs have been identified, out of which 144 proteins have been confirmed using gene expression methods (Yadav et al., 2006). In recent times, numerous genome projects have accelerated the sequence compilation of CYPs, and as a result of that, the sequence database of P450s has greatly enlarged (Park et al., 2008; Nelson, 2009). Classification of the CYP proteins is based primarily on amino acid sequence similarity and phylogenetic relationships. Furthermore, classifications based on clans, which represent higher order grouping of CYP families have been proposed in the CYP community. "CYP clans are the deepest branching clades on a CYP dendrogram" (Nelson, 2011). Genes within a clan are most likely to diverge from a common ancestor gene and may have the same functions (Nelson, 1998). Although clan structure has been suggested in fungi (Yadav et al., 2006), the parameters for clan membership have not been clearly defined. Although a reasonable number of CYP genes has been predicted in *H. annosum* s.l., the biological roles and the evolutionary mechanisms driving the diversification are not fully known. In addition, no study has reported a detailed transcriptional response of the *H. annosum* s.l. CYPs during interaction of the pathogen with the biocontrol fungus, *Phlebiopsis gigantea* (Fir.) Jülich. *P. gigantea* is a saprotrophic white rot

fungus that competes for nutrients and shares the same ecological niche with *H. annosum* s.l (Mgbeahuruike et al., 2012). Despite the fact that the interaction between the conifer pathogen and the biocontrol agent has been well studied (Adomas et al., 2006; Mgbeahuruike et al., 2011; Sun et al., 2009a; Sun et al., 2009b), the genetics and the molecular mechanisms underlying the biological control process are still at its infancy.

In the present study, we used both bioinformatics and phylogenetic approaches to identify and classify all *H. irregulare* CYP genes into families, subfamilies and clans. We also used microarrays to study the expression pattern of the identified CYPs when the fungal pathogen is grown in the presence of-culture filtrates from *P. gigantea*. Understanding the regulatory pattern of CYPs in *H. annosum* when grown in culture filtrate from *P. gigantea* will give further insight into the molecular mechanisms of biocontrol of the conifer pathogen by *P. gigantea*. It is possible that the CYPs could be relevant target proteins for application of the biocontrol agent during the conifer disease control process. Furthermore, a comprehensive survey of *H. irregulare* CYPs at the transcriptional level provides important information on the metabolic diversity of the CYPs in this economically important basidiomycete.

2.0 Materials and Methods

2.1 Gene mining and annotation

Amino acid sequences of the putative CYP proteins encoded in the genome of *H. irregulare* were obtained in FASTA format from JGI after text search for “CYP”. The 187 hits were examined and non-CYP sequences were removed. The remaining sequences were BLAST searched against a private set of about 5000 named CYPs from fungi. Sequences that were short

or misaligned were examined at the JGI using the "view nucleotide and three frame translation tool". Missing exons or incorrect intron boundaries were manually corrected as needed. Names were officially assigned to these sequences based on their sequence relatedness to known fungal CYPs by the Committee on Standardized CYP Nomenclature. Sequences that did not fit in existing families or subfamilies were assigned to new families or subfamilies according to nomenclature rules for CYP sequences (> 40% amino acid identity for assigning a family and > 55% for a subfamily). The families were then grouped into clans, based on consistent clustering in phylogenetic trees.

2.2 Sequence alignment and phylogenetic analysis.

Phylogenetic analysis was performed with the program package MEGA5.2 (Tamura et al., 2011). Multiple sequence alignments of *H. irregulare* CYP sequences were constructed with the MUSCLE algorithm integrated in MEGA5.2 package using default settings. Predicted pseudogenes were excluded from the alignments. Alignments were quality trimmed with Gblocks 0.91 b (Talavera and Castresana, 2007) to eliminate poorly aligned positions and divergent regions of alignments in order to make them more suitable for phylogenetic analysis. Both full-length and trimmed alignments were used to produce phylogenetic trees. To have a clearer picture of the actual topology of the phylogenetic relationships among the CYP genes and for the purpose of comparison, the phylogenetic trees were reconstructed using different algorithms. Maximum-likelihood trees were obtained using Jones-Taylor-Thornton model with 100 bootstrap replications. Neighbor-joining trees were constructed using Poisson substitution model with 500 bootstrap replications. Minimum evolution trees were constructed using Poisson substitution model with 500 bootstrap replications. Maximum parsimony trees were constructed

using Subtree-Pruning-Regrafting search method with 500 bootstrap replications. Although the topologies of the phylogenetic trees produced with the different algorithms showed some minor differences, the same major groups of CYP proteins were recognized in all reconstructions.

2.3 Homology modeling

To understand the protein structure and the functional significance of the different *H. irregulare* CYP families, homology modeling was carried out. Similar structures to *H. irregulare* cytochrome P450 (CYP63A22, CYP5150S3, CYP5037B16, CYP5144BJ1, CYP5144M12 and CYP5344D1) deposited in the Protein Data Bank (PDB) (Altschul et al., 1997; Berman et al., 2000) were located using BLAST searches (Altschul, et al., 1997), obtained, then superimposed and compared with the programs LSQMAN (Kleywegt et al., 2001) and O (Jones et al., 1991). The sequences of these structures and *H. irregulare* cytochrome P450 were aligned using CLUSTAL W (Thompson et al., 1994). Based on the sequence identity, the homology models were built using the structures of human cytochrome P450 46A1 (CYP46A2) (PDB entry 2Q9F.pdb (Mast et al., 2008)) human microsomal P450 3A4 (CYP3A4) (PDB entry 1TQN.pdb (Yano, et al., 2004)) and human microsomal P450 1A2 (CYP1A2) (PDB entry 2HI4.pdb (Sansen et al., 2007)). PDB entries 2Q9F and 1TQN were used to model CYP63A22 (sequence identity 32%) and CYP5150S3 (sequence identity 35%) as templates respectively. The homology models of CYP5037B16 (sequence identity 32%), CYP5144BJ1 (sequence identity 30%), CYP5144M12 (sequence identity 29%) and CYP5344D1 (sequence identity 31%) were built on the PDB entry 2HI4. The program SOD (Kleywegt et al., 2001) was used for homology modelling. Then, the models were adjusted in O, using rotamers that would improve packing in the interior of the protein, and accounting for insertions and deletions in loop regions. The models are available

upon request from the authors. The figures were prepared using O, MOLSCRIPT (Kraulis 1991) and Molray (Harris and Jones 2001).

2.4. Transcriptional profiling of *H. irregulare* CYP genes

2.4.1 Growth of *H. irregulare* in culture filtrate of *P. gigantea*

Isolates of *P. gigantea* (Rotstop F) and *H. annosum* s. str. (strain FP5) used in the present study were kindly provided by Kari Korhonen (Finnish Forest Research Institute, [METLA], Vantaa, Finland). Agar plugs of about 3mm in diameter were cut from *P. gigantea*-overgrown agar plate. They were used to inoculate three 300 ml Erlenmayer flasks containing 100 ml liquid malt extract (ME: 2g/l, Sigma Aldrich, US) each, and the cultures were incubated for 10 days at 20°C in a static condition. Mycelia were separated from each culture using a 0.2 μm sterile filter paper (Munktell Filter AB, Sweden). Agar plugs (3 mm in diameter) containing freshly grown *H. annosum* s.s. were inoculated into the filtrates from each flask after overnight heating at 70 °C to kill any remaining particles of mycelia from the *P. gigantea*. The flasks were incubated for another 10 days at 20°C. Three positive controls containing 3mm of agar plugs of *H. annossum* s.s in fresh liquid ME medium were also set up. The cultures were incubated at 20°C under static condition and harvested at 10 days post inoculation (d.p.i). The harvested mycelia were frozen in liquid nitrogen, homogenized with mortar and pestle and stored for further processing.

2.4.2. RNA Processing and cDNA synthesis

RNA extraction was done following the method described by Chang et al., (1993). RNA integrity and concentrations were assessed with RNA 6000 Nano kit using an Agilent Bioanalyzer (Agilent, CA) and the measurement was repeated using NanoDrop ND-1000

Spectrophotometer. The purity of the samples was estimated by the OD ratios (A260/A280, ranging within 1.8–2.2). RNA samples (2μg) were DNase treated using DNaseI according to the recommendations from the manufacturer (Fermentas, Canada). The DNase-treated samples were purified with RNeasy® MinElute Cleanup kit (QIAGEN) according to the protocol. The cDNA was synthesized from 100ng of total RNA samples, using the TransPlex® Complete Whole Transcriptome Amplification Kit according to the manufacturer's protocol (SIGMA, USA). The generated cDNA was purified with GenElute PCR Clean-Up kit (Sigma-Aldrich, Finland) and run in 1.5% agarose gel to-assess the integrity and size of the fragments.

2.4.3. Array design and Oligonucleotide synthesis

The oligonucleotides for the microarray analysis were based on the *H. irregulare* genome sequence (Olson et al., 2012). The genome sequence which has about 12,299 genes was used to construct a high-density *H. irregulare* microarray based on the Nimblegen (Nimblegen Systems, Inc., Madison, WI, USA) 4x72K design format (Mgbeahuruike et al., 2013). About 12,199 probe sets representing all annotated ORFs and genetic elements were designed from the 12,299 gene sequences identified. Five non-identical replicates of 60-mers probe per gene model coding sequence were used to represent each of the annotated gene sequences. 19 sequences had no probes, and 81 sequences with probes identical to several gene models were identified. Furthermore, about 916 random 60mers control probes and labeling controls were added in the experiment. Also, 2032 probes were included in the array as technical duplicates for internal check. However, for the purpose of this study, we restricted our analysis to the gene sequences encoding CYP proteins.

2.4.4. Microarray hybridization and data analysis

For the microarray hybridization, 5μg of the cDNA was sent to Nimblegen (Roche Nimblegen Systems, Iceland) for expression analysis. The cDNA samples were hybridized on *H. irregulare* customized arrays according to the Nimblegen standard protocols. Washing, scanning, data acquisition, background correction and normalization of the generated data were done by Nimblegen (Roche Nimblegen Systems, Iceland) following the standard procedure. Nonspecific oligos were filtered from the microarray data (an oligo was considered non-specific if it shares more than 90% homology with a gene model different from the one it was made for) and normalization was done with ARRAYSTAR software (DNASTAR, Inc. Madison, WI, USA) using nonparametric variable selection and approximation (NVSA). The NVSA identifies genes exhibiting no differential expression, and uses them as the basis for normalization. Filtering of the non-specific oligos produced 11,578 gene models in the expression data, out of which 135 were CYP gene transcripts. The mean expression value was calculated from 250 random oligos present on the array. To minimize gene expression measurement errors, statistical significance was assessed with Student t-test using *P*-values adjusted for multiple test correction using the Benjamini–Hochberg false discovery rate (FDR) method (Benjamini and Hochberg, 1995). Fold changes were calculated as a ratio of the expression values of the experimental sample (*H. annosum* s.s. grown in liquid culture filtrate from *P. gigantea*) divided by the control (*H. annosum* s.s. grown on liquid medium). A stringency of ≤ -2.0 to $\geq +2.0$ for fold changes of the down- and up-regulated genes respectively, was applied to the 135 CYP dataset. A gene was considered significantly up-or down-regulated if the P value is 0.05 or below, and the fold change is $\geq +2.0$ or ≤ -2.0.

3.0 Results

3.1 Annotation, classification and comparative analysis of *H. irregulare* CYPome

H. irregulare genome features 121 CYP genes and 17 predicted CYP pseudogenes (Tables 1 and 2, Supplementary Table 1). Using the CYP nomenclature criteria, the *H. irregulare* CYPs were classified into 11 clans, 36 families and 64 subfamilies (Table 2). Among the clans, CYP64 includes the largest number of CYP members (80 authentic CYPs and 6 pseudogenes) followed by clans CYP56 (23 authentic CYPs and 8 pseudogenes), CYP53 (6 authentic CYPs and 1 pseudogene) (Table 1, Supplementary Table 1). Among the 35 families, CYP5144 includes the greatest number of CYPs (45 CYPs classified in 11 subfamilies and 5 pseudogenes) followed by CYP5344 (10 CYPs classified in 2 subfamilies) (Table 1, Supplementary Table 1). Family CYP5144 had the highest number of subfamilies (11 subfamilies) followed by CYP5037 (6 subfamilies). *H. irregulare* CYPome is somewhat smaller compared with the number of CYPs in *P. chrysosporium* (149 CYPs classified into 32 families and 70 subfamilies) (Table 2).

3. 2 Phylogenetic analysis

We have analyzed the phylogenetic relationships between *H. irregulare* CYP proteins using a number of different phylogenetic algorithms. Despite some small differences, the same major groups were recognized in all the trees generated (Figure 1, Supplementary Figures 1-3). Therefore, we will only discuss the trees generated with the Maximum Likelihood algorithm (Figure 1). Most of the recognized clans and families of *H. irregulare* CYP proteins were identified as monophyletic groups, which received high bootstrap support (>60%), with two exceptions. First, the bootstrap support for the family CYP5037 was below 50%. Second, the members of the two small families, CYP5348 and CYP5358 appeared nested within the largest

family, CYP5144, but the corresponding clade received very low bootstrap support. None of the applied algorithms could satisfactorily resolve the relationships within these three subfamilies. However, all the subfamilies recognized within the family CYP5144 received bootstrap support of ≥90% in this analysis. It should be noted that the CYP families CYP5144, CYP5348 and CYP5358 belong to the same clan and are related. The difficulties in the phylogenetic reconstruction of the family CYP5144 might be due to the presence of a high number of closely related paralogous sequences (i.e., subfamily CYP5144M was represented by 14 genes). Large CYP families have this problem of expanding and absorbing closely related families.

3.3 Location of the CYP genes and potential pseudogenes in *H. irregulare* genome

Substantial numbers of the CYP-encoding genes in the *Heterobasidion* genome are organized in arrays of tandem repeats (Figure 2). This is particularly the case for the genes belonging to the family CYP5144. The largest arrays are located on scaffold 5 (four adjacent genes in a 14 kbp region) and on scaffold 10 (two arrays made of three adjacent genes each in a 29 kbp region). Some of the identified arrays included putative pseudogenes. The identified CYP arrays consist of closely related genes showing a high degree of sequence similarity and belonging to the same family and subfamily. Moreover, the genes within the CYP arrays have the same orientation and are located on the same DNA strand, and the arrays are not interrupted by other predicted genes. Taken together, these facts support the idea that the CYP arrays in the *H. irregulare* genome originate from relatively recent gene duplications. The underlying mechanisms of the array formation are not entirely clear, but may involve unequal crossover events.

3.4 Protein Modeling

3.4 Protein Modeling

Only 2 residues (Ala388 and Val170) in CYP63A22 and 4 residues (Ala395, Ala331, Phe330 and Thr335) in CYP5150S3 are conserved among the catalytically important residues described for human cytochrome CYP46A1 (Mast et al., 2008) and human microsomal CYP3A4 (Yano et al., 2004) respectively (Figure 3). The catalytic acid (Asp320) described for human microsomal P450 1A2 (CYP1A2) (Sansen et al., 2007) is the only conserved residue in CYP5037B16, CYP5144BJ1, and CYP5144M12 among the catalytically important residues where as in CYP5344D1 it is replaced with a Glutamate (Glu265) (Figure 4). Having a conserved acid residue in this exact position in the structure confirms that all modeled CYPs possess the ability to convert cholesterol to 24S-hydroxycholesterol. Several of the heme binding residues are conserved in all the homology models supporting the fact that heme binding is essential to these enzymes (Yano et al., 2004; Sansen et al., 2007; Mast et al., 2008). Therefore, these *H. irregulare* CYPs probably show different dynamics in the function compared to the template structures.

3.5 Transcriptional response of *H. irregulare* CYPome to culture filtrate from *P. gigantea*

Filtering out of the non-specific oligos produced 11,578 gene models in the expression data, out of which 130 were CYP genes. Out of the 130 CYP genes, 13 transcripts were pseudogenes and two transcripts could not show any detectable signal (Supplementary Table 2). Furthermore, 59 transcripts were differentially expressed, 16 were up regulated (Figure 5) and 43 were down regulated (Figure 6). Statistical analysis showed that five CYP genes were significantly differentially expressed with expression fold change $\geq +2.0$ or ≤ -2.0. Three CYP genes

322 (CYP5144BK8, CYP5139N1 and CYP5148B12) were down-regulated whereas two genes
323 (CYP502B9 and CYP5150S4) were significantly up-regulated (Table 3).
324
325 **4.0 Discussion**
326 Root and butt rot disease of the conifer trees caused by members of *H. annosum* species complex
327 is the most economically important disease of conifer trees in the Northern Hemisphere (Asiegbu
328 et al., 2005). The mechanisms by which the fungus colonizes living conifers has not been fully
329 understood. In addition, the molecular mechanisms underlying the interaction between the
330 conifer pathogen and the biocontrol fungus, *P. gigantea* are not well known. In the present study,
331 a comprehensive annotation, characterization and investigation of the response of the *H.*
332 *annosum* CYPs to culture filtrates from the biological control agent was conducted. The *H.*
333 *annosum* s.l. genome features 121 CYP genes and 17 pseudogenes which are grouped into
334 different families, sub-families and clans. The presence of a large repertoire of CYP
335 monooxygenases in *H. annosum* s.l. genome suggests a potential role of these CYPs in various
336 endogenous and xenobiotic metabolic processes. However, except for the conserved fungal
337 CYPs such as CYP51 and CYP61 (ergosterol biosynthesis), CYP52 (alkane/fatty acid
338 hydroxylation), and CYP505 (fusion P450s), the majority of the CYPs identified in the *H.*
339 *annosum* s.l. genome are orphan with no known function. Functional analysis of the available
340 basidiomycete CYPs has suggested that the highly conserved CYPs that are common across
341 different fungal phyla viz. CYP51 and CYP61 play roles in basic processes (ergosterol synthesis)
342 (Črešnar and Petrič, 2011). Other CYPs which are fairly conserved across the fungal phyla viz.
343 CYP52 and CYP505 participate in cellular metabolism of aliphatic compounds-fatty acids and
344 alkanes (Yadav et al., 2006; Syed et al., 2010; Syed et al., 2011a; Syed et al., 2011b; Syed and

Yadav, 2012). Considering that *H. annosum* is capable of colonizing fresh wood, several of the CYPs may be expected to be involved in the colonization process via degradation or assimilation of plant defense chemicals and wood extractives. A considerable variation exists in the number of CYP encoding genes across different lignolytic basidiomycetes. For instance, the *H. annosum* s.l. CYPs features a CYP count somewhat lower than that reported for the model white-rot basidiomycete *P. chrysosporium* (149 CYPs) (Martinez et al., 2004; Syed & Yadav, 2012) and much lower as compared to the model brown rot basidiomycete *P. placenta* (250 CYPs) (Martinez et al., 2009). Furthermore, in *Ganoderma lucidum*, a much more expanded CYP repertoire has been reported, 219 CYP protein sequences classified into 42 families (Chen et al., 2012). In comparison to *P. chrysosporium*, which colonizes dead wood–*H. irregulare*, a fresh wood-degrading species showed an 18% reduction in CYP count (27 CYPs). In contrast, our analysis showed that *H. annosum* s.l. features a CYPome that is greater in number than those found in non-wood-degrading basidiomycetes like *Cryptococcus neoformans* and *Tremella mesenterica*, which contain 8 and 10 CYPs respectively (Syed et al., 2014). In the present study, some CYP families like CYP5144 (49 genes), CYP5150 (10 genes) and CYP5344 (10 genes) were found to have expanded. CYP5144 is a basidiomycete-specific CYP family and it has shown considerable expansion in most basidiomycetes. For example, in the mushroom fungi, *Coprinus cinereus* (61 CYP genes) and *Pleurotus osteratus* (60 CYP genes), the model white rot fungus *P. chrysosporium* (56 CYP genes), and 55 CYP genes in the brown rot fungus, *Serpula lacrymans* (Ide et al., 2012). It is important to note that the enrichment of this CYP family in most basidiomycetes could suggest a key role in the physiology of the associated fungi especially during metabolism. However, in ascomycetes, lower numbers of CYP5144 count have been reported, *Aspergillus flavus* (8), *A. oryzae* (8) and *A. niger* (5). Furthermore, CYP512 was found

368 to have only a few members in this study (3 CYPs), although CYP512 has been reported to have
369 undergone expansion in some polyporales like *G. lucidum* (23 CYPs), *P. chrysosporium* (14) and
370 *P. placenta* (14) (Otrosina et al., 1993). Using the genome sequence of *H. irregulare,* a
371 preliminary automated P450 analysis based primarily on an automated BLAST analysis against
372 David R Nelson's online P450 database indicated the presence of 140 CYPs classified into 29
373 CYP families and an unassigned group (Park et al., 2008). Considering that eight of the CYPs
374 were listed under the unclassified group by Park et al., 2008 and their preliminary CYP family
375 assignment does not separate the pseudogenes from the authentic P450 genes, an accurate
376 analysis and direct comparison with Nelson's P450 database was not possible.
377 The most prominent feature of the phylogenetic grouping of the CYPs from *H. irregulare* is the
378 separation of the sequences into monophyletic groups. The only family that was not recovered as
379 monophyletic in our analysis is CYP5144, which is by far the largest in *H. irregulare* genome.
380 Two small families CYP5348 and CYP5358, each with a single representative, nested within the
381 family CYP5144. However, the bootstrap support for the branch encompassing CYP5144, CYP
382 5348 and CYP5358 was far below 50% in all our phylogenetic reconstructions, and no definitive
383 conclusions about relationships of these three families can be made based on the obtained results.
384 We also observed that CYP genes in *H. irregulare* (in particular, members of the family
385 CYP5144) are often adjacent on the chromosomes, forming small arrays of tandem repeats.
386 Although the biological relevance of this clustering is not properly understood, the arrangement
387 could be a result of unequal crossing over, resulting in tandem duplications. A tandem gene
388 duplication in turn increases the probability of additional unequal crossing over events (due to
389 the repeated sequence), thereby resulting in paralog gene clusters. Emergence of new copies of
390 CYP genes via gene duplication has been reported for other basidiomycete species (Doddapaneni

et al., 2004). Functional diversification of the duplicated genes observed within the identified tandem arrays could be one of the driving forces for *H. irregulare* CYPome diversification. Several pseudogenes were found in the *H. irregulare* CYPome. The pseudogenisation is predicted to be a frequent outcome of gene duplication, as both copies originating from the duplication event are identical and therefore functionally redundant. However, CYP51, CYP53 and CYP63 appeared to be single proteins both at the family and clan levels. A similar result has been reported in CYP53 protein of *P. chrysosporium* (Doddapaneni et al., 2004). The presence of a single CYP53 in *P. chrysosporium* and *H. irregulare* contrasts with the multiple CYP53 proteins observed in ascomycetes. A total of 170 CYP proteins (including CYP53 and CYP58) that are grouped into 4 sub-classes (A, B, C and D) have been assigned to the CYP53 clan in different ascomycetous fungi http://drnelson.utmem.edu/53clan2.pdf. The clustering patterns and the phylogenetic groupings of the different CYP proteins could partly be explained by the conserved nature of some important residues in the CYP family. The conserved catalytic acidic amino acid residue and some very important residues for heme binding shows that the modeled enzymes possess the catalytic activity confirming that these CYPs play roles during the life cycle of the fungus. The changes observed in the models especially the non-conserved amino acids in the catalytic site as well as in the heme binding site may hint us about possible diverse binding properties and substrate specificities (Yano et al., 2004; Sansen et al., 2007; Mast et al., 2008). This idea is further proved by the suggestion /observation of structural plasticity or conformational changes during the catalytic activity (Hargrove et al., 2012; Chen et al., 2014).

From the gene expression analysis, 130 CYP transcripts were represented on the *Heterobasidion* custom microarray (Supplementary table 2), each showing a different regulatory pattern during growth on *P. gigantea* culture filtrate. Thirteen of the transcripts belonged to the

predicted pseudogenes. The identified pseudogenes in the expression data were probably spliced variants of the neighboring CYP genes. In our study, only five CYP genes showed statistically significant changes in their expression level. Members of three CYP families, CYP5144, CYP5139 and CYP5148, were significantly down-regulated, whereas members of the families CYP502 and CYP5150 were significantly up-regulated. These results indicate, on one hand, that culture filtrate of *P. gigantea* has a limited effect on the expression of *H. annosum* CYP genes. On the other hand, expression level of the genes CYP5148B12 and CYP5150S4 changed over 80- and 130-fold, respectively, suggesting a strong specific response of these two genes to the compounds produced by *P. gigantea*.

The scarcity of the data on the biological role of CYP genes in basidiomycetes complicates the interpretation of the obtained results. None of *Heterobasidion* CYP genes has been hitherto characterized experimentally. Nevertheless, culture filtrate from *P. gigantea* had been previously shown to repress the genes involved in nutrient processing and acquisition, signal transduction and transport during competition with *H. annosum* on artificial media (Mgbeahuruike et al., 2012). Karlsson et al., 2008 reported the up-regulation of CPM2 protein, a member of CYP64 family during the growth of *H. parviporum* on living bark of spruce. They concluded that the CPM2 protein could play an important biological role during *H. parviporium* infection. In addition, some CYP genes located within secondary metabolism gene clusters were found to be differentially expressed during intersterility (IS) incompatibility studies in *H. annosum* s.l. (Van der Nest et al., 2014). Further analysis of the gene clusters using phylogeny separated them into 6 major clans belonging to 3 CYP families, CYP53, CYP534 and CYP64 (Van der Nest et al., 2014). CYP63 was also reported in other studies to be involved in the degradation of xenobiotic compounds during IS (Syed and Yadav, 2012). Other studies also

found that CYPs were differentially expressed during heterospecific interactions between *P. gigantea* and *H. parviporum* (Mgbeahuruike et al., 2012; Adomas et al., 2006; Hansson et al., 2012a, b).

Interactions between organisms using artificial media to elucidate the basis for biological control and the dynamics of fungal competition have analyzed in other studies (Carruthers & Rayner, 1979; Magan & Lacey, 1984a, b). These studies have usually used paired cultures of fungi in Petri dishes of agar medium. Although Dowding (1978) has expressed doubts about extrapolating results obtained from such surface culture to natural situations. Magan & Lacey (1984a, b) on the other hand have suggested that such methods were the best available for analyzing the interaction between fungi. Other studies have shown strong relationship between the combative ability of fungi in dual cultures and their ecological roles (Rayner, 1978; Carruthers & Rayner, 1979; Boddy & Rayner, 1983). Secondary metabolites synthesis confer a competitive advantage to the producer of the compounds, for example, CYP64 members in *Aspergillus* species produce aflatoxins (Bhatnagar et al., 2003). The highly expressed *H. irregulare* CYPs in our study may also be involved in detoxification of secondary metabolites and toxins from the *P. gigantea* during combative interaction on the stump.

5.0 Conclusions

The result from this work has provided a comprehensive survey of *H. irregulare* CYPs at the genomic and transcriptional level and has also generated important information on the role of *H. irregulare* CYPs in the antagonistic interaction between the pathogen and the biocontrol fungus. The diverse number of CYPs observed in this study could be due to extensive gene duplication resulting from the high metabolic demands of this fungus in its ecological niche.

460 **Availability of Supporting Data**

461 Study Accession URL: http://purl.org/phylo/treebase/phylows/study/TB2:S16928

462

463 **6.0 Acknowledgement**

464 The authors acknowledge the support from Yadav's laboratory and the members of
465 Mgbeahuruike's Molecular Biology Laboratory, Department Microbiology, University of
466 Nigeria Nsukka. The efforts of the staff of the Department of Cell and Molecular Biology,
467 Uppsala University are well acknowledged.

468

469 **References**

470

471 Adomas A, Eklund M, Johansson M, Asiegbu FO, 2006. Identification and analysis of
472 differentially expressed cDNAs during nonself-competitive interaction between *Phlebiopsis*
473 *gigantea* and *Heterobasidion parviporum***.** *FEMS Microbiology Ecology* **57**:26-39.

474 Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ, 1997. Gapped
475 BLAST and PSI BLAST: a new generation of protein database search programs**.** *Nucleic Acids*
476 *Research* **25**:3389–3402.

477 Asiegbu FO, Adomas A, Stenlid J, 2005. Conifer root and butt rot caused by *Heterobasidion*
478 *annosum* (Fr.) Bref. s.l. *Molecular Plant Pathology* **6**:395-409.

479 Benjamini Y, Hochberg Y, 1995. Controlling the FALSE Discovery Rate: A Practical and
480 Powerful Approach to Multiple Testing. *Journal of Royal Statistical Society*. Series B
481 (Methodological) Vol. 57, No 289-300.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE, 2000. The Protein Data Bank**.** *Nucleic Acids Research* **28**:235–242.

Bezalel L, Hadar Y, Fu PP, Freeman JP, Cerniglia C, 1996. Initial oxidation products in the metabolism of pyrene, anthracene, fluorene, and dibenzothiophene by the white rot fungus *Pleurotus ostreatus*. *Applied Environmental Microbiology* **62**:2554–2559

Bezalel L, Hadar Y, Fu PP, Freeman JP, Cemiglia CE, 1996a. Metabolism of phenanthrene by the white rot fungus *Pleurotus ostreatus*. *Applied Environmental Microbiology* **62**:2547–2553.

Bhatnagar D, Ehrlich KC, Cleveland TE, 2003. Molecular genetic analysis and regulation of aflatoxin biosynthesis. *Applied Microbiology and Biotechnology* **61**: 83–93.

Black SD, Coon MJ, 1987. Cytochromes P450: structure and function**.** *Advances in Enzymology* **60:**35–87.

Boddy L, Rayner ADM, 1983. Mycelial interactions, morphogenesis and ecology of Phlebia radiata and P. rufa from oak. *Transactions of British Mycological Society* **80**: 437–448

Chang S, Puryear J, Cairney J, 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molocular Biology Report* **11**:113–116

Chen W, Lee MK, Jefcoate C, Kim SC, Chen F, Y JH, 2014. Fungal Cytochrome P450 Monooxygenases: Their Distribution, Structure, Functions, Family Expansion, and Evolutionary Origin**.** *Genome Biology Evolution* **6**:1620–1634.

Chen S, Xu J, Liu C, Zhu Y, Nelson DR, Zhou S, Li C, Wang L, Guo X, Sun Y, Luo H, Li Y, Song J, Henrissat B, Levasseur A, Qian J, Li J, Luo X, Shi L, He L, Xiang L, Xu X, Niu Y, Li Q, Han MV, Yan H, Zhang J, Chen H, Lv A, Wang Z, Liu M, Schwartz DC, Sun C, 2012. Genome sequence of the model medicinal mushroom *Ganoderma lucidum*. *Nature Communication* | 3:913 | DOI: 10.1038/ncomms1923.

505 Črešnar B, Petrič S, 2011. Cytochrome P450 enzymes in fungal kingdom. *Biochimica et*
506 *Biophysica Acta* **1814**: 29-35.
507 Carruthers S, Rayner ADM, 1979. Fungal communities in decaying wood branches.
508 *Transactions of British Mycological* Society **72**: 283–289.
509 Denison MS, Whitlock JP, 1995. Xenobiotic-inducible Transcription of Cytochrome P450
510 Genes. *The Journal of Biological Chemistry* **270**: 18175-18178.
511 Deng J, Carbone I, 2007. The evolutionary history of Cytochrome P450 genes in four
512 filamentous Ascomycetes. *BMC Evolutionary Biology* **7**:30.
513 DiGuistini S, Wang Y, Liao NY, Taylor G, Tanguay P, Feau N, Henrissat B, Chan SK, Hesse-
514 Orce U, Alamouti SM, Tsui CKM, Docking RT, Levasseur A, Haridas S, Robertson G, Birol I,
515 Holt RA, Marra MA, Hamelin RC, Hirst M, Jones SJM, Bohlmann J, Breuil C, 2011. Genome
516 and transcriptome analyses of the mountain pine beetle-fungal symbiont *Grosmannia clavigera*,
517 a lodge pole pine pathogen. *Proceedings of the National Academy of Sciences of the United*
518 *States of America* **108**: 2504–2509.
519 Doddapaneni H, Chakraborty R, Yadav JS, 2004. Genome-wide structural and evolutionary
520 analysis of the P450 monooxygenase genes (CYPome) in the white rot fungus *Phanerochaete*
521 *chrysosporium*: Evidence for gene duplications and extensive gene clustering. *BMC Genomics*
522 **6**:92.
523 Dowding P, 1978. Methods for studying microbial interactions. *Annals of Applied Biology* **89**:
524 166–171.
525 Eastwood DC, Floudas D, Binder M, Majcherczyk A, Schneider P, Aerts A, Asiegbu FO, Baker
526 SE, Barry K, Bendiksby M, Blumentritt M, Coutinho PM, Cullen D, de Vries RP, Gathman A,
527 Goodell B, Henrissat B, Ihrmark K, Kauserud H, Kohler A, LaButti K, Lapidus A, Lavin JL, Lee

528 YH, Lindquist E, Lilly W, Lucas S, Morin E, Murat C, Oguiza JA, Park J, Pisabarro AG, Riley
529 R, Rosling A, Salamov A, Schmidt O, Schmutz J, Skrede I, Stenlid J, Wiebenga A, Xie X, Ku¨es
530 U, Hibbett DS, Hoffmeister D, Ho¨gberg N, Martin F, Grigoriev IV, Watkinson SC, 2011. The
531 plant cell wall-decomposing machinery underlies the functional diversity of forest fungi**.** *Science*
532 **333**:762–765.
533 Fernandez-Fueyo E, Ruiz-Duen˜as FJ, Ferreira P, Floudas D, Hibbett DS, Canessa P, Larrondo
534 LF, James TY, Seelenfreund D, Lobos S, Polanco R, Tello M, Honda Y, Watanabe T, Watanabe
535 T, Ryu JS, Kubicek CP, Schmoll M, Gaskell J, Hammel KE, St John FJ, vanden Wymelenberg
536 A, Sabat G, Splinter BonDurant S, Syed K, Yadav JS, Doddapaneni H, Subramanian V, Lavı´n
537 JL, Oguiza JA, Perez G, Pisabarro AG, Ramirez L, Santoyo F, Master E, Coutinho PM,
538 Henrissat B, Lombard V, Magnuson JK, Ku¨es U, Hori C, Igarashi K, Samejima M, Held BW,
539 Barry KW, LaButti KM, Lapidus A, Lindquist EA, Lucas SM, Riley R, Salamov AA,
540 Hoffmeister D, Schwenk D, Hadar Y, Yarden O, de Vries RP, Wiebenga A, Stenlid J, Eastwood
541 D, Grigoriev IV, Berka RM, Blanchette RA, Kersten P, Martinez AT, Vicuna R, Cullen D**,** 2012**.**
542 Comparative genomics of *Ceriporiopsis subvermispora* and *Phanerochaete chrysosporium*
543 provide insight into selective ligninolysis. *Proceedings of the National Academy of Sciences of*
544 *the United States of America* **109**: 5458–5463.
545 Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, Martı´nez AT,Otillar R,
546 Spatafora JW, YadavJS, Aerts A, Benoit I, Boyd A, Carlson A, Copeland A, Coutinho PM, de
547 Vries RP, Ferreira P, Findley K, Foster B,Gaskell J, Glotzer D, Go´recki P, Heitman J, Hesse C,
548 Hori C, Igarashi K, Jurgens JA, Kallen N, Kersten P, Kohler A,Ku¨es U, Kumar TK, Kuo A,
549 LaButti K, Larrondo LF, Lindquist E, Ling A, Lombard V, Lucas S, Lundell T, Martin
550 R,McLaughlin DJ,Morgenstern I,Morin E,Murat C, Nagy LG, NolanM, Ohm RA,

551 Patyshakuliyeva A, Rokas A, Ruiz-Duen˜as FJ, Sabat G, Salamov A, Samejima M, Schmutz J,
552 Slot JC, St John F, Stenlid J, SunH, Sun S, Syed K, Tsang A,Wiebenga A, Young D, Pisabarro
553 A, Eastwood DC, Martin F, Cullen D, Grigoriev IV, Hibbett DS, 2012. The Paleozoic origin of
554 enzymatic lignin decomposition reconstructed from 31 fungal genomes**.** *Science* **336**: 1715–
555 1719.
556 Garbelotto M, Guglielmo F, Mascheretti S, Croucher PJP, Gonthier P, 2013. Population genetic
557 analyses provide insights on the introduction pathway and spread patterns of the North American
558 forest pathogen *Heterobasidion irregulare* in Italy. *Molecular Ecology* **22**: 4855-4869.
559 Gonzalez FJ, Lee YH, 1996. Constitutive expression of hepatic cytochrome P450 genes. *FASEB*
560 *Journal* **10**: 1112–1117.
561 Hansson D, Menkis A, Himmelstrand K, Thelander M, Olson Å, Stenlid J, Karlsson M, Broberg
562 A, 2012a. Sesquiterpenes from the conifer root rot pathogen *Heterobasidion occidentale*.
563 *Phytochemistry* **82:** 158–165.
564 Hansson D, Menkis A, Olson Å, Stenlid J, Broberg A, Karlsson M, 2012b. Biosynthesis of
565 fomannoxin in the root rotting pathogen *Heterobasidion occidentale*. *Phytochemistry* **84**: 31–39.
566 Hargrove TY, Wawrzak Z, Liu J, Waterman MR, Nes WD, Lepesheva GI, 2012. Structural
567 complex of sterol 14a-demethylase (CYP51) with 14a-methylenecyclopropyl-_7-24, 25-
568 dihydrolanosterol. *Journal of Lipid Research* **53**:311–320.
569 Harris M, Jones T, 2001. "Molray-a web interface between O and the POV-Ray ray tracer." *Acta*
570 *Crystallographica Section D Biological Crystallography* **57**: 1201-1203.
571 Hirosue S, Tazaki M, Hiratsuka N, Yanai S, Kabumoto H, Shinkyo R, Arisawa A, Sakaki T,
572 Tsunekawa H, Johdo O, Ichinose H, Wariishi H, 2011. Insight into functional diversity of
573 cytochrome P450 in the white-rot basidiomycete *Phanerochaete chrysosporium***:** involvement of

574 versatile monooxygenase. *Biochemical and Biophysical Research Communications* **407**:118–
575 1233.
576 Ide M, Ichinose H, Wariishi H, 2012. Molecular identification and functional characterization of
577 cytochrome P450 monooxygenases from the brown-rot basidiomycete *Postia placenta*. *Arch*
578 *Microbiology* **194**:243–253.
579 Jones TA, Zou JY, Cowan SW, Kjeldgaard M, 1991. Improved methods for building protein
580 models in electron density maps and the location of errors in these models**.** *Acta*
581 *Crystallographica Section A* **47**:110–119.
582 Karlsson M, Elfstrand M, Stenlid J, Olson A, 2008. A fungal cytochrome P450 is expressed
583 during the interaction between the fungal pathogen *Heterobasidion annosum* sensu lato and
584 conifer trees. *DNA Sequence* **19**: 115–120
585 Kleywegt GJ, Zo JY, et al, 2001. Around O. International Tables for Crystallography. In
586 Crystallography of Biological Macromolecules**.** Edited by Rossmann MG, Arnold E. *Dordrecht:*
587 *Kluwer Academic* 353–356.
588 Kraulis PJ, 1991."Molscript - a Program to Produce both Detailed and Schematic Plots of Protein
589 Structures." *Journal of Applied Crystallography* **24**: 946-950.
590 Liu D, Gong J, Dai W, Kang X, Huang Z, Zhang HM, Liu W, Liu L, Ma J, Xia Z, Chen Y, Chen
591 Y, Wang D, Ni P, Guo AY, Xiong X, 2012. The Genome of *Ganderma lucidum* Provide Insights
592 into Triterpense Biosynthesis and Wood Degradation *PLoS One* **7(5**):e36146. doi:
593 10.1371/journal.pone.0036146.
594 MacDonald J, Doering M, Canam T, Gong Y, Guttman GS, Campbell MM, Master ER, 2011.
595 Transcriptome responses of the softwood-degrading white-rot fungus *Phanerochaete carnosa*

during growth on coniferous and deciduous wood. *Applied Environmental Microbiology* **77**: 3211–3218.

Magan N, Lacey J, 1984a. Effect of temperature and pH on water relations of field and storage fungi. *Transactions of British Mycological Society* **82**: 71–81.

Magan N, Lacey J, 1984b. Effect of water activity, temperature and substrate on interactions between field and storage fungi. *Transactions of British Mycological Society* **82**: 83–91.

Martinez D, Larrondo LF, Putnam N, Sollewijn Gelpke MD, Huang K, Chapman J, Helfenbein KG, Ramaiya P, Detter JC, Larimer F, Coutinho PM, Henrissat B, Berka R, Cullen D, Rokhsa D, 2004. Genome sequence of the lignocellulose degrading fungus *Phanerochaete chrysosporium* strain RP78. *Nature Biotechnology* **22**: 695-700.

Martinez D, Challacombe J, Morgenstern I, Hibbett D, Schmoll M, Kubicek CP, Ferreira P, Ruiz-Duenas FJ,Martinez AT, Kersten P, Hammel KE, vanden Wymelenberg A, Gaskell J, Lindquist E, Sabat G, Bondurant SS, Larrondo LF, Canessa P, Vicuna R, Yadav J, Doddapaneni H, Subramanian V, Pisabarro AG, Lavı´n JL, Oguiza JA, Master E, Henrissat B, Coutinho PM, Harris P, Magnuson JK, Baker SE, Bruno K, Kenealy W, Hoegger PJ, Ku¨es U, Ramaiya P, Lucas S, Salamov A, Shapiro H, Tu H, Chee CL, Misra M, Xie G, Teter S, Yaver D, James T, Mokrejs M, Pospisek M, Grigoriev IV, Brettin T, Rokhsar D, Berka R, Cullen D, 2009. Genome, transcriptome and secretome analysis of the fungus *Postia placenta* supports unique mechanisms of lignocellulose conversion *Proceedings of the National Academy of Sciences of the United States of America* **106**: 1954–1959.

Masaphy S, Levanon D, Henis Y, Venkateswarlu K, Kelly SL, 1996. Microsomal and cytosolic cytochrome P450 mediated benzo(a)pyrene hydroxylation in *Pleurotus pulmonarius*. *Biotechnology Letter* **17**:967–974

Mast NMA White et al, 2008. "Crystal structures of substrate-bound and substrate-free cytochrome P450 46A1, the principal cholesterol hydroxylase in the brain." *Proceedings of the National Academy of Sciences of the United States of America* **105**: 9546-9551.

Mgbeahuruike AC, Sun H, Fransson P, Kasanen R, Daniel G, Karlsson M, Asiegbu FO, 2011. Screening of *Phlebiopsis gigantea* isolates for traits associated with biocontrol of the conifer pathogen *Heterobasidion annosum*. *Biological Control* **57**: 118–129.

Mgbeahuruike AC, Karlsson M, Asigbu FO, 2012. Differential expression of two hydrophobins (*Pgh1* and *Pgh2*) from the biocontrol fungus *Phlebiopsis gigantea*. *Fungal Biology* **116**:620–629.

Mgbeahuruike AC, Kovalchuk A, Chen H, Ubhayasekera Asiegbu FO**,** 2013**.** Evolutionary analysis of hydrophobin gene family in two wood-degrading basidiomycetes, *Phlebia brevispora* and *Heterobasidion annosum* s.l. *BMC Evolutionary Biology* **13**:240.

Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K, 1993. The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biology* **12**:1–51.

Nelson DR, 2009. The cytochrome p450 homepage. *Human Genomics* **4**: 59–6

Nelson DR, 2011. Progress in tracing the evolutionary paths of cytochrome P450. *Biochimica et Biophysica Acta* **1814**: 14-8.

Nelson DR, 1998. Metazoan cytochrome P450. *Pharmacology Toxicology Endocrinol*ogy **121**: 15-22.

Olson Å, Aerts A, Asiegbu F, Belbahri L, Bouzid O, Broberg A, Canbäck B, Coutinho PM, Cullen D, Dalman K, Deflorio G, van Diepen LTA, Dunand C, Duplessis S, Durling M, Gonthier

642 P, Grimwood J, Fossdal CG, Hansson D, Henrissat B, Hietala A, Himmelstrand K, Hoffmeister
643 D, Högberg N, James TY, Karlsson M, Kohler A, Kues U, Lee YH, Yao-Cheng L, Lind M,
644 Lindquist E, Lombard V, Lucas S, Lunde´n K, Morin Claude ME, Park J, Raffaello T, Rouze´ P,
645 Salamov A, Schmutz J, Solheim H, Sta°hlberg J, Ve´le¨z H, de Vries RP, Wiebenga A,
646 Woodward S, Yakovlev I, Garbelotto M, Martin F, Grigoriev IV, Stenlid J, 2012. Insight into
647 trade-off between wood decay and parasitism from the genome of a fungal forest pathogen. *New*
648 *Phytologist* **194**: 1001–1013.

649 Ortiz de Montellano, PR, 2005. Cytochrome P450: structure, mechanism, and biochemistry, 3rd
650 edn. Kluwer Academic/Plenum, New York.

651 Otrosina WJ, Garbelotto M, 2010. *Heterobasidion occidentale* sp. nov. and *Heterobasidion*
652 *irregulare* nom. nov. A disposition of North American *Heterobasidion* biological species.
653 *Fungal biology* **114**: 16-25.

654 Park J, Park J, Jang S, Kim S, Kong S, Choi J, Ahn K, Kim J, Lee S, Kim S, Park B, Jung K,
655 Kim S, Kang S, Lee YH, 2008. FTFD: An Informatics Pipeline Supporting Phylogenomic
656 Analysis of Fungal Transcription Factors. *Bioinformatics* **24**:1024-1025.

657 Piri T, Korhonen K, Sairanen A, 1990. Occurrence of *Heterobasidion annosum* in pure and
658 mixed stands in Southern Finland. *Scandinavian Journal of Forest Research* **5**: 113-125.

659 Sansen S, JK Yano et al**,** 2007. "Adaptations for the oxidation of polycyclic aromatic
660 hydrocarbons exhibited by the structure of human P450 1A2." *The Journal of biological*
661 *chemistry* **282**: 14348-14355.

662 Rayner ADM, 1978. Interactions between fungi colonizing hardwood stumps and their possible
663 role in determining patterns of colonization and succession. *Annals of Applied Biology* **89**: 131–
664 134.

665 Sono M, Roach RP, Coulter E.D, Dawson J.D, 1996. Heme-containing oxygenases, *Chemical*
666 *Reviews* **96**: 2841-2888.
667 Stenlid J, 1987. Controlling and predicting the spread of *Heterobasidion annosum* from infected
668 stumps and trees of *Picea abies*. *Scandinavian Journal of Forest Research* **2**: 187-198.
669 Sun H, Korhonen K, Hantula J, Kasanen R, 2009a. Variation in properties of *Phlebiopsis*
670 *gigantea* related to biocontrol against infection by *Heterobasidion spp*. in Norway spruce
671 stumps. *Forest Pathology* **39**: 133–144.
672 Sun H, Korhonen K, Hantula J, Asiegbu FO, Kasanen R, 2009b. Use of a breeding approach for
673 improving biocontrol efficacy of *Phlebiopsis gigantea* strains against *Heterobasidion* infection
674 of Norway spruce stumps. *FEMS Microbiology Ecology* **69**:266 -273.
675 Syed K, Doddapaneni H, Subramanian V, Lam YW, Yadav JS, 2010. Genome-to-function
676 characterization of novel fungal P450 monooxygenases oxidizing polycyclic aromatic
677 hydrocarbons (PAHs). *Biochemical and Biophysical Research Communications* **399**: 492-497.
678 Syed K, Nelson DR, Riley R, Yadav JS, 2013. Genomewide annotation and comparative
679 genomics of cytochrome P450 monooxygenases (P450s) in the polypore species *Bjerkandera*
680 *adusta*, *Ganoderma sp*. and *Phlebia brevispora*. *Mycologia* **105:**1445-55.
681 Syed K, Kattamuri C, Thompson T.B, Yadav JS, 2011a. Cytochrome b5 reductase-cytochrome
682 b5 as an active P450 redox enzyme system in *Phanerochaete chrysosporium*: Atypical properties
683 and in vivo evidence of electron transfer capability to CYP63A2. *Archives of Biochemistry and*
684 *Biophysics* 26-32.
685 Syed K, Porollo A, Lam YW, Yadav JS, 2011b. A fungal P450 (CYP5136A3) capable of
686 oxidizing polycyclic aromatic hydrocarbons and endocrine disrupting alkylphenol: Role of
687 Trp129 and Leu 324. *PLoS One* 6 (12) e28286. doi:10.1371/journal.pone.0028286

688 Syed K, Yadav JS, 2012. P450monooxygenases (P450ome) of the model white rot
689 fungus *Phanerochaete chrysosporium*. *Critical Review of Microbiology* **38**: 339–363
690 Syed K, Yadav JS, 2012. P450 monooxygenases (CYPome) of the model white rot fungus
691 *Phanerochaete chrysosporium*. *Critical Reviews in Microbiology* **38**: 339-363.
692 Syed K, Shale K, Pagadala NS, Tuszynski J, 2014. Systematic identification and evolutionary
693 analysis of catalytically versatile cytochrome p450 monooxygenase families enriched in model
694 basidiomycete fungi. *PLoS One* **22**:9(1):e86683.
695 Woodward S, Stenlid J, Karjalainen R, Huttermann A, 1998. *Heterobasidion annosum*: Biology
696 Impact and Control. Wallingford, UK and New York, USA, *CAB International* pp 589.
697 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S, 2011. MEGA 5: Molecular
698 evolutionary genetics analysis using maximum likelihood, evolutionary distance and maximum
699 parsimony methods**.** *Molecular Biology and Evolution* **28**:2731–2739
700 Thompson JD, Higgins DG, Gibson TJ, 1994. CLUSTAL W, Improving the sensitivity of
701 progressive multiple sequence alignment through sequence weighting, position-specific gap
702 penalties and weight matrix choice. *Nucleic Acids Research* **22**:4673–4680.
703 Van der Nest MA, Olson Å, Karlsson M, Lind M, Dalman K, Brandström-Durling M, Elfstrand
704 M, Wingfield BD, Stenlid J, 2014. Gene expression associated with intersterility in
705 Heterobasidion. *Fungal Genetics and Biology* **73**: 104–119
706 Yadav JS, Doddapaneni H, Subramanian V, 2006. CYPome of the white rot fungus
707 *Phanerochaete chrysosporium*: Structure, evolution and regulation of expression of genomic
708 P450 clusters *Biochemical Society Transactions* **34**: 1165-1169.

709 Yano JK, Wester MR et al, 2004. "The structure of human microsomal cytochrome P450 3A4
710 determined by X-ray crystallography to 2.05-A resolution." *The Journal of biological chemistry*
711 **279**(37): 38091-38094.

712
713

714 **Figure Legends**

715 **Figure 1**: Maximum-likelihood phylogenetic tree of the CYP proteins from the genome of
716 *Heterobasidion irregulare*. The nomenclature of CYP proteins follows the one shown in the
717 **Supplementary Table I**. Predicted pseudogene sequences were excluded from the analysis. All
718 CYP families are indicated as well as subfamilies within the family CYP5144 represented by at
719 least two sequences. Numbers next to the branching points indicate the support from 100
720 bootstrap replicates (only values above 60 are shown). Most of the families (except for the
721 CYP5144) were recovered as monophyletic groups. The tree is drawn to scale, with branch
722 lengths measured in the number of substitutions per site. The analysis involved 121 amino acid
723 sequences. Evolutionary analyses were conducted in MEGA5.

724

725 **Figure 2:** The organization of CYP-encoding genes in the genome of *Heterobasidion*
726 *irregulare*. (**a**) *H.irregulare* scaffold 3 (**b**) *H. irregulare* scaffold 5 (**c**) *H. irregulare* scaffold 8
727 (**d**) *H. irregulare* scaffold 11 (**e**) *H. irregulare* scaffold 10. The scheme illustrates the genomic
728 organization of representative set of CYP-encoding genes. Corresponding genes are shown as
729 open arrows with their names indicated; predicted pseudogenes are indicated with a letter ψ.
730 Note that all illustrated tandem arrays are formed by genes belonging to the same subfamily and
731 located on the same DNA strand; the clusters of CYP-encoding genes are not interrupted by any
732 other unrelated genes.

Figure 3: Ribbon cartoons of the homology models of *Heterobasidion irregulare* CYPs (**A**) CYP63A22 in gold and (**B**) CYP5150S3 in moccasin. Conserved catalytic site residues and heme binding residues are shown in royal blue and light gray respectively. Modeled cholesterol-3-sulfate to visualize the catalytic pocket is shown in magenta. Heme is shown as a line drawing.

Figure 4: Ribbon cartoons of the homology models of *Heterobasidion irregulare* CYP (**A**) CYP5144M12 in pale green (**B**) CYP5144BJ1 in light turquoise (**C**) CYP5037B16 in aquamarine and (**D**) CYP5344D1 in chartreuse. Conserved catalytic site residues and heme binding residues are shown in brick-red and light gray respectively. Modeled alpha-naphthoflavone to visualize the catalytic pocket is shown in orange. Heme is shown as a line drawing.

Figure 5: Microarray analysis of CYP transcripts up-regulated during growth of *H. irregulare* in culture filtrate produced by *P. gigantea*. Expression data were normalized to liquid ME media and calculated as fold changes between *H. annosum* s.s grown in liquid culture filtrate from *P. gigantea*) over the control (*H. annosum* s.s grown in fresh liquid ME media), but the values are presented in percentages. CYP transcripts in the pie chart represent different CYP families.

Figure 6: Microarray analysis of CYP transcripts that were down-regulated during growth of *H. irregulare* in culture filtrate produced by *P. gigantea*. Expression data were normalized to liquid ME media and calculated as fold changes between *H. annosum* s.s grown in liquid culture filtrate from *P. gigantea*) over the control (*H. annosum* s.s grown in fresh liquid ME media), but the

values are presented in percentages. CYP transcripts in the pie charts represent different CYP families.

Additional Files

Suppl. Table 1: Inventory and classification of *H. irregulare* v 2.0 (Hetan 2.0) CYPs. Classification was done based on sequence relatedness to known fungal CYPs in consultation with the Committee on Standardized CYP Nomenclature. Sequences were assigned to families and subfamilies according to nomenclature rules for P450 sequences (> 40% homology for assigning a family and > 55% for a subfamily). The families were then grouped into clans.

Suppl. Table 2: Microarray expression analysis of cytochrome P450 in *H. irregulare.* Expression data was obtained by determining the fold change between the experimental sample (*H. annosum* s.s grown in liquid culture filtrate from *P. gigantea*) and the control (*H. annosum* grown in fresh liquid ME media).

Supplementary Figures

Suppl. Figure I: Neighbor-joining phylogenetic tree of the CYP proteins from the genome of *H. irregulare*. The nomenclature of CYP proteins follows the one shown in the Supplementary Table I. Predicted pseudogene sequences were excluded from the analysis. Numbers next to the branching points indicate the relative support from 500 bootstrap replicates. Most of the families (except for the CYP5144) were recovered as monophyletic groups. The analysis involved 121

amino acid sequences. All positions with less than 0% site coverage were eliminated. There were a total of 416 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.2.

Suppl. Figure 2: Maximum-parsimony phylogenetic tree of the CYP proteins from the genome of *H. irregulare*. The nomenclature of CYP proteins follows the one shown in the Supplementary Table I. Predicted pseudogene sequences were excluded from the analysis. Numbers next to the branching points indicate the support from 100 bootstrap replicates. Most of the families (except for the CYP5144) were recovered as monophyletic groups. The analysis involved 121 amino acid sequences. All positions with less than 0% site coverage were eliminated. There were a total of 416 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.2.

Suppl. Figure 3: Minimum-evolution phylogenetic tree of the CYP proteins from the genome of *H. irregulare*. The nomenclature of CYP proteins follows the one shown in the Supplementary Table I. Predicted pseudogene sequences were excluded from the analysis. Numbers next to the branching points indicate the support from 100 bootstrap replicates. Most of the families (except for the CYP5144) were recovered as monophyletic groups. The analysis involved 121 amino acid sequences. All positions with less than 0% site coverage were eliminated. There were a total of 416 positions in the final dataset. Evolutionary analyses were conducted in MEGA5.2.

Table 1. **P450ome annotation and classification in *H. irregulare***

S/N	Family	Subfamilies	Member P450 genes[a]	Clan

1	CYP5035	X, Y	CYP5035X1, CYP5035Y2P, CYP5035Y1	53
2	CYP51	F	CYP51F1	51
3	CYP53	C	CYP53C8	53
4	CYP61	A	CYP61A1	51
5	CYP63	A, C, J	CYP63A22, CYP63C4, CYP63C3, CYP63J1	52
6	CYP502	B, C	CYP502B9, CYP502C3P, CYP502C4, CYP502C1, CYP502C2	64
7	CYP505	D	CYP505D20	505
8	CYP512	P	CYP512P5P, CYP512P3, CYP512P4P	54
9	CYP5035	X, Y	CYP5035X1, CYP5035Y2P, CYP5035Y1	53
10	CYP5037	B, R, S, T, U V	CYP5037B15, CYP5037B16, CYP5037R1, CYP5037S1, CYP5037T1, CYP5037U1, CYP5037V1	64
11	CYP5065	B, C, D	CYP5065B2, CYP5065B3, CYP5065C1, CYP5065C2, CYP5065D1	64
12	CYP5136	F	CYP5136F4, CYP5136F3, CYP5136F2, CYP5136F1	56
13	CYP5138	A	CYP5138A6	56
14	CYP5139	D, M, N	CYP5139D9, CYP5139D10, CYP5139M1, CYP5139N1	56
15	CYP5140	A	CYP5140A8	53
16	CYP5141	A	CYP5141A14	534
17	CYP5143	C, D	CYP5143C1, CYP5143D1	53
18	CYP5144	BG, BH, BJ, BK, BL, BM, BN, BQ, K,	CYP5144BG2,CYP5144BG1, CYP5144BG3, CYP5144BG4, CYP5144BH1, CYP5144BH2, CYP5144BJ1,	64

		M	CYP5144BJ2, CYP5144BJ3, CYP5144BJ4, CYP5144BJ5, CYP5144BJ6, CYP5144BJ7, CYP5144BJ8, CYP5144BJ9, CYP5144BK8, CYP5144BK5, CYP5144BK4, CYP5144BK3, CYP5144BK2, CYP5144BK1P, CYP5144BK6, CYP5144BK7, CYP5144BK10P, CYP5144BK9, CYP5144BL2P, CYP5144BL3, CYP5144BL4, CYP5144BL1, CYP5144BM1, CYP5144BM2P, , CYP5144BN1, CYP5144BP1, CYP5144BQ1, CYP5144K2, CYP5144M9, CYP5144M10, CYP5144M11, CYP5144M12, CYP5144M13, CYP5144M14, CYP5144M5, CYP5144M6, CYP5144M7, CYP5144M15, CYP5144M16, CYP5144M4, CYP5144M17, CYP5144M8	
19	CYP5148	B	CYP5148B12	64
20	CYP5150	B, S, T, U, V	CYP5150S6, CYP5150S5P , CYP5150S4, CYP5150S3, CYP5150S1, CYP5150S2, CYP5150T1 , CYP5150T2P, CYP5150U1, CYP5150V1,	56
21	CYP5151	A	CYP5151A8	56
22	CYP5152	E	CYP5152E2, CYP5152E1	64
23	CYP5153	C	CYP5153C1	5153
24	CYP5156	B	CYP5156B2	56
25	CYP5340	C	CYP5340C1P, CYP5340C2P, CYP5340C3,	56

			CYP5340C4P, CYP5340C5	
26	CYP5341	B	CYP5341B4, CYP5341B3	56
27	CYP5344	C, D	CYP5344C6, CYP5344C5, CYP5344C4, CYP5344C1, CYP5344C2, CYP5344C3, CYP5344C7, CYP5344C8, CYP5344D2P, CYP5344D1	64
28	CYP5348	W	CYP5348W1	64
29	CYP5352	A,C	CYP5352A5P, CYP5352A4, CYP5352C1	64
30	CYP5358	B	CYP5358B1,	64
31	CYP5416	D	CYP5416D1	64
32	CYP5429	B	CYP5429B1	534
33	CYP5430	A	CYP5430A1	56
34	CYP5431	A	CYP5431A1	534
35	CYP5432	B	CYP5432B1P	5432
36	CYP6005	A, G	CYP6005A3, CYP6005G1	6001

Family-, subfamily- and clan- level classification of the P450ome of *H. irregulare.* Classification was done based on sequence relatedness to known fungal P450s in consultation with the Committee on Standardized Cytochrome P450 Nomenclature. Sequences that did not fit into existing families or subfamilies were assigned to new families or subfamilies according to nomenclature rules for P450 sequences (> 40% homology for assigning a family and > 55% for a subfamily). The families were then grouped into clans.

[a]= CYP protein-encoding genes in each family

Table 2. Overview of the *H. irregulare* P450ome and its comparison with the P450ome of the model white rot basidiomycete *P. chrysosporium*.

Species	***H. irregulare***	***P. chrysosporium***
Authentic P450s:		
Clans	11	10
Families	35	32
Subfamilies	64	70
Member P450s	**121**	**149**
Pseudogenes	17	10

Comparison of P450omes of the fresh wood-degrading basidiomycete *H. irregulare* and the dead wood-degrading basidiomycete *P. chrysospori*

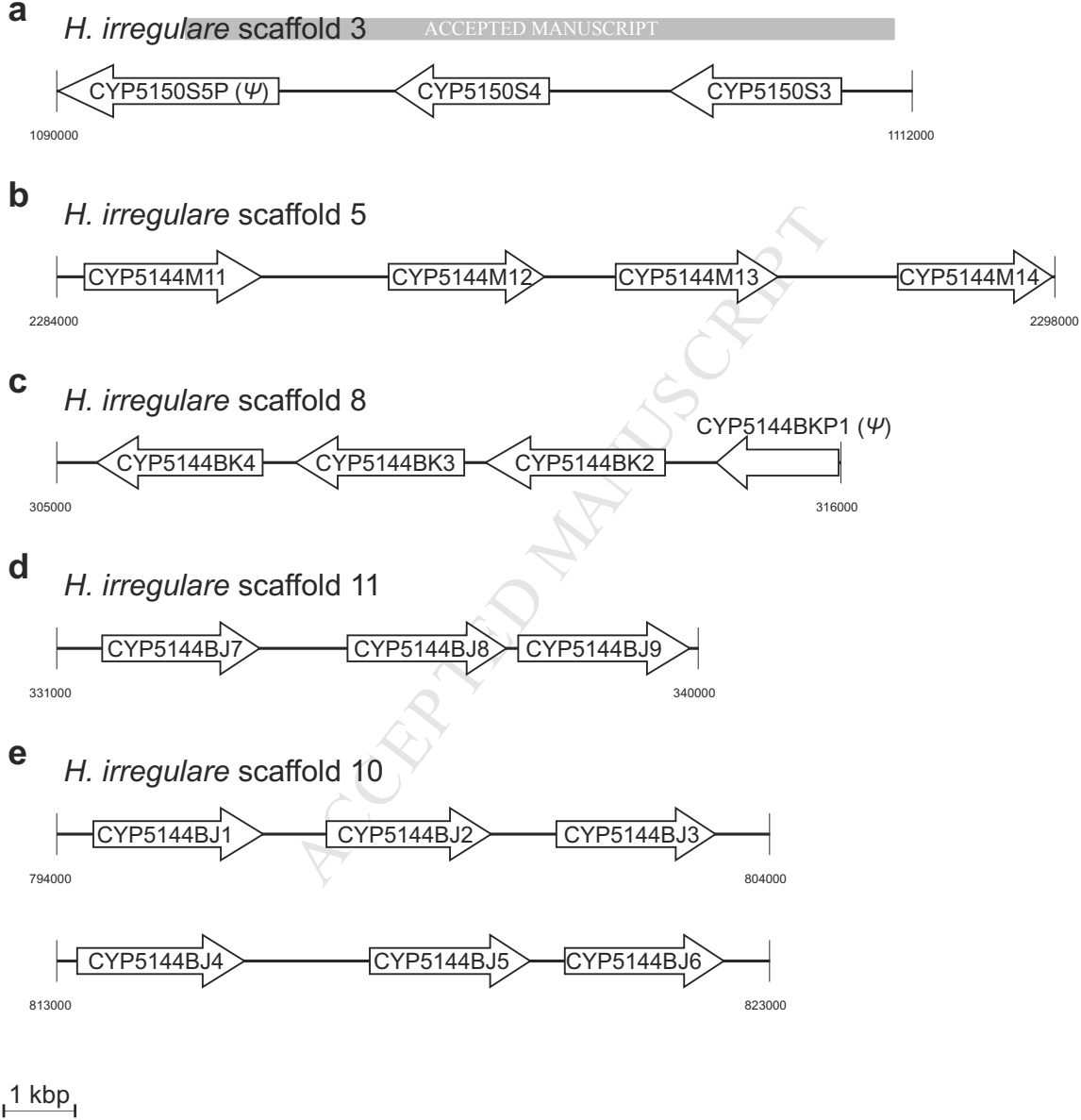
Table 3: **Microarray expression[a] data on statistically significant up-regulated *H. irregulare* CYP genes during growth in culture filtrate from *P. gigantea*.**

Expresion data[a]	**S/N**	**Protein ID**	**CYP Name**	**CYP family**	**Fold change[b]**	**P-value**
Up-regulated	1	442518	CYP502B9	CYP502	2.37	0.04
genes	2	170468	CYP5150S4	CYP5150	132.11	0.001
Down regulated	1	124408	CYP5144BK8	CYP5144	-3.53	0.04
genes	2	37362	CYP5139N1	CYP5139	-3.20	0.05

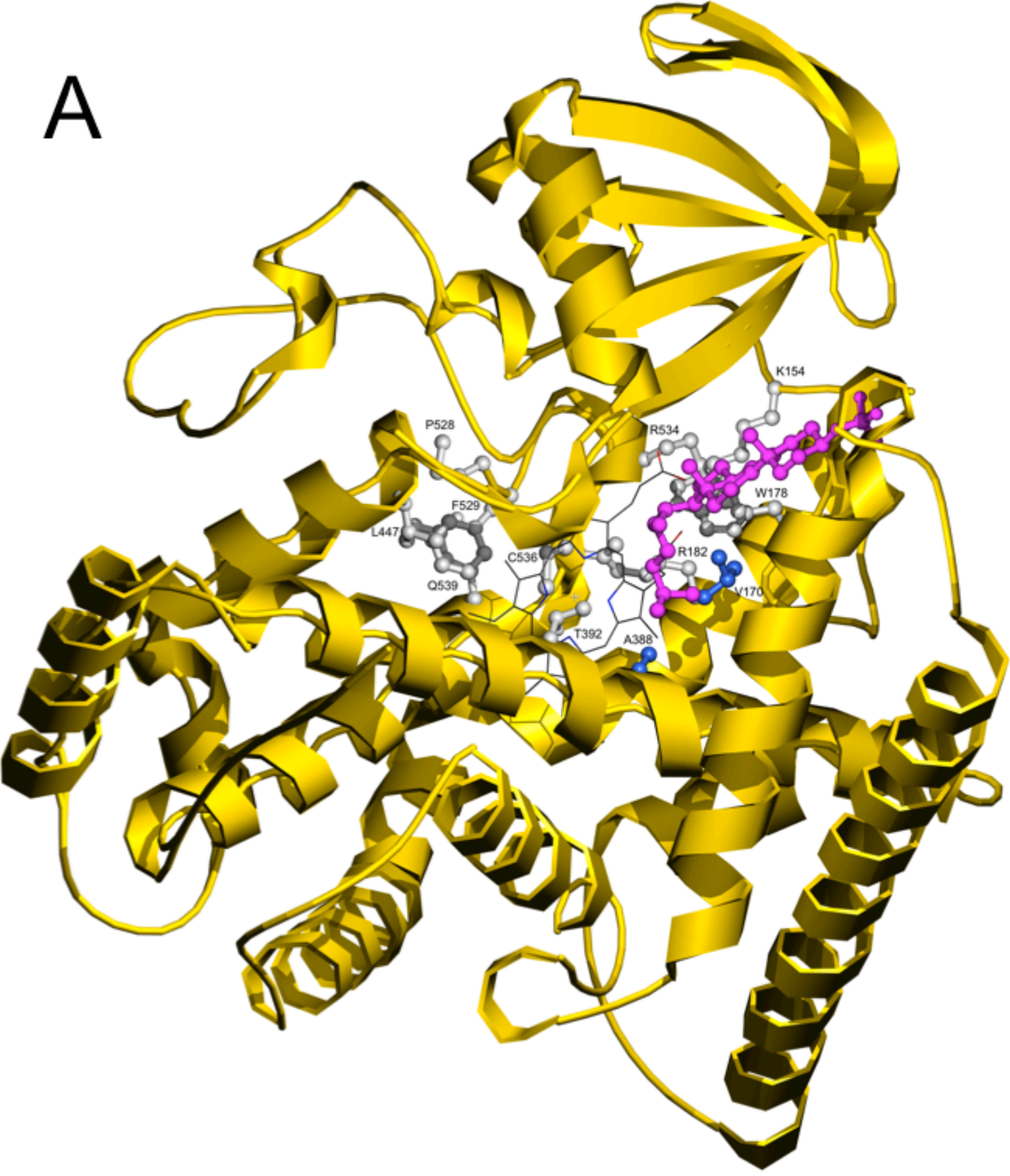
3	126733	CYP5148B12	CYP5148	-81.10	0.01

a= Expression data normalized to liquid ME media.

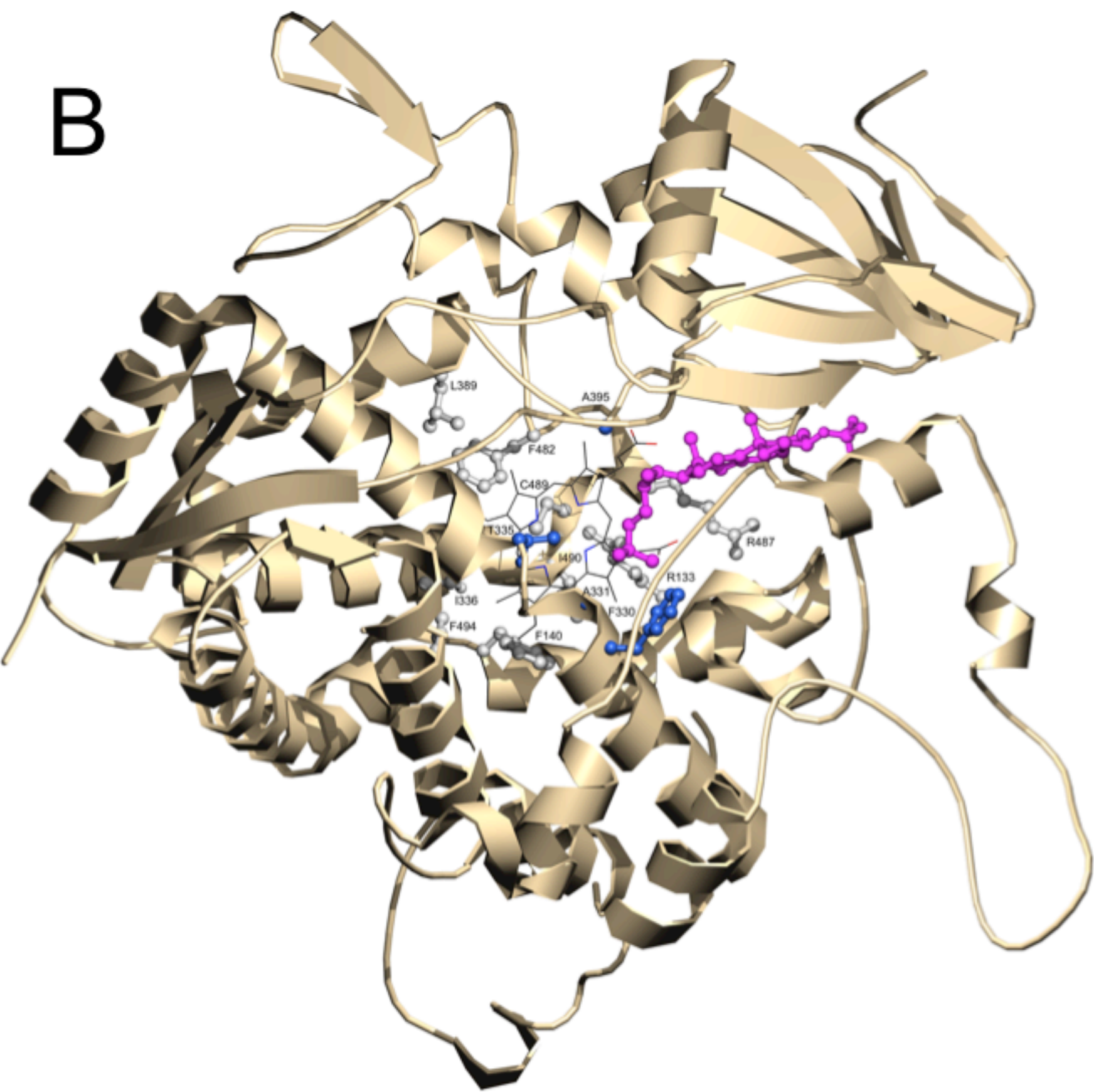
b= Fold changes calculated as the expression value of the experimental sample (*H. annosum* s.s grown in liquid culture filtrate from *P. gigantea*) over the control (*H. annosum* s.s grown in fresh liquid ME media).



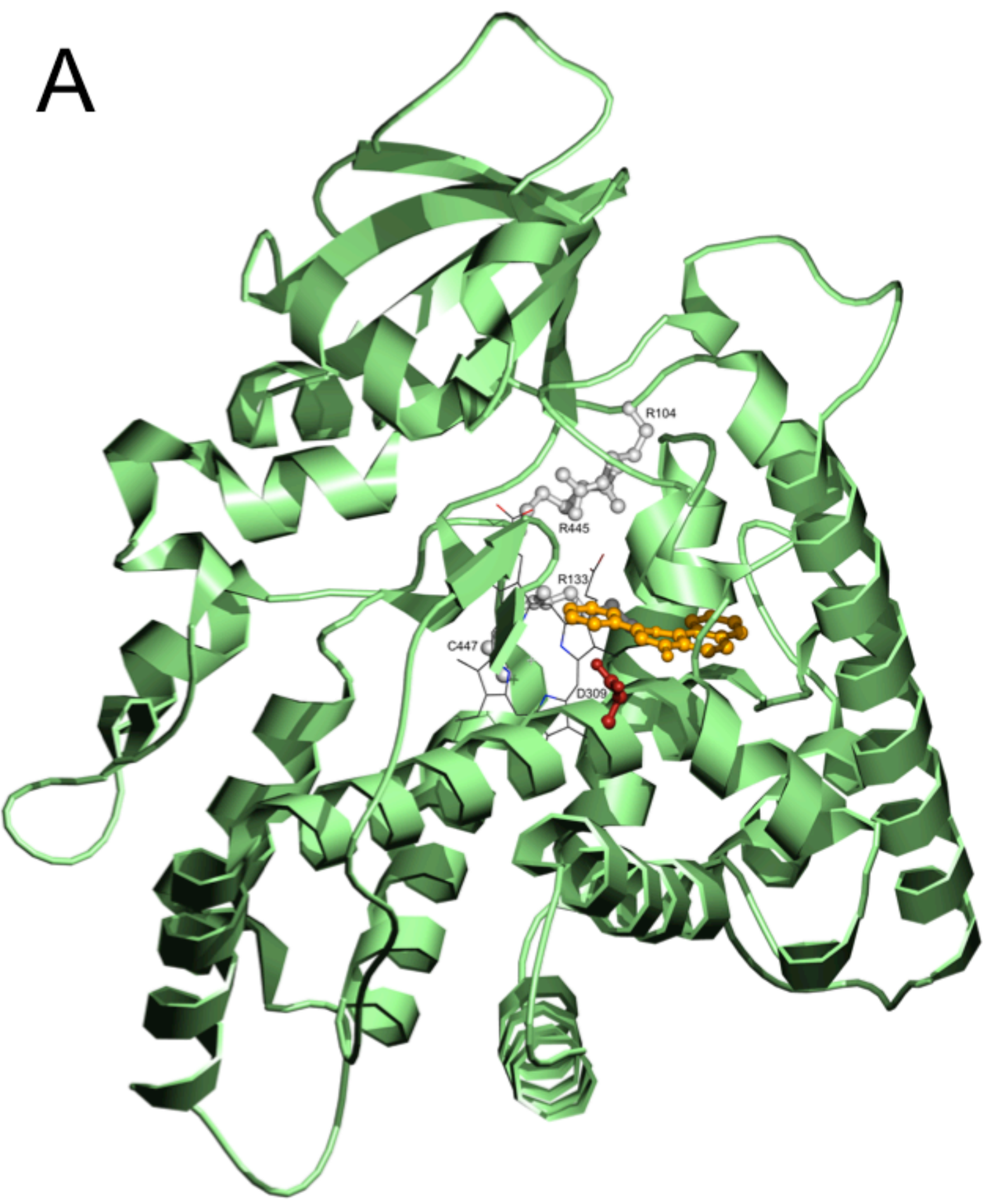
A


K154
P528
R534
W178
F529
L447
C536
R182
Q539
V170
T392
A388

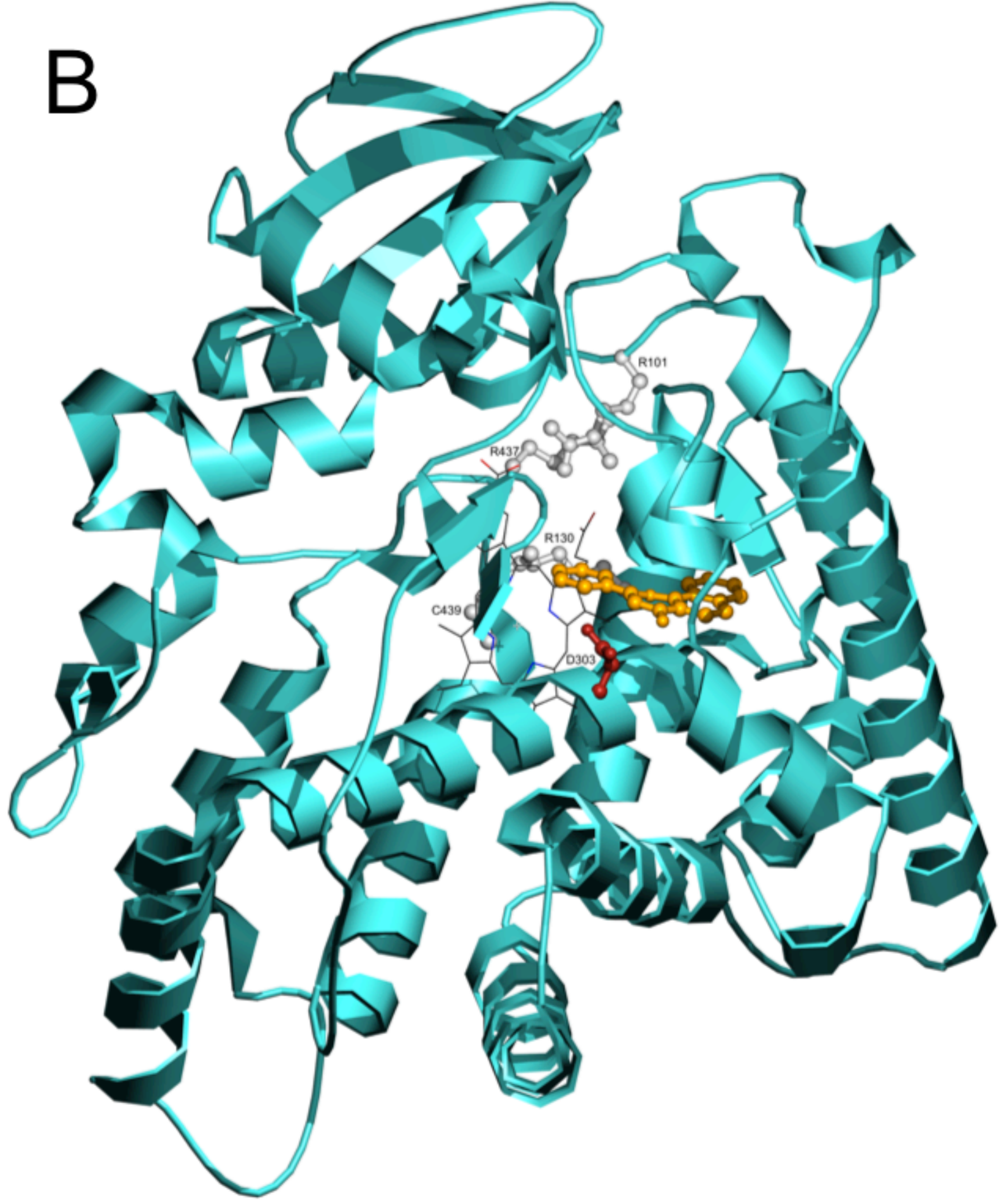
B


L389
A395
F482
C489
T338
I490
R487
R133
I336
A331
F330
F494
F140

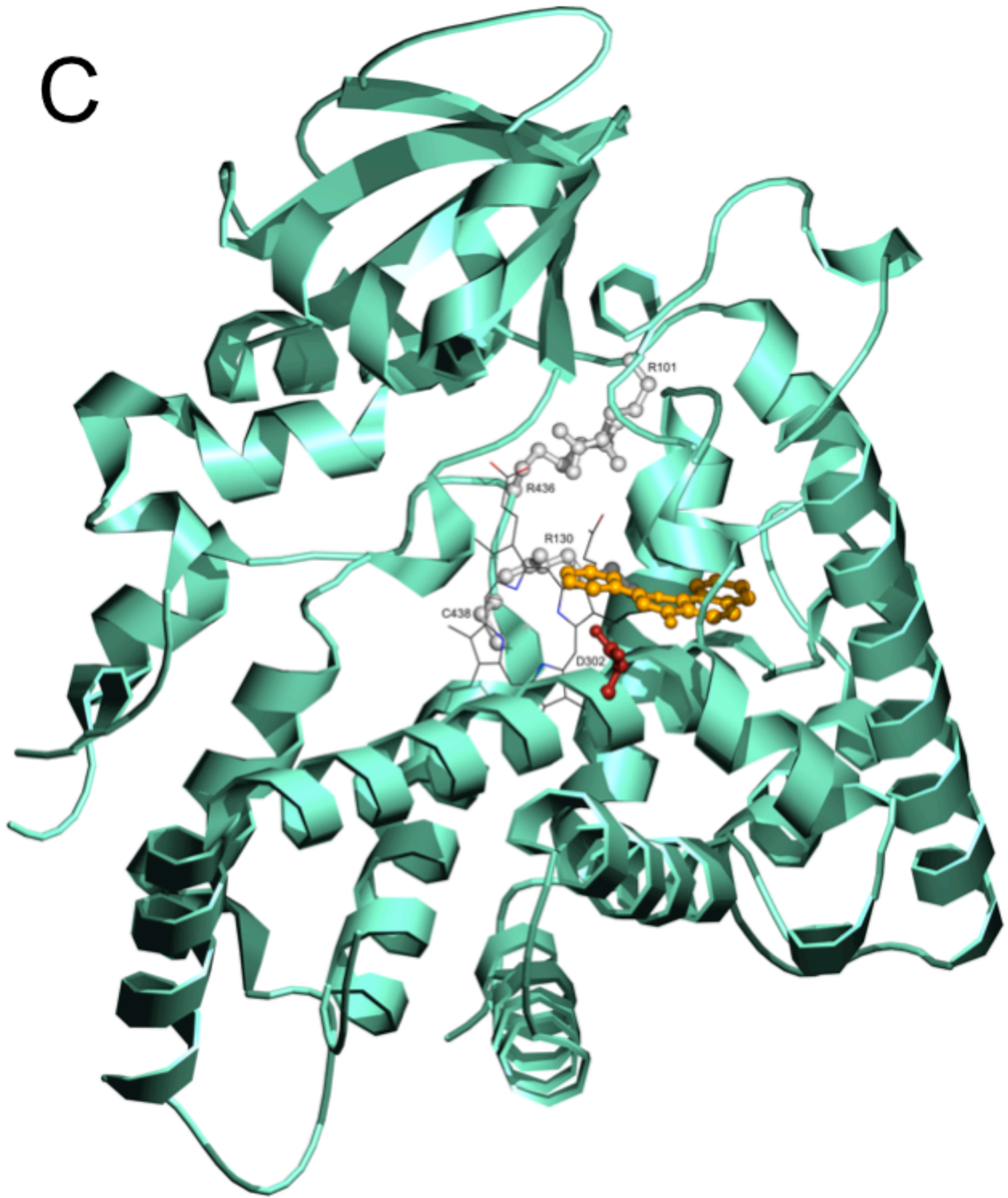
A


R104
R445
R133
C447
D309

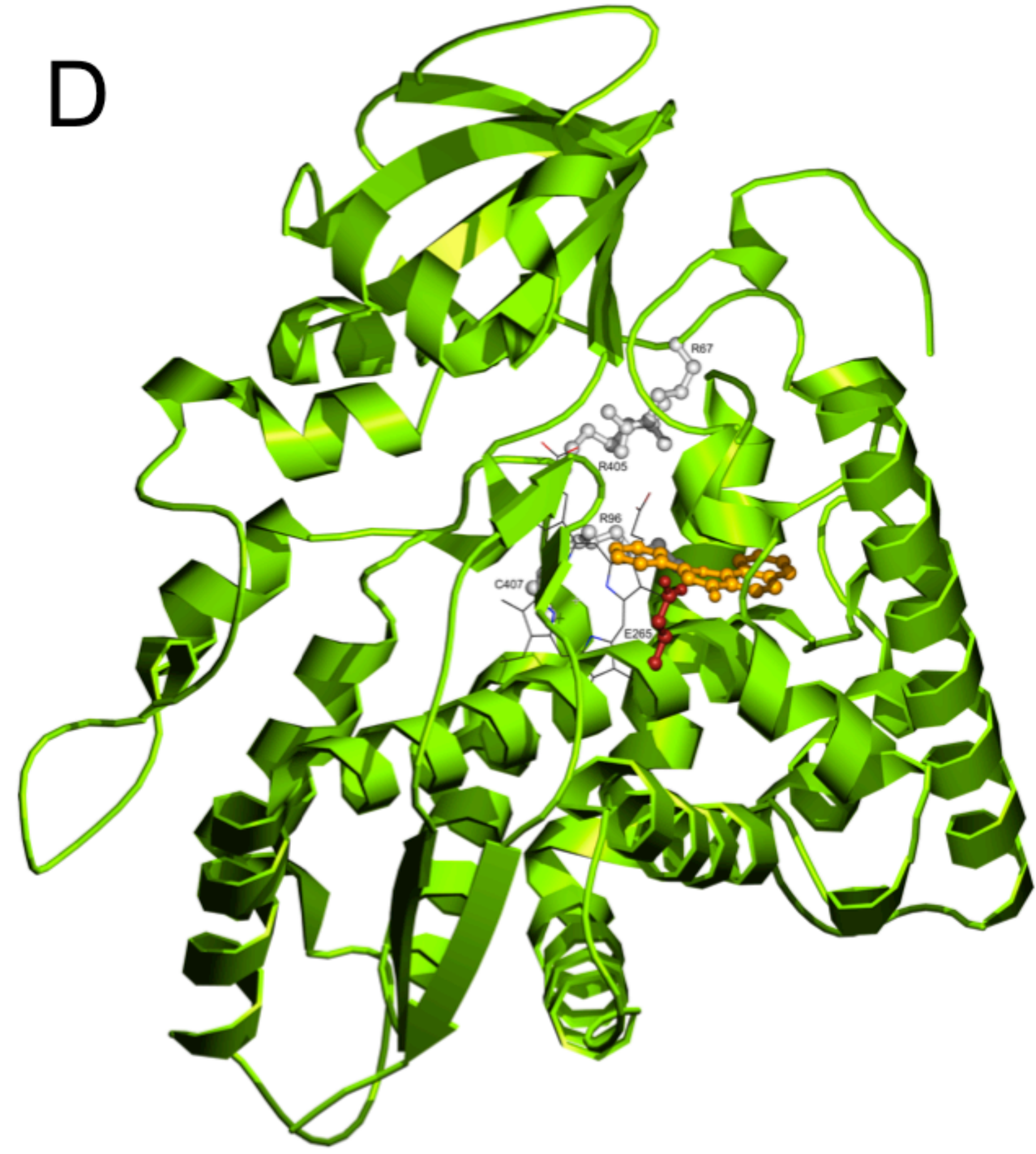
B

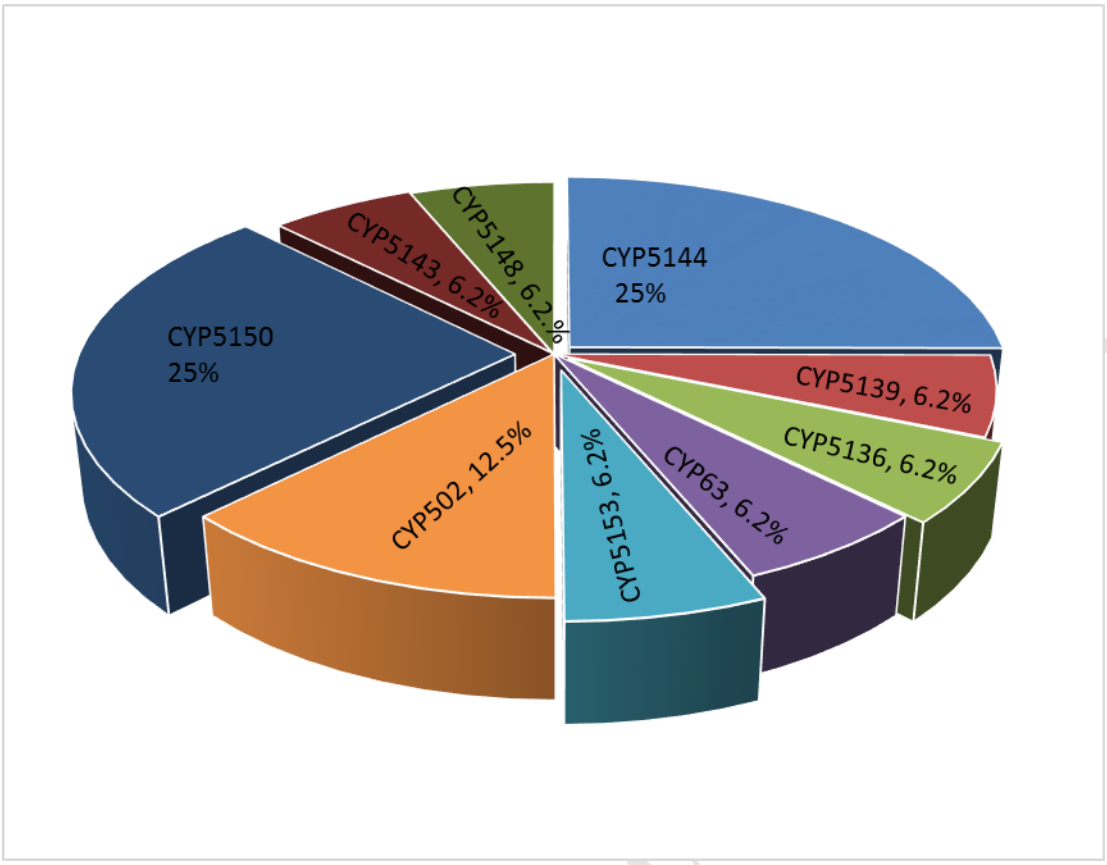

R101
R437
R130
C439
D303

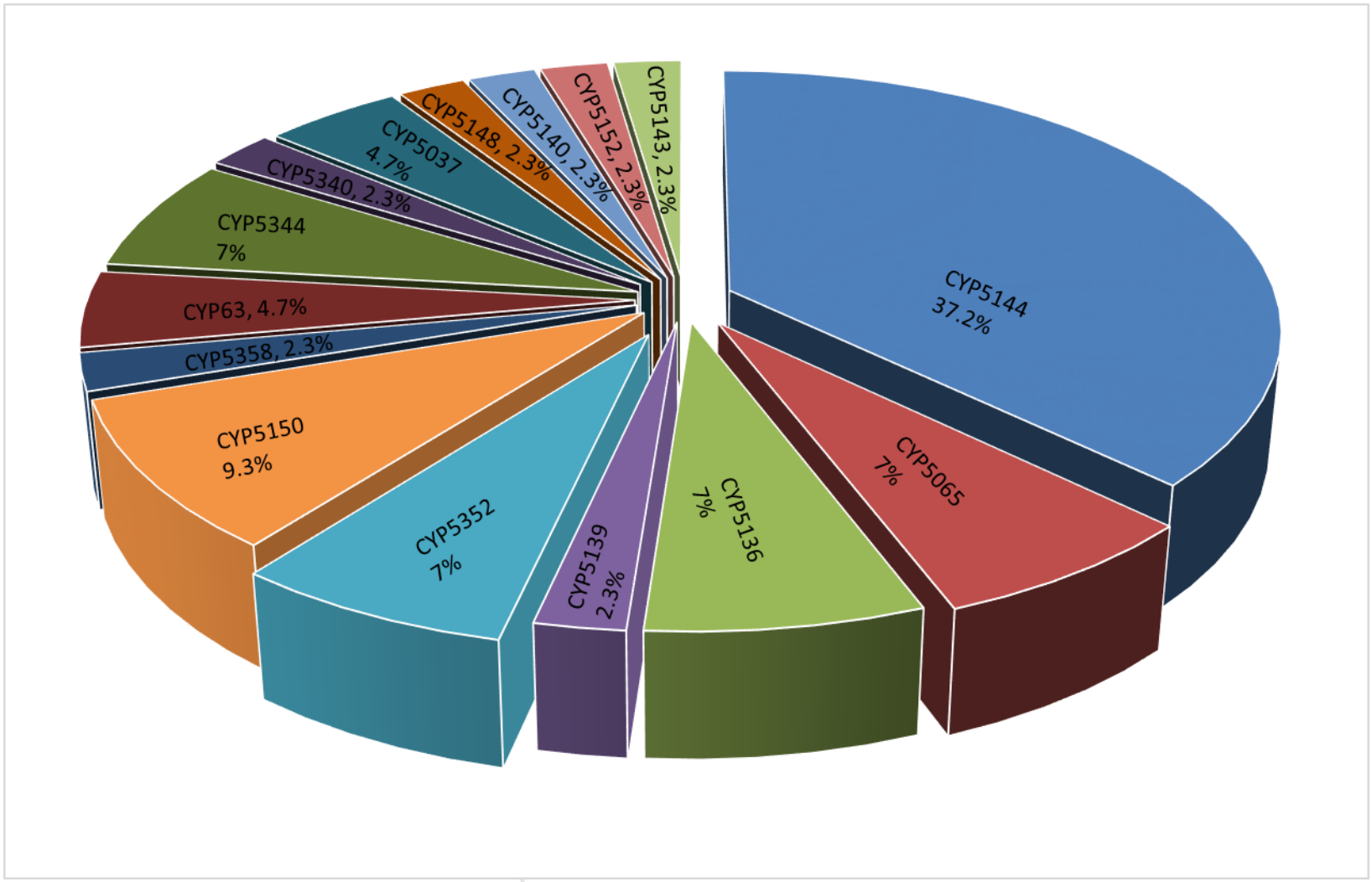
C


R101
R436
R130
C438
D302

D


R67
R405
R96
C407
E265




CYP5144
37.2%
CYP5065
7%
CYP5136
7%
CYP5139
2.3%
CYP5352
7%
CYP5150
9.3%
CYP5358, 2.3%
CYP63, 4.7%
CYP5344
7%
CYP5340, 2.3%
CYP5037
4.7%
CYP5148, 2.3%
CYP5140, 2.3%
CYP5152, 2.3%
CYP5143, 2.3%

Research Highlights

- *H. irregulare* genome has 122 CYP genes.
- The CYPs were classified into 11 clans, 35 families and 64 subfamilies.
- The largest cluster was on scaffold 5 in subfamily M (CYP5144).
- Microarray analysis identified 130 transcripts of P450 encoding genes.
- The *H. irregulare* CYPs showed different expression patterns.