Genome-Wide Analysis of the Cytochromes P450 Gene Family in Cordyceps militaris

Gongbo Lv\textsuperscript{a}, Xinping Liu\textsuperscript{b}, Bin Zeng\textsuperscript{a}, Bin He\textsuperscript{c}

College of Life Science Jiangxi S & T Normal University, Nanchang, Jiangxi, China

\textsuperscript{*}Corresponding author e-mail: Zengtx001@aliyun.com, \textsuperscript{a}Lv1367613373@163.com, \textsuperscript{b}nckdliuxp@sina.com, \textsuperscript{c}hebin.li@foxmail.com

Abstract. Cytochromes P450 (CYP450) gene family has been shown to play significant roles in various physiological processes of some species, including participate in the metabolism of various exogenous substances, synthesis of endogenous substances, abiotic and biotic stress responses and stress signaling. In this study, the members of CYP450 gene family in Cordyceps militaris (C.militaris) were identified and analyzed on the whole genome level using bioinformatics-based methods. The phylogenetic tree, gene structure, motifs and gene expression level under pH stress were analyzed. In total, 54 putative CYP450 gene family members were identified in C.militaris. The phylogenetic analysis indicated that all of the C.militaris CYP450 (CmCYP) genes, except CmCYP19, were grouped into 3 groups. The 54 CmCYP genes were randomly distributed on 7 chromosomes. The expression analysis of CmCYP under different pH environments showed the mechanism of CmCYP genes family involved in stress resistance regulation is complex and diversified. Systematic investigation of a gene family, would provide important insights into gene function and application. The results of this study will provide reference information for studying other abiotic and biotic stress response regulation mechanism of C.militaris.

1. Introduction

CYP450 is a heme protein superfamily terminal monooxygenase, or mixed-function oxidases, which were presented in all forms of life (plants, bacteria, mammals even in viruses) \cite{1}. It plays a key role in the oxidative transformation of endogenous and exogenous molecules \cite{2}. The name P450 refers to the belief that the reduced CYP450 has the largest ultraviolet absorption peak at 450nm after it is combined with CO \cite{3, 4}. The location and state of CYP450 are different in prokaryotic and eukaryotic cells. In prokaryotic cells, CYP450 is a soluble protein mainly distributed in the cytoplasm, while CYP450 is mainly distributed on microsomes, endoplasmic reticulum, and mitochondrial inner membrane in eukaryotic cells. In addition, studies have shown that the family of the CYP450 gene is characterized by a preserved heme-binding region in its structure. On the near surface of heme, the hallmark sequence of CYP450 is the sequence FXXGXXXCG/A and highly conserved Cys \cite{5}. In 1958, Kingenberg \cite{6} reported the first CYP450 in mammalian liver microsomes. The development and maturity of next generation sequencing has highly facilitated the identification, classification, and expression analysis of gene families that have been generated by evolutional divergence from a single
ancestral. There were 57 cytochrome P450s in humans, 15 cytochrome P450s in *Helicoverpa armigera* while the plant *Arabidopsis thaliana* has 286 cytochrome P450s [7].

*Cordyceps militaris* (*C.militaris*), one of the most important traditional Chinese medicines, is gaining more and more attention due to its diverse bioactive ingredients. Modern chemical and pharmacological studies showed that *C.militaris* contains a variety of chemical components, such as cordycepin, cordyceps polysaccharide, cordycepic acid, protein, amino acid [8], nucleosides [9], sterols [10], etc. Therefore, it has immune regulation, anti-bacterial [11], hypoglycaemic, anti-oxidant, antimalarial, anti-fatigue, anti-inflammatory, anti-tumor, anti-infective and other pharmacological effects [12]. Based on the above, *C.militaris* has become a hot spot for scientists with a variety of biological activities, broad application prospects and microscopic fields. In addition, *C.militaris* is considered as a substitution for *Cordyceps sinensis* due to the same bioactive ingredients and pharmacological effects [13]. *C.militaris* is often subjected to some biotic or abiotic stress during the cultivation process, such as pH press. However, there are few reports on related genes of *C.militaris* in response to pH stress.

Here, we used bioinformatics methods to identify CYP450 genes from *C.militaris* genome, and analyzed the sequence features, chromosomal locations, phylogenetic relationships and gene expression levels under pH stress. The purpose of this study is to obtain comprehensive information of the CYP450 genes family of *C.militaris*. The results would provide basis for revealing the mechanism of cytochrome family genes in *C.militaris* in response to pH stress.

2. Materials and Methods

2.1. Identification of CYP450 Gene Family members in *C. militaris*

All of the CYP450 protein sequences of *C.militaris*, including names CmCYP1-54, and the whole genome sequence of CYP450 gene were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/). Then, the whole genome sequence of *C.militaris* was constructed into a native BLAST database using ncbi-blast-2.2.31+ [14]. To avoid missing probable CYP450 gene family members in *C.militaris* (CmCYP) because of incomplete conserved CYP450 protein domain, a BLASTP-algorithm based search using Arabidopsis P450 amino acid sequences as queries was conducted with an e-value ≤1e−5 [15]. Additionally, keywords “P450” were employed to search against BLAST database. Finally, after removing all of the redundant sequences, the output the amino acid sequence of the candidate CYP450 gene family in *C.militaris* were submitted to CDD (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi), PFAM (http://pfam.xfam.org/) and SMART (http://smart.embl-heidelberg.de/) by a conserved CYP450 protein domain to determine CYP450 gene family members of *C.militaris*. All of the non-redundant genes were assigned as CmCYP. These CmCYP genes were named on the basis of their positions on chromosomes.

2.2. Multiple Sequence Alignment, Phylogenetic tree Construction and Classification of CmCYP genes

All CmCYP protein sequences were subjected to multiple sequence alignment using the Muscle of MEGA5 with parameters as default [16]. The phylogenetic tree was constructed using Maximum Likelihood (ML) and Neighbor-Joining (NJ). The phylogenetic tree was evaluated by bootstrap method with replicate value set as 1000 in order to choose the best. The phylogenetic tree was visualized by Figtree program (v. 1.4.3).

2.3. Motif analysis and Chromosomal localization of CmCYP genes

Motif is a secondary structural polymer that has a specific function of a protein molecule or is similar to a part of an independent domain. Also, motif as the binding site of transcription factors is of great significance for studying the transcriptional expression of genes. To obtain insights into the diversity of motif compositions in CmCYP proteins, putative motifs were analyzed by MEME software (http://memesuite.org/tools/meme, V 4.11.2). The parameters were set as zero or one occurrence (of a contributing motif site) per sequence; the numbers of motif were chosen six motifs; the motif width
was set 6 to 50. The results are visualized by TBtools software. To determine the chromosomal locations of all CYP450 genes in *C. militaris*, the information of locus coordinates was obtained from the genome sequences. Map Chart software [17] was used for the mapping of CmCYP genes’ chromosomal positions and relative distances.

2.4. Expression analysis of CYP450 gene family in *C. militaris* under different pH stress

In order to study the function of CYP450 gene further, HemI [18] software was used to generate the heatmap, mainly by mapped the transcriptome data of *C. militaris* at different pH onto the identified 54 candidate CYP450 gene family members and the final result was converted to FPKM (Fragments Per Kilobase of transcript per Million fragments) value to measure the amount of gene expression. The full data sets have been submitted to NCBI SRA databases under Accession PRJNA576297, BioSample: from SAMN12989487 to SAMN12989493.

3. Results

3.1. Genome-Wide Identification of CYP450 Gene Family in *C. militaris*

| Names     | Accession number in NCBI | Protein length(aa) | Names     | Accession number in NCBI | Protein length(aa) |
|-----------|--------------------------|--------------------|-----------|--------------------------|--------------------|
| CmCYP1    | ATY67502                 | 519                | CmCYP28   | ATY60276                 | 517                |
| CmCYP2    | ATY67406                 | 529                | CmCYP29   | ATY59929                 | 516                |
| CmCYP3    | ATY66410                 | 511                | CmCYP30   | ATY61323                 | 786                |
| CmCYP4    | ATY66855                 | 550                | CmCYP31   | ATY60304                 | 530                |
| CmCYP5    | ATY66817                 | 600                | CmCYP32   | ATY60306                 | 526                |
| CmCYP6    | ATY65842                 | 820                | CmCYP33   | ATY60298                 | 487                |
| CmCYP7    | ATY65841                 | 508                | CmCYP34   | ATY61100                 | 795                |
| CmCYP8    | ATY65919                 | 523                | CmCYP35   | ATY61047                 | 1066               |
| CmCYP9    | ATY65879                 | 563                | CmCYP36   | ATY60876                 | 534                |
| CmCYP10   | ATY65882                 | 555                | CmCYP37   | ATY60716                 | 519                |
| CmCYP11   | ATY66050                 | 513                | CmCYP38   | ATY61019                 | 456                |
| CmCYP12   | ATY59283                 | 517                | CmCYP39   | ATY61066                 | 531                |
| CmCYP13   | ATY58633                 | 569                | CmCYP40   | ATY60896                 | 566                |
| CmCYP14   | ATY58249                 | 624                | CmCYP41   | ATY60012                 | 507                |
| CmCYP15   | ATY58569                 | 488                | CmCYP42   | ATY60052                 | 516                |
| CmCYP16   | ATY58321                 | 489                | CmCYP43   | ATY60336                 | 526                |
| CmCYP17   | ATY59094                 | 554                | CmCYP44   | ATY60394                 | 801                |
| CmCYP18   | ATY64267                 | 549                | CmCYP45   | ATY61295                 | 521                |
| CmCYP19   | ATY64628                 | 523                | CmCYP46   | ATY60728                 | 509                |
| CmCYP20   | ATY63901                 | 527                | CmCYP47   | ATY61956                 | 1483               |
| CmCYP21   | ATY64296                 | 538                | CmCYP48   | ATY63302                 | 559                |
| CmCYP22   | ATY65337                 | 560                | CmCYP49   | ATY63239                 | 540                |
| CmCYP23   | ATY64412                 | 401                | CmCYP50   | ATY62011                 | 521                |
| CmCYP24   | ATY64872                 | 638                | CmCYP51   | ATY62537                 | 629                |
| CmCYP25   | ATY64778                 | 526                | CmCYP52   | ATY61879                 | 539                |
| CmCYP26   | ATY61214                 | 540                | CmCYP53   | ATY63525                 | 506                |
| CmCYP27   | ATY61274                 | 527                | CmCYP54   | ATY63732                 | 508                |

A total of 60 CYP450 gene family members were identified in *C. militaris* by Blast homology alignment. After removing the repetitive sequences, 54 sequences were reserved and submitted to

3
CDD, Pfam and SMART to confirm the conserved CYP protein domain. Finally, 54 sequences were confirmed as *C.militaris* CYP450 gene and named based on their chromosomal locations. Gene names, accession number in NCBI, protein length, were listed in TABLE I. The functional domain verifies of CmCYP1-CmCYP54 indicating that the 54 family members all contained p450 domain. The analysis of protein biochemical properties showed that the longest CmCYP protein in *C.militaris* was CmCYP47 containing 1483 amino acid residues and the shortest was CmCYP23 containing 401 amino acid residues.

### 3.2. Evolutionary Analysis

In order to better understand the genetic relationship among members of CYP450 gene family in *C.militaris*, phylogenetic tree was constructed based on the results of multiple sequences alignment. The *C.militaris* CYP450 genes were classified into different groups according to the topology of phylogenetic tree. As shown in Fig. 1, the 54 members in *C.militaris* were clustered into 3 groups (Group A, B and C), except for CmCYP19. In the three groups, near half of CmCYP gene existed on Group B (24/54), and the least is group A, only containing 13 CmCYP genes. In addition, Group A appeared to be distinct from the Group C which grouped closely with Group B.

![Phylogenetic tree of C. militaris CYP450 proteins](image)

**Table 1.** The phylogenetic relationship in *C. militaris* CYP450 proteins. The maximum likelihood tree was constructed by the MEGA program with 1000 bootstrap replications. The 3 subfamilies were distinguished in different colors, and the unclassified CmCYP19 were colored in black.
3.3. Motif Analysis

Figure 2. Conserved motifs of the CYP450 gene family in C. militaris.
The motifs were identified by MEME software based on the maximum expectation (EM) algorithm, which alternates the execution of steps E (expectation) and M (maximum). A total of 6 conserved motifs were identified in the CYP450 gene family of *C. militaris*. The relative location of these motifs within the protein is represented in Fig. 2. CmCYP3, CmCYP14, CmCYP27 and CmCYP38 comprised motif1-4 except for motif5 and motif6. CmCYP6 and CmCYP24 contained motif1-3 and motif6, CmCYP15 contained motif2-6, CmCYP32 contained motif1, 3, 4, 6 while CmCYP12 including motif1, 2, 4, 6. CmCYP1, CmCYP4, CmCYP8, CmCYP11, CmCYP13, CmCYP16, CmCYP19, CmCYP26, CmCYP29, CmCYP30, CmCYP31, CmCYP33, CmCYP34, CmCYP37, CmCYP45, CmCYP48, CmCYP52, CmCYP53 and CmCYP54 contained all motifs. The other 25 CmCYP genes contained motif1-4 and motif 6 while CmCYP23 only contain motif2 and motif4. This result further confirmed the result of evolutionary analysis. The results of frequency analysis of codon usage in each motif showed that the codon usage in motif 6 were significantly distinct in different genes and the codon usage in motif 1, 2 and 5 are relatively conservative (Fig. 3). Combination of multiple sequence alignment and pfam analysis showed that the CYP450 gene family was conservative.

3.4. The chromosomal positions of CmCYP genes
The CYP450 genes located in this research are all non-repetitive genes. MapChart software was used to locate 54 CYP450 family genes of *C. militaris* on 7 chromosomes. It can be seen from Fig. 4 that the 54 CYP450 genes of *C. militaris* are distributed on 7 chromosomes of the whole genome randomly and unevenly. The majority of CmCYP genes were located on the proximate or the distal ends of the chromosomes. Among them chromosomes VI harbored the most (21 of 54) CmCYP genes and densely distributed whereas only two CmCYP gene existed on chromosome I.
Figure 3. Frequency analysis of codon usage in each motif. Six conserved functional amino acid residues in each P-loop are shown from top to bottom. Height of letter displays relative frequency of each amino acid residue; Abscissa and ordinate demonstrate the number of amino acid residues and relative frequency of each amino acid residue, respectively.
3.5. Expression Analysis of CYP450 Gene Family in C. militaris under pH stress

To further explore the expression changes in the CmCYP genes under abiotic stresses, such as heat, pH and salt, we investigated the expression patterns of CmCYP gene family under pH stress (cultivated in PDA medium supplied with phosphate buffer with pH of 5, 5.5, 6, 7, 8, 8.5 and 9 respectively). According to the transcriptome data, heatmap of 54 CYP450 genes, represented by FPKM values in different pH, was established by Heml shown in Fig. 5. From the results, we found that most of CYP450 genes were highly expressed in pH of 5 and 9, especially in pH of 5, which shows that these genes are sensitive to acids. On the contrary, lots of CYP450 genes were with extremely low expression in pH of 8 and 8.5, it indicates that the expression of these genes is inhibited when the pH is 8 and 8.5. In other pH, such as pH of 6 to 8, only CmCYP19, CmCYP40, CmCYP43, CmCYP9, CmCYP42, CmCYP54 and CmCYP52 showed very high expression levels, the rest of
genes are either not expressed or down-regulated. Interestingly, No gene is up- or down-regulated at all pH. The altered expression pattern under different pH stress suggested that the mechanism of CmCYP genes family involved in stress resistance regulation is complex and diversified.

Figure 5. Expression profile of CYP450 gene at different pH. The heatmap was constructed by HemI software. CMs_5, 5.5, 6, 7, 8, 8.5, and 9 indicated at pH 5-9, respectively.
4. Discussion
Cytochrome P450, an important heme protein monooxygenase, were existed in all forms of life (plants, bacteria, mammals even in viruses). In recent years, with the maturity of redox partner engineering, metabolic engineering, and synthetic biology, it has become possible to obtain the P450 biocatalysts with the desired properties to meet the industrial requirements, via rational design and direct evolution of P450 enzymes. Hence, it has been expanded greatly the application scope of P450 enzymes in biosynthesis.

CYP450 enzymes act as biocatalysts in nature, can catalyse multiple reaction types, including hydroxylation reaction, dealkylation reaction, functional group migration reaction, and anti-Markovnikov addition, etc., and mainly involved in two major functions of biosynthesis and bio detoxification [19]. For instance, it can catalyze various metabolic reactions in plants and involve in the biosynthesis of some biologically active substances. Human CYPs are primarily membrane-associated proteins located either in the inner membrane of mitochondria or in the endoplasmic reticulum of cells. CYP450 in animal cells mainly plays the role of oxidative metabolism, including the metabolism of endogenous substances (fatty acids, etc.), and exogenous substances (environmental pollutants, etc.). CYP450 in microorganisms can be utilized to degrade organic substances. This process aims to reduce from oxygen molecules to oxygen atoms, is mainly via the catalysis of CYP450, and using NADPH as a coenzyme. In bacteria, the distribution of P450s is very variable with many bacteria having no identified P450s (e.g. *Escherichia coli*). Therefore, *E. coli* can be used as a host for cloning and expression of exogenous CYP450 [20], which is conducive to detailed research on the function and properties of CYP450. CYP450 enzymes are involved in many complex fungal biotransformation processes. The commonly used azole class antifungal drugs work by inhibition of the fungal cytochrome P450 14α-demethylase. It interrupts the conversion of lanosterol to ergosterol, a component of the fungal cell membrane [21]. From the above, the function of CYP450 enzymes is relatively diverse within the family and among various organisms. Furthermore, due to the differences in structural characteristics, substrate types, catalytic reactions, and distribution, it has a wide range of applications in many fields such as medicine, food, biological control, environmental protection, and agriculture [19].

With the development of genomics, the identification, classification and naming of different CYP450 gene families has been greatly facilitated. It has become possible to regulate and modify the structure and expression activity of CYP450. Nowadays, many CYP450 genes were successfully identified, such as *Arabidopsis*, *Populus* and grapes [22]. However, little is known about CYP450 gene family in *C. militaris*.

The current study identified 54 CmCYP genes, and analyzed their structure, chromosomal location, phylogeny, and motif. Moreover, we explored the expression level for all putative CmCYP genes in different pH stresses (at pH 5, 5.5, 6, 7, 8, 8.5 and 9). From the results, we found that CmCYP genes involved in the response of pH stress, which may influence the stress resistance regulation mechanism of *C. militaris*.

Acknowledgments
This article is sponsored by NSFC (Grant NO. 31460447 and31171731) and Doctor and master specific projects of Honghe University (XJ17B09).

References
[1] S. Kanamura, and J. Watanabe, “Cell biology of cytochrome P450 in the liver,” Int Rev Cytol, vol. 198, 2000.
[2] J. Shin, J. E. Kim, Y. W. Lee, and H. Son, “Fungal Cytochrome P450s and the P450 Complement (CYPome) of Fusarium graminearum,” Toxins (Basel), vol. 10, no. 3, Mar 7, 2018.
[3] R. W. Estabrook, “The remarkable P450s: a historical overview of these versatile hemeprotein catalysts,” FASEB J, 1996.
[4] I. G. Denisov, T. M. Makris, S. G. Sligar, and I. Schlichting, “Structure and Chemistry of Cytochrome P450,” Chem. Rev, vol. 105, pp. 2253 – 2277, 2005.

[5] C.A. Hasemann, K.G. Ravichandran, J.A. Peterson, and J. Deisenhofer, “Crystal structure and refinement of cytochrome P450terp at 2.3 A resolution,” J Mol Biol, vol. 236, no. 4, Mar 4, 1994.

[6] M. Klingenberg, “Pigments of rat liver microsomes,” Arch Biochem Biophys, vol. 75, pp. 376-386, 1958.

[7] B. Meunier, S. l. P. d. Visser, and S. Shaik, “Mechanism of Oxidation Reactions Catalyzed by Cytochrome P450 Enzymes,” Chem. Rev, pp. 3947–3980, 2004.

[8] J. P. LV, Y. ZHANG, M. J. YAO, H. ZHANG, and J.M. SUN, “Chemical constituents and pharmacological effects of C.militaris,” Jilin Journal of Traditional Chinese Medicine, pp. 38310-312, 2018.

[9] B. J. Ma, Y. Ruan, and J.K. Liu, “Chemical study of nucleosides on C.militaris,” Chin. Med. Mater, vol. 8, pp. 957 – 958, 2007.

[10] Z.M. Lv, Y.T. Jiang, L. J. Wu, and K. Liu, “Chemical constituents from dried sorophore of cultured C.militaris,” Chin. Mater. Med, vol. 33, pp. 2914-2917, 2008.

[11] B. T. Park, K. H. Na, E. C. Jung, J. W. Park, and H. H. Kim, “Antifungal and Anticancer Activities of a Protein from the Mushroom Cordyceps militaris,” Korean Journal of Physiology & Pharmacology, vol. 13, no. 1, pp. 49 - 54, Feb, 2009.

[12] J. M. Park, J. S. Lee, K. R. Lee, S. J. Ha, and E. K. Hong, “Cordyceps militaris extract protects human dermal fibroblasts against oxidative stress-induced apoptosis and premature senescence,” Nutrients, vol. 6, no. 9, pp. 3711-26, Sep 16, 2014.

[13] W. S. Jo, Y. J. Choi, H. J. Kim, J. Y. Lee, B. H. Nam, J. D. Lee et al., “The Anti-inflammatory Effects of Water Extract from Cordyceps militaris in Murine Macrophage,” Mycobiology, vol. 38, no. 1, pp. 46 - 51, Mar, 2010.

[14] B. Fröhlich, and J. Plate, “The cubic mouse: a new device for three-dimensional input. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,” (The Hague, The Netherlands, April 01-06, 2000). CHI ’00 . ACM, New York, NY, 526 - 531.

[15] P. Zhao, D. Wang, R. Wang, N. Kong, C. Zhang, C. Yang et al., “Genome-wide analysis of the potato Hsp20 gene family: identification, genomic organization and expression profiles in response to heat stress,” BMC Genomics, vol. 19, no. 1, pp. 61, Jan 18, 2018.

[16] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, “MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods,” Mol Biol Evol, vol. 28, OCT, 2011.

[17] R.E. Voorrips, “Mapchart: software for the graphical presentation of linkage maps and qtls,” J Hered, vol. 93, 2002.

[18] W. K. Deng, Y. B. Wang, Z.X. Liu, H. Cheng, and Y. Xue, “Hemi: a toolkit for illustrating heatmaps,” PLoS One, vol. 9, Nov 5, 2014.

[19] X. F. Leng, X. H. Qiu, “Structure, function and application prospect of cytochrome P450 enzyme system,” Beijing: Sci-ence Press. 2001.

[20] B. W. Biggs, C. G. Lim, K. Sagliani, S. Shankar, G. Stephanopoulos, M. De Mey et al., “Overcoming heterologous protein interdependency to optimize P450-mediated Taxol precursor synthesis in Escherichia coli,” Proc Natl Acad Sci U S A, vol. 113, no. 12, pp. 3209-14, Mar 22, 2016.

[21] H. M. Girvan, T. N. Waltham, and R. Neeli, et al., “Flavocytochrome P450 BM3 and the origin of CYP102 fusion species,” Biochem Soc T, vol. 34, pp. 1173 - 1177, JUN, 2006.

[22] D. R. Nelson, R. Ming, M. Alam, and M.A. Schuler, “Comparison of Cytochronme P450 Genes from Six Plant Ge-nomes,” Tropical Plant Biology I, pp. 216 - 235, 2008.