

茯苓细胞色素 P450 还原酶基因的克隆与生物信息学分析

何海, 郭继云, 舒少华, 王沫*

华中农业大学 药用植物研究所, 湖北 武汉 430070

摘要:目的 从茯苓 *Poria cocos* 中克隆出细胞色素 P450 还原酶 (CPR) 基因, 并对其进行生物信息学分析。方法 利用茯苓转录组注释信息, 通过 RACE 扩增得到茯苓 CPR 基因 (PcCPR) 全长, 并 PCR 得到对应基因组 DNA 序列。通过生物信息学方法, 对该基因编码蛋白的特征进行分析, I-TASSER 模拟出蛋白三级结构, MEGA 构建蛋白系统进化树。结果 获得含有完整开放阅读框 (ORF) 的 cDNA 全长 2 514 bp (GenBank 号为 KP768251), PcCPR 的基因组 DNA 5 292 bp (GenBank 号为 KP896487), 含 4 个外显子、3 个内含子。基因编码 732 个氨基酸的蛋白, 相对分子质量 81 147, 等电点 (pI) 为 5.39, 为不稳定性亲水蛋白, 非分泌蛋白, 且不具有信号识别功能。蛋白定位于内质网上, 第 7~22 位氨基酸为跨膜区; 具有 2 个黄素结合的保守结构域, FAD、NADP 的结合位点各自在蛋白三级结构空间上相邻近。同源性分析显示, PcCPR 与担子菌 CPR 的相似性明显高于子囊菌 CPR。结论 成功从茯苓中克隆得到 PcCPR 基因, 为进一步研究 PcCPR 及相关代谢提供基础。

关键词: 茯苓; 细胞色素 P450 还原酶; 基因克隆; 生物信息学分析; RACE

中图分类号: R282.12 文献标志码: A 文章编号: 0253-2670(2016)16-2909-07

DOI: 10.7501/j.issn.0253-2670.2016.16.023

Cloning and bioinformatics analysis of cytochrome P450 reductase gene in *Poria cocos*

HE Hai, GUO Ji-yun, SHU Shao-hua, WANG Mo

Institute of Medicinal Plant, Huazhong Agricultural University, Wuhan 430070, China

Abstract: Objective To clone cytochrome P450 reductase (PcCPR) gene from *Poria cocos* and to characterize with bioinformatics methods. **Methods** According to annotated transcriptome of *P. cocos*, the PcCPR gene was cloned through RACE, and the genomic DNA sequence was further obtained through PCR. The characteristics of the encoded protein were analyzed using bioinformatics, the 3D structure of the protein was modeled with I-TASSER server, and phylogenetic tree of CPR was carried out with MEGA. **Results** The 2 514 bp full-length cDNA sequence (GenBank Accession No. KP768251) and the 5 292 bp genomic DNA sequence (GenBank Accession No. KP896487) of PcCPR was obtained, which contained four exons and three introns. PcCPR encoded a protein with 732 amino acids. The protein was predicted to be an unstable hydrophilic protein with calculated molecular weight of 81 147 and isoelectric point 5.39. PcCPR does not have a signal peptide but has a transmembrane segment (aa residues 7 to 22), and it is anchored to endoplasmic reticulum. There are two flavin binding domains and many predicted FAD and NADP binding sites, these sites are adjacent respectively at 3D structure level. The homologous analysis indicates that PcCPR has a higher similarity with CPR from basidiomycetes than CPR from ascomycetes. **Conclusion** The PcCPR was successfully cloned, which will provide a foundation for researches on PcCPR and PcCPR associated metabolic.

Key words: *Poria cocos* (Schw.) Wolf; cytochrome P450 reductase; gene cloning; bioinformatics analysis; rapid-amplification of cDNA ends

细胞色素 P450 还原酶 (EC 1.6.2.4, cytochrome P450 reductase, 简称 CPR 或 POR) 是一种在许多代谢进程中必需的酶, 主要参与细胞色素 P450 氧化酶系 (cytochrome P450 oxidoreductase system,

P450s) 催化的初级及次级代谢反应, 如萜类物质代谢、药物代谢、外源化合物代谢等^[1-2]。在此氧化系统中, CPR 的功能是从底物 NADPH 获取电子, 通过 FMN 和 FAD 辅因子传递给细胞色素 P450, 氧化

收稿日期: 2015-09-16

基金项目: 国家“十二五”科技支撑计划项目 (2011BAI06B03)

作者简介: 何海 (1990—), 男, 在读硕士研究生, 研究方向为中药资源学。E-mail: he-hai@outlook.com

*通信作者 王沫, 男, 教授, 研究方向为药用植物学。E-mail: wangmo@mail.hzau.edu.cn

底物；CPR 与 P450s 间的电子传递是 P450s 氧化反应的限速步骤^[3]。此外，CPR 还能向细胞色素 b5、鲨烯单氧化酶、血红素加氧酶提供电子，能直接催化一些前体药物的单原子还原活化^[2,4]。

茯苓 *Poria cocos* (Schw.) Wolf 是一种腐生或寄生在松科植物赤松和马尾松根部的高等真菌，隶属于担子菌门 (Basidiomycetes) 多孔菌科 (Polyporaceae)^[5]。茯苓菌核是一种传统中药，约 10% 的中成药 (约 300 种) 原料配方中含有茯苓，也被用于食品添加剂、膳食、保健品、化妆品等^[6-7]。茯苓具有抗肿瘤、抗氧化、抗惊厥等多种生物活性，其主要化学成分为多糖及三萜类化合物^[5]。

P450 参与萜类物质代谢已有大量研究报告，如 CYP71AV1 参与青蒿素的生物合成^[8]、CYP88D6 与 CYP72A154 参与甘草酸的生物合成^[9-10]、CYP76AH1 参与丹参酮的生物合成^[11]、在灵芝中大量 CYP 基因被推测与灵芝三萜合成相关^[12]。此外，研究发现担子菌基因组中存在大量 CYP 基因，且数量较其他真菌多；另有认为棕腐菌比白腐菌拥有丰富的 CYP 基因，这都让 CYP 被认为与担子菌的生物学、生态、进化密切相关^[13-14]。

本实验室通过对茯苓转录组测序及分析，提出了茯苓中主要活性成分三萜骨架的生物合成途径，并注释出了 249 个 Unigene 为 CYP，且推测多个参与茯苓酸、土莫酸等三萜化合物的生物合成^[15]。本研究在上述研究基础之上，利用茯苓的高通量测序结果，挖掘出茯苓 CPR (PcCPR) 基因相关序列，通过 RACE (rapid-amplification of cDNA ends) 技术扩增得到 PcCPR 基因 cDNA 全长，PCR 得到对应基因组 DNA 序列。通过生物信息学方法分析 PcCPR 基因及其编码蛋白，为阐释 PcCPR 基因功能提供理论依据。从而为研究茯苓 CYP 基因功能乃至解析茯苓三萜的生物合成途径奠定基础。

1 材料与方法

1.1 材料

茯苓 *Poria cocos* (Schw.) Wolf 菌核于 2012 年采自湖北省英山县石头咀镇的栽培田间，并通过分离纯化得到纯培养。28 °C 下，在马铃薯葡萄糖培养基 (PDA) 上培养 7 d，收集得到菌丝体。

1.2 方法

1.2.1 核酸提取 茯苓菌丝总 RNA 的提取使用 RNAiso Plus (Takara, 大连) 试剂提取，提取方法依据其说明书进行。茯苓 DNA 的提取采用

CTAB 法。

1.2.2 PcCPR 基因的克隆 根据茯苓的转录组数据^[15]中注释为 CPR 基因的 Unigene 设计 RACE 引物 CPR-31、CPR-32、CPR-51、CPR-52 (表 1)。3'RACE 方法^[16]使用经典 3'RACE 方法，使用 PrimeScript II 1st strand cDNA synthesis 试剂盒 (Takara, 大连)，以 Q_T 引物对提取的茯苓总 RNA 进行反转录，得到第 1 链 cDNA 产物用 Tricine-EDTA 稀释 10 倍作为 PCR 模版。3'RACE 第 1 轮 PCR 反应体系：0.2 μmol/L dNTPs，0.25 μmol/L Q₀，0.25 μmol/L CPR-31，1 U Ex Taq 酶，1 μL cDNA 模版，20 μL 体系。PCR 程序为 94 °C、5 min 预变性；94 °C、30 s，67 °C、30 s，72 °C、2 min，每个循环退火温度降 1 °C，共 10 个循环；后为 25 个循环 94 °C、30 s，57 °C、30 s，72 °C、2 min 扩增；最后加 72 °C 10 min 延伸。巢式 PCR 第 2 轮使用 CPR-32、Q_I 引物，以 1 μL 稀释 20 倍的第 1 轮 PCR 产物为模版，从 66 °C 开始进行降落 PCR 反应。PCR 产物使用胶回收试剂盒 (Axygen, 杭州) 回收，TA 克隆连接至 pMD18-T 载体 (Takara, 大连)，转化 Trans1-T1 *E. coli* 感受态细胞 (全式金, 北京)。单克隆进行菌落 PCR 鉴定，阳性菌株测序。

5' RACE 利用 SMARTer RACR 5'/3' 试剂盒 (Clontech, 日本)，依据其说明书进行 5' RACE 的

表 1 实验中使用的引物
Table 1 Primers used in this study

名称	序列 (5'→3')
CPR-31	GAATGTGGAGGAGGGTCAGGG
CPR-32	ACGATGCAAAGAACCCGTACC
CPR-51	GATTACGCCAAGCTTGACGACCGCGACGGCGCTAATG
CPR-52	GATTACGCCAAGCTTCCGCATTGGACGGCCAAACACC
CPR-g1	CCATGCTGAATGGAGATTCTGAC
CPR-g2	GTTGATCCTGCGATCATTGTGG
CPR-1	ATGGCGTCGTCTTCGGACA
CPR-2	TCAGCTCCAGACATCGACA
Q _T ^a	CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAG- CTTTTTTTTTTTTTTTTT
Q ₀ ^a	CCAGTGAGCAGAGTGACG
Q _I ^a	GAGGACTCGAGCTCAAGC

^a 引物序列参考文献报道^[16]

^aPrimers' sequences from reference^[16]

反转录, 得到相应的第一链 cDNA。根据试剂盒说明书使用引物 CPR-51、CPR-52 分别进行第 1 轮 PCR 和巢式第 2 轮 PCR。PCR 产物胶回收, In-Fusion 克隆至 pRACE 载体中, 转化。菌落 PCR 验证单克隆, 阳性菌落测序。

根据 3' RACE 及 5' RACE 测序拼接结果设计引物 CPR-1、CPR-2, PCR 并测序验证。利用转录组数据, 在茯苓的基因组数据^[17]中找到 PcCPR 基因位置, 在其上下游设计引物 CPR-g1、CPR-g2。以提取的茯苓 DNA 为模版, 克隆 PcCPR 基因。PCR 体系: 在 GC buffer I 中加入 0.4 μmol/L dNTPs, 0.2 μmol/L CPR-g1, 0.2 μmol/L CPR-g2, 1.5 U LA Taq 酶, 1 μL DNA 模版, 共 25 μL 体系。反应程序: 95 °C 预变性 1 min; 98 °C 变性 20 s; 57 °C 退火 30 s; 68 °C 延伸 7 min, 且延伸时间每个循环加 10 s; 共 35 个循环反应, 后加 72 °C 反应 15 min。

1.2.3 PcCPR 基因的生物信息学分析 使用 DNAMAN 8 软件对 3' 和 5' RACE 得到的基因序列进行拼接, 得到基因全长 mRNA 序列。利用 NCBI BlastX 比对, 以确定得到的为 PcCPR 基因, 并得到其开放阅读框 (ORF)。利用 DNAMAN 8 翻译得到 PcCPR 蛋白质氨基酸序列。从 GenBank 下载真菌 CPR 蛋白序列, 使用 MAFFT v7, 在默认参数下进行序列多重比对, 比对结果使用 MEGA 6 进行系统发育分析。使用邻接法 (Neighbor-Joining, NJ), Poisson 模型构建系统发育树, 建树可靠性使用 Bootstrap, 重复 1 000 次进行验证。

利用 GSDS 2.0 (Gene Structure Display Server, <http://gsds.cbi.pku.edu.cn/>) 在线分析 PcCPR 的基因结构。使用 ExPASy-ProtParam tool (<http://web.expasy.org/protparam/>) 分析 PcCPR 蛋白的基本理化性质, 应用 ExPASy-ProtScale (<http://web.expasy.org/protscale/>) 分析蛋白质氨基酸的亲疏水性。PcCPR 蛋白质信号肽预测采用 SignalP 4.1 server (<http://www.cbs.dtu.dk/services/SignalP/>); 跨膜区预测采用 TMHMM Server v. 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>); 蛋白亚细胞定位预测则采用 ProtComp v. 9.0 (<http://linux1.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc>) 进行。应用 I-TASSER (<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>) 进行蛋白质三级模拟。应用 NCBI 的 CDD (Conserved Domain Database) 数据库 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb>。

cgi) 进行蛋白质保守结构域分析。

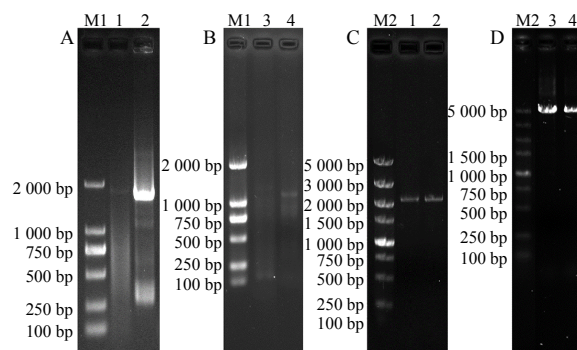
2 结果与分析

2.1 PcCPR 基因的克隆

利用茯苓转录组数据中注释为 cytochrome P450 oxidoreductase 的 Unigene 序列直接设计引物进行 3' RACE 及 5' RACE, 结果见图 1。3' RACE 得到 1 647 bp 含有 17 bp Poly (A) 尾的序列; 5' RACE 得到 1 002 bp 序列。使用 DNAMAN 拼接得到 2 514 bp 的基因全长 mRNA 序列, 并通过基因全长序列 PCR 测序验证正确。通过 Blast 确认所得为 CPR 基因序列, 并将其命名为 PcCPR。利用分离自美国佛罗里达州阿拉楚瓦郡的茯苓单孢培养菌株 MD-104 SS10 的基因组测序结果设计的引物扩增得到 5 292 bp 的 PcCPR 基因的基因组 DNA 序列 (图 1-D)。将 PcCPR 基因的 mRNA 及基因组 DNA 序列提交至 Genbank, 获得 Genbank 登录号分别为 KP768251、KP896487。利用 GSDS 2.0 在线工具分析 PcCPR 基因结构, 如图 2 所示, 其含有 4 个外显子、3 个内含子。PcCPR 基因含 2 199 bp 的编码区 (CDS), 编码 732 个氨基酸; 含有 80 bp 的 5' 非编码区 (5' UTR)、218 bp 的 3' 非编码区 (3' UTR)。

2.2 PcCPR 蛋白基本性质预测

利用克隆得到的 PcCPR mRNA 序列, 翻译得到 PcCPR 的氨基酸序列。使用 ExPASy-ProtParam tool 预测 PcCPR 蛋白的性质, 见表 2。预测该蛋白为



A-PcCPR 基因 3' RACE B-PcCPR 基因 5' RACE C-PcCPR 基因全长扩增 D-PcCPR 基因组 DNA 扩增 M1、M2-Marker, 1、2-3' RACE 第 1、2 轮产物, 3、4-5' RACE 第 1、2 轮产物 A-3' RACE product of PcCPR B-5' RACE product of PcCPR C-CDS amplification of PcCPR D-genomic DNA amplification M1, M2-Marker 1, 2-first and second round product of 3' RACE 3, 4-first and second round product of 5' RACE

图 1 PcCPR 基因扩增产物电泳结果

Fig. 1 Agarose gel electrophoretic analysis of PCR products of PcCPR gene



图 2 PcCPR 基因结构

Fig. 2 Gene structure of PcCPR gene

相对分子质量 81 147 的蛋白，等电点 (pI) 为 5.39，不稳定系数 (II) 40.02，平均疏水性 (grand average of hydropathy, GRAVY) 为 -0.337。根据工具定义，不稳定系数 (Instability index, II) 值大于 40 时，蛋白预测为不稳定，小于 40 时为稳定蛋白；GRAVY 值范围为 -2~2，正值表明此蛋白为疏水蛋白，负值表明为亲水蛋白。因此，PcCPR 为不稳定的亲水蛋白。此外，利用 ProtScale 工具预测蛋白一级结构上的亲疏水性，使用默认的 Hphob./Kyte & Doolittle 标度打分，窗口大小 (window size) 为 9，线性加权模型 (linear weight variation model)。从预测结果 (图 3) 可知，PcCPR 在一级结构上，多数区域的打分为负，即表现亲水性，其结果与 ExPASy-ProtParam tool 预测结果一致。其中，第 11 位具有最高正值 3.122，疏水性最强；第 195 位具有最低负值 -2.967，亲水性最强。

表 2 PcCPR 理化性质预测

Table 2 Predicted physicochemical properties of PcCPR

理化指标	预测结果
氨基酸数/个	732
相对分子质量	81 147
pI	5.39
负电荷残基数 (Asp+Glu) /个	101
正电荷残基数 (Arg+Lys) /个	82
分子式	C ₃₆₁₅ H ₅₆₄₅ N ₉₇₇ O ₁₀₉₇ S ₂₅
II	40.02
GRAVY	-0.337
脂肪系数 (AI)	82.21

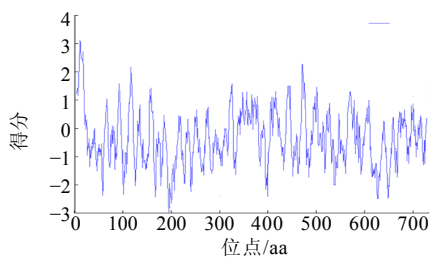


图 3 PcCPR 氨基酸亲水性/疏水性预测

Fig. 3 Predicted hydrophobicity/hydrophilicity of the amino acid sequence of PcCPR

2.3 PcCPR 信号肽预测

使用在线工具 SignalP 4.1 server 预测 PcCPR 的 N 端信号肽，使用默认参数设置，预测结果如图 4 所示。预测结果显示，在第 18 位氨基酸有最高的 C 分值 (原始剪切位点的分值, raw cleavage site score) 0.325，即蛋白可能在此位置被切割；而最大 S 分值 (0.627, 信号肽分值, signal peptide score) 在第 1 位氨基酸，蛋白第 1~17 位氨基酸的平均 S 分值为 0.387；几何平均分 C 值与 S 分值打分的斜率得到的 Y 分值 (combined cleavage site score) 最大在第 18 位氨基酸，为 0.342。对第 1~17 位氨基酸平均 S 分值及最大 Y 分值加权平均后得到的 D 分值 (discrimination score) 为 0.360，小于设定阈值 0.500。因此，预测 PcCPR 不含有信号肽，说明此蛋白为非分泌蛋白且不具有信号识别功能。

2.4 PcCPR 跨膜区预测

使用 TMHMM Server v. 2.0 预测 PcCPR 的跨膜区，结果 (图 5) 表明 PcCPR 有一段跨膜螺旋区，

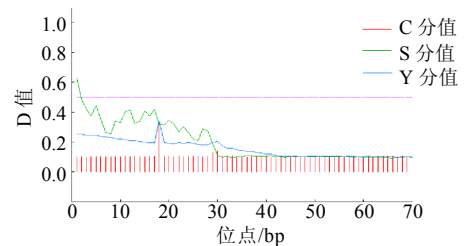


图 4 PcCPR 信号肽预测

Fig. 4 Signal peptide prediction of PcCPR

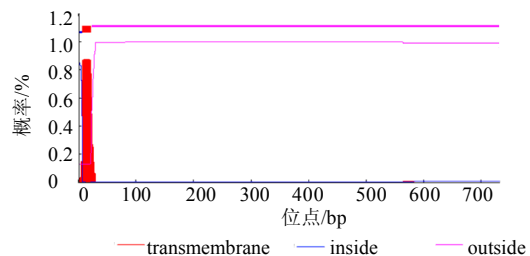


图 5 PcCPR 跨膜区预测

Fig. 5 Transmembrane helices prediction of PcCPR

即蛋白的第 7~22 位氨基酸。蛋白第 1~6 位氨基酸在膜内侧，第 23~732 位氨基酸在膜外侧。

2.5 PcCPR 亚细胞定位预测

ProtComp 9.0 是一种结合多种方法对蛋白进行亚细胞定位预测的软件。通过直接对比已知亚细胞定位的同源性蛋白，得出总分为 10 分的打分结果 LocDB；而对比数据库中具有有力理论推测定位的蛋白结果，则得到总分为 5 分的打分结果 (PotLocDB)；基于神经网络分析预测的打分则定义为总分 3 分的 Neural Nets 分值；基于序列与数据库五聚体分布对比计算打分结果得到总分 5 分的 Pentamers 分值；最后，对这 4 项打分进行加权后，得到总分为 10 分的最终结果 (Integral)。对 PcCPR 使用 ProtComp 9.0 进行亚细胞定位预测，结果如表 3 所示。预测结果表明，PcCPR 应定位于内质网上，其软件综合打分结果最高，为 9.90。

表 3 PcCPR 亚细胞定位预测

Table 3 Sub-cellular location prediction of PcCPR

亚细胞	分值				
	LocDB	PotLocDB	Neural Nets	Pentamers	Integral
细胞核	0.0	0.0	0.23	0.00	0.02
质膜	0.0	0.0	0.37	0.00	0.06
胞外	0.0	0.0	0.20	0.00	0.00
细胞质	0.0	0.0	0.31	0.00	0.01
线粒体	0.0	0.0	1.11	0.17	0.00
内质网	10.0	3.0	0.19	3.20	9.90
过氧化物酶体	0.0	0.0	0.19	0.00	0.01
溶酶体	0.0	0.0	0.14	0.00	0.00
高尔基体	0.0	0.0	0.02	0.29	0.00
液泡	0.0	0.0	0.25	0.00	0.00

2.6 PcCPR 三级结构模拟

I-TASSER 是一款蛋白高级结构及功能预测软件，能利用蛋白序列模拟预测出高质量、高分辨的蛋白 3D 结构。利用 I-TASSER 对 PcCPR 的高级结构进行模拟预测，模型结果利用 Swiss-PdbViewer 4.1.0 软件查看。模型预测结果如图 6 所示，PcCPR 蛋白的空间结构以螺旋和无规则卷曲为主。模型预测结果与 PDB 数据库中酿酒酵母 (2bf4A)、褐家

鼠 (1ja1A、1tllA) CPR 蛋白的结构相似度高。

2.7 PcCPR 保守结构域分析

PcCPR 的保守结构域分析预测使用 NCBI 中的 CDD 数据库 (Conserved Domain Database)，结果如图 7 所示。该蛋白 N 端第 65~208 位与 NADPH 依赖的 FMN 还原酶超家族 (FMN_red superfamily) 蛋白保守相似，第 264~731 位氨基酸的 C 端则与类似铁氧还蛋白还原酶超家族 (FNR_like superfamily) 蛋白保守相似；属于 CysJ 多保守域蛋白。结果显示，PcCPR 蛋白的第 264~731 位极类似于 NADPH cytochrome p450 reductase (CYPOR) 保守域，且在此区域中预测出多个 FAD 结合位点、NADP 结合位点、催化残基等。

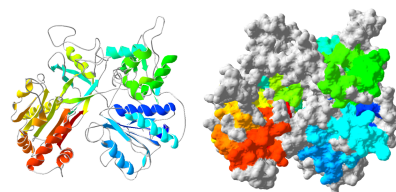


图 6 应用 I-TASSER 模拟 PcCPR 蛋白三级结构

Fig. 6 Predicted three-dimensional structure of PcCPR with I-TASSER

2.8 PcCPR 系统进化分析

从 Genbank 中下载不同分类下的真菌的 18 条 CPR 蛋白质序列，以及黏菌盘基网柄菌的序列。使用 BioEdit 7 对 PcCPR 蛋白与其他物种蛋白进行同源性分析显示 (表 4)，PcCPR 氨基酸序列与同属于担子菌亚门的黄孢原毛皮革菌、异孔异担子菌、毛韧革菌的序列相似度较高，相似性分别为 91%、87%、86%；序列一致性为 82%、79%和 76%；而与黑曲霉、裂殖酵母等子囊菌的 CPR 蛋白的相似度相对较低为 54%~60%，序列一致性为 37%~44%；与分类为黏菌的盘基网柄菌的相似为 49%，一致性为 32%。对这 20 条蛋白序列使用 MAFFT 软件进行多序列比对。使用 MEGA 软，采用 NJ 法构建系统发育树。结果显示 (图 8)，PcCPR 蛋白与黄孢原毛皮革菌中该蛋

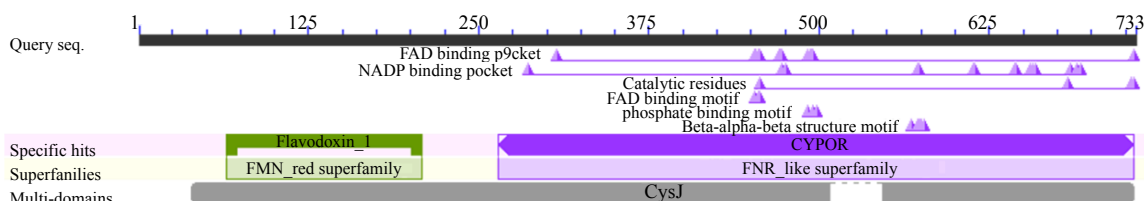


图 7 PcCPR 保守结构域分析

Fig. 7 Conserved domain prediction of PcCPR

表 4 PcCPR 与不同真菌 CPR 蛋白序列一致性及相似性

Table 4 Identities and similarities between PcCPR and CPR proteins from different fungi

物种名	序列登录号	序列一致性/%	序列相似性/%
茯苓 <i>Poria cocos</i>	KP768251	—	—
黄孢原毛平革菌 <i>Phanerochaete chrysosporium</i>	Q9HDG2.1	82	91
异孔异担子菌 <i>Heterobasidion irregular</i>	XP_009545907.1	79	87
毛韧革菌 <i>Stereum hirsutum</i>	XP_007304804.1	76	86
黑曲霉 <i>Aspergillus niger</i>	A2QS05.1	44	59
构巢曲霉 <i>Aspergillus nidulans</i>	Q5BFT5.1	44	59
烟曲霉 <i>Aspergillus fumigatus</i>	Q4WM67.1	44	59
灰葡萄孢 <i>Botrytis cinerea</i>	CAE76653.1	43	60
米曲霉 <i>Aspergillus oryzae</i>	Q2UHA7.1	43	59
土曲霉 <i>Aspergillus terreus</i>	Q0CMM0.1	43	59
稻瘟菌 <i>Magnaporthe oryzae</i>	ELQ61488.1	43	58
蛹虫草 <i>Cordyceps militaris</i>	XP_006671968.1	42	59
禾谷镰刀菌 <i>Fusarium graminearum</i>	ESU16412.1	42	59
粗糙脉孢菌 <i>Neurospora crassa</i>	EAA35207.1	42	59
里氏木霉 <i>Trichoderma reesei</i>	XP_006962366.1	41	59
麦芽糖假丝酵母 <i>Candida maltosa</i>	P50126.1	38	57
热带假丝酵母 <i>Candida tropicalis</i>	P37201.1	38	56
酿酒酵母 <i>Saccharomyces cerevisiae</i>	P16603.3	38	55
裂殖酵母 <i>Schizosaccharomyces pombe</i>	P36587.1	37	54
盘基网柄菌 <i>Dictyostelium discoideum</i>	XP_646400.1	32	49

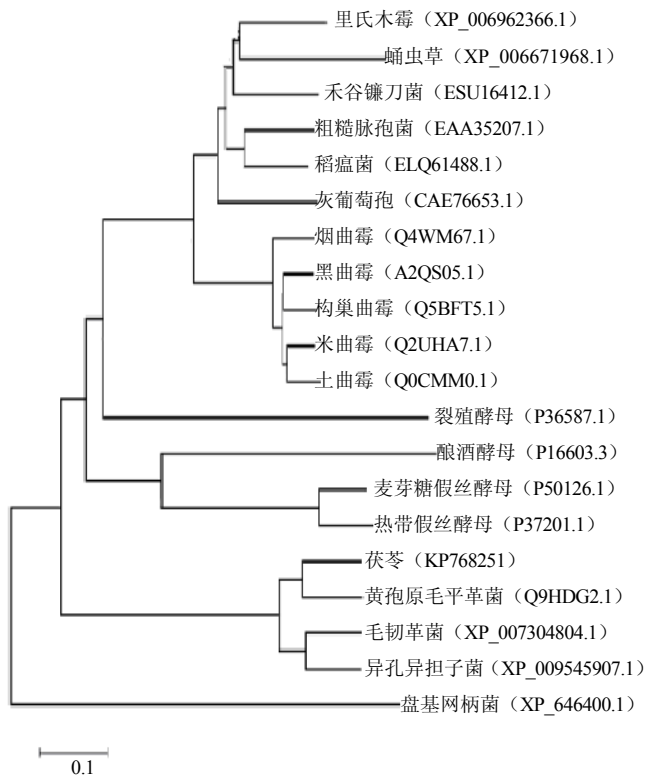


图 8 不同真菌 CPR 蛋白系统发育关系

Fig. 8 Phylogenetic relationships among CPR proteins from different fungi

白的亲缘关系最近，同时与另 2 种担子菌异孔异担子菌、毛韧革菌聚为一支。这与它们的分类地位一致，茯苓与黄孢原毛皮革菌同属于多孔菌目

(Polyporales); 而后两者则属于红菇目(Russulales)。不仅如此，真菌 CPR 蛋白的聚类结果与其分类基本一致。

3 讨论

在真核生物体内，CPR 是唯一必需的中间体黄素蛋白，是 P450s 的重要组成部分^[3]，而后者因其参与了生物体的许多重要代谢途径，在各物种里均成为了研究热点。CPR 基因已在许多动物、植物、酵母中克隆出来，而在丝状真菌仅有黄孢原毛皮革菌、黑曲霉等少量报道^[1,3,18]，PcCPR 基因尚未见报道。

本研究从茯苓中克隆出 PcCPR 基因，基因 cDNA 全长 2 514 bp，其中 2 199 bp 的编码区编码 732 个氨基酸，与黄孢原毛皮革菌的 CPR 同源性最高。同时获得相应的含有上下游序列的基因组基因序列 5 292 bp，PcCPR 基因的结构与报道的近缘种黄孢原毛皮革菌相似，均有 3 个内含子^[18]。与分离自美国的茯苓菌株 MD-104 SS10 的高通量测序结果对比分析，基因组基因的序列一致性为 93.56%，MD-104 SS10 菌株基因序列长度增加 115 bp，主要是在基因上游约-1 030 bp 位置多出一段 78 bp 的序列；基因编码区序列一致性为 98.59%，在 30 个位点有 31 bp 的碱基差异，这些碱基差异造成编码蛋白的共 7 个氨基酸的变异。

生物信息学分析结果显示，PcCPR 是一个相

对分子质量为 81 147, PI 为 5.39 的不稳定性亲水蛋白。蛋白为非分泌蛋白, 且不具有信号识别功能。蛋白的第 7~22 位氨基酸被预测为跨膜区, 这与报道的黄孢原毛平革菌的前 22 个氨基酸为跨膜区的结果相似^[18]。同时, 预测出蛋白定位于内质网上, 这与一般报道的 CPR 蛋白的定位结果一致; 且研究表明这种定位是 P450s 功能所必需的, 若将 N 端锚定序列去除, CPR 蛋白不能定位在膜上则 CPR 不能支撑 P450 反应, 只能向细胞色素 C、小分子化合物等提供电子^[2]。对蛋白的亲水性、信号肽及跨膜区的预测也支持了蛋白定位于膜上的预测。

CPR 蛋白的活性需要 FMN 和 FAD 2 个辅因子才能从底物 NADPH 获取并传递电子。PcCPR 蛋白预测出多个 FAD、NADP 的结合位点, 这些位点在模拟的蛋白三级结构空间上是各自相邻近的。并且预测得到 2 个黄素结合的保守结构域, 分别与 FMN 还原酶超家族、类似铁氧还蛋白还原酶 (ferredoxin NADP oxidoreductase, FNR) 超家族保守相似, 其中 FMN 与前者结合, FAD 结合至后者上。这一预测结果与学者研究认为的 CPR 基因是由 2 个分别编码黄素氧还蛋白 (flavodoxin, Fld) 和 FNR 基因融合而成的结果相一致^[2,18]。

综上所述, 本研究首次从茯苓中克隆出 PcCPR 基因, 通过各种生物信息学分析方法对蛋白序列进行预测和分析, 为深入研究该酶提供基础及参考。同时, CPR 在细胞色素 P450 酶系中具有关键作用, 也为研究茯苓细胞色素 P450 酶系奠定基础。

参考文献

- [1] Zhao C Q, Tang T, Feng X Y, *et al.* Cloning and characterisation of NADPH-dependent cytochrome P450 reductase gene in the cotton bollworm, *Helicoverpa armigera* [J]. *Pest Manage Sci*, 2014, 70(1): 130-139.
- [2] Pandey A V, Fluck C E. NADPH P450 oxidoreductase: structure, function, and pathology of diseases [J]. *Pharm Therap*, 2013, 138(2): 229-254.
- [3] Liu D, Zhou X J, Li M, *et al.* Characterization of NADPH-cytochrome P450 reductase gene from the cotton bollworm, *Helicoverpa armigera* [J]. *Gene*, 2014, 545(2): 262-270.
- [4] Riddick D S, Ding X X, Wolf C R, *et al.* NADPH-cytochrome P450 oxidoreductase: roles in physiology, pharmacology, and toxicology [J]. *Drug Metab Dispos*, 2013, 41(1): 12-23.
- [5] Wang Y Z, Zhang J, Zhao Y L, *et al.* Mycology, cultivation, traditional uses, phytochemistry and pharmacology of *Wolfiporia cocos* (Schwein.) Ryvarden *et Gilb.*: A review [J]. *J Ethnopharmacol*, 2013, 147(2): 265-276.
- [6] Wang W H, Dong H J, Yan R Y, *et al.* Comparative study of lanostane-type triterpene acids in different parts of *Poria cocos* (Schw.) Wolf by UHPLC-Fourier transform MS and UHPLC-triple quadrupole MS [J]. *J Pharm Biomed Anal*, 2015, 102: 203-214.
- [7] Xia B, Zhou Y, Tan H S, *et al.* Advanced ultra-performance liquid chromatography-photodiode array-quadrupole time-of-flight mass spectrometric methods for simultaneous screening and quantification of triterpenoids in *Poria cocos* [J]. *Food Chem*, 2014, 152: 237-244.
- [8] Paddon C J, Westfall P J, Pitera D J, *et al.* High-level semi-synthetic production of the potent antimalarial artemisinin [J]. *Nature*, 2013, 496(7446): 528-532.
- [9] Seki H, Ohyama K, Sawai S, *et al.* Licorice beta-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin [J]. *PNAS*, 2008, 105(37): 14204-14209.
- [10] Seki H, Sawai S, Ohyama K, *et al.* Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin [J]. *Plant Cell*, 2011, 23(11): 4112-4123.
- [11] Guo J, Zhou Y J, Hillwig M L, *et al.* CYP76AH1 catalyzes turnover of miltiradiene in tanshinones biosynthesis and enables heterologous production of ferruginol in yeasts [J]. *PNAS*, 2013, 110(29): 12108-12113.
- [12] Chen S L, Xu J, Liu C, *et al.* Genome sequence of the model medicinal mushroom *Ganoderma lucidum* [J]. *Nat Commun*, 2012, 3: 913.
- [13] Syed K, Shale K, Pagadala N S, *et al.* Systematic identification and evolutionary analysis of catalytically versatile cytochrome p450 monooxygenase families enriched in model basidiomycete fungi [J]. *PLoS One*, 2014, 9(1): e86683.
- [14] Ide M, Ichinose H, Wariishi H. Molecular identification and functional characterization of cytochrome P450 monooxygenases from the brown-rot basidiomycete *Postia placenta* [J]. *Arch Microbiol*, 2012, 194(4): 243-253.
- [15] Shu S H, Chen B, Zhou M C, *et al.* De novo sequencing and transcriptome analysis of *Wolfiporia cocos* to reveal genes related to biosynthesis of triterpenoids [J]. *PLoS One*, 2013, 8(8): e71350.
- [16] Scotto-lavino E, Du G, Frohman M A. 3' end cDNA amplification using classic RACE [J]. *Nat Prod*, 2006, 1(6): 2742-2745.
- [17] Floudas D, Binder M, Riley R, *et al.* The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes [J]. *Science*, 2012, 336(6089): 1715-1719.
- [18] Yadav J S, Loper J C. Cytochrome P450 oxidoreductase gene and its differentially terminated cDNAs from the white rot fungus *Phanerochaete chrysosporium* [J]. *Curr Genet*, 2000, 37(1): 65-73.