

# 亚麻 *COBRA* 基因家族的鉴定与生物信息学分析

齐燕妮, 李闻娟, 王利民, 赵 玮, 党 照, 谢亚萍, 张建平

(甘肃省农业科学院作物研究所, 甘肃 兰州 730070)

**摘要:** *COBRA* 基因编码糖基磷脂酰肌醇锚定蛋白, 是次生壁加厚和纤维素含量的调节因子。亚麻纤维为韧皮纤维, 纤维产量与次生壁中纤维素含量及次生壁加厚密切相关。通过 BLASTP 和 Pfam 分析, 鉴定到 24 个亚麻 *COBRA* 家族成员, 并对其进行了基因结构、蛋白理化性质及系统进化等生物信息学分析。结果表明, 有 22 个成员具有跨膜结构域, 21 个成员具有 N-端信号肽, 16 个成员具有 GPI 锚定位点, 所有蛋白均具有 CCVS 保守结构域而且都定位在细胞膜。基因结构和系统进化分析表明, 亚麻 *COBRA* 基因家族分为 2 个亚家族: *COBRA*-I 和 *COBRA*-II, 2 个亚家族的基因结构具有明显的差异。

**关键词:** 亚麻; 纤维; *COBRA* 基因家族; 序列分析; 系统进化分析

**中图分类号:** S565.9 **文献标志码:** A **文章编号:** 1001-1463(2019)09-0033-06

**doi:** 10.3969/j.issn.1001-1463.2019.09.009

## Identification and Bioinformatics Analysis of the *COBRA* Gene Family in Flax

QI Yanni, LI Wenjuan, WANG Limin, ZHAO Wei, DANG Zhao, XIE Yaping, ZHANG Jianping  
(Institute of Crops, Gansu Academy of Agricultural Sciences, Lanzhou Gansu 730070, China)

**Abstract:** The *COBRA* gene encodes the glycosylphosphatidylinositol anchoring protein, which is a regulator of secondary wall thickening and cellulose content. Flax fiber is bast fiber, and fiber yield is closely related to the content of cellulose in secondary wall and the thickening of secondary wall. Through BLASTP and Pfam analysis, 24 members of the *COBRA* family were identified in this study, and the gene structure, physicochemical properties of proteins, systematic evolution were conducted. The results showed that 22 members had transmembrane domain, 21 members had N-terminal signal peptide, 16 members had GPI anchor site, and all proteins had CCVS conserved domain and were located in cell membrane. The results of gene structure analysis and systematic evolution showed that the *COBRA* gene family was divided into two subfamilies, *COBRA*-I and *COBRA*-II, the gene structure of the two subfamilies were significantly different. This study laid a foundation for the in-depth study on the function of *COBRA* gene and the use of *COBRA* gene to increase the fiber yield.

**Key words:** Flax; Fiber; *COBRA* gene family; Sequence analysis; Phylogeny analysis

亚麻 (*Linum usitatissimum* L.) 是一年生自花授粉的植物, 是亚麻科中唯一重要的农作物<sup>[1]</sup>。亚麻根据其用途, 分为油用、纤维

用两种, 也有油纤两用亚麻<sup>[2]</sup>。纤维亚麻株高可达到 80 ~ 120 cm、分枝少、籽粒小, 主要分布西欧、俄罗斯、中国<sup>[3]</sup>。亚麻纤维

收稿日期: 2019-07-23

基金项目: 国家自然科学基金项目(31760426)资助。

作者简介: 齐燕妮(1987—), 女, 甘肃庆阳人, 研究实习员, 硕士, 主要从事亚麻分子育种研究工作。Email: qiyanni@gsagr.ac.cn。

通信作者: 张建平(1972—), 男, 陕西蓝田人, 研究员, 主要从事亚麻育种研究工作。Email: zhangJPzw3@gsagr.ac.cn。

是纺织产品制造的原材料, 中国亚麻纤维需求较大, 而且是世界上最大的亚麻纤维出口国, 黑龙江占亚麻纤维总产量的 85% [4]。提高纤维产量是亚麻育种的主要目标之一, 明确纤维合成及调控基因可以为纤维产量的提高提供理论基础。

*COBRA* 基因编码糖基磷脂酰肌醇锚定蛋白 (glycosyl-phosphatidyl inositol-anchored protein), 一般具有 N-端信号肽、碳水化合物结合模块(CBM)、高度保守的中心半胱氨酸富集区 (CCVS) 以及将蛋白靶向内质网的 C-端 GPI 锚定位点和疏水尾巴 [5]。*COBRA* 蛋白是细胞定向生长和纤维素结晶度的关键调节因子, 其突变会导致细胞次生壁厚度和纤维素含量降低 [6-7]。已经在拟南芥、番茄、水稻及玉米中分别鉴定到 12、17、11、9 个 *COBRA* 基因家族成员, 拟南芥中 *COBRA* 家族分为 2 个亚家族 [8-11]。我们利用生物信息学方法, 在亚麻全基因组中对 *COBRA* 基因家族进行鉴定, 并对其序列及系统进化进行分析, 以期深入研究 *COBRA* 家族各成员功能, 进而提高亚麻纤维产量提供依据。

## 1 材料与方法

### 1.1 亚麻 *COBRA* 基因家族的鉴定

从 NCBI 数据库中下载拟南芥 *COBRA* 基因家族各成员的蛋白序列。使用 BlastP 在亚麻全基因组蛋白数据库中进行搜索, 阈值小于等于  $1E-5$ , 得到候选序列后用 Pfam 对结构域进行确认。

### 1.2 *COBRA* 基因家族的序列分析

利用 Expasy 在线分析蛋白质的分子量、等电点, 采用 SignalP-5.0 对亚麻 *COBRA* 蛋白进行 N-端信号肽预测。使用 big-PI 分析蛋白的 C 端疏水末端锚定位点, 用 SOSUI 在线预测跨膜结构域, 用 MEME 分析 *COBRA* 蛋白的保守基序, 使用 Plant-mPloc 进行亚细胞定位预测。利用 IBS 绘制基因结

构图。

### 1.3 *COBRA* 基因家族的进化分析

利用 MEGA6.0 软件中的 ClustalW, 采用默认参数, 对拟南芥和亚麻 *COBRA* 蛋白的氨基酸序列进行多重比对, 比对结果采用 NJ (Neighbor-Joining) 法构建系统进化树, Bootstrap 校验值为 1 000 次。

## 2 结果与分析

### 2.1 亚麻 *COBRA* 基因家族的鉴定

利用 12 条拟南芥 *COBRA* 基因氨基酸序列在亚麻全基因组进行 BLASTP 搜索, 获得亚麻 *COBRA* 蛋白的候选序列。用 Pfam 对候选序列进行注释, 最终确定亚麻基因组 *COBRA* 家族含有 24 个成员, 均含有 *COBRA* 家族的特有结构域 *COBRA*, 命名为 *LuCOBL1~LuCOBL24* (表 1)。

### 2.2 亚麻 *COBRA* 基因家族的序列和结构域特征

对亚麻 *COBRA* 基因家族进行系统进化分析, 并对其外显子-内含子结构进行分析 (图 1, 表 1)。从进化树可以看出, 亚麻 *COBRA* 蛋白与拟南芥类似, 也被分为 2 个亚家族 (*COBRA-I* 和 *COBRA-II*)。*COBRA-I* 有 18 个成员, 包括 *LuCOBL-3*、*LuCOBL-4*、*LuCOBL-6~LuCOBL-16*、*LuCOBL-20~LuCOBL-24*, 蛋白质长度为 375~616 aa, 大部分成员的氨基酸数量在 400 多个, 外显子数量为 4~8 个, 内含子数量为 3~7 个; 其中有 14 个基因均含有 6 个外显子, 5 个内含子, 说明同一亚家族的基因具有相似的基因结构; *COBRA-II* 包含 *LuCOBL-1*、*LuCOBL-2*、*LuCOBL-5*、*LuCOBL-17~LuCOBL-19* 共 6 个基因, 蛋白质长度范围在 662~709 aa, 外显子数量 2~3 个; 其中有 4 个基因均含有 2 个外显子。从以上结果看出, 亚麻 *COBRA-II* 亚家族的蛋白长度明显大于 *COBRA-I*, 但外显子数量却比较少, 2 个亚家族具有明显的差异。

对亚麻 24 个 COBRA 蛋白的理化性质进行分析表明,不同亚家族之间,同一亚家族不同成员之间均有差异(表1)。分子量为 41 648.89~77 575.57 Da,等电点为 5.61~9.12,跨膜结构域为 0~3,大部分蛋白含有 2 个跨膜结构域。信号肽预测结果表明,除 COBRA-I 亚家族的 LuCOBL-8、LuCOBL-16 和 LuCOBL-3 无 N-端信号肽外,其余蛋白质均含有信号肽序列。COBRA-I 亚家族除 LuCOBL-9、LuCOBL-21、LuCOBL-22 和 LuCOBL-24 外,其余成员都有 GPI 锚定位点,而 COBRA-II 只有 2 个成员具有 GPI 锚定位点(LuCOBRA-1 和 LuCOBRA-17)。对 COBRA 蛋白进行亚细胞定位预测,所有基因都定位在细胞膜。

亚麻 COBRA 蛋白多重序列比对结果表明,所有亚麻 COBRA 蛋白均具有保守结构域 C CVS(图2)。利用 MEME 对亚麻 COBRA 蛋白进行 Motif 预测,所有蛋白都具有 Motif1、Motif3 和 Motif5,说明这 3 个 Motif 为 COBRA 家族的特征性 Motif。此外,有 23 个基因含有 Motif2,有 17~20 个基因含有 Motif4、Motif6、Motif15、Motif20,说明这些 Motif 在 COBRA 蛋白中也相当保守(图3)。

### 2.3 亚麻 COBRA 基因家族的系统进化分析

选择模式植物拟南芥 COBRA 基因和亚麻 COBRA 基因,利用 MEGA6.0 以 NJ 法构建系统进化树(图4)。从进化树上可以看出,亚麻和拟南芥的 COBRA 基因都被聚类在两

表 1 亚麻 COBRA 基因家族成员及其序列特征

基因名称	蛋白长度 /aa	分子量 /Da	等电点	外显子 数量	信号肽	跨膜结构	GPI-锚定位点	亚细胞定位
<i>LuCOBL1</i>	674	73 189.02	5.61	3	有	0	有	细胞膜
<i>LuCOBL2</i>	696	76 577.63	8.95	2	有	1	无	细胞膜
<i>LuCOBL3</i>	424	46 938.50	6.40	6	有	2	有	细胞膜
<i>LuCOBL4</i>	443	49 793.16	9.05	6	有	2	有	细胞膜
<i>LuCOBL5</i>	677	73 572.68	6.40	3	有	2	无	细胞膜
<i>LuCOBL6</i>	440	49 232.59	8.21	6	有	3	有	细胞膜
<i>LuCOBL7</i>	436	48 983.43	9.07	5	有	2	有	细胞膜
<i>LuCOBL8</i>	436	49 218.43	8.73	6	无	1	有	细胞膜
<i>LuCOBL9</i>	454	51 014.95	8.82	6	有	2	无	细胞膜
<i>LuCOBL10</i>	616	68 330.59	9.12	8	有	2	有	细胞膜
<i>LuCOBL11</i>	437	48 869.07	7.79	6	有	2	有	细胞膜
<i>LuCOBL12</i>	442	49 615.81	9.01	6	有	2	有	细胞膜
<i>LuCOBL13</i>	455	50 610.94	8.62	6	有	2	有	细胞膜
<i>LuCOBL14</i>	375	42 029.40	8.84	6	有	1	有	细胞膜
<i>LuCOBL15</i>	455	50 602.88	8.70	6	有	2	有	细胞膜
<i>LuCOBL16</i>	377	41 648.89	8.49	4	无	0	有	细胞膜
<i>LuCOBL17</i>	665	72 283.13	6.04	2	有	2	有	细胞膜
<i>LuCOBL18</i>	709	77 575.57	8.90	2	有	1	无	细胞膜
<i>LuCOBL19</i>	662	72 038.92	5.87	2	有	2	无	细胞膜
<i>LuCOBL20</i>	435	48 744.25	9.07	4	有	2	有	细胞膜
<i>LuCOBL21</i>	462	52 031.52	8.90	6	无	1	无	细胞膜
<i>LuCOBL22</i>	454	50 988.90	8.84	6	有	2	无	细胞膜
<i>LuCOBL23</i>	424	47 046.62	6.44	6	有	2	有	细胞膜
<i>LuCOBL24</i>	444	49 854.18	6.25	6	有	2	无	细胞膜





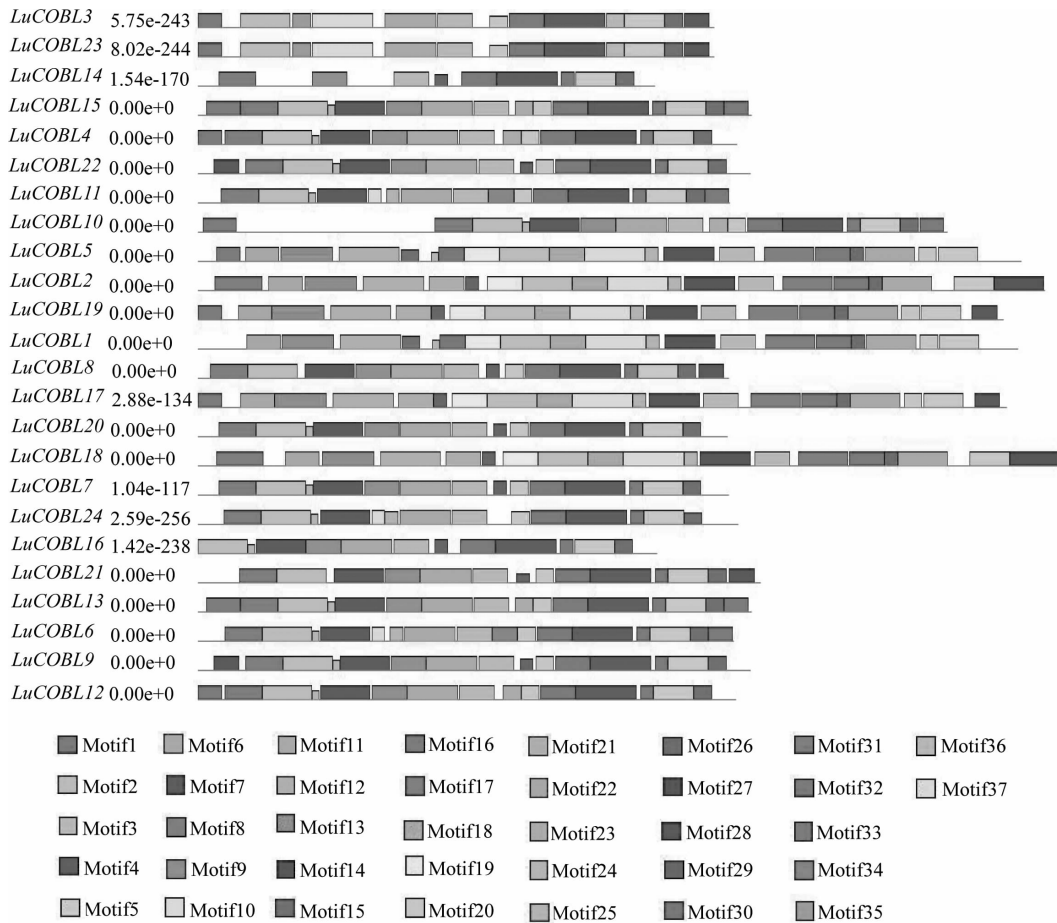


图 3 亚麻 COBRA 基因家族 Motif 分析

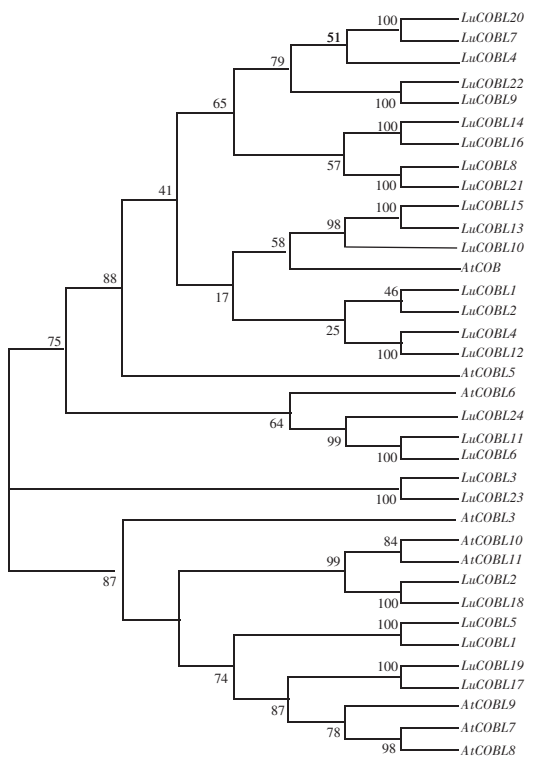


图 4 亚麻与拟南芥 COBRA 基因家族进化分析

大支上，与单独用亚麻蛋白序列构建进化树的聚类结果一致。在 COBRA-I 亚家族中，*AtCOBL4* 与 *LuCOBL7* 和 *LuCOBL20*，*AtCOBL* 与 *LuCOBL10*、*LuCOBL13* 和 *LuCOBL15*，*AtCOBL3* 与 *LuCOBL6*、*LuCOBL11* 和 *LuCOBL24* 进化距离较近；在 COBRA-II 亚家族中，*AtCOBL10*、*AtCOBL11* 与 *LuCOBL2* 和 *LuCOBL18*，*AtCOBL7*、*AtCOBL8*、*AtCOBL9* 与 *LuCOBL17* 和 *LuCOBL19* 遗传距离比较相近。

### 3 结论与讨论

共鉴定到 24 个亚麻 COBRA 基因家族成员。对其基因及蛋白序列进行分析发现，所有蛋白都具有 CCVS 结构域，有 21 个蛋白含有 N-端信号肽，22 个蛋白含有跨膜结构域，16 个蛋白具有 GPI 锚定位点，这符合 COBRA 蛋白的主要特征<sup>[5]</sup>。*LuCOBL2*、

*LuCOBL8*、*LuCOBL16* 等基因编码蛋白虽缺少信号肽或 GPI 锚定位点,但由于其含有 *COBRA* 家族特有结构域 COBRA (PF04833.1012),亲缘关系与其他成员又比较接近,所以将其作为亚麻 *COBRA* 基因。从亚麻 *COBRA* 基因家族整体分析结果看,亚麻 *COBRA* 基因分为 2 个亚家族,与拟南芥分类结果相同,同一亚家族的基因具有相似的基因结构,蛋白长度,不同亚家族之间差异明显。

目前对 *COBRA* 基因家族的研究在拟南芥、水稻、番茄、玉米、杨树等植物中有过报道,也研究过其中一些成员的功能<sup>[8-12]</sup>,亚麻 *COBRA* 基因家族的研究尚未有报道,因此本研究对亚麻 *COBRA* 基因家族的鉴定及一系列生物信息学分析,为进一步挖掘各成员的功能奠定了基础。然而,本研究只是初步鉴定了亚麻 *COBRA* 基因,对于无 *COBRA* 家族主要特征的个别成员需进行更深入的分析验证,同时需要利用 CRISPR/Cas9、过表达、反义 RNA、遗传转化等技术对各成员的具体功能进行研究,以便更好的利用 *COBRA* 基因进行亚麻育种。

#### 参考文献:

- [1] ZOHARY D, MONOPHYLETIC V S. Polyphyletic origin of the crops on which agriculture was founded in the near east [J]. *Gen. Resour. Crop Evol.*, 1999(46): 133-142.
- [2] 张运晖, 赵 瑛, 罗俊杰. 甘肃胡麻产业发展浅议[J]. *甘肃农业科技*, 2013(7): 54-55.
- [3] GREEN A G, CHEN Y, SINGH S P, et al. Flax. In *Compendium of transgenic crop plants* [D]. Oxford: Blackwell Publishing Ltd, 2008: 199-226.
- [4] WU J, ZHAO Q, ZHANG L, et al. QTL mapping of fiber-related traits based on a high-density genetic map in flax (*Linum usitatissimum* L.) [J]. *Front. Plant. Sci.*, 2018(9): 885.
- [5] BORNER G H, LILLEY K S, STEVENS T J, et al. Identification of glycosyl phosphatidylinositol-anchored proteins in *Arabidopsis*, a proteomic and genomic analysis [J]. *Plant Physiol.*, 2003(132): 568-577.
- [6] BROWN, D M, ZEEF L A H, ELLIS J, et al. Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics [J]. *Plant Cell*, 2005 (17): 2281-2295.
- [7] SATO K, SUZUKI R, NISHIKUBO N, et al. Isolation of a novel cell wall architecture mutant of rice with defective *Arabidopsis* COBL4 ortholog BC1 required for regulated deposition of secondary cell wall components [J]. *Planta*, 2010(232): 257-270.
- [8] ROUDIER F, SCHINDELMAN G, DESALLE R, et al. The COBRA family of putative GPI-anchored proteins in *Arabidopsis* a new fellowship in expansion [J]. *Plant Physiol.*, 2002(130): 538-548.
- [9] CAO Y, TANG X F, GIOVANNONI J, et al. Functional characterization of a tomato COBRA-like gene functioning in fruit development and ripening [J]. *BMC Plant Biol.*, 2012(12): 211.
- [10] DAI X X, YOU C J, WANG L, et al. Molecular characterization, expression pattern, and function analysis of the OsBC1L family in rice [J]. *Plant Mol. Biol.*, 2009(71): 469-481.
- [11] BRADY S M, SONG S, DHUGGA K S, et al. Combining expression and comparative evolutionary analysis. The COBRA gene family [J]. *Plant Physiol.*, 2007(143): 172-187.
- [12] YE X, KANG B G, OSBURN L D, et al. The COBRA gene family in populus and gene expression in vegetative organs and in response to hormones and environmental stresses [J]. *Plant Growth Regul.*, 2009(58): 211-223.

(本文责编: 陈 珩)