百脉根细胞亲环素电子克隆和生物信息学分析

闫 嵩，任伟超，刘振鹏，张开雪，刘秀波，马 伟*
黑龙江中医药大学药学院，黑龙江 哈尔滨 150040

摘 要：目的 通过电子克隆技术对百脉根细胞亲环素（cyclophilin）基因进行预测。**方法** 以大豆 cyclophilin 序列为探针，基于 NCBI 中百脉根的 EST 数据库和 CAP3 在线软件进行序列拼接，利用生物信息学数据库及相关软件对其结构和功能进行预测分析。**结果** 百脉根 cyclophilin 基因全长 1 346 bp，包含 771 bp 的开放阅读框，编码 256 个氨基酸，该蛋白为亲水性蛋白。**结论** 为进一步解释基因的分子功能奠定理论及实验基础。

关键词： 百脉根；细胞亲环素；电子克隆；生物信息学；分子功能
中图分类号： R282.12 **文献标志码：** A **文章编号：** 0253 - 2670(2016)19 - 3481 - 05
DOI: 10.7501/j.issn.0253-2670.2016.19.021

In silico cloning and bioinformatics analysis of cyclophilin from *Lotus corniculatus*

YAN Song, REN Wei-chao, LIU Zhen-peng, ZHANG Kai-xue, LIU Xiu-bo, MA Wei
College of Pharmaceutical Sciences, Heilongjiang University of Chinese Medicine, Harbin 150040, China

Abstract: Objective Using electronic cloning technology to predict cyclophilin gene of *Lotus corniculatus*. **Methods** Using glycine max cyclophilin sequence as probe sequence, based on EST sequence from NCBI and assembled by CAP3 sequence assembly programme, using bioinformatic database and related software, the structure prediction and function analysis were performed. **Results** The full length of cyclophilin gene was 1 346 bp, it contained a 771 bp ORF, encoding 256 amino acids, and the protein was a hydrophilic protein. **Conclusion** The study is intended to further explain the molecular genetic function theory and experimental basis.

Key words: *Lotus corniculatus* L.; cyclophilin; in silico cloning; bioinformatics; molecular function

电子克隆（in silico cloning）是近年来伴随着基因组计划和 EST 计划发展起来的基因克隆方法。电子克隆技术是通过序列的拼接和组装，克隆新基因的技术，具有高效、快速、投入低，并可以为实验克隆提供精准的参考序列等优点[1-3]。

百脉根 *Lotus corniculatus* L. 别名黄花草、牛角花等，是豆科百脉根属多年生草本植物，以全草入药。具有清热解毒、止咳平喘的功效。同时，百脉根也可用于防止水土流失及优良的豆科牧草[4-5]，广泛应用于中医药、环保及畜牧业等行业。

细胞亲环素（cyclophilin）是能够与免疫抑制剂环孢霉素 A 特异结合，普遍存在于细菌、真菌、植物和动物等有机体中的一个较大的蛋白家族，具有肽脯氨酰顺反异构酶活性，能够催化脯氨酸肽键的顺反异构化[6-8]。此外，在植物中细胞亲环素促进 HSP90 调节 RNA 诱导沉默复合体的装配，同时在细胞分裂、转录调节、信号转导及胁迫应答等多种生理过程中发挥重要作用[9]。目前，已有对水稻[10]、小麦[11]、水芹[12]、大豆[13]、棉花[14]、茶树[15]等经济类作物细胞亲环素的研究，而在中药方面的研究报道甚少。

1 材料与方法

1.1 电子克隆获得新基因序列

以大豆 AtCYP20-2 基因（BW675370）作为探针，序列使用 Blastn 工具检索 NCBI 中百脉根 EST 序列，得到与探针序列同源性较高的百脉根 EST 序列，使用在线工具 CAP3[16]进行拼接，以拼接好的

收稿日期：2016-03-19
基金项目：国家自然基金资助项目（81274010）；黑龙江省杰出青年基金（JC201101）；黑龙江中医药大学“优秀创新人才支持计划”（2012001）；哈尔滨市优秀学科带头人基金（2014RFXXJ122）；黑龙江省教育厅科学技术研究项目（12541743）
作者简介：闫 嵩（1989—），男，在读硕士研究生，研究方向为药用植物生物工程。E-mail: zleztme@163.com
*通信作者 马 伟，女，研究员，博士生导师，研究方向为药用植物生物工程。Tel: (0451)82193430 E-mail: mawei@hljucm.net

重叠群（Contig）为探针，再次 Blast 检索，直到没有新的 EST 序列可供拼接，Contig 不能延伸为止。

1.2　生物信息学分析

基于在线生物信息学软件对新基因序列进行分析。本研究对该基因进行开放阅读框分析、理化性质分析、结构预测、信号肽分析、亲疏水性分析、亚细胞定位预测。本研究在线分析工具见表 1。

2　结果与分析

2.1　新基因的识别

以大豆细胞亲环素基因作为探针，利用 NCBI 中百脉根 EST 数据库进行 Blastn 搜索，将得到的 EST 序列使用 CAP3 进行拼接，获得 1 条全长为 1 346 bp 的 Contig。通过 ORF Finder 软件分析结果显示，其开放阅读框长度为 771 bp，编码 256 个氨基酸，见图 1。

表 1　生物信息学预测项目及相关网站

Table 1　Bioinformatic prediction and websites

研究内容	相关网站
相似性搜索	Blast (http://blast.ncbi.nlm.nih.gov/)
重叠群组装	CAP3 (http://doua.prabi.fr/software/cap3)
开放阅读框	ORF finder (http://www.ncbi.nlm.nih.gov/gorf/orfig.cgi)
蛋白质理化性质预测	ProtParam (http://web.expasy.org/protparam/)
信号肽预测	SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/)
蛋白质疏水性/亲水性分析	ProtScale (http://web.expasy.org/protscale/)
蛋白质跨膜结构预测	TMpred (http://www.ch.embnet.org/software/TMPRED_form.html)
蛋白质亚细胞定位	Psort (http://www.genscript.com/psort.html)
蛋白质二级结构预测	SOPMA (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)
蛋白质功能分类预测	Protfun 2.2 (http://www.cbs.dtu.dk/services/ProtFun/)
蛋白质三级结构预测	SWISS-MODEL (http://www.swissmodel.expasy.org/)

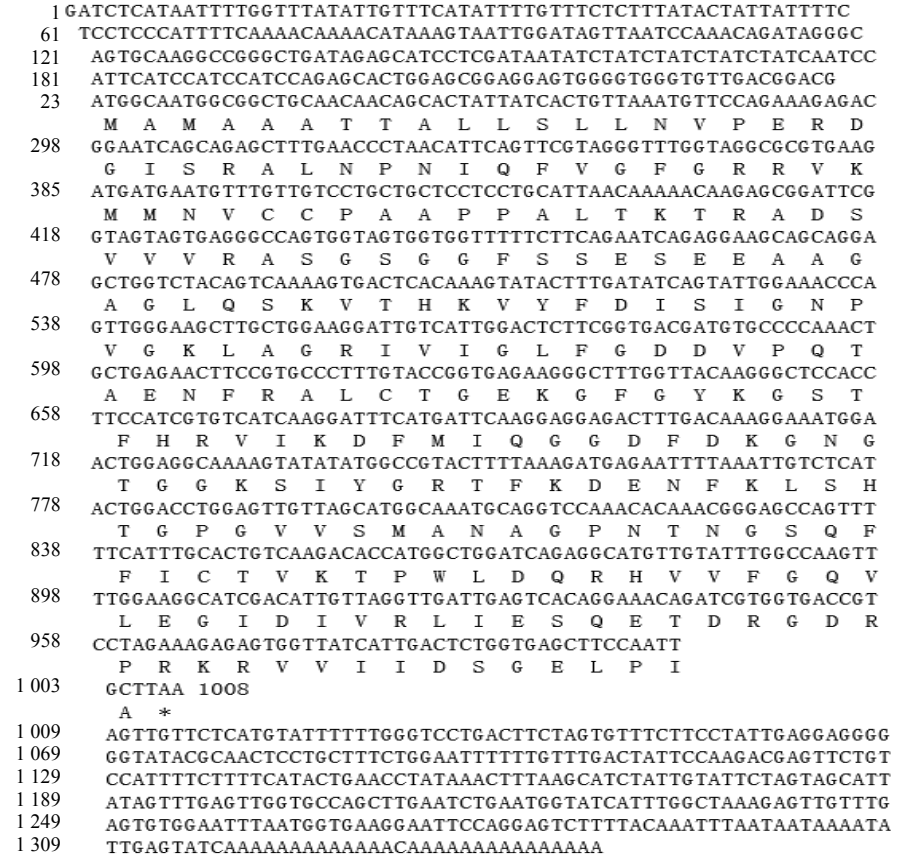

图 1　百脉根细胞亲环素基因电子克隆 cDNA 序列和编码氨基酸序列

Fig. 1　cDNA sequence and coding amino acid sequence of cyclophilin in *L. corniculatus* using in silico cloning

2.2 百脉根细胞亲环素基因编码氨基酸一级结构预测

通过在线软件 ProtParam，基于蛋白质数据库，对百脉根细胞亲环素基因编码的氨基酸的一级结构预测见表 2。

表 2 百脉根细胞亲环素蛋白的一级结构预测分析

Table 2 Primary structure analysis of cyclophilin in *L. corniculatus*

一级结构特征	预测结果
编码氨基酸个数	256
等电点（pI）	9.08
相对分子质量（M_W）	27 353.2
正电荷残基（Arg＋Lys）	30
负电荷残基（Asp＋Glu）	25
分子式	$C_{1205}H_{1922}N_{346}O_{361}S_{10}$
不稳定系数（II）	30.93
平均疏水性（GRAVY）	−0.131
脂肪系数（AI）	79.96

2.3 百脉根细胞亲环素信号肽预测和分析

采用 SignaIP4.1 Server[17]，预测百脉根细胞亲环素的信号肽，结果如图 2 和表 3。基因所编码的蛋白质在氨基酸序列的 26 位置上有一个潜在的信号肽断裂位点，概率为 0.189。由于氨基酸残基的原始剪切位点和信号肽的分值均较小，可推测该基因所编码的蛋白可能不存在信号肽，为非分泌蛋白，该蛋白在细胞质中合成后，不进行蛋白转运。

2.4 百脉根细胞亲环素蛋白疏水性/亲水性分析

利用 ProScale 在线软件，对百脉根细胞亲环素编码的氨基酸进行亲疏水性预测。根据正值越大表示蛋白疏水性越强，负值越大表示蛋白亲水性越强，介于+0.5～−0.5 的主要为两性氨基酸的规律，预测结果如图 3。因此推测细胞亲环素编码的蛋白为亲水性蛋白质[18]。

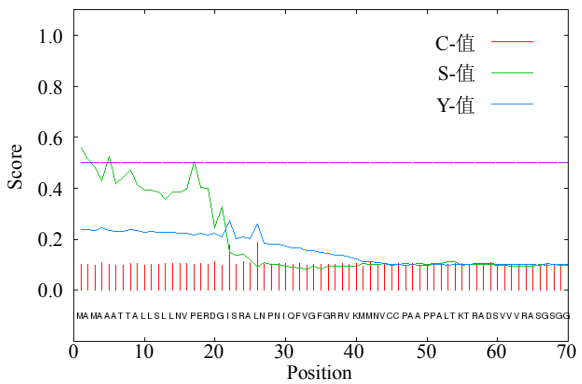

图 2 百脉根细胞亲环素信号肽预测结果

Fig. 2 Signal P-NN prediction for cyclophilin in *L. corniculatus*

表 3 百脉根细胞亲环素信号肽预测

Table 3 Signal peptide prediction for cyclophilin in *L. corniculatus*

预测	位置	分值	剪切位点	信号肽
最大原始剪切	26	0.189		
最大综合剪切	22	0.275		
信号肽最大值	1	0.557		
信号肽平均值	1～21	0.421		
加权平均值	1～21	0.354	0.450	NO

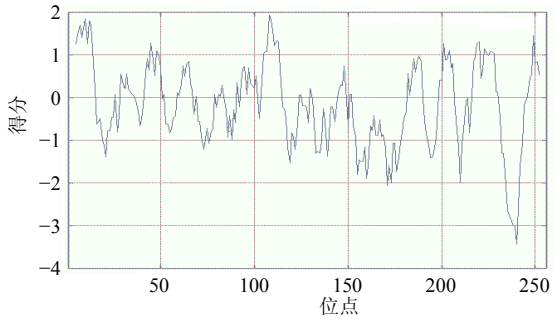

图 3 百脉根细胞亲环素蛋白疏水性/亲水性分析结果

Fig. 3 Hydrophobicity/hydrophilicity analysis on cyclophilin in *L. corniculatus*

2.5 百脉根细胞亲环素蛋白的二级结构预测

蛋白质二级结构是指蛋白质分子中某一段肽链的局部空间结构，也就是该段肽链主链骨架原子的相对空间位置，并不涉及氨基酸残基侧链的构象[19]。通过在线软件 SOPMA[20]，基于蛋白数据库，对百脉根细胞亲环素蛋白进行二级结构预测见图 4，结果表明该蛋白质的二级结构主要由 4 种形式组成，即由无规卷曲占 42.19%，延伸链占 28.52%，α-螺旋占 19.53%，β-折叠占 9.77%。据此推测，无规卷曲是百脉根细胞亲环素蛋白二级结构中最大量的结构元件。

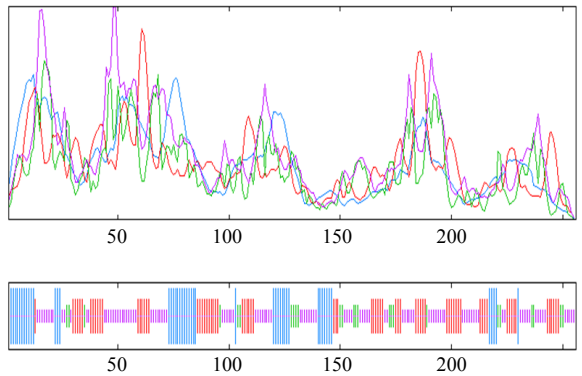

图 4 百脉根细胞亲环素蛋白二级结构预测分析

Fig. 4 Secondary structure prediction of cyclophilin protein

2.6　百脉根细胞亲环素蛋白质跨膜结构预测

对于跨膜结构域的预测和分析，有助于对蛋白的结构和功能以及其他基本信息的理解。采用在线跨膜蛋白结构预测 TMpred 软件，对该蛋白氨基酸的跨膜结构域进行预测的结果见图 5。依据图 5 可发现有 2 处跨膜结构域分别是 1～17、181～200 氨基酸位置。但一般将纵坐标分值大于 500 被认为是可能性较高的跨膜结构域，1～17 处的跨膜结构域的分值为 803，而 181～200 处的跨膜结构域的分值仅为 77，据此推测仅有 1 处跨膜结构域即 1～17 处的跨膜结构域。

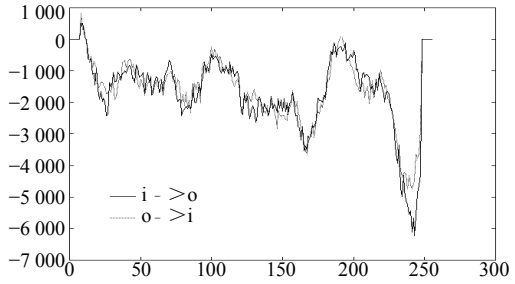

图 5　基于 TMpred 软件预测蛋白质的跨膜区域预测

Fig. 5　Predicted transmembrane domains of deduced amino acid sequence by TMpred software

2.7　百脉根细胞亲环素蛋白的亚细胞定位

使用 Psort 在线软件[21]，基于蛋白数据库，对百脉根细胞亲环素蛋白进行亚细胞定位。蛋白质在细胞内定位是了解蛋白生物学功能的基础，存在于不同位置，相应的靶蛋白的定位过程也不同。结果显示该蛋白在细胞质的概率最大，达到 60.9%，在线粒体的概率是 26.1%，在细胞核的概率是 8.7%，在细胞液中有 4.3%的概率，说明百脉根细胞亲环素蛋白可能位于细胞质中。

2.8　百脉根细胞亲环素蛋白质功能分类预测和分析

通过 CBS 的 Protfun 软件预测编码蛋白的功能[22]，其功能分类如表 4。该蛋白为酶蛋白，根据酶的分类，很可能属于转运蛋白。电压门控离子通道功能的概率最高为 0.279。据此推断该蛋白是一种在电压门控离子通道中起转运作用的转运蛋白。

2.9　百脉根细胞亲环素蛋白质三级结构预测

蛋白质三级结构指整条多肽链中全部氨基酸残基的相对空间位置，也就是整条肽链所有原子在三维空间的排布位置。将百脉根细胞亲环素编码的蛋白质序列提交至 SWISS-MODEL 在线软件[23]，采用同源建模法，见图 6。该蛋白主要由无规卷曲、α-螺旋、β-折叠组成，基本与二级结构预测结果一致。

表 4　百脉根细胞亲环素蛋白的功能预测

Table 4　Functions prediction of cyclophilin protein

基因本体论类别	可能性（概率）	机率/%
信号转导	0.205	0.958
受体	0.007	0.041
激素	0.001	0.154
结构蛋白	0.003	0.107
转运蛋白	0.025	0.229
离子通道	0.169	2.965
电压门控离子通道	0.279	12.682
阳离子通道	0.146	3.174
转录	0.219	1.711
转录调控	0.111	0.888
应激蛋白	0.073	0.830
免疫反应	0.011	0.129
生长因子	0.005	0.357
金属离子运输	0.018	0.039

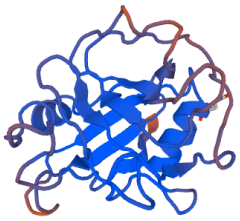

图 6　百脉根 cyclophilin 蛋白质空间构象模拟图

Fig. 6　Confonnation simulated maps of cyclophilin protein

3　讨论

电子克隆最初是随着人类基因组和 EST 计划而产生和发展的。与传统的基因克隆方法相比，电子克隆主要有以下优点：速度快、成本低、技术要求低、常规设备即可、针对性强[24]。电子克隆在植物基因克隆中也起到不可替代的作用，随着转录组、基因组等生物信息的丰富，电子克隆技术将会在药用植物种质资源等方面具有广阔的应用前景[25]。

通过电子克隆得到百脉根细胞亲环素基因并进行生物信息学分析，推测该基因的开放阅读框长度为 771 bp，编码 256 个氨基酸。编码的蛋白为亲水性蛋白，亚细胞定位显示可能位于细胞质中，二级结构中最大量的结构元件是无规卷曲，该蛋白在电压门控离子通道中起转运作用。电子克隆虽然在基因克隆效率上有很大的优势，但也存在一些弊端。一方面受到已有的 EST 数目的限制，同时电子克隆不适用于种间保守性差的基因和外显子数目多而且每个外显子短的基因，使得电子克隆技术应用的普遍性受到一定程度的限制[3]。由于生物大分子结构与功能的复杂性，很

多分析软件的输出结果存在偏差，因此通过生物信息学软件对相关基因电子克隆的分析结果仍然需要通过实验来进行验证。本研究为百脉根 cyclophilin 基因的实验克隆及功能验证提供理论基础，同时也为电子克隆技术在中药的育种及基因修饰等方面提供参考。

参考文献

[1] Huminiecki L, Bicknell R. In silico cloning of novel endothelial-specific genes [J]. *Genome Res*, 2000, 10(11): 1796-1806.

[2] Gill R W, Sanseau P. Rapid in silico cloning of genes using expressed sequence tags (ESTs) [J]. *Biotechnol Ann Rev*, 2000, 5: 25-44.

[3] 王冬冬, 朱延明, 李 勇, 等. 电子克隆技术及其在植物基因工程中的应用 [J]. 东北农业大学学报, 2006, 37(3): 403-408.

[4] 张鸭关, 匡崇义, 陈 功. 云南引进帝国百脉根的研究 [J]. 四川草原, 2004, 109(12): 9-11.

[5] 张振霞, 储成才, 符义坤. GA20-氧化酶基因转化豆科牧草百脉根的研究 [J]. 草业学报, 2002, 11(3): 97-100.

[6] 王保明, 谭晓风. 植物亲环素基因的结构, 功能及表达调控 [J]. 中南林业科技大学学报: 自然科学版, 2008, 28(1): 168-174.

[7] 白宇杰, 马华升, 王火旭, 等. 植物细胞亲环素研究进展 [J]. 植物生理学通讯, 2010, 46(9): 881-889.

[8] 王保明, 谭晓风. 植物亲环素基因的结构, 功能及表达调控 [J]. 中南林业科技大学学报: 自然科学版, 2008, 28(1): 168-174.

[9] Iki T, Yoshikawa M, Meshi T, *et al*. Cyclophilin 40 facilitates HSP90-mediated RISC assembly in plants [J]. *EMBO J*, 2012, 31(2): 267-278.

[10] Lee S S, Park H J, Yoon D H, *et al*. Rice cyclophilin OsCYP18-2 is translocated to the nucleus by an interaction with SKIP and enhances drought tolerance in rice and *Arabidopsis* [J]. *Plant Cell Env*, 2015, 38(10): 2071-2087.

[11] Sekhon S S, Kaur H, Dutta T, *et al*. Structural and biochemical characterization of the cytosolic wheat cyclophilin TaCypA-1 [J]. *Acta Crystallogr D*, 2013, 69(4): 555-563.

[12] Chen A P, Wang G L, Qu Z L, *et al*. Ectopic expression of ThCYP1, a stress-responsive cyclophilin gene from Thellungiella halophila, confers salt tolerance in fission yeast and tobacco cells [J]. *Plant Cell Rep*, 2007, 26(2): 237-245.

[13] Kan Y, Liu S, Guo Z, *et al*. Characterization of a cyclophilin cDNA from soybean cells [J]. *Acta Bot Sin*, 2001, 44(2): 173-176.

[14] 吴 莉, 王雅琴, 张新宇, 等. 棉花亲环素基因 GhCYP2 的克隆及实时定量表达分析 [J]. 新疆农业科学, 2011, 48(9): 1569-1575.

[15] 张亚丽, 赵丽萍, 马春雷, 等. 茶树亲环素基因 cDNA 全长的分析鉴定与原核表达 [J]. 茶叶科学, 2007, 27(2): 120-126.

[16] PHuang X, Madan A. CAP3: A DNA sequence assembly program [J]. *Genome Res*, 1999, 9(9): 868-877.

[17] Petersen T N, Brunak S, von Heijne G, *et al*. SignalP 4. 0: discriminating signal peptides from transmembrane regions [J]. *Nat Methods*, 2011, 8(10): 785-786.

[18] Kyte J, Doolittle R F. A simple method for displaying the hydropathic character of a protein [J]. *J Mol Biol*, 1982, 157(1): 105-132.

[19] Nakai K, Horton P. PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization [J]. *Trends Biochem Sci*, 1999, 24(1): 34-35.

[20] 熊 伟, 张晓娟, 张海洋, 等. 基于生物信息学方法预测人线粒体转录终止因子 3 蛋白的结构与功能 [J]. 生物技术通讯, 2015, 26(3): 367-373.

[21] Geourjon C, Deleage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments [J]. *Computer Appl Biosci*: CABIOS, 1995, 11(6): 681-684.

[22] Jensen L J, Gupta R, Blom N, *et al*. Prediction of human protein function from post-translational modifications and localization features [J]. *J Mol Biol*, 2002, 319(5): 1257-1265.

[23] Biasini M, Bienert S, Waterhouse A, *et al*. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information [J]. *Nucl Acids Res*, 2014, 42(W1): 252-258.

[24] 胡 骁, 萧浪涛. 生物信息学在新基因全长 cDNA 电子克隆中的应用 [J]. 生物技术通报, 2007(4): 93-96.

[25] Gill B S, Appels R, Botha-Oberholster A M, *et al*. A workshop report on wheat genome sequencing international genome research on wheat consortium [J]. *Genetics*, 2004, 168(2): 1087-1096.