

文章编号: 1001-1498(2010)03-0349-06

麻疯树磷酸烯酮式丙酮酸羧化酶 *pepc* 基因全长 cDNA 克隆及序列分析

范正琪¹, 李纪元¹, 田敏¹, 李辛雷¹, 陈东亮¹, 卢孟柱²

(1. 中国林业科学研究院亚热带林业研究所, 浙江 富阳 311400; 2. 中国林业科学研究院林业研究所, 北京 100091)

摘要: 应用 RT-PCR 和 RACE 法从麻疯树总 RNA 中分离出 *pepc* 基因全长 cDNA, 长度为 3 142 bp, 阅读框 2 898 bp, 编码 965 个氨基酸, 在 GenBank 中登录, 序列号为 EU069413。它的氨基酸序列与蓖麻、陆地棉、橙、大豆、花生、烟草、油菜、拟南芥的氨基酸同源性分别为 94.94%、92.46%、90.60%、90.50%、90.50%、88.33%、84.61%、82.44%。该基因编码了 *pepc* 基因家族中的 *pepc1*, 蛋白属于 C₃ 型 PEPC。推测了 *pepc* 编码蛋白分子量为 110.6 kD, 并对其稳定性、二级结构、疏水性等特性进行了分析, 最后确定了该蛋白的功能位点和结构域。

关键词: 麻疯树; *pepc* 基因; 克隆

中图分类号: S718.46

文献标识码: A

Cloning and Sequence Analysis of Full-length cDNA of *pepc* Gene from *Jatropha curcas*

FAN Zheng-qi¹, LI Ji-yuan¹, TIAN Min¹, LI Xin-lei¹, CHEN Dong-liang¹, LU Meng-zhu²

(1. Research Institute of Subtropical Forestry, Chinese Academy Forestry, Fuyang 311400, Zhejiang, China;

2. Research Institute of Forestry, Chinese Academy Forestry, Beijing 100091, China)

Abstract: Applying RT-PCR and RACE method, 3 142 bp *pepc* gene was isolated from *Jatropha curcas*. The sequence analysis shows that the *pepc* gene including an open reading frame is 2 898 bp in length, and encodes a protein of 965 amino acid. Its amino acid sequence shares 94.94%, 92.46%, 90.60%, 90.50%, 90.50%, 88.33%, 84.61%, 82.44% identity with those of *pepc* from *Ricinus communis*, *Gossypium hirsutum*, *Citrus sinensis*, *Glycine max*, *Arachis hypogaea*, *Nicotiana tabacum*, *Brassica napus*, *Arabidopsis thaliana*, respectively. The *pepc* gene has been accepted by GenBank, and the accession number of gene sequence is EU069413. The deduced protein molecular weight of *pepc* is 110.6 kD, and belong to *pepc1* in gene family and C₃ form. Its stability, secondary structure, hydrophobicity profile were analysed, and the functional sites and characterized domains were confirmed.

Key words: *Jatropha curcas*; *pepc* gene; cloning

麻疯树 (*Jatropha curcas* L.) 又名小桐子, 为大戟科 (Euphorbiaceae) 麻疯树属 (*Jatropha* L.) 落叶灌木或小乔木, 我国主要分布在广东、广西、云南、贵州、四川、台湾、福建、海南等省区的热带及干热

河谷地区, 为栽培或半野生状态。麻疯树种仁中含油量可达 60% 以上, 是制备生物柴油的理想材料, 同时也是一个优秀的脂肪酸基因供体植物。

磷酸烯醇式丙酮酸羧化酶 (PEPCase, 简称

收稿日期: 2009-03-04

基金项目: 浙江省重大科技专项重点项目 (2005C12003、2006C12030)

作者简介: 范正琪 (1976—), 女, 浙江桐乡人, 助理研究员, 主要从事植物基因工程研究。

PEPC) 主 参与植物 C_4 代谢途径 CO_2 固定作用, 最终提高 C_4 植物的光合效率。目前, 对其在拟南芥 (*Arabidopsis thaliana* L.)^[11]、玉米 (*Zea mays* L.)^[12]、高粱 (*Sorghum vulgare* Pers.)^[13]、蓖麻 (*Ricinus communis* L.)^[14] 等植物和一些 C_4 型草类^[15-17] 中的功能都有较深入的研究。利用生物技术将编码 PEPC 的基因 *pepc* 转入 C_3 植物, 为改善 C_3 植物光合作用提供了一条重要的研究途径。另一方面, 在提高植物脂肪酸含量上, PEPC 近年来也越来越受到研究者的关注。蛋白质是光合作用的主要产物, 据 Sugimoto 等^[8] 报道, 大豆 (*Glycine max* (L.) Merrill) 的油脂含量与蛋白质含量呈高度的负相关, 油菜 (*Brassica napus* L.) 中也存在这种情况, 根据陈锦清等^[9-11] 提出的“底物竞争”假说, PEPC 活性的抑制有利于更多的碳源用于脂肪酸的合成。此后, 应用 *pepc* 基因反义导入植物、微生物调控脂肪酸合成的研究也陆续报道。候李君等^[12] 将蓝藻的 *pepcA* 基因正反义导入大肠杆菌, 结果显示, 转反义插入 *pepcA* 基因的宿主菌脂类含量增加, 蛋白质含量下降, 而转正义 *pepcA* 基因的宿主菌出现了相反的结果, 进一步支持了“底物竞争”假说。此外, 大豆^[13]、水稻 (*Oryza sativa* L.)^[14] 也获得了反义 *pepc* 转基因材料, 提高了籽粒的油脂含量。因此, 通过反义抑制技术控制麻疯树 *pepc* 基因的表达, 是提高麻疯树种子含油量的有效途径。本研究从麻疯树克隆了 *pepc* 基因全长 cDNA, 并对其所编码的氨基酸序列进行了生物信息学分析, 为鉴定麻疯树 *pepc* 基因功能和进一步培育高油麻疯树新品种奠定了前期基础。

1 材料与方 法

1.1 试验材料

植物材料: 麻疯树未成熟种子采自贵州省贞丰县, 为防止种子 RNA 在运输途中降解, 于 5 月至 8 月在原产地摘下果实后迅速剥出种仁, 切碎, 浸泡在 TaKaRa 公司生产的 Sample Protector 中, 邮寄收货后弃保存液, 立即放入 -80 冰箱冷藏。

菌株和质粒: 大肠杆菌 (*Escherichia coli* (T. Escherich 1885)) DH5 购自 TaKaRa 公司, pGEM-T vector 购自 Promage 公司。

试剂与酶: 总 RNA 提取试剂植物 RNAout 购自天泽基因公司, 反转录酶 (AMV)、PCR 试剂、限制

性内切酶、 T_4 DNA 连接酶和其它工具酶购自 TaKaRa 公司, RACE 试剂盒购自 Clontech 公司, PCR 引物由上海生工合成, 凝胶回收试剂盒购自博大泰克公司。

1.2 试验方法

1.2.1 种子 RNA 提取 将种子碎块在液氮中迅速研磨成粉末状, 取 50 ~ 100 mg 转入 1.5 mL 离心管中, 具体步骤按植物 RNAout 操作手册进行。提取完毕后取 5 μ L RNA 用 1% 的琼脂糖凝胶电泳检测。

1.2.2 引物设计与 RT-PCR 扩增 cDNA 第一链合成按 TaKaRa AMV 反应步骤进行, *pepc* 基因 3' 端的 PCR 扩增所用引物根据 GenBank 中麻疯树、绿玉树 (*Euphorbia tirucalli* L.)、陆地棉 (*Gossypium hirsutum* L.)、油菜等基因序列, 利用 Primer 5.0 软件设计: 上游引物: 5' GCAATACCATGGATCTTCG3; 下游引物: 5' TTAACCGGTATTTTGCATTC3。20 μ L 反应体系, 94 预变性 4 min, 94 1 min, 49 1 min, 72 1 min, 经 30 个循环后, 72 延伸 7 min。

1.2.3 5 RACE 扩增全长 cDNA 根据获得的 3' 端 cDNA 片段设计基因特异性引物 (GSP1) 和巢式基因特异性引物 (GSP2)。GSP1: 5' CCAGCAGCAATGCCCTTCATGGTCAA 3; GSP2: 5' TGCTGCACCAAAGC-CAAGCCAGACTG 3。按 Clontech 的 SMARTTM RACE cDNA Amplification Kit 操作手册进行 5' RACE 扩增。

1.2.4 PCR 产物的克隆 利用凝胶回收试剂盒电泳回收 PCR 产物, 并与 TA 克隆载体 pGEM-T vector 连接, 4 过夜, 转化大肠杆菌 DH5 感受态细胞, 在含 X-gal 的 Amp 平板上筛选重组子, 重组子经酶切后作电泳鉴定。

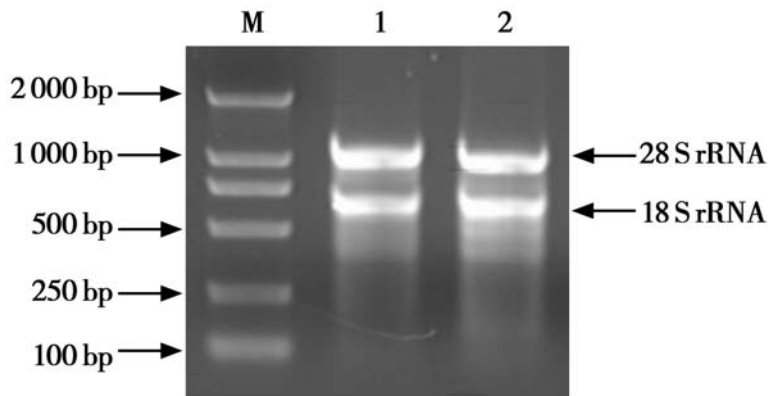
1.2.5 核酸、氨基酸生物信息学分析 提取重组子质粒后送上海生工测序, 用 Primer5.0 分析基因序列, 用 DNAMAN 进行同源性序列比对。运用 NCBI 中保守功能域 (CDD) 搜索程序分析结构功能域。在数据库 (<http://cn.expasy.org/prosite/>) 搜索蛋白质 Motif。

2 结果与分析

2.1 麻疯树未成熟种子总 RNA 提取

植物组织中的 RNA 非常容易降解, 最好用鲜样提取, 作者的实验材料来自贵州省, 因此运输中的样

品保存是个十分关键的问题。经反复试验, TAKARA 的 Sample Protector 可在常温下保存样品 1 周。将麻疯树种仁浸泡在 Sample Protector 中, 经过贵州贞丰至浙江富阳的长途运输, 提取总 RNA, 其 OD_{260} / OD_{280} 为 1.91, 电泳显示 28 S rRNA 和 18 S rRNA 都比较清晰明亮(图 1), RNA 完整性较好。

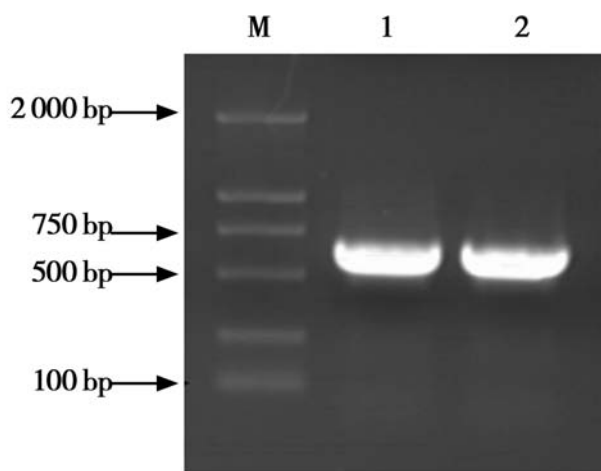


M: 标准分子量; 1、2: 麻疯树总 RNA

图 1 麻疯树种总 RNA

2.2 *pepc* 基因 cDNA 的分离

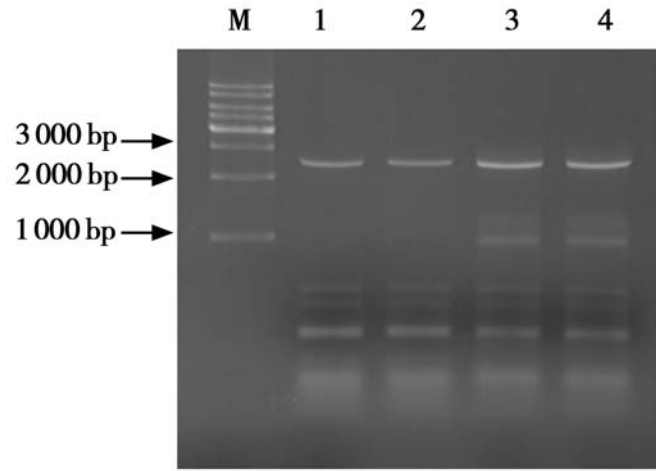
通过两步法 RT-PCR 反应, 电泳后出现一条特异性的亮带(图 2), 大小在 600 bp 左右, 与预期片段一致, TA 克隆到 pGEM-T vector 后测序, 测序结果与其它物种的 *pepc* 基因进行比对, 发现该片段长 597 bp, 为 *pepc* 基因的 3' 末端, 最末端碱基截至到终止密码子。将该片段在 GenBank 中登录, 登录号为 EF158047。



M: 标准分子量 DL2000; 1、2: RT-PCR 产物

图 2 *pepc* 基因 3' 端 RT-PCR 扩增

根据 3' 端的测序结果设计了一对 5' RACE 特异引物 GSP1 和 GSP2, 以总 RNA 为模板, 反转录第 1 链 cDNA, 第 1 次 PCR 反应后, 将反应产物稀释 20 倍, 再进行第 2 次巢式 PCR, 电泳结果显示: 第 1 次反应出现了 1 条分子量较大的带和几条较小的带(图 3), 较大的带在 2 000 ~ 3 000 bp 之间, 经第 2 次 PCR 之后, 较大的带继续存在, 回收克隆到 pGEM-T vector 后测序。



M: 标准分子量; 1、2: 巢式 PCR 反应产物;

3、4: 第一次 PCR 反应产物

图 3 *pepc* 基因 5' RACE 反应

2.3 PCR 产物测序与分析

测序得知该片段长 2 617 bp, 应用 InforMax 软件中的 AlignX 与其他物种进行比对, 表明该片段是麻疯树 *pepc* 基因 cDNA 片段。将 5' 端和 3' 端拼接后, 所得 cDNA 长 3 142 bp, 包括一个 2 898 bp 的开放阅读框, 在起始密码子前还有一段 244 bp 的序列。将核苷酸序列在 NCBI 服务器上用 Blast 搜索软件进行同源基因检索, 结果表明: 与数据库其他植物的磷酸烯醇式丙酮酸羧化酶基因序列有较高的同源性, 选择部分植物的基因运用 DNAMAN 进行多重序列比对, 与蓖麻、油菜、烟草(*Nicotiana tabacum* L.)、拟南芥的序列同源性分别为 89%、75%、79%、77%, 证明克隆的片段为麻疯树的磷酸烯醇式丙酮酸羧化酶基因 cDNA 全长。在 GenBank 中登录, 序列号为 EU069413。

2.4 PEPC 蛋白氨基酸序列同源性及系统进化分析

运用 Primer5.0 软件将 cDNA 序列翻译成氨基酸序列, 结果得到麻疯树 *pepc* 基因编码 965 个氨基酸, 与油菜、大豆、烟草、花生(*Arachis hypogaea* L.)、橙(*Citrus sinensis*(L.) Osbeck)、蓖麻、陆地棉、拟南芥的编码序列进行多重比对分析, 结果表明: 它与蓖麻的氨基酸同源性最高, 可达 94.94%; 其次是陆地棉, 为 92.46%; 与橙、大豆、花生、烟草、油菜、拟南芥的同源性分别为 90.60%、90.50%、90.50%、88.33%、84.61%、82.44%。

对这几个氨基酸序列进行进一步的系统进化分析(图 4), 可发现麻疯树与蓖麻的最先聚合, 接着与陆地棉聚合, 然后与花生和玉米聚合, 与麻疯树 PEPC 进化关系较远的是拟南芥和油菜。在进化树上最先聚合的是亲缘关系较近的物种, 如油菜和拟南芥都是十字花科(Brassicaceae)植物, 因此它们遗

传距离最近,而麻疯树和蓖麻同属于大戟科(Euphorbiaceae),相对于上述其它物种遗传距离也应该是最近的。

此外,在图 4 中可以看出:麻疯树 PEPC 与蓖麻 2 个基因家族中的 PEPC1 先聚合,而其他 2 个成员

PEPC3 和 PEPC4 距离较远,其他含有多个 PEPC 基因家族成员的植物如陆地棉、大豆、拟南芥,它们的 PEPC1 较其他成员而言与本研究克隆的 *pepc* 基因编码的蛋白遗传距离更近,因此,本研究克隆的 *pepc* 基因属于基因家族中的 PEPC1。

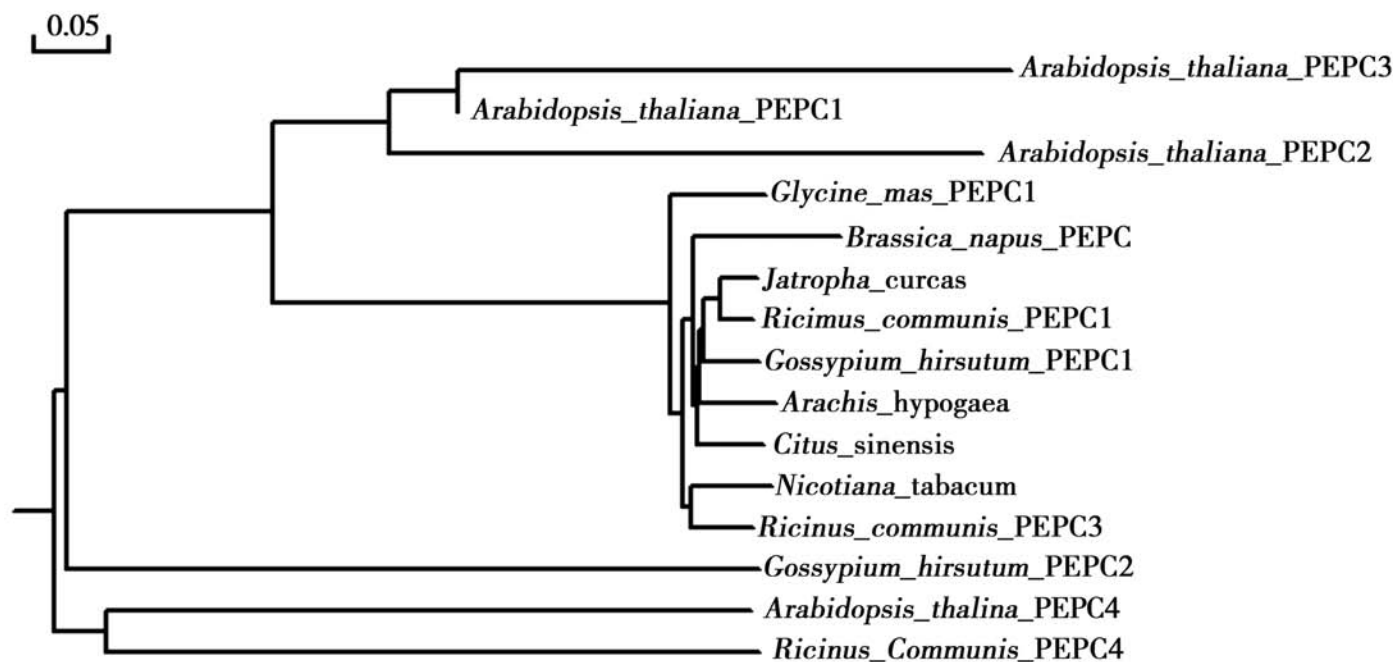


图 4 麻疯树 PEPCase 的进化树

研究表明: C_4 型 PEPC 的 C 末端对应于第 774 位或左右附近的一个氨基酸残基都为丝氨酸(S),而所有 C_3 型植物 PEPC 相应位置均为丙氨酸(A)。本研究克隆的麻疯树 PEPC 在 C 末端第 774 位氨基酸是丙氨酸,表明是一个 C_3 型基因。

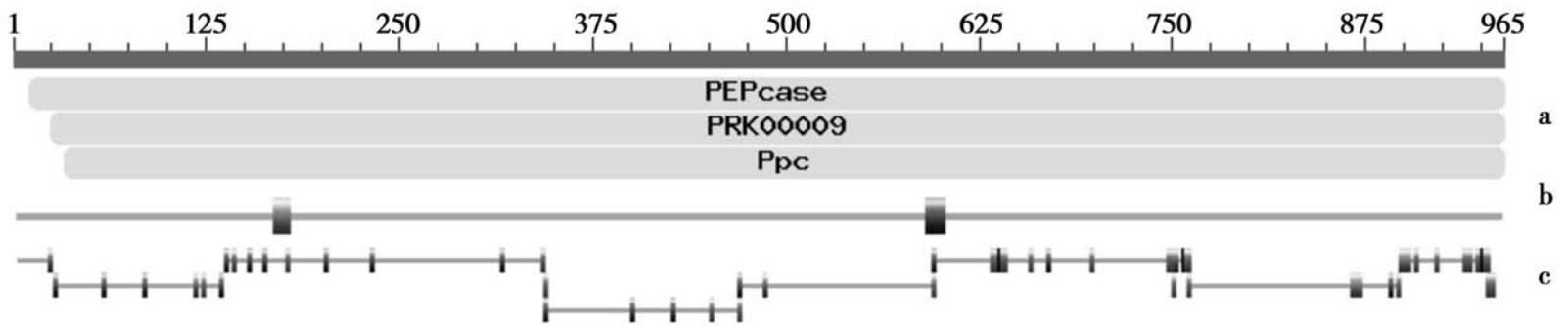
2.5 麻疯树 *pepc* 基因编码的蛋白质特性及功能位点、结构功能域确定

麻疯树 *pepc* 基因编码的蛋白质推测其分子量为 110.6 kD, 该蛋白酸性氨基酸占总氨基酸的 14.6%, 碱性氨基酸占总氨基酸的 13.5%, 理论等电点 6.18, 稳定系数为 46.5, 单条肽链属于不稳定蛋白。用 DNAMAN 分析蛋白的二级结构, 螺旋(-helix, 373 个 AA) 占 38.56%, 折叠片(-pleated sheet, 107 个 AA) 占 11.09%, 无规卷曲(Coil, 485 个 AA) 占 50.26%, 说明麻疯树 *pepc* 基因编码的蛋白质主要以无规卷曲为主, 间或螺旋和折叠。经亲水性/疏水性及跨膜蛋白分析、信号肽分析, 该蛋白无跨膜区及信号肽。

所有的 PEPC 氨基酸序列由 2 个不同的区域构成, 1 个保守区域(C 末端区) 和 1 个可变区域(N 末端区), C 末端有 100 个氨基酸总是保守不变的。本文选择的几个物种 PEPC 都为 C_3 型, 其氨基酸序列具有多个高度保守的序列, 在麻疯树 PEPC 的第 159 至 312 个氨基酸之间这些物种仅有 12 个氨基酸略

有差异, 其它都是不变的保守序列, 另一处较大的保守区在 745 和 789 个氨基酸之间, 除拟南芥以外, 这 45 个氨基酸在这几个物种间是共有的序列。

由于 C 末端的高度保守, 来源不同的 PEPC 长度的差异似乎由在 N 末端或内部区域中添加或缺失额外的序列引起的。有几个高度保守的残基和基序可能与酶的活性位点和调控有关。通过 CDD 搜索, 发现在 11 ~ 965 氨基酸位点之间是一个典型的大 Ppc 结构域(图 5-a)。根据 PROSITE 数据库分析, 其氨基酸序列含有 2 个大的 PEPC 活性位点, 分别位于 168 ~ 179 氨基酸位点(PS00781, VITAHPTQs-vRR) 和 591 ~ 603 氨基酸位点(PS00393, VMIGYS-DSgKDAG)(图 5-b)。搜索蛋白质的 Motif, 发现 46 个功能位点(图 5-c), 分为 7 种模式: 3 个依赖于 cAMP 和 cGMP 的蛋白激酶磷酸化位点(cAMP- and cGMP-dependent protein kinase phosphorylation site, PS00004), 15 个酪蛋白激酶磷酸化位点(Casein kinase phosphorylation site, PS00006), 16 个蛋白激酶 C 磷酸化位点(Protein kinase C phosphorylation site, PS00005), 2 个 N-端糖基化位点(N-glycosylation site, PS00001), 1 个酰胺化位点(Amidation site, PS00009), 7 个 N-端十四烷酰化位点(N-myristoylation site, PS00008), 2 个酪氨酸激酶磷酸化位点(Tyrosine kinase phosphorylation site, PS00007)。



a: PEPC 的保守区; b: 2 个大的 PEPC 活性位点; c: EPC 功能位点

图5 麻疯树 PEPC 蛋白的保守功能结构区域分析

3 讨论

PEPC 是植物特别是 C_4 和 CAM 植物光合作用的关键酶,而在 C_3 植物中,该酶为 Krebs 循环补充四碳二羧酸水平起着回补反应的功能^[15]。菊科 (Asteraceae) 黄菊属 (*Flaveria* Juss.) 植物是研究 C_4 PEPC 进化的一个独特例子,黄菊属的一些物种分别含有 C_3 、 C_4 及 C_3 - C_4 中间型 PEPC,研究证明了 C_4 PEPC 基因是由 C_3 PEPC 基因进化而来的^[16]。已有大量的研究投入在应用 *pepc* 基因的超表达提高 C_3 植物如水稻^[17-18]、小麦^[19-20] 等经济作物的光合速率,以期达到提高产量的目的。

目前,麻疯树 *pepc* 基因只克隆到了 3' 末端的片段^[21],*pepc* 基因在 C 末端的高度保守为 *pepc* 基因的克隆提供了可靠的序列信息,本研究是根据多个物种 *pepc* 基因的比对后,首先克隆了麻疯树 *pepc* 基因 3' 末端的 cDNA 序列片段,然后才进行 5' RACE 最后获得全长的。它的氨基酸序列在进化树中最先与同一个科大戟科的蓖麻聚合,遗传距离最近,然后才与锦葵科 (Malvaceae) 的落地棉、芸香科 (Rutaceae) 的橙、豆科 (Leguminosae) 的大豆和花生聚合,与十字花科的油菜和拟南芥距离较远,表明 PEPC 在进化过程中既具有较强的保守性,又存在不同的种属间的差异。

籽粒脂肪酸含量的提高一直是油料植物分子育种的难点。自从提出“底物竞争”假说以后,人们开始利用 *pepc* 基因的反义抑制使光合作用产物从蛋白质方向到脂肪酸方向转化,从而提高脂肪酸的含量,已经获得成功并得到高油新品种的有油菜^[11]、大豆^[13]、水稻^[14] 等作物。本研究以麻疯树未成熟种子总 RNA 为模板,利用 RT-PCR 和 5' RACE 克隆了 *pepc* 基因全长 cDNA,而且作者已经构建了麻疯树 *pepc* 基因的正义、反义表达载体,试图将该基因正向、反向插入植物基因组,为“底物竞争”假说奠定坚实的实验基础,并为培育高油麻疯树新品种提供参考依据。

参考文献:

- [1] Gregory A L, Hurley B A, Tran H T, *et al.* In vivo regulatory phosphorylation of the phosphoenolpyruvate carboxylase AtPPC1 in phosphate-starved *Arabidopsis thaliana* [J]. *Biochem J*, 2009, 420(1): 57 - 65
- [2] Takahashi-Terada A, Kotera M, Ohshima K, *et al.* Maize phosphoenolpyruvate carboxylase: mutations at the putative binding site for glucose 6-phosphate caused desensitization and abolished responsiveness to regulatory phosphorylation [J]. *J Biol Chem*, 2005, 280(12): 11798 - 11806
- [3] Cousins A B, Adam N R, Wall G W, *et al.* Development of C_4 photosynthesis in sorghum leaves grown under free-air CO_2 enrichment (FACE) [J]. *Journal of Experimental Botany*, 2003, 54(389): 1969 - 1975
- [4] Gennidakis S, Rao S, Greenham K, *et al.* Bacterial- and plant-type phosphoenolpyruvate carboxylase polypeptides interact in the heterooligomeric Class-2 PEPC complex of developing castor oil seeds [J]. *Plant J*, 2007, 52(5): 839 - 849
- [5] Alvarez R, Castillo J M, Mateos-Naranjo E, *et al.* Ecotypic variations in phosphoenolpyruvate carboxylase activity of the cordgrass *Spartina densiflora* throughout its latitudinal distribution range [J]. *Plant Biology*, 2009, 12(1): 154 - 160
- [6] Wang X, Gowik U, Tang H, *et al.* Comparative genomic analysis of C_4 photosynthetic pathway evolution in grasses [J]. *Genome Biol* 2009, 10(6): 68
- [7] 张桂芳, 赵明, 丁在松, 等. 稗草磷酸烯醇式丙酮酸羧化酶 (PEPCase) 基因的克隆与分析 [J]. *作物学报*, 2005, 31(10): 1365 - 1369
- [8] Sugimoto T, Tanaka K. Phosphoenolpyruvate carboxylase level in soybean seed highly correlates to its contents of protein and lipid [J]. *Agric Biol Chem*, 1989, 53(3): 885 - 887
- [9] 陈锦清, 黄锐之, 郎春秀, 等. 油菜 PEP 基因的克隆及 PEP 反义基因的构建 [J]. *浙江大学学报: 农业与生命科学版*, 1999, 25(4): 365 - 367
- [10] 郎春秀, 胡张华, 刘智宏, 等. 油菜农杆菌转基因体系的建立及转 PEP 反义基因油菜的获得 [J]. *浙江农业学报*, 1999, 11(2): 55 - 58
- [11] 陈锦清, 郎春秀, 胡张华, 等. 反义 PEP 基因调控油菜籽粒蛋白质/油脂含量比率的研究 [J]. *农业生物技术学报*, 1999, 7(4): 316 - 320
- [12] 候李君, 施定基, 蔡泽富, 等. 蓝藻正反义 *pepcA* 基因导入对大肠杆菌中脂类合成的调控 [J]. *中国生物工程杂志*, 2008, 28

- (5): 52 - 58
- [13] 赵桂兰, 陈锦清, 尹爱萍, 等. 获得转反义 PEP 基因超高油大豆新材料[J]. 分子植物育种, 2005, 3(6): 792 - 796
- [14] 吴关庭, 郎春秀, 胡张华, 等. 应用反义 PEP 基因表达技术提高稻米脂肪含量[J]. 植物生理与分子生物学学报, 2006, 32(3): 339 - 344
- [15] 郝乃斌, 戈巧英, 杜维广. 大豆高光效育种光合生理研究进展[J]. 植物学通报, 1991, 8(2): 13 - 19
- [16] Westhoff P, Gowik U. Evolution of C₄ phosphoenolpyruvate carboxylase genes and proteins: a case study with the genus *Flaveria*[J]. *Annals of Botany*, 2004, 93: 13 - 23
- [17] 张边江, 华 春, 周 峰, 等. 转 PEPC + PPKK 双基因水稻的光合特性[J]. 中国农业科学, 2008, 41(10): 3008 - 3014
- [18] 丁在松, 赵 明, 荆玉祥, 等. 玉米 *ppc* 基因过表达对转基因水稻光合速率的影响[J]. 作物学报, 2007, 33(5): 717 - 722
- [19] 张 彬, 丁在松, 张桂芳, 等. 根癌农杆菌介导获得稗草 *Ecppc* 转基因小麦的研究[J]. 作物学报, 2007, 33(3): 356 - 362
- [20] 陈绪清, 张晓东, 梁荣奇, 等. 玉米 C₄ 型 *pepc* 基因的分子克隆及其在小麦的转基因研究[J]. 科学通报, 2004, 49(19): 1976 - 1982
- [21] 王治涛. 麻疯树 (*Jatropha curcas* L.) 胚乳 cDNA 文库的构建及磷酸烯醇式丙酮酸羧化酶基因克隆的研究[D]. 成都: 四川大学, 2006