

文章编号: 1673-1689(2011)05-0728-06

南方红豆杉 GGPP 合酶基因的克隆与 生物信息学分析

韩飞^{1,2}, 康林芝^{1,2}, 郭丽琼^{1,2}, 陈晓阳², 林俊芳^{* 1,2}

(1. 华南农业大学 食品学院, 广东 广州 510640; 2. 华南农业大学 生物质能研究所, 广东 广州 510640)

摘 要: 根据 GeneBank 中公布的 GGPPS 基因保守区设计特异引物, 克隆南方红豆杉 GGPPS 基因的 DNA 和 cDNA 序列, 并通过生物信息学方法对其核苷酸序列和蛋白质结构进行分析预测。结果表明, GGPPS 基因 DNA 和 cDNA 序列都为 1 182 bp, DNA 序列中无内含子, cDNA 序列为 1 个编码 393 个氨基酸残基的开放式阅读框; GGPPS 的理论相对分子质量为 42 590, 等电点为 5.68。核苷酸同源性分析显示, 它与 *Taxus canadensis* (AF081514) 同源性为 98.8%; 推导出的氨基酸序列与 *Taxus canadensis* (AAD16018) 同源性达 99.0%。利用 ProtParam、SignalP、ProtScale、SOPMA、Swiss-Modeling、Scan Prosite 和 MEGA4.0 等生物信息学工具分别对其理化性质、信号肽、疏水性、亲水性、二级结构、三级结构和进化树进行分析, 为其功能研究和生物合成紫杉醇研究提供分子基础。

关键词: 南方红豆杉; GGPPS; 克隆; 序列分析

中图分类号: Q 785

文献标识码: A

Molecular Cloning and Sequence Analysis of Geranylgeranyl Pyrophosphate Synthase from *Taxus wallichiana* var. *Mairei*

HAN Fei^{1,2}, KANG Lin-zhi^{1,2}, GUO Li-qiong^{1,2},
CHEN Xiao-yang², LIN Jun-fang^{* 1,2}

(1. College of Food Science, South China Agriculture University, Guangzhou 510640, China; 2. Institute of Biomass Energy, South China Agriculture University, Guangzhou 510640, China)

Abstract: In this study, the DNA sequence and cDNA sequence of ggpps were cloned from *Taxus wallichiana* var. *Mairei*. Results showed that the length of both DNA and cDNA sequences is 1182 bp, without introns, contained an opening reading frame of 1182 bp, which coded for 393 amino acid residues; theoretical molecular weight is 42.59 kDa, and isoelectric point is 5.68. Alignment showed that cDNA sequence exhibited 98.8% similarity to the ggpps gene of *Taxus canadensis* (AF081514) which has been reported in NCBI, and its protein sequence has 99.0% similarity to GGPPS of *Taxus canadensis* (AAD16018). The bioinformatics of GGPPS, including

收稿日期: 2010-10-10

基金项目: 国家自然科学基金项目(31071837)。

作者简介: 韩飞(1984-), 男, 安徽肥东人, 食品生物技术博士研究生。Email: fei.han@163.com

* 通信作者: 林俊芳(1962-), 男, 福建莆田人, 农学博士, 研究员, 博士研究生导师, 主要从事生物活性物质研究与开发方面的研究。Email: linjf@scau.edu.cn; junfanglin2003@yahoo.com.cn

physical and chemical properties, the signal peptide, hydrophobicity, hydrophilicity, secondary structure, tertiary structure, and phylogenetic trees, were analyzed with bioinformatics softwares and database, such as ProtParam, SignalP, ProtScale, SOPMA, Swiss-Modeling, Scan Prosite, MEGA4.0 and so on.

Key words: *Taxus wallichiana* var. *Mairei*; GGPPS; gene cloning; sequence analysis

紫杉醇是从红豆杉属植物中分离出的一种二萜类生物碱,因具有广谱的抗癌活性和独特的抗癌机制而受人青睐。自紫杉醇的抗癌功效和作用机制被确定以来,尤其是获得批准用于临床治疗卵巢癌和乳腺癌等治疗中,人们就希望通过研究紫杉醇生物合成过程的某些关键性酶的发现、克隆和高效表达,实现紫杉醇高效、稳定生产。紫杉醇的合成首先是通过=牛儿基=牛儿基焦磷酸(Geranylgeranyl pyrophosphate, GGPP)的环化建立紫杉烷类物质的骨架结构 taxane 4(5), 11(12)-diene, 然后进一步经过氧化修饰和侧链添加,最终得到具有活性的紫杉醇^[1]。而 GGPP 是由=牛儿基=牛儿基焦磷酸合酶(Geranylgeranyl pyrophosphate synthase, GGPPS) (EC: 2.5.1.29) 催化异戊烯焦磷酸与甲基丙烯基焦磷酸生成的。Gregroy 等^[2]研究发现,GGPPS 活性和紫杉醇生成量具有线性相关性。GGPPS 虽不是催化紫杉醇生物合成过程中的限速步骤,但它在紫杉醇生物合成代谢中的功能仍十分重要,因为维持高水平的 GGPP 供应能力将会为位于此酶下游的限速酶提供充足的底物,因而 GGPPS 被认为是红豆杉生物合成紫杉醇过程中的关键酶之一^[3]。

GGPPS 的克隆和表达特征的分析对于阐明紫杉醇生物合成机理及其基因工程的应用非常重要。作者通过在 GenBank 数据库进行同源比对,并通过 CLUSTX 软件分析,获得保守性区域,然后以 9 年生南方红豆杉(*Taxus wallichiana* var. *Mairei*) 植株为材料,提取其 DNA 和 RNA 作为模板,采用 primer 5 软件设计引物,经 PCR 获得 GGPPS 基因全长序列,并对其 DNA 序列和氨基酸序列进行了分析,为进一步阐明紫杉醇生物合成途径中的分子调控机理以及构建高产紫杉醇的转基因菌株奠定了基础。

1 材料与方法

1.1 供试材料

9 年生南方红豆杉(*Taxus wallichiana* var. *Mairei*) 购自广东省韶关市金山地红豆杉科技有

限公司; *E. coli* Top10: 由华南农业大学生物质能研究所保藏。

1.2 酶与试剂

ExTaq 酶、dNTP、Taq DNA、PrimeScript™ RT-PCR Kit、*EcoR* I 限制性内切酶: 购自大连宝生物工程有限公司; DNA 凝胶回收试剂盒: 购自北京天跟生物公司; pGEM-T Easy 载体: 购自 Promega 公司; Plant RNAout 提取试剂盒: 购自北京天恩泽基因科技有限公司; PCR 引物: 由上海生物工程公司合成; DNA 测序: 由华大基因完成; 其他试剂: 均为国产分析纯。上游引物(P₁): 5'-GACTAGT-CATGGCTTACACGGC AATGGCAG-3'; 下游引物(P₂): 5'-CATGTACATCAGTTTTGCCT-GAATGCAATGTAAATG-3'。

1.3 实验方法

1.3.1 南方红豆杉总 DNA 和 RNA 的提取 取南方红豆杉的幼嫩叶片,液氮迅速冷冻后研磨成粉末,用改良的 CTAB 法提取总 DNA^[4],参照天泽基因的 Plant RNAout 试剂盒说明书提取总 RNA。用微量紫外分光光度计测定 DNA 和 RNA 浓度,选取 A₂₆₀/A₂₈₀ 值在 1.8~2.0 的样品于 -20℃ 保存备用。

1.3.2 GGPPS 全长序列的克隆 以南方红豆杉的总 DNA 为模板进行 PCR 扩增。PCR 反应体系含有: 10× buffer 5 μL, 200 μmol/L dNTPs, 2.0 mmol/L MgCl₂, Taq 酶 2.5 U, P₁ 10×10⁻⁹ mol, P₂ 10×10⁻⁹ mol, DNA 模板 100 ng, 最后用无菌超纯水补足至 50 μL。扩增程序为: 94℃ 预变性 5 min, 94℃ 40 s, 55℃ 45 s, 72℃ 2 min, 35 个循环, 72℃ 10 min。PCR 扩增后,产物经 1.0 g/dL 琼脂糖凝胶电泳检测。

以南方红豆杉的总 RNA 为模板,参照 Prime-Script RT-PCR kit 的说明书,反转录合成 cDNA 后,用设计的特异引物 P₁ 和 P₂, 进行 RT-PCR。RT-PCR 反应体系: 10× PCR buffer II 5 μL, 400 μmol/L dNTPs, TaKaRa Ex Taq HS 2.5 U, P₁ 10×10⁻⁹ mol, P₂ 10×10⁻⁹ mol, cDNA 模板 5 μL, 最后用无菌超纯水补足至 50 μL。扩增程序为: 94℃

预变性 5 min, 94 ℃ 40 s, 55 ℃ 45 s, 72 ℃ 2 min, 进行 35 个循环, 72 ℃ 10 min。RT-PCR 扩增后, 产物经 1.0 g/dL 琼脂糖凝胶电泳检测。

1.3.3 目的片段的回收、连接和转化 采用胶回收试剂盒回收 PCR 扩增中的特异条带, 将 PCR 产物与 pGEM-T Vector 连接, 连接反应条件参见 pGEM-T Vector 操作手册。然后将连接产物转化 *E. coli* Top10 感受态细胞。

1.3.4 重组质粒的鉴定与序列测定 分别挑选白色单菌落, 接种于 LB/Amp 液体培养基中, 37 ℃ 剧烈振荡过夜。质粒提取方法参照天根的碱裂解法小量制备质粒 DNA 说明书。以重组质粒 DNA 为模板, P₁和 P₂为引物, 进行 PCR 鉴定; 用 *EcoR* I 单酶切对重组质粒进行酶切鉴定。将阳性质粒提交华大基因公司进行序列测定。

1.3.5 序列分析 利用 DNASTAR 软件中 Edit-Seq 查找 GGPPS 基因的 ORF, 并推测出氨基酸序列; 利用 Blast 工具分别对获得的核苷酸序列和推测的氨基酸序列进行比对, 并利用 Spidey 工具分析该基因的内含子和外显子。用 Protparam 分析蛋白质的一级结构及其等电点、相对分子质量和半衰期; 用 SignalP 工具进行信号肽预测; 用 TargetP 进行导肽分析; 用 ProtScale 进行蛋白的疏水性分析; 用 SOPMA 和 Predictprotein 预测其二级结构; 用 SWISS-MODEL 工具自动建模分析预测其三维结构; 用 MEGA 4 软件将推导的氨基酸序列与其他不同物种来源的 GGPPS 序列进行多重比对, 并构建系统进化树, 预测进化关系。

2 结果与分析

2.1 南方红豆杉总 DNA 和 RNA 的提取

CTAB 法提取南方红豆杉基因组 DNA 的结果见图 1。从图 1a 中可以看出, 南方红豆杉基因组 DNA 序列大于 23 kb, 提取的基因组 DNA 条带清晰, 无严重拖尾降解, 蛋白质及杂质含量较少, A₂₆₀/A₂₈₀ 值在 1.8~2.0 之间, 该样品纯度较高。

用 RNAout 抽提试剂盒抽提南方红豆杉总 RNA(图 1b), 28S 和 18S 的 RNA 条带清晰, 比例适当, 没有明显的降解痕迹, 说明 RNA 样品完整性良好。A₂₆₀/A₂₈₀ 值>1.9, RNA 纯度较高。因此, 所获得的 DNA 和 RNA 在纯度和完整性方面均符合要求, 可用于后续试验。

2.2 GGPPS 基因全长序列的克隆

以南方红豆杉嫩叶中提取总 DNA 和总 RNA 为模板, 用设计的引物 P₁和 P₂进行 PCR/RT-PCR

扩增, 经凝胶电泳可见 PCR 产物均接近 1 200 bp, 见图 2。其片断大小与预期特异条带相符, 且 PCR 扩增产物纯度较高, 无其它非特异条带。

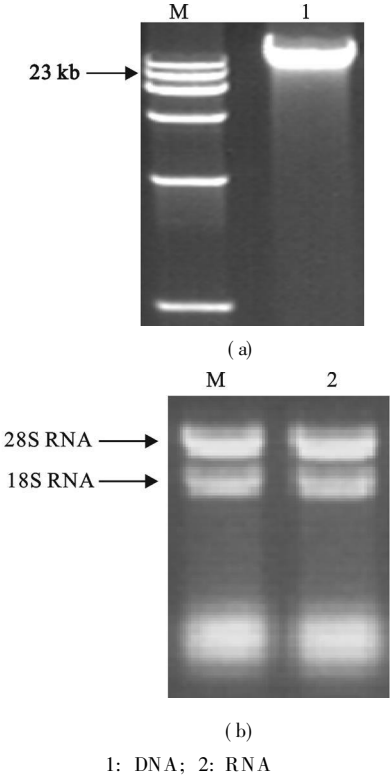
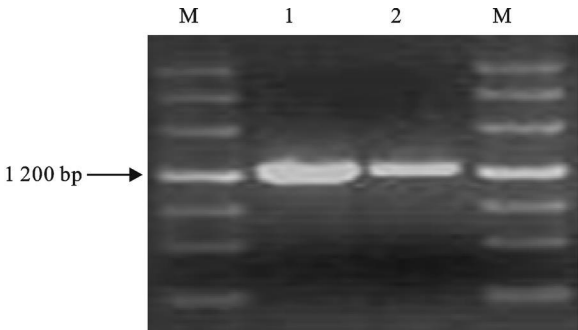


图 1 南方红豆杉总 DNA(a) 和 RNA(b) 的电泳分析
Fig. 1 Electrophoresis analysis of DNA and RNA of *Taxus chinensis* var. *mairei*



1: *ggpps* gene segment; 2: *ggpps* cDNA segment
图 2 南方红豆杉 *ggpps* 的 RCP 扩增和 RT-PCR 扩增
Fig. 2 Products of PCR and RT-PCR of *ggpps* from *Taxus chinensis* var. *mairei* M: Molecular mass standards

2.3 阳性克隆质粒的鉴定

将 PCR 产物和 RT-PCR 产物分别回收纯化后, 与 pGEM-T Easy 载体连接, 转化 *E. coli* Top10 感受态细胞, 利用蓝白斑菌落筛选阳性重组克隆 pGEM-*ggpps* 和 pGEM-*ggppsc*。阳性克隆质粒经 *EcoR* I 单酶切鉴定, 都出现一条带约 1 200 bp, 另一条约为 3 000 bp, 见图 3。对阳性克隆的质粒 PCR 鉴定, 扩增片段约为 1 200 bp。

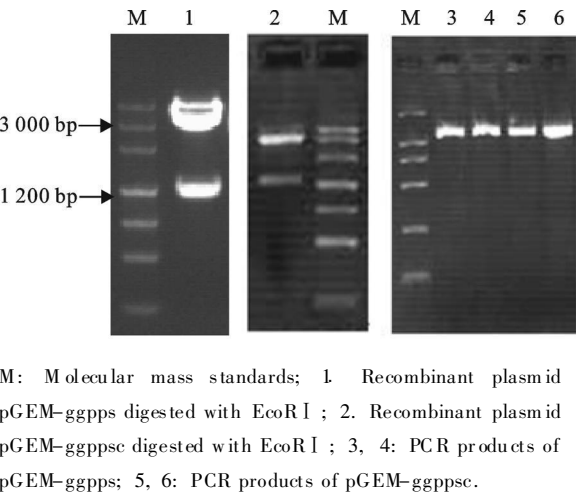


图 3 重组质粒的酶切鉴定

Fig. 3 Digestion and identification of recombinant plasmid

2.4 ggpps 序列比对

测序结果表明, ggpps 基因全长为 1 182 bp, 见

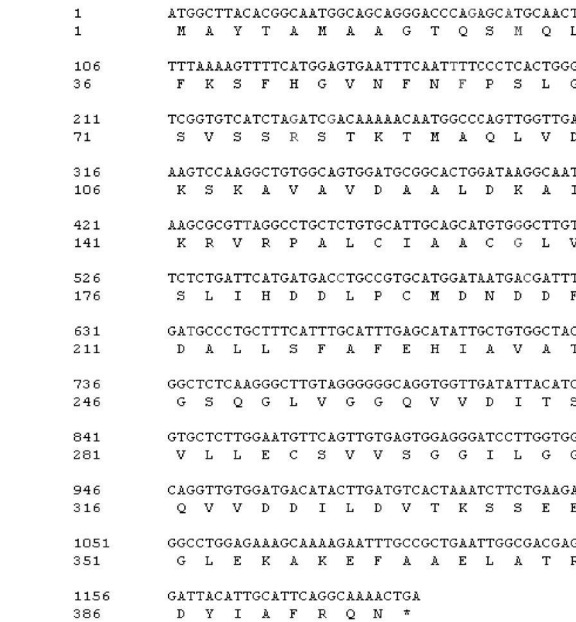


图 4 ggpps cDNA 序列及其推导的氨基酸序列

Fig. 4 ggpps nucleotide sequences and its deduced amino acid sequence

2.5.2 信号肽预测和分析 分泌蛋白质及细胞膜蛋白质均以前体物质多肽的形式合成, 其 N 末端含有作为通过膜时信号的氨基酸序列, 这种氨基酸序列称信号肽或信号序列。由 15~ 25 个氨基酸组成。用 SignalP 3.0 Server 分析此蛋白质, 信号肽可能性值为 0.041, 信号结合可能性值为 0, 得知 ggpps 基因表达的蛋白质不是分泌蛋白质。

2.5.3 亚细胞定位和分析 在核糖体上合成的蛋白质或者留在细胞质中, 或者运送到其他细胞器

图 4。经 BLAST 比对, 其与公布的 ggpps 基因 (GenBank Accession No: AF081514) 的同源性是 98.8%, 共有 14 个碱基存在差异, 其中 4 个碱基影响其蛋白质编码。克隆的 ggpps 基因包含一个开放式阅读框 (ORF), 编码含有 393 个氨基酸残基的蛋白质 (含起始 Met)。将推测的氨基酸序列进行 BLAST, 结果显示与其它红豆杉品种的 GGPPS 序列高度同源, 其中与 Genbank 中的 GGPPS (AAD16018) 序列的同源性为 99.0%。克隆的 ggpps DNA 序列和 cDNA 序列经 Spidey 工具分析发现, 克隆的南方红豆杉 ggpps 基因内部无内含子。

2.5 GGPPS 的生物信息学分析

2.5.1 理化性质预测 ProtParam 工具预测表明: GGPPS 蛋白质的等电点和相对分子质量大小分别为 5 680 和 42 590, 半衰期理论值为 30 h, 不稳定参数为 41.65, 属于不稳定蛋白质, 这也许可能使是功能尚不完全明确的原因之一。

中, 并形成有功能的空间构象后才能发挥其功能。导肽是一段引导新合成的肽链进入细胞器的识别序列。因此, 预测和分析导肽, 对正确认识蛋白质的亚细胞定位和功能作用的途径与机制有一定的意义。为此, 利用 TargetP 工具预测此蛋白质的亚细胞定位结果表明: 此蛋白质为叶绿体转运蛋白质 (分值为 0.174) 和分泌蛋白质 (分值为 0.046, 这与信号肽分析结果一致) 的可能性较小, 推测该蛋白质可能为定位于线粒体内 (分值为 0.276)。

2 5 4 疏水性和亲水性的预测与分析 氨基酸是蛋白质的构件分子,其亲水性、疏水性在形成和保持蛋白质的三级结构上起作用。通过了解肽链中不同肽段的疏水性,还可以对蛋白质的跨膜结构域进行预测。因此,疏水性、亲水性的预测和分析,为蛋白质次级结构的预测及功能分析提供理论参考。采用 ProtScale 分析此蛋白质的疏水性、亲水性,结果见图 5。第 195 位的脯氨酸残基的亲水性最强(分值为 2.911),第 151 位的丙氨酸残基的疏水性最强(分值为 2.722)。总体来看,亲水性氨基酸较为均匀的分布在整条肽链中,且略多于疏水性氨基酸中。因此,除了局部的疏水区域,整个肽链表现为弱亲水性,由此,可推测是此酶是亲水蛋白质。

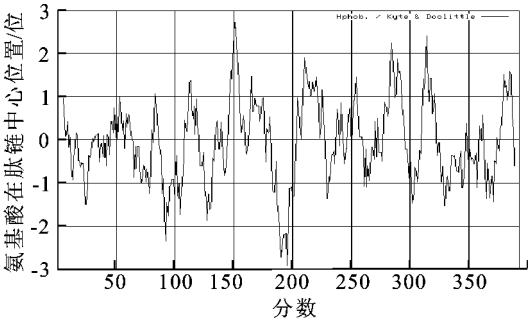


图 5 *ggpps* 基因编码蛋白氨基酸序列的疏水性/亲水性预测

Fig. 5 Predicted hydrophobicity and hydrophilicity of the amino acid sequence of GGPPS

2 5 5 二级结构的预测和分析 多肽链借助氢键排列成沿一维方向而呈现有规则的重复构象的二级结构,是氨基酸顺序与三维构象之间的桥梁。二级结构借助范德华力、氢键、静电和疏水等相互作用形成蛋白质的三级结构,从而发挥正常的生物学功能。用 SOPMA 对此蛋白质的二级结构预测见图 6。此蛋白质 α 螺旋占 60.05%, β 折叠占 4.07%, β 转角占 6.11%,无规则卷曲占 29.77%,和延伸链占 21.71%。

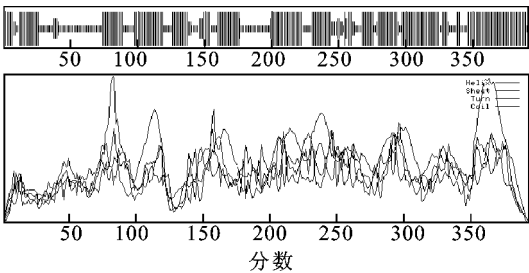


图 6 南方红豆杉 *ggpps* 基因编码蛋白的二级结构预测

Fig. 6 Protein secondary structure prediction of the amino acid sequence of GGPPS

2 5 6 三级结构预测与分析 利用 SWISS-MOD-

EL 同源建模,得到南方红豆杉 GGPPS 氨基酸序列的预测三级结构,见图 7。与预测蛋白质三级结构匹配度最高的是薄荷的 GGPPS(Parent PDB: Chain: D),相似度为 67.80%。从图 7 中可以更直观的看出,预测结构中含有较多的 α 螺旋和无规则卷曲,这与用二级结构的预测结果一致。

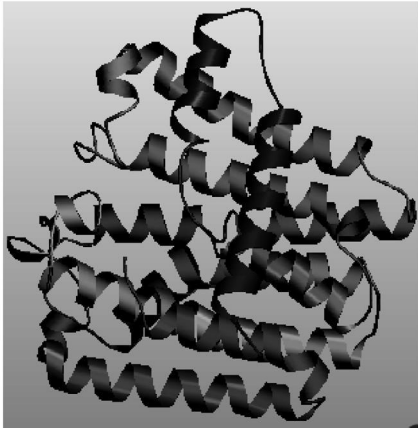


图 7 GGPPS 蛋白三级结构预测图

Fig. 7 Three-dimensional structure of the putative GGPPS

2 5 7 结构域分析 将序列提交至 Scan Prosite 数据库 (<http://www.Expasy.org/scanprosite>),进行蛋白质结构域分析预测,结果显示:它含有多个与蛋白质的修饰作用有关的活性位点,包括 9 处蛋白激酶 C 的磷酸化位点(26-28, 73-75, 76-78, 89-91, 131-133, 225-227, 231-233, 234-236, 341-343);1 处酪蛋白激酶位点(127-134);7 处 N 豆蔻酰化位点(41-46, 50-55, 153-158, 249-254, 290-295, 291-296, 351-356);5 处为酪蛋白激酶 II 磷酸化位点(20-23; 260-263; 297-300; 327-330; 371-374);1 处酰胺化位点(139-142)。同时,利用 Sanger 中心的 Pfam 同源蛋白家族数据库分析,此蛋白质中含有 2 个结构域(128-385 和 129-375),属于聚戊烯合成酶家族。

2 6 同源性和分子进化分析

在序列比对的基础上,采用 MEGA4.0 软件的 NJ(neighbor-joining)法构建系统进化树,以确定南方红豆杉 GGPPS 在系统进化中的地位,见图 8。由 NJ 法构建的系统树^[5],枝长代表分歧度,枝上的数字为 Bootstrap 10000 次的支持率。中可以看出,克隆的南方红豆杉与其它的红豆杉属植物聚为一小类(IV),与云杉属和冷杉属物种的 GGPPS 相似性较近。红豆杉属、云杉属和冷杉属中不同物种的 GGPPS 在进化上具有较近的亲缘关系,这可以为 3 个属种的其它物种的 GGPPS 克隆与后续研究提供有价值的参考。

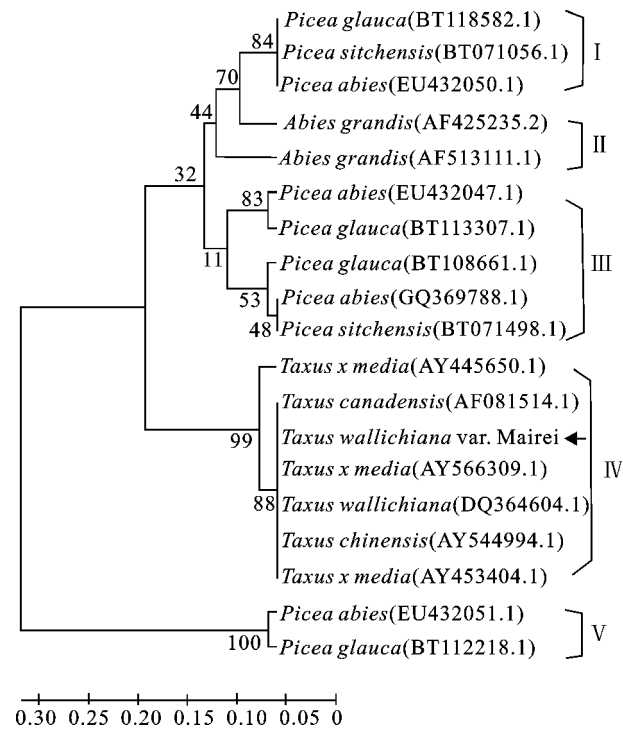


图 8 GGPPS 氨基酸与相似蛋白的同源树

Fig. 8 Homology phylogeny tree about amino acid sequences from ggpps and some similar protein sequences

3 结 语

GGPPS 属异戊烯基转移酶, 催化 FPP 与 IPP 发生亲电子耦合形成 GGPP。虽然该酶促反应并非紫杉醇生物合成的限速步骤, 但它合成了紫杉烷类的共同前体物 GGPP, 并成为二萜和四萜生物合成途径中的分支点, 而且启动了紫杉醇的生物合成。因此, 研究 GGPPS 对于阐明南方红豆杉产生紫杉醇的机理具有重要意义。作者以南方红豆杉为材料, 克隆获得了其 GGPPS 基因的全长序列, 用推断的蛋白质序列作同源比较, 结果显示它与不同物种来源的 GGPPS 都有同源性, 尤其是与 NCBI 中公布的加拿大红豆杉 GGPPS 的同源性高达 99.0%。本研究还发现, 南方红豆杉的 GGPPS 基因中没有插入序列, 这与一些报道的曼地亚红豆杉研究结果相同^[6], 但与其他报道的红豆杉属、云杉属和冷杉属植物的 GGPPS 基因则不同^[7-9], 这说明对南方红豆杉中 GGPPS 基因的研究还需进一步深入。

参考文献 (References):

[1] Walker K, Croteau R. Molecules of interest taxol biosynthetic genes[J]. **Phytochemistry**, 2001, 58(1): 1- 7.

[2] Gregory L, Camiel F D, Mondher J, et al. Geranylgeranyl diphosphate synthase activity and taxane production in *Taxus baccata* cells[J]. **Phytochemistry**, 1999, 50(1): 939- 946.

[3] Hefner J, Ketchum E B R, Croteau R. Cloning and functional expression of a cDNA encoding geranylgeranyl diphosphate synthase from *taxus canadensis* and assessment of the role of this prenyltransferase in cells induced for taxol production [J]. **Archives of Biochemistry and Biophysics**, 1998, 360(1): 62- 74.

[4] 茹文明, 秦永燕, 张桂萍, 等. 濒危植物南方红豆杉遗传多样性的 RAPD 分析[J]. 植物研究, 2008, 28(6): 698- 704. RU Wen-ming, QIN Yong-yan, ZHANG Gui-ping, et al. Genetic diversity of rare and endangered plant *Taxus chinensis* var. *mairei* [J]. **Bulletin of Botanical Research**, 2008, 28(6): 698- 704. (in Chinese)

[5] 王玲军, 陶妍, 党静, 等. 鲢鱼骨骼肌轻酶解肌球蛋白基因的 cDNA 克隆[J]. 食品与生物技术学报, 2010, 29(4): 609- 616. WANG Ling-jun, TAO Yan, DANG Jing, et al. cDNA cloning of silver carp fast skeletal light meromyosin gene[J]. **Journal of Food Science and Biotechnolog**, 2010, 29(4): 609- 616. (in Chinese)

[6] Liao Z H, Gong Y F, Kai G Y, et al. An intron-free methyl jasmonate inducible geranylgeranyl diphosphate synthase gene from *Taxus media* and its functional identification in yeast[J]. **Molecular Biology**, 2005, 39(1): 11- 17.

[7] Schmidt A, Gershenzon J. Cloning and characterization of isoprenyl diphosphate synthases with farnesyl diphosphate and geranylgeranyl diphosphate synthase activity from Norway spruce (*Picea abies*) and their relation to induced oleoresin formation[J]. **Phytochemistry**, 2007, 68(21): 2649- 2659.

[8] Schmidt A, Gershenzon J. Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (*Picea abies*) [J]. **Phytochemistry**, 2008, 69(1): 49- 57.

[9] Burke C, Croteau R. Geranyl diphosphate synthase from *Abies grandis*: cDNA isolation, functional expression, and characterization[J]. **Archives of Biochemistry and Biophysics**, 2002, 405(1): 130- 136.