

doi:10.3969/j.issn.1672-5565.2016.03.01

牦牛儿基牻牛儿基焦磷酸合成酶(GGPS) 的生物信息学分析

张鹏飞,王 宁,周 蔓,李 萍*
(西南交通大学生命科学与工程学院,成都 610031)

摘 要:GGPS 是萜类化合物合成的一个重要分支点酶,同时也是合成 GGPP 的催化酶。本课题共选择了 9 个不同科的植物,对其 GGPS 核苷酸及氨基酸序列的理化性质、蛋白质结构功能以及亲缘进化关系等进行详细的分析。结果表明,九个科植物 GGPS 核苷酸序列长度均大于 1 000 bp 小于 2 000 bp;GGPS 氨基酸都不含有信号肽,不存在跨膜结构域;GGPS 蛋白的二级结构中 α 螺旋和无规卷曲是主要结构,且无 β -折叠结构,延伸链和 β -转角则分散于整个蛋白中。通过生物信息学方法进行合理预测,为进一步研究 GGPS 的酶学特性提供了重要理论依据。

关键词:生物信息学;GGPS;分子进化
中图分类号:Q55 **文献标志码:**A **文章编号:**1672-5565(2016)03-127-07

Bioinformatics analysis of geranylgeranyl pyrophosphate synthase in plants

ZHANG Pengfei, WANG Ning, ZHOU Man, LI Ping*
(School of Life Science and Engineering, Southwest Jiaotong University, Chengdu 610031, China)

Abstract:GGPS is an important branch point enzyme for terpenoids synthesis, but also it's catalyzing enzyme of GGPS. This study select 9 different families of plants, implement a series of detailed analyses to the physicochemical properties of GGPS nucleotide and amino acid sequence, protein structure and function and phylogenetic and so on. The results indicate that full length of DNA sequence from 9 species is in the range from 1 000 bp to 2 000 bp. All the GGPS amino acid sequence don't contain signal peptide and transmembrane structure; α -helix and random coil are the primary structure of GGPS secondary structure and don't have any β -sheet structure, extension chain and β -corner are scattered throughout the protein. Reasonably prediction based on bioinformatics methods provide important theoretical basis for further research GGPS enzymatic properties.

Keywords: Bioinformatics; GGPS; Molecular evolution

高等植物中类胡萝卜素的合成途径理论已经成熟,牻牛儿基牻牛儿基焦磷酸(GGPP)是类胡萝卜素最直接的前体物质^[1]。GGPS 催化 3,3-二甲烯丙基焦磷酸(DMAPP)与异戊烯基焦磷酸(IPP)合成 GGPP,两个 GGPP 分子通过八氢番茄红素合成酶缩合形成八氢番茄红素。GGPS 基因最早从辣椒^[2]中分离得出,之后又在番茄^[3]、丹参^[4]、烟草^[5]、银杏^[6]等植物中分离得到。

近年来,GGPS 基因在功能研究上取得了一些进展。在一些植物中发现 GGPS 对类胡萝卜素的合

成积累具有极为重要的作用^[3,7]。但在啤酒酵母中 GGPS 基因过表达,会导致 β -胡萝卜素含量显著增加^[8]。在法夫酵母和卷枝毛霉中,GGPS 过表达也能显著提高类胡萝卜素产量^[9-10]。在 cDNA 文库中,通过使用该酶抗体完成筛选,从而得到 GGPS cDNA^[11]。另外研究还发现,GGPS 对于调节碳流还有着重要的作用^[12]。在植物体中,GGPS 经常是多基因家族化合物。其表达的蛋白质存在于不同的亚细胞结构中^[13]。本课题共选择了以下 9 种不同科且已知 GGPS 基因序列的物种:银杏(银杏

科)、烟草(茄科)、长春花(夹竹桃科)、丹参(唇形科)、蔷薇(蔷薇科)、万寿菊(菊科)、甘薯(薯蓣科)、茉莉(木犀科)、加拿大红豆杉(红豆杉科)。对 GGPS 的理化性质、核酸序列、蛋白质结构功能等进行详细的分析,研究其进化关系,进而更深层次的了解 GGPS 的性质,为该基因对植物的生长调控所产生的影响提供更可靠的理论依据。

1 材料与方法

1.1 实验材料

本实验数据来源于 NCBI 核酸和蛋白质数据库中已经注册的 GGPS 核酸及相应的氨基酸序列,包括银杏(*Ginkgobiloba*, AAQ72786.1)、万寿菊(*Tagetes erecta*, AAG10424.1)、甘薯(*Ipomoea batatas*, ACF37217.1)、茉莉(*Jasminum sambac*, AIY24421.1)、加拿大红豆杉(*Taxus canadensis*, AAD16018.1)、烟草(*Nicotiana tabacum*, ADD49734.1)、长春花(*Catharanthus roseus*, AGL91648.1)、丹参(*Salvia miltiorrhiza*, ACR19637.1)、蔷薇(*Cistus creticus*, AAM21639.1)、甜菊(*Stevia rebaudiana*,

ABD92926.2)、蓖麻(*Ricinus communis*, XP_002531191.1)、麻疯树(*Jatropha curcas*, ADD82422.1)、橡胶树(*Hevea brasiliensis*, BAF98302.1)。

1.2 方法

GGPS 氨基酸序列的理化性质的分析使用 Expasy 软件中的 ProtParam 在线工具;ORF 的预测分析使用 NCBI 数据库中 ORF 查询工具;使用 SignalP 4.1 在线工具分析 GGPS 是否含有信号肽;通过 NetPhos 2.0 和 NetNGlyc 1.0 在线工具分析糖基化和磷酸化位点;通过 NCBI 数据库中的 CDD 工具对 GGPS 的保守结构域进行预测分析;通过 SOPMA 软件可以预测 GGPS 二级结构,通过 SWISS-MODEL 在线分析其三级结构;通过 Clustal X 采用渐进比较算法进行多序列比对^[14];通过 MEGA 建树工具构建进化树。

2 结果和分析

2.1 GGPS 氨基酸序列理化性质分析

使用在线分析软件 ProtParam 分析 9 种植物 GGPS 氨基酸序列的理化性质(见表 1)。

表 1 GGPS 的氨基酸组成及理化性质分析
Table 1 Physical and chemical characteristics of GGPS amino acid in different plants

植物种类	银杏	万寿菊	甘薯	茉莉	加拿大红豆杉	烟草	长春花	丹参	蔷薇
基因全长/(bp)	1 657	1 176	1 368	1 410	1 889	1 109	1 491	1 563	1 311
氨基酸数目	391	361	363	360	393	366	371	364	369
原子总数	5 976	5 537	5 548	5 552	5 990	5 576	5 700	5 520	5 591
分子量(D)	42 533.8	39 263.0	39 162.2	39 274.2	42 571.9	39 535.6	40 202.3	39 035.9	39 735.4
正电荷残基 (Arg + Lys)	45	38	43	39	43	40	40	38	37
负电荷残基 (Asp+ Glu)	51	49	44	46	52	48	48	45	48
pI	5.98	5.38	6.84	5.91	5.58	5.51	5.51	5.90	5.35
含量最丰富 氨基酸/(%)	Leu(10.7)	Leu(11.1)	Ala(11.3)	Leu(11.7)	Ala(11.2)	Ala(12.6)	Ala(11.3)	Ala(13.5)	Ala(10.3)
	Ala(10.2)	Ala(9.4)	Leu(11.3)	Ala(8.6)	Leu(10.7)	Leu(10.1)	Leu(10.8)	Leu(10.4)	Leu(10.0)
	Ser(9.0)	Ser(8.0)	Gly(8.0)	Lys(7.5)	Gly(7.1)	Val(8.5)	Val(7.5)	Val(8.2)	Val(10.0)
	Glu(7.9)	Glu(7.5)	Glu(7.4)	Ser(7.2)	Leu(7.1)	Lys(7.7)	Lys(7.3)	Gly(7.1)	Ser(9.8)
	Gly(7.2)	Gly(7.5)	Lys(7.4)	Gly(6.9)	Lys(7.1)	Glu(6.8)	Gly(7.0)	Asp(6.9)	Gly(7.3)
	Lys(7.2)	Lys(6.6)	Val(7.4)		Ser(7.1)				
不含有氨基酸	Pyl,Sec	Pyl,Sec	Pyl,Sec	Pyl,Sec	Pyl,Sec	Pyl,Sec	Pyl,Sec	Pyl,Sec	Pyl,Sec
稳定指数	46.03 (不稳定)	46.49 (不稳定)	36.49 (稳定)	40.21 (不稳定)	41.44 (不稳定)	36.19 (稳定)	40.93 (不稳定)	36.32 (稳定)	38.55 (稳定)
脂肪指数	87.62	93.27	96.25	94.31	90.18	98.93	99.43	97.36	95.39

由表可知,所选植物 GGPS 序列在氨基酸数目、预测分子量、Ip 都表现出一致性,他们的基因序列全长在 1 000 bp 到 2 000 bp 之间;氨基酸数目除银杏和加拿大红豆杉在 390 左右外,其余的在 365 左右;

I_p 值除甘薯 6.84 以外,其余均小于 6,不稳定指数 40 左右,脂肪指数在 94 左右;Ala、Leu、Glu、Lys、Gly 是所选植物 GGPS 序列中含量最为丰富的氨基酸,而 Pyl 和 Sec 为所选植物中都不存在的氨基酸。

2.2 GGPS 核酸序列 ORF 分析

通过 ORF Finder 在线软件分析 9 种植物 GGPS 核酸序列的开放阅读框(见表 2)。

表 2 不同植物 GGPS 核酸序列 ORF 分析									
Table2 Nucleic acid sequence analysis of GGPS ORF in different plants									
植物	银杏	万寿菊	甘薯	茉莉	红豆杉	烟草	长春花	丹参	蔷薇
基因全长/(bp)	1 657	1 176	1 368	1 410	1 889	1 109	1 491	1 563	1 311
起始密码	A ₂₂₅	A ₅₀	A ₈₆	A ₂₄₅	A ₃₀₈	A ₄	A ₁₈₁	A ₇₃	A ₁₁₆
终止密码	A ₁₄₀₀	A ₁₁₃₅	A ₁₁₇₇	A ₁₃₂₇	A ₁₄₈₉	A ₁₁₀₄	A ₁₂₉₆	A ₈₁₉	A ₁₂₂₅
ORF 长度/(bp)	1 176	1 086	1 092	1 083	1 182	1 101	1 116	747	1 110
5'/3' 非编码区长度	224/257	49/41	85/191	244/83	307/400	3/5	180/195	72/744	115/86

在线分析软件 ORF Finder 提供了六组参考数据,包含了不同阅读顺序可能出现的情况。由于 GGPS 氨基酸序列长度已经知道,所以每一个 ORF 基本都对应着相应编码蛋白质的大小,结果显示除了丹参以外其余植物 GGPS 的开放阅读框大小均在 1 100 bp 左右,而丹参可能由于蛋白编辑剪切不同,

所以稍有差异。从这些也可以看出,GGPS 序列存在一定的保守性。

2.3 蛋白质信号肽的预测分析

使用 SignalP 4.1 Server 在线工具以银杏为研究对象对所选 GGPS 未知氨基酸序列中包含的信号肽进行分析(见图 1)。

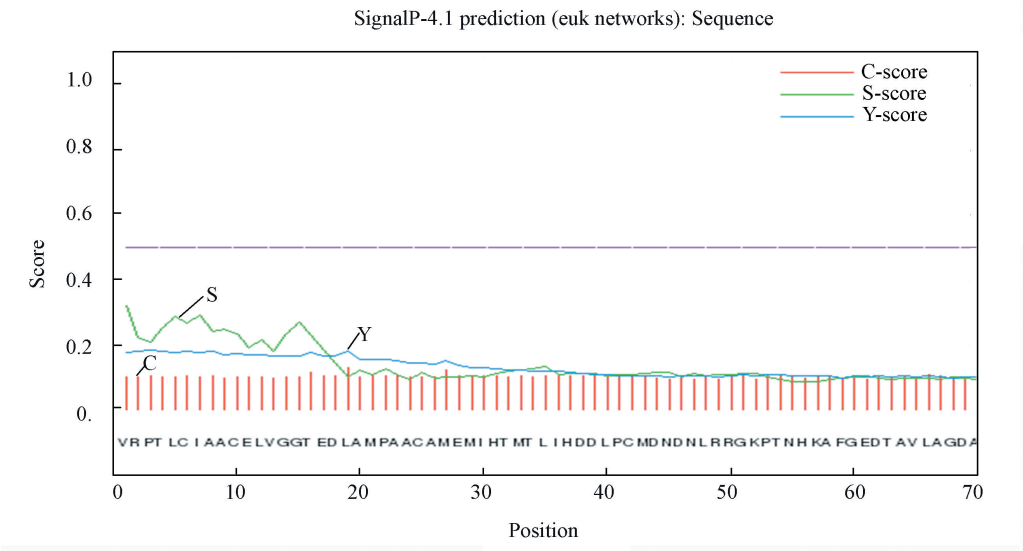


图 1 银杏 GGPS 的信号肽分析结果

Fig. 1 Signal peptide analysis result of Ginkgo GGPS

结果表明,银杏的 GGPS 蛋白序列都不存在信号肽。且由图可以看出,S 值和 Y 值均比较低,因此可以推测以上植物的 GGPS 蛋白通过核糖体合成之后,生成的蛋白质属于非分泌蛋白。通过对另外 8 种植物分析可得到相近的结果。

2.4 糖基化和磷酸化位点分析

使用 NetNGlyc 1.0 以万寿菊为例对所选 GGPS 序列的糖基化位点进行分析(见图 2)。

如图所示,红色横线(Threshold)表示阈值,阈值设定为 0.5,大于阈值的蓝色竖线(Potential)表示

此点有被糖基化的可能。万寿菊 GGPS 氨基酸序列糖基化分析结果显示,在 5 和 24 位点的天冬酰胺处存在蓝色竖线,值分别为 0.782 8 和 0.715 6,因此这两处可能被糖基化。对其他植物 GGPS 氨基酸序列的糖基化预测分析得出,除银杏、茉莉和红豆杉没有糖基化位点外,其余植物 GGPS 蛋白均存在不同程度的糖基化。

使用 NetPhos 2.0 在线工具以银杏为研究对象分别预测了丝氨酸、苏氨酸、酪氨酸 3 种不同的磷酸化位点(见图 3)。

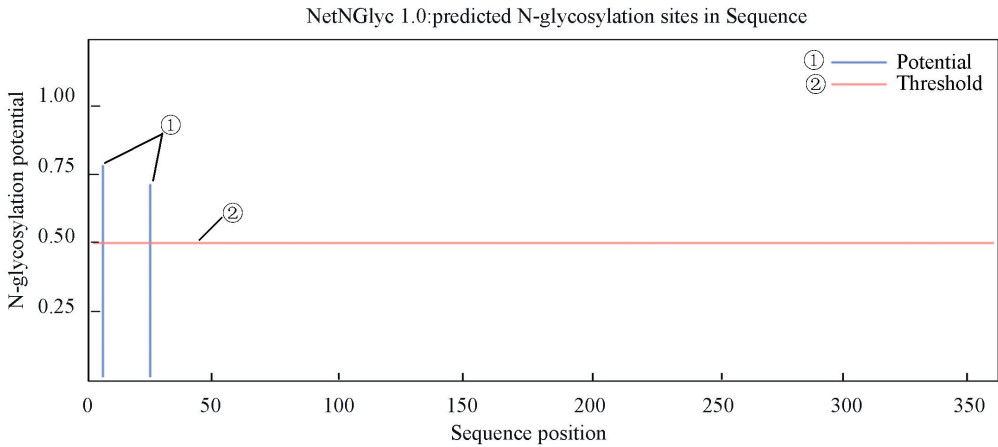


图 2 万寿菊 GGPS 氨基酸序列的糖基化分析
Fig. 2 Glycosylation prediction of ginkgo GGPS amino acid sequence

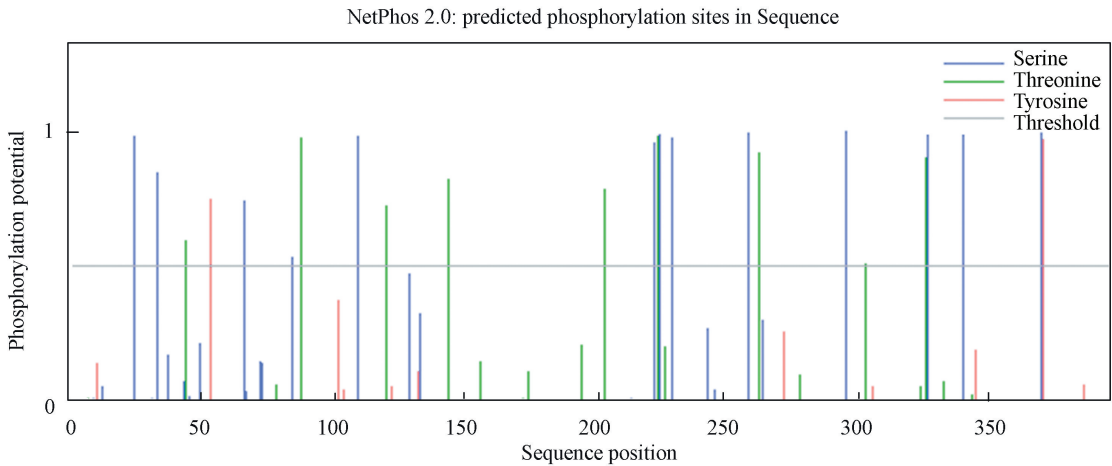


图 3 银杏 GGPS 氨基酸序列的磷酸化位点预测*
Fig. 3 Phosphorylation site prediction of ginkgo GGPS amino acid sequence
注：* 彩图见电子版 (<http://swxxx.alljournals.cn/ch/index.aspx>) (2016 年第 3 期 doi:10.3969/j.issn.1672-5565.2016.03.01)。

结果表明,总共有 24 个位点被磷酸化,其中有 13 个丝氨酸磷酸化位点(分别是第 24、33、66、84、109、222、224、229、258、295、326、339 和 369 位)、9 个苏氨酸磷酸化位点(分别是第 44、88、120、144、203、223、262、302 和 325 位)以及 2 个酪氨酸磷酸化位点(分别是第 53 和 370 位),在以上 24 个磷酸化

位点中 S369、S258、S295 预测分值最高,均在 0.990 以上。通过对另外 8 种植物 GGPS 氨基酸序列的分析,可以得到同样的结果。

2.5 蛋白质二级结构预测分析

以银杏为研究对象,使用 SOPMA 在线工具分析 GGPS 氨基酸序列的二级结构(见图 4)。

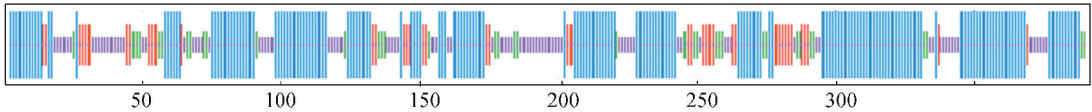


图 4 银杏 GGPS 氨基酸序列的二级结构预测*
Fig. 4 Secondary structure prediction of ginkgo GGPS amino acid sequence
注：* 彩图见电子版 (<http://swxxx.alljournals.cn/ch/index.aspx>) (2016 年第 3 期 doi:10.3969/j.issn.1672-5565.2016.03.01)。

由图可知,在银杏多肽链中二级结构中 α -螺旋(蓝色线*)在链中出现的概率是 51.92%、无规卷曲(紫色线*)在链中出现的概率是 26.09%、延伸链(红色线*)在链中出现的概率是 12.02%、 β -转角(绿色线*)在链中出现的概率是 9.97%,可以看出银

杏 GGPS 二级结构中没有 β -折叠结构。可见 α -螺旋是 GGPS 多肽链中大量存在的结构元件而散布于整个肽链中。另外 8 种植物的 GGPS 二级结构中也有类似的情况。

2.6 保守结构域的预测和分析

使用 NCBI 中的 CDD 工具以银杏为例进行蛋白

质结构域的预测分析(见图 5)。

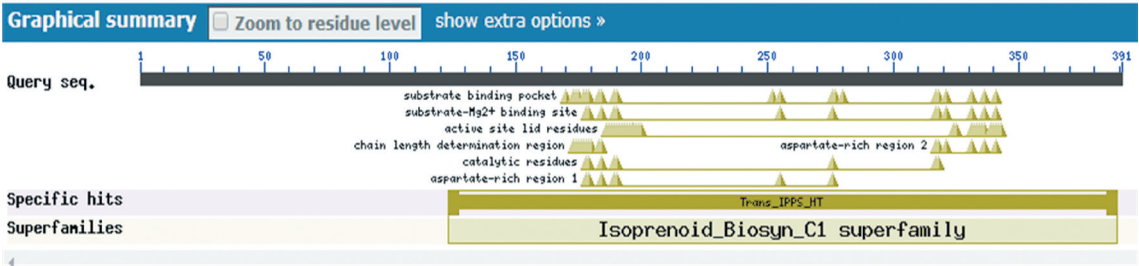


图 5 银杏 GGPS 保守结构域分析

Fig. 5 conserved domains analysis of Ginkgo GGPS

由 NCBI 中 CDD 工具中的详细解释可以得出,银杏 GGPS 氨基酸序列有一个异戊二烯合酶 C1 超家族,总共有六个不同的保守结构域,分别是底物结合保守结构域,结合 Mg^{2+} 的结构域,活性位点残基结合结构域,决定链长度的保守结构域,催化残基的

保守结构域和富含天冬氨酸的保守结构域。

2.7 蛋白质三级结构预测分析

使用 Swiss-Model 对九个科植物 GGPS 氨基酸序列的三级结构进行预测分析(见图 6)。

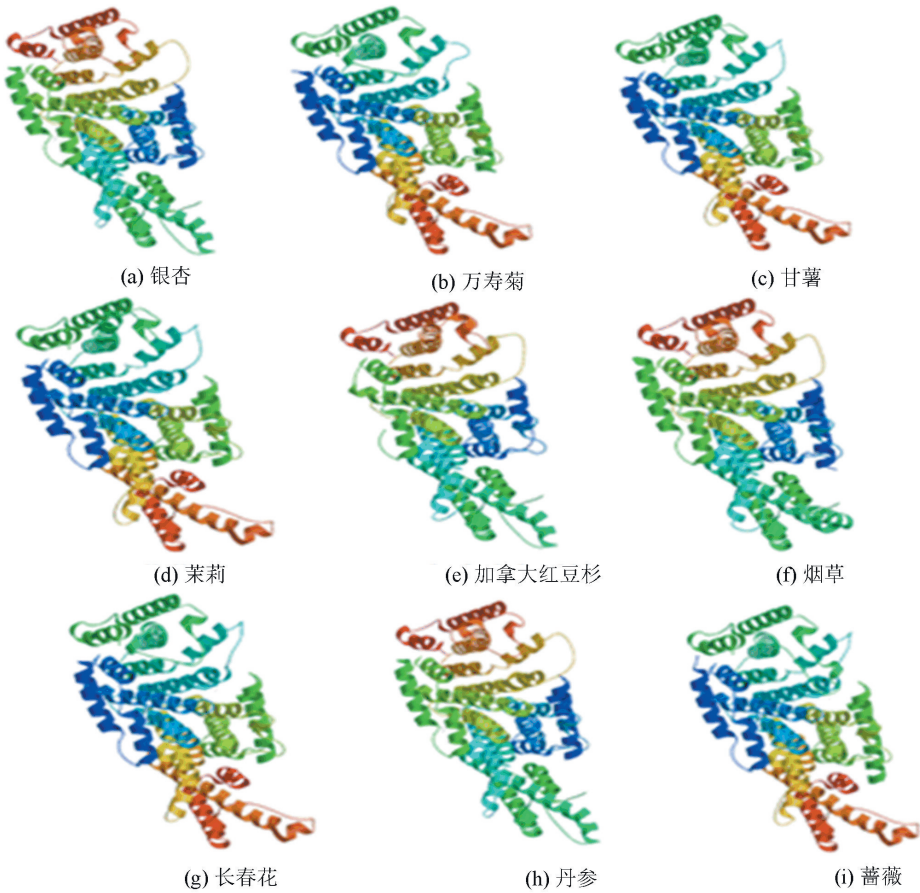


图 6 不同植物 GGPS 氨基酸序列的三级结构预测

Fig. 6 Tertiary structure prediction of GGPS in different plants

通过上面九个三级结构图可以看出,这些植物 GGPS 蛋白的三级结构极为相似,这同它们核苷酸序列与氨基酸序列的相似性有很大关系,相似的空间结构往往具有相似的功能,同时也可以看出 GGPS 在进化过程中是相对保守的。

2.8 核苷酸多序列比对

从前面九个不同科植物 GGPS 序列的分析可以看出,它们的序列相似性相对较高,理化性质差异也不是很大,因此将进一步从序列上分析它们的同源性,为这些植物构建进化树,从而较直观的反映出它们之间的亲缘关系。使用 Clustal X 软件对九个

不同科植物进行多序列比对(见图 7)。

通过 Clustal 软件对九个科植物 GGPS 氨基酸序列进行比对的结果,可以明显的看到这些序列在

很大程度上都是相似的,只有少部分存在差异,因此可以得知 GGPS 在进化过程中是相对保守的。

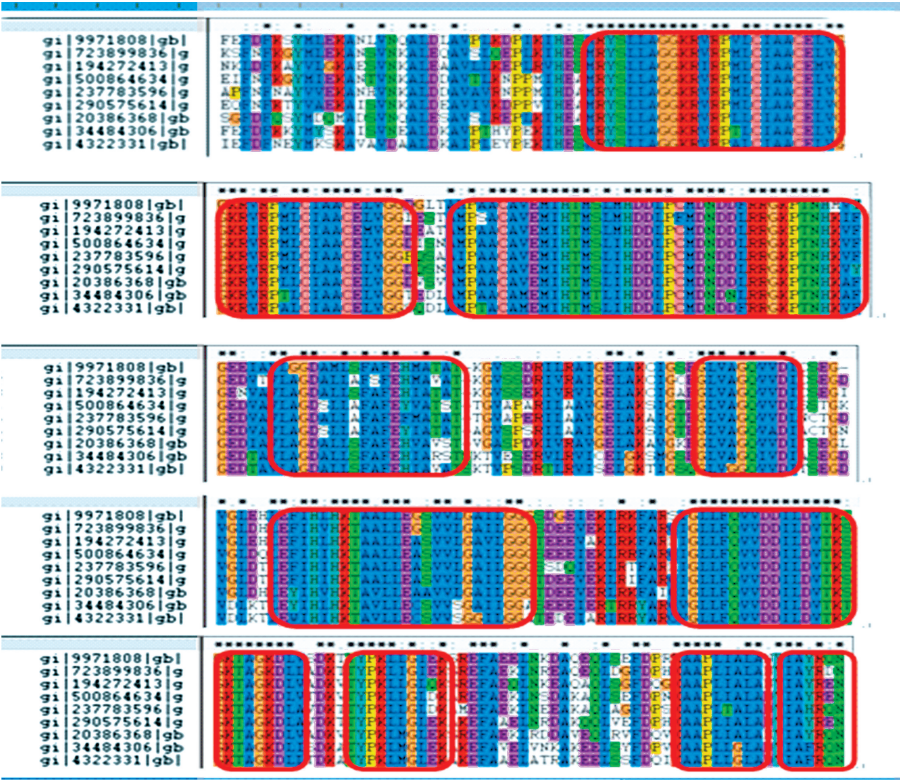


图 7 不同植物 GGPS 的氨基酸多序列对比分析*

Fig. 7 Multiple sequence alignment of GGPS in different plants

注: * 彩图见电子版 (<http://swxxx.alljournals.cn/ch/index.aspx>) (2016 年第 3 期 doi:10.3969/j.issn.1672-5565.2016.03.01)。

2.9 GGPS 进化树的构建

对九种植物的 GGPS 氨基酸序列和一些相同科的植物如甜菊、蓖麻、麻疯树、橡胶树使用 MEGA4.0 软件进行分子系统演化分析,获得 GGPS 核苷酸序列的分子进化树,更进一步地分析它们的亲缘进化关系(见图 8)。从图中可看出 13 种植物 GGPS 氨基酸序列明显被分为三大类,万寿菊和甜菊同属于菊科因此亲缘关系最近,蓖麻和麻疯树同是大戟科,亲缘

关系也很近,由于有九个不同的科所以进化树的小分支较多,而氨基酸序列和基因序列进化树在大体结构上是相似的,但其分类也有稍微差异,可能是由于在进化过程中某些核苷酸的变异导致其氨基酸发生变化,但结合其序列比对结果发现,总体上 GGPS 序列还是较为保守的。因此 GPSS 是牻牛儿基牻牛儿基焦磷酸生物合成途径中一个相对保守的基因,可作为生物遗传分化和分子进化研究中的重要依据。

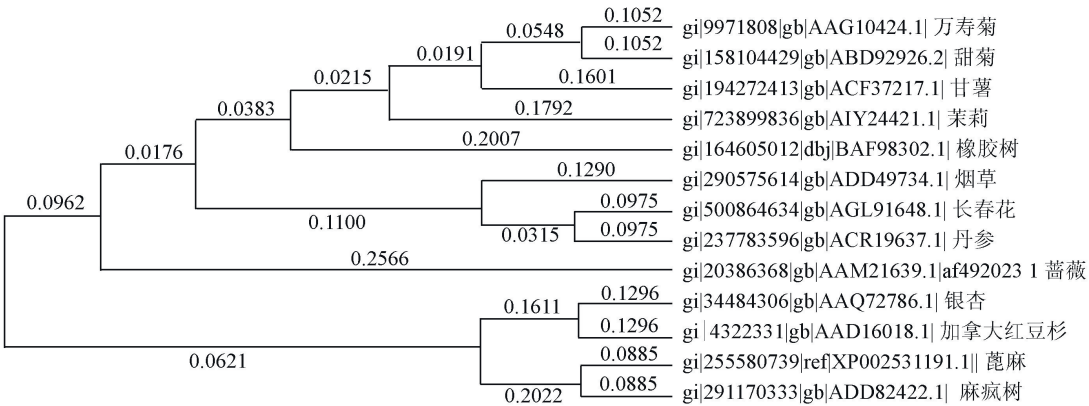


图 8 不同植物 GGPS 的氨基酸序列分子进化分析

Fig. 8 Phylogenetic tree analysis of GGPS in different plants

3 结 论

本研究使用生物信息学的方法对来源不同植物 GGPS 的氨基酸序列进行了系统分析。试验结果表明,植物 GPSS 属于大分子蛋白;氨基酸序列理化性质分析显示植物 GGPS 富含 Ala 和 Leu;通过植物 GGPS 氨基酸序列进行比对的得知 GGPS 在进化过程中是相对保守的;将 13 种植物的 GGPS 进行分子进化分析,结果表明万寿菊和甜菊同属于菊科因此亲缘关系最近,蓖麻和麻疯树同是大戟科,亲缘关系也很近,由于有九个不同的科所以进化树的小分支较多,而氨基酸序列和基因序列进化树在大体结构上是相似的,但其分类也有稍微差异,可能是由于在进化过程中某些核苷酸的变异导致其氨基酸发生变化,但结合其序列比对结果发现,总体上 GGPS 序列还是较为保守的,因此可作为生物遗传分化和分子进化研究中的重要依据;二级结构分析推测 α -螺旋是多肽链中的主要结构元件。该研究结果可为深入开展 GGPS 酶学特性的分子机理研究提供重要理论依据。

参考文献(References)

[1]CAZZONELLI C I,POGSON B J. Source to sink : regulation of carotenoid biosynthesis in plants [J]. Trends Plant Science,2010,15(5) :266-274.

[2]KUNTZ M,ROMER S,SUIRE C. Identification of a cDNA for the plastid-located geranylgeranyl pyrophosphate synthase from Capsicum annuum: correlative increase in enzyme activity and transcript level during fruit ripening [J]. The Plant Journal,1992,2(1) :25-34.

[3]钟秋月,国艳梅,梁燕,等.两个不同来源的番茄 GGPS 基因克隆和序列分析 [J].华北农学报,2009,24(3) :15-22.

ZHONG Qiuyue, GUO Yanmei, LIANG Yan, et al. Cloning and sequence analysis of GGPS gene in two sources of tomato [J]. Acta Agriculturae Boreali-Sinica,2009, 24(3) : 15-22.

[4]张蕾,戴住波,崔光红,等.丹参牦牛儿基牦牛儿基焦磷酸合酶基因的克隆与分析 [J].中国中药杂志,2009,34(21) :2704-2708.

ZHANG Lei, DAI Zhuho, CUI Guanghong, et al. Cloning and characterization of geranylgeranyl diphosphate synthase gene of Salvia miltiorrhiza [J]. China Journal of Chinese Materia Medica,2009,34(21) :2704-2708.

[5]李锋,李明,金立锋,等.烟草牦牛儿基牦牛儿基焦磷酸合

酶基因的克隆及分析 [J].烟草科技,2012(5) :60-64.

LI Feng, LI Ming, JIN Lifeng, et al. Cloning and characterization of a new gene encoding geranylgeranyl pyrophosphate synthase from Nicotiana tabacum [J]. Tobacco Agronomy,2012(5) :60-64.

[6]张洪娟,谭碧月,曹福亮.银杏 GbGGPS 基因的克隆及序列分析 [J].南京林业大学学报:自然科学版,2013,37(4) :9-12.

ZHANG Hongjuan, TAN Biye, CAO Fuliang. Cloning and characterization of the geranylgeranyl pyrophosphate synthase gene from Ginkgo biloba Linn [J]. Journal of Nanjing Forestry University : Natural Sciences Edition, 2013, 37(4) :9-12.

[7] ZHAO W E, LV P, GU H H. Studies on carotenoids in watermelon flesh [J]. Agricultural Science,2013,4(7A) :13-20.

[8] UKIBE K, HASHIDA K, YOSHIDA N, et al. Metabolic engineering of Saccharomyces cerevisiae for astaxanthin production and oxidative stress tolerance [J]. Applied Environmental Microbiology,2009,75(22) :7205-7211.

[9] BREITENBACH J, VISSER H, VERDOES J C, et al. Engineering of geranylgeranyl pyrophosphate synthase levels and physiological conditions for enhanced carotenoid and astaxanthin synthesis in Xanthophyllomyces dendrorhous [J]. Biotechnology Letters,2011,33(4) :755-761.

[10] CSERNETICS A, NAGY G, ITURRIAGA E A, et al. Expression of three isoprenoid biosynthesis genes and their effects on the carotenoid production of the zygomycete Mucor circinelloides [J]. Fungal Genetics and Biology, 2011,48(7) :696-703.

[11] KUNTZ M, ROMER S, SUIRE C, et al. Identification of a cDNA for the plastidlocated geranylgeranyl pyrophosphate synthase from Cap sicumannum: Correlative increase in enzyme activity and trGGPScript level during fruit ripening [J]. The Plant Journal, 2002 , 2(1) : 25-34.

[12] LASKARIS G, BOUNKHAY M, THEODORIDIS G, et al. Induction of ger Anylgeranyl diphosphate synthase activity and taxane accumulation inTaxus baccatacell cultures afterelicitation bymethyl jasmonate [J]. Plant Science, 1999, 147(1) :1-8.

[13] LUCKER J, BOUWMEESTER H J, SCHWAB W, et al. Expression of Clarkia S-linalool synthase in trGGPSgenic petunia plants results in the accumulation of Dglucopyranoside [J]. The Plant Journal, 2001, 27(4) : 315-324.

[14] CHIZZOLA R. Regular monoterpenes and sesquiterpenes (Essential oils) [J]. In Natural Products, 2013, 2973 - 3008.DOI: 10.1007/978-3-642-22144-6_130.