

石笔木 CHS 基因家族成员的分析*

杨俊波¹, 田欣¹, 李德铎^{1**}, 顾红雅², 杨世雄¹

(1 中国科学院昆明植物研究所, 云南 昆明 650204; 2 北京大学生命科学学院, 北京 100871)

摘要: 用 PCR 方法从石笔木 (*Tutcheria spectabilis* Dunn) 的总 DNA 中扩增 CHS 基因外显子 2 的部分序列以代表该基因进行研究, 经克隆后测序, 得到长约 740~780bp 的序列共 12 个。以 EMBL 数据库中得到的紫花苜蓿和欧洲赤松各一个序列作为参照, 进行排序和系统树的构建分析。结果显示, 所有被测定的序列同源性均高于 70%, 为 CHS 基因家族的成员, 且这些序列由 3 大类共 5 种不同的基因拷贝组成: 第一类家族成员因碱基的插入和缺失改变了读码框, 推测已失去了 CHS 基因的功能, 成为假基因, 其中个别拷贝存在较大缺失, 这在以前的研究中未见报道; 第二类家族成员因活性位点的氨基酸发生突变, 可能具有新的基因功能; 第三类家族成员则具有原 CHS 基因的功能。综上结果, 可预测, 山茶科 CHS 基因家族较大, 且有着复杂的进化样式。

关键词: 山茶科; 石笔木; CHS 基因

中图分类号: Q 75 文献标识码: A 文章编号: 0253-2700(2002)02-0209-06

An Analysis of CHS Gene Family Based on *Tutcheria spectabilis* (Theaceae)

YANG Jun-Bo¹, TIAN Xin¹, LI De-Zhu¹, GU Hong-Ya², YANG Shi-Xiong¹

(1 *Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650204;*

2 College of Life Sciences, Peking University, Beijing 100871)

Abstract: By using PCR method, partial sequences of chalcone synthase (CHS) gene exon 2 were amplified from genomic DNA of *Tutcheria spectabilis* (Theaceae). After cloning and sequencing, 12 sequences ranging from 740bp to 780bp were obtained. Comparing these sequences, with the ones of *Medicago sativa* and *Pinus sylvestris* from EMBL database, it reveals a similarity higher than 70% at the nucleotide level, and they are the members of CHS gene family. The result indicates that these sequences consist of five different gene copies in three groups. In group 1, the CHS genes lost their functions because of indels of bases, then turning into pseudo-genes, which is reported here for the first time; In group 2, the gene copies provide new functions because of the mutation of the amino acid in active sites; In group 3, the copies have the same function of CHS genes. It is suggested that the CHS gene family in Theaceae is large, which has complex evolutionary mode.

* 基金项目: 国家自然科学基金重点项目资助(编号 39830020)

** 通讯作者 DZL@mail.kib.ac.cn

收稿日期: 2001-05-15, 2001-07-04 接受发表

作者简介: 杨俊波(1970-)男, 云南人, 助研, 主要从事生物化学与分子生物学研究。

Key words : Theaceae ; *Tutcheria spectabilis* ; CHS genes

查尔酮合酶 (Chalcone Synthase, CHS) 是类黄酮物质生物合成途径的关键酶, 它催化 3 个分子的丙酰 CoA (malonyl CoA) 与 1 个分子的香豆酰 CoA (Coumaryl CoA) 生成查尔酮。CHS 基因是一个较大的多基因家族, 其编码区比较保守, 长约 1.2 kb, 它的家族成员有查尔酮合酶基因, 1, 2-二苯乙烯合酶基因及参与雄蕊发育的基因 (Helasiutta 等, 1996; Qu 等, 1997), 科间的氨基酸同源性在 70% ~ 90%。CHS 基因编码区被不同长度的内含子分隔成外显子 1 和外显子 2, 其中外显子 1 较短, 只编码约 60 个氨基酸, 且长度变异较大; 外显子 2 编码约 340 个氨基酸, 在进化中较为保守, 易于排序, 提供的进化信息较多。王金玲等 (2000) 的研究也表明, 可用 CHS 基因外显子 2 代表全基因进行研究。

最近几年, 随着已知 CHS 基因及其调控序列的迅速增加, 已有人将其用于分子进化及系统发育研究。Ursula 等 (1987) 首次尝试用 CHS 基因编码区的 DNA 序列来研究物种进化关系, 当时他们只分析了 7 个种 8 个序列。王金玲等 (2000) 的工作较为系统, 共分析了 19 个科的 83 个序列。从目前的研究来看, CHS 基因在不同科中发生的重复和丢失情况不同, 所含 CHS 基因的拷贝数也差异较大: 欧芹 (Herrmann 等, 1988)、拟南芥 (Feinbaum 等, 1988)、金鱼草 (Wienand 等, 1982) 中似乎只存在一个拷贝, 而其它物种中均有多个, 其中豆科拷贝数最多, 故很难确定直系同源的 CHS 基因成员 (王金玲等, 2000)。因此, 搞清每个类群中 CHS 基因家族的组成, 并确定出直系同源序列及相互间的进化关系是进行分子系统学研究的关键。

本文用 PCR 方法, 从山茶科石笔木属的石笔木 (*Tutcheria spectabilis* Dunn) 总 DNA 中扩增出 CHS 基因的外显子 2 部分序列, 克隆筛选后进行序列测定, 对拷贝数、不同成员的同源性、进化关系等进行分子系统学分析, 最后对该基因在山茶科的分子系统学研究的意义及可行性进行评估。

1 材料与方法

1.1 材料

石笔木 (*Tutcheria spectabilis* Dunn): 广西, Yang S X 97749, 凭证标本存于昆明植物所标本馆 (KUN)。紫花苜蓿 (*Medicago sativa*) 的 GMACHS2、欧洲赤松 (*Pinus sylvestris*) 的 PSCHS 两个序列: 来自 EMBL 数据库。

大肠杆菌菌株 DH5 α : 来自北京大学蛋白质工程及植物基因工程国家重点实验室。

实验所用酶购自华美公司, 克隆试剂盒购自华舜公司, 自动测序试剂盒购自美国 ABI 公司。

1.2 方法

PCR 扩增 以改良 CTAB 法 (Doyle & Doyle, 1987) 提取的叶片总 DNA 为模板, 用王金玲等 (2000) 的简并引物: 5' CCK TCH YTC GCN MGR CAR GAC3 和 5' GG BCC RAA NAR MAR MAC ACC3 对 CHS 基因的外显子 2 的部分序列进行 PCR 扩增。反应条件为 97 $^{\circ}$ C 预变性 4 min, 接 94 $^{\circ}$ C 变性 1 min、50 $^{\circ}$ C 退火 1 min、72 $^{\circ}$ C 延伸 1 min 共 40 个循环后, 72 $^{\circ}$ C 延伸 10 min, 回收目的条带。经纯化后, 克隆到 PGEM-TV 载体, 转入 DH5 α 菌培养后筛选, 用 310 DNA 自动测序仪 (ABI 公司) 测序。

序列分析 应用 DNA Star 软件 (Dnastar Inc., Madison, Wisconsin) 进行序列校对和排序。基因树的构建由 PAUP 4.02B (Swofford, 2000) 完成。

2 结果

2.1 PCR 扩增及克隆结果

由于引物的简并性和植物中类 CHS 基因的存在，用 PCR 方法从石笔木 (*T. spectabilis*) 的总 DNA 中往往扩增到多带产物，并有较多的引物二聚体。通过增加反应循环数，调整引物用量，采用热启动 PCR 反应等办法，能较大改善扩增效果，得到以目的产物为主的 PCR 产物。目的产物经克隆后，共得到 37 个阳性克隆斑，我们随机测定了其中的 12 个。序列名由该种的属名和种加词的前两个字母组成，末尾的数字为随机测定克隆的顺序。

2.2 CHS 外显子 2 核苷酸及相应氨基酸序列的排序

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
GMACHS2	—	69.7	65.1	74.2	68.3	71.2	74.4	75.2	74.3	69.9	74.0	70.8	68.3	74.0	1
PINUS	34.3	—	60.7	69.9	58.6	63.4	70.1	71.1	70.1	64.1	69.9	63.3	58.4	69.7	2
TUSP1	32.8	41.2	—	78.1	72.1	76.9	78.5	76.9	78.2	77.5	78.1	77.1	72.1	77.9	3
TUSP10	26.8	33.8	18.3	—	78.3	84.0	99.7	91.1	99.9	87.5	99.7	84.0	78.1	99.6	4
TUSP11	32.7	47.3	25.6	21.9	—	75.6	78.4	77.5	78.3	76.2	78.1	75.3	99.6	78.1	5
TUSP12	27.6	38.9	18.3	15.3	24.4	—	84.2	82.4	84.1	94.4	84.0	99.6	75.6	83.9	6
TUSP2	26.4	33.6	17.9	0.3	21.7	15.0	—	91.4	99.9	87.5	99.7	84.2	78.3	99.6	7
TUSP3	26.2	31.7	20.4	9.0	22.7	16.8	8.7	—	91.3	83.7	91.1	82.2	77.4	91.0	8
TUSP4	26.6	33.6	18.1	0.1	21.9	15.2	0.1	8.8	—	87.7	99.9	84.1	78.1	99.7	9
TUSP5	30.0	39.3	19.1	12.4	24.6	4.5	12.4	16.4	12.2	—	87.5	94.6	76.2	87.4	10
TUSP6	26.8	33.8	18.3	0.3	22.1	15.4	0.3	9.0	0.1	12.4	—	84.0	78.0	99.6	11
TUSP7	28.0	38.9	17.9	15.3	24.8	0.4	15.0	17.1	15.2	4.4	15.4	—	75.3	83.9	12
TUSP8	32.7	47.9	25.7	22.1	0.4	24.6	21.9	22.9	22.1	24.8	22.3	25.0	—	78.0	13
TUSP9	26.8	34.0	18.5	0.4	21.9	15.5	0.4	9.1	0.3	12.5	0.4	15.5	22.1	—	14

图 1 石笔木及两外类群 CHS 外显子 2 的部分核苷酸序列的同源性比较 (对角线上部为成对序列的相似百分比率, 对角线下部为成对序列的差异百分比率, 阴影部分表示出同源性极高的两个序列)

Fig. 1 A comparison of homology among the CHS exon 2 partial nucleotide sequences of *Tutcheria spectabilis* (12 sequences) and its 2 outgroups (GMACHS2 and PSCHS) (Right upper: Percent pairwise sequence similarity. Left lower: Percent pairwise sequence divergence. Shadow: High homology between 2 sequences)

对所测定的 12 个 CHS 基因外显子 2 部分核苷酸序列进行排序，并以紫花苜蓿 (*M. sativa*) GMACHS2 和欧洲赤松 (*P. sylvestris*) PSCHS 的氨基酸序列作为参照，进行相应氨基酸序列分析。发现有 4 个序列存在碱基的缺失，其中之一在所测序列的两端附近各有一段较长缺失，另外 3 个序列在所测序列的中后部分有一段四碱基的缺失，这些缺失均导致了该基因读码框的改变，在这 4 个序列中 Tusp12、Tusp5 和 Tusp7 是同一基因拷贝；序列 Tusp8 和 Tusp11 的部分活性位点的氨基酸残基已发生了突变，并且两者同源性很高，可能是同一个具有新功能的基因拷贝；剩余的 6 个序列具有原 CHS2 基因功能，其中 Tusp2、Tusp4、Tusp6、Tusp9、Tusp10 是同一基因拷贝。根据以上比较分析，可确定所有被测定的 12 个核苷酸序列是由 5 种不同的基因拷贝组成，并具有较高的同源性 (图 1)。

***GTTGAAGTATCAAA* <u>TAGCA</u> AAAGAGCTCC*AA*AA*GCAATCAAA GAA TGG GG*CAACAA	
* V E V P K L G * E A A * * A I K E W G Q P	22
AA*TCAAAAT*ACC UAL TGGGT*TTTTCG ACC AT*AGT GAT*TTGACATU UCCGG*GCGGAT*TT	
K S K I T H L V F C T T S G * D M P G A D *	44
C**CH*TCAM*CT*CT* GR*CT*CT* CCA TC*GTC AA*G*TC ATG ATG TACCAA CAA GGTTCG	
* L * * L L G L * P S V * R * M M Y Q Q G C	66
TTTTCCTGGG GGC WGGT*CTTGGCCTAGCTAA* GA*TT*GC*GAG AAC AAC AAA GG*GC*CG*GT*	
F A G G T V I R I A K D * A E N N K G A R V	88
TTAGTCCT*TTTTC*GAGATC*AT*GC* * * * * TT*CT*GGGCTT AGTGA*TC*ATCT*CT*CT*	
L V V C S E I T A * * F * G P S D * * * * *	110
CT*TTGG*CA*GA*TT* TT*TC*GA* GG*GC*GCTCCTATTAIA*TTGGG TGGGACCA*AT*CC*	
L * * Q A L T G D G A A A I I * G S D * * P	132
G**GTTGAGAA*CT*TTT TT*GAG*TG LIT TC*GGG GGC CAA ACU A**CTCCT*GATAG*GA*GT*	
* V E K P L F E * V S A A Q T * I * D S * *	154
GR*AT*AA*CA*TT* TGTTAAAGT* GGTCTTACA T*CA*TT* * * AAG * * TTTCT*GRTT*	
A I * * * L R E * G L T F H L * K D V P G L	176
*L TTTAA*AA*TT*TA* AAG A**TT TGGT*TA*TT* CA*CT**TCGCCAT*TC*GAT TGGAA*	
* S * N * E K * L V E A F Q * * G I S D W N	198
TC*TT TTRG*ATTGA* CA*CTTGGT GAGCTTACTAT TIG GAC CAG*TT GA* TTA AA* TTG*GA	
S L F W I * H P G G P A I L D Q V E L K L *	220
TTAAG ***GAG AAG*TA *GG*TC A** AAG CA*GT*CTAAG*GAGTA* GGTAA*TT*CAAG*GT*	
L K * E K L * * * R H V L * E Y G N * * S *	242
TT*TTG*TT*TA *TT GATGAGA*GAG*AG AA* TC*CT***A*GG	
C V I F * L D E M R * K S * * *	258

图2 石笔木 CHS 外显子 2 的部分核苷酸序列和由核苷酸序列推导出的氨基酸序列 (* 表示有变异的位点, 用框住的位点表示与紫花苜蓿 CHS2 基因对应的催化位点, 有阴影的位点表示与紫花苜蓿 CHS2 基因对应的活性位点, 有下划线的位点表示已失去 CHS 基因功能的拷贝的碱基缺失位点)

Fig. 2 The sequenced nucleotide and deduced amino acid sequences of CHS exon 2 of *T. spectabilis* (Variable sites are denoted by asterisks. Catalytic sites and active sites, correspond to *M. sativa*, are boxed and in the shadow, respectively.

The sites underlined represent the deleted bases of the copies which lose the function of CHS genes)

图2为对所测12个CHS基因外显子2部分核苷酸序列及相应的8个具有CHS基因功能的氨基酸序列进行排序后得到的一致序列。两种序列的变异位点分布都比较均匀,在784个核苷酸序分析位点中,变异位点数为399个,占总数的50.9%,信息位点为257个,占总数的32.7%。氨基酸序列较短,且更加保守,能提供的信息位点相当有限,故本文未作进一步的讨论。

2.3 基因树的构建

基于核苷酸序列的简约性分析得到两个步长为654最简约树。其一致性指数(CI)和维持性指数(RI)分别为0.7844和0.7618。图3是对最简约树进行Bootstrap检测的结果。各分支均有较高的Bootstrap支持率。

从图3可看出,被测定的石笔木的12个序列分为3个Clade, Clade A由4个有碱基缺失,推测已失去基因功能的序列构成, Clade B由6个具有原基因功能的序列构成, Clade C

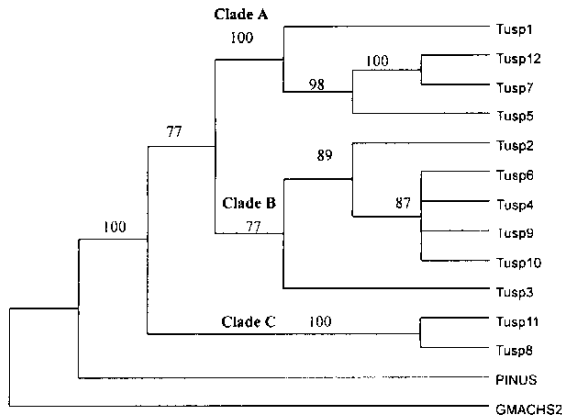


图3 基于最简约法对石笔木及两外类群 14 个基因外显子 2 部分 DNA 序列构建的系统树进行自展分析的结果
(各分支上的数字为 1000 次重复抽样检测的 bootstrap 值)

Fig. 3 The phylogenetic tree based on DNA sequences of partial CHS exon 2 using the most parsimonious method.
(Numbers above branch points are bootstrap estimates for clades 1000 replicates)

由剩余两个序列构成，它们的部分活性位点的氨基酸序列已发生了突变。从该图我们也可看出，石笔木中 CHS 基因家族由三大类共 5 种不同的基因拷贝组成：其中 Tusp5，Tusp7 和 Tusp12，Tusp8 和 Tusp11，以及 Tusp2，Tusp4，Tusp6，Tusp9 和 Tusp10 分别来自同一个基因拷贝。

3 讨论

3.1 CHS 基因是一个多基因家族，研究表明，在不同的植物类群中它的拷贝数及其序列的重复和丢失情况不同，就拷贝数而言，从单一拷贝到多个不同拷贝差异较大，因此，搞清每个类群中 CHS 基因家族的组成及其特点是进行分子系统学研究的关键。本研究工作表明，石笔木 CHS 基因家族的组成较为复杂并有自己的一些特点。具有由于碱基的缺失而产生的非基因功能的拷贝，较长的碱基缺失情况在其它类群的研究中尚无报道。拷贝数目较多，但依据其氨基酸序列的功能改变与否可分为三类，每一类中拷贝数目较为有限（1~2 种），且各类序列间差异较大，易于识别。在山茶科中确定不同属、种的 CHS 基因的直系同源序列以进行分子系统学研究有较大的可行性，并且可根据该科整个 CHS 基因家族组成的复杂程度，选用其中一类或全部进行分子系统学研究。

3.2 目前的研究表明，直系同源序列的确定具有较大的困难，一个较为可行的方法是尽量克隆筛选到各种植物所有的基因拷贝进行研究。这样一方面可研究各种植物该基因的组成和进化，另一方面可用该基因家族的不同成员共同进行相关植物的系统发育重建工作，并能作为相互印证。在本研究中，我们即采用了该方法。这种方法在进行整个科的研究中，会有一些困难，主要是测序花费昂贵，费时、费力。一种可行的方法是在测序前，根据该科的序列特点，应用相应的限制性内切酶进行酶切鉴定，排除相同的拷贝后，再进行测序工作。

3.3 石笔木 CHS 外显子 2 系列的碱基缺失与以往的研究有所不同, 以前的工作均没有较长的插入/缺失突变的报道, 在本研究中发现存在有长达 35 个碱基的缺失, 并且由于这些缺失包括较少碱基的缺失大都是非 3 碱基或其倍数的缺失, 这样的缺失使缺失位点后的氨基酸读码框发生变化, 往往导致基因原功能的丧失, 至于这样的缺失在基因家族的进化中是如何形成的且有何意义, 目前还没有定论, 有待于进一步研究探讨。

〔参 考 文 献〕

- 王金玲, 瞿礼嘉, 陈军等, 2000. CHS 基因外显子 2 的进化规律及其用于植物分子系统学研究的可行性 [J]. 科学通报, **45** (9): 942—950
- Doyle JJ, Doyle JL, 1987. A rapid DNA isolation procedure for smaller quantities of fresh leaf tissue [J]. *Phytochemical Bulletin*, **19**: 11—15
- Feinbaum RL, Ausubel FM, 1988. Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene [J]. *Mol Cell Biol*, **8**: 1985—1992
- Helariutta Y, Kotilainen M, Elomaa P, *et al*, 1996. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: Evolution with substrate change and catalytic simplification [J]. *Proc Natl Acad Sci USA*, **93**: 9033—9038
- Herrmann A, Schulz W, Hahlbrock K, 1988. Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element [J]. *Mol Gen Genet*, **12** (1): 93—98
- Qu LJ, Zhang Yi, Xie Ming, *et al*, 1997. A chalcone synthase-like cDNA from rice anther [J]. *Sex Plant Reprod*, **10**: 181—183
- Swofford DL, 2000. PAUP: phylogenetic analysis using parsimony, Ver. 4.0b5 [M]. Sinauer, Associates, Massachusetts, USA
- Ursula NK, Barzen E, Bernhardt J, *et al*, 1987. Chalcone synthase genes in plants: a tool to study evolutionary relationship [J]. *J Mol Evol*, **26**: 213—225
- Wienand U, Sommer H, Schwarz U, *et al*, 1982. A general method to identify plant structural genes among genomic DNA clones using transposable element induced mutations [J]. *Mol Gen Genet*, **187**: 195—201