

查尔酮合酶超家族(chalcone synthase superfamily) 基因重复和分化的式样

杨 继 顾红雅

(北京大学生命科学学院, 北京 100871. E-mail: jiyang@pku.edu.cn)

摘要 查尔酮合酶(chalcone synthase, CHS)是类黄酮类物质生物合成途径中的关键酶, 对植物的花色素形成以及抗虫和抗病有重要意义. 本文综述了有关查尔酮合酶超家族基因重复和分化式样的研究进展. 有资料显示, *CHS* 基因在不同植物谱系中发生了不同程度的重复和分化, 导致在多数植物基因组中存在具不同表达特性的 *CHS* 重复基因, 并在部分植物基因组中出现由 *CHS* 基因分化形成的类 *CHS* (*CHS*-like) 基因. 不同学者对该家族基因序列及其表达式样进行比较分析, 揭示了该家族重复基因表达式样的分化在很大程度上受启动子和顺式调节元件的影响, 结果导致重复基因的亚功能化(subfunctionalization); 而发生在 *CHS* 活性位点附近的碱基替换则导致重复基因的新功能化(neofunctionalization), 形成不同的类 *CHS* 基因, 且多数类 *CHS* 基因是在不同植物谱系中由 *CHS* 基因多次独立进化形成, 是趋同进化的结果. 探讨 *CHS* 超基因家族的进化历史对深入了解基因重复-分化的过程和机制有很大帮助.

关键词 查尔酮合酶 基因重复 分化

查尔酮合酶(chalcone synthase, CHS)是类黄酮类物质生物合成途径中的关键酶, 它催化 3 个分子的丙酚 CoA (malonyl CoA) 与 1 个分子的香豆酰 CoA (coumaryl CoA) 生成 4,5,7-三羟基黄酮(narigenin chalcone). 类黄酮类物质是广泛存在于高等植物中的次生代谢物, 与植物的花色形成密切相关, 并在植物与环境的相互作用中起重要作用, 如防止 UV 损伤、抗病、影响豆科植物的根瘤形成等^[1].

目前已从苔藓、蕨类、裸子植物和被子植物中克隆了约 650 个 *CHS* 基因及其相关基因的序列^[2], 并发现在多数植物基因组中存在 *CHS* 重复基因. 从基因结构看, 多数 *CHS* 基因包含 1 个内含子和 2 个外显子, 并且内含子的位置在所有已知序列中相同. Ferre 等人^[3]通过对紫花苜蓿(*Medicago sativa*)的一个 *CHS* 晶体结构进行研究, 证实其活性位点为 Cys₁₆₄, Phe₂₁₅, His₃₀₃ 和 Asn₃₃₆, 这 4 个氨基酸残基在至今发现的所有 *CHS* 和类 *CHS* (*CHS*-like) 酶中均相同. 这种高度一致的基因结构说明所有 *CHS* 和类 *CHS* 基因可能来源于一个共同的祖先基因. 但在进化过程中, 伴随着物种分化和新类群的产生, *CHS* 基因也发生了不同程度的重复和分化, 导致在不同植物基因组中产生数目和分化程度不等的重复基因^[4-9]. 这些结构相似、功能相关、但表达式样及编码产物的催化特性有一定差异的基因组成了一个庞大的基因家族——*CHS* 基

因超家族(chalcone synthase gene superfamily)^[10].

基因重复和分化一直被看作是新基因起源或基因功能分化的主要途径^[11]. 基因重复事件导致产生冗余的重复基因, 由于选择压力的松弛, 重复基因得以在加速碱基替换速率的同时, 各自积累不同的遗传变异(突变), 进而改变基因的结构和功能^[12-14]. 以往一直认为重复基因主要有两种不同的进化命运^[11,15]: () 由于随机突变的积累, 导致基因沉默或失活, 即非功能化(nonfunctionalization); () 由于突变而获得新的功能, 并被选择作用所保留, 即新功能化(neofunctionalization). 但近期研究结果表明, 除了蜕变成假基因(pseudogene)或发展成新基因以外, 多数重复基因的命运是亚功能化(subfunctionalization), 亦即基因重复以后所有拷贝都可能由于选择压力的松弛而发生突变, 且突变的效应常常最早体现在转录调控区域, 致使不同拷贝在表达时间和器官特异性方面产生明显分化, 分担了祖先基因的功能, 因而都被选择作用所保留^[15,16].

系统发育分析是生物学研究中一种常用的分析方法, 目前不仅被用于探讨不同生物类群间的系统演化关系, 而且被广泛用于研究基因家族中不同重复基因的起源和进化历史^[17], 即通过对来自不同类群生物的基因序列进行分析和比较, 了解其基本变异式样和速率, 选择恰当的模式, 构建系统树, 一方

面揭示基因家族中不同重复拷贝的分歧时间、进化关系;另一方面,基于对基因树(gene tree)和物种树(species tree)的比较,可以帮助了解该家族基因重复和分化的历史与不同生物类群产生和分化的关系。

CHS 基因在多数植物基因组中都存在不同的拷贝,并在分化速率、方向以及表型效应等方面表现不同的式样,不仅对研究基因重复-分化的过程和机制有重要意义,而且能为探讨不同类群植物的演化关系提供信息。本文综述了有关查尔酮合酶超家族基因重复和分化式样的研究进展,为进一步深入研究 *CHS* 超基因家族的进化历史、以及从系统发育角度研究其他基因家族的进化历程提供参考。

1 *CHS* 基因的重复和分化

在多数植物基因组中存在数量不等的 *CHS* 重复基因。矮牵牛(*Petunia hybrida*) *CHS* 基因家族最早得到深入的研究^[18,19],从该植物基因文库中共获得 8 个 *CHS* 基因全长序列(*CHSA*, *CHSB*, *CHSC*, *CHSD*, *CHSE*, *CHSF*, *CHSG*和*CHSJ*)和 4 条 *CHS* 基因片段序列(*CHSH*, *CHSI*, *CHSK*和*CHSL*)。序列分析的结果显示, *CHSB* 在序列组成上已与其他拷贝发生了明显的分歧,推测其产生于一次较早的基因重复事件,并随后发生了加速进化的过程;而 *CHSC*, *CHSG*和*CHSL* 无论是编码序列还是内含子序列都表现很高的相似性(>90%),因而很可能是近期基因重复事件的产物。从表达式样上看, *CHSA*和*CHSJ*主要在花器官中表达,但也在经UV处理的苗中表达;而 *CHSB*和*CHSG*只在经UV处理的苗中表达,在花中检测不到。但从DNA序列一致性看, *CHSB*与*CHSG*的相似性为 67%,而 *CHSB*与*CHSA*及*CHSJ*的相似性分别为 71%和 69%, *CHSG*与*CHSA*及*CHSJ*的相似性分别为 80%和 81%^[11]。可见,该类群不同基因家族成员的表达式样与其序列相似性程度关系不大。至今在花中和经UV处理的苗中尚未检测到*CHSD*的表达,序列分析的结果显示,由于*CHSD*基因的终止子发生突变,导致编码区延长约 30 个密码子,这可能是造成*CHSD*失活的主要原因,也代表了重复基因进化的另一种命运。

目前在番薯属(*Ipomoea*)植物中至少已发现 6 个 *CHS* 重复基因,其中 *CHSA*, *CHSB*, *CHSC*和*CHSPS*(假基因)组成一个亚家族(subfamily), *CHSD*和*CHSE* 组成另一个亚家族,根据序列分歧程度推测 2 个亚家族的分歧发生在约 100 万年前^[4]。不同拷贝表达的时间和

部位差异很大, *CHSD*在花冠檐部大量表达,而*CHSA*, *CHSB*, *CHSC*和*CHSE*则主要在花萼、无色的花冠管、花药以及子叶等部位选择性表达^[4],且*CHSA*, *B*和*C*与*CHSD*及*E*编码的酶在功能上也发生了一定程度的分化^[20]。通过比较不同亚家族基因的同义和非同义替换速率,发现*CHSA*, *B*和*C*亚家族的绝对进化速率(非同义替换速率)是*CHSD*和*E*亚家族的 2.7 倍^[4],说明基因重复以后发生了明显的不对称进化(asymmetric evolution)。有证据表明,正选择作用(positive selection)驱动了*CHSA*, *B*和*C*亚家族基因的加速进化^[21]。通过序列比较发现,在*CHSD*的启动子和内含子中含有多个转座因子(mobile element)^[22,23],但这些转座因子在*CHSE*中不是位于 5'端的调控区域和内含子中,而是分布在 3'端;在*CHSA*和*CHSB*的内含子以及上、下游调控区域均未发现有转座因子的存在。显然,基因重复以后,不同拷贝经历了不同的转座因子插入的过程,而转座因子在不同拷贝中独特的活动式样可能正是导致不同拷贝表达式样分化的重要因素。

Ito等人^[24]研究豌豆(*Pisum sativum*)*CHS*基因家族时发现其 8 个 *CHS* 基因可以依据彼此间序列相似性的高低分为 3 组,第 1 组与第 2 组的序列相似性为 90%,而第 3 组与第 1 组及第 2 组的相似性低于 80%。有趣的是,这种划分与它们的表达模式一致,第 1 组包括*PSCHS1*, *PSCHS2* 和*PSCHS8*,主要在根和花中表达,并可在UV及elicitor处理后大量表达;第 2 组包括*PSCHS3*, *PSCHS4*和*PSCHS5*,与第 1 组相似之处在于它们同样可被UV及elicitor诱导表达,但只在根中表达;第 3 组包括*PSCHS6*和*PSCHS7*,它们只在根中微弱表达,在花中检测不到,且不能被UV及elicitor处理诱导表达。但对*CHS*重复基因而言,并非序列相似性程度越高,表达式样就越接近。在菜豆(*Phaseolus vulgaris*)中目前已发现有 6 个 *CHS* 基因,其中*CHS1*, *CHS4*, *CHS14*和*CHS17*的DNA相似性高达 96%~98%,但它们的表达模式却明显不同, *CHS1*, *CHS4*和*CHS14* 受elicitor诱导表达,而*CHS17* 却不受elicitor影响;在菜豆的下胚轴中, *CHS1* 受白光诱导,而*CHS4*, *CHS14*和*CHS17* 则不受影响^[25]。比较豌豆 *CHS* 基因 5'端的转录调控序列发现, *PSCHS1*, *PSCHS2*, *PSCHS3*, *PSCHS4*和*PSCHS5* 具很高的相似性,并都含有与elicitor诱导相关的 Box , Box 和 G-Box 等顺式作用元件,而*PSCHS6* 和

PSCHS7的启动子序列与其他拷贝差异很大,无法进行很好的比对(align),并缺乏与elicitor诱导相关的顺式作用元件^[24];在菜豆中也发现有类似的现象^[26]。由此可见,重复基因表达式样的分化在很大程度上受启动子和顺式调节元件的影响。这些数据也进一步证实了伴随着基因重复事件的发生,不同拷贝的转录调控序列往往会发生迅速的进化改变,结果导致重复基因的亚功能化,在以后的进化过程中,亚功能化的基因有可能进一步发展成具有不同功能的新基因。

2 类CHS(CHS-like)基因的起源和分化

基因重复和功能分化一直被看作是新基因起源的主要途径。有证据表明,植物类黄酮代谢途径中有相当一部分酶是由CHS通过重复-分化的途径产生,其中最典型的是1,2-二苯乙烯合酶(stilbene synthase, STS)。当植物受到真菌侵染或其他逆境伤害时, STS基因会诱导表达,催化合成与1,2-二苯乙烯相关的化合物,以提高植物的抗性^[2,10]。STS与CHS有很高的相似性,氨基酸序列相似程度高达70%~90%,并有实验证明只需突变CHS中少数几个氨基酸,就可使其具有STS活性^[5]。CHS与STS催化反应的底物相同,但产物不同,前者为4,5,7-三羟基黄酮,后者为白藜芦醇(resveratrol)或其他与1,2-二苯乙烯相关的化合物,如松属植物中的赤松素(pinosylvin)和二氢赤松素(dihydropinosylvin)^[2]。对来自不同类群植物STS的氨基酸序列进行比较,发现除了CHS超家族成员共有的保守位点外,没有一个STS特有的保守位点,说明CHS与STS的功能分化并非缘自少数关键位点上某个特殊氨基酸的替换。有实验证明STS功能的获得主要是由于催化位点附近少数氨基酸的替换改变了活性中心的氢键骨架,从而导致环化反应的特性发生改变^[2,27]。系统发育分析的结果则显示,来自不同类群植物的STS基因在分子系统树上并不形成一个独立的分支,而是分别与同一科植物的CHS基因结合在一起^[5],但在科以下水平,来自不同种、属的STS基因聚集成一个独立的分支,CHS基因聚成另一支。由此推测:() STS基因是在不同植物线系中由CHS基因多次独立进化而来;() 同一科植物的STS基因享有共同的祖先,导致产生STS祖先基因的基因重复事件发生在不同种、属分化之前。

2-吡喃酮合酶(2-pyrone synthase, 2-PS)是另一种

由类CHS基因编码的酶,最早在菊科大丁草(*Gerbera*)属植物中发现^[6,28]。该类植物的CHS基因分成3个不同的亚家族, *GCHS1* 和 *GCHS3* 亚家族基因的编码产物具有正常的CHS的功能,但在不同发育阶段、在花的不同部位选择性表达^[6],而*GCHS2* 亚家族基因不仅表达式样与前两个亚家族不同,而且表达产物的功能和催化特性都发生了明显分化,实验证明*GCHS2* 编码的酶以苯甲酰CoA (benzoyl CoA)为催化底物,通过2次脱羧反应和1次环化反应合成6-甲基-4-羟基-2-吡喃酮(6-methyl-4-hydroxy-2-pyrone),因而被命名为2-吡喃酮合酶^[29]。比较CHS与2-PS的氨基酸序列及其活性中心的结构,发现2-PS催化位点的氨基酸残基与CHS相同,但3个参与形成香豆酰(coumaryl)结合口袋和环化反应口袋的氨基酸残基发生了替换,导致活性中心的体积缩小,反应过程及产物均发生改变^[30]。有资料显示2-PS不仅存在于大丁草属植物,在其他菊科植物中也有分布^[6,31],且在分子进化树中,来自不同种、属的2-PS基因聚集成一个独立的分支,而*GCHS1* 和 *GCHS3* 亚家族基因形成另外2个独立的分支,说明*GCHS1*, *GCHS2* 和 *GCHS3* 的分化发生在菊科种属分化以前^[6]。对不同亚家族基因的进化速率进行比较,发现3个亚家族基因的同义替换速率没有显著差异, *GCHS1* 和 *GCHS3* 的非同义替换速率也相接近,惟有*GCHS2* 亚家族基因的非同义替换速率明显高于其他两个亚家族,且*GCHS2* 基因某些位点上非同义替换速率与同义替换速率的比率明显大于1 ($d_N/d_S > 1$)^[31],表明在这些位点上发生的氨基酸替换并非由于基因重复以后选择压力的松弛所致,而是由正选择所驱动,具有显著的适应意义。目前有资料证明,2-PS的催化产物是菊科植物特有的能有效阻止细菌或真菌侵害的成份^[29]。

除STS和2-PS外,植物体中其他一些与类黄酮代谢有关的酶也都是由CHS通过重复-分化的途径产生,包括主要分布在芸香科(Rutaceae)植物中的丫啉酮合酶(acridone synthase, ACS)^[7]、兰科(Orchidaceae)植物普遍含有的bibenzyl synthase(BBS)^[8]、大麻科(Cannabaceae)植物含有的与合成啤酒花苦味酸(hop bitter acids)有关的phlorisovalerophenone synthase (PVPS)^[9]等;并有资料显示,CHS通过重复、分化成STS或PVPS的过程在苔类植物*Marchantia paleacea* var. *diptera*^[32], *Lunularia cruciata*^[33]和原始的蕨类植物松叶兰(*Psilotum nudum*)^[34]中就已经发生。但

与STS及2-PS一样,这些来自不同植物类群、但具有相同功能的类CHS酶在分子进化树上并不形成一个独立的分支,而是与相应科的CHS结合在一起,表明尽管它们的功能相同,但并非来自一个共同的祖先,而是多次独立进化的结果。这些酶催化反应的产物多数与抵抗不良环境或病虫害的侵害有关,因而具有显著的适应意义,这也为类CHS基因的起源和分化提供了强大的选择压力。在绝大多数植物基因组中都存在CHS重复基因,当它们处在相同的选择压力下时,就可能发生趋同进化的过程,问题是这些在功能上表现趋同进化式样的基因在DNA序列及其所编码的氨基酸序列中都存在明显的差异,而且彼此之间常常缺乏特异性的保守位点,那么它们的功能通过什么途径来实现?这是研究CHS超基因家族进化式样有待深入探讨的问题。此外,目前主要依据序列相似性程度确定CHS和类CHS基因的同源关系,但有证据表明,部分根据序列相似程度确定的CHS基因实际编码的是具不同催化功能的类CHS酶,加之在进化过程中伴随着不同植物类群的产生,在不同植物谱系中发生了多次独立的基因重复和丢失事件,使我们在确定基因同源关系方面面临很大的困难,因此需要积累更多的资料,尤其是有关结构和功能的信息,为重建CHS超家族基因重复和分化的历史提供依据。

致谢 本工作受国家重点基础研究发展计划(批准号:2003CB715904)和国家自然科学基金重点项目(批准号:39830020)资助。

参 考 文 献

- Koes R E, Quattrocchio F, Mol J N M. The flavonoid biosynthetic pathway in plants: Function and evolution. *BioEssays*, 1994, 16: 123—132
- Austin M B, Noel J P. The chalcone synthase superfamily of type polyketide synthases. *Nat Prod Rep*, 2003, 20: 79—110[DOI]
- Ferrer J L, Jez J M, Bowman M E, et al. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. *Nature Struct Biol*, 1999, 6: 775—784[DOI]
- Durbin M L, McCaig B, Clegg M T. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. *Plant Mol Biol*, 2000, 42: 79—92[DOI]
- Tropf S, Lanz T, Rensing S A, et al. Evidence that stibene synthases have developed from chalcone synthases several times in the course of evolution. *J Mol Evol*, 1994, 38: 610—618
- Helariutta Y, Kotilainen M, Elomaa P, et al. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: Evolution with substrate change and catalytic simplification. *Proc Natl Acad Sci USA*, 1996, 93: 9033—9038[DOI]
- Lukacin R, Springob K, Urbanke C, et al. Native acridone synthases and from *Ruta graveolens* L. form homodimers. *FEBS Lett*, 1999, 448: 135—140[DOI]
- Preisig-Muller R, Gnau P, Kindl H. The inducible 9,10-dihydro-phenanthrene pathway: Characterization and expression of bibenzyl synthase and s-adenosylhomocysteine hydrolase. *Arch Biochem Biophys*, 1995, 317: 201—207[DOI]
- Paniego N B, Zuurbier K W M, Fung S, et al. Phlorisovalerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. *Eur J Biochem*, 1999, 262: 612—616[DOI]
- Schroder J. A family of plant-specific polyketide synthases: Facts and predictions. *Trends Plant Sci*, 1997, 2: 373—378[DOI]
- Ohno S. *Evolution by Gene Duplication*. New York: Springer-Verlag, 1970
- Goodman M, Moore G W, Matsuda G. Darwinian evolution in the genealogy of hemoglobin. *Nature*, 1975, 253: 603—608
- Li W H, Gojobori T. Rapid evolution of goat and sheep globin genes following gene duplication. *Mol Biol Evol*, 1983, 1: 94—108
- Zhang L, Vision T J, Gaut B S. Patterns of nucleotide substitution among simultaneously duplicated gene pairs in *Arabidopsis thaliana*. *Mol Biol Evol*, 2002, 19: 1464—1473
- Lynch M, Conery J S. The evolutionary fate and consequences of duplicate genes. *Science*, 2000, 290: 1151—1155[DOI]
- Force A, Lynch M, Pickett F B, et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 1999, 151: 1531—1545
- Thornton J W, DeSalle R. Gene family evolution and homology: Genomics meets phylogenetics. *Annu Rev Genom Hum Genet*, 2000, 1: 41—73[DOI]
- Koes R E, Spelt C E, van den Elzen P J M, et al. Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida*. *Gene*, 1989, 81: 245—257[DOI]
- Koes R E, Spelt C E, Mol J N M. The chalcone synthase multigene family of *Petunia hybrida* (V30): Differential, light-regulated expression during flower development and UV light induction. *Plant Mol Biol*, 1989, 12: 213—225[DOI]
- Clegg M T, Durbin M L. Flower color variation: A model for the experimental study of evolution. *Proc Natl Acad Sci USA*, 2000, 97: 7016—7023[DOI]
- Yang J, Gu H, Yang Z. Likelihood analysis of the chalcone synthase genes suggests the role of positive selection in morning glories (*Ipomoea*). *J Mol Evol*, 2004, 58: 54—63[DOI]
- Durbin M L, Denton A L, Clegg M T. Dynamics of mobile elements activity in chalcone synthase loci in the common morning glory (*Ipomoea purpurea*). *Proc Natl Acad Sci USA*, 2001, 98: 5084—5089[DOI]
- Hoshino A, Johzuka-Hisatomi Y, Iida S. Gene duplication and mobile genetic elements in the morning glories. *Gene*, 2001, 265: 1—10[DOI]
- Ito M, Ichinose Y, Kato H, et al. Molecular evolution and func-

- tional relevance of the chalcone synthase genes of pea. *Mol Gen Genet*, 1997, 255: 28—37[DOI]
- 25 Ryder T B, Hedrick S A, Bell J N, et al. Organization and differential activation of a gene family encoding the plant defense enzyme chalcone synthase in *Phaseolus vulgaris*. *Mol Gen Genet*, 1987, 210: 219—233[DOI]
- 26 Loake G J, Faktor O, Lamb C J, et al. Combination of H-box [CCTACC(N)₇CT] and G-box (CACGTG) *cis*-elements is necessary for feed-forward stimulation of a chalcone synthase promoter by the phenylpropanoid-pathway intermediate *p*-coumaric acid. *Proc Natl Acad Sci USA*, 1992, 89: 9230—9234
- 27 Austin M B, Bowman M E, Ferrer J L, et al. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type polyketide synthases. *Chem Biol*, 2004, 11: 1179—1194[DOI]
- 28 Helariutta Y, Elomaa P, Kotilainen M, et al. Chalcone synthase-like genes active during corolla development are differentially expressed and encode enzymes with different catalytic properties in *Gerbera hybrida* (Asteraceae). *Plant Mol Biol*, 1995, 28: 47—60[DOI]
- 29 Eckermann S, Schröder G, Schmidt J, et al. New pathway to polyketides in plants. *Nature*, 1998, 396: 397—390
- 30 Jez J M, Austin M B, Ferrer J L, et al. Structural control of polyketide formation in plant-specific polyketide synthases. *Chem Biol*, 2000, 7: 919—930[DOI]
- 31 Yang J, Huang J, Gu H, et al. Duplication and adaptive evolution of the chalcone synthase genes of *Dendranthema* (Asteraceae). *Mol Biol Evol*, 2002, 19: 1752—1759
- 32 Harashima S, Takano H, Ono K, et al. Chalcone synthase-like gene in the liverwort, *Marchantia paleacea* var. *diptera*. *Plant Cell Rep*, 2004, 23: 167—173
- 33 Huang J, Qu L, Yang J, et al. A preliminary study on the origin and evolution of chalcone synthase (*CHS*) gene in angiosperms. *Acta Bot Sin*, 2004, 46: 10—19
- 34 Yamazaki Y, Suh D, Sitthithaworn W, et al. Diverse chalcone synthase superfamily enzymes from the most primitive vascular plant, *Psilotum nudum*. *Planta*, 2001, 214: 75—84

(2005-11-08 收稿, 2006-01-04 接受)