

植物二氢吡啶二羧酸合成酶的功能分歧与进化分析

朱红霞¹, 胡利宗^{2,3}, 李 慧¹, 张家洋¹

(1. 新乡学院, 河南 新乡 453003; 2. 中国科学院 遗传发育所, 北京 100101; 3. 中国科学院 研究生院, 北京 100039)

摘要: 二氢吡啶二羧酸合成酶(DHDPS) 是赖氨酸生物合成途径的关键酶, 探讨 DHDPS 在不同植物中所受到的选择压力及其功能分歧, 以期阐明赖氨酸积累的分子机制奠定基础, 并为植物逆境适应性研究提供线索。利用模式植物基因组数据库, 通过生物信息学手段, 鉴定并获得水稻、拟南芥和大豆等 11 个物种的 DHDPS 基因的核酸和蛋白全序列; 利用 Clustal X、PAML 和 DIVERGE 等工具剖析了 21 条 DHDPS 基因的序列、功能分歧、选择压力及进化关系, 并计算了相关参数。系统进化树揭示了植物 DHDPS 基因能被划分为 5 个类群, 其中, 杨树的 2 个 DHDPS 蛋白单独形成了 E 类群, 剩余的 4 个类群至少包括 2 个物种的 DHDPS 蛋白。基于位点模型的选择压力检测显示, M8 模型虽然能鉴定出正选择位点, 但这些位点均未达到显著水平, 说明植物 DHDPS 基因主要受控于纯净选择。类群间功能分歧研究显示: 类群之间存在 I 型功能分歧, 不存在 II 型功能分歧; 总共鉴定了 273 个功能分歧位点, 其中, I 型功能分歧位点 214 个, II 型功能分歧位点 59 个; 似然比检测表明, I 型功能分歧位点均达到显著水平, 但 II 型功能分歧位点均不显著。因此, 在长期进化过程中, DHDPS 基因类群间发生 I 型功能分化, 同时, 单个类群内纯净选择起着主导作用。

关键词: 植物; 基因; 功能分歧; 适应性进化; 分歧时间; 选择压力

中图分类号: Q75 **文献标志码:** A **文章编号:** 1004-3268(2012)04-0034-04

Functional Divergence and Evolutionary Analysis of Dihydrodipicolinate Synthase in Plants

ZHU Hong-xia¹, HU Li-zong^{2,3}, LI Hui¹, ZHANG Jia-yang¹

(1. Xinxiang College, Xinxiang 453003, China; 2. The Institute of Genetics and Developmental Biology of Chinese Academy of Sciences, Beijing 100101, China;
3. Graduate School of Chinese Academy of Sciences, Beijing 100039, China)

Abstract: Dihydrodipicolinate synthase (DHDPS) is the key enzyme for the pathway of lysine biosynthesis. This paper aimed at investigation of selective pressure and functional divergence in DHDPS proteins in different plants, which would pave the way for adequate understanding of molecular mechanism for lysine accumulation, and provide useful clues for stress tolerance in plants. Based on *in silico* methods, nucleic acid and protein sequences of DHDPS genes in 11 model plants, including rice, *Arabidopsis*, soybean, and so on, were identified from the corresponding genome databases. The sequence alignment, functional divergence, selective pressure, and phylogenetic relationship of 21 DHDPS genes were analyzed using Clustal X, PAML and DIVERGE, and the related parameters were estimated. All the DHDPS genes were classified into five groups

收稿日期: 2011-12-27

基金项目: 河南省政府决策研究招标课题(B578)

作者简介: 朱红霞(1981-), 女, 河南长葛人, 在读硕士研究生, 研究方向: 分子生物学。E-mail: lg10077@163.com

based on the phylogenetic tree, among which E group only included two *DHDPS* genes in *Populus*, but other groups included *DHDPS* genes from at least two plants. The site specific model implemented in Codeml from PAML package revealed that *DHDPS* genes in each of five groups were mainly controlled by purifying selection, and there were no significant positive sites. Functional divergent analyses between five groups indicated that there was type I but no type II functional divergence and there were 273 functional divergence sites including 214 type I and 59 type II functional divergence sites. Likelihood ratio test showed that 59 type II functional divergence sites were not significant at 0.05 level, but 214 type I functional divergence sites were significant at 0.01 level. Taken together, there was type I functional divergence between five groups, and purifying selection played a dominant role in each of five groups during *DHDPS* gene evolutionary process.

Key words: plant; gene; functional divergence; adaptive evolution; divergence time; selective pressure

植物赖氨酸合成相关酶活性和赖氨酸合成调控机制的研究揭示了天冬氨酸激酶(AK)和二氢吡啶二羧酸合成酶(DHDPS)基因是控制植物种子赖氨酸合成的关键基因^[1]。有研究表明,这2种酶的活动性都受到末端产物的反馈抑制,因此,它们已成为限制赖氨酸积累的瓶颈^[2]。近年来,反馈抑制不敏感的AK和DHDPS的发现,为利用基因工程手段提高植物赖氨酸含量开辟了新途径。目前,已在大肠杆菌、烟草等中发现不受赖氨酸反馈调节的DHDPS基因^[3-4]。许多研究已将高赖氨酸蛋白基因成功转化到水稻^[5]、生菜^[6]和马铃薯^[7]等物种中,结果显示,转基因植物种子总赖氨酸含量有不同程度提高。尽管对DHDPS基因已有较多研究,但未见植物DHDPS基因的适应性进化和功能分歧研究的报道。本研究以赖氨酸合成途径关键酶DHDPS的基因为研究对象,系统分析了植物DHDPS基因的选择压力、功能分歧及亲缘进化关系,以期探索植物赖氨酸合成的分子机制奠定基础,为植物DHDPS基因功能研究提供线索。

1 材料和方法

1.1 数据获取与进化树构建

本研究涉及的拟南芥、水稻与衣藻等11种植物DHDPS基因的相关序列来自公共数据库GreenPhyl(<http://greenphyl.cirad.fr/v2/>)和CoGe(<http://synteny.cnr.berkeley.edu/CoGe/>)。植物DHDPS基因氨基酸序列的多重比对由Clustal X软件完成,参数为默认值^[8]。采用邻接法(Neighbor-Joining method)构建系统发生树^[9],其输出借助于MEGA软件完成^[10]。此外,以中性进化率 6.03×10^{-9} 为标准,推算了同一进化支上同源基因对的分歧时间。

1.2 植物DHDPS基因的选择压力检测

按照PAML 3.15([http://abacus.gene.ucl.](http://abacus.gene.ucl.ac.uk/software/paml.html)

[ac.uk/software/paml.html](http://abacus.gene.ucl.ac.uk/software/paml.html))软件的格式要求,建立正确的序列比对和进化树文件。然后,利用CODEML程序中的位点特异模型分别对植物DHDPS基因不同类群的选择压力进行检测,并寻找受到正选择作用的密码子位点^[11]。

1.3 植物DHDPS基因的功能分化研究

根据前面所构建的系统发生树,对植物DHDPS基因进行分类。然后,利用DIVERGE 2.0软件对不同类群之间的功能分歧系数进行评价,并鉴定发生功能分歧的氨基酸位点^[12]。

2 结果与分析

2.1 植物DHDPS基因的进化分析

为了解不同模式植物DHDPS基因成员间的进化关系,将莱茵衣藻、拟南芥、水稻等11个物种的DHDPS蛋白序列用于构建基因进化树。基因进化树揭示了植物DHDPS基因具有5个类群,分别被命名为A、B、C、D和E类群(图1)。尽管基因进化树并不能完全揭示物种之间的进化关系,但它可以揭示基因之间的系统发生关系。从图1可以看出,有7对同物种的DHDPS同源基因均位于进化树的同一分支,这说明这些基因均起源于物种分化之后(表1)。此外,还发现大豆2个DHDPS基因和拟南芥该基因聚在一起,可以推断这些基因起源于物种分化之前。

为推算同一进化支上的同源基因对的分歧时间,利用计算公式 $T = Ks / (2 \times \lambda)$ ($\lambda = 6.03 \times 10^{-9}$)^[13]推算植物DHDPS同源基因对的分歧时间(表1)。对同源基因对的分歧时间和物种形成的大致时间进行比较分析,结果显示,除了Glyma01g34610.1和Glyma03g02550.1起源于物种形成前以外,剩余的6个同源基因对很可能起源于物种形成后。

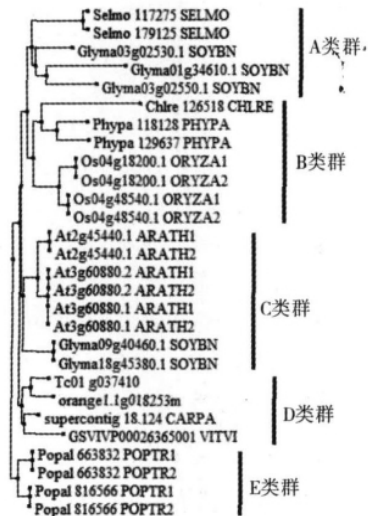


图 1 11 个模式物种 DHDPS 氨基酸序列的进化树

表 1 不同植物同源 DHDPS 基因对的分歧时间

基因 1	基因 2	同义替代率	分歧时间 / 百万年
Selmo_117275	Selmo_179125	0.018 4	1.53
Glyma09g40460.1	Glyma18g45380.1	0.109 7	9.10
Os04g18200.1	Os04g48540.1	0.363 1	30.11
Popal_663832	Popal_816566	0.378 9	31.42
Glyma01g34610.1	Glyma03g02550.1	0.634 6	52.62
At2g45440.1	At3g60880.1	0.925 3	76.73
Phypa_118128	Phypa_129637	1.105 4	91.66

2.2 植物 DHDPS 基因的选择压力

利用 M0&M3 和 M7&M8 2 对模型, 分别对植物 DHDPS 基因进化树的 5 个类群进行选择压力分析, 结果见表 2。似然比检验表明: M3 显著优于 M0,

说明位点间承受的选择压力具有异质性; 但 M8 并不优于 M7, 这说明 DHDPS 基因没有经历正选择。M3 模型在 5 个类群中都能检测到正选择位点, 均达到极显著水平; M8 模型在这些类群中虽然能检测到正选择位点, 但均未达到显著水平(表 2)。M8 模型在 5 个类群中没有检测到正选择位点。选择压力检测结果显示, 不同类群 A、B、C、D 和 E 所包括的 DHDPS 基因均受制于净化选择, 具有较高的保守性。

2.3 植物 DHDPS 基因的功能分化

为了保证不同类群中序列大于或等于 4 个序列, 本研究对水稻、拟南芥和杨树 DHDPS 基因进行重复处理(不会影响基于进化速率相关的功能分歧系数)。将上述进化树的 5 个类群进行两两组合, 利用 DIVERGE 软件对任意 2 个组之间的功能分歧系数进行计算(表 3)。I 型功能分歧分析显示: 除了 B 与 C 或 D 类群之间的 I 型功能分歧系数较小以外, 其他类群之间的 I 型功能分歧系数均大于 0.5; 尽管 B 与 C 或 D 类群之间 I 型功能分歧度较小, 但似然比检验表明它们之间 I 型功能分歧已达显著水平; 大多数类群组合之间都能够检测到 I 型功能分歧位点, 尽管 B 与 D、C 与 D 类群之间 I 型功能分歧已达显著水平, 但并没有检测到相应的 I 型功能分歧位点。II 型功能分歧分析显示: 具有 II 型功能分歧位点的类群组合的 II 型功能分歧系数都较小, 虽然 DIVERGER 软件并没有给出显著性检验, 但是可以看出它们的标准差都相对较大, 因此, II 型功能分歧位点的识别需要进一步确证(表 3)。

表 2 检测 DHDPS 基因编码区位点的选择压力

类群	序列数	M0 模型的 ω	M3 和 M0 似然比 $2\Delta l = 2(\ln L_{M3} - \ln L_{M0})$	M8 和 M7 似然比 $2\Delta l = 2(\ln L_{M8} - \ln L_{M7})$	M8 模型评价	正选择位点 (M8 模型)
A	5	0.260	47.782**	0.122	$p1=0.0428, \omega=3.379$ $\beta(p=0.357, q=0.491)$	58F, 59V, 74G, 78R
B	5	0.0675	170.624**	1.023	$p1=0.00554, \omega=9.226$ $\beta(p=0.269, q=2.804)$	40A, 115A, 124T, 190K, 247S
C	4	0.0688	45.407**	0	$p1=0.00001, \omega=1.0$ $\beta(p=0.317, q=3.244)$	21V, 176L, 311V
D	4	0.0962	65.714**	0.337	$p1=0.047, \omega=1.0$ $\beta(p=0.277, q=2.928)$	1M, 4N, 7A, 10K, 14L, 8C, 19K, 26K, 189S, 245D, 269K, 277A
E	3	0.0754	56.836**	0	$p1=0.00001, \omega=1.0$ $\beta(p=0.210, q=1.60)$	29S, 38L, 177C, 342N, 358Y

注: $\ln L$ 是最大似然值的对数, $2\Delta l$ 是 M3 和 M0 及 M8 和 M7 间 $\ln L$ 之差的 2 倍, * 表示似然比检验显著, ** 表示似然比检验极显著。

表 3 植物 DHDPS 基因组间的功能分化

组间比较	I 型功能分歧系数				II 型功能分歧系数	
	$\theta_1 \pm S.E.^a$	LRT	P	$Q_k \geq 0.95^b$	$\theta_{II} \pm S.E.^a$	$Q_k \geq 0.95^b$
A/B	0.999 ± 0.199	25.316	$P < 0.01$	67	0.580 ± 0.139	0
A/C	0.999 ± 0.291	11.754	$P < 0.01$	67	0.234 ± 0.149	0
A/D	0.999 ± 0.353	7.998	$P < 0.01$	67	0.182 ± 0.171	14
A/E	0.818 ± 0.235	12.065	$P < 0.01$	2	0.340 ± 0.124	0

续表 3 植物 DHDPS 基因组间的功能分化

组间比较	I 型功能分歧系数				II 型功能分歧系数	
	$\theta_1 \pm S. E. ^a$	LRT	P	$Q_k \geq 0.95^b$	$\theta_{II} \pm S. E. ^a$	$Q_k \geq 0.95^b$
B/C	0.186±0.084	4.884	$P<0.05$	1	0.071±0.147	7
B/D	0.418±0.188	4.971	$P<0.05$	0	0.074±0.162	9
B/E	0.790±0.167	22.373	$P<0.01$	2	0.178±0.126	11
C/D	0.510±0.168	9.253	$P<0.01$	0	0.048±0.093	5
C/E	0.778±0.172	20.390	$P<0.01$	5	0.091±0.070	7
D/E	0.819±0.220	13.917	$P<0.01$	3	0.095±0.080	6

注:a 为类群间的功能分歧系数与标准误,b 为满足后验概率 $Q_k \geq 0.95$ 的氨基酸位点个数。

3 讨论

基因重复在植物进化过程中扮演着十分重要的角色^[14]。正如孙红正等所综述的那样,尽管不同的重复基因进化模型被提出,但每一种模型都不可能很好地解释所有重复基因的进化模式^[15-17]。无论重复基因采用哪一种进化模型,最终人们关注的是基因的进化命运,即选择压力和功能分歧。显然,本研究对象 DHDPS 基因在进化过程中经历了重复扩增,尤其是在突变体植物中。因此,考察植物 DHDPS 基因的选择压力和功能分歧成为本研究的重点内容。DHDPS 基因的功能歧化分析揭示了不同类群之间具有较高的 I 型功能分歧系数,这充分说明植物 DHDPS 基因不同类群之间已经经历了 I 型功能分化。进化树显示,除了大豆 DHDPS 同源基因对属于 2 个不同类群外,其他植物 DHDPS 同源基因对均被划入同一类群,并位于进化树末端的同一分支,因此,推测植物 DHDPS 基因功能分化主要原因是物种分化引起的,但大豆的两类基因之间的功能分化则是由基因重复后适应性进化引起的。同一类群内正选择压力分析表明,虽然 M8 模型在植物 DHDPS 基因中检测到正选择位点,但是没有达到显著水平,故而该基因主要受制于纯净选择作用,具有较高的保守性。虽然本研究结论需要试验证据的支持,但它为植物 DHDPS 基因功能研究提供了线索。

参考文献:

[1] Galili G. Regulation of lysine and threonine synthesis [J]. Plant Cell,1995,7:899-906.
[2] Bryan J K. Synthesis of the aspartate family and branchedchain amino acids[M]// The biochemistry of plants. New York:Academic Press,1980:403-452.
[3] Cassan M,Parsot C,Cohen G N,*et al.* Nucleotide sequence of the *lysC* gene encoding the lysine sensitive aspartokinase III of *Eschench coli* K12: Evolutionary pathway leading to three isofunctional enzymes[J]. J Biol Chem,1986,261:1052-1057.
[4] Ghislain M, Frankard V, Jacobs M A. Dinucleotide mu-

tation in dihydrodipicolinate synthase of *Nicotiana sylvestris* leads to lysine overproduction [J]. Plant J, 1995,8(5):733-743.
[5] 蒋家焕,刘峰,许明,等. 高赖氨酸蛋白基因遗传转化水稻的研究[J]. 福建农林大学学报,2006,35(6):615-618.
[6] 李兴涛,李霞,张金文,等. 高赖氨酸蛋白基因在转基因生菜中的表达和遗传转化[J]. 应用与环境生物学报,2006,12(4):472-475.
[7] Perl A,Shaul O,Galili G. Regulation of lysine synthesis in transgenic potato plants expressing a bacterial dihydrodipicolinate synthase in their chloroplasts [J]. Plant Molecular Biology,1992,19:815-823.
[8] Thompson J D,Gibson T J,Plewniak F. The CLUST-AL_X windows interface;Flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. Nucleic Acids Res,1997,25:4876-4882.
[9] Saitou N,Nei M. The neighbor-joining method;A new method for reconstructing phylogenetic trees[J]. Mol Biol Evo,1987,4(4):406-425.
[10] Tamura K,Dudley J,Nei M,*et al.* MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0[J]. Molecular Biology and Evolution,2007,24:1596-1599.
[11] Yang Z. PAML4:Phylogenetic analysis by maximum likelihood[J]. Molecular Biology and Evolution,2007,24(8):1586-1591.
[12] Gu X. Statistical methods for testing functional divergence after gene duplication [J]. Molecular Biology and Evolution,1999,16(12):1664-1674.
[13] Blanc G, Wolfe K H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes[J]. Plant Cell,2004,16(7):1667-1678.
[14] Ohno S. Evolution by gene duplication [M]. Berlin: Springer-Verlag,1970:1-160.
[15] 孙红正,葛颂. 重复基因的进化——回顾与进展[J]. 植物学报,2010,45(1):13-22.
[16] Hughes A L. The evolution of functionally novel proteins after gene duplication[J]. Proc Biol Sci,1994,256:119-124.
[17] Force A,Lynch M,Pickett F B,*et al.* Preservation of duplicate genes by complementary, degenerative mutations[J]. Genetics,1999,151:1531-1545.