# 黄花蒿 HDS 基因的克隆与功能分析

张祖荣<sup>1,2</sup>,廖志华<sup>2</sup>,彭梅芳<sup>2</sup>

1. 重庆文理学院 生命科学系, 重庆 402168

2. 西南大学生命科学学院 三峡库区生态环境教育部重点实验室, 重庆 400715

摘 要:目的 克隆获得黄花蒿 MEP 途径中必需关键酶——羟甲基丁烯基-4-磷酸合成酶基因 (HDS),并进行生物信息 学分析和功能互补分析研究。方法 对已知的其他种子植物 HDS 基因的核苷酸序列进行多重序列比对,选取保守区域设 计简并引物,利用同源扩增和 cDNA 末端快速扩增技术从黄花蒿中获得目的基因;利用 BLAST 进行序列比对,ORF Finder 寻找开放阅读框,并用 MEGA3.0 中的临位相联法构建进化树。结果 得到 1 条长 2 324 bp 的 HDS cDNA 序列,其 ORF 框长 1 854 bp,编码 617 个氨基酸残基的蛋白;生物信息学分析显示,黄花蒿 HDS 基因 AaHDS 与其他种子植物来源的 HDS 高度同源;功能互补分析表明,AaHDS 能互补突变菌株 Escherichia coli MG1655 ara <>HDS 中缺失的 HDS 功能,使突变菌株恢复生长,证明 AaHDS 具有典型的 HDS 基因功能。结论 首次克隆获得黄花蒿 HDS 基因,为青蒿素的代谢 工程研究提供相应的基础。

关键词:黄花蒿;羟甲基丁烯基-4-磷酸合成酶基因;基因克隆;生物信息学分析;青蒿素 中图分类号:R282.21 文献标志码:A 文章编号:0253-2670(2012)01-0148-07

# Cloning and function analyses of HDS gene from Artemisia annua

ZHANG Zu-rong<sup>1, 2</sup>, LIAO Zhi-hua<sup>2</sup>, PENG Mei-fang<sup>2</sup>

- 1. Department of Life Science, Chongqing University of Arts and Sciences, Chongqing 402168, China
- Key Laboratory of Eco-environments in Three Gorges Reservoir Region, Ministry of Education, School of Life Sciences, Southwest University, Chongqing 400715, China

**Abstract: Objective** To obtain the indispensable key enzyme—hydroxide methyl enylamino 4-cyclodiphosphate synthase (*HDS*) gene involved in the MEP pathway cloned from *Artemisia annua* and conduct bioinformatic and functional complementation analysis. **Methods** To perform multiple sequence alignment for the nucleotide acid sequence of the other reported seed plants' *HDS* gene, to select conservative areas for designing degenerate primers, and to gain the aim gene from *A. annua* through homologous expanding and cDNA bottom speedily expanding technique. To perform sequence alignment using BLAST, to identify open reading frame (ORF) using ORF Finder, and to construct phylogenetic tree using neighbor joining (NJ) ways in MEGA3.0. **Results** The obtained *HDS* cDNA sequence was 2 324 bp containing a 1 854 bp ORF and encoding a 617-amino acid protein. Bioinformatic analysis showed that *AaHDS* was homologous with *HDS* derived from other seed plant species. Functional complementation analysis indicated that *AaHDS* could make up the short *HDS* function of mutant *Escherichia coli* MG1655 ara<>*HDS*. It could make the mutant get back to upgrowth, which showed *AaHDS* had typical *HDS* gene function. **Conclusion** The cloning *HDS* gene from *A. annua* for the first time provides a good basis for further study on the metabolization project of artemisinin.

Key words: Artemisia annua L.; HDS; gene cloning; bioinformatic analysis; artemisinin

青蒿素是从黄花蒿 Artemisia annua L.中提取的 一种治疗疟疾的特效药物<sup>[1-3]</sup>,近年来的研究还发现 青蒿素具有免疫抑制、抗血吸虫、抗病毒及抗肿瘤 等多方面的药理作用<sup>[4]</sup>。但天然药源黄花蒿中青蒿 素量很低(0.01%~0.6%)<sup>[5]</sup>,大大限制了商业化的 生产,青蒿素的生产远远不能满足国际市场的需 求<sup>[6]</sup>。因此寻找和扩大青蒿素药源是目前亟待解决 的问题。化学合成青蒿素,成本高、难度大;组织 培养方法也不能有效提高青蒿素的产量,目前青蒿 素来源仍是依赖从栽培黄花蒿中提取<sup>[7-9]</sup>。

收稿日期: 2011-03-31

基金项目:重庆市教委自然科学研究资助项目(kj071204)

作者简介:张祖荣(1966—),男,重庆江津人,重庆文理学院副教授,长期从事药用植物的教学与科研工作,现为西南大学访问学者。 Tel:13658304114 E-mail:412587791@qq.com

近年来,随着青蒿素生物合成途径相关酶基因的克隆,基因工程成为提高青蒿素产量的有效途径之一<sup>[10]</sup>。为研究青蒿素生物合成的分子机制,以及为青蒿素代谢工程提供候选基因和作用靶点,本研究采用 RACE 技术<sup>[11]</sup>从黄花蒿中克隆获得了青蒿素代谢途径中的一个关键酶基因——羟甲基丁烯基-4-磷酸合成酶基因(*HDS*),并进行了相应的生物信息学分析和功能验证。

## 1 材料与方法

# 1.1 材料

黄花蒿全株采集于西南大学生命科学学院的重 庆酉阳科研教学基地,用液氮速冻后在-70 ℃保存 备用。大肠杆菌 *Escherichia coli* T. Escherich 菌株 DH5 和 M15、质粒 pQE30、突变菌株 *E. coli* MG1655 ara<>HDS 以及其他实验材料均由该实验室廖志华 教授鉴定,由西南大学生命科学学院天然产物与代 谢工程实验室保存和提供。

### 1.2 方法

**1.2.1** 黄花蒿总 RNA 的提取 取-70 ℃冰箱中保存的黄花蒿叶、茎混合材料 0.1 g,用 RNAplant 试剂盒(天根生物,北京)提取总 RNA,保存于-70 ℃ 冰箱备用。

**1.2.2** 黄花蒿 HDS 基因 (AaHDS) 核心片段的获得 在美国国家生物技术信息中心 (NCBI) 网站上分别 下载已报道的 HDS 基因的核苷酸序列[如甜菊 Stevia rebaudiana (Bert.) Hemsl.、 拟南芥 Arabidopsis thaliana (L.) Heynh.、细长聚球藻 Synechococcus elongates Nag. 等], 使用 Vector NTI 8.0 进行多重序 列比对,选取最保守核苷酸序列设计简并引物<sup>[12]</sup>, 正向引物为dfAaHDS: 5'-TATGG(A/G)(A/C)G(A/G/-T/C)GCAATGCG(C/A/T)ATTG-3',反向引物为 drAaHDS: 5'-CAATCTTTCC(A/G)GG(A/T)G(C/A)-ACCACC-3'。cDNA 第一链合成使用 TaKaRa RNA PCR Kit(AMV), 按照试剂盒操作手册合成 cDNA 链。以 cDNA 为模板进行 PCR 扩增,反应条件为: 94 ℃预变性 4 min, 94 ℃变性 45 s, 50~60 ℃温 度梯度退火 45 s (每个循环降低 0.5 ℃), 72 ℃延 伸 1 min, 共 30 个循环; 最后 72 ℃总延伸 10 min。 PCR 产物纯化按照上海赛百盛基因技术有限公司 的 PCR 产物纯化试剂盒说明操作,将回收产物与 pMD18-T 连接后转化 DH5α,进行蓝白斑筛选,挑 取白斑进行菌落 PCR 检测, 阳性克隆送至上海英骏 生物技术有限公司测序,测序结果进行 BLAST 分 析以获得 AaHDS 核心片段。

**1.2.3** *AaHDS* 全长 cDNA 的获得 根据 AaHDS 核心片段序列设计 RACE 引物。AaHDS3-1: 5'-TGCGGGTGCCAATACTGGTG-3'; AaHDS3-2: 5'-CTGCTACAAGGATGCAGGATG-3'; AaHDS5-1: 5'-GAGCCATCACGGTGGAGTACACC-3'; AaHDS5-2: 5'-CCTCTTCACCCTCCTTCTGCAC-3'。 RACE-PCR 扩增按照 SMART RACE cDNA(Clontech,美国) 扩增试剂盒的说明书进行。RACE 产物的纯化、克 隆及测序与"1.2.2"相关操作相同。将 3'RACE、 5'RACE 和核心片段进行电子拼接(Vector NTI Suite 8.0),获得全长 cDNA 序列,根据拼接序列设计全 长特异性引物,正向引物 fiAaHDS: 5'-TTCTGAAGGGAGTCCTCTGTT-3';反向引物 riAaHDS: 5'-CCATCACCTTTAACTAGACGGT-3', 克隆获得物理全长 cDNA 序列。

**1.2.4** *AaHDS* 序列分析 序列比对均使用 Vector NTI Suite 8.0 软件和 BlastP2.2.3 (http://www.ncbi. nlm. nih. gov) 进行分析;开放性阅读框 (open reading frame, ORF) 的查找和核苷酸的翻译在 http:// www. ncbi. nlm.nih.gov 网站的 ORF Finder 上完成;用 CLUXTALX<sup>[13]</sup>进行多重序列比对,用 MEGA3.0 中的 临位相联法 (neighbor-joining, NJ)<sup>[14]</sup>构建进化树,重复次数为 1 000 次;蛋白质基本性质及转运肽的分 析使用 http://www.expasy.org 网站提供的相关生物信 息学分析软件进行<sup>[15]</sup>。

1.2.5 AaHDS 基因的功能验证 大肠杆菌突变菌株 MG1655 ara <>HDS中,内源HDS基因被一个卡那 霉素抗性基因所替代,染色体中仅含有一个单拷贝 的受到 PBAD 启动子控制的 HDS 基因<sup>[16]</sup>。MEP 途径 存在于大肠杆菌中, HDS 基因是大肠杆菌生存所必 须的<sup>[17]</sup>, MG1655 突变菌株只能在含有阿拉伯糖 (Ara)的培养基上生长,而不能在含有葡萄糖 (Glc) 的培养基上生长<sup>[16]</sup>。为了验证 AaHDS 的催化活性, 在大肠杆菌突变体 MG1655 中进行互补试验<sup>[18]</sup>: AaHDS的编码区被构建到原核表达载体 pQE30上, 并转化至大肠杆菌突变体 MG1655 中,获得 AaHDS 功能表达的工程菌,同时以空质粒 pQE30 转化大肠 杆菌突变菌株 MG1655 ara <>HDS 感受态作为阴性 对照,采用划线培养的方式将 MG1655 ara <> HDS 分别接种于 LB+Kan+0.2% Ara 与 LB+Kan+ 0.2%Glc 平板上,将携带 pQE30-AaHDS 质粒和空质 粒的 MG1655 ara <>HDS 分别接种于 LB+Kan+

Amp+0.5 mmol/L IPTG+0.2% Glc 平板上,过夜培养后,观察生长情况以验证 *AaHDS* 的功能。

# 2 结果与分析

## 2.1 总 RNA 的提取

以黄花蒿茎、叶混合取样为材料,按照 RNAplant buffer 试剂盒操作说明提取黄花蒿总 RNA,将抽提的 RNA 取 5 μL 进行电泳检测, RNA 明显呈 3 个条带, 无降解现象,无蛋白杂质, 28S 条带明亮(图 1),说 明获得了高质量的 RNA;再用紫外法进行定量测 定,*A*<sub>260</sub>/*A*<sub>280</sub>值均在 1.8~2.0,表明所得 RNA 的纯 度达到了相应实验的要求。

### 2.2 基因 AaHDS 的克隆

黄花蒿总 RNA 反转录为 cDNA 后,以 df*AaHDS*和 dr*AaHDS*为引物, cDNA 为模板进行 PCR 扩增。将获得的扩增条带进行目的片段回收, 亚克隆后的测序结果表明该片段长1 420 bp。经 BLAST 分析初步鉴定为黄花蒿 *HDS* 基因核心片 段,并命名为 AaHDS。根据所得片段设计基因特 异性引物用于扩增 AaHDS 的 cDNA 末端,通过 3'-RACE 和 5'-RACE 分别获得 602 bp 的 3'-末端 和 1 045 bp 5'-末端。将 AaHDS 基因 3 个片段拼 接后获得其全长 cDNA,并最终扩增获得其物理 全长,测序结果表明 AaHDS 基因全长为 2 324 bp (图 2)。

### 2.3 基因 AaHDS 的生物信息学分析

AaHDS的 cDNA 全长 2 324 bp, 包含一段长



#### 图 1 黄花蒿总 RNA 电泳图

#### Fig. 1 Electrophoresis of total RNA from A. annua

acgcgggatcgcggaacttcggcctgcttc tgaagggagtoctotgttagttoccgtacagaagtattgtgaatccacatataaaaccatcaggaggaaaaccogcacagtaatggttgg 31 ag atgt acc tot tggt ago gat cacc oca taag aat toa aaca atg acac aac tga tact aag gaog tgg otgo aac tgt ag as 1 AT CCAA AT COCT GACA GAG GAG CTGA CAT TGTT CGA ATA ACTG TOC AAGG AAA AAG AG AG GCA GAT GCAT GTTT TGA AAT CAA GAAC ACC D R G A D I V R I T V Q G K R E A D A C MQIA F E Τ ĸ N 01 CT TGTT CAGAAGAATT ATAACA TACC TCT OGTGGCCA GAC ATTC ATT TTOC TCC ATC AGTT CCA CT CG AGTTCC TGA ATGCTT TGACAAA QKN YNIP VADIHF A P S V ALRV L E D 91 AT TOGT GTC AAC COOGGAGATT TTGC TGA TOGC OGA GCT CAGT TTGAGCA GTT GGA GTAT ACA GAAGATGATTA TCA ACA GGA ACTT GAG NP GDE D R Q F D 81 CA TATT GAA AAG GTTT TTACTC CATT GGT T GAA AAA T GC CAAGA AAT AT GG GAAGGC CAAT GC GAA CAA ACCA T GG GAG T CTT T CA MR TPL VEKC кк Y G R A 71 GA TOGT ATC ATG AGT T ACT ATG GTGA TTC TOCT AGG GGA ATGG TTG AGT T GGA T TC GCAA GAA TCTG T AGG AGT AGAC T TC R MS Y YGDSP R GM v E S A F E F A R 61 CA TAAC TITI GTA TITT CAA TGA AAGC AAGC AAT CCAGTGATCA TGG TTCAAGC ATA TCGT CTT CTAG TGG CCGA AAT GAA TGT TCAGGGA V F S M K A S N P V I M V Q A Y R L L HNE v E M N 51 TGOGAC TATCCT TTGCACT TOGGOGT TAC TGAA GCT GGC GAGGGTGAGGA TGGACG TATGAAA TCTGCAA TTGGCAT TGGGAC ACTTCTT D Y P L H L G V T E A G E G E D G R M K S 41 CAGGAT GGT CTT GGTGATA CTA TCAGGGT TTOCCTC ACAGAACCACCAGA GGA AGA AATA GACCOCT GTA GGAAATT GGCCAA CCTT GGT P ТЕ P E E D R E 31 AT GAAA GCA TOC CAAC TTC AAC AAGGAGT GOCA COG TTT GAAGAAA AGCA CAGGOG TTA T TTT GATT TTC CAACGGAGAAC TOG TGA T TTG M K A S Q L Q Q G V A P F E E K H R R Y F D F 0 RR 21 OC TGT OCAGAAGGAGGGTGAAGAGGT AGA TTA CAGA GGT GTAC TOCACOG TGA TOGCTC TGTT ATCA TGT CTG TTAC CCT OGA TCAA CTC KEGEEVDYRGVLHRDGSVIM S w. 11 AA GACCOCT GAACTC TTCT ACAGATC ATT AGC AGCA AAACTTG TTC TTGG TAT GOC ATT TAAGGATC TTGC CAA CAGT TGA TTC TATC TTA K T P E L F Y R S L A A K L V L G M P F K D L A T V D S I 01 TTAAGAGAGGCTTOCACCAGCAGAGAGA TAAAGA TGCTCGTTTAGCTCCTTAAGAGGGTTGATCGATGTAAGTATGGGAATTATTACCCCT TTG ADDKDARLAL LPP KRL D MG 91 TCAGAACAGTTGACAAAGCCTT TGCCCAA OGC TATTGTC TTGG TAAACTT GAA TGAACT ATCCACAGGAGCTCATAAGCT TTT GCCAGAA Q L T K P L P N A I V L V N L N E L S T G нк R A LL P 81 OGCACAOGT TTGGTGGTGTCAGTAOGTGGTGA TGAGCCT TATGAGGAGCT TGA TAT TCT TAAAAOCA CTGATGCTACAATGAT TCTTCAT VSVRGDRPYRRLDT ктт TR w. т D 71 GAACTOCCTTATOCTGAAGAAAAAACTOGCAGAGTACATOCTGCTAGCAGGTTGTTTGAGTACCTTTCTGAAAATTCACTGAA TTTCCCT EEKT GR HAARRL E A L 6] GT TCT TCACCACATAAAGT TCCCAAAAGGAAT TCCCACGGATGATT TGGT TAT CAG TGCCGGTGCCCAATGCTGGTGC TTT AT TAGTTGAC V L H H I K F P K G I P R D D L V I S A G A N A G A L L V D 51 OGACT TOGA GAT OGT ATT TAT TAGA AGC TOC OGAT CAA GACT TOGAATT TAT CAGGAA TACA TCTT TCAATC TOCT ACA AGGATOC AGG ILLEAPDODFEFIRNTSF LG DG NL 4] A TOCGCAAC ACT AAGACAGAAT A TGT CTC ATGCOCA TCT TGTGGTAGAAC TCT ATT TGA TCTT CAAG TGA TTA GTGC CGA AAT TAGA GAA F VSCP S С G R T. F D Q 3.5 3] AAAACATCACAT TTGOCTGGTG TCTCGAT TGCGATTATGGGTT GCA TAGTCAA TGGACCAGGGGAGA TGGCTGACGCAGA TTT TGGA TAT SHLPG v s IA IMGCIVNGP GEMA DA D G 2] GTCOGTGGTGCTCCTCGAAAGA TTGA TCT TTA TGTTGGAAAGACOGTGGTACAACGACGAGGAATTGCAA TGGAAGGTGCAACCGA TGGA TTG V G G A P G K I D L Y V G K T V V Q R G I A M E G A T D A L 11 ATTCAACTAATCAAACAACCATGGCCGCTGGGTCGATCCTCCTGTTGAAGAGTAGaggagtgcatgctccatgtataa I Q L I K D H G R W V D P P V E E \* 2281 gatat gaaaag gage caagaat teaaat caaaaaaaaaaaaaaaaaa 编码区和翻译的氨基酸用大写字母表示,非编码区(UTR)用小写字母表示,终止密码子用\*表示

Coding and its deduced amino acid sequences are shown by capital letters, UTR is shown by lowercase letters; \* indicates stop codon

### 图 2 AaHDS cDNA 全长序列和由此推测的氨基酸序列

Fig. 2 AaHDS cDNA full-length sequence and its deduced amino acid sequence

度为 1 854 bp 的 ORF,还有长度为 210 bp 的 5'-端非翻译区(untranslated region, UTR)和 260 bp 的 3'-UTR,终止密码子为 TAG (图 2)。*AaHDS* 编码含 617 个氨基酸残基的蛋白,其相对分子质量预测为 6.87×10<sup>4</sup>,等电点预测为 5.3。在 Expasy 中进一步预测 *AaHDS* 的二级结构<sup>[19]</sup>,显示其中的随机卷曲占 45.38%,α-螺旋占 40.19%,延伸链占 14.42%(图 3)。将 *AaHDS* 与来源于其他植物 以及藻类、细菌的 *HDS* 进行多重序列比对。

结果图 4 表明,来源于植物与来源于藻类和细菌

的 HDS 差异较大,尤其是在 N 端,植物中 HDS 的 N 端比细菌和藻类多出一段大约含 80 个氨基酸残基 的区域,这段序列的功能已经在拟南芥 A. thaliana 中得到了证实<sup>[20]</sup>,是引导 HDS 定位于质体的植物 特有的转运肽,这与在植物中 MEP 途径定位于质 体中的观点一致<sup>[21]</sup>;同是来源于植物的 HDS 在 N 端的氨基酸序列差异也较大,主要因为这部分序列 是非催化区域,而催化区域的 HDS 氨基酸序列非常 保守,其保守一致性高达 80%~95%。

用 Clustalx 和 MEGA3.0 软件对 AaHDS 进行分

> h-α-螺旋, e-延伸链, c-随意卷曲 h-alpha helix e-extended strand c-random coil

#### 图 3 AaHDS 的二级结构预测

#### Fig. 3 Secondary structure prediction of AaHDS

	(1) 1		1	10		20	20		30			40			50			66		
拟南芥	(1) M	ATGV	LPAPV	SGI	KIPD	SKVGF	GKSM	NLVR	ICDV	RSLF	SARR		-RV	SVI	RNSN	1QGS	DLA	AELQ	P	
大肠杆菌	(1)																			
细长聚球藻	(1)																		-	
黄花蒿	(1)																		-	
甜菊	(1) M	ATGA	APTSF	MSLI	KGREI	NGLGF	AKTS	DFVK	VSDL	KRVF	FHRT		-KI	FVF	KNSF	SGS	DIA	AEFK	Р	
保守区	(1) M	A G	P	L	K	LGF	KS		V		R		K	v	ΚS	G	E	E	Ρ	
101-1-1-1-	(67) 6	7	-		80		,90		1	00		,110	_	-	,120	,		13	32	
扒用介 十匹打击	(62) A	SEGS	PLLVE	RŐK	YCES	LHKIV	RRKT	RTVM	VGDV	ALGS	EHPI	RIQT	MTT	SDT	KDIT	IGIN	DEN	/MR I	A	
入 肋 杆 困 如 上 取 球 遠	(1) =				-MIN	QAPIQ	RRRS	IKII	VGNV	PIGL	GAPI	AVUD	MIN	TRI		LAIV	IN Q 1	TRAL	-1-	
出 C 永 小 孫 昔 花 蒼	(1) =	m	QTEP3	SPVQ.	AIPI	LIAIV	REEL	RPVP	1624	AT GG	GUEV	AV	MITIN		LENI	LG SN	AAI	MOT	л. Л	
甜菊	(62) A	SEGS	PLLVE	VOK	YCES	THETI	RRKT	RTVM	VGDV	ALGS	DHPI	RIOT	MTT	SDT	KDV2	ATA	/E ON	ZMOI	A	
保守区	(67) A	S GS	LLVI	QK	YCES	H TV	RRKT	RTVM	VGNV	LGS	DHPI	RIQT	MTT	TDT	KDV	ATV	/EQV	/MR I	A	
	(133)	133	1	40		150		16	0		170		18	30				1	98	
拟南芥	(128)	DKG	ADIVR	ITVQ	GKKE	ADACE	EIKD	KLVQ	LNYN	IPLV	ADIH	FAP-	:	TVA	LRVA	ECF	DKI	RWN	2G	
大肠杆菌	(53)	RVG	ADIVR	VSVP	TMDA	AEAFI	(LIK-		QQVN	VPLV	ADIH	FDYR	:	IAL	KVAE	YGV	DCL	RINI	?G	
细长聚球藻	(63)	EIG	CEIVR	VIVP	SLAH	AKAMI	EIRD	RLYK	TYKP	VPLV	ADVH	HNGM	1	KIA	LEVA	KYV	DNV	RINI	?G	
黄花蒿	(5)	DRG	ADIVR	ITVQ	GKRE	ADACE	E <mark>I</mark> KN	TLVQ	KN YN	IPLV	ADIH	FAP-	:	SVA	LRVA	ECF	DKI	RVNI	?G	
甜菊	(128)	DRG	ADLVR	ITVQ	GKRE	ADACI	E <mark>I</mark> KN	TLVQ	KNYN	IPLV	ADIH	FAP-	:	SVA	LRVA	ECF	DKI	RVNI	?G	
保守区	(133)	D G	ADIVR	ITVQ	GKKE	ADACE	EIKN	TLVQ	KNYN	IPLV	ADIH	FAP		VA	LRVA	ECF	DKI	RVN	2G	
	(199)	199		2	10		220		23(	)		240		é	250			2	54	
拟南芥	(190)	NFAI	RRAQI	FETI	DYTE	DEYQK	ELQH	IEQV	FTPL	VEKC	KKY G.	RAMR	IGIN	NHG:	SLSD	RIM	SYY(	GDS-		
大肠杆菌	(111)	NIGN	1					-EER	IRMV	VDCA	RDKN	IPIR	I GVI	NA GS	5LEK	DLQI	EKY	GEPI	- :	
细长聚球藻	(128)	LYVE	EKPKI	PNRT	EYTQ	AEFDE	IGAK	IRET	LE PL	VISL	RDQG	KSMR.	I GVI	1HGS	SLAE	RMLI	FTY	GDT-	·	
黄花蒿	(67)	DFAI	ORRAQI	FEQL	EYTE	DDYQQ	ELEH	IEKV	FTPL	VEKC	KKYG	RAM <mark>R</mark>	IGIN	NHG:	SL SDI	RIM	5 Y Y (	GDS-		
甜菊	(190)	NFAI	RRAQI	FEQL	EYTD	DDYQK	ELEH	IEKV	FVPL	VEKC	KKYG	RAMR	IGIN	¶⊞GS	SL SDI	RIM	SYY(	GDS-		
保守区	(199)	NFAL	RRAQ	FE L	EYTD	DDYQK	ELEH	IE V	FTPL	VEKC	KKYG:	RAMR	IGTN	THGS	SLSD	RIM	SYY(	GDS		

	(265)	265	270		28	0	2	90		300		310		32	0	330	
拟南芥	(254)			F	RGM	/FSAFE	FART	CRKU	YHN	FVESM	ASNE	V T MV	OA YR	T.T.VAF	MYVHG	WDYP	
大肠杆菌	(154)			Î	OALI	ESAME	RHVDH	LDRL	IEDO	FKVSV	ASDV	FLAN	ESY-	F	LLAKO	IDOB	
细长寒球藻	(190)			8	EGMV	ESALE	FIRI	CESL	JEYN	LEISL	ASRV	PVMI	AANR	LMVKE	MDELG	MDYP	
- 山 () 永 小 休 去 士 吉	(131)				RGM	TESAFE	FART	CRKL	FHN	EVESM	ASNE	VTMV	OA YR	LUVAR	MNVOG	WDYP	
<b>央化</b> 尚 田井	(254)			F	RGM	TESAFE	FART	CRKL	FHN	EVESM	ASNE	VTMV	OA YR	I.I.VAF	MYVOG	WDYP	
<b>甜</b> 匊	(285)			Ť	DDCM	TEGATE	TOKE	CDVII	THN	FUESM	ZA SNID	V T MU	ON VD.	TTVAR	MVU C	WDVD	
保守区	(207)	397		-	410	LJATE	42	0	) E II IA	130	ADME	440	YAIK.	450	SMIV G	46 MD 1 F.	2
카루슈	(374)		FUUD	VED	FODD	TODID	VOVE	CEEVE	VDN		CTT M	eter	DOLK	100F, 1 TEGA	VD GT 7	TRIA	-
<b>拟</b> 闱介 士 昭 红 靑	(252)	AFEE	ENHA	11110	I XILL	TODE	V QUUE		TRU	VEHREC	12141214	0105	LQLIG	NE 1111	INCOMP	I IND V	ř
八肋竹 困 如 上 取 球 蕩	(289)	,															_
出 <b>以</b> 來 小 傑	(251)		FIZUDI	VED	FORD	TODID	VOVE	OFFUE	VDC			OTTT	DOT	TOFT	VD GT 7	7.727.72	-
胡菊	(274)	APPE	FRUDI	NILD	FORD	TCDLP	VQLE	CEEVE	VDC	VERKEG	SVDM	SVIL	FOLK	IPELI VDEL I	TROLF	ARLV	
保守区	(307	APPE	ENDR	VED	FORR	TOULD	VQRE	CEEVE	VD	VERKEG	SUPP	SVIL	POLK	DEL	VDGT 7	ARLV	<u>لل</u>
	(463)	463	250 EKIR 47	1110	LŐKK	480	VQRE	490	IR	50 SI	0	2421	510	PEL	IROLF	52	Ř
扣责茨	(440)	GMPFF	DLAT	VDSI	LLR	SLPEVI		ARLAL	KRLI	DVSMG	VIAPI	LSEC	LTKPI	LPNAN	IVLVNI	KELS	Ē
大肠杆菌	(252)																_
细长聚球藻	(289)																_
黄花蒿	(317)	GMPFF	KDLAT	VDSI	ILLRE	ELPPAI	DEKD	ARLAL	KRL	DVSMG	IITP	LSEQ	LTKPI	LPNAI	VLVNI	NELS	Т
甜匊 俣字区	(440)	GMPF	XDLAT	VDSI	IFLRE	SLPPVI	DEKD	ARLAL	KRL	DVSMG	VITP	LSEC	LTKPI	IPNAI	VLVNI	NELS	т
体可应	(403)	GMPFr 520	SDLAI	VD21	10	LPPVI	550	AKLAL	KKLI 560	DISMG	VI PI 570	LSEQ	LIKPI	LPNAP	IVLV I	. ELS	4
拟南茶	(506)	GAYKI	T.PFG	TRU	UVST.	RGDEP	YFFU	FTLKN	TR-2	TMTTH	TRUC	, NGTT	VSRVI	HAARF	⊺ननन्⊺ ≲	SENS	-
大肠杆菌	(252)																_
细长聚球藻	(289)												IP				_
黄花蒿	(383)	GAHKI	LLPEG	TRL	7VSVI	RGDEP	YEELI	DILKT	TD-7	TMILH	ELPY	AEEK	TGRVH	laarf	RLFEYL	SENS	L
甜菊	(506)	GAHKI	LLPEG	TRL	/VSLE	RGDEP	YEEL	EVLKS	ID-7	TMILH	DLPY	SEEK	TGRVH	AARF	RLFEYL	SENS	Ľ
保守区	(529)	GAHKI 595	600	TRL	VVSLI 6	RGDE 10	YEEL	EILK 620	D	TMILH 630	LPY	EEK 64	1GRVI 0	HAARE	ALFEYI 550	SENS	L 50
	(555)	NEDIZ	TUUTN		~ T U D		UZCT	VACCE	T 37D			DDOE		DNIEGI		CDM	2.7
拟南芥	(071)	NEPV.	THHI	IEE.L.	GIHR	DEPAT	HAGT.	IAGGE	LVD(	יטעטעני	MLEA	PDQL	EDEL	RNTSI	IN LEQU	JURMR	
大肠杆菌	(202)														DINKS	LRIR	2
细长聚球澡	(291)	NEDIZ			CTDD		C2C2	NACAT	T 17TN			DDOD				TGPR	N
黄花高	(571)	NEPVI		PRO	GIPR		DAGA	NAGAL		ST CDC1	LLEA	PDQL		RN 151	TNLEQU	CRMR	IN N
<b>胡匊</b> 4 京臣	(505)	NEPVI	LUUT	EPA	JUPR	DELVI	NAGA	AGAL		ST CDC1		PDQL		EN 151		CRMR	IN N
保守区	(661)	NE PV.	IUUT	670	IR	680 DDD01	AG	AGAL	90 11 A Di	נטעטענ	700	DQL	710	китэт г	SINTTAGO	JCRMR 72	28
	(001)	001		.070		.001		0	20		,00		1/11			,,,	
拟南芥											·						
十肠杆菌	(637)	TKTEY	(VS <mark>CE</mark>	SCG	RTL	F	CLQE:	ISAEI	REK	CSHLPG		v	SIAI	ØGCIN	NGPGE	MADA	.D
八成竹西	(637) (263)	TKTEY RGINE	VSCF FIACF	SCGI	RTL RQE	F	DLQE DVIG	ISAEI TVNAL	REK: EQRI	SHLPG LEDIII	 P	v m	SIAI DVSI	AGCIN IGOV	VNGPGE VNGPGE	MADA ALVS	T
细长聚球藻	(637) (263) (304)	TKTEN RGINH TMVEN	INSCE IACE VACE	SCGI TCSI SCGI	RTL-· RQE-· RTL-·	F F	DLQE DVIG NLEE	ISAEI TVNAL VLHKV	REK: EQRI REA:	SHLPG LEDIII KHLTG	 P	V M L	SIAI IDVSI NIAV	MGCIN IGOVN MGCIN	VNGPGE VNGPGE VNGPGE	MADA ALVS MADA	T
知长聚球藻 黄花蒿	(637) (263) (304) (514)	TKTEN RGINE TMVEN TKTEN	(VSCP FIACP (VACP (VSCP	SCGI TCSI SCGI SCGI	RTL RQE RTL RTL	F	DLQE DVIG NLEE DLQV	ISAEI TVNAL VLHKV ISAEI	REK: EQRI REA: REK:	ISHLPG LEDIIT IKHLTG ISHLPG	 P 	V M L	SIAI DVSI NIAV SIAI	MGCIN IGOVI MGCIN MGCIN	VNGPGE VNGPGE VNGPGE VNGPGE	MADA ALVS MADA MADA	D T D D
(1) (1) (1) (1) (1) (1) (1) (1) (1) (1)	(637) (263) (304) (514) (637)	TKTEN RGINE TMVEN TKTEN TKTEN	(VSCF FIACF (VACF (VSCF (VSCF		RTL RQE RTL RTL RTL	F	DLQE DVIG NLEE DLQV DLQV	ISAEI TVNAL VLHKV ISAEI ISAEI	REK EQRI REA REK RDK	CSHLPG LEDIIT CKHLTG CSHLPG CSHLPG	P	V M L V	SIAI DVSI NIAV SIAI SIAI	MGCIN IGOVI MGCIN MGCIN MGCIN	VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE	MADA ALVS MADA MADA MADA	
知行困 细长聚球藻 黄花蒿 甜菊 保守区	(637) (263) (304) (514) (637) (661)	TKTEN RGINE TMVEN TKTEN TKTEN	VSCF FIACE VACE VSCF VSCF	SCGI SCGI SCGI SCGI SCGI	RTL RQE RTL RTL RTL RTL	F F F	DLQE DVIG NLEE DLQV DLQV DLQE	ISAEI TVNAL VLHKV ISAEI ISAEI ISAEI	REK EQRI REA REK RDK RDK	ISHLPG LEDIIT IKHLTG ISHLPG ISHLPG	P	V M L V V	SIAI DVSI NIAV SIAI SIAI	MGCIN IGOVI MGCIN MGCIN MGCIN	VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE	MADA ALVS MADA MADA MADA	
细长聚球藻 黄花蒿 甜菊 保守区	(637) (263) (304) (514) (637) (661) (727)	TKTEN RGINH TMVEN TKTEN TKTEN TKTEN	IVSCP FIACE IVACE IVSCP IVSCP	SCGI SCGI SCGI SCGI SCGI	RTL- RQE- RTL- RTL- RTL- RTL RTL	F F F F F 40	DLQE DVIG NLEE DLQV DLQV DLQE	ISAEI TVNAL VLHKV ISAEI ISAEI ISAEI _750	REK EQRI REA REK RDK REK	ISHLPG LEDIII IKHLTG ISHLPG ISHLPG ISHLPG	P P  50	V L V V V	SIAI DVSI NIAV SIAI SIAI SIAI 770	MGCIN IGOV MGCIN MGCIN MGCIN	VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE	MADA ALVS MADA MADA MADA MADA	
细长聚球藻 黄花蒿 甜菊 保守区 拟南芥	(637) (263) (304) (514) (637) (661) (727) (691)	TKTEN RGINH TMVEN TKTEN TKTEN 727 ) F <b>G</b> YV	IVSCP FIACE IVACE IVSCP IVSCP IVSCP		RTL RTL RTL RTL RTL (74)	F F F F 40 LYVGF	CLQE DVIG DLQV DLQV DLQE CLQV	ISAEI TVNAL VLHKV ISAEI ISAEI ISAEI _750 KRGIZ	REK EQRI REA REK RDK REK	ISHLPG LEDIIT IKHLTG ISHLPG ISHLPG ISHLPG .76 ATDAI	P P  50	V L V V V	SIAI DVSI NIAV SIAI SIAI SIAI 770	MGCI IGOV MGCI MGCI MGCI MGCI MGCI	VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE	MADA ALVS MADA MADA MADA MADA 78	
(加) (加) (小) (小) (小) (小) (小) (小) (小) (小) (小) (小	(637) (263) (304) (514) (637) (661) (727) (691) (318)	TKTEN RGINE TMVEN TKTEN TKTEN 727 FGYV LGVI	VSCF FIACE VACE VSCF VSCF VSCF VSCF	SCG SCG SCG SCG SCG SCG	RTL- RQE- RTL- RTL- RTL- RTL RTL RTL RTL RTL RTL RKS(	F F F F 40 LYVGF GLYEI	DLQE DVIG NLEE DLQV DLQV DLQE (TVV) QVR)	ISAEI TVNAL VLHKV ISAEI ISAEI ISAEI 750 KRGI# KDRLI	REK EQRI REA REK RDK REK	ISHLPG LEDIIT IKHLTG ISHLPG ISHLPG ISHLPG .70 ATDAI MIDQ	P   50 LIGLI	V L V V IKEH IRAK	SIAI DVSI NIAV SIAI SIAI 770 GRWV	MGCI IGOV MGCI MGCI MGCI MGCI MGCI DEP	VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE VNGPGE R I DV	MADA ALVS MADA MADA MADA MADA 78 	
<ul> <li>(加长花蒿 黄花菊 保守区</li> <li>(加大花蒿 秋守区</li> <li>(加大市)</li> <li>(加大市)</li></ul>	(637) (263) (304) (514) (637) (661) (727) (691) (318) (358)	TKTES RGINE TMVES TKTES TKTES 727 FCYV ICVI	YVSCF PIACF YVACF YVSCF YVSCF YGGSI GGN- YGKQF		RTL RTL RTL RTL RTL RTL  RTL  RTL  RTL  RTL	F F F 40 LYVGF LYRGF	DLQE DVIG NLEE DLQV DLQV DLQE (TVV) GVR REEVI	ISAEI TVNAL VLHKV ISAEI ISAEI 750 KRGIZ KDRLI	REK EQRI REA REK RDK REK	ISHLPG LEDIIT IKHLTG ISHLPG ISHLPG ISHLPG ISHLPG ISHLPG ISHLPG ISHLPG ISHLPG ISHLPG ISHLPG	P  50 LIGLI LEAR VEL	V L V V V IKEH IRAK IRAK	SIAI) DVSI NIAV SIAI) SIAI (SIAI) (SIAI) (SIAI) (SIAI) (SIAI) (SIAI) (SIAI) (SIAI) (SIAI) (SIAI) (SIAI)	MGCI IGOV MGCI MGCI MGCI MGCI MGCI DEAH DEAH		MADA ALVS MADA MADA MADA MADA 78 	
八细黄甜保 拟大细黄甜 保 拟大细黄甜 保 刺肠杆浆 蒿 國家 南新杆浆蒿 國家 前期 化二乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基乙基	(637) (263) (304) (514) (637) (661) (727) (691) (318) (358) (568) (691)	TKTEY RGINH TKTEY TKTEY 727 FCYV ICVI YCYV	VVSCF VACF VVSCF VVSCF VVSCF VS		RTL RTL	40 LYVGF LYVGF	DLQE DVIG NLEE DLQV DLQV DLQE CTVVI CGVRI REEVI	ISAEI TVNAL VLHKV ISAEI ISAEI ISAEI SAEI KRGIZ KRGIZ KRGIZ QRGIZ	REK EQRI REA REK RDK REK AMTE ONNI PEAF	ISHLPG LEDIIT IKHLTG ISHLPG IS	P p 50 1GL LEAR VEL 10L	V L V V IKEH IRAK IRAD IKAD	SIAI DVSI SIAI SIAI SIAI 770 GRWV ASQI GRWV GRWV	MGCIY IGOV MGCIY MGCIY MGCIY MGCIY MGCIY DPPY TDPPY TDPPY		MADA ALVS MADA MADA MADA MADA 2 78 2 20 2 2	

一致氨基酸残基采用白色字体黑色背景表示,保守氨基酸残基采用黑色字体灰色背景表示,其他氨基酸残基均用黑色字体白色背景表示;★ 表示半胱氨酸残基保守位点

Identical amino acid residues were shown in white with black background and conserved amino acid residues were shown in black with gray background, other amino acid residues were shown in black with white background;  $\star$  indicates position of conserved cysteine residues

### 图 4 来自不同物种的 HDS 氨基酸序列多重比对

#### Fig. 4 Multi-alignment of HDS amino acid sequences from various plant species

子系统发育树分析,结果显示,来源于植物、藻类、细菌的 HDS 在发育进化树上明显各聚为一支(图5),并 且在进化树上的距离也与其所属类别的亲缘关系相符。 其中 AaHDS 与甜菊 S. rebaudiana 的 HDS 亲缘关系最接 近,这与黄花蒿和甜菊同属菊科植物的事实相符。

# 2.4 基因 AaHDS 的遗传功能互补分析

由于 MEP 途径对细菌的生长是必需的, 所以

HDS 基因对大肠杆菌的生长是必需的。而 E. coli MG1655 ara<>HDS 突变菌株经过人工改造后, 其的内源性 HDS 基因由卡拉霉素抗性表达盒取 代,因此其只能在含有 Ara 的培养基上生长,而 不能在含有 Glc 的培养基上生长。将携带 AaHDS 基因的重组表达质粒 pQE30-AaHDS 转化进突变 菌株 E. coli MG1655 ara<>HDS 后,该菌株获得

• 152 •



# 图 5 在 MEGA3.0 平台上用 NJ 方法构建的 HDS 分子进化树 Fig. 5 HDS phylogenetic tree constructed by NJ method on MEGA 3.0

外源性的 HDS 基因,在含有葡萄糖的培养基上可以 生长,而对照菌株则不能生长,从而证明 AaHDS 互 补了大肠杆菌缺失的 HDS 的功能。

大肠杆菌突变菌株 MG1655 ara <>HDS 能在含 有 0.2% Ara 的 LB 培养基上生长,却不能在含有 0.2% Glc 的培养基上生长 (图 6-a, b); 当将携带 AaHDS 编码区的质粒 pQE30-AaHDS 转入 E.edt MG1655 ara <>HDS 之后,能使大肠杆菌 HDS 突变菌株恢复 生长,在含有 0.2% Glc 和 0.5 mmol/L IPTG 的培养基 上能生长 (图 6-d),而作为对照的空载体 pQE30 则 不能在此培养基上生长 (图 6-c)。

### 3 讨论

青蒿素作为治疗疟疾的特效药,在国际市场上供不应求。自从青蒿素被发现以来,研究者们不断尝试 各种方法以期提高青蒿素量:采用化学合成青蒿素, 需要经过一系列繁杂的反应,且成本高、得率低、毒 性大;组织培养也不能有效提高青蒿素的产量。近年 来,随着基因转化技术的广泛运用,不少研究者致



图 6 AaHDS 基因的功能验证 Fig. 6 Functional demonstration of AaHDR gene

力于研究利用转基因技术和次生代谢工程手段来 提高黄花蒿中青蒿素的量,但其前提是必需对青 蒿素生物合成代谢途径的分子生物学和生物化学 有深入的了解<sup>[22]</sup>。

到目前为止, 青蒿素生物合成途径的研究已 取得了非常可喜的进展,其代谢途径上多个编码 关键酶的基因已被克隆<sup>[7]</sup>。HDS 催化 ME-cPP 转 化生成 HMBPP, 是青蒿素生物合成上游途径 MEP 途径中必需的一个关键酶<sup>[23]</sup>。本研究在前人 研究的基础上,采用 RACE 技术,以重庆酉阳的 优质高产黄花蒿为研究材料, 克隆获得了编码黄 花蒿 HDS 的 AaHDS 基因。对其编码的氨基酸序 列进行生物信息学分析,结果表明, AaHDS 与来 源于其他植物的 HDS 同源性高达 80%~95%, 且 具有典型的 HDS 催化活性位点,在发育进化树 上, AaHDS 与来源于植物甜菊的 HDS 亲缘关系 最接近。所有这些证据都表明,本研究对 AaHDS 基因的克隆是非常成功的。该成果为更加深入地 研究青蒿素生物合成代谢途径奠定了重要的分子 生物学基础,也为研究青蒿素的代谢工程提供了 更多可能的候选基因和调控靶点。

### 参考文献

- Klayman D L. Qinghaosu (artemisinin): an antimalarial drug from China [J]. *Science*, 1985, 228: 1049-1055.
- [2] Lin C X, Xiao P G, Peng Y, et al. Challenges in research and development of traditional Chinese medicines [J]. Chin Herb Med, 2009, 1(1): 1-28.
- [3] 王鸿博,肖 皖,华会明,等. 黄花蒿的化学成分研 究进展 [J]. 现代药物与临床, 2011, 26(6): 430-434.
- [4] 陆金健. 青蒿素类化合物抗肿瘤研究进展 [J]. 中国 药理学通报, 2010, 26(6): 818-820.
- [5] 赵 兵, 王玉春, 欧阳藩. 青蒿素生物合成机理研究 现状 [J]. 广西植物, 1999, 19(2): 154-158.
- [6] 刘春朝, 王玉春, 欧阳藩, 等. 青蒿素研究进展 [J]. 化学进展, 1999, 11(1): 41-48.
- [7] 孔建强, 王 伟, 朱 平, 等. 青蒿素生物合成研究 进展 [J]. 中国药学杂志, 2008, 43(11): 804-808.
- [8] 唐 煜,朱建华,于荣敏.黄花蒿悬浮培养细胞对二
   氢青蒿酸的生物转化研究 [J].中草药, 2010, 41(8):
   1358-1361.
- [9] 周 洁,张 案,郭兰萍,等.稀土元素镧对黄花蒿 光合作用及青蒿素积累的影响 [J].中草药,2010, 41(8):1371-1374.
- [10] Brown G D. Cadinanes from Artemisia annua that may be intermediates in the biosythesis of artemisinin [J].

Phytochemistry, 1994, 36(3): 637-641.

- [11] 魏 群编译. 基因克隆和 DNA 分析 (第5版) [M]. 北京: 高等教育出版社, 2007.
- [12] 刘万宏.紫杉醇前体合成途径两个关键酶基因克隆和分析 [D].重庆:西南大学,2008.
- [13] Thompson L D, Gibson T J, Plewniak F, et al. The clustal X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools [J]. Nucl Acids Res, 1997, 24: 4876-4882.
- [14] Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment [J]. *Brief Bioinform*, 2004, 5: 150-163.
- [15] Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees [J]. *Mol Biol Evol*, 1987, 4: 406-425.
- [16] McAteer S, Coulson A, McLennan N, et al. The lytB gene of Escherichia coli is essential and specifies a product needed for isoprenoid biosynthesis [J]. J Bacteriol, 2001, 183: 7403-7407.
- [17] Hsieh M H, Goodman H M. The arabidopsis IspH homolog is involved in the plastid nonmevalonate pathway of isoprenoid

biosynthesis [J]. Plant Physiol, 2005, 138: 641-653.

- [18] 刘秋云,李 蕾,蒋一帆. 粗糙脉孢霉瓜氨酸需求型 突变株的分离 [J]. 中山大学学报:自然科学版, 2000, 39(1): 118-120.
- [19] Garnier J, Gibrat J, FRobson B. GOR secondary structure prediction method version IV [J]. *Methods Enzymol*, 1996, 266: 540-553.
- [20] Querol J, Campos N, Imperial S, et al. Functional analysis of the Arabidopsis thaliana GCPE protein involved in plastid isoprenoid biosynthesis [J]. FEBS Lett, 2002, 514: 343-346.
- [21] Lichtenthaler H K, Rohmer M, Schwender J. Two independent biochemical pathways for isopentenyl diphosphate and isoprenoid biosynthesis in higher plants [J]. *Plant Physiol*, 1997, 101: 643-52.
- [22] 刘春明译. 植物代谢途径的生物工程和分子生物学[M]. 北京: 科学出版社, 2009.
- [23] Rohmer M, Knani M, Simonin P, et al. Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate [J]. Biochem J, 1993, 295: 517-524.