

杜仲转录组基因密码子使用偏好性分析

刘慧敏^{1,2}, 乌云塔娜^{1,2}, 杜红岩^{1,2}

(1. 中国林业科学研究院 经济林研究开发中心, 河南 郑州 450003; 2. 国家林业局 杜仲工程技术研究中心, 河南 郑州 450003)

摘要:以杜仲转录组数据为试材,采用 CodonW 和 SPSS 软件相结合的方法,研究了杜仲基因密码子组成和长度对基因表达量的影响,以期将来改造外源基因及利用基因工程技术改良杜仲提供理论基础。结果表明:杜仲转录组中基因的表达量与 G3s、C3s、G+C 和 GC3s 含量均呈极显著正相关($P < 0.01$),与 A3s 含量、T3s 含量、单条基因总密码子数(L_aa)和有效密码子数 ENC 均呈极显著负相关($P < 0.01$),与 Aro 含量呈显著负相关($P < 0.05$),表明高表达基因的密码子使用偏好性较强,且偏好于使用以 G/C 结尾的密码子;ENC 与 C3s、G3s、G+C 和 GC3 含量均呈极显著负相关($P < 0.01$),与 A3s、T3s、Aro 含量呈极显著正相关($P < 0.01$);确定了 15 个杜仲最优密码子,分别为 AAC、AAG、ACC、AGG、AUC、CAC、CAG、CCG、CUC、CUG、GCC、GGC、GUC、UAC 和 UCC。果实和叶片特异表达基因在氨基酸 Leu、Ile、Ser、Pro、Thr、Asn、Glu 和 Cys 密码子的使用上存在差异。

关键词:杜仲;转录组;密码子;偏好性

中图分类号:S 567.1⁺9 **文献标识码:**A **文章编号:**1001-0009(2016)13-0085-05

密码子是生命信息传递的基本单位,在长期进化过程中,不同物种的基因都会逐渐适应宿主的基因组环境,形成特定的符合宿主基因组的密码子用法,因此在某一物种或某一类基因中普遍存在同义密码子非均衡使用的现象,即密码子偏好性,其中被优先使用的同义密码子称为最优密码子。密码子使用偏好性在不同的生物体中存在很大的差异^[1]。密码子的使用偏好性不仅在基因表达水平方面发挥重要调节作用^[2],而且有利于提高翻译的准确性和效率^[3]。密码子的使用除了受选择和突变^[4-5]的影响外,还受到细胞内 tRNA 丰度^[6-7]、碱基组成^[8]、基因在染色体上的位置^[9]、基因长度及表达水平^[10-11]、氨基酸疏水性^[12]、mRNA 二级结构^[13]等影响。

杜仲(*Eucommia ulmoides*)是我国特有的名贵中药材,杜仲皮作为药用植物最早记载于《神农本草经》上。近 20 年来,随着对杜仲叶、雄花和果实研究的不断深入,研究者发现,杜仲皮、叶、雄花、果实中含有多种药用及保健成分,如木脂素、环烯醚萜类、黄酮类、绿原酸、松脂

醇二葡萄糖苷、 α -亚麻酸等^[14]。除了药用,杜仲的果实、皮、叶中含有丰富的杜仲胶,其物理性质特殊,开发潜力巨大^[15]。

该研究基于杜仲叶片和果实的转录组数据,解析杜仲表达基因的密码子组成,探讨了杜仲表达基因的密码子使用偏好性和影响因素及叶片、果实特异表达基因的编码特点,以期根据杜仲基因特点优化和改造外源基因密码子,从而有效抑制转基因沉默现象^[16-17],利用基因工程技术改良杜仲提供理论依据。

1 材料与方法

1.1 试验材料

以果用杜仲良种“华仲 6 号”叶片和果实的转录组数据为原始数据,转录组数据来自于中国林业科学研究院经济林研究开发中心高通量测序结果,鉴定出 4 271 条具有完整阅读框的序列,其中在果实中特异表达的序列有 36 条,在叶子中特异表达的序列有 4 条,这些序列是长度 ≥ 300 bp,均以 ATG 为起始密码子,以 TAA、TAG 或 TGA 为终止密码子,且序列中间无 N 或者终止密码子。

1.2 试验方法

利用 CodonW 1.4.2 对杜仲转录组 4 271 条具有完整阅读框的序列,果实中特异表达的 36 条序列和叶子中特异表达的 4 条序列,分别进行密码子组成和使用偏好性分析。密码子组成的度量指标包括 G+C 含量(鸟

第一作者简介:刘慧敏(1989-),女,河南辉县人,博士研究生,研究方向为林木遗传育种。E-mail:313324495@qq.com.

责任作者:乌云塔娜(1975-),女,内蒙古通辽人,博士,教授,现主要从事经济林育种与栽培等研究工作。E-mail:tanatana@163.com.

基金项目:国家“十二五”科技支撑计划资助项目(2012BAD21B0502)。

收稿日期:2016-02-15

嘌呤和胞嘧啶含量); A3s、T3s、G3s、C3s(同义密码子在第3位上腺嘌呤、胸腺嘧啶、鸟嘌呤和胞嘧啶的出现频率); GC3s(同义密码子第3位的G+C含量); Aro(芳香族氨基酸频率)等。密码子使用偏好性的度量指标包括RSCU(同义密码子相对使用度)、ENC(有效密码子数)等。

同义密码子相对使用度(relative synonymous codon usage, RSCU)是指对于某一特定的密码子,在编码对应氨基酸的同义密码子间的相对概率,去除了氨基酸组成对密码子使用的影响。该值的计算方法为某一密码子所使用的频率与其在无偏好性使用时预期频率之间的比值^[18-19],如果密码子使用无偏好性,则RSCU值为1;如果该密码子比其它密码子使用更频繁,则其RSCU值大于1;反之亦然^[19]。

有效密码子数(effective number of codons, ENC or Nc)是检测单个基因的密码子偏好程度,取值范围为20(每个氨基酸只使用1个密码子)到61(各个密码子被均衡使用)。其值越低,表明该基因的密码子使用偏好性越强^[20-21]。

1.3 数据分析

运用SPSS 20.0统计软件,绘制ENC与GC3s的关系图,并进行密码子组成和使用偏好性各参数A3s、T3s、G3s、C3s、ENC、GC3s、G+C含量、Aro和基因表达量RPKM间相关性分析,以及叶片和果实特异表达基因密码子组成和使用偏好性各参数的差异显著性分析。

2 结果与分析

2.1 转录组的密码子组成和使用性参数分析

杜仲转录组4 271个全长基因序列的长度变化范围为300~11 736 bp,平均数为1 275 bp,所包含基因的G+C含量的变化范围为0.37~0.64,平均为0.47(表1);同义密码子第3位上腺嘌呤A3s、胸腺嘧啶T3s、鸟嘌呤G3s、胞嘧啶C3s出现的频率变化范围分别为0.06~0.53、

表1 基于杜仲转录组全长基因密码子的组成及使用参数

Table 1 Composition and parameters of codon usage in genes identified from full-length cDNA transcriptome

密码子组成和使用参数	变化范围	平均数	标准差
T3s(同义密码子第3位上胸腺嘧啶出现的频率)	0.08~0.59	0.37	0.08
C3s(同义密码子第3位上胞嘧啶出现的频率)	0.11~0.76	0.28	0.08
A3s(同义密码子第3位上腺嘌呤出现的频率)	0.06~0.53	0.30	0.06
G3s(同义密码子第3位上鸟嘌呤出现的频率)	0.09~0.72	0.31	0.07
ENC(有效密码子数)	30.35~61.00	54.31	3.93
GC3s(同义密码子第3位的G+C含量)	0.26~0.86	0.46	0.10
G+C(鸟嘌呤和胞嘧啶含量)	0.37~0.64	0.47	0.04
Aro(芳香族氨基酸频率)	0.00~0.23	0.08	0.03
L_aa(氨基酸序列的长度/aa)	100~3 912	425.00	279.00

0.08~0.59、0.09~0.72、0.11~0.76,平均数分别为0.30、0.37、0.31、0.28(表1);同义密码子第3位G+C的含量变化范围为0.26~0.86,平均数为0.46;芳香族氨基酸(Aro)的频率范围为0.00~0.23(表1)。

2.2 密码子使用参数的相关性分析

以GC3s值为X轴,ENC值为Y轴,用SPSS 20.0绘制散点图(图1),其中连续曲线反映了无选择压力条件下二者之间的关系。然而,由图1可以看出,大多数基因位点的分布偏离期望曲线,表明除核苷酸组成偏好外,自然选择等其它因素对密码子的使用也有一定的影响。位于曲线下方的基因,由于其GC3s含量高,趋向于使用以G或C结尾的同义密码子,具有较强的密码子使用偏好性,因此其ENC值偏低;而曲线上方的基因,密码子偏好性较弱,更倾向于随机使用密码子。

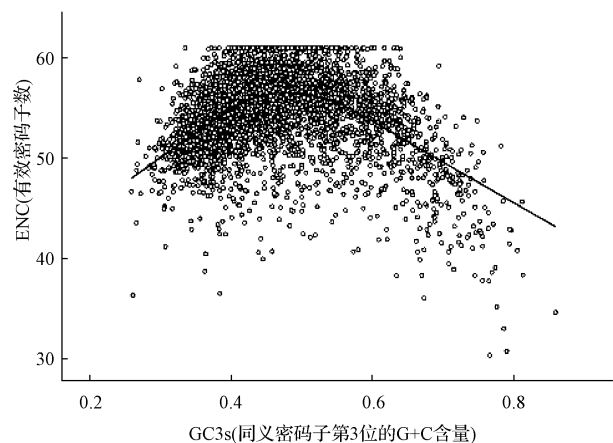


图1 杜仲密码子的ENC-plot曲线

Fig. 1 ENC-plot of *Eucommia* codons

相关性分析表明,基因的表达量(RPKM)与同义密码子第3位碱基含量(G3s和C3s)、G+C含量、GC3s含量均呈极显著正相关($P < 0.01$),与同义密码子第3位碱基含量(A3s)、基因编码氨基酸的长度(L_aa)和ENC均呈极显著负相关($P < 0.01$),与芳香族氨基酸含量(Aro)呈显著正相关($P < 0.05$)。有效密码子数(ENC)与同义密码子第3位碱基含量(C3s和G3s)、G+C含量、GC3s均呈极显著负相关($P < 0.01$),与同义密码子第3位碱基含量(A3s和T3s)和芳香族氨基酸(Aro)呈极显著正相关($P < 0.01$)(表2)。由各参数间的相关性关系可知,同义密码子第3位的碱基含量直接影响着基因的表达水平和密码子使用偏好性程度的大小。密码子使用偏好性越强(ENC值较小)的基因,越偏好于使用G+C含量高尤其是以G/C结尾(GC3s较大)的密码子,同时其表达量(RPKM值)也越高。

将ENC值按大小排序,选取上、下限区域5%各213条基因序列作为高表达基因样本和低表达基因样本,分别计算其各密码子的RSCU值。如二者之间相应的密

表 2 转录组全长基因密码子使用偏好性参数及表达量之间的相关系数

Table 2 Correlation coefficients among the parameters of codon usage and RPKM in full-length transcriptome

	T3s	C3s	A3s	G3s	ENC	GC3s	G+C	L_aa	RPKM
C3s	-0.84**								
A3s	0.56**	-0.67**							
G3s	-0.58**	0.35**	-0.61**						
ENC	0.05**	-0.06**	0.14**	-0.08**					
GC3s	-0.91**	0.88**	-0.82**	0.72**	-0.10**				
G+C	-0.78**	0.72**	-0.76**	0.50**	-0.12**	0.83**			
L_aa	0.31**	-0.32**	0.22**	-0.20**	0.01	-0.32**	-0.26**		
RPKM	-0.02	0.13**	-0.20**	0.09**	-0.05**	0.13**	0.11**	-0.12**	
Aro	-0.06**	0.18**	-0.09**	-0.08**	0.08**	0.07**	-0.10**	-0.05**	0.04*

注: ** 在 0.01 水平(双侧)上显著相关, * 在 0.05 水平(双侧)上显著相关。

Note: ** indicates significant correlation with $P<0.01$ and bilateral; * indicates significant correlation with $P<0.05$ and bilateral.

码子的 RSCU 值达到极显著差异($P<0.01$), 2 个基因样本之间的差值大于等于 0.08, 并且在高表达基因样本中该密码子的 RSCU 值大于 1, 则确定其为高表达最优密码子, 最终确定了 AAC、AAG、ACC、AGG、AUC、CAC、CAG、CCG、CUC、CUG、GCC、GGC、GUC、UAC 和 UCC 为最优密码子, 均以 G/C 结尾, 见图 2。

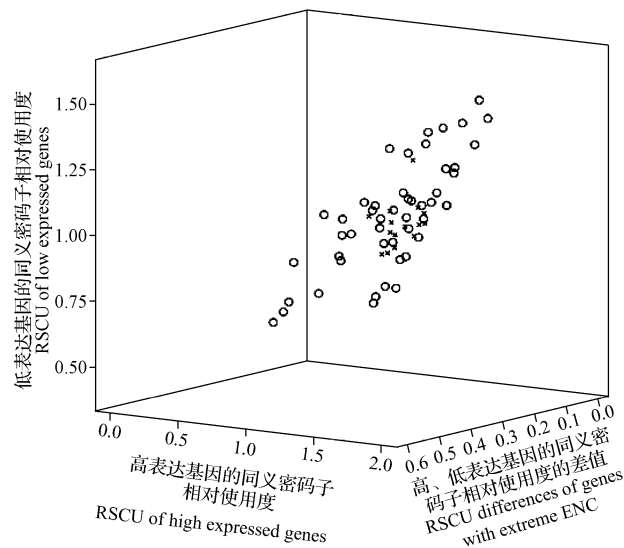


图 2 杜仲密码子分布图

Fig. 2 Distribution of *Eucommia* codons

2.3 组织特异性密码子分析

以 4 条叶片特异表达基因和 36 条果实特异表达基因为研究对象, 分别计算这 2 类组织特异表达基因中各密码子的组成和同义密码子相对使用度(RSCU 值), 并将其与整个转录组比较。结果表明, 叶片和果实特异表达基因与转录组基因在密码子的使用上分别有 14、21 个密码子具有差异; 果实特异表达基因和叶片特异表达基因在密码子的使用上有 14 个密码子具有差异, 二者对 Leu、Ile、Ser、Pro、Thr、Asn、Glu 和 Cys 的使用偏性存在差异; 对于终止密码子, 叶片偏好使用 UGA, 而果实偏好使用 UGA 和 UAA。

表 3 杜仲同义密码子在转录组和叶片、果实特异表达基因中的使用频率

Table 3 Frequency of synonymous codons in

Eucommia transcriptome, leaf-specific and fruit-specific genes

氨基酸	密码子	转录组		叶片特异表达		果实特异表达	
		数目	RSCU	数目	RSCU	数目	RSCU
Phe	UUU	39 074	1.05	34	0.99	170	0.92
	UUC	35 439	0.95	35	1.01	201	1.08
	UUA	19 350	0.68	8	0.44	72	0.58
	UUG	43 243	1.51	33	1.83	181	1.45
Leu	CUU	39 688	1.38	21	1.17	119	0.95
	CUC	28 353	0.99	22	1.22	171	1.37
	CUA	16 547	0.58	8	0.44	70	0.56
	CUG	24 787	0.86	16	0.89	137	1.10
Ile	AUU	43 060	1.36	22	1.74	169	1.08
	AUC	28 693	0.90	9	0.71	178	1.14
Met	AUA	23 470	1.00	7	0.55	123	0.79
	AUG	44 248	1.00	29	1.00	243	1.00
Val	GUU	46 810	1.56	20	1.48	167	1.12
	GUC	23 115	0.77	6	0.44	127	0.85
Ser	GUA	16 486	0.55	9	0.67	82	0.55
	GUG	33 402	1.12	19	1.41	220	1.48
	UCU	38 162	1.44	28	1.63	121	1.04
	UCC	24 295	0.91	12	0.70	128	1.10
Pro	UCA	31 373	1.18	18	1.05	125	1.08
	UCG	20 934	0.79	17	0.99	101	0.87
	CCU	30 525	1.33	16	0.96	124	1.03
	CCC	16 205	0.71	12	0.72	95	0.79
Thr	CCA	26 879	1.17	25	1.49	140	1.16
	CCG	17 977	0.79	14	0.84	124	1.03
	ACU	26 600	1.23	14	1.12	87	0.84
	ACC	21 087	0.98	14	1.12	133	1.28
Ala	ACA	25 267	1.17	12	0.96	106	1.02
	ACG	13 393	0.62	10	0.80	89	0.86
	GCU	46 727	1.46	20	1.14	166	1.01
	GCC	28 499	0.89	15	0.86	194	1.18
Tyr	GCA	35 455	1.11	18	1.03	166	1.01
	GCG	17 534	0.55	17	0.97	132	0.80
	UAU	27 611	1.10	5	0.42	101	0.83
	UAC	22 622	0.90	19	1.58	143	1.17
TER	UAA	1 275	0.90	1	0.75	14	1.17
	UAG	1 009	0.71	1	0.75	6	0.50
His	CAU	24 986	1.19	7	0.82	86	0.88
	CAC	17 038	0.81	10	1.18	109	1.12

表 3(续)

Talbe 3(Continued)

氨基酸	密码子	转录组		叶片特异表达		果实特异表达	
		数目	RSCU	数目	RSCU	数目	RSCU
Gln	CAA	33 864	1.03	26	1.18	185	1.21
	CAG	32 090	0.97	18	0.82	120	0.79
Asn	AAU	45 648	1.19	26	1.16	191	0.98
	AAC	31 177	0.81	19	0.84	199	1.02
Lys	AAA	51 731	0.94	23	0.96	208	0.92
	AAG	58 367	1.06	25	1.04	243	1.08
Asp	GAU	66 407	1.35	45	1.43	205	1.23
	GAC	31 759	0.65	18	0.57	127	0.77
Glu	GAA	63 047	1.04	38	1.10	242	0.97
	GAG	57 749	0.96	31	0.90	256	1.03
Cys	UGU	16 066	1.04	9	1.06	84	0.97
	UGC	14 920	0.96	8	0.94	89	1.03
TER	UGA	1 987	1.40	2	1.50	16	1.33
Trp	UGG	22 308	1.00	13	1.00	127	1.00
	CGU	13 263	0.77	3	0.27	34	0.46
Arg	CGC	9 029	0.53	4	0.36	45	0.61
	CGA	12 660	0.74	8	0.72	33	0.45
	CGG	12 770	0.74	7	0.63	51	0.69
	AGU	23 384	0.88	12	0.70	94	0.81
Ser	AGC	21 301	0.80	16	0.93	127	1.09
	AGA	28 538	1.66	18	1.61	122	1.66
Arg	AGG	26 622	1.55	27	2.42	156	2.12
	GGU	34 760	1.14	14	0.75	156	0.88
Gly	GGC	24 001	0.78	13	0.69	166	0.94
	GGA	37 677	1.23	27	1.44	208	1.17
	GGG	25 894	0.85	21	1.12	180	1.01

注:斜体加重数值表明对于某一特定的密码子,该密码子比其它密码子使用更频繁(RSCU>1)。

Note: The numbers with bold and italic indicate to a specific codon, this codon was used more frequently(RSCU>1).

3 结论与讨论

该研究主要以杜仲叶片和果实转录组数据为研究对象,鉴别出 4 271 条具有完整阅读框的基因序列,其中在果实中特异表达的基因有 36 条,在叶子中特异表达的基因有 4 条。利用 CodonW 和 SPSS 软件对基因的密码子组成和使用性参数进行分析,杜仲偏好使用以 G/C 结尾的同义密码子,其基因的表达量与 G3s、C3s、G+C、GC3s 含量均呈极显著正相关,并确定了 AAC、AAG、ACC、AGG、AUC、CAC、CAG、CCG、CUC、CUG、GCC、GGC、GUC、UAC 和 UCC 为最优密码子。果实和叶片的组织特异性密码子分析发现,二者在个别密码子的使用上存在差异,叶片偏好使用 UGA 终止密码子,而果实偏好使用 UGA 和 UAA 终止密码子。

在鉴别基因的过程中发现,具有全长阅读框的基因占组装出来的基因总数的叶片和果实特异表达的基因比较少,可能是因为二代测序片段比较短,导致组装出来的大量基因片段较短或者存在 N,如果能结合第三代测序技术,将会对转录组的数据质量带来很大的改善^[22-23]。

杜仲是一种古老的植物,栽培历史悠久,在进化上除了自然选择和基因突变外,人工选择在其进化上也有

影响,该研究首次利用杜仲转录组数据初步研究了杜仲密码子的使用偏性,对基因工程中外源基因的改良及其表达研究有重要的参考价值。

参考文献

- [1] ZHANG G Y, GE H H, FANG B S. Analysis of synonymous codon usage in two hyperthermophiles[J]. Journal of Huaqiao University, 2006, 22(2): 87-90.
- [2] CARLINI D B, STEPHAN W. *In vivo* introduction of unpreferred synonymous codons into the *Drosophila Adh* gene results in reduced levels of ADH protein[J]. Genetics, 2003, 163: 239-243.
- [3] SHARP P M, MATASSI G. Codon usage and genome evolution[J]. Curr Opin Genet Dev, 1994(4): 851-860.
- [4] LIU Q P, FENG Y, ZHAO X A, et al. Synonymous codon usage bias in *Oryza sativa*[J]. Plant Science, 2004, 67: 101-105.
- [5] ROMERO H, MUSTO H, ZAVALA A. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces[J]. Nucleic Acids Res, 2000, 28(10): 2084-2090.
- [6] STENICO M, LLOYD A T, SHARP P M. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases[J]. Nucl Acid Res, 1994, 22: 2437-2446.
- [7] OLEJNICZAK M, UHLENBECK O C. tRNA residues that have co-evolved with their anticodon to ensure uniform and accurate codon recognition[J]. Biochimie, 2006, 88: 943-950.
- [8] 张文娟. 基于密码子水平的生物信息学分析及进化研究[D]. 上海: 复旦大学, 2006.
- [9] ROMERO H, ZAVALA A, MUSTO H, et al. The influence of translational selection on codon usage in fishes from the family Cyprinidae[J]. Gebe, 2003, 317: 141-147.
- [10] HIRAOKA Y, KAWAMATA K, HARAGUCHI T, et al. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*[J]. Genes Cells, 2009, 14(4): 499-509.
- [11] 孙铮, 马亮, MURPHY R, 等. Wolbachia *pepientis* wME1 基因组水平上的密码子使用分析[J]. 中国科学 C 辑(生命科学), 2009, 39(10): 948-953.
- [12] KNIGHT R D, FREELAND S J, LANDWEBER L F. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes[J]. Genome Biol, 2001(2): RESERCH0010.
- [13] GU W J, ZHOU T, MA J M, et al. Folding type specific secondary structure propensities of synonymous codons[J]. IEEE Trans Nanobiosci, 2003, 2(3): 150-157.
- [14] 杜红岩. 杜仲活性成分与药理研究的新进展[J]. 经济林研究, 2003, 21(2): 58-61.
- [15] 杜红岩. 我国的杜仲胶资源及其开发潜力与产业发展思路[J]. 经济林研究, 2009, 28(3): 1-5.
- [16] 曹慧颖, 张锐, 郭三堆. 串联的人胸腺素 $\alpha 1$ 基因在番茄中的高效表达[J]. 中国农业科学, 2009, 42(7): 2291-2296.
- [17] 邹永梅, 施季森, 诸葛强, 等. 遗传转化植物中沉默基因的消除[J]. 分子植物育种, 2004, 4(1): 95-102.
- [18] SHARP P M, HANEY T M F, MOSURSKI K R. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes[J]. Nucl Acids Res, 1986(14): 5125-5143.
- [19] SAU K, GUPTA S K, SAU S, et al. Analysis of factors shaping codon usage in the mitochondrion genome of *Oryza sativa*[J]. Mitochondrion, 2004(4): 313-320.
- [20] WRIGHT F. The effective number of codons used in a gene[J]. Gene, 1990, 87: 23-29.

三种食用百合种质资源限制生长保存及遗传稳定性研究

万珠珠, 牛来春, 谭秀梅, 刘 敏, 吴 亮, 董 草

(云南师范大学 文理学院, 云南 昆明 650222)

摘 要:以3种食用百合试管苗为试材,采用限制生长保存法,研究了不同浓度生长抑制剂对食用百合试管苗保存效果的影响,并对保存后的试管苗进行遗传稳定性检测,以期建立食用百合种质资源离体保存体系。结果表明:龙芽百合试管苗用 $10 \text{ mg} \cdot \text{L}^{-1}$ 防落素(PCPA)处理保存300 d后,试管苗生长缓慢,存活率达96.00%;川百合试管苗用 $2.0 \text{ mg} \cdot \text{L}^{-1}$ 青鲜素(MH)处理保存150 d后,试管苗的抑制生长效果明显,结鳞率和存活率高达100.00%;兰州百合试管苗用 $10 \text{ mg} \cdot \text{L}^{-1}$ 多效唑(PP₃₃₃)处理保存300 d后,试管苗抑制作用明显,结鳞率达100.00%,存活率达94.40%。比较限制生长保存后与未保存植株的可溶性蛋白和酯酶同工酶图谱,各处理和对照图谱带相似,初步证明了以上保存方法的可行性,较好的保持了遗传稳定性。

关键词:食用百合;限制生长保存;遗传稳定性

中图分类号:S 644.102.4 **文献标识码:**B **文章编号:**1001-0009(2016)13-0089-04

食用百合是集食用、药用、观赏兼用的多年生草本植物,近年来市场需求日益增加。作为无性繁殖蔬菜,

传统的种质保存方法需每年进行种质更新,不仅耗费大量的人力、物力和财力,还容易受到病虫害和自然灾害的影响^[1]。为了克服传统保存方法所面临的问题,现以龙芽百合(*Lilium browerzii* F. E. Brown var. *viridulum* Baker)、川百合(*Lilium davidii* Duchartre)和兰州百合(*Lilium davidii* Duch. var. *unicolor*(Hoog)Cotton)3种食

第一作者简介:万珠珠(1982-),女,硕士研究生,讲师,现主要从事观赏植物无性繁殖保存等研究工作。E-mail:42998072@qq.com.
收稿日期:2016-02-14

[21] GUPTA S K, BHATTACHARYYA T K, GHOSH T C. Synonymous codon usage in *Lactococcus lactis*; mutational bias versus translational selection[J]. *Biomol Struct Dyn*, 2004(21):1-9.
[22] DONALD S, HAGEN T, FABIAN G, et al. A single-molecule long-read survey of the human transcriptome[J]. *Nature Biotechnology*, 2013(31):

1009-1014.

[23] HAGEN T, FABIAN G, DONALD S, et al. Defining a personal, allele-specific, and single molecule long-read transcriptome[J]. *PNAS*, 2014, 111: 9869-9874.

Analysis of Characteristic of Codon Usage of *Eucommia ulmoides* Transcriptome

LIU Huimin^{1,2}, WUYUN Tana^{1,2}, DU Hongyan^{1,2}

(1. Non-timber Research and Development Center, Chinese Academy of Forestry Science, Zhengzhou, Henan 450003; 2. The *Eucommia* Engineering Research Center of State Forestry Administration, Zhengzhou, Henan 450003)

Abstract: *Eucommia ulmoides* was used as test materials, the correlation between codon composition and gene expression was analyzed based on the transcriptome data by using CodonW and SPSS software, in order to offer a theoretical foundation for transforming exogenous gene and *Eucommia* genetic engineering. The results showed that the gene expression had a significant and positive correlation to G3s, C3s, G+C and GC3s ($P < 0.01$) content and a negative correlation to A3s content, T3s content, gene coding sequence length and effective codon number (ENC) ($P < 0.01$) and Aro content ($P < 0.05$). These results indicated that genes with a high expression had a stronger codon usage bias and likely end with G/C. ENC had a significant positive correlation to A3s, T3s and Aro content ($P < 0.01$) and a significant negative correlation to C3s, G3s, G+C, and GC3s content ($P < 0.01$). Fifteen optimal codons were identified, namely, AAC, AAG, ACC, AGG, AUC, CAC, CAG, CCG, CUC, CUG, GCC, GGC, GUC, UAC, and UCC. Fruit and leaf special genes had difference on the codon usage of Leu, Ile, Ser, Pro, Thr, Asn, Glu and Cys.

Keywords: *Eucommia ulmoides*; transcriptome; codon; bias