

胡振民, 万青, 李欢, 等. 茶树 *CsNRT1.1* 基因密码子使用特性分析[J]. 江苏农业学报, 2019, 35(4): 896-903.

doi: 10.3969/j.issn.1000-4440.2019.04.021

## 茶树 *CsNRT1.1* 基因密码子使用特性分析

胡振民<sup>1</sup>, 万青<sup>1</sup>, 李欢<sup>1</sup>, 李荣林<sup>1</sup>, 王枫<sup>2</sup>, 杨亦扬<sup>1</sup>

(1.江苏省农业科学院休闲农业研究所, 江苏省高效园艺作物遗传改良重点实验室, 江苏 南京 210014; 2.南京农业大学园艺学院, 作物遗传与种质创新国家重点实验室, 江苏 南京 210095)

**摘要:** 运用 CUSP、Condow 等程序对茶树 *CsNRT1.1* 基因、茶树基因组及不同物种 *NRT1.1* 基因进行密码子使用特性分析, 并比较了 *CsNRT1.1* 基因与不同模式物种密码子使用偏好性差异。结果显示, 茶树 *CsNRT1.1* 基因编码区偏好使用 A/U, 密码子结尾偏好使用 G/C, *RSCU* 值>1 的偏好性密码子有 26 个, 其中偏好性较强的只有 AGG (*RSCU*>2), 并偏好以 G/C 结尾; *CsNRT1.1* 基因和茶树基因组中密码子偏好性存在差异, 共计 25 个密码子使用偏好性存在明显差异, 5 个氨基酸密码子偏好性完全一致; *CsNRT1.1* 等少数 *NRT1.1* 基因的 *ENc* 值较高, *CAI*、*CBI* 值较低, 说明其密码子使用偏好性较低, 同时基因表达水平可能较低; *CsNRT1.1* 基因与双子叶植物拟南芥、烟草及酵母菌密码子偏好性差异较小; 对 *NRT1.1* 基因基于编码基因序列 (CDS) 组成和基于密码子偏性的聚类分析结果存在一定差异, 但在一定程度上都能反映物种间的亲缘关系。

**关键词:** 茶树; 密码子; 偏好性; 硝态氮; *CsNRT1.1* 基因

**中图分类号:** S571.101 **文献标识码:** A **文章编号:** 1000-4440(2019)04-0896-08

## Analysis of codon usage features of *CsNRT1.1* gene in *Camellia sinensis*

HU Zhen-min<sup>1</sup>, WAN Qing<sup>1</sup>, LI Huan<sup>1</sup>, LI Rong-lin<sup>1</sup>, WANG Feng<sup>2</sup>, YANG Yi-yang<sup>1</sup>

(1. Institute of Leisure Agriculture, Jiangsu Academy of Agricultural Sciences, Jiangsu Key Laboratory for Horticultural Crop Genetic Improvement, Nanjing 210014, China; 2. College of Horticulture, Nanjing Agricultural University, State Key Laboratory of Crop Genetics and Germplasm Enhancement, Nanjing 210095, China)

**Abstract:** The codon usage characteristics of *CsNRT1.1* gene, tea plant (*Camellia sinensis*) genome and *NRT1.1* gene of different species were analyzed by CUSP and Condow programs, and the codon preference differences between *CsNRT1.1* gene and different model species were also compared. The results showed that the *CsNRT1.1* gene coding region and codon end of tea plant preferred A/U and G/C respectively, there were 26 preferred codons with *RSCU* value>1, among which AGG (*RSCU*>2) was strongly preferred. There were differences in codon preference between *CsNRT1.1* gene and tea plant genome, and a total of 25 codon usage preferences were obviously different, five amino acids had the same codon

收稿日期: 2018-11-30

**基金项目:** 国家自然科学基金项目 (31400587、31600558、31800590); 江苏省自然科学基金项目 (BK20160590); 江苏省农业科技自主创新基金项目 [CX(17)2018]

**作者简介:** 胡振民 (1986-), 男, 山东临沂人, 博士, 助理研究员, 主要从事茶树栽培与茶园土壤养分管理研究。(Tel) 025-84391693; (E-mail) hzm861@sina.com

**通讯作者:** 杨亦扬, (Tel) 025-84391693; (E-mail) yangyiyang\_yyy@126.com

preference. A few *NRT1.1* genes including *CsNRT1.1* had higher *ENc*, lower *CAI* and *CBI* values, indicating lower codon usage preference and lower gene expression level. The difference of codon preference between *CsNRT1.1* gene with dicotyledonous *Arabidopsis thaliana*, tobacco and yeast was relatively small. Cluster analysis results of *NRT1.1* gene based on CDS composition and codon preference showed that there were some differences, but these results could reflect the relationship between species to a

certain extent.

**Key words:** *Camellia sinensis*; codon; preference; nitrate nitrogen; *CsNRT1.1* gene

密码子在生物体内遗传信息传递过程中不可或缺。每种氨基酸有 1~6 个密码子,编码同一种氨基酸的密码子称为同义密码子(Synonymous condon)。生物体在同义密码子的使用上,会优先使用其中的某些密码子,即存在密码子偏性(Codon bias)<sup>[1]</sup>。这些使用次数最多的同义密码子,被称为最优密码子(Optimal codon)。不同物种或同一物种的基因家族间在密码子使用模式上存在不同,进化过程中承受的不同选择压力和(或)选择方式可能导致了这些差异<sup>[2-4]</sup>。密码子偏性分析的应用较为广泛,不仅可以用来描述基因进化规律、基因表达过程中的调控机制<sup>[3,5]</sup>,也可用于预测外源基因的最适宿主,通过密码子优化提高宿主中外源基因的表达量以及评估基因的表达水平等<sup>[6-8]</sup>。

硝态氮转运蛋白(Nitrate transporters, NRTs)参与植物根系对  $\text{NO}_3^-$  的吸收和转运过程,由于在植物氮代谢中起到重要作用,其基因功能研究正受到越来越多的关注<sup>[9-12]</sup>。*CsNRT1.1* 是从茶树中克隆到的硝态氮转运蛋白基因,编码 595 个氨基酸,叶片中 *CsNRT1.1* 表达受高浓度  $\text{NO}_3^-$  诱导<sup>[11]</sup>。

茶树[*Camellia sinensis* (L.)]是中国重要的经济作物之一。氮肥是茶园中的主要施用肥料,占茶园施肥总量的一半以上,是茶树生长和茶叶品质的主要限制因子之一<sup>[13]</sup>。尽管茶树对铵态氮( $\text{NH}_4^+$ )吸收量更大,且具有偏性吸收特性<sup>[14-15]</sup>,但茶园中施用的铵态氮肥( $\text{NH}_4^+$ )在土壤硝化细菌的作用下迅速转化为硝态氮( $\text{NO}_3^-$ )<sup>[16]</sup>,实际上茶园土壤中  $\text{NO}_3^-$  占较大比例,故 *CsNRT1.1* 可能对于茶树根系吸收氮素( $\text{NO}_3^-$ )起到较大的贡献,有必要针对 *CsNRT1.1* 基因开展研究。

近年来,大量植物的密码子偏性研究相继被报道,例如柑橘<sup>[17]</sup>、玉米<sup>[18]</sup>、苜蓿<sup>[19]</sup>、橄榄<sup>[20]</sup>、拟南芥<sup>[2]</sup>及兰科植物<sup>[21]</sup>等。目前茶树密码子偏性研究报道尚不多。赵洋等分析了茶树基因组及肌动蛋白基因密码子偏性<sup>[22-23]</sup>,郭秀丽等分析了茶树低温应答转录因子 *CBF1* 基因密码子使用特性<sup>[24]</sup>,时慧等分析了茶树抗寒转录因子 *ICE* 基因密码子偏性<sup>[25]</sup>。这些报道阐述了茶树不同基因的密码子使用特点。

本研究分析 *CsNRT1.1* 基因的密码子偏好性,并

比较其与不同物种基因组、不同物种 *NRT1.1* 基因的密码子偏好性差异,为 *CsNRT1.1* 基因外源表达宿主的选择及高效表达,以及 *CsNRT1.1* 基因功能验证提供基础。

## 1 材料与方法

### 1.1 目的基因获取及筛选

*CsNRT1.1* 基因为江苏省农业科学院休闲农业研究所茶研究室前期克隆所得,cDNA 全长为 1 880 bp,开放阅读框(ORF)1 788 bp,编码 595 个氨基酸。在 GenBank 中筛选 95 个茶树基因编码序列(CDS),作为茶树基因组密码子偏好性分析的样本,同时在 GenBank 中筛选 25 个其他植物物种的 *NRT1.1* 基因 CDS 序列。CDS 筛选按照 Sharp 等的方法<sup>[26]</sup>。

### 1.2 密码子偏好性分析

同义密码子相对使用度(Relative synonymous codon usage, RSCU)是指某一密码子使用的次数与其无偏性使用次数的比值,能直接反映出密码子使用的偏好性程度;有效密码子数(Effective number of codons, ENc)是基因的密码子使用频率与同义密码子平均使用频率偏差的量化值,此值反映同义密码子非均衡使用的偏好程度;GC 表示基因中所有密码子 G 和 C 的总含量,T3s、C3s、A3s、G3s 和 GC3s 分别表示密码子第 3 位是 T、C、A、G 和 G/C 的含量,GC 反映方向性突变压力的强弱,GC3s 则和密码子偏性有密切关系;密码子适应性指数(Codon adaptation index, CAI)用来衡量同义密码子与密码子最佳使用相符合的程度;密码子偏性指数(Codon bias index, CBI)也反映密码子偏性,CBI 值范围 0~1,CBI 为 1 时表示密码子偏性最大,为 0 时表示密码子无偏性。本研究应用 CondonW(version 1.4.2)程序对每条 CDS 序列进行偏好性参数(包括 RSCU、ENc、GC、GC3s、CAI、CBI 等)计算,并分别比较 *CsNRT1.1* 基因与茶树基因组、其他物种中 *NRT1.1* 基因密码子偏好性<sup>[5]</sup>。

密码子使用频率(Frequency)表示密码子在编码基因总密码子中出现的频率。密码子使用频率的比值可用来衡量物种间密码子偏好性差异,比值 0.5~2.0 表示密码子偏好性较为接近,比值  $\geq 2.0$

或 $\leq 0.5$ 则表示密码子偏好性差异较大。拟南芥(*Arabidopsis thaliana*)、烟草(*Nicotiana tabacum*)、小麦(*Triticum aestivum*)、玉米(*Zea mays*)、酵母菌(*Saccharomyces cerevisiae*)和大肠杆菌(*Escherichia coli*)基因组的密码子使用频率数据从 Codon Usage Database (<http://www.kazusa.or.jp/codon/>)中检索得到。使用 EMBOSS (<http://emboss.toulouse.inra.fr/>)在线程序中的 CUSP 程序计算 *CsNRT1.1* 基因的密码子使用频率,并分别计算 *CsNRT1.1* 与各物种密码子使用频率比值。

### 1.3 基于密码子使用偏性和 CDS 的聚类分析

对包括茶树在内的 26 个不同物种 *NRT1.1* 基因进行聚类分析,排除编码色氨酸(Trp)、甲硫氨酸(Met)密码子及 3 个终止密码子,以 *RSCU* 值作为变量,基于欧式平方距离采用 SPSS18.0 离差平方和(Ward 法)进行密码子偏性聚类分析。基于 *NRT1.1* 基因 CDS 的聚类分析采用 MEGA5.1 分析软件,采用 Neighbour-joining 法进行聚类分析。

## 2 结果与分析

### 2.1 *CsNRT1.1* 基因和茶树基因组有效密码子数(*ENc*)和 GC 含量

茶树 *CsNRT1.1* 基因的 *ENc*、GC 和 *GC3s* 值分别为 55.89、0.479 和 0.527(表 1)。密码子 *ENc* 值越低,偏好性越强。*CsNRT1.1* 基因 *ENc* 值偏大,说明 *CsNRT1.1* 密码子没有明显的偏好性;同时 GC 值为 0.479,*GC3s* 值为 0.527,说明其编码区偏好使用 A/U,密码子结尾偏好使用 G/C。

表 1 *CsNRT1.1* 基因和茶树基因组的 GC、GC3s 和 *ENc* 值

Table 1 The GC, GC3s and *ENc* values of *CsNRT1.1* gene and genome of *Camellia sinensis*

基因	CAI	CBI	<i>ENc</i>	<i>GC3s</i>	GC
<i>CsNRT1.1</i>	0.193	-0.001	55.89	0.527	0.479
茶树基因组	0.212	-0.029	56.12	0.459	0.469

CAI:密码子适应性指数;CBI:密码子偏性指数;*ENc*:有效密码子数;*GC3s*:密码子第 3 位是 G/C 的含量;GC:基因组中所有密码子 G 和 C 的总含量。

茶树基因组 *ENc*、GC 和 *GC3s* 值分别为 56.12、0.469 和 0.459(表 1),说明茶树基因组密码子使用没有明显偏好性,编码区密码子、末尾密码子较 *CsNRT1.1* 基因都更偏好使用 A/U。

### 2.2 *CsNRT1.1* 基因和茶树基因组密码子偏好性

对 *CsNRT1.1* 基因的密码子使用次数和 *RSCU*

值进行分析,结果表明,*RSCU* 值 $>1$ 的密码子有 26 个(即为 *CsNRT1.1* 偏好性密码子),其中以 A/U 结尾的有 11 个,G/C 结尾的有 15 个,表现出密码子对 G/C 结尾的偏好(这与 *CsNRT1.1* 编码区密码子以 A/U 结尾的偏好不同),其中偏好性较强的只有 AGG(*RSCU* 值 $>2$ ,精氨酸编码密码子)(表 2)。*CsNRT1.1* 中 AAA、AAG、CCU、AUG 和 UGG 的 *RSCU* 值为 1,说明其使用没有偏好性(表 2)。

表 2 *CsNRT1.1* 基因和茶树基因组密码子数量和同义密码子相对使用度

Table 2 Codon number and relative synonymous codon usage in *CsNRT1.1* gene and genome of *Camellia sinensis*

氨基酸	密码子	<i>CsNRT1.1</i> 基因		茶树基因组	
		数量	相对使用度( <i>RSCU</i> )	数量	相对使用度( <i>RSCU</i> )
A(Ala)	GCA *	10	0.89	511	1.21
	GCC *	16	1.42	371	0.88
	GCG	5	0.44	180	0.43
	GCU	14	1.24	626	1.48
C(Cys)	UGC	5	1.25	202	1.01
	UGU	3	0.75	197	0.99
D(Asp)	GAC *	12	1.14	435	0.75
	GAU *	9	0.86	727	1.25
E(Glu)	GAA	9	0.95	681	0.89
	GAG	10	1.05	853	1.11
F(Phe)	UUC *	24	1.33	475	0.98
	UUU *	12	0.67	490	1.02
G(Gly)	GGA *	6	0.57	477	1.18
	GGC *	14	1.33	318	0.79
	GGG *	13	1.24	333	0.83
	GGU *	9	0.86	483	1.20
H(His)	CAC	4	0.80	251	0.93
	CAU *	6	1.20	290	1.07
I(Ile)	AUA	8	0.67	233	0.58
	AUC *	14	1.17	397	0.98
	AUU	14	1.17	585	1.44
K(Lys)	AAA	14	1.00	635	0.78
	AAG *	14	1.00	986	1.22
L(Leu)	CUA	10	0.82	195	0.55
	CUC	14	1.15	406	1.15
	CUG	5	0.41	294	0.83

续表 2 Continued

氨基酸	密码子	<i>CsNRT1.1</i> 基因		茶树基因组	
		数量	相对使用度 ( <i>RSCU</i> )	数量	相对使用度 ( <i>RSCU</i> )
M (Met)	CUU *	10	0.82	482	1.37
	UUA	10	0.82	202	0.57
	UUG	24	1.97	536	1.52
	AUG	17	1.00	539	1.00
N (Asn)	AAC *	12	1.26	416	0.91
	AAU *	7	0.74	498	1.09
P (Pro)	CCA	9	1.50	410	1.39
	CCC	5	0.83	218	0.74
	CCG	4	0.67	182	0.62
	CCU *	6	1.00	371	1.26
Q (Gln)	CAA	18	1.57	391	1.05
	CAG	5	0.43	357	0.95
R (Arg)	AGA *	3	0.86	278	1.45
	AGG	10	2.86	331	1.72
	CGA	1	0.29	143	0.74
	CGC	2	0.57	119	0.62
	CGG	3	0.86	111	0.58
S (Ser)	CGU	2	0.57	170	0.89
	AGC *	9	1.29	237	0.84
	AGU *	8	1.14	233	0.82
	UCA	10	1.43	342	1.21
	UCC *	4	0.57	276	0.98
	UCG	5	0.71	158	0.56
T (Thr)	UCU *	6	0.86	450	1.59
	ACA *	12	0.91	352	1.17
	ACC	21	1.58	333	1.11
	ACG	6	0.45	113	0.38
V (Val)	ACU	14	1.06	403	1.34
	GUA	4	0.34	202	0.48
	GUC *	22	1.87	335	0.80
	GUG *	7	0.60	557	1.32
W (Trp)	GUU	14	1.19	589	1.40
	UGG	14	1.00	243	1.00
Y (Tyr)	UAC *	8	0.94	336	1.07
	UAU *	9	1.06	293	0.93
TER	UAA *	1	3.00	28	0.88
	UAG	0	0	28	0.88
	UGA *	0	0	39	1.23

密码子使用偏性差异明显的用 \* 标记。TER 表示终止密码子。

在 GenBank 中筛选到 95 个茶树基因, 共计 22 932 个密码子, 其中 *RSCU* 值 >1 的密码子有 29 个 (即为茶树基因组偏好性密码子), 其中以 A/U 结尾的有 20 个, G/C 结尾的有 9 个 (表 2), 说明茶树基因组整体偏好使用以 A/U 结尾的密码子。茶树基因组中 AUG、UGG 的 *RSCU* 值为 1, 密码子使用没有偏好性 (表 2)。

*CsNRT1.1* 基因和茶树基因组中共计 25 个密码子 (GCA、GCC、GAC、GAU、UUC、UUU、GGA、GGC、GGG、GGU、AUC、AAG、CUU、AAC、AAU、CCU、AGA、AGC、AGU、UCU、ACA、UAC、UAU、UAA、UGA) 使用偏性存在明显差异。密码子偏好性一致的氨基酸有 5 种 (Cys、Glu、His、Gln 和 Val); 密码子偏好性完全不同的氨基酸有 8 种 (Asp、Phe、Gly、Lys、Asn、Pro、Tyr 和 TER), 密码子偏性完全不同的 8 种氨基酸中有 5 种 (Asp、Phe、Lys、Asn、Tyr) 有 2 个同义密码子, 且大部分密码子以 U/C 结尾 (表 2)。

### 2.3 *CsNRT1.1* 基因与其他植物 *NRT1.1* 基因密码子偏好性比较

*CsNRT1.1* 基因 *CAI*、*CBI* 和 *ENc* 值分别为 0.193、-0.001 和 55.89 (表 3), 说明 *CsNRT1.1* 密码子偏好性较低。对 25 种其他植物物种中 *NRT1.1* 基因进行密码子分析 (表 3), 发现 25 个物种中 *NRT1.1* 基因的 *ENc* 值介于 49.50 和 60.97 之间, 平均值为 54.62, 说明各物种中 *NRT1.1* 基因均一性较好, 密码子偏好性较低; *CAI* 值介于 0.166 和 0.242 之间, 平均值为 0.192, *CAI* 值用来评价同义密码子与密码子最佳使用相符合的程度, *CBI* 值介于 -0.103 和 0.108 之间, 均值为 -0.023, 这说明 *NRT1.1* 基因密码子偏好性较低; *GC* 值介于 0.431 和 0.563 之间, 均值为 0.466, *GC3s* 值介于 0.391 和 0.675 之间, 均值为 0.475, 说明大部分 *NRT1.1* 基因密码子及密码子结尾都偏好是用 A/U, 这与 *CsNRT1.1* 中偏性相一致。

### 2.4 *CsNRT1.1* 基因与其他物种基因组密码子偏好性的比较

比较 *CsNRT1.1* 基因与拟南芥、烟草、小麦、玉米、大肠杆菌和酵母菌密码子使用频率, 结果显示 *CsNRT1.1* 基因与各物种密码子使用频率比值  $\geq 2.0$  或者  $\leq 0.5$  的数量分别为 13、14、18、16、15 和 20 种 (表 4), 说明 *CsNRT1.1* 基因与这些物种间存在不同程度的密码子偏好性差异。其中 *CsNRT1.1* 基因与



表 3 *CsNRT1.1* 基因和其他物种 *NRT1.1* 基因密码子偏好性分析Table 3 Codon preference analysis of *CsNRT1.1* gene and *NRT1.1* genes in other plant species

物种	序列登录号	<i>T3s</i>	<i>C3s</i>	<i>A3s</i>	<i>G3s</i>	<i>CAI</i>	<i>CBI</i>	<i>ENc</i>	<i>GC3s</i>	<i>GC</i>
大豆( <i>Glycine max</i> )	NM_001251041.2	0.333	0.299	0.365	0.224	0.180	-0.021	52.49	0.420	0.432
高粱( <i>Sorghum bicolor</i> )	XM_021463813.1	0.285	0.338	0.298	0.283	0.189	0.015	55.91	0.508	0.493
胡萝卜( <i>Daucuscarota</i> )	XM_017379391.1	0.292	0.342	0.284	0.312	0.213	0.057	60.97	0.520	0.474
大麦( <i>Hordeum vulgare</i> )	AB251946.1	0.368	0.256	0.337	0.251	0.191	-0.053	52.15	0.408	0.448
蔓花生( <i>Arachisdurenensis</i> )	XM_016333201.2	0.230	0.365	0.317	0.306	0.191	0.028	52.99	0.544	0.483
黄瓜( <i>Cucumissativus</i> )	XM_004147569.2	0.287	0.291	0.341	0.300	0.180	-0.070	56.09	0.474	0.461
棉花( <i>Gossypiumhirsutum</i> )	XM_016850186.1	0.327	0.270	0.377	0.237	0.172	-0.076	49.75	0.411	0.433
南瓜( <i>Cucurbita moschata</i> )	XM_023101987.1	0.215	0.357	0.285	0.362	0.189	-0.060	53.60	0.581	0.500
雷蒙德氏棉( <i>Gossypiumraimondii</i> )	XM_023067507.1	0.333	0.343	0.292	0.240	0.205	-0.007	59.41	0.475	0.470
拟南芥( <i>Arabidopsis thaliana</i> )	NM_101083.4	0.292	0.346	0.304	0.261	0.203	0.021	57.38	0.498	0.488
葡萄( <i>Vitisvinifera</i> )	XM_019222641.1	0.338	0.306	0.328	0.242	0.198	-0.008	55.15	0.443	0.461
绒毛状烟草( <i>Nicotianatomentosiformis</i> )	XM_018773585.1	0.357	0.255	0.368	0.229	0.178	-0.063	54.82	0.391	0.431
树棉( <i>Gossypiumarboreum</i> )	XM_017783494.1	0.317	0.279	0.386	0.228	0.172	-0.072	49.50	0.412	0.433
栓皮栎树( <i>Quercussuber</i> )	XM_024072909.1	0.336	0.285	0.286	0.302	0.194	-0.011	54.48	0.474	0.472
水稻( <i>Oryza sativa</i> )	XM_015776737.1	0.360	0.266	0.329	0.277	0.220	0.016	53.59	0.428	0.447
笋瓜( <i>Cucurbita maxima</i> )	XM_023114122.1	0.267	0.320	0.307	0.318	0.166	-0.103	54.91	0.516	0.481
桃树( <i>Prunuspersica</i> )	AB089677.1	0.252	0.390	0.287	0.290	0.190	-0.036	54.29	0.552	0.488
狭叶羽扇豆( <i>Lupinusangustifolius</i> )	XM_019591155.1	0.320	0.277	0.369	0.269	0.177	-0.042	52.42	0.433	0.434
向日葵( <i>Helianthus annuus</i> )	XM_022144642.1	0.325	0.287	0.332	0.263	0.191	-0.043	56.71	0.447	0.460
小麦( <i>Triticumaestivum</i> )	FB829196.1	0.387	0.248	0.333	0.262	0.201	-0.077	51.49	0.401	0.445
油菜( <i>Brassica rapa</i> )	XM_009112421.2	0.294	0.332	0.320	0.262	0.209	0.028	56.29	0.485	0.477
橄榄( <i>Olea europaea</i> )	XM_023020390.1	0.335	0.279	0.364	0.236	0.178	-0.042	55.72	0.416	0.447

*CAI*、*CBI*、*ENc*、*GC3s*、*GC* 见表 1 注, *T3s*、*C3s*、*A3s*、*G3s* 分别表示密码子第 3 位是 T、C、A、G 的含量。

小麦的密码子偏好性差异最大,这与赵洋等<sup>[23]</sup>分析结果一致,同时, *CsNRT1.1* 基因与玉米密码子偏好性也大于双子叶植物拟南芥、烟草。结果表明,双子叶植物密码子偏好性与 *CsNRT1.1* 基因更接近,更适合 *CsNRT1.1* 的外源表达。酵母菌、大肠杆菌分别是最常应用的真核和原核基因表达系统,比较 *CsNRT1.1* 与这 2 类表达系统的密码子使用频率比值,  $\geq 2.0$  或  $\leq 0.5$  的密码子分别为 15、20 个,说明 *CsNRT1.1* 基因与酵母菌密码子偏好性更接近。

## 2.5 不同物种 *NRT1.1* 基因聚类分析

基于 *NRT1.1* 基因密码子使用偏性的聚类分析结果表明,所有序列归为 5 个大类,其中茶树与大部分物种都归为一类,黄瓜、笋瓜等归为一类,拟南芥、

油菜和胡萝卜等归为一类,大麦、小麦和水稻等单子叶植物归为一类,高粱、葡萄和大豆归为一类(图 1)。而在基于基因 CDS 序列的聚类分析中,所有序列归为 5 大类,高粱、葡萄和大豆各为一类,玉米、水稻、大麦和小麦聚为一类,包括茶树在内的其他物种聚为一类(图 2)。可见,两种聚类方法得到的结果存在部分相同之处,说明相对 CDS 聚类,密码子偏好性聚类也能部分反映物种间亲缘关系的远近;同时两种方法得到的结果也存在一定差异,其中基于 CDS 的聚类较基于密码子偏好性的聚类能更好地反映这些物种的系统分类,例如玉米在 CDS 聚类分析中与水稻等单子叶植物归为一类,而在密码子偏好性聚类中则与油菜等聚在一类中。

表 4 *CsNRT1.1* 基因和不同模式物种密码子使用频率比值

Table 4 The ratio of codon usage frequency between *CsNRT1.1* gene and other species

氨基酸	密码子	C/A	C/N	C/T	C/Z	C/S	C/E
A(Ala)	GCA	0.96	0.73	1.09	1.00	1.04	0.81
	GCC	2.61	2.15	0.82	0.86	2.13	1.07
	GCG	0.93	1.45	0.38	0.36	1.35	0.26
C(Cys)	GCT	0.83	0.75	1.43	1.12	1.11	1.53
	TGC	1.17	1.17	0.62	0.69	1.75	1.31
	TGT	0.48	0.51	0.99	0.90	0.62	0.97
D(Asp)	GAC	1.17	1.19	0.69	0.63	1.00	1.05
	GAT	0.41	0.41	0.88	0.66	0.40	0.46
E(Glu)	GAA	0.44	0.42	0.99	0.76	0.33	0.38
	GAG	0.52	0.57	0.44	0.41	0.87	0.90
F(Phe)	TTC	1.95	2.24	1.62	1.60	2.19	2.53
	TTT	0.92	0.80	1.64	1.60	0.77	0.91
G(Gly)	GGA	0.42	0.43	0.68	0.75	0.92	1.13
	GGC	2.55	2.10	0.76	0.78	2.40	0.84
	GGG	2.14	2.08	1.22	1.42	3.64	1.85
	GGT	0.68	0.68	1.10	1.07	0.63	0.62
H(His)	CAC	0.77	0.77	0.49	0.45	0.86	0.71
	CAT	0.73	0.75	1.20	1.00	0.74	0.79
I(Ile)	ATA	1.07	0.96	2.00	1.60	0.75	2.44
	ATC	1.27	1.69	0.97	1.03	1.37	0.98
	ATT	1.09	0.84	2.01	1.70	0.78	0.79
K(Lys)	AAA	0.76	0.72	2.24	1.56	0.56	0.69
	AAG	0.72	0.70	0.62	0.60	0.76	2.14
L(Leu)	CTA	1.69	1.79	2.27	2.30	1.25	4.30
	CTC	1.46	1.91	0.86	0.92	4.35	2.24
	CTG	0.86	0.82	0.38	0.33	0.80	0.16
	CTT	0.70	0.70	1.36	1.07	1.36	1.47
	TTA	1.32	1.25	4.53	2.94	0.64	1.22
M(Met)	TTG	1.93	1.81	3.33	3.10	1.48	3.10
	ATG	1.16	1.14	1.17	1.18	1.36	1.05
N(Asn)	AAC	0.96	1.12	0.94	0.91	0.81	0.93
	AAT	0.53	0.42	1.12	0.87	0.33	0.61
P(Pro)	CCA	0.94	0.76	0.65	1.09	0.83	1.80
	CCC	1.58	1.27	0.57	0.62	1.23	1.50
	CCG	0.78	1.34	0.41	0.42	1.27	0.30
	CCT	0.54	0.54	0.86	0.80	0.75	1.40
Q(Gln)	CAA	1.56	1.46	0.72	2.29	1.11	2.05
	CAG	0.55	0.56	0.22	0.35	0.69	0.29
R(Arg)	AGA	0.26	0.31	0.75	0.57	0.24	1.74
	AGG	1.53	1.38	1.29	1.13	1.82	9.32
	CGA	0.27	0.32	0.56	0.38	0.56	0.44
	CGC	0.88	0.86	0.26	0.23	1.29	0.16

续表 4 Continued

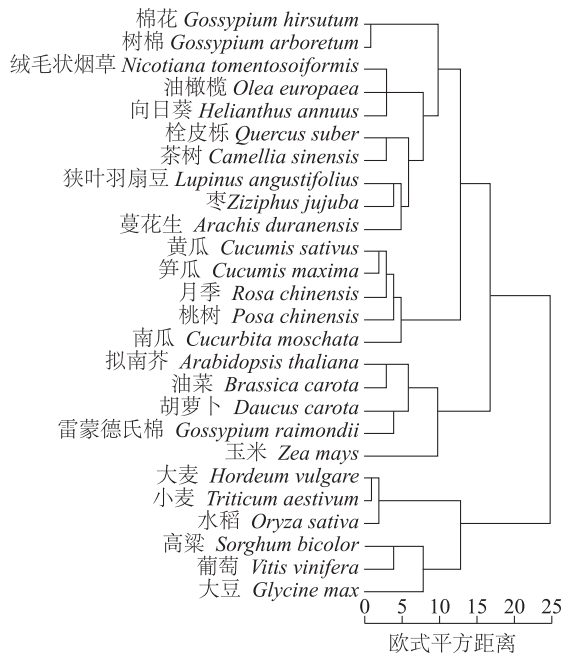
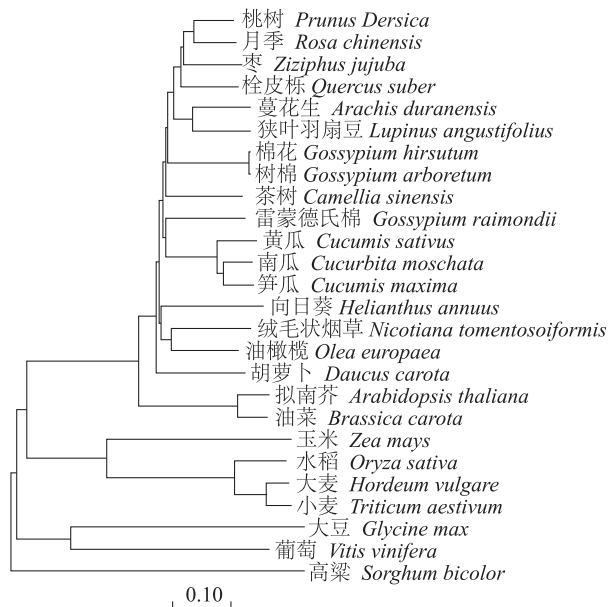
氨基酸	密码子	C/A	C/N	C/T	C/Z	C/S	C/E
S(Ser)	CGG	1.03	1.36	0.57	0.53	2.96	0.81
	CGT	0.37	0.45	0.60	0.56	0.52	0.17
	AGC	1.34	1.51	0.91	0.92	1.54	0.94
	AGT	0.96	1.01	2.03	1.72	0.95	1.43
	TCA	0.92	0.95	1.58	1.53	0.90	2.07
T(Thr)	TCC	0.60	0.66	0.37	0.41	0.47	0.75
	TCG	0.90	1.58	0.78	0.78	0.98	0.95
	TCT	0.40	0.50	0.98	0.84	0.43	1.16
	ACA	1.28	1.16	2.21	1.92	1.13	2.49
	ACC	3.42	3.63	1.84	2.12	2.77	1.55
V(Val)	ACG	1.31	2.24	1.04	0.92	1.26	0.67
	ACT	1.34	1.16	2.55	2.20	1.16	2.58
	GTA	0.68	0.59	1.27	1.05	0.57	0.62
	GTC	0.92	1.06	0.55	0.56	1.00	0.79
	GTG	2.12	2.21	1.48	1.45	3.42	1.41
W(Trp)	GTT	0.86	0.88	1.64	1.50	1.06	1.30
	TGG	1.88	1.93	1.96	1.81	2.26	1.54
Y(Tyr)	TAC	0.98	0.99	0.65	0.69	0.91	1.09
	TAT	1.03	0.85	1.82	1.59	0.80	0.92
*	TAA	1.86	1.53	2.80	3.36	1.53	0.84
*	TAG	0	0	0	0	0	0
*	TGA	0	0	0	0	0	0

\* 表示终止密码子。C/A、C/N、C/T、C/Z、C/S 和 C/E 分别表示 *CsNRT1.1* 基因与拟南芥、烟草、小麦、玉米、酵母菌和大肠杆菌的每种密码子使用频率的比值。

### 3 讨论

*CsNRT1.1* 基因负责茶树根系 NO<sub>3</sub><sup>-</sup> 吸收及转运,可能具有独特的密码子偏好性。对 *CsNRT1.1* 基因及相关基因密码子使用特征进行分析,发现 *CsNRT1.1* 基因偏好使用以 A/U 结尾的密码子,这与前人发现的茶树基因组密码子偏好使用 A/U 结尾的研究结果相一致<sup>[23]</sup>。*CsNRT1.1* 基因中 29 个密码子具有偏好性并偏好以 G/C 结尾,其中使用频率最高的是 AGG(*RSCU* 值=2.86)。比较茶树基因组与 *CsNRT1.1* 基因的密码子使用特性,发现基因组中 26 个密码子具有偏好性,并偏好以 G/C 结尾,没有密码子表现出较强的偏好性(*RSCU* 值>2)。

*CsNRT1.1* 基因和茶树基因组中共计 25 个密码子使用偏好性存在明显差异,而密码子偏好性一致的氨基酸只有 5 种。Lys 的偏好密码子在植物中具

图1 基于 *NRT1.1* 密码子使用偏性的聚类分析Fig.1 Cluster analysis dendrogram of codon usage preference of *NRT1.1*图2 基于 *NRT1.1* 基因 CDS 组成的聚类分析Fig.2 Cluster analysis dendrogram of CDS sequences of *NRT1.1*

有较高的保守性,包括在茶树基因组中普遍偏好使用 AAG<sup>[17,23]</sup>。本研究中茶树基因组中 Lys 也偏好使用 AAG,而 Lys 在 *CsNRT1.1* 中没有偏好性,这与以往研究结果有所不同。可见,*CsNRT1.1* 基因与茶

树基因组在密码子偏好性上存在差异,这可能受 *CsNRT1.1* 表达水平、功能等影响,具体原因有待进一步研究。

*CsNRT1.1* 及 26 个其他 *NRT1* 基因密码子的 CAI、CBI 值都较低,说明 *NRT1.1* 基因密码子偏好性不强;同时大部分基因 *ENc* 值小于 55,只有包括 *CsNRT1.1* 在内的少数 *NRT1.1* 基因密码子 *ENc* 值大于 55。*ENc* 值越接近 20,则表达水平越高,*ENc* 值 <30 或 >55 的基因分别被预测为高表达或低表达基因<sup>[5]</sup>。*CsNRT1.1* 基因和油菜、胡萝卜、黄瓜、油橄榄、枣、月季和向日葵 *NRT1.1* 基因的 *ENc* 值都大于 55(占总体的 31.8%),说明这些基因表达水平可能较低。

在基因功能研究中,转基因异源表达是经常使用的研究手段,而密码子偏好性会影响外源基因在宿主中的表达效率。由于不同物种间密码子偏好性差异较大,分析外源基因与宿主的密码子使用特性差异,采用符合宿主基因组的优化改造的密码子,有助于提高基因表达水平<sup>[18]</sup>。比较 *CsNRT1.1* 基因与常用宿主的密码子偏好性,均存在不同程度的偏好性差异,其中与双子叶植物拟南芥、烟草及酵母菌偏好性差异较小,说明这些物种更适合作为 *CsNRT1.1* 基因的外源表达宿主。同时,对不同物种 *NRT1.1* 基因进行聚类分析,基于 CDS 组成的聚类结果较基于密码子使用偏性的聚类分析结果更接近物种的植物学分类,在其他物种密码子偏好性研究中也存在类似报道<sup>[20,22,27]</sup>。在基于密码子使用偏性的聚类分析中,喜铵态氮植物茶树和喜硝态氮植物棉花、黄瓜、向日葵、烟草归为一类,同时喜铵态氮植物水稻和喜硝态氮植物小麦归为一类,喜铵态氮植物拟南芥和喜硝态氮植物玉米、油菜归为一类,这说明植物喜硝态氮与否与 *NRT1.1* 密码子使用偏性并无直接相关性。

本研究对 *CsNRT1.1* 基因密码子特性进行了分析,发现其密码子偏好性与茶树基因组、其他物种 *NRT1.1* 基因都存在差异,同时发现 *CsNRT1.1* 基因与常用宿主表达系统的密码子偏好性也存在差异,考虑到 *CsNRT1.1* 基因可能为低表达基因,若要实现 *CsNRT1.1* 在其他宿主中的高效表达,有必要对 *CsNRT1.1* 密码子进行优化改造以提高基因表达水平。近年来研究发现,*NRT1.1* 在植物氮素吸收中起到非常重要的作用。例如在低  $\text{NO}_3^-$  条件下,拟南芥

*NRT1.1* 基因通过苏氨酸磷酸化修饰,表现  $\text{NO}_3^-$  双亲和转运功能来促进氮素吸收<sup>[28]</sup>。籼稻中 *NRT1.1* 碱基变异和基因表达与氮肥利用效率密切相关,该基因过表达后显著提高水稻生物量和产量<sup>[29-30]</sup>。可见 *NRT1.1* 基因在提高植物氮素利用效率方面具有很好的应用前景。本研究密码子偏好性分析结果为茶树中 *CsNRT1.1* 基因表达水平和活性提升提供了参考,为茶树氮高效利用的分子设计育种提供了理论基础,这对于茶树氮素利用效率提升及茶园减肥增效,具有一定的参考意义。

### 参考文献:

- [1] 赵 洋, 刘 振, 杨培迪, 等. 密码子偏性分析方法及茶树中密码子偏性研究进展[J]. 茶叶通讯, 2016, 43(2): 3-7.
- [2] CHIAPELLO H, LISACEK F, CABOCHE M, et al. Codon usage and gene function are related in sequences of *Arabidopsis thaliana* [J]. Gene, 1998, 209(1): 1-38.
- [3] POWELL J R, SEZZI E, MORIYAMA E N, et al. Analysis of a shift in codon usage in *Drosophila* [J]. J MolEvol, 2003, 57 (Suppl 1): 214-225.
- [4] GUO Y, LIU J, ZHANG J, et al. Selective modes determine evolutionary rates, gene compactness and expression patterns in *Brassica* [J]. The Plant Journal, 2017, 91(1): 34-44.
- [5] SHARP P M, LI W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. Journal of Molecular Evolution, 1986, 24(1/2): 28-38.
- [6] LI F, YANG S Y, ZHAO L G, et al. Synonymous condon usage bias and overexpression of a synthetic Xynb gene from *Aspergillus Niger* N1-1 in *Pichia Pastoris* [J]. Bioresources, 2012, 7(2): 2330-2343.
- [7] GUSTAFSSON C, GOVINDARAJAN S, MINSHULL J. Codon bias and heterologous protein expression [J]. Trends in Biotechnology, 2004, 22(7): 346-353.
- [8] 曲俊杰, 黄 羽, 卢 江, 等. 香蕉基因组密码子使用偏好性分析[J]. 南方农业学报, 2017, 48(1): 14-19.
- [9] 冯素花, 王丽鸳, 陈常颂, 等. 茶树硝酸根转运蛋白基因 *NRT2.5* 的克隆及表达分析 [J]. 茶叶科学, 2014, 34(4): 364-370.
- [10] 汪 进, 添先凤, 江昌俊, 等. 茶树硝酸盐转运蛋白基因的克隆和表达分析[J]. 植物生理学报, 2014, 50(7): 983-988.
- [11] 杨亦扬, 胡云飞, 万 青, 等. 茶树硝态氮转运蛋白 *NRT1.1* 基因的克隆及表达分析 [J]. 茶叶科学, 2016, 36(5): 505-512.
- [12] 侯 美, 喻春明. 植物 *NRT1.1* 的研究进展及其在苧麻研究中的展望[J]. 中国麻业科学, 2017, 39(4): 207-214.
- [13] KAMAU D M, SPIERTZ J H, OENEMA O, et al. Productivity and nitrogen use of tea plantations in relation to age and genotype [J]. Field Crops Research, 2008, 108(1): 60-70.
- [14] YANG Y Y, LI X H, RATCLIFFE R G, et al. Characterization of ammonium and nitrate uptake and assimilation in roots of tea plants [J]. Russian Journal of Plant Physiology, 2013, 60(1): 91-99.
- [15] RUAN J Y, GERENDAS J, HARDTER R, et al. Effect of nitrogen form and root-zone pH on growth and nitrogen uptake of tea (*Camellia sinensis*) plants [J]. Annals of Botany, 2007, 99 (2): 301-310.
- [16] KRISHNAPILLAI S. Inhibition of nitrification by waste tea (tea fluff) [J]. Plant and Soil, 1979, 51(4): 563-569.
- [17] 胡桂兵, 张上隆, 徐昌杰, 等. 柑橘密码子用法分析[J]. 果树学报, 2006, 23(3): 479-485.
- [18] 赵 耀, 刘汉梅, 顾 勇, 等. 玉米 *waxy* 基因密码子偏好性分析[J]. 玉米科学, 2008, 16(2): 16-21.
- [19] 杨国锋, 苏昆龙, 赵怡然, 等. 蕹菜苜蓿叶绿体密码子偏好性分析[J]. 草业学报, 2015, 24(12): 171-179.
- [20] 赖瑞联, 冯 新, 陈 瑾, 等. 橄榄查尔酮异构酶基因 *CHI* 的密码子偏好模式[J]. 应用与环境生物学报, 2017, 23(5): 945-951.
- [21] 李 蓉, 谢析颖, 王雪晶, 等. 兰科植物 *FNR* 基因的密码子偏好性分析[J]. 热带作物学报, 2018, 39(6): 1137-1145.
- [22] 赵 洋, 刘 振, 杨培迪, 等. 茶树 *CsActin1* 基因密码子偏性分析[J]. 茶叶通讯, 2014, 41(4): 13-17.
- [23] 赵 洋, 杨 阳, 刘 振, 等. 茶树密码子用法分析[J]. 茶叶科学, 2011, 31(4): 319-325.
- [24] 郭秀丽, 王 玉, 杨路成, 等. 茶树 *CBF1* 基因密码子使用特性分析[J]. 遗传, 2012, 34(12): 1614-1623.
- [25] 时慧, 王 玉, 杨路成, 等. 茶树抗寒调控转录因子 *ICE1* 密码子偏性分析[J]. 园艺学报, 2012, 39(7): 1341-1352.
- [26] SHARP P M, COWE E. Synonymous codon usage in *Saccharomyces cerevisiae* [J]. Yeast, 1991, 7(7): 657-678.
- [27] 孙继峰, 韩太利, 徐立功, 等. 萝卜 *ICE1* 密码子使用偏性分析[J]. 核农学报, 2018, 32(3): 477-485.
- [28] PARKER J L, NEWSTEAD S. Molecular basis of nitrate uptake by the plant nitrate transporter *NRT1.1* [J]. Nature, 2014, 507: 68.
- [29] HU B, WANG W, OU S J, et al. Variation in *NRT1.1B* contributes to nitrate-use divergence between rice subspecies [J]. Nature Genetics, 2015, 47: 834.
- [30] WANG W, HU B, YUAN D Y, et al. Expression of the nitrate transporter gene *OsNRT1.1A/OsNPF6.3* confers high yield and early maturation in rice [J]. Plant Cell, 2018, 30(3): 638-651.

(责任编辑:张震林)