

DOI:10.13275/j.cnki.lykxyj.2021.02.011

马尾松 *PmDXR* 基因密码子偏好性分析

朱灵芝, 朱沛煌, 李 荣, 陈 好, 季孔庶*

(南京林业大学, 林木遗传与生物技术省部共建教育部重点实验室, 南方现代林业协同创新中心, 江苏 南京 210037)

摘要: [目的] 对马尾松 *PmDXR* 基因密码子偏好性进行分析, 为后续 *PmDXR* 基因的功能验证及遗传转化等研究提供理论依据。[方法] 利用 CodonW、EMBOSS、SPSS 等软件及在线服务器, 分析马尾松与其他 37 个物种 *DXR* 基因密码子的偏好性。对不同物种 *DXR* 基因密码子进行中性分析、PR2-plot 分析、ENc-plot 分析, 推测其偏好性形成的主要影响因素; 同时分析比较了马尾松 *PmDXR* 基因与 4 种模式生物的密码子使用频率。[结果] 马尾松 *PmDXR* 基因的密码子偏好性较弱, 偏好使用 A/T 结尾的密码子。比较分析不同物种 *DXR* 基因的密码子偏好性参数, 发现不同物种 *DXR* 基因密码子偏好性存在一定差异, 整体看, 裸子植物 *DXR* 基因密码子偏好性比被子植物强。与其他物种 *DXR* 基因的 CDS 序列及 RSCU 值聚类分析表明, 马尾松与其它裸子植物 *DXR* 基因表现出相近的密码子使用偏好性。中性分析、PR2-plot 分析、ENc-plot 分析均表明: 自然选择是影响 *DXR* 基因密码子偏好性形成的主要因素。与 4 种模式生物的基因组密码子使用频率比较发现, 在异源转化受体选择中, 酵母真核表达系统较大肠杆菌原核表达系统更适合马尾松 *PmDXR* 异源表达, 拟南芥和烟草均可作为马尾松 *PmDXR* 基因的遗传转化受体, 其中, 烟草更理想。[结论] 本研究初步揭示了马尾松 *PmDXR* 基因密码子使用规律, 对后续开展基因功能验证有重要的指导作用, 同时也可为研究马尾松分子进化提供一定的科学依据。

关键词: 马尾松; *PmDXR* 基因; 密码子偏好性; 异源表达; 外源宿主

中图分类号: S791.248

文献标志码: A

文章编号: 1001-1498(2021)02-0102-12

马尾松 (*Pinus massoniana* Lamb.) 在我国松属树种中分布最广泛, 具有适应性强、经济价值高等特点。它分泌的大量以萜类化合物为主的次生代谢物称为松脂, 主要由单萜、倍半萜和二萜组成, 贮存在针叶树的根、茎、叶和球果中^[1]。松脂是松香、松节油工业的基础原料, 松香、松节油深加工产品又可作为其他相关工业的中间原料, 广泛用于油墨、胶黏剂、香料、涂料、纺织、电子印染和医药等工业生物产品中^[2]。

萜类化合物是植物次生代谢产物中最大的一类天然化合物, 广泛存在于动植物体内, 目前已研究报道过的萜类化合物有 5 万多种。萜类化合物有重要的生物学功能和应用价值, 在植物体内, 单萜和倍半萜可看作挥发油的成分, 大多数具有香气, 是食品、医药、化妆品的重要原料。超过 20 个碳的

萜类物质一般为非挥发性物质, 树脂酸和赤霉素是二萜中最常见的物质, 其中, 树脂酸在马尾松中占所有代谢物的 70% 以上^[3]。1-脱氧-d-木酮糖-5-磷酸还原异构酶 (*DXR*) 是马尾松萜类化合物合成途径中的重要限速酶, 它催化 1-脱氧-d-木酮糖-5-磷酸 (*DXP*) 发生还原反应生成 2-C-甲基-D-赤藓糖醇-4-磷酸 (*MEP*)^[4], 同时 *DXP* 也是维生素 B1 和 B6 的前体物质^[5-6], 故该催化反应被认为是 *MEP* 途径中的关键调控位点。目前, 已从多种植物中克隆出 *DXR* 基因, 并对其功能作了相关研究; 但关于 *DXR* 基因密码子偏好性的研究较少, 仅在豆科植物中有报道, 李凌焯等发现豆科植物 *DXR* 基因对 A 或 U 结尾的密码子表现出较强的偏好性^[7]。

密码子是自然界中遗传信息准确识别和传递的基本载体, 在生物遗传和变异中起着关键作用。构

收稿日期: 2020-07-28 修回日期: 2020-12-23

基金项目: 国家重点研发计划 (2017YFD0600304); 江苏高校优势学科建设工程资助项目 (PAPD)

* 通讯作者: 季孔庶, 教授, 博士生导师, Email: ksj32@vip.163.com

成天然蛋白质的 20 种氨基酸中除甲硫氨酸和色氨酸外, 其余氨基酸都对应 2~6 种密码子, 编码同一种氨基酸的密码子被称为同义密码子^[8]。研究发现, 不同生物体在编码氨基酸时对密码子的选择并不是完全随机的, 而是偏好使用某些特定的同义密码子^[9]; 另外, 同种生物中不同基因对密码子的使用频率也存在差异, 这种现象即为密码子使用偏好性^[10]。同义密码子的选择使用不改变氨基酸序列, 有利于翻译的准确性, 但可能会对蛋白表达水平产生影响^[11-12]。密码子偏好性会影响蛋白质翻译的速度和折叠、mRNA 水平的转录调控及外源基因的表达量^[13-14]。如果外源基因密码子与宿主系统密码子的使用频率相比有明显差异, 将会影响外源基因的翻译与表达^[15]。因此, 分析密码子的偏好性对研究基因功能、蛋白结构和提高外源基因表达效率有重要意义。目前, 科学界尚无关于马尾松 *PmDXR* 基因在密码子偏好性方面的报道, 对其相关功能尚不清楚。本研究将分析马尾松 *PmDXR* 基因密码子使用偏好性, 讨论其形成的主要因素, 再与拟南芥 (*Arabidopsis thaliana* (L.) Heynh.)、烟草 (*Nicotiana tabacum* L.)、酿酒酵母 (*Saccharomyces cerevisiae* Meyen. ex Hansen.) 和大肠杆菌 (*Escherichia coli* (Migula.) Castellani & Chalmers.) 基因组密码子进行比较分析, 以期为马尾松 *PmDXR* 功能验证相关研究提供基础数据。

1 材料与方 法

1.1 序列来源

马尾松 *PmDXR* 基因序列由本实验室克隆获得, 该序列全长 1 994 bp (GenBank 登录号为 MK969119.1), 其开放阅读框 (ORF) 为 1 584 bp, 编码 527 个氨基酸。银杏 (*Ginkgo biloba* L.)、赤松 (*Pinus densiflora* Sieb. et Zucc.)、火炬松 (*Pinus taeda* L.) 等 38 个物种的 *DXR* 基因 CDS 序列来自 GenBank (<https://www.ncbi.nlm.nih.gov/>), 登录号见表 1。本文所采用的拟南芥、烟草、酿酒酵母和大肠杆菌的基因组密码子使用频率数据来自密码子使用数据库 (<http://www.kazusa.or.jp/codon>)。

1.2 数据分析方法

1.2.1 *DXR* 基因同义密码子偏好性分析 整理 38 个物种 *DXR* 基因的 CDS 序列, 借助 CodonW1.4.2 软件计算密码子的使用特性参数, 包括 A3s、

C3s、U3s、G3s、相对同义密码子使用度 (RSCU)、密码子适应指数 (CAI) 和有效密码子数 (ENc) 等。运用 EMBOSS 中的 CUSP 和 CHIP 在线程序计算 38 个物种 *DXR* 基因密码子 G 和 C 总含量 (GC)、密码子第 3 位上的 GC 含量 (GC3s)、使用频率 (Frequency) 等; 同时运用 SPSS20.0 软件分析 38 个物种 *DXR* 基因密码子使用特性参数的相关性, 再进一步对 *DXR* 基因密码子进行中性绘图分析、奇偶偏好性 (PR2-plot) 分析以及 ENc-plot 分析。

CAI 值范围为 0~1, 越接近 1 表明偏好性越强, 而越接近 0 则表明偏好性越弱^[16]。ENc 值范围为 20~61, 可反映特定基因密码子家族中同义密码子非均衡使用的偏好程度^[17]。RSCU 为某一密码子的使用频率与其无偏性时期望频率的比值。一般, 若 RSCU 值小于 1, 表明该密码子使用偏好性较低; 若其值等于 1, 表明该密码子无偏好性; 若其值大于 1, 则表明该密码子使用偏好性较强^[18]。

1.2.2 基于 *DXR* 基因 CDS 序列和密码子使用偏好性的聚类分析 通过 MEGA5.1 软件中的邻接法 (Neighbor-Joining), 基于 38 个物种 *DXR* 基因的 CDS 序列构建系统进化树, 检验方法为 Bootstrap method, 检验次数为 1 000, Gaps 处理方法为 Complete deletion。并利用在线绘图软件 EvolView (<https://evolgenius.info/evolview-v2/#login>) 可视化展示进化树。将每个物种的 *DXR* 基因作为一个对象, 以 59 个同义密码子的 (去除终止密码子 TAA、TAG 和 TGA, 起始密码子 ATG 和编码甲硫氨酸的密码子 TGG) 的 RSCU 值为变量对密码子使用偏好性进行分析。

2 结果分析

2.1 马尾松 *DXR* 基因密码子偏性分析

2.1.1 GC、GC3s、ENc 和 CAI 分析 利用 CondonW 程序分析马尾松 *PmDXR* 基因密码子的 CAI 值、ENc 值, 同时运用 EMBOSS 中的 CUSP 在线程序计算总 GC 含量以及 GC3s 有效密码子数。结果 (表 1) 显示: *PmDXR* 基因中密码子的 GC 和 GC3s 分别为 43.75% 和 34.47%, 说明其密码子比较偏好以 A/T 结尾。ENc 值为 35 可作为区分偏好性强弱的标准^[19], *PmDXR* 基因密码子 ENc 值为 48.93, 说明 *PmDXR* 基因在编码氨基酸时密码子使用偏好性较弱, 其表达水平可能相对较低。CAI 值为 0.184, 进一步表明 *PmDXR* 基因对密码子的选择偏好性较弱。

表1 不同物种间DXR基因密码子选择偏好性相关参数

Table 1 Preference related parameters of DXR gene codons among different species

类别 Classes	物种 Species	登录号 Genbank accession number	A3s	C3s	G3s	U3s	ENc	CAI	GC3s/ %	GC/ %
裸子植物	银杏 <i>Ginkgo biloba</i>	AY494186.1	0.350	0.177	0.256	0.435	51.05	0.200	36.82	45.26
	赤松 <i>Pinus densiflora</i>	EU439294.1	0.390	0.178	0.231	0.425	49.32	0.192	34.58	44.17
	火炬松 <i>Pinus taeda</i>	EU862299.1	0.396	0.171	0.226	0.426	47.89	0.188	33.75	44.24
	马尾松 <i>Pinus massoniana</i>	MK969119.1	0.410	0.186	0.222	0.402	48.93	0.184	34.47	43.75
	东北红豆杉 <i>Taxus cuspidata</i> Sieb. et Zucc.	AY575140.1	0.383	0.176	0.242	0.420	47.99	0.216	35.56	44.70
单子叶植物	枫茅 <i>Cymbopogon winterianus</i> Jowitt.	KJ749651.1	0.346	0.228	0.269	0.374	55.97	0.203	41.60	47.69
	盾叶薯蓣 <i>Dioscorea zingiberensis</i> C. H. Wright.	KY131955.1	0.326	0.257	0.252	0.384	53.92	0.208	42.89	47.49
	浦竹仔 <i>Indosasa hispida</i> McClure.	KX819309.1	0.319	0.229	0.301	0.380	53.41	0.235	43.64	47.88
	水稻 <i>Oryza sativa</i> L.	AF367205.1	0.330	0.240	0.279	0.375	54.75	0.214	43.46	48.03
双子叶植物	砂仁 <i>Amomum villosum</i> Lour.	FJ459894.1	0.349	0.221	0.245	0.415	52.84	0.193	38.90	46.23
	金鱼草 <i>Antirrhinum majus</i> L.	AY770406.1	0.280	0.224	0.281	0.450	54.68	0.227	41.53	46.12
	沉香 <i>Aquilaria sinensis</i> (Lour.) Spreng	KF498704.1	0.264	0.221	0.291	0.450	52.71	0.227	42.17	47.11
	拟南芥 <i>Arabidopsis thaliana</i>	AF148852.1	0.326	0.178	0.250	0.473	49.92	0.212	35.98	44.56
	黄花蒿 <i>Artemisia annua</i> L.	AF182287.2	0.313	0.190	0.241	0.474	51.65	0.229	36.36	44.05
	艾草 <i>Artemisia argyi</i> Lévl. et Van.	MG780999.1	0.323	0.203	0.235	0.455	50.48	0.228	37.00	44.33
	苍术 <i>Atractylodes lancea</i> (Thunb.) DC.	MG859909.1	0.263	0.265	0.282	0.421	54.13	0.244	44.94	47.19
	胭脂树 <i>Bixa orellana</i> L.	KT358987.1	0.279	0.205	0.261	0.483	52.04	0.233	38.89	45.62
	积雪草 <i>Centella asiatica</i> (L.) Urban.	JQ965955.1	0.298	0.214	0.255	0.461	50.52	0.217	38.95	45.54
	樟树 <i>Cinnamomum camphora</i> (L.) Presl	KU886266.1	0.297	0.220	0.296	0.413	53.46	0.208	42.68	47.13
	枇杷 <i>Eriobotrya japonica</i> (Thunb.) Lindl.	JX089590.1	0.295	0.217	0.256	0.446	55.01	0.216	39.66	46.84
	栀子花 <i>Gardenia jasminoides</i> Ellis.	KY631919.1	0.326	0.233	0.229	0.439	52.50	0.211	38.57	45.56
	滇龙胆草 <i>Gentiana rigescens</i> Franch. ex Hemsl.	KF941189.1	0.276	0.196	0.282	0.475	50.44	0.227	39.37	44.98
	橡胶树 <i>Hevea brasiliensis</i> (Willd. ex A. Juss.) Muell. Arg.	DQ437514.1	0.323	0.180	0.229	0.490	52.05	0.200	34.53	44.42
	薰衣草 <i>Lavandula angustifolia</i> Mill.	JX630151.1	0.217	0.307	0.329	0.382	53.97	0.209	51.68	49.23
	番茄 <i>Lycopersicon esculentum</i> Mill.	AF331705.2	0.309	0.202	0.264	0.444	49.67	0.191	38.87	45.59
	桑 <i>Morus alba</i> L.	KX387390.1	0.297	0.262	0.258	0.403	55.14	0.218	43.43	47.53
	水仙 <i>Narcissus tazetta</i> var. <i>chinensis</i> Roem.	GU574805.1	0.313	0.196	0.273	0.439	51.40	0.204	39.28	46.28
	木樨 <i>Osmanthus fragrans</i> (Thunb.) Lour.	KX400843.1	0.313	0.218	0.240	0.458	53.99	0.213	38.11	45.19
	夏枯草 <i>Prunella vulgaris</i> L.	MK272811.1	0.269	0.309	0.276	0.372	59.49	0.235	48.52	49.16
	西洋梨 <i>Pyrus communis</i> L.	KM030333.1	0.290	0.213	0.273	0.442	54.71	0.210	40.55	46.99
	萝芙木 <i>Rauvolfia verticillata</i> (Lour.) Baill.	DQ779286.1	0.321	0.203	0.233	0.475	50.94	0.219	36.21	44.14
	地黄 <i>Rehmannia glutinosa</i> (Gaert.) Libosch. ex Fisch. et Mey.	KX058461.1	0.289	0.215	0.273	0.453	55.23	0.206	40.42	45.89
	玫瑰 <i>Rosa rugosa</i> Thunb.	JX518616.1	0.324	0.190	0.248	0.452	52.32	0.201	37.00	45.88
	丹参 <i>Salvia miltiorrhiza</i> Bunge.	DQ991431.1	0.274	0.269	0.309	0.388	55.65	0.205	47.37	48.35
	半枝莲 <i>Scutellaria barbata</i> D. Don.	MK035042.1	0.265	0.251	0.285	0.425	56.63	0.209	44.26	47.69
	阿育魏实 <i>Trachyspermum ammi</i> (L.) Sprague	MG762013.1	0.320	0.175	0.235	0.500	48.84	0.216	34.11	43.65
	雷公藤 <i>Tripterygium wilfordii</i> Hook. f.	KJ174341.1	0.301	0.179	0.261	0.496	53.61	0.232	36.38	44.40
	睡茄 <i>Withania somnifera</i> (L.) Dunal	GQ921844.1	0.297	0.199	0.274	0.446	51.30	0.197	39.58	45.96

2.1.2 同义密码子相对使用度分析 CUSP 和 CodonW 计算结果 (表 2) 表明: 有 25 个密码子的 RSCU 值大于 1 (除终止密码子 TGA), 其中, GCA、GCT、AGA、AAT、GGA、CCA、CCT、TCA、TCT、ACA、TAT、GTT 共 12 个密码子的

RSCU 值大于 1.5, 且编码精氨酸的 AGA 与编码丝氨酸的 TCT 密码子 RSCU 值分别为 3.60、2.17, 说明 *PmDXR* 基因中共有 25 个密码子具有偏好性, 其中, 有 12 个密码子偏好性较强, 2 个偏好性极强。TGC、TGT、TCC、TGG 共 4 个密码子 (除起

表 2 马尾松 *PmDXR* 基因同义密码子相对使用度

Table 2 RSCU of *PmDXR* gene in *P. massoniana*

密码子 Codon	氨基酸 Amino acid	比例 Fraction	频率 Frequency	数目 Number	相对密码子使用度 RSCU	密码子 Codon	氨基酸 Amino acid	比例 Fraction	频率 Frequency	数目 Number	相对密码子使用度 RSCU
<u>GCA</u>	Ala	0.377	37.879	20	1.51	<u>CTT</u>	Leu	0.235	22.727	12	1.41
GCC		0.189	18.939	10	0.75	<u>TTA</u>		0.176	17.045	9	1.06
GCG		0.038	3.788	2	0.15	<u>TTG</u>		0.196	18.939	10	1.18
<u>GCT</u>		0.396	39.773	21	1.58	<u>AAA</u>	Lys	0.543	35.985	19	1.09
<u>AGA</u>	0.600	22.727	12	3.60	AAG	0.457		30.303	16	0.91	
AGG	0.150	5.682	3	0.90	ATG	Met	1.000	18.939	10	1.00	
CGA	Arg	0.150	5.682	3	0.90	TTC	Phe	0.316	11.364	6	0.63
CGC		0.000	0.000	0	0.00	<u>TTT</u>		0.684	24.621	13	1.37
CGG		0.100	3.788	2	0.60	<u>CCA</u>	0.441	28.409	15	1.76	
CGT		0.000	0.000	0	0.00	CCC	Pro	0.088	5.682	3	0.35
AAC	Asn	0.222	3.788	2	0.44	CCG		0.059	3.788	2	0.24
<u>AAT</u>		0.778	13.258	7	1.56	<u>CCT</u>		0.412	26.515	14	1.65
GAC	Asp	0.385	18.939	10	0.77	AGC	0.056	3.788	2	0.33	
<u>GAT</u>		0.615	30.303	16	1.23	AGT	0.083	5.682	3	0.50	
TGC	Cys	0.500	7.576	4	1.00	<u>TCA</u>	Ser	0.278	18.939	10	1.67
TGT		0.500	7.576	4	1.00	TCC		0.167	11.364	6	1.00
<u>CAA</u>	Gln	0.600	11.364	6	1.20	TCG		0.056	3.788	2	0.33
CAG		0.400	7.576	4	0.80	<u>TCT</u>	0.361	24.621	13	2.17	
GAA	Glu	0.471	30.303	16	0.94	<u>ACA</u>	0.500	24.621	13	2.00	
<u>GAG</u>		0.529	34.091	18	1.06	<u>ACC</u>	Thr	0.269	13.258	7	1.08
<u>GGA</u>	0.400	26.515	14	1.60	ACG	0.038		1.894	1	0.15	
GGC	Gly	0.143	9.470	5	0.57	ACT		0.192	9.470	5	0.77
GGG		0.200	13.258	7	0.80	TGG	Trp	1.000	15.152	8	1.00
<u>GGT</u>		0.257	17.045	9	1.03	TAC	Tyr	0.091	1.894	1	0.18
<u>CAC</u>		0.563	17.045	9	1.13	<u>TAT</u>		0.909	18.939	10	1.82
CAT	His	0.438	13.258	7	0.88	GTA	0.244	20.833	11	0.98	
<u>ATA</u>	Ile	0.390	30.303	16	1.17	GTC	Val	0.089	7.576	4	0.36
ATC		0.146	11.364	6	0.44	GTG		0.222	18.939	10	0.89
<u>ATT</u>		0.463	35.985	19	1.39	<u>GTT</u>		0.444	37.879	20	1.78
CTA	Leu	0.157	15.152	8	0.94	TAA		0.000	0.000	0	0.00
CTC		0.098	9.470	5	0.59	TAG	TER	0.000	0.000	0	0.00
CTG		0.137	13.258	7	0.82	TGA	1.000	1.894	1	3.00	

注: 下划线表示该密码子 RSCU > 1。

Note: The underline means that the value of RSCU > 1.

始密码子 ATG) 的 RSCU 值等于 1, 说明 *PmDXR* 基因中有 4 个密码子没有偏好性。29 个密码子的 RSCU 值小于 1, 说明 *PmDXR* 基因对这些密码子的使用偏好性较弱。编码 Arg 的密码子 CGC 和 CGT 的 RSCU 值为 0, 说明它们可能不参与马尾松 *PmDXR* 基因的翻译过程。此外, 马尾松中 NCG 型密码子的 RSCU 值 (GCG: 0.15, CCG: 0.24, TCG: 0.33, ACG: 0.15) 较低, 推断马尾松可能是一种甲基化程度较高的植物。

2.2 38 个物种 *DXR* 基因密码子使用偏好性比较

2.2.1 相关参数比较分析

38 个物种 *DXR* 基因密码子的 A3s、C3s、U3s、G3s、ENc 值、CAI 值、GC 和 GC3s 含量见表 1。分析比较 A3s、C3s、U3s 和 G3s 发现, 38 种植物 *DXR* 的 A3s 和 U3s 普遍高于 C3s 和 G3s, 表明 *DXR* 编码区优先使用 A 或 U 结尾的同义密码子。38 个物种 *DXR* 基因的 ENc 值为 47.89~59.49, 平均值为 52.59, 其中, 裸子植物 *DXR* 的 ENc 平均值为 49.04, 单子叶植物 *DXR* 的 ENc 平均值为 54.51, 双子叶植物 *DXR* 的 ENc 平均值为 52.94, 表明裸子植物 *DXR* 基因的密码子偏好性相对单子叶和双子叶植物较高。由此可见, *DXR* 基因在各物种中表现出较弱的密码子偏好性, 表达水平普遍不高, 但物种间还是存在一定差异。38 个物种 *DXR* 基因的 CAI 值介于 0.184~0.244, 进一步说明 *DXR* 基因的密码子偏好性较弱。裸子植物 *DXR* 的 GC 含量介于 43.75%~45.26%, GC3s 介于 33.75%~36.82%。单子叶植物 *DXR* 的 GC 含量介于 47.49%~48.03%, GC3s 介于 41.60%~43.64%。双子叶植物 *DXR* 的 GC 含量介于 43.65%~49.23%, GC3s 介于 34.11%~51.68%。整体看, 裸子植物 *DXR* 的 GC 含量和 GC3s 低于双子叶植物和单子叶植物, 且 38 种植物 *DXR* 基因均偏好以 A/T 结尾的密码子。

2.2.2 *DXR* 同义密码子相对使用度分析

RSCU 是衡量密码子偏好性的重要指标, 用 38 个物种 *DXR* 基因密码子的 RSCU 值制作热图 (图 1), 结果显示: 裸子植物与单子叶植物和双子叶植物 *DXR* 的 RSCU 值存在一定差异。整体看, 裸子植物 *DXR* 中 RSCU 值大于 1.5 的密码子较单子叶植物和双子叶植物多, 东北红豆杉 RSCU 大于 1.5 的密码子有 16 个, 丹参和夏枯草只有 6 个。说明裸子植物密码子使用偏好性比其他植物强, 这与基于 ENC 分析的结果相同。在所分析的绝大部分物种的 *DXR* 基因密码子中, GCT、AGA、GAT、ATT、

CTT、TTG、CCT、TCT、GTT 的 RSCU 值都大于 1, 说明这几个密码子出现频率较高, 属于共同偏好使用的密码子, 且这些密码子偏向以 A/T 结尾。除此之外, 通过对马尾松、火炬松、赤松 *DXR* 的密码子分析发现, AGA、TCT 的 RSCU 值均大于 2, 由此推测松科植物对这 2 个密码子的偏好性极强。

2.2.3 *DXR* 密码子成分相关分析

用 SPSS20.0 对 *DXR* 密码子成分 GC、GC3s、ENc、CAI、密码子偏爱指数 (CBI)、最优密码子使用频率 (Fop)、A3s、U3s、C3s、G3s 进行两两 Pearson 关联分析 (表 3)。结果显示: ENc 和 GC、GC3s、CAI、CBI、Fop、C3s、G3s 参数呈显著或极显著正相关, 与 U3s、A3s 参数呈极显著负相关, 说明 *DXR* 密码子偏好性越强, 其碱基中 A、T 含量越高。

2.2.4 不同物种 *DXR* 基因中性绘图分析

密码子第 1、2 位 GC 含量平均值 (GC12) 与 GC3s 相关性分析表明相关不显著, 对不同物种 *DXR* 基因进行中性分析 (图 2) 发现: GC3s 的范围为 33.75%~51.68%, GC12 的范围为 47.79%~50.74%。GC12 的范围相对较小, 绝大部分基因靠近回归线分布。GC12 与 GC3s 的线性回归系数为 0.0227, 远小于 1, 说明 *DXR* 基因密码子第 1、2 位碱基与第 3 位碱基组成差异较大。由此推断, 在进化过程中 *DXR* 基因密码子的偏好性主要受自然选择影响^[20]。

2.2.5 不同物种 *DXR* 基因 PR2-plot 分析

图 3 显示: 在 PR2-plot 分析中, A3/(A3 + T3) 或 G3/(G3 + C3) 值都偏离了 0.5, 大都分布在平面图的右下方, 表明密码子第 3 位碱基 T 的频率高于 A, G 的频率高于 C, 表明多数物种 *DXR* 密码子偏好模式很大程度上受自然选择的影响^[21], 与中性分析的结果一致。

2.2.6 不同物种 *DXR* 基因 ENc-plot 分析

ENc-plot 分析结果 (图 4) 发现: 38 个物种 *DXR* 基因位点均分布在标准曲线下方, 实际 ENc 值与理论 ENc 值存在差异, 且大部分离标准曲线较远, 说明大部分 *DXR* 基因密码子偏好性形成受自然选择的作用较大; 但其中夏枯草 *DXR* 基因位点距离标准曲线较近, 说明夏枯草 *DXR* 基因偏好性主要受突变影响。

2.3 *DXR* 基因 CDS 序列和密码子使用偏好性的系统聚类分析

采用邻接法构建 38 个物种 *DXR* 基因的 CDS 序列系统进化树 (图 5)。38 个物种聚为 2 大类, 第 1 类由裸子植物组成, 第 2 类由单子叶植物和双

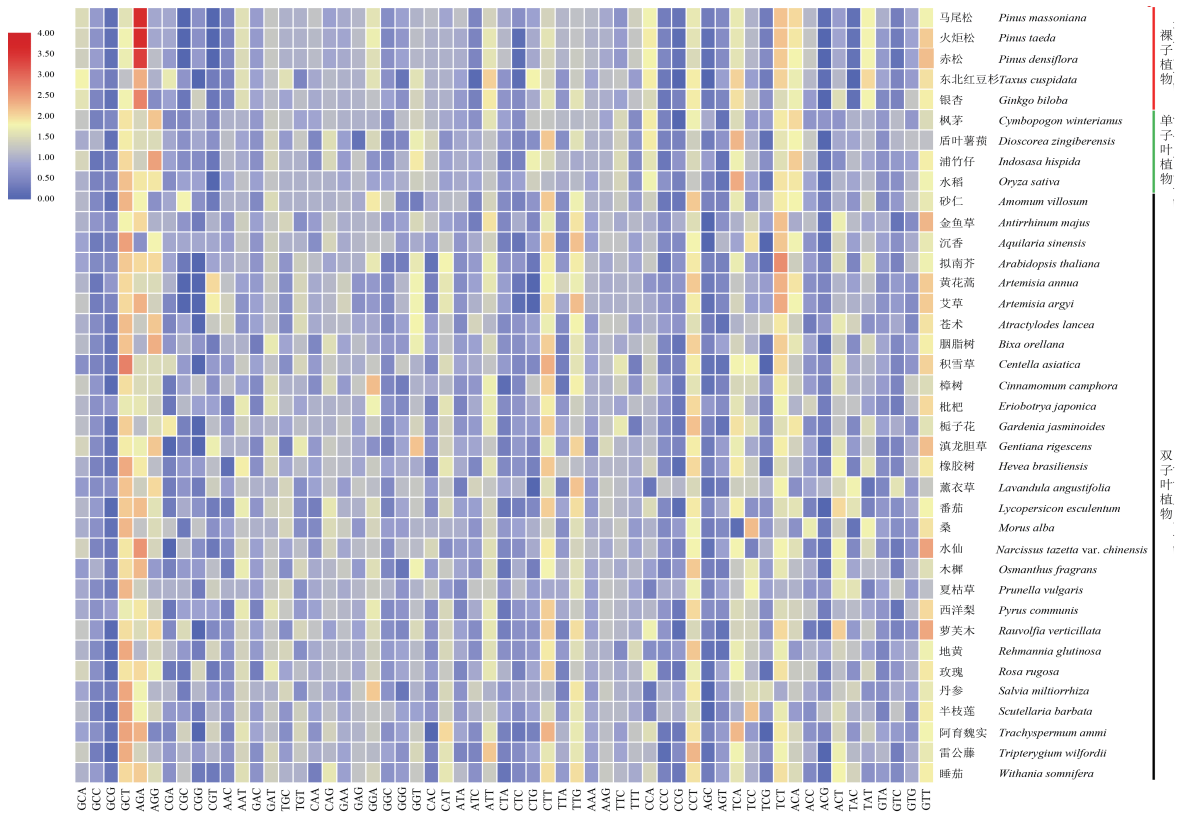


图 1 38 个物种 *DXR* 基因的 RSCU 热图

Fig. 1 Heat map of RSCU values of *DXR* gene from 38 species

表 3 *DXR* 密码子成分相关分析

Table 3 Correlation analysis of *DXR* related parameters

参数 parameter	GC	GC3s	ENc	CAI	CBI	Fop	A3s	U3s	C3s	G3s
GC	1.000									
GC3s	0.945**	1.000								
ENc	0.796**	0.763**	1.000							
CAI	0.191	0.313	0.331*	1.000						
CBI	0.587**	0.671**	0.590**	0.730**	1.000					
Fop	0.572**	0.663**	0.580**	0.756**	0.997**	1.000				
A3s	-0.583**	-0.719**	-0.595**	-0.544**	-0.659**	-0.658**	1.000			
U3s	-0.733**	-0.651**	-0.443**	0.162	-0.248	-0.236	-0.053	1.000		
C3s	0.877**	0.940**	0.776**	0.302	0.662**	0.662**	-0.625**	-0.661**	1.000	
G3s	0.791**	0.851**	0.556**	0.299	0.538**	0.524**	-0.741**	-0.419**	0.628**	1.000

注: **表示在0.01水平上显著相关, *表示在0.05水平上显著相关。

Notes: ** means significant correlations at 0.01 level, * means significant correlations at 0.01 leve.

子叶植物组成。在单子叶植物中, 3 种禾本科植物的 CDS 序列同源性较高, 聚在一个小的分支上。使用 SPSS20.0 软件对 38 个物种 *DXR* 基因的 RSCU 值进行聚类分析 (图 6), 同样将 38 个物种聚为 2 大类, 第 1 大类为裸子植物, 其中, 同属于松科

的马尾松、赤松和火炬松的 *DXR* 基因聚在一个小支上, 与银杏、东北红豆杉的 2 个小支共聚在一个大支上, 表明具有相近亲缘关系的物种在密码子使用模式上较为相似。第 2 大类中, 禾本科植物仍聚为一个类别, 同属蔷薇科的枇杷、西洋梨和玫瑰

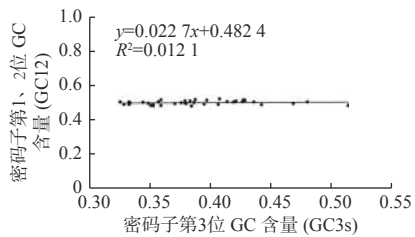


图2 不同物种DXR基因中性绘图

Fig. 2 Neutrality plot of DXR genes in different species

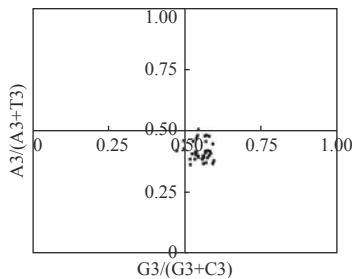
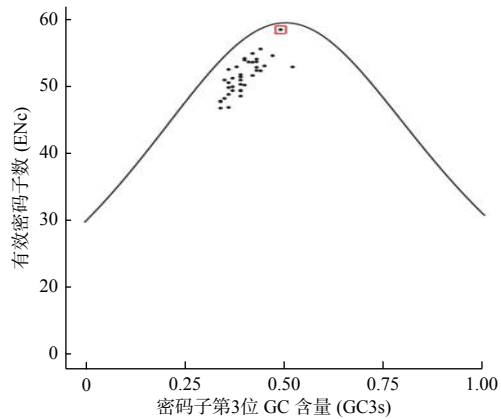


图3 不同物种DXR基因PR2-plot分析

Fig. 3 PR2-plot analysis of DXR genes in different species



注：红色方框表示夏枯草DXR基因

Note: The red box means DXR gene of *Prunella vulgaris*

图4 不同物种DXR基因ENc-GC3s关联绘图

Fig. 4 ENc-plot of DXR genes in different species

聚为一小类，菊科植物黄花蒿、艾草与属于龙胆科的滇龙胆草聚为一小类，但是也发现基于RSCU

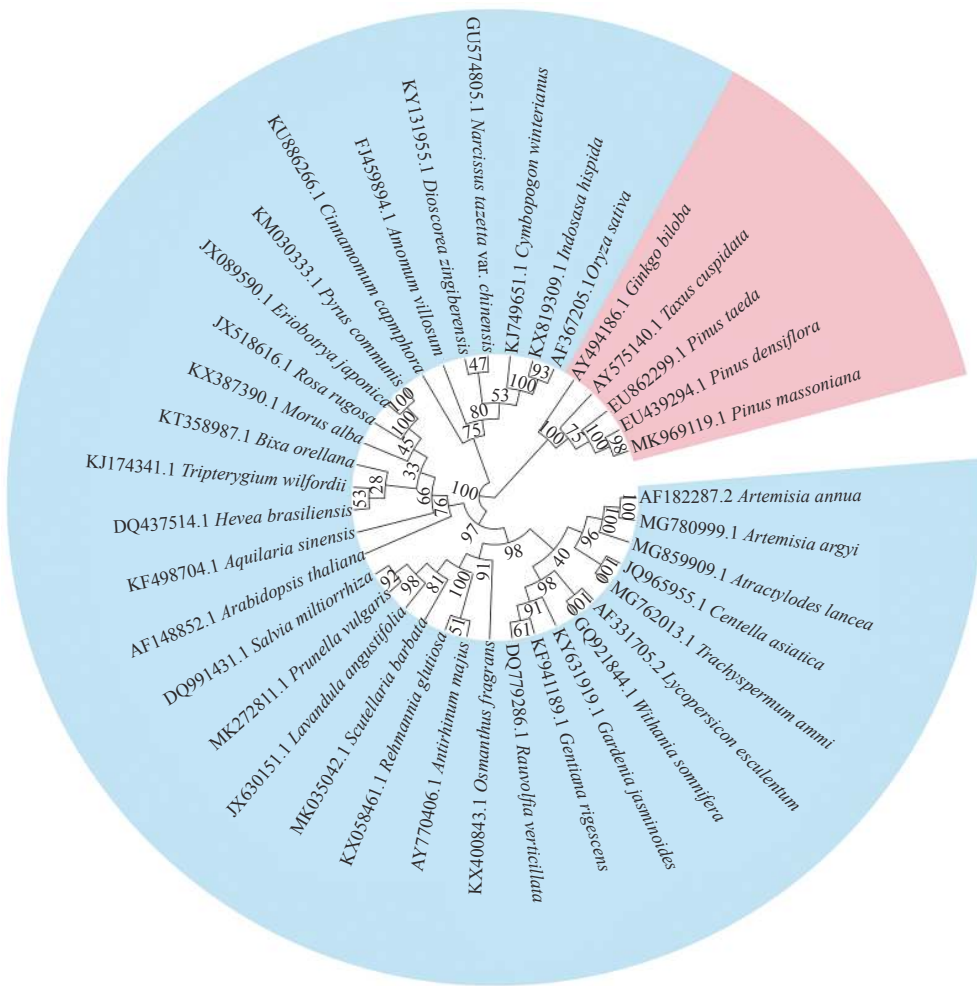
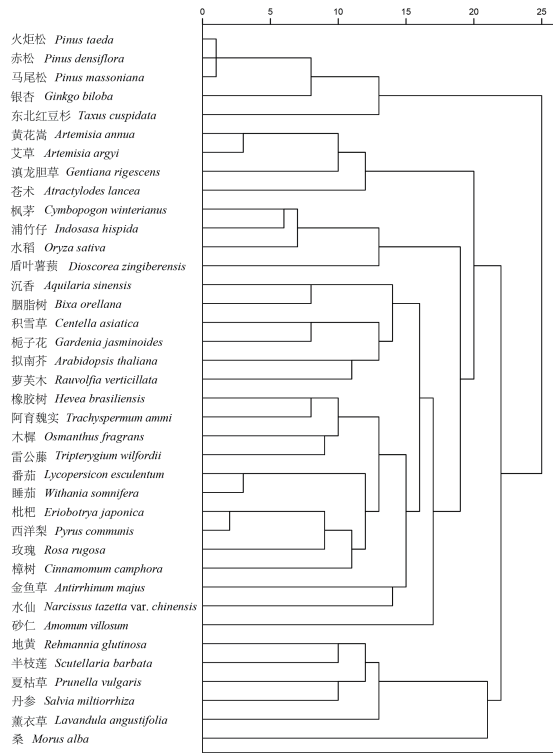


图5 DXR系统进化树

Fig. 5 evolutionary tree of DXR

图6 *DXR* 的 RSCU 值聚类分析Fig. 6 RSCU cluster analysis of *DXR*

值的聚类分析结果与 CDS 序列构建的进化树存在差异。由此可见,并非所有亲缘关系较近的物种其 *DXR* 基因的密码子使用偏好性都具有相似性,部分物种在长期进化过程中可能会形成独特的密码子使用偏好性。

2.4 马尾松 *DXR* 基因的外源宿主选择

将马尾松 *PmDXR* 密码子使用频率分别与拟南芥、烟草、酿酒酵母、大肠杆菌基因组密码子的使用频率进行比较(表4)。密码子使用频率比值是评价物种间密码子使用偏好性差异的指标之一,若比值为 0.5~2.0,表示偏好性差异较小,若比值 ≤ 0.5 、 ≥ 2.0 ,则表示偏好性差异较大^[22]。在异源表达过程中,应尽量选择密码子使用偏好性差异较小的作为受体,以提高转录、翻译效率,进而提高表达量。分析结果表明:*PmDXR*与酿酒酵母基因组有 15 个密码子的偏好性差异较大,与大肠杆菌基因组密码子偏好性差异较大有 25 个,可见,酵母更适合做 *PmDXR* 基因的真核表达系统。*PmDXR* 基因与拟南芥、烟草基因组分别有 12、9 个密码子的使用频率差异较大,说明在 *PmDXR* 基因遗传转化试验中,烟草相较拟南芥更适合作为异源表达受体。

3 讨论

在漫长的自然选择和进化进程中,各物种形成了特定的密码子使用模式,其使用偏好性对研究基因进化和功能鉴定具有重要的生物学意义。本研究用生物信息学方法对马尾松 *PmDXR* 基因密码子偏好模式进行系统分析,确定了该基因有 12 个高频率密码子,其中,AGA、TCT 具有极强的偏好性;同时发现编码 Arg 的 CGC、CGT 可能不参与 *PmDXR* 基因的编码过程。*PmDXR* 基因中密码子的 GC 和 GC3s 含量分别为 43.75% 和 34.47%,说明其密码子比较偏好以 A/T 结尾,这一结果符合马尾松总体上偏好使用第 3 位为 A/T 碱基的密码子的特征^[23]。38 个物种 *DXR* 基因的 CAI 值介于 0.184~0.244, ENc 值为 47.89~59.49,平均值为 52.59,其中,裸子植物 *DXR* 基因的 ENc 平均值为 49.04,较单子叶植物和双子叶植物低,表明裸子植物 *DXR* 基因的密码子偏好性相对双子叶和单子叶植物较高。由此可见,*DXR* 基因密码子偏好性较弱,在各物种的表达水平普遍不高,但物种间还是存在一定差异。

密码子偏好性的形成受突变压力和自然选择等因素的综合影响,对不同物种 *DXR* 基因的中性分析、ENc-plot 分析、PR2 分析结果均表明, *DXR* 基

表 4 马尾松 *PmDXR* 基因与部分模式生物基因组密码子使用偏好性比较Table 4 Comparison of codon usage preference between *PmDXR* gene in *P. massoniana* and other model organisms

氨基酸 Amino acid	密码子 Codon	马尾松 <i>DXR</i> 基因 <i>PmDXR</i>	拟南芥基因组 At	烟草基因组 Nt	酵母基因组 Sc	大肠杆菌基因组 Ec	<i>PmDXR</i> /At	<i>PmDXR</i> /Nt	<i>PmDXR</i> /Sc	<i>PmDXR</i> /Ec
Ala	GCA	37.879	17.5	23.1	16.2	20.6	<u>2.16</u>	1.64	<u>2.34</u>	1.84
	GCC	18.939	10.3	12.5	12.6	25.1	1.84	1.52	1.50	0.75
	GCG	3.788	9.0	5.8	6.2	31.7	<u>0.42</u>	0.65	0.61	<u>0.12</u>
	GCT	39.773	28.3	31.2	21.2	15.6	1.41	1.27	1.88	<u>2.55</u>
Arg	AGA	22.727	19.0	16.0	21.3	2.9	1.20	1.42	1.07	<u>7.84</u>
	AGG	5.682	11.0	12.2	9.2	1.9	0.52	<u>0.47</u>	0.62	<u>2.99</u>
	CGA	5.682	6.3	5.3	3.0	3.9	0.90	1.07	1.89	1.46
	CGC	0.000	3.8	3.9	2.6	21.0	0.00	0.00	0.00	0.00
	CGG	3.788	4.9	3.7	1.7	6.3	0.77	1.02	<u>2.23</u>	0.60
	CGT	0.000	9.0	7.5	6.4	20.3	0.00	0.00	0.00	0.00
Asn	AAC	3.788	20.9	17.9	24.8	21.4	<u>0.18</u>	<u>0.21</u>	<u>0.15</u>	<u>0.18</u>
	AAT	13.258	22.3	28.0	35.7	18.6	0.59	<u>0.47</u>	<u>0.37</u>	0.71
Asp	GAC	18.939	17.2	16.9	20.2	18.6	1.10	1.12	0.94	1.02
	GAT	30.303	36.6	36.9	37.6	32.1	0.83	0.82	0.81	0.94
Cys	TGC	7.576	7.2	7.2	4.8	6.9	1.05	1.05	1.58	1.10
	TGT	7.576	10.5	9.8	8.1	5.5	0.72	0.77	0.94	1.38
Gln	CAA	11.364	19.4	20.7	27.3	15.0	0.59	0.55	<u>0.42</u>	0.76
	CAG	7.576	15.2	15.0	12.1	29.5	<u>0.50</u>	0.51	0.63	<u>0.26</u>
Glu	GAA	30.303	34.3	36.0	45.6	38.2	0.88	0.84	0.66	0.79
	GAG	34.091	32.2	29.4	19.2	17.7	1.06	1.16	1.78	1.93
Gly	GGA	26.515	24.2	23.2	10.9	9.0	1.10	1.14	<u>2.43</u>	<u>2.95</u>
	GGC	9.470	9.2	11.2	9.8	27.9	1.03	0.85	0.97	<u>0.34</u>
	GGG	13.258	10.2	10.5	6.0	11.3	1.30	1.26	<u>2.21</u>	1.17
	GGT	17.045	22.2	22.3	23.9	24.4	0.77	0.76	0.71	0.70
His	CAC	17.045	8.7	8.7	7.8	9.8	1.96	1.96	<u>2.19</u>	1.74
	CAT	13.258	13.8	13.4	13.6	13.6	0.96	0.99	0.97	0.97
Ile	ATA	30.303	12.6	14.0	17.8	5.4	<u>2.41</u>	<u>2.16</u>	1.70	<u>5.61</u>
	ATC	11.364	18.5	13.9	17.2	24.2	0.61	0.82	0.66	<u>0.47</u>
	ATT	35.985	21.5	27.8	30.1	29.8	1.67	1.29	1.20	1.21
Leu	CTA	15.152	9.9	9.4	13.4	4.0	1.53	1.61	1.13	<u>3.79</u>
	CTC	9.470	16.1	12.3	5.4	11.0	0.59	0.77	1.75	0.86
	CTG	13.258	9.8	10.2	10.5	50.9	1.35	1.30	1.26	<u>0.26</u>
	CTT	22.727	24.1	24.0	12.3	11.7	0.94	0.95	1.85	1.94
	TTA	17.045	12.7	13.4	26.2	13.9	1.34	1.27	0.65	1.23
	TTG	18.939	20.9	22.3	27.2	14.0	0.91	0.85	0.70	1.35
Lys	AAA	35.985	30.8	32.6	41.9	33.2	1.17	1.10	0.86	1.08
	AAG	30.303	32.7	33.5	30.8	10.7	0.93	0.90	0.98	<u>2.83</u>
Met	ATG	18.939	24.5	25.0	20.9	27.0	0.77	0.76	0.91	0.70

续表 4

氨基酸 Amino acid	密码子 Codon	马尾松 <i>DXR</i> 基因 <i>PmDXR</i>	拟南芥基因组 At	烟草基因组 Nt	酵母基因组 Sc	大肠杆菌基因组 Ec	<i>PmDXR</i> /At	<i>PmDXR</i> /Nt	<i>PmDXR</i> /Sc	<i>PmDXR</i> /Ec
Phe	TTC	11.364	20.7	18.0	18.4	16.9	0.55	0.63	0.62	0.67
	TTT	24.621	21.8	25.1	26.1	23.2	1.13	0.98	0.94	1.06
Pro	CCA	28.409	16.1	19.8	18.3	8.5	1.76	1.43	1.55	<u>3.34</u>
	CCC	5.682	5.3	6.6	6.8	5.8	1.07	0.86	0.84	0.98
	CCG	3.788	8.6	5.0	5.3	21.8	<u>0.44</u>	0.76	0.71	<u>0.17</u>
	CCT	26.515	18.7	18.7	13.5	7.3	1.42	1.42	1.96	<u>3.63</u>
Ser	AGC	3.788	11.3	10.0	9.8	16.0	<u>0.34</u>	<u>0.38</u>	<u>0.39</u>	<u>0.24</u>
	AGT	5.682	14.0	13.3	14.2	9.5	<u>0.41</u>	<u>0.43</u>	<u>0.40</u>	<u>0.60</u>
	TCA	18.939	18.3	17.6	18.7	7.8	1.03	1.08	1.01	<u>2.43</u>
	TCC	11.364	11.2	10.2	14.2	8.9	1.01	1.11	0.80	1.28
	TCG	3.788	9.3	5.3	8.6	8.7	<u>0.41</u>	0.71	<u>0.44</u>	<u>0.44</u>
	TCT	24.621	25.2	20.0	23.5	8.7	0.98	1.23	1.05	<u>2.83</u>
Thr	ACA	24.621	15.7	17.4	17.8	8.2	1.57	1.42	1.38	<u>3.00</u>
	ACC	13.258	10.3	9.7	12.7	22.8	1.29	1.37	1.04	0.58
	ACG	1.894	7.7	4.5	8.0	14.8	<u>0.25</u>	<u>0.42</u>	<u>0.24</u>	<u>0.13</u>
	ACT	9.470	17.5	20.3	20.3	9.1	0.54	<u>0.47</u>	<u>0.47</u>	1.04
Trp	TGG	15.152	12.5	12.2	10.4	15.2	1.21	1.24	1.46	1.00
Tyr	TAC	1.894	13.7	13.5	14.8	12.1	<u>0.14</u>	<u>0.14</u>	<u>0.13</u>	<u>0.16</u>
	TAT	18.939	14.6	17.8	18.8	16.5	1.30	1.06	1.01	1.15
Val	GTA	20.833	9.9	11.4	11.8	11.1	<u>2.10</u>	1.83	1.77	1.88
	GTC	7.576	12.8	11.1	11.8	15.1	0.59	0.68	0.64	<u>0.50</u>
	GTG	18.939	17.4	16.7	10.8	25.5	1.09	1.13	1.75	0.74
	GTT	37.879	27.2	26.8	22.1	18.5	1.39	1.41	1.71	<u>2.05</u>
*	TAA	0.000	0.9	1.1	1.1	2.0	0.00	0.00	0.00	0.00
*	TAG	0.000	0.5	0.5	0.5	0.3	0.00	0.00	0.00	0.00
*	TGA	1.894	1.2	1.0	0.7	1.1	1.58	1.89	<u>2.71</u>	1.72

注: 下划线表示密码子使用频率差异较大, *表示终止密码子。

Notes: The underline indicates that the codon usage frequency is quite different, * indicates the stop codon.

因密码子偏好性形成的主要影响因素为选择压力。基于 RSCU 值的聚类分析与基于 CDS 的进化分析结果表明, 具有相近亲缘关系的物种其密码子使用偏好性较为相似, 但部分物种基于 RSCU 值的聚类分析与基于 CDS 序列的进化分析结果存在差异, 推测部分物种在长期的进化过程中可能形成了独特的密码子使用模式, 这在前人的研究结果中也曾出现^[24-28]。本研究中, 松科植物在 2 种聚类结果中均聚在一个大支上, 3 种禾本科植物基于 CDS 序列聚类分析与基于 RSCU 值的聚类结果也相似, 双子叶植物中同属一个科植物的种聚类分析结果具有一定相似性, 这表明物种间亲缘关系越近,

其 *DXR* 基因的密码子使用差异越小。

在异源表达过程中, 为实现外源基因的成功表达并提高其表达量, 应尽量选择密码子使用偏好性差异较小的作为受体。*DXR* 作为萜类合成途径中的关键酶基因, 了解其密码子使用特性并选择合适的异源表达受体对研究其功能具有重要指导作用。本研究通过比较 *PmDXR* 基因与拟南芥、烟草、酿酒酵母、大肠杆菌的基因组密码子使用频率的差异, 发现相比于大肠杆菌原核表达系统, 酵母真核表达系统更适合作为 *PmDXR* 基因的表达系统。在 *PmDXR* 遗传转化功能验证中, 模式植物拟南芥、烟草均可作为其遗传转化受体, 其中, 烟草更为理

想。本研究结果初步揭示了 *PmDXR* 基因密码子的使用规律, 可为后续开展基因功能验证等相关研究提供理论基础。目前, 马尾松还没有成熟的遗传转化体系, 对 *PmDXR* 基因的研究通常需要借助遗传转化体系成熟的受体, 今后可依据此研究结果为 *PmDXR* 基因选择合适的遗传转化受体, 进一步了解其在 MEP 途径中的调控机制。

4 结 论

研究表明: 马尾松 *PmDXR* 基因的密码子偏好性较弱, 偏好使用 A/T 结尾的密码子。不同物种 *DXR* 基因的密码子偏好性存在一定差异, 整体看, 裸子植物 *DXR* 基因密码子偏好性比被子植物强。基于 RSCU 值和基于 CDS 的聚类分析结果存在一定差异, 且基于 CDS 序列的进化分析结果更能反映物种亲缘关系的远近。中性分析、PR2-plot 分析、ENc-plot 分析结果均表明 *DXR* 基因密码子偏好性形成主要受选择压力的影响。酵母真核表达系统较大肠杆菌原核表达系统更适合马尾松 *PmDXR* 异源表达, 模式植物烟草较拟南芥更适合做马尾松 *PmDXR* 遗传转化受体。

参 考 文 献:

- [1] Trapp S, Croteau R. Defensive resin biosynthesis in conifers[J]. Annual Review of Plant Physiology and Plant Molecular Biology, 2001, 52(1): 689-724.
- [2] Kelkar V M, Geils B W, Becker D R, et al. How to recover more value from small pine trees: Essential oils and resins[J]. Biomass and Bioenergy, 2006, 30(4): 316-320.
- [3] 陈晓明. 马尾松产脂相关基因挖掘及表达规律研究[D]. 北京: 中国林业科学研究院, 2018.
- [4] Takahashi S, Kuzuyama T, Watanabe H, et al. A 1-deoxy-d-xylulose 5-phosphate reductoisomerase catalyzing the formation of 2-c-methyl-d-erythritol 4-phosphate in an alternative nonmevalonate pathway for terpenoid biosynthesis[J]. Proceedings of the National Academy of Sciences, 1998, 95(17): 9879-9884.
- [5] Mahmoud S S, Croteau R B. Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase[J]. Proceedings of the National Academy of Sciences, 2001, 98(15): 8915-8920.
- [6] Carretero-Paulet L, Ahumada I, Cunillera N, et al. Expression and molecular analysis of the Arabidopsis *DXR* Gene encoding 1-Deoxy-d-xylulose 5-phosphate reductoisomerase, the first committed enzyme of the 2-C-methyl-D-erythritol 4-phosphate pathway[J]. Plant Physiology, 2002, 129(4): 1581-1591.
- [7] 李凌炬, 陈安琪, 黄 凯, 等. 豆科植物 *dxr* 基因密码子偏好性分析[J]. 生物学杂志, 2020, 37 (1): 30-34.
- [8] Liu H, Lu Y, Lan B, et al. Codon usage by chloroplast gene is bias in *Hemiptelea davidii*[J]. Journal of Genetics, 2020, 99(1): 8.
- [9] Plotkin J B, Kudla G. Synonymous but not the same: the causes and consequences of codon bias[J]. Nature Reviews Genetics, 2010, 12(1): 32-42.
- [10] 李慧姬, 吉雪花, 朱冉冉, 等. 10种植物 *PSY* 基因密码子使用偏好性分析[J]. 西北农业学报, 2020, 29 (2): 276-284.
- [11] Tuller T, Waldman Y Y, Kupiec M, et al. Translation efficiency is determined by both codon bias and folding energy[J]. Proceedings of the National Academy of Sciences, 2010, 107(8): 3645-3650.
- [12] 赖瑞联, 冯 新, 陈 瑾, 等. 橄榄转录组密码子使用偏好性及其影响因素[J]. 核农学报, 2019, 33 (1): 31-38.
- [13] Chen S, Li K, Cao W, et al. Codon-resolution analysis reveals a direct and context-dependent impact of individual synonymous mutations on mRNA level[J]. Molecular Biology and Evolution, 2017, 34(11): 2944-2958.
- [14] Zhou Z, Dang Y, Zhou M, et al. Codon usage biases co-evolve with transcription termination machinery to suppress premature cleavage and polyadenylation[J]. eLife, 2018, 7: e33569.
- [15] 吴宪明, 吴松峰, 任大明, 等. 密码子偏性的分析方法及相关研究进展[J]. 遗传, 2007, 29 (4): 420-426.
- [16] Sharp P M, Li W H. The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications[J]. Nucleic Acids Research, 1987, 15(3): 1281-1295.
- [17] 严子成, 蒋瑞平, 梁浩伟, 等. 川芎咖啡酸-O-甲基转移酶基因密码子偏好性与进化分析[J]. 应用与环境生物学报, 2020, 26 (4): 1-13.
- [18] 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.
- [19] Guan D, Ma L, Khan M S, et al. Analysis of codon usage patterns in *Hirudinaria manillensis* reveals a preference for GC-ending codons caused by dominant selection constraints[J]. BMC Genomics, 2018, 19(1): 542.
- [20] 周承哲, 朱 晨, 李小桢, 等. 茶树密码子使用偏性分析方法及其研究进展[J]. 分子植物育种, 2020, 18 (5): 1480-1488.
- [21] Li G, Ren Y, Pan H, et al. Comprehensive analysis and comparison on the codon usage pattern of whole *Mycobacterium tuberculosis* coding genome from different area[J]. BioMed Research International, 2018, 2018: 1-7.
- [22] 李 平, 白云凤, 冯瑞云, 等. 籽粒苋苹果酸酶(*NAD-ME*)基因密码子偏好性分析[J]. 应用与环境生物学报, 2011, 17 (1): 12-17.
- [23] 朱沛煌, 陈 好, 朱灵芝, 等. 马尾松转录组密码子使用偏好性及其影响因素[J]. 林业科学, 2020, 56 (4): 74-81.
- [24] 李 翔, 范作义, 王井源, 等. 红松查尔酮合成酶基因 *CHS* 密码子偏好性分析[J]. 植物研究, 2020, 40 (3): 447-457.
- [25] Christianson M L. Codon usage patterns distort phylogenies from or

- of DNA sequences[J]. American Journal of Botany, 2005, 92(8): 1221-1233.
- [26] 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.
- [27] 晁岳恩, 吴政卿, 杨会民, 等. 11种植物 *psbA* 基因的密码子偏好性及聚类分析[J]. 核农学报, 2011, 25(5): 927-932.
- [28] 刘汉梅, 赵耀, 顾勇, 等. 几种植物 *waxy* 基因的密码子用法特性分析[J]. 核农学报, 2010, 24(3): 476-481.

Analysis on Codon Bias of *PmDXR* Gene in *Pinus massoniana* Lamb.

ZHU Ling-zhi, ZHU Pei-huang, LI Rong, CHEN Yu, JI Kong shu

(Key Laboratory of Forest Genetics & Biotechnology of Ministry of Education, Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, Jiangsu, China)

Abstract: [Objective] To analyze the codon bias of *PmDXR* gene in *Pinus massoniana* in order to provide a theoretical basis for the subsequent functional verification and genetic transformation of *PmDXR* gene. [Method] The codon bias of *DXR* gene of *P. massoniana* and other 37 species was analyzed by CodonW, EMBOSS and SPSS online program. Neutral analysis, PR2-plot analysis and ENC-plot analysis were conducted to speculate the main factors for the formation of *DXR* gene codon bias. Then the codon frequency of the *PmDXR* gene of *P. massoniana* were compared with those of the 4 model organisms' genome codon usage bias. [Result] The results showed that the codon bias level of *PmDXR* was low and biased toward the synonymous codons with A or T. The *DXR* codon bias analysis of different species showed that different species had different codon preference patterns, and gymnosperms *DXR* genes had stronger codon preference than angiosperm. The CDS sequences and RSCU cluster analysis showed that the preference of *PmDXR* gene was similar to other gymnosperms. Neutral analysis, PR2-plot analysis and ENC-plot analysis results indicated that the codon bias of *DXR* gene was mainly affected by natural selection. The codon usage frequency comparison and analysis showed that eukaryotic expression system of yeast was more suitable for heterologous expression of *PmDXR* gene in the selection of heterologous transforming receptors. The model plants *Arabidopsis thaliana* and *Nicotiana tabacum* can be used as genetic transformation receptors for *PmDXR* gene, but *N. tabacum* is the most ideal genetic transformation receptor. [Conclusion] This study initially revealed the codon preference of the *PmDXR* gene, the results provides an important guidance for the subsequent gene function verification, and also provides a scientific basis for the study of molecular evolution of *P. massoniana*.

Keywords: *Pinus massoniana*; *PmDXR* gene; codon usage bias; heterologous expression; exogenous host

(责任编辑: 张 研)