

DOI:10.12171/j.1000-1522.20200216

旱柳扩展蛋白基因家族与基因组分化

杨锐霞 尹 鹏 刘 晓 王 艳 刘家福 徐吉臣

(北京林业大学林木育种国家工程实验室, 北京 100083)

摘要:【目的】扩展蛋白是植物中重要的基因家族之一, 在生长发育和逆境抗性过程中发挥重要的作用。本研究拟探讨旱柳中的扩展蛋白基因组成变化, 为理解旱柳 AA、BB 基因组的分化提供理论依据, 也为柳树分子设计育种提供参考。【方法】本研究基于毛果杨扩展蛋白基因氨基酸序列, 鉴定旱柳中的扩展蛋白基因; 通过生物信息学软件, 分析旱柳 AA、BB 基因组扩展蛋白基因家族的特征和变化。【结果】在异源四倍体旱柳基因组中鉴定了 65 个扩展蛋白基因, 其中 AA 基因组 32 个、BB 基因组 33 个, 共有的扩展蛋白基因 28 个, 有 3 个基因产生了重复; AA、BB 基因组中各有 3 个扩展蛋白基因缺失, 并分别有 2 个和 1 个基因结构域发生了缺失。基因结构分析表明, 部分 AA、BB 基因组的对应基因在内含子数量和内含子剪切位点等方面存在明显变化。在密码子使用选择上, AA、BB 基因组间各含有 5 个高频密码子, 分别含有 5 个和 10 个最优密码子, 部分密码子如 AGG 和 UAG 等的使用频率有很大的差别。比较蛋白理化性质, AA、BB 基因组的扩展蛋白在亲疏水性、结构稳定性方面存在不同程度的变化。进化分析显示, 两套基因组中 *EXP423* 基因经历了正选择, 其他同源基因经历了纯化选择, 但 *Ka/Ks* 值变化较大。【结论】旱柳 AA、BB 基因组扩展蛋白家族具有各自的组成、结构及表达特征, 是基因组分化的基础, 体现了扩展蛋白在物种分化和分类中的重要地位。

关键词: 旱柳; 扩展蛋白; 基因家族; 密码子偏好性; 分化分析

中图分类号: S718.46 文献标志码: A 文章编号: 1000-1522(2021)01-0037-12

引文格式: 杨锐霞, 尹鹏, 刘晓, 等. 旱柳扩展蛋白基因家族与基因组分化 [J]. 北京林业大学学报, 2021, 43(1): 37-48.
Yang Ruixia, Yin Peng, Liu Xiao, et al. Expansin gene family in association with the genome differentiation of *Salix matsudana* [J]. Journal of Beijing Forestry University, 2021, 43(1): 37-48.

Expansin gene family in association with the genome differentiation of *Salix matsudana*

Yang Ruixia Yin Peng Liu Xiao Wang Yan Liu Jiafu Xu Jichen

(National Engineering Laboratory for Tree Breeding, Beijing Forestry University, Beijing 100083, China)

Abstract: [Objective] Expansin, an important gene family in plants, plays a critical role in plant growth and development and resistance to various biotic and abiotic stresses. The purpose of this study was to investigate the variation of expansin genes in *Salix matsudana* so as to provide theoretical basis for understanding the AA and BB genome differentiation in *Salix matsudana*, and to contribute greatly for molecular design and breeding of *Salix*. [Method] According to the expansin gene sequences of *Populus trichocarpa*, the expansin family genes in *Salix matsudana* were individually identified. The characteristics and variations of the expansin genes in AA and BB genome were analyzed and evaluated by the bioinformatics softwares. [Result] A total of 65 expansin genes were identified in allotetraploid *Salix matsudana* based on the conservation of sequence and structure of the expansin genes. Both genomes shared 28 genes with 3 genes duplicated. Moreover, 3 expansin genes were missed and 2 genes were with a domain deficient in AA genome, while 3 expansin genes were missed and 1 gene was with a domain deficient in BB

收稿日期: 2020-07-13 修回日期: 2020-08-28

基金项目: 国家自然科学基金面上项目 (31870648)。

第一作者: 杨锐霞。主要研究方向: 林木基因组与分子育种。Email: 2413536685@qq.com 地址: 100083 北京市海淀区清华东路 35 号北京林业大学生物科学与技术学院。

责任作者: 徐吉臣, 博士, 教授。主要研究方向: 林木基因组与分子育种。Email: jcxu282@sina.com 地址: 同上。

本刊网址: <http://j.bjfu.edu.cn>; <http://journal.bjfu.edu.cn>

genome, respectively. Gene structure analysis showed the considerable changes in the number and cutting sites of intron between the relative gene pairs of both genomes. Besides, AA and BB expansin genes both had 5 high-frequency codons, and separately had 5 and 10 optimal codons. Some codons had significant usage bias such as AGG and UAG. Evaluation of the protein physicochemical properties showed that the AA and BB expansins changed a lot, especially in terms of hydrophilicity and structural stability. Ka/Ks calculation displayed that *EXPA23* gene had experienced positive selection during AA and BB genome differentiation, while the other genes underwent purification selection with Ka/Ks value changed widely. [Conclusion] The report here characterized the composition, structure and expression of the expansin genes in AA and BB genomes, which probably advanced the genome differentiation and further confirmed the important status of expansins in species classification.

Key words: *Salix matsudana*; expansin; gene family; codon usage bias; differentiation analysis

扩展蛋白(expansin)作为一种细胞壁松弛蛋白,可以疏松细胞壁、增加细胞的柔韧性。研究表明,扩展蛋白参与了植物许多生长发育过程,如过表达 *OsEXPA8* 基因的水稻植株,初生根的长度和侧根的数量显著增加,旗叶平均长度和宽度分别比野生型增加 17.77% 和 58.42%^[1]。 *NiEXPA5* 转基因烟草的叶长比野生型增加了 10% ~ 22%, 叶面积增加了 10% ~ 21%, 叶肉细胞增加了 20% ~ 40%^[2]。除此之外,扩展蛋白还参与了种子的萌发^[3]、气孔开放^[4]、腋生分生组织的发育^[5]、果实成熟^[6]、籽粒大小^[7]等。扩展蛋白也参与了植物的胁迫应答,如低温胁迫^[8]、镉胁迫^[9]、病毒防御^[10]等过程。

不同植物基因组进化过程中衍生出成员数量不等的扩展蛋白基因家族,依据序列差异分为 4 个亚家族(*EXPA*、*EXPB*、*EXLA* 和 *EXLB*)。目前已在烟草(*Nicotiana tabacum*)^[11]、小麦(*Triticum aestivum*)^[12]、杨树(*Populus trichocarpa*)^[13]、枣(*Ziziphus jujuba*)^[14]、甘蔗(*Saccharum officinarum*)^[15]等物种中进行了扩展蛋白基因的鉴定和生物信息学分析。研究表明,扩展蛋白具有明显的种属特征,如双子叶植物较单子叶植物 *EXPA* 亚家族扩展蛋白比例高, *EXPB* 比例较低^[16]。对 8 个物种扩展蛋白基因的高频密码子统计显示,2 个单子叶植物(水稻(*Oryza sativa*)和玉米(*Zea mays*))各有 24 个高频密码子,双子叶植物(番茄(*Solanum lycopersicum*)和巨桉(*Eucalyptus grandis*))分别有 20 个和 15 个高频密码子,其余双子叶植物高频密码子数介于 4 ~ 9 之间。并且,单子叶植物的高频密码子均以 C/G 结尾,而双子叶植物大多以 A/T 结尾^[17]。因此,扩展蛋白在物种分类及进化中具有非常重要的作用。

柳树(*Salix*)是重要的能源、造林和绿化树种,具有很好的抗逆性,如耐盐^[18]、耐旱^[19]、耐涝^[20]以及耐重金属胁迫^[21]等,在生态修复中发挥着重要的作用^[22]。柳树种植资源丰富,染色体倍性变化较大,其中灌木

柳多为二倍体,乔木柳多为异源多倍体^[23]。旱柳(*Salix matsudana*)为异源四倍体高大乔木,抗逆性好,适应性强,在中国各地广泛种植。目前,已经开展了旱柳遗传作图^[24]、转录组分析等^[25]工作,近期也完成了旱柳基因组的测序及 AA 和 BB 基因组的组装工作(文章待发表)。本文拟利用毛果杨扩展蛋白基因氨基酸序列,通过同源性和结构域分析,鉴定旱柳中的扩展蛋白基因,探讨旱柳 AA 和 BB 基因组扩展蛋白基因的变异,揭示其对旱柳基因组分化的影响,为进一步探讨旱柳扩展蛋白基因的功能奠定基础。

1 材料与方法

1.1 旱柳扩展蛋白基因的鉴定

毛果杨(*Populus trichocarpa*)扩展蛋白基因序列源于扩展蛋白研究网站 EXPANSIN CENTRAL(<http://www.personal.psu.edu/fsl/ExpCentral/index.htm>)。以毛果杨扩展蛋白基因氨基酸序列为探针, BlastP (E-value = 0.001) 比对旱柳基因组序列库(待发表), 筛选同源的扩展蛋白基因。比对杨树和柳树扩展蛋白基因连锁的基因排序, 进一步确认二者之间的同源关系, 并命名旱柳扩展蛋白基因。

1.2 扩展蛋白基因序列分析

利用在线工具 SMART(<http://www.ncbi.nlm.nih.gov/structure/cdd/wrpsb.cgi>)和 Pfam(<http://pfam.xfam.org/search>)进行扩展蛋白两个特征结构域 Pollen_allerg_1(PF01357)和 DPBB_1(PF03330)的验证。使用在线网站 SignalP v5.0(<http://www.cbs.dtu.dk/services/SignalP/>)对旱柳扩展蛋白的氨基酸序列进行信号肽预测。利用软件 DNAMAN v7.0 进行多序列比对。利用软件 CodonW v1.4.4 分析扩展蛋白基因的碱基组成和密码子偏好性, 使用频率高于 60% 或高于平均频率的 1.5 倍的同义密码子被定义为高频密码子^[26]。根据 ΔRSCU 值 ≥ 0.3 ($\Delta\text{RSCU} = 4$ 个高 ENC

基因的 RSCU 值 - 4 个低 ENC 基因的 RSCU 值 (RSCU 为相对同义密码子使用度, ENC 为有效密码子数), 且高表达基因的 RSCU 值 > 1 和低表达基因的 RSCU 值 < 1, 确定旱柳扩展蛋白基因的最优密码子^[27]。利用 ProtParam tool 蛋白质组学在线工具 (<https://web.expasy.org/protparam/>) 获取旱柳扩展蛋白的分子量、等电点、脂溶指数等相关理化性质参数。

1.3 进化树构建和基因结构分析

利用软件 MEGA v7.0 里的 MUSCLE 程序进行氨基酸序列比对, 利用最大似然法 (ML) 进行进化树的构建, “bootstrap” 对进化树进行评估, 重复值设为 1 000。通过软件 TBtools 分析基因的外显子-内含子结构。

1.4 基因进化分析

利用 ClustalW 比对工具和 KaKs_Calculator2.0 软件 (http://code.google.com/p/kaks-calculator/wiki/KaKs_Calculator) 计算 Ka/Ks (每个位点的非同义替换/同义替换) 的比值, 分析扩展蛋白基因的进化选择。

2 结果与分析

2.1 旱柳扩展蛋白基因的鉴定

利用 36 个毛果杨扩展蛋白基因氨基酸序列搜索旱柳测序基因组, 确定旱柳 AA 和 BB 基因组中的扩展蛋白基因并分别命名为 *SmA_EXP* 和 *SmB_EXP*。结果显示, 旱柳中有 33 个毛果杨的扩展蛋白同源基因, 其中 *EXPA16*、*EXPA17*、*EXPA18* 在柳树中存在基因重复, 分别以 a、b 区分。另外, 旱柳中存在 1 个新基因, 其附近基因排序与杨树第 3 号染色体上的区段一致, 但杨树缺少该基因, 命名为 *EXPA28*。

旱柳两套基因组中均有部分染色体区段发生了缺失, 导致 AA 基因组中第 2、9、19 染色体以及 BB 基因组中的第 8、17、18 染色体上对应的扩展蛋白基因 *SmA_EXP17b*、*SmA_EXP21*、*SmA_EXP16b*、*SmB_EXP44*、*SmB_EXP46*、*SmB_EXP18b* 缺失。另外, AA 基因组中的 *SmA_EXP19* 和 *SmA_EXLA1*, BB 基因组中的 *SmB_EXP11* 等发生结构域缺失。这些缺失基因或不完整的基因将不用于后续的分析。

综合上述结果, AA 基因组中含有 32 个结构完整的扩展蛋白基因, *EXPA*、*EXPB*、*EXLA*、*EXLB* 等 4 个亚家族各含 25、3、1、3 个成员; BB 基因组中含有结构完整的扩展蛋白基因 33 个, 4 个亚家族各含 25、3、2、3 个成员。所有扩展蛋白均包含 3 个结构域, 信号肽长度介于 45 ~ 111 bp 之间, 外显子长度

介于 528 ~ 861 bp 之间, 内含子长度介于 146 ~ 5 796 bp 之间。EXPA 亚家族基因多含有 3 个外显子, *EXPB* 亚家族基因多含有 4 个外显子, *EXLA*、*EXLB* 亚家族基因多含有 5 个外显子。AA、BB 基因组间扩展蛋白亚家族组成基本特征相似 (表 1)。

2.2 扩展蛋白基因碱基组成分析

序列比较显示, 同一亚家族内的基因序列一致性较高, 不同亚家族间的基因序列一致性较低, 呈现明显的亚家族特征, AA 和 BB 基因组具有相似的趋势。如旱柳 AA 基因组中, A 亚家族成员的氨基酸序列一致性均在 41.1% 以上, 核苷酸一致率高于 48.3%, A 与 B 亚家族成员间氨基酸序列的一致性则低于 38.7%, 核苷酸序列的一致性均低于 45.3%。旱柳 AA 基因组与 BB 基因组的 A 亚家族成员间, 氨基酸序列一致性均在 43.6% 以上, 核苷酸一致率高于 52.0%, 而旱柳 AA 基因组的 A 亚家族成员与 BB 基因组的 B 亚家族成员间, 氨基酸序列一致性均在 42.9% 以下, 核苷酸一致率低于 46.9% (表 2)。

AA 基因组和 BB 基因组的扩展蛋白基因 GC 含量差异不明显, 但 CT/AG 值和 AT/GC 值存在一定的差异, AA 基因组较 BB 基因组更趋向使用嘧啶类碱基 CT (表 3), 而 BB 基因组较 AA 基因组更趋向使用二氢键连接的碱基 AT, 在 DNA 复制、转录时更容易解链, 呈现一定程度的表达优势。

密码子碱基组成分析显示, AA 和 BB 基因组中密码子第三个碱基的使用选择比较吻合, 较多使用 T 而较少使用 G。对扩展蛋白基因密码子的第二和第三碱基组合选择分析显示 (表 4), AA 和 BB 两个基因组在多数组合的选择上类似, 在个别组合的选择上存在一定的差别, 如 BB 基因组较 AA 基因组更喜欢选择 *TA, 前者是后者的 1.12 倍; AA 基因组较 BB 基因组更喜欢选择 *CC 等, 前者是后者的 1.06 倍。

2.3 旱柳扩展蛋白基因结构分析

依据 65 个旱柳扩展蛋白的全长氨基酸序列构建系统发育树 (图 1), 所有的扩展蛋白均按照亚家族聚类, 并形成三大类群, 其中 A、B 亚家族成员各自成群, LA 与 LB 亚家族组合成一群, 显示 LA 与 LB 亚家族间具有更近的亲缘关系。

AA、BB 基因组对应的基因间亲缘关系近, 更多聚合在一起。个别基因存在例外, 如 *SmB_EXP49* 与 *SmA_EXP46* 更近, 与 *SmA_EXP49* 稍远, 表明两基因间相近的起源关系。类似的偏离表现也出现在几个重复基因间, 如 *SmA_EXP16a* 与 *SmB_EXP16b* 更近, 与 *SmB_EXP16a* 稍远; *SmB_EXP17a* 与 *SmB_EXP17b* 更近, 与 *SmA_EXP17a* 稍远; *SmA_EXP18a* 与 *SmA_EXP18b* 更近, 与 *SmB_EXP18a*

表 1 旱柳扩展蛋白基因信息

Tab. 1 Information of expansin genes in *Salix matsudana*

基因名称 Gene name	染色体 Chromosome	信号肽 长度 Signal peptide length/bp	外显子 长度 Length of exon/bp	内含子 长度 Length of intron/bp	外显子数 Number of exon	基因名称 Gene name	染色体 Chromosome	信号肽 长度 Signal peptide length/bp	外显子 长度 Length of exon/bp	内含子 长度 Length of intron/bp	外显子数 Number of exon
<i>SmA_EXPA1</i>	1A	75	789	1 136	3	<i>SmB_EXPA1</i>	1B	75	789	1 087	3
<i>SmA_EXPA2</i>	Contig01890	69	753	237	3	<i>SmB_EXPA2</i>	13B	69	753	562	3
<i>SmA_EXPA3</i>	10A	60	750	712	3	<i>SmB_EXPA3</i>	10B	60	750	715	3
<i>SmA_EXPA4</i>	8A	66	750	277	3	<i>SmB_EXPA4</i>	8B	片段缺失 Fragment deletion			
<i>SmA_EXPA5</i>	9A	78	792	733	3	<i>SmB_EXPA5</i>	9B	78	792	668	3
<i>SmA_EXPA6</i>	17A	60	726	218	3	<i>SmB_EXPA6</i>	17B	片段缺失 Fragment deletion			
<i>SmA_EXPA7</i>	8A	63	777	796	3	<i>SmB_EXPA7</i>	Contig01117	63	777	821	3
<i>SmA_EXPA8</i>	Contig02775	93	789	1 070	2	<i>SmB_EXPA8</i>	4B	93	819	1 045	2
<i>SmA_EXPA9</i>	17A	60	726	205	3	<i>SmB_EXPA9</i>	17B	60	528	397	4
<i>SmA_EXPA10</i>	16A	60	744	225	3	<i>SmB_EXPA10</i>	16B	60	816	936	4
<i>SmA_EXPA11</i>	13A	60	729	469	3	<i>SmB_EXPA11</i>	13B	结构域缺失 Domain deletion			
<i>SmA_EXPA12</i>	19A	69	753	437	3	<i>SmB_EXPA12</i>	Contig02348	69	753	285	3
<i>SmA_EXPA13</i>	4A	45	741	321	4	<i>SmB_EXPA13</i>	4B	72	756	286	3
<i>SmA_EXPA15</i>	10A	63	777	497	3	<i>SmB_EXPA15</i>	10B	63	777	578	3
<i>SmA_EXPA16a</i>	6A	69	783	751	3	<i>SmB_EXPA16a</i>	6B	69	783	730	3
<i>SmA_EXPA16b</i>	19A	片段缺失 Fragment deletion				<i>SmB_EXPA16b</i>	19B	69	750	758	4
<i>SmA_EXPA17a</i>	2A	72	771	185	3	<i>SmB_EXPA17a</i>	2B	66	771	181	3
<i>SmA_EXPA17b</i>	2A	片段缺失 Fragment deletion				<i>SmB_EXPA17b</i>	2B	66	726	223	4
<i>SmA_EXPA18a</i>	5A	72	771	181	3	<i>SmB_EXPA18a</i>	5B	72	771	191	3
<i>SmA_EXPA18b</i>	18A	72	771	191	3	<i>SmB_EXPA18b</i>	18B	片段缺失 Fragment deletion			
<i>SmA_EXPA19</i>	2A	结构域缺失 Domain deletion				<i>SmB_EXPA19</i>	2B	81	765	5 796	4
<i>SmA_EXPA20</i>	1A	84	765	394	3	<i>SmB_EXPA20</i>	1B	84	765	386	3
<i>SmA_EXPA21</i>	9A	片段缺失 Fragment deletion				<i>SmB_EXPA21</i>	9B	63	753	228	3
<i>SmA_EXPA22</i>	4A	63	753	208	3	<i>SmB_EXPA22</i>	4B	63	753	233	3
<i>SmA_EXPA23</i>	17A	111	843	1 067	2	<i>SmB_EXPA23</i>	17B	102	834	1 072	2
<i>SmA_EXPA24</i>	16A	87	807	366	3	<i>SmB_EXPA24</i>	16B	87	807	436	3
<i>SmA_EXPA25</i>	Contig01569	72	783	458	4	<i>SmB_EXPA25</i>	Contig01569	72	861	427	3
<i>SmA_EXPA27</i>	17A	66	774	146	2	<i>SmB_EXPA27</i>	17B	75	783	228	2
<i>SmA_EXPA28</i>	3A	63	747	224	3	<i>SmB_EXPA28</i>	3B	63	747	220	3
<i>SmA_EXPB1</i>	6A	78	828	540	4	<i>SmB_EXPB1</i>	14B	78	825	569	4
<i>SmA_EXPB2</i>	19A	60	789	1 260	4	<i>SmB_EXPB2</i>	Contig01569	75	789	1 171	4
<i>SmA_EXPB3</i>	13A	66	798	1 507	4	<i>SmB_EXPB3</i>	13B	60	789	1 107	4
<i>SmA_EXLA1</i>	Contig00810	结构域缺失 Domain deletion				<i>SmB_EXLA1</i>	4B	51	852	546	4
<i>SmA_EXLA2</i>	9A	96	825	729	5	<i>SmB_EXLA2</i>	9B	51	780	925	5
<i>SmA_EXLB1</i>	3A	72	750	533	4	<i>SmB_EXLB1</i>	3B	72	750	581	4
<i>SmA_EXLB3</i>	Contig01873	69	774	981	5	<i>SmB_EXLB3</i>	16B	69	774	672	5
<i>SmA_EXLB4</i>	3A	72	771	673	5	<i>SmB_EXLB4</i>	3B	72	741	803	6

表 2 旱柳扩展蛋白亚家族成员间氨基酸及核苷酸 (括号) 序列一致性比较

Tab. 2 Sequence identity of amino acid and nucleotides (in brackets) among the expansin genes in <i>S. matsudana</i>										%
		AA				BB				
		EXPA	EXPB	EXLA	EXLB	EXPA	EXPB	EXLA	EXLB	
AA	EXPA	41.1 ~ 99.4 (48.3 ~ 97.5)								
	EXPB	28.0 ~ 38.7 (38.5 ~ 45.3)	51.2 ~ 91.7 (49.2 ~ 86.3)							
	EXLA	21.4 ~ 32.1 (33.6 ~ 42.6)	33.9 ~ 35.1 (40.5 ~ 44.2)	— —						
	EXLB	20.8 ~ 32.7 (36.1 ~ 42.3)	31.0 ~ 39.3 (39.8 ~ 44.0)	41.7 ~ 46.4 (43.0 ~ 45.3)	49.4 ~ 81.5 (45.5 ~ 85.1)					
BB	EXPA	43.6 ~ 100.0 (52.0 ~ 99.7)	29.3 ~ 41.4 (37.5 ~ 47.1)	21.4 ~ 32.9 (36.0 ~ 40.8)	23.6 ~ 35.0 (34.3 ~ 42.3)	42.5 ~ 99.3 (51.5 ~ 97.0)				
	EXPB	32.9 ~ 42.9 (38.0 ~ 46.9)	54.3 ~ 97.1 (55.2 ~ 97.5)	35.7 ~ 38.6 (41.6 ~ 46.3)	35.7 ~ 44.3 (40.1 ~ 44.3)	30.1 ~ 41.2 (39.2 ~ 47.5)	54.9 ~ 90.2 (55.9 ~ 88.7)			
	EXLA	21.4 ~ 32.9 (35.3 ~ 41.8)	32.1 ~ 35.0 (40.8 ~ 46.1)	89.3 ~ 95.7 (87.9 ~ 96.7)	41.4 ~ 49.3 (52.4 ~ 55.2)	22.2 ~ 32.7 (36.2 ~ 42.1)	32.7 ~ 36.6 (43.1 ~ 46.3)	87.6 (86.2)		
	EXLB	24.3 ~ 35.0 (34.8 ~ 41.6)	35.7 ~ 44.3 (39.0 ~ 44.8)	43.6 ~ 49.3 (51.6 ~ 54.4)	51.4 ~ 96.8 (54.4 ~ 97.5)	23.5 ~ 34.0 (34.2 ~ 42.9)	35.9 ~ 44.4 (41.4 ~ 45.6)	41.2 ~ 48.4 (51.0 ~ 53.4)	52.9 ~ 86.9 (48.0 ~ 89.4)	

表 3 旱柳扩展蛋白密码子碱基组成

Tab. 3 Codon base composition of expansin genes in <i>S. matsudana</i>		
参数 Parameter	AA	BB
T3	0.38	0.38
C3	0.31	0.30
A3	0.29	0.30
G3	0.23	0.23
ENC	53.92	54.12
CAI	0.21	0.21
GC3	0.44	0.43
GC	0.49	0.48
CT/AG	1.33	1.29
AT/GC	1.25	1.28

稍远。很明显,基因组的分化导致了基因更多的变异。

基因内含子/外显子结构分析显示, AA、BB 基因组对应的基因间基本趋势一致, 如 *SmA_EXP420/SmB_EXP420*。部分基因则存在明显的差别, 如 *SmA_EXP43/SmB_EXP43*、*SmA_EXP49/SmB_EXP49* 虽然均含有相同的外显子数目, 但内含子剪切模式明显不同(图 1); *SmA_EXP410/SmB_EXP410* 则分别含有 3 个和 4 个外显子。

2.4 密码子偏好性分析

密码子使用频率统计分析显示, 编码旱柳扩展

表 4 旱柳扩展蛋白密码子中第二和第三位碱基组成的比例

Tab. 4 Second and third base combination proportions of expansin gene codons in <i>S. matsudana</i>			%
密码子组成 Codon combination	AA	BB	
*TT	7.81	7.87	
*TC	7.52	7.36	
*TA	2.73	3.06	
*TG	6.86	6.99	
*CT	9.59	9.38	
*CC	5.68	5.35	
*CA	7.15	7.21	
*CG	1.88	2.08	
*AT	8.69	8.82	
*AC	6.77	6.51	
*AA	4.71	5.20	
*AG	4.84	4.83	
*GT	6.47	6.44	
*GC	6.59	6.45	
*GA	6.14	5.90	
*GG	6.58	6.55	

蛋白氨基酸的同义密码子间的使用存在偏好性。如 AA 基因组中编码 Leu 的 UUA 和 UUG 的相对使用频率(relative frequency of synonymous codon, RFSC)

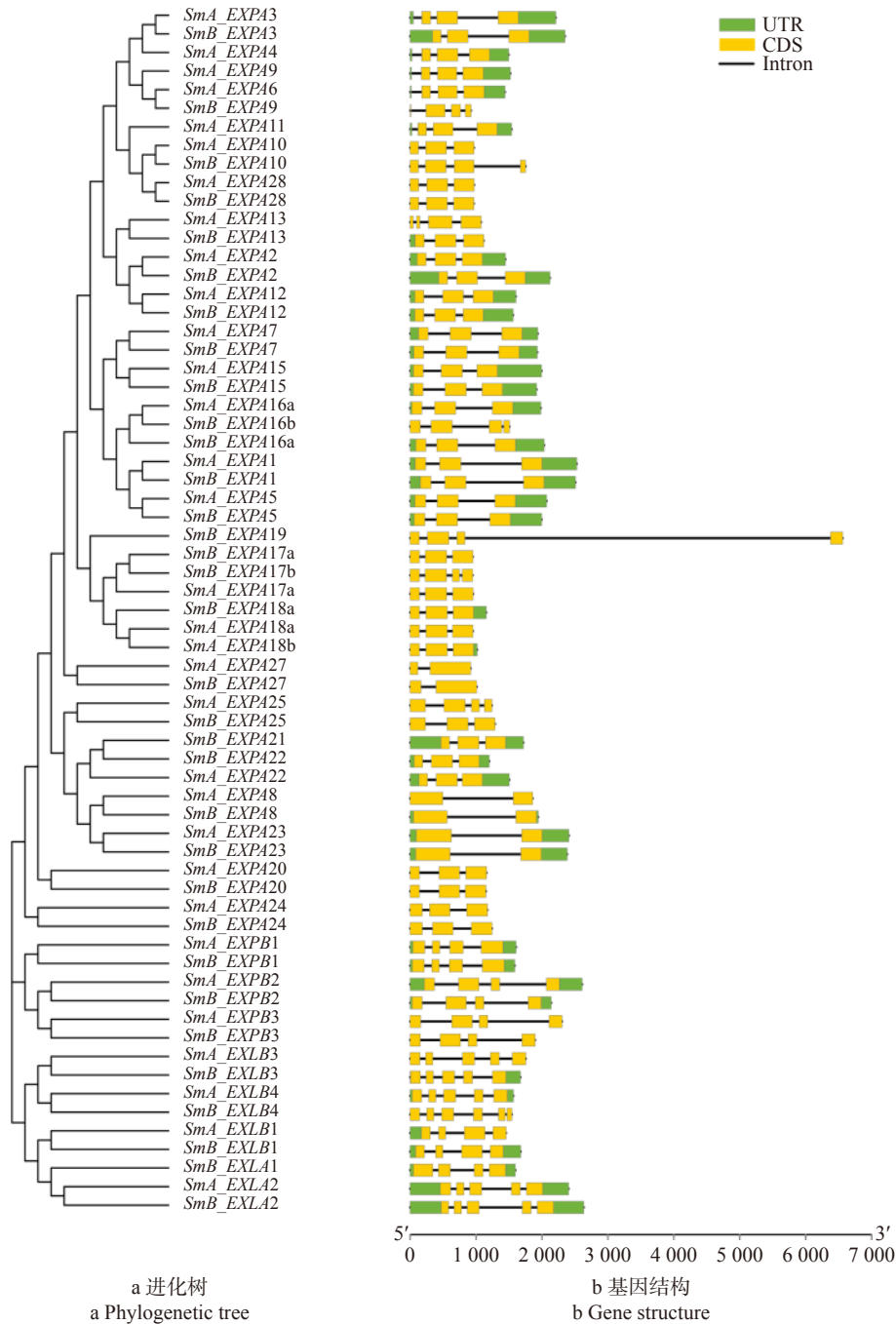


图 1 旱柳扩展蛋白进化关系及基因结构分析

Fig. 1 Phylogenetic relationship and gene structure of expansin genes in *S. matsudana*

分别为 9.96% 和 21.22%，BB 基因组中编码 Ser 的 UCU 和 UCG 的相对使用频率分别为 21.45% 和 5.40%。AA、BB 基因组间，相同密码子的相对使用频率多数表现类似，个别密码子的使用表现不同，如 AGG 在 AA、BB 基因组中的 RFSC 值分别为 29.30% 和 38.87%，UAG 在 AA、BB 基因组中的 RFSC 值分别为 46.88% 和 36.36%，其他表现差异的密码子还包括 ACG、CGU、AGA 等。

扩展蛋白所有 64 个密码子中，AA、BB 基因组均含有相同的 5 个高频密码子(表 5)，分别为 CCU (编码 Pro)、GCU(编码 Ala)、GAU(编码 Asp)、AGA

和 AGG(编码 Arg)。另外，AA、BB 基因组还分别含有 7 个和 10 个最优密码子，共享的最优密码子为 CUG、ACC、GCC、UGC 和 CGU(表 6)。AA、BB 基因组中 4 个高表达基因(ENC 值均为 58.03)的最优密码子占比分别为 8.62% 和 12.21%，而 4 个低表达基因(ENC 值分别为 48.40 和 48.09)的最优密码子占比分别为 4.80% 和 6.10%。很明显，最优密码子的数量与基因的表达量呈正相关，BB 基因组扩展蛋白具有更多的最优密码子，具有一定的表达优势。

2.5 蛋白质组成及理化特征分析

氨基酸组成分析发现，旱柳 AA、BB 基因组扩

表 5 旱柳扩展蛋白基因密码子的相对使用频率

Tab. 5 RFSC values of expansin genes in *S. matsudana* %

氨基酸 Amino acid	密码子 Codon	RFSC (AA)	RFSC (BB)	氨基酸 Amino acid	密码子 Codon	RFSC (AA)	RFSC (BB)
Phe	UUU	42.99	43.25	Tyr	UAU	50.16	52.99
	UUC	57.01	56.75		UAC	49.84	47.01
	UUA	9.96	11.21		UAA	34.38	42.42
	UUG	21.22	22.24		UAG	46.88	36.36
Leu	CUU	18.45	18.79	His	CAU	54.97	53.74
	CUC	21.77	21.03		CAC	45.03	46.26
	CUA	9.23	8.45		CAA	51.78	52.51
	CUG	19.37	18.28		CAG	48.22	47.49
Ile	AUU	49.18	48.88	Asn	AAU	54.76	56.34
	AUC	34.70	34.41		AAC	45.24	43.66
	AUA	16.12	16.71		AAA	45.79	49.52
	AUG	100.00	100.00		AAG	54.21	50.48
Met	GUU	33.14	32.46	Asp	GAU	65.77	66.23
	GUC	23.43	21.83		GAC	34.23	33.77
	GUA	11.81	14.74		GAA	51.33	54.60
	GUG	31.62	30.97		GAG	48.67	45.40
Val	UCU	23.20	21.45	Cys	UGU	47.14	46.64
	UCC	15.84	15.48		UGC	52.86	53.36
	UCA	18.10	19.46		UGA	18.75	21.21
	UCG	5.37	5.40		UGG	100.00	100.00
Ser	CCU	42.93	44.53	Trp	CGU	11.56	9.92
	CCC	18.48	17.07		CGC	9.41	10.70
	CCA	29.89	29.60		CGA	7.26	7.31
	CCG	8.70	8.80		CGG	5.11	5.22
Pro	ACU	34.79	36.38	Arg	AGU	14.85	15.77
	ACC	28.83	26.07		AGC	22.63	22.44
	ACA	29.22	27.24		AGA	37.37	51.59
	ACG	7.16	10.31		AGG	29.30	38.87
Thr	GCU	42.53	40.78	Gly	GGU	26.11	27.09
	GCC	20.75	20.53		GGC	20.64	19.87
	GCA	29.61	31.28		GGA	34.47	33.47
	GCG	7.11	7.40		GGG	18.78	19.56

注：粗体字的密码子为高频密码子。Note: codons in bold mean high frequency codons.

展蛋白的氨基酸组成趋势类似，但个别氨基酸也存在量的变化。变化最大的为 Glu，在 AA 基因组中的含量较 BB 基因组降低 11.16%(表 7)，其次为 Lys，降幅为 10.11%。AA 基因组扩展蛋白的氨基酸含量较 BB 基因组升幅最高的为 His，达到 5.86%。相对而言，与扩展蛋白功能相关的 Cys 和 Trp 的含量变化较小，在 AA 基因组中的含量较 BB 基因组分别降低 3.17% 和提高 0.04%，从而有助于扩展蛋白功能的稳定性。

蛋白理化指标分析显示，AA、BB 基因组扩展蛋白总体表现一致，但个别基因不同指标存在明显的差异(表 8)。与 BB 基因组比较，AA 基因组扩展蛋白分子量的变化幅度(AA 基因组数据/BB 基因组数据 - 1)在-10.11% 和 14.57% 之间，理论等电点的变幅在-14.89% 和 10.57% 之间，脂溶指数的变幅在-12.73% 和 3.40% 之间，总平均疏水性的变幅在-313.89% 和 325.00% 之间，蛋白不稳定指数的变幅在-20.12% 和 20.31% 之间。很明显，蛋白的亲疏水

表 6 旱柳扩展蛋白最优密码子
Tab. 6 Optimal codons of expansin
genes in *S. matsudana*

基因组 Genome	最优密码子 Optimal codon
AA	CUG ACC GCC GCA UGC CGU AGG
BB	UUC UUA CUG CCA ACC GCC UAG UGC CGU CGC

表 7 旱柳扩展蛋白基因编码蛋白氨基酸组分
Tab. 7 Amino acid composition of expansin
proteins in *S. matsudana*

氨基酸 Amino acid	AA	BB	变幅 Variation
Gly	11.80	11.30	4.46
Ser	8.61	8.32	3.50
Ala	8.39	8.46	-0.83
Leu	6.60	6.86	-3.69
Val	6.40	6.34	0.94
Asn	6.14	6.06	1.25
Thr	6.13	6.08	0.85
Phe	5.38	5.17	4.24
Arg	4.53	4.53	0.10
Pro	4.48	4.43	1.13
Ile	4.46	4.74	-5.94
Gln	4.12	4.01	2.75
Tyr	3.91	3.95	-0.95
Asp	3.63	3.64	-0.29
Cys	3.41	3.52	-3.17
Lys	3.33	3.70	-10.11
Trp	2.83	2.83	0.04
Met	2.18	2.28	-4.42
His	1.84	1.74	5.86
Glu	1.83	2.06	-11.16

注: 变幅 = AA 基因组氨基酸含量/BB 基因组氨基酸含量 - 1。Note:
variation = AA genome amino acid content/BB genome amino acid
content - 1.

性变化最大, 其次为蛋白稳定性, 这些指标的变化对于蛋白的功能效率将产生一定的影响。

2.6 进化分析

对 AA、BB 基因组中对应的扩展蛋白基因进行 Ka、Ks 计算, 其中基因对 *SmA_EXP23&SmB_EXP23* 的 Ka/Ks 值大于 1, 表明 *EXP23* 基因在 AA、BB 基因组分化时受到强烈正选择。其余同源基因对的 Ka/Ks 值均小于 1, 但不同基因对之间的 Ka/Ks 值变化较大, 分布范围介于 0.029 5 ~ 0.878 6 之间, 说明这些基因在柳树基因组进化过程中经历了不同程度的纯化选择(表 9)。

3 讨 论

旱柳为异源四倍体植物, 本研究从旱柳 AA、BB 两个基因组中鉴定了 65 个扩展蛋白基因(AA 基因组 32 个、BB 基因组 33 个), 同一亚家族内的基因同源性较高, 不同亚家族间的基因同源性较低, 呈现明显的亚家族特征。AA、BB 基因组间扩展蛋白基因家族组成、基因结构、高频密码子数量、氨基酸组成及蛋白理化特征等趋势一致。Ka/Ks 值进化分析显示, 旱柳两套基因组扩展蛋白同源基因主要经历了纯化选择, 说明在基因组分化中高度保守, 也间接说明了扩展蛋白在植物中不可或缺的生物學功能。

物种进化过程中, 经历了复杂的基因组变化, 如基因的突变、基因组的重排、基因组的加倍等^[28]。变异的结果导致基因组间亲和性降低, 杂交时染色体不能正常配对联会, 成为异源基因组^[29]。异源基因组可经染色体加倍形成异源四倍体, 可产生可育的兼具杂交优势的后代。六倍体小麦、八倍体小黑麦等皆源于此过程。因此, 解析异源基因组的基因组成和序列变异, 对于理解异源基因组的分化机制和异源多倍体的形成过程具有重要的意义。旱柳作为异源四倍体植物, 包含 AA、BB 两套不亲和的基因组。本研究通过基因组间扩展蛋白基因序列比较分析发现, 柳树基因组分化过程中导致了扩展蛋白家族基因发生了特异性的变化, 如基因的缺失、密码子的使用偏好性、碱基突变等。换言之, 扩展蛋白基因家族的变异对柳树基因组的分化产生了一定的影响, 进一步佐证了扩展蛋白在物种分化中的地位和作用。

同源分析显示, 旱柳基因组包含 37 个扩展蛋白基因, 但 AA、BB 基因组中各有部分染色体区段发生了缺失, 导致分别丢失了 3 个不同的扩展蛋白基因, 另有 2 个和 1 个扩展蛋白基因发生了结构域的缺失, 基因组中发生变异的扩展蛋白基因位点占总数的 24.32%(9/37)。如此高频度的基因家族变异导致的染色体组成结构变化, 对基因组的亲和性将产生很大的影响, 并由此推动基因组的分化^[30]。另外, 杨树与柳树同属杨柳科植物, 均由 19 条染色体组成单倍体基因组, 基因序列及基因排序一致性较高^[31]。扩展蛋白基因分析显示, 柳树缺失了 *EXP14* 和 *EXP26* 基因, 增加了 *EXP28* 基因, 预测扩展蛋白可能参与了杨柳基因组的分化过程。

Grantham 等^[32-33]指出, 每个基因组都有特定的密码子使用特征, 这种特异性也是基因组进化的压力。基于碱基兼并性的同义密码子在物种甚至基因间特定的使用模式, 也一定程度反映了进化过程

表 8 旱柳扩展蛋白的理化特征

Tab. 8 Physicochemic properties of expansin proteins in *S. matsudana*

扩展蛋白 Expansin	分子量 Molecular weight/ kDa	理论等电点 Isoelectric point (pI)	脂溶指数 Aliphatic index (AI)	总平均疏水性 Grand average of hydropathicity (GRAVY)	蛋白不稳定指数 Instability index (II)	扩展蛋白 Expansin	分子量 Molecular weight/ kDa	理论等电点 Isoelectric point (pI)	脂溶指数 Aliphatic index (AI)	总平均疏水性 Grand average of hydropathicity (GRAVY)	蛋白不稳定指数 Instability index (II)
SmA_EXPA1	28.40	9.63	67.44	-0.053	48.87	SmB_EXPA1	28.44	9.82	66.72	-0.068	43.99
SmA_EXPA2	26.59	7.53	64.84	-0.086	37.79	SmB_EXPA2	26.63	8.04	63.28	-0.123	36.91
SmA_EXPA3	26.53	9.36	67.75	-0.108	36.02	SmB_EXPA3	26.53	9.36	66.59	-0.096	35.55
SmA_EXPA4	26.55	9.50	70.08	-0.100	40.95						
SmA_EXPA5	28.60	9.62	67.57	-0.041	41.92	SmB_EXPA5	28.64	9.53	66.08	-0.031	41.81
SmA_EXPA6	26.03	8.94	66.76	-0.066	36.98						
SmA_EXPA7	27.62	9.45	69.57	-0.033	30.24	SmB_EXPA7	27.74	9.63	72.64	-0.065	31.84
SmA_EXPA8	28.23	7.57	73.74	-0.042	30.49	SmB_EXPA8	29.54	8.60	74.60	-0.066	31.97
SmA_EXPA9	26.11	9.06	67.97	-0.077	36.63	SmB_EXPA9	22.79	8.63	69.15	0.036	35.82
SmA_EXPA10	26.36	9.34	67.53	-0.170	26.51	SmB_EXPA10	29.08	9.16	77.38	-0.040	26.68
SmA_EXPA11	25.71	9.35	68.97	0.013	37.83						
SmA_EXPA12	26.57	6.86	63.28	-0.124	29.47	SmB_EXPA12	26.53	8.06	61.72	-0.144	31.33
SmA_EXPA13	27.02	9.06	69.76	-0.125	26.92	SmB_EXPA13	27.00	8.94	72.05	-0.092	31.23
SmA_EXPA15	28.15	9.33	69.88	-0.049	35.32	SmB_EXPA15	28.07	9.30	70.66	-0.080	36.13
SmA_EXPA16a	28.09	9.50	71.27	-0.047	31.41	SmB_EXPA16a	28.22	9.45	71.27	-0.040	31.28
						SmB_EXPA16b	27.18	9.34	73.25	0.002	30.90
SmA_EXPA17a	27.85	9.19	64.14	-0.055	30.07	SmB_EXPA17a	27.75	9.07	63.36	-0.048	29.57
						SmB_EXPA17b	26.10	8.96	64.07	-0.031	28.55
SmA_EXPA18a	27.65	9.17	63.71	-0.060	33.09	SmB_EXPA18a	27.68	9.17	64.10	-0.059	33.09
SmA_EXPA18b	27.68	9.17	64.10	-0.059	33.09						
						SmB_EXPA19	27.56	9.24	78.78	-0.019	32.75
SmA_EXPA20	28.16	10.13	67.20	-0.120	42.97	SmB_EXPA20	28.05	10.13	69.88	-0.114	44.59
						SmB_EXPA21	27.92	8.08	76.84	-0.124	39.07
SmA_EXPA22	27.57	8.37	75.68	-0.133	32.68	SmB_EXPA22	27.68	7.57	74.92	-0.152	31.42
SmA_EXPA23	30.46	9.01	80.46	-0.068	37.22	SmB_EXPA23	30.18	8.90	79.57	-0.095	38.82
SmA_EXPA24	28.97	9.30	60.49	-0.015	35.5	SmB_EXPA24	29.23	9.78	64.81	0.028	31.03
SmA_EXPA25	28.54	9.25	68.69	-0.354	35.64	SmB_EXPA25	31.75	9.33	70.66	-0.390	35.47
SmA_EXPA27	28.24	8.35	71.79	-0.114	27.69	SmB_EXPA27	28.44	8.54	75.08	-0.087	25.98
SmA_EXPA28	26.65	8.88	69.64	-0.103	21.72	SmB_EXPA28	26.64	9.00	68.47	-0.120	23.03
SmA_EXPB1	29.04	5.29	71.64	-0.067	42.36	SmB_EXPB1	28.44	5.53	72.96	-0.044	42.72
SmA_EXPB2	28.59	8.55	79.27	-0.139	30.37	SmB_EXPB2	28.70	8.79	80.34	-0.105	28.82
SmA_EXPB3	29.08	8.60	74.34	-0.235	32.10	SmB_EXPB3	28.70	8.35	76.30	-0.225	27.88
						SmB_EXLA1	30.80	8.86	83.07	0.083	34.02
SmA_EXLA2	29.85	8.88	74.42	-0.051	33.37	SmB_EXLA2	28.24	8.73	71.97	-0.099	30.84
SmA_EXLB1	27.49	6.42	74.10	-0.157	31.04	SmB_EXLB1	27.45	6.30	76.83	-0.127	25.80
SmA_EXLB3	27.84	5.00	79.69	-0.168	30.42	SmB_EXLB3	28.01	4.72	78.17	-0.214	38.08
SmA_EXLB4	28.14	4.70	74.30	-0.172	38.18	SmB_EXLB4	26.99	5.37	78.54	-0.093	36.23
均值 Average value	27.94	8.40	70.36	-0.106	33.78	均值 Average value	27.96	8.49	71.36	-0.098	33.49

表 9 旱柳中旁系同源扩展蛋白基因对 Ka/Ks 比值

Tab. 9 Ka/Ks ratios between the paralogous expansin gene pairs in *S. matsudana*

同源基因 Homologous gene	Ka	Ks	Ka/Ks	选择压 Selective pressure
<i>SmA_EXPA1</i> & <i>SmB_EXPA1</i>	0.010 5	0.081 4	0.128 5	纯化选择 Purify selection
<i>SmA_EXPA2</i> & <i>SmB_EXPA2</i>	0.005 6	0.190 8	0.029 5	纯化选择 Purify selection
<i>SmA_EXPA3</i> & <i>SmB_EXPA3</i>	0.013 3	0.099 8	0.133 0	纯化选择 Purify selection
<i>SmA_EXPA5</i> & <i>SmB_EXPA5</i>	0.010 7	0.127 0	0.084 6	纯化选择 Purify selection
<i>SmA_EXPA7</i> & <i>SmB_EXPA7</i>	0.016 2	0.118 6	0.137 0	纯化选择 Purify selection
<i>SmA_EXPA8</i> & <i>SmB_EXPA8</i>	0.019 7	0.090 9	0.216 9	纯化选择 Purify selection
<i>SmA_EXPA9</i> & <i>SmB_EXPA9</i>	0.965 1	1.098 5	0.878 6	纯化选择 Purify selection
<i>SmA_EXPA10</i> & <i>SmB_EXPA10</i>	0.018 3	0.116 4	0.157 5	纯化选择 Purify selection
<i>SmA_EXPA12</i> & <i>SmB_EXPA12</i>	0.010 6	0.195 6	0.054 2	纯化选择 Purify selection
<i>SmA_EXPA13</i> & <i>SmB_EXPA13</i>	0.023 2	0.129 1	0.180 1	纯化选择 Purify selection
<i>SmA_EXPA15</i> & <i>SmB_EXPA15</i>	0.013 0	0.084 1	0.154 2	纯化选择 Purify selection
<i>SmA_EXPA16a</i> & <i>SmB_EXPA16a</i>	0.012 5	0.111 0	0.112 3	纯化选择 Purify selection
<i>SmA_EXPA17a</i> & <i>SmB_EXPA17a</i>	0.017 9	0.087 9	0.203 7	纯化选择 Purify selection
<i>SmA_EXPA18a</i> & <i>SmB_EXPA18a</i>	0.003 8	0.083 1	0.046 2	纯化选择 Purify selection
<i>SmA_EXPA20</i> & <i>SmB_EXPA20</i>	0.012 3	0.077 1	0.160 0	纯化选择 Purify selection
<i>SmA_EXPA22</i> & <i>SmB_EXPA22</i>	0.012 9	0.120 2	0.107 2	纯化选择 Purify selection
<i>SmA_EXPA23</i> & <i>SmB_EXPA23</i>	0.084 3	0.073 8	1.142 0	正选择 Purify selection
<i>SmA_EXPA24</i> & <i>SmB_EXPA24</i>	0.051 3	0.156 5	0.327 8	纯化选择 Purify selection
<i>SmA_EXPA25</i> & <i>SmB_EXPA25</i>	0.151 8	0.183 9	0.825 5	纯化选择 Purify selection
<i>SmA_EXPA27</i> & <i>SmB_EXPA27</i>	0.012 3	0.028 7	0.427 9	纯化选择 Purify selection
<i>SmA_EXPA28</i> & <i>SmB_EXPA28</i>	0.020 6	0.039 2	0.523 7	纯化选择 Purify selection
<i>SmA_EXPB1</i> & <i>SmB_EXPB1</i>	0.015 3	0.157 3	0.097 3	纯化选择 Purify selection
<i>SmA_EXPB2</i> & <i>SmB_EXPB2</i>	0.027 7	0.065 9	0.420 5	纯化选择 Purify selection
<i>SmA_EXPB3</i> & <i>SmB_EXPB3</i>	0.025 6	0.070 3	0.363 6	纯化选择 Purify selection
<i>SmA_EXLA2</i> & <i>SmB_EXLA2</i>	0.017 4	0.106 4	0.163 4	纯化选择 Purify selection
<i>SmA_EXLB1</i> & <i>SmB_EXLB1</i>	0.018 0	0.065 4	0.275 4	纯化选择 Purify selection
<i>SmA_EXLB3</i> & <i>SmB_EXLB3</i>	0.028 0	0.089 9	0.312 0	纯化选择 Purify selection
<i>SmA_EXLB4</i> & <i>SmB_EXLB4</i>	0.023 1	0.089 7	0.257 4	纯化选择 Purify selection

注: Ka/Ks > 1、= 1 和 < 1 分别表示基因受正选择、中性选择和纯化选择压力。Notes: Ka/Ks > 1, = 1 and < 1 indicate that the genes are subject to positive selection, neutral selection and purify selection pressure, respectively.

中的突变偏好压力、自然选择和随机遗传漂变之间的平衡^[34-36]。在旱柳 AA 和 BB 基因组中, 部分扩展蛋白基因密码子使用频率差别较大, 如 AA 基因组使用终止密码子 UAA 和 UAG 的频率分别比 BB 基因组低 8.04% 和高 10.52%, 不同终止密码子对于蛋白质的翻译终止效率存在一定差异。Korkmaz 等^[37]在细菌中发现, 两个释放因子 RF1 和 RF2 参与了蛋白翻译终止过程, 但终止密码子 UAG 只识别 RF1, UGA 只识别 RF2, 但 UAA 却能被 RF1 和 RF2 识别。因此, UAG 在低表达基因中的使用频率较高, 而高表达基因主要以 UAA 结尾。旱柳两个基因组中终止密码子频率的差异, 也从一定程度上影响着各

自基因表达的效率。

最优密码子在蛋白质翻译延伸和蛋白质合成动力学中起重要作用^[38]。有研究表明, 最优密码子比非最优密码子更快、更准确地将核糖体解码, 后者被认为可延缓翻译延伸^[39]。因此, 最优密码子的数量和分布与基因蛋白的表达速度有密切的关系。旱柳 AA 和 BB 基因组分别含有 7 个和 10 个最优密码子, 两者之间共有 5 个最优密码子。很明显, BB 基因组的扩展蛋白表达更有优势。BB 基因组的这种表达优势在其碱基组成上也有体现。对扩展蛋白基因密码子的第二和第三碱基组合选择分析显示, BB 基因组较 AA 基因组更喜欢选择*TA, AT 为二

氢键连接,较三氢键连接的GC更容易断裂,因此BB基因组AT含量高有助于DNA解链,促进BB扩展蛋白基因的转录过程。此外,氨基酸组成分析也显示,BB基因组中Cys含量较AA基因组提高3.17%,而Cys是扩展蛋白功能实施的核心氨基酸,对于植物的逆境抗性等过程有举足轻重的作用,未来需要对BB基因组的基因进行更多的关注。

另外,基因进化过程中的非同义密码子替换(Ka)可造成氨基酸变化,导致蛋白质的构象和功能变化,从而响应自然选择效应。而同义替换(Ks)没有改变蛋白质的组成,与自然选择效应无关。因此,Ka/Ks值是基因进化评价的一个重要指标,Ka/Ks值增加是物种适应环境及物种分化的压力或动力^[40]。对旱柳扩展蛋白同源基因Ka/Ks计算分析显示,AA、BB基因组主要经历了纯化选择,但Ka/Ks值变化较大,如*Sma_EXPA2&SmB_EXPA2*间的Ka/Ks值为0.029 5,*Sma_EXPA9&SmB_EXPA9*间的Ka/Ks值为0.878 6,说明不同基因经历的选择压力不同。甚至,*Sma_EXPA23&SmB_EXPA23*的Ka/Ks值达到1.142 0,说明*EXPA23*基因在基因组分化中受到强烈的正选择,未来需要更多的关注,以探讨其对于基因组分化的作用。

参 考 文 献

- [1] Ma N, Wang Y, Qiu S, et al. Overexpression of *OsEXP48*, a root-specific gene, improves rice growth and root system architecture by facilitating cell extension[J/OL]. PLoS One, 2013, 8(10): e75997 (2013-10-04) [2019-03-13]. <https://doi.org/10.1371/journal.pone.0075997.eCollection 2013>.
- [2] Kuluev B, Safiullina M, Knyazev A, et al. Effect of ectopic expression of *NeEXP45* gene on cell size and growth of organs of transgenic tobacco plants[J]. Russian Journal of Developmental Biology, 2013, 44: 28-34.
- [3] Yan A, Wu M, Yan L, et al. *AtEXP2* is involved in seed germination and abiotic stress response in *Arabidopsis*[J/OL]. PLoS One, 2014, 9(1): e85208 (2014-01-03) [2019-04-07]. <https://doi.org/10.1371/journal.pone.0085208>.
- [4] Ouyang K X, Liu M Q, Pian R Q, et al. Methodology Isolation and analysis of α -expansin genes in the tree *Anthocephalus chinensis* (Rubiaceae)[J]. Genetics and Molecular Research, 2013, 12(2): 1061-1073.
- [5] Zenoni S, Fasoli M, Tornielli G B, et al. Overexpression of *PhEXPA1* increases cell size, modifies cell wall polymer composition and affects the timing of axillary meristem development in *Petunia hybrida*[J]. New Phytologist, 2011, 191(3): 662-677.
- [6] Perini M A, Sin I N, Villarreal N M, et al. Overexpression of the carbohydrate binding module from, *Solanum lycopersicum*, expansin 1 (*Sl-EXP1*) modifies tomato fruit firmness and Botrytis cinerea susceptibility[J]. Plant Physiology and Biochemistry, 2017, 113: 122-132.
- [7] Castillo F M, Canales J, Claude A, et al. Expansin genes expression in growing ovaries and grains of sunflower are tissue-specific and associate with final grain weight[J/OL]. BMC Plant Biology, 2018, 18(1): 327 (2018-12-10) [2019-01-14]. <https://doi.org/10.1186/s12870-018-1535-7>.
- [8] Feng X, Xu Y, Peng L, et al. *TaEXPB7-B*, a β -expansin gene involved in low-temperature stress and abscisic acid responses, promotes growth and cold resistance in *Arabidopsis thaliana* [J/OL]. Journal of Plant Physiology, 2019, 240: 153004 (2019-06-25) [2019-09-07]. <https://doi.org/10.1016/j.jplph.2019.153004>.
- [9] Ren Y Q, Chen Y H, An J, et al. Wheat expansin gene, *TaEXPA2*, is involved in conferring plant tolerance to Cd toxicity[J]. Plant Science, 2018, 270: 245-256.
- [10] Chen L J, Zou W S, Wu G, et al. Tobacco alpha-expansin *EXPA4* plays a role in *Nicotiana benthamiana* defence against tobacco mosaic virus[J]. Planta, 2018, 247(2): 355-368.
- [11] Ding A, Marowa P, Kong Y. Genome-wide identification of the expansin gene family in tobacco (*Nicotiana tabacum*)[J]. Molecular Genetics & Genomics, 2016, 291(5): 1891-1907.
- [12] Han Z S, Liu Y L, Deng X, et al. Genome-wide identification and expression analysis of expansin gene family in common wheat (*Triticum aestivum* L.)[J/OL]. BMC Genomics, 2019, 20(1): 101 (2019-02-01) [2019-11-25]. <https://doi.org/10.1186/s12864-019-5455-1>.
- [13] 李昊阳, 施杨, 丁亚娜. 杨树扩展蛋白基因家族的生物信息学分析[J]. 北京林业大学学报, 2014, 36(2): 59-67.
- [14] Li H Y, Shi Y, Ding Y N. Bioinformatics analysis of expansin gene family in poplar genome[J]. Journal of Beijing Forestry University, 2014, 36(2): 59-67.
- [15] Hou L, Zhang Z Y, Dou S H, et al. Genome-wide identification, characterization, and expression analysis of the expansin gene family in Chinese jujube (*Ziziphus jujuba* Mill.)[J]. Planta, 2019, 249(3): 815-829.
- [16] Santiago T R, Pereira V M, de Souza W R, et al. Genome-wide identification, characterization and expression profile analysis of expansins gene family in sugarcane (*Saccharum* spp.)[J/OL]. PLoS One, 2018, 13(4): e0196140 (2018-04-17) [2019-05-17]. <https://doi.org/10.1371/journal.pone.0196140.eCollection2018>.
- [17] Li N, Pu Y, Gong Y, et al. Genomic location and expression analysis of expansin gene family reveals the evolutionary and functional significance in *Triticum aestivum*[J]. Genes & Genomics, 2016, 38(5): 1021-1030.
- [18] Zhang H, Li J, Wang R X, et al. Comparative analysis of expansin gene codon usage patterns among eight plant species[J]. Journal of Biomolecular Structure and Dynamics, 2019, 37(4): 910-917.
- [19] Quiñones Martorello A S, Fernández M E, Monterubbianesi M G, et al. Effect of combined stress (salinity + hypoxia) and auxin rooting hormone addition on morphology and growth traits in six *Salix* spp. clones[J]. New Forests, 2020, 51: 61-80.
- [20] 耿云红. 干旱胁迫对绿化木本植物抗逆性研究[J]. 山东农业大学学报(自然科学版), 2019, 50(1): 12-18.

- Geng Y H. Study on resistance of woody plants under drought stress[J]. Journal of Shandong Agricultural University (Natural Science), 2019, 50(1): 12–18.
- [20] 刘春风. 淹水对 15 个树种苗木生长和形态特征的影响 [D]. 南京: 南京林业大学, 2009.
- Liu C F. Effects of artificial flooding on the growth and morphological characteristics of 15 trees species seedlings[D]. Nanjing: Nanjing Forestry University, 2009.
- [21] Shu Y, Li K L, Song J F, et al. Single and competitive adsorption of Cd (II) and Pb (II) from aqueous solution by activated carbon prepared with *Salix matsudana* Koidz[J]. *Water Science and Technology*, 2016, 74 (12): 2751–2761.
- [22] Gullberg U. Towards making willows pilot species for coppicing production[J]. *The Forestry Chronicle*, 1993, 69(6): 721–726.
- [23] Argus G W. Infrageneric classification of *Salix* (Salicaceae) in the new world[J]. *Systematic Botany Monographs*, 1997, 52: 1–121.
- [24] Zhang J, Yuan H, Li M, et al. A high-density genetic map of tetraploid *Salix matsudana* using specific length amplified fragment sequencing (SLAF-seq)[J/OL]. PLoS ONE, 2016, 11(6): e0157777 (2016–06–21) [2019–02–12]. <https://doi.org/10.1371/journal.pone.0157777.eCollection2016>.
- [25] Zhang J, Yuan H, Yang Q, et al. The genetic architecture of growth traits in *Salix matsudana* under salt stress[J/OL]. Horticulture Research, 2017, 4: 17024 (2017–06–14) [2019–01–21]. <https://doi.org/10.1038/hortres.2017.24>. eCollection2017.
- [26] Prabha R, Singh D P, Sinha S, et al. Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes[J]. *Marine Genomics*, 2017, 32: 31–39.
- [27] Duret L, Mouchiroud D. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*[J]. *Proceedings of the National Academy of Sciences*, 1999, 96: 4482–4487.
- [28] 王兰伟. 异源多倍体物种的形成和演化: 来自菊科蓍草属的证据 [D]. 开封: 河南大学, 2011.
- Wang L W. Allotetraploid speciation and evolution: evidence from the eastern asia *Achillea* species[D]. Kaifeng: Henan University, 2011.
- [29] Tayale A, Parisod C. Natural pathways to polyploidy in plants and consequences for genome reorganization[J]. *Cytogenetic & Genome Research*, 2013, 140: 2–4.
- [30] Liu B, Wendel J F. Non-Mendelian phenomena in allopolyploid genome evolution[J]. *Current Genomics*, 2002, 3(6): 489–505.
- [31] Wei S, Yang Y, Yin T. The chromosome-scale assembly of the willow genome provides insight into Salicaceae genome evolution [J/OL]. Horticulture Research, 2020, 7: 45 (2020–04–01) [2020–04–25]. <https://www.nature.com/articles>.
- [32] Grantham R, Gautier C, Gouy M, et al. Codon catalog usage and the genome hypothesis[J]. *Nucleic Acids Research*, 1980, 8(1): 49–62.
- [33] Grantham R, Gautier C, Gouy M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type[J]. *Nucleic Acids Research*, 1980, 8(9): 1893–1912.
- [34] Lloyd A T, Sharp P M. Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*[J]. *Nucleic Acids Research*, 1992, 20(20): 5289–5295.
- [35] Grocock R J, Sharp P M. Synonymous codon usage in *Cryptosporidium parvum*: identification of two distinct trends among genes[J]. *International Journal for Parasitology*, 2001, 31(4): 402–412.
- [36] Povolotskaya I S, Kondrashov F A, Ledda A, et al. Stop codons in bacteria are not selectively equivalent[J/OL]. Biology Direct, 2012, 7: 30 (2012–09–13) [2019–06–19]. <https://link.springer.com/article/10.1186/1745-6150-7-30>.
- [37] Korkmaz G, Holm M, Wiens T, et al. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance[J]. *Journal of Biological Chemistry*, 2014, 289(44): 30334–30342.
- [38] Krisko A, Copic T, Gabaldón T, et al. Inferring gene function from evolutionary change in signatures of translation efficiency [J/OL]. Genome Biology, 2014, 15: 44 (2014–03–03) [2019–06–08]. <https://link.springer.com/article/10.1186/gb-2014-15-3-r44>.
- [39] Presnyak V, Alhusaini N, Chen Y H, et al. Codon optimality is a major determinant of mRNA stability[J]. *Cell*, 2015, 160(6): 1111–1124.
- [40] Navarro A, Barton N H. Chromosomal speciation and molecular divergence-accelerated evolution in rearranged chromosomes[J]. *Science*, 2003, 300: 321–324.

(责任编辑 赵 勃
责任编委 康向阳)