

· 研究报告 ·

## 基于转录组的阿拉善地区多枝怪柳多态 Genic-SSRs的识别与开发

张雅楠<sup>1</sup>, 黄蕾<sup>1</sup>, 李佳彬<sup>1</sup>, 张雷<sup>2,3</sup>, 党振华<sup>1\*</sup>

<sup>1</sup>内蒙古大学生态与环境学院, 蒙古高原生态与资源利用教育部重点实验室/内蒙古草地生态学重点实验室, 呼和浩特 010021

<sup>2</sup>内蒙古自治区林业科学研究院, 呼和浩特 010010; <sup>3</sup>内蒙古大青山森林生态系统定位观测研究站, 呼和浩特 010010

**摘要** 基因内部的简单重复序列(Genic-SSR)可在植物适应环境变化中发挥重要作用。通过对阿拉善5个样点多枝怪柳(*Tamarix ramosissima*)的转录组进行测序、组装和比较, 经CandiSSR软件分析, 共鉴定出代表157个基序类型的1 185个多态性Genic-SSR位点, 位于1 123个转录本中。其中, 三核苷酸重复序列(596, 50.30%)最多, 其次是二核苷酸重复序列(486, 41.01%)。定位分析表明, 分别有411、239和163个Genic-SSRs位于相应基因的CDSs、5'UTRs和3'UTRs; 78.47%的三核苷酸重复SSRs位于基因的CDSs, 94.07%的二核苷酸重复SSRs位于基因的UTRs; 在CDS中, AGC/GCT、AGG/CCT、AAG/CTT、CCG/CGG和ATC/GAT重复相对丰富, 占有Genic-SSRs的64.48%; AG/CT和AT/AT是UTRs中最丰富的重复类型, 占UTR中所有Genic-SSRs的55.22%。功能注释表明, 含有多态Genic-SSRs的基因可注释到多个与植物逆境应答相关的GO条目和KEGG通路中。在随机选取的15个多态性SSR位点中, 14个被成功扩增, 共检测到64个等位基因。遗传多态性估算表明, 它们的期望杂合度(*He*)、观测杂合度(*Ho*)和多态性信息含量(*PIC*)平均值分别为0.553、0.421和0.493, 均属于中、高多态SSR标记, 表明利用RNA-seq技术开发SSR标记可行。

**关键词** Genic-SSR, 多态性, 转录组测序, 多枝怪柳

张雅楠, 黄蕾, 李佳彬, 张雷, 党振华 (2021). 基于转录组的阿拉善地区多枝怪柳多态Genic-SSRs的识别与开发. 植物学报 56, 433–442.

评价物种的遗传多样性和识别适应过程中的遗传变异并鉴定其功能, 是适应性进化研究的2个重要方向(Lynch and Lande, 1993)。分子标记是遗传多样性分析的有效手段, 具有数量丰富、多态性高、直接以DNA的形式表现等优势, 比形态学标记、细胞学标记和化学标记的精准度更高。SSR (simple sequence repeat)又称微卫星DNA, 是由1–6个核苷酸串联重复而成的DNA序列, 广泛分布于真核生物基因组中, 是种群遗传学、目标基因定位及农作物种质资源等研究中常用的分子标记。研究表明, 位于基因内部的SSR (Genic-SSR或EST-SSR)变异对染色质组成、基因活性调节、DNA复制及重组、细胞周期调控以及错配修复等有重要影响, 被称为生物适应外界环境的分子进化装置, 在适应性进化过程中扮演重要角色(Trifonov,

1995; Li et al., 2002)。同时, Genic-SSR在遗传多样性评估、遗传作图及分子标记辅助育种研究中也同样有效(陈雨等, 2008; Wen et al., 2010)。可见, Genic-SSR能满足适应性进化研究的多方面需要, 极具开发潜力和研究价值。

多枝怪柳(*Tamarix ramosissima*)是怪柳科(Tamaricaceae)、怪柳属植物, 大多分布于中亚和西亚地区, 是中国干旱和半干旱区的重要建群种, 常见于草地、滩涂及海滨沙地, 对于干旱和盐土等环境具有极强的适应性, 是防风固沙、改造盐碱地和绿化环境的优良树种(叶春秀等, 2015)。此外, 该植物枝叶可作为治疗痘疹、感冒、咳嗽和风湿骨痛等疾病的药材, 具有一定的药用价值。近年来, 国内外学者对该物种开展了广泛研究, 涉及诸多领域, 如地理学(温月仙等,

收稿日期: 2020-11-20; 接受日期: 2021-05-27

基金项目: 国家林业和草原局林业科技发展项目(No.KJZXSA2019047)和内蒙古自然科学基金(No.2020MS03005)

\* 通讯作者。E-mail: zhdang\_1982@aliyun.com

2020)、形态解剖学(王慧等, 2020)、生理生态学(陈敏等, 2012; 李永涛等, 2017)、分子生态学(伍明江等, 2020)和药用植物化学(毕江涛等, 2013)。叶春秀等(2015)利用SSR标记技术, 发现新疆塔河流域的柽柳(*T. chinensis*)种群具有较丰富的遗传多样性。蒋志敏等(2011)利用ISSR (inter-simple sequence repeat) 技术, 揭示黄河三角洲3个柽柳种群的遗传变异主要存在于种群内。然而, 关于该物种更广泛自然种群的遗传特性及其在不同生境中的生存策略报道相对较少, 亟须开展相关调查研究。

本研究对阿拉善额济纳地区5个样点的多枝柽柳转录组进行测序、组装和比较, 系统识别和分析了该物种多态Genic-SSRs的类型、频率及其在相关基因中的位置, 并开发了14个多态性SSR分子标记, 为开展多枝柽柳的种群遗传学研究及进一步解析与Genic-SSR相关的适应性进化机制奠定基础。

## 1 材料与方法

### 1.1 植物材料

2019年9月, 在内蒙古自治区阿拉善盟额济纳旗西部(97°10'E–106°53'E, 37°24'N–42°47'N) 5个样地采集多枝柽柳幼嫩叶片, 样地间平均距离为18 km。每个样地采集20个个体, 个体间隔5–10 m。样本采集后置于液氮中速冻, 带回实验室备用。采样位置的详细地理信息见表1。

表1 采样点具体信息

Table 1 Detailed information of the sampling locations

Population	Longitude (E)	Latitude (N)	Altitude (m)	Habitats
TR1	101°00'06"	41°52'16"	939	Wetland
TR2	101°11'44"	41°57'27"	927	Sandy land
TR3	101°03'36"	42°07'02"	915	Wetland
TR4	101°12'26"	41°58'45"	924	Wetland
TR5	101°16'25"	42°01'55"	921	Wetland

### 1.2 RNA制备、质量检测、cDNA文库构建及转录组测序

每个采样点选取1个个体的材料, 在干冰环境下寄送至安诺优达基因科技有限公司(中国)进行RNA提取、质量检测及测序文库构建。RNA质量检测通过Agilent 2100 Bioanalyzer (安捷伦科技有限公司, 美国)完成,

完整性(RIN值)需大于7.0。检测合格的RNA进行测序文库构建, 最终采用Illumina HiSeq™ 4000进行测序。

### 1.3 数据过滤及组装

对测序所得原始序列进行过滤, 分别去除接头污染、低质量及模糊碱基(N)比例大于5%的Reads, 得到高质量Clean Reads。最终, 高质量Clean Reads的Q30 (错误率小于0.1%)比例达到总数据量的90%以上。使用Trinity 2.4软件分别对各样本的Clean Reads进行*de novo*组装。使用TGICL 2.1对组装序列进行聚类, 形成不可延伸的Unigene。最后, 对上述5个样本组装的Unigenes进一步拼接、去冗余和聚类, 得到非冗余的All-unigene序列。

### 1.4 多态Genic-SSRs的识别

利用 CandiSSR (<https://github.com/xiaenhua/CandiSSR>)对5个Unigenes数据集中的潜在多态Genic-SSRs进行识别和分析, 搜索目标是2–6核苷酸重复的SSR, 识别最低重复次数分别为6、5、5、4和4次, 不检测单核苷酸重复和大于6核苷酸重复的SSR。SSR位点识别后, 利用Primer 3.0进行引物批量设计。引物设计参数为: GC含量为40%–60%, 上下游引物GC含量相差小于10%, 引物间不形成发卡结构, 退火温度范围50–60°C, 扩增产物长度在100–210 bp之间。

### 1.5 多态Genic-SSRs的定位及功能注释

使用BLASTX软件将含有Genic-SSR的Unigene序列与Nr、Swiss-Prot、KEGG和COGs数据库进行比对, 并按照优先级顺序进行蛋白编码序列(CDS)的预测。对于预测出全长CDS的Unigenes, 通过起始密码子(ATG)和终止密码子(TAA、TAG和TGA)的相对位置定位Genic-SSRs在相应基因中的位置; 对于预测出非全长CDS的Unigenes, 在GenBank (<https://www.ncbi.nlm.nih.gov/>)中搜索其全长同源基因, 并根据上述方法对相应Genic-SSRs在基因中的位置进行判定。采用Trinotate 3.0.2软件对上述功能注释信息进行整合, 得到Unigenes的功能注释结果。

### 1.6 SSR多态性验证

利用植物基因组DNA提取试剂盒(DP305, 天根生化科技有限公司, 中国)对28个多枝柽柳个体叶片的基

因组DNA进行分离。用1%琼脂糖凝胶电泳和Nano Drop 2000c (基因有限公司, 中国)对所提取DNA的质量进行检测。在ABI2720中进行PCR扩增。反应体系总体积为25  $\mu\text{L}$ , 包括1  $\mu\text{L}$  50 ng· $\mu\text{L}^{-1}$ 模板DNA, 12.5  $\mu\text{L}$  Premix Taq酶(宝生物工程有限公司, 中国), 0.5  $\mu\text{L}$ 正向引物, 0.5  $\mu\text{L}$ 反向引物, 10.5  $\mu\text{L}$  ddH<sub>2</sub>O。通过温度梯度(50–60°C) PCR, 确定每对引物的最佳退火温度。PCR扩增程序如下: 94°C5分钟; 94°C30秒, 每对引物的最佳退火温度下反应30秒, 72°C30秒, 35个循环; 72°C10分钟。对于扩增成功的引物, 在上游引物的5'端添加荧光标记(6-carboxy-fluoresceine、hexachloro-6-carboxy-fluoresceine或6-carboxy-X-rhodamine), 然后用相同的PCR程序进行扩增, 扩增完成后进行毛细管电泳检测(擎科新业生物技术有限公司, 中国)。使用GeneMarker 2.6 (SoftGenetics, 宾夕法尼亚州立大学, 美国)进行基因型判读。使用GenAlEx 6.502 (<http://biology.anu.edu.au/GenAlEx>)和PopGene 32 (<http://www.ualberta.ca/~fyeh/>)分析每个SSR的等位基因数( $N_a$ )、观测杂合度( $H_o$ )和期望杂合度( $H_e$ )等遗传学参数。用PowerMarker 3.0 (Liu and Muse, 2005)测定各SSR的多态性信息含量( $PI_C$ )。查看各GenicSSR位点是否符合哈迪-温伯格平衡和连锁不平衡。

## 2 结果与讨论

### 2.1 测序输出和从头组装

我们从5个cDNA文库中共测得51.48 Mb的Clean Reads, 平均GC含量为40.49%, 平均Q30为93.06%。

从头组装后, 5个个体分别得到37 790、35 818、39 252、32 175和34 364条Unigenes。进一步组装和去冗余后, 共得到81 728条All-unigene, 平均长度约为823 bp, 平均N50为1 364 bp (表2)。

### 2.2 多态Genic-SSRs的识别及频率分析

经CandiSSR分析, 我们共检测到1 185个多态Genic-SSRs位于1 123个All-unigene序列中, 这些SSR代表157个重复单元类型。其中, 最常见的是三核苷酸重复(596, 50.30%), 其次是二核苷酸重复(486, 41.01%); 四、五和六核苷酸重复的数目相对较少。在三核苷酸重复中, AGC/GCT (94, 15.77%)最为丰富, 其次是AAG/CTT (90, 15.10%)、AGG/CCT (86, 14.43%)和AAT/ATT (80, 13.42%)。在二核苷酸重复中, AG/CT (222, 45.70%)最为丰富, 其次是AT/AT (188, 38.68%)和AC/CT (76, 15.64%)。

### 2.3 多态Genic-SSRs的定位及频率分析

在含有SSR的基因序列中, 818个Genic-SSRs对应761个基因序列, 与已知蛋白有比对结果。对位于756个Unigenes的813个Genic-SSRs进行定位分析, 结果表明, 分别有411、239和163个SSR位于相应基因的CDSs、5'UTRs和3'UTRs区。大部分(379, 78.47%)三核苷酸重复SSR位于CDSs, 而大部分(254, 94.07%)二核苷酸重复SSR位于UTRs。在CDSs中, AGC/GCT、AGG/CCT、AAG/CTT、CCG/CGG和ATC/GAT相对丰富, 占该区域分布SSR的64.48%; AG/CT在UTRs中最为丰富, AT/AT次之, 共占有该位置SSR的55.22% (图1)。

表2 测序与组装结果统计

Table 2 Statistics of sequencing and assembly results

Sample	CR (No.)	Q30 (%)	GC (%)	Ug (No.)	ML (bp)	N50 (bp)
TR1	47893646	93.12	41.95	37790	1053.09	1953
TR2	43861774	93.06	41.96	35818	1083.87	1983
TR3	46194046	93.04	42.33	39252	995.67	1812
TR4	44817976	93.17	42.87	32175	1099.48	1899
TR5	44190698	92.92	42.49	34364	1082.91	1930
All-unigene				81728	823.46	1364

CR代表Clean Reads; Q30代表错误识别率小于0.1%的碱基数目; GC、Ug和ML分别代表组装所得Unigene的GC含量、数目和平均长度; N50代表将Unigene从长到短依次累加, 碱基数为Unigene总碱基数的50%时Unigene的长度。

CR represents Clean Reads; Q30 represents the Clean Reads that had Phred Quality Scores at the level of an error probability of less than 0.1%; GC, Ug and ML represent the GC content, number and the mean length of the assembled unigenes, respectively; N50 represents the length of unigene when the base number reaches 50% of the total base number of unigenes by adding the unigenes from long to short in turn.

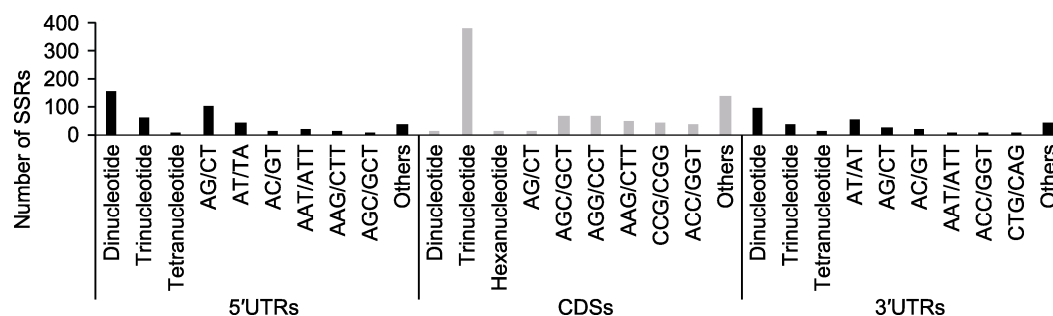


图1 基因内部简单重复序列(Genic-SSRs)的分布

x轴表示多态性Genic-SSRs的分布和motif序列类型; y轴表示Genic-SSRs的数目。

Figure 1 Distribution of the identified Genic-SSRs

The x-axis represents the distribution and motif types of polymorphic Genic-SSRs; The y-axis represents the number of Genic-SSRs.

## 2.4 含多态SSR序列的功能注释

将626个Genic-SSRs对应的582个Unigenes在GO数据库中比对后得到4 947个比对结果, 涵盖3个GO分类中的1 209个子类别。在位于基因5'UTRs的Genic-SSRs中, 最丰富的GO条目是nucleus、transcription、DNA-templated、transcription factor activity, sequence-specific DNA binding和regulation of transcription, DNA-templated。在位于基因CDSs的Genic-SSRs中, 最丰富的GO条目是nucleus, 其次是transcription、DNA-templated、regulation of transcription, DNA-templated和transcription factor activity, sequence-specific DNA binding。在位于基因3'UTRs的Genic-SSRs中, nucleus、integral component of membrane和transcription factor activity, sequence-specific DNA binding最为丰富(图2A)。

KEGG注释显示, 105个含有Genic-SSR的基因序列可匹配到73个KEGG通路中(图2B)。在位于基因5'UTRs、CDSs和3'UTRs的Genic-SSRs中, Metabolic pathways是最丰富的条目, 其次是plant hormone signal transduction、biosynthesis of secondary metabolites、plant-pathogen interaction和ubiquitin mediated proteolysis。

## 2.5 多态性Genic-SSRs的验证

在所识别的Genic-SSR中, 1 176个可以设计出引物。对随机选取的15个候选多态Genic-SSRs进行验证, 发现14个能扩增出预期条带, 并具有多态性(图3)。在这些SSR位点中, 共检测到64个等位基因, 每个基因

座位有2–8个等位基因(平均为4.5); 期望杂合度( $H_e$ ) 在0.223–0.758之间, 观测杂合度( $H_o$ ) 在0.214–0.714之间, 平均值分别为0.553和0.421; 多态性信息含量( $PIC$ )值在0.195–0.711之间, 平均值为0.493(表3)。所有标记均符合哈迪-温伯格平衡, 且未检测到连锁不平衡现象。

## 2.6 讨论

高通量测序可快速获得研究对象的海量遗传信息, 使分子标记的开发步入全新时代(Bräutigam and Gowik, 2010)。转录组测序是二代测序技术体系中的重要组成部分, 研究人员可从转录组数据中识别成百上千个SSR位点, 克服了传统SSR标记开发较为繁琐的不足, 推动了该标记在相关研究领域中的应用。基于转录组数据识别SSR已在植物研究中获得成功, 如芫荽(*Coriandrum sativum*) (Tulsani et al., 2020)、樟树(*Cinnamomum camphora*) (Li et al., 2018)和野菊(*Chrysanthemum indicum*) (Han et al., 2018)。然而, 在众多候选SSR位点中开发多态性标记仍为尚未突破的瓶颈。例如, Liu等(2021)利用11个苔草(*Carex* sp.)个体, 对96个SSR位点进行验证, 开发出42个多态性SSR标记, 成功率为44%。Weng等(2020)验证了梨锈病病原菌(*Gymnosporangium asiaticum*)转录组中的72个SSR位点, 其中22个有多态性, 成功率为31%。为提高多态性标记的开发效率, 研究人员相继开发出多款生物信息学分析工具, 如CandiSSR (Xia et al., 2015, 2019; Zhang et al., 2019)和SSRMMD (Gou et al., 2020)。利用这些工具

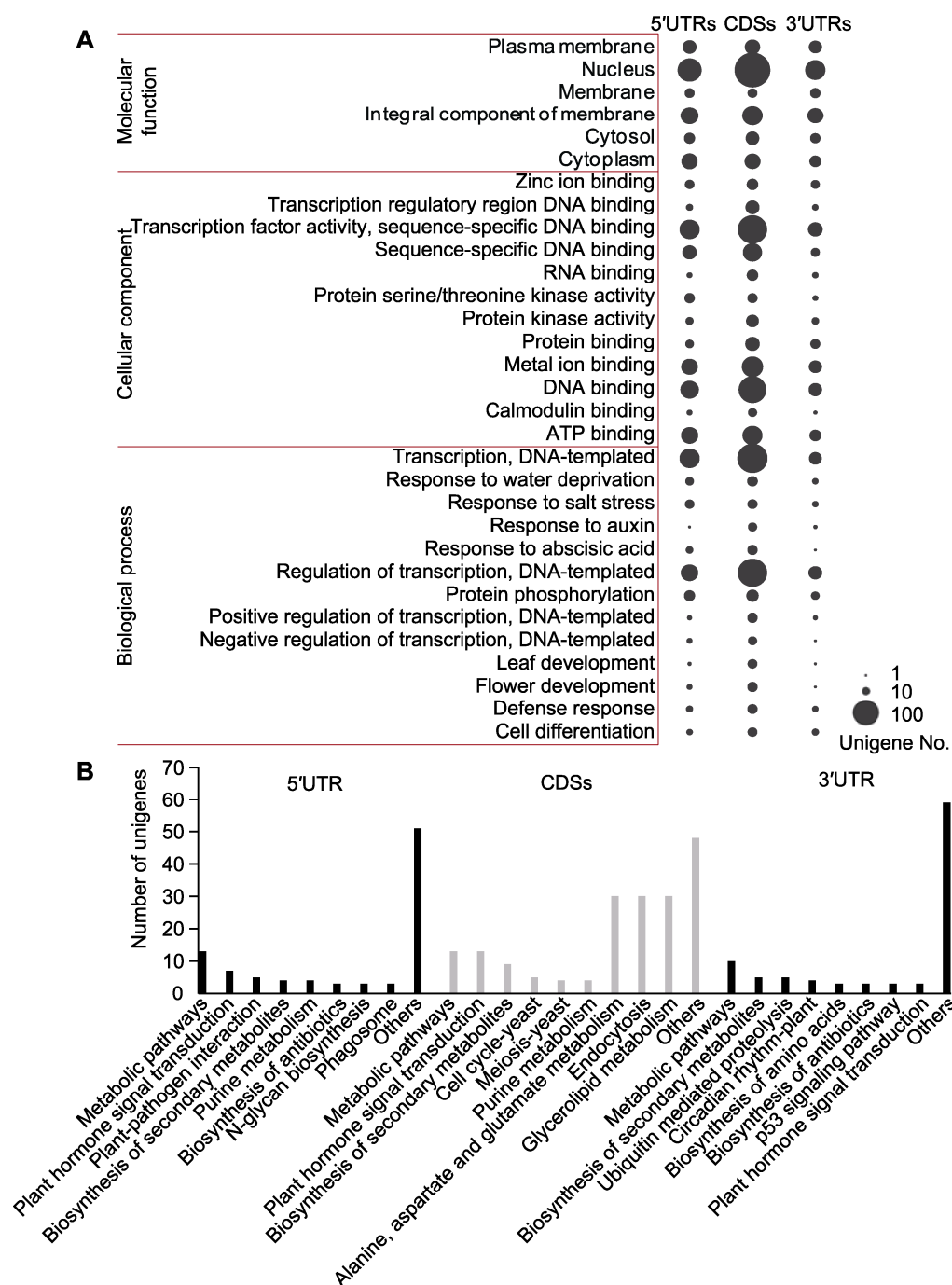


图2 含Genic-SSR序列基因的GO和KEGG富集分析

(A) GO分析, 3列气泡分别代表5'UTRs、CDSs和3'UTRs含有Genic-SSRs基因的GO富集分析结果, GO条目的基因数目由气泡大小表示; 每个基因区域显示基因数目 $\geq 10$ 的GO条目; (B) KEGG富集分析, x轴表示5'UTRs、CDSs和3'UTRs含有Genic-SSRs基因富集的代谢通路; y轴表示富集在KEGG通路中的基因数目。

Figure 2 GO and KEGG enrichment of the Genic-SSR-containing sequences

(A) GO analysis, the three columns of the bubbles represent GO enrichment analysis of transcripts that containing Genic-SSRs in the 5'UTRs, CDSs and 3'UTRs, respectively, and the number of unigenes assigned to each term is indicated by the size of each bubble; GO terms that contained unigenes more than or equal to ten in one of the gene regions are shown; (B) KEGG enrichment analysis, the x-axis indicates the enriched pathways assigned to the 5'UTRs, CDSs and 3'UTRs Genic-SSR-containing sequences; the y-axis represents the number of unigenes enriched in KEGG pathways.



**表3** 14个多态Genic-SSRs标记的基本信息及遗传学参数**Table 3** Basic information and genetic parameters of the 14 polymorphic Genic-SSRs

SSR ID	PS (5'→3')	RM	Tm (°C)	AS (bp)	Na	PIC	Ho	He
104	F: CAAGGAGGAGCTGTCGTCTG R: GCAGACACGAAGTTTGCAT	AG	59.9	123–139	8	0.516	0.286	0.558
221	F: TGAAGCAGCTGTGTTGGTGA R: TCCTCCTCGATTCCCTACTGA	AT	53.4	158–162	3	0.562	0.321	0.645
300	F: GAAGGGTTTGGGTGTTTTCAGA R: AAAACGCACCTCTCAGCAG	AT	59.5	155–161	4	0.641	0.643	0.708
392	F: CGCAACAAGCACAACATCCA R: GTTAACCGGTGCGACAACCTG	CAG	57.9	100–109	4	0.407	0.608	0.489
438	F: GGCACCGATACACAAGGACA R: CCCACAGGCTTACACCCATT	CCTT	55.4	192–204	4	0.512	0.500	0.593
441	F: GCCACCGCCACATTTATCTT R: GGAGCTTGACAGGTACAGCA	CGAG	50.8	106–122	5	0.520	0.286	0.570
563	F: TCCTGTGCAACGAAGTGAAGT R: GGGTTTAGCCATGGGTGACA	GA	51.7	176–186	6	0.593	0.536	0.666
668	F: ATGGCGATGATGGAGCAACA R: AGGGATTGGCGGAAGTGAAG	GAG	59.9	146–173	7	0.711	0.429	0.758
670	F: GAAGACACAGCACCAAGGGA R: CGTCTTCCCCAAGTCCGATC	GAG	56.7	186–201	5	0.546	0.714	0.627
947	F: TCCCCACACGTAATCCCTTC R: GCGGATGGAAGGAGAAGAGG	TC	50.8	139–149	5	0.445	0.607	0.488
1024	F: TGCGCATTTCTTGATTGCCA R: GGGTTGATGCGGCTTGATTG	TCT	52.1	168–180	5	0.476	0.464	0.516
1053	F: GTCGACACTGCAAGCATCAC R: GCACGTACGAGCACATCTCT	TG	56.7	202–204	2	0.293	0.036	0.363
1093	F: GAATAGTGGTGGCGGGAGTC R: GCTTGGCATCGTACCCCTAT	TGC	54.7	117–126	2	0.195	0.250	0.223
1114	F: GGTCTCAGTGTGGCCATAG R: GTGTTAGAGATGGCGGGCAA	TGG	54.1	156–171	4	0.487	0.214	0.543
Mean					4.5	0.493	0.421	0.553

PS: 引物序列; RM: 重复基元; Tm: 退火温度; AS: 等位基因大小; Na: 等位基因数; PIC: 多态性信息含量; Ho: 观测杂合度; He: 期望杂合度

PS: Primer sequences; RM: Repeat motif; Tm: Annealing temperature; AS: Allele size; Na: Number of allele; PIC: Polymorphism information content; Ho: Observed heterozygosity; He: Expected heterozygosity

AG/CT重复SSRs分布在基因5'UTR, 大部分AT/AT重复SSRs位于基因3'UTR, 而AGC/GCT、AGG/CCT、AAG/CTT、CCG/CGG以及ATC/GAT等三核苷酸重复SSRs则集中分布在CDS中。上述结果在一定程度上反映了本研究结果的可靠性, 并暗示这些SSR对相关基因的影响可能存在功能相似性。通过比较应用CandiSSR在其余物种中识别多态SSRs的结果, 发现多枝怪柳的多态Genic-SSR信息较为丰富,

说明其可能具有较高的遗传多样性。在云南山茶(*Camellia reticulata*)及其近缘种的转录组中鉴定出835个多态Genic-SSRs (姚秋阳, 2015); 通过比较6种异质生境下的四合木(*Tetraena mongolica*)转录组, 鉴定出811个多态性Genic-SSR位点(Dang et al., 2020)。需要指出的是, 基于此方法识别潜在多态SSR信息, 与不同研究中的样本采集距离以及种群生境间的异质性有很大关系。本研究选择地理距离相



对较远的多枝桤柳个体进行测序,以期检测到更多的变异位点,取得了较好的效果,可为同类研究提供借鉴。尽管Genic-SSR在进化上相对保守,较基因组SSR多态性偏低,但在上述验证的SSR位点中,共检测到64个等位基因( $N_a$ ),它们的 $H_e$ 和 $H_o$ 平均值分别为0.553和0.421,  $PIC$ 平均值为0.493,均属于中、高多态性SSR标记,表明内蒙古阿拉善地区多枝桤柳种群可能具有丰富的遗传多样性及遗传结构。

基因编码区和非编码区都存在大量高度可变的SSRs,其变异可能与特定基因的活性和表达相关,尤其是处于基因UTR区中的二核苷酸和三核苷酸重复SSRs,可能在植物快速适应环境中发挥重要作用(Varshney et al., 2005; Sawaya et al., 2013)。例如,Ranathunge等(2020)发现,向日葵(*Helianthus annuus*)中479个Genic-SSRs的变异与其关联基因在不同自然种群的差异表达相关;在这些SSR中,70.4%位于相关基因的非翻译区。Santi等(2003)报道了植物基因调控区存在大量的(GA) $n$ 序列,可被特定的DNA结合蛋白识别,从而参与基因转录调控。刘林等(2010)发现,灰色大角间座壳菌(*Magnaporthe grisea*)蛋白激酶基因中SSR的扩张或收缩会改变其编码产物的空间结构,进而影响其功能。在本研究识别到的多态Genic-SSRs中,分别有94.07%的二核苷酸和21.53%的三核苷酸重复SSR分布于相应基因的UTR区,表明它们对相关基因具有潜在的调节作用。桤柳以其极强的耐旱和耐盐碱能力成为我国西北干旱地区的主要植被类型。据报道,桤柳可通过根系的“提水作用”将水分从深层土壤输送到地表,水分和养分富集形成“沃岛效应”,间接为栖息地中的动植物提供适宜环境(Zhang et al., 2018)。在潮湿河岸地带,多枝桤柳凭借其有效的水分竞争能力和泌盐能力限制草本植物的生存,形成单物种优势植物群落(Natale et al., 2010);干旱时,桤柳内的脱落酸(ABA)、吲哚乙酸(IAA)及赤霉素(GA)含量均显著增加(Chen et al., 2013)。对上述逆境适应的分子机理研究表明,植物内源激素调节和次生代谢物积累在桤柳属植物抗逆性形成过程中发挥重要作用。Zheng等(2013)发现,ThWRKY4转录因子能提高桤柳对盐、渗透和脱落酸胁迫的抗性。刚毛桤柳(*T. hispida*)中乙烯应答转录因子ThCRF1可调节海藻糖和脯氨酸的生物合成,从而增强其耐盐性(Qin et al., 2017)。本研究

中,对含有多态Genic-SSR基因的功能富集分析显示,在1 123个基因中,767个获得了注释信息。有意思的是,其中大量基因富集分布在与桤柳植物适应生境和逆境应答密切相关的细胞过程和代谢通路中,GO条目如transcription、transcription factor activity、sequence-specific DNA binding、response to water deprivation和response to salt stress,KEGG代谢通路如 plant hormone signal transduction、biosynthesis of secondary metabolites和diterpenoid biosynthesis(图2),暗示这些SSR可能与植物的抗逆适应性相关。然而,Genic-SSR的多态性是否对相关基因的表达或功能有影响?它们是否能驱动植物对特定环境做出响应?这些问题仍需要进一步研究。

## 参考文献

- 毕江涛,马萍,杨志伟,关晓庆(2013). 药用植物桤柳内生真菌分离及其抑菌活性初步研究. 草业学报 22(3), 132–138.
- 陈敏,刘林德,张莉,王丽娟(2012). 黑河中游和烟台海滨中国桤柳的传粉生态学研究. 植物学报 47, 264–270.
- 陈雨,潘大建,曲延英,范芝兰,陈建西,李晨(2008). 广东高州7个普通野生稻居群遗传结构的SSR分析. 植物学通报 25, 430–436.
- 蒋志敏,陈玉霞,包颖(2011). 黄河三角洲桤柳居群的遗传结构和遗传分化. 植物分类与资源学报 33, 403–408.
- 李珊,周天华,赵桂仿,朱云国,杨晓伶,程舟(2010). 马蹄香表达序列标签资源的SSR信息分析. 中草药 41, 464–468.
- 李永涛,王霞,魏海霞,杨庆山,周健,刘德玺,刘忠杰,魏文千(2017). 盐碱生境模拟下两种桤柳的生理特性研究. 山东农业科学 49, 53–58.
- 刘林,李进斌,张悦,苏源,杨静,陈梦琪,李成云(2010). 灰色大角间座壳菌(稻瘟病菌)蛋白激酶编码基因中SSR变异及其对蛋白结构的潜在影响. 菌物学报 29, 698–706.
- 王慧,刘宁,刘金龙,姚延椿,王林(2020). 晋北干旱区盐碱地桤柳形态特征及其与土壤养分的关系. 中南林业科技大学学报 40, 37–48.
- 温月仙,甘红豪,史胜青,江泽平,吴利禄,褚建民(2020). 基于叶绿体和核基因片段序列的甘蒙桤柳谱系地理研究. 林业科学 56, 55–64.
- 伍明江,张德芹,李盼,石旭柳,刘璐(2020). 桤柳黄素对3T3-L1脂肪细胞胰岛素抵抗的影响及AMPK信号通路的作用机制. 天然产物研究与开发 32, 953–960.



- 杨成君, 王军 (2008). 人参EST资源的SSR信息分析. 植物生理学通讯 **44**, 69–73.
- 姚秋阳 (2015). 利用RNA-seq技术在云南山茶中解析重要分子通路与开发多态性EST-SSR. 博士学位论文. 昆明: 云南大学. pp. 38–96.
- 叶春秀, 姜继元, 董鹏, 庄振刚, 陈奇凌, 谢宗铭 (2015). 基于SSR标记的新疆塔里木河流域怪柳指纹图谱构建及遗传多样性分析. 分子植物育种 **13**, 2566–2571.
- Bräutigam A, Gowik U (2010). What can next generation sequencing do for you? Next generation sequencing as a valuable tool in plant research. *Plant Biol* **12**, 831–841.
- Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000). Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics* **156**, 847–854.
- Chen CX, Zhou P, Choi YA, Huang S, Gmitter FG Jr (2006). Mining and characterizing microsatellites from citrus ESTs. *Theor Appl Genet* **112**, 1248–1257.
- Chen YN, Zhou HH, Chen YP (2013). Adaptation strategies of desert riparian forest vegetation in response to drought stress. *Ecohydrology* **6**, 956–973.
- Dang ZH, Huang L, Jia YY, Lockhart PJ, Fong Y, Tian YY (2020). Identification of Genic SSRs provide a perspective for studying environmental adaptation in the endemic shrub *Tetraena mongolica*. *Genes* **11**, 322.
- Fraser LG, Harvey CF, Crowhurst RN, Silva HN (2004). EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theor Appl Genet* **108**, 1010–1016.
- Gao LF, Tang JF, Li HW, Jia JZ (2003). Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Mol Breed* **12**, 245–261.
- Gou XJ, Shi HR, Yu SF, Wang ZQ, Li CX, Liu SH, Ma J, Chen GD, Liu T, Liu YX (2020). SSRMMD: a rapid and accurate algorithm for mining SSR feature loci and candidate polymorphic SSRs based on assembled sequences. *Front Genet* **11**, 706.
- Han ZZ, Ma XY, Wei M, Zhao T, Zhan RT, Chen WW (2018). SSR marker development and intraspecific genetic divergence exploration of *Chrysanthemum indicum* based on transcriptome analysis. *BMC Genomics* **19**, 291.
- Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol Ecol* **11**, 2453–2465.
- Li YC, Korol AB, Fahima T, Nevo E (2004). Microsatellites within genes: structure, function, and evolution. *Mol Biol Evol* **21**, 991–1007.
- Li ZT, Zhong YD, Yu FX, Xu M (2018). Novel SSR marker development and genetic diversity analysis of *Cinnamomum camphora* based on transcriptome sequencing. *Plant Genet Res* **16**, 568–571.
- Liu KJ, Muse SV (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* **21**, 2128–2129.
- Liu LY, Fan XF, Tan PH, Wu JY, Zhang H, Han C, Chen C, Xun LL, Guo WE, Chang ZH, Teng K (2021). The development of SSR markers based on RNA-sequencing and its validation between and within *Carex* L. species. *BMC Plant Biol* **21**, 17.
- Lynch M, Lande R (1993). Biotic Interactions and Global Change. Sunderland MA: Sinauer Assocs. Inc. pp. 234–250.
- Morgante M, Hanafey M, Powell W (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* **30**, 194–200.
- Natale E, Zalba SM, Oggero A, Reinoso H (2010). Establishment of *Tamarix ramosissima* under different conditions of salinity and water availability: implications for its management as an invasive species. *J Arid Environ* **74**, 1399–1407.
- Qin LP, Wang LQ, Guo Y, Li Y, Ümüt H, Wang YC (2017). An ERF transcription factor from *Tamarix hispida*, ThC-RF1, can adjust osmotic potential and reactive oxygen species scavenging capability to improve salt tolerance. *Plant Sci* **265**, 154–166.
- Ranathunge C, Wheeler GL, Chimahusky ME, Perkins AD, Pramod S, Welch ME (2020). Transcribed microsatellite allele lengths are often correlated with gene expression in natural sunflower populations. *Mol Ecol* **29**, 1704–1716.
- Santi L, Wang YM, Stile MR, Berendzen K, Wanke D, Roig C, Pozzi C, Müller K, Müller J, Rohde W, Salamini F (2003). The GA octodinucleotide repeat binding factor BBR participates in the transcriptional regulation of the homeobox gene *Bkn3*. *Plant J* **34**, 813–826.
- Sawaya S, Bagshaw A, Buschiazio E, Kumar P, Chowdhury S, Black MA, Gemmell N (2013). Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PLoS One* **8**, e54710.
- Trifonov EN (1995). Segmented structure of protein sequences and early evolution of genome by combinatorial fusion of DNA elements. *J Mol Evol* **40**, 337–342.

## Identification and Development of Polymorphic Genic-SSRs in *Tamarix ramosissima* in Alxa Region Based on Transcriptome

Yanan Zhang<sup>1</sup>, Lei Huang<sup>1</sup>, Jiabin Li<sup>1</sup>, Lei Zhang<sup>2,3</sup>, Zhenhua Dang<sup>1\*</sup>

<sup>1</sup>Ministry of Education Key Laboratory of Ecology and Resource Use of the Mongolian Plateau/Inner Mongolia Key Laboratory of Grassland Ecology, School of Ecology and Environment, Inner Mongolia University, Hohhot 010021, China

<sup>2</sup>Inner Mongolia Academy of Forestry Sciences, Hohhot 010010, China; <sup>3</sup>Daqing Mountains Forest Ecosystem Research Station, Hohhot 010010, China

**Abstract** Simple sequence repeats located in gene transcribed regions (Genic-SSR) can play important roles in plant adaptation to environmental changes. In this study, the transcriptomes of *Tamarix ramosissima* from five different locations in Alxa were sequenced, assembled, and compared. By using CandiSSR software, a total of 1 185 polymorphic Genic-SSRs representing 157 motif types were identified in 1 123 transcripts. Among them, the trinucleotide repeats (596, 50.30%) were the most abundant, followed by dinucleotide repeats (486, 41.01%). Location analysis showed that 411, 239, and 163 Genic-SSRs were located in CDSs, 5'UTRs, and 3'UTRs of the relevant transcripts, respectively; 78.47% of the trinucleotide SSRs were located in CDSs, and 94.07% of the dinucleotide SSRs were located in UTRs. Among SSRs distributed in CDSs, AGC/GCT, AGG/CCT, AAG/CTT, CCG/CGG, and ATC/GAT were relatively abundant, accounting for 64.48% of all the Genic-SSRs; AG/CT and AT/AT were the most abundant repeat types in UTRs, which together account for 55.22% of all the Genic-SSRs in UTRs. Functional annotation showed that polymorphic Genic-SSRs containing genes enriched in a wide range of GO terms and KEGG pathways that highly related to stress response in *T. ramosissima*. Of the 15 randomly selected Genic-SSRs, 14 were successfully amplified by using polymerase chain reaction technology and 64 alleles were found in these SSR loci. Genetic polymorphism estimation showed that the mean of expected and observed heterozygosity ( $H_e$ ,  $H_o$ ), polymorphism information content ( $PIC$ ) of these SSRs were 0.553, 0.421, and 0.493, respectively, demonstrating the feasibility of developing SSR markers by RNA-seq.

**Key words** Genic-SSR, polymorphism, RNA-seq, *Tamarix ramosissima*

**Zhang YN, Huang L, Li JB, Zhang L, Dang ZH** (2021). Identification and development of polymorphic Genic-SSRs in *Tamarix ramosissima* in Alxa region based on transcriptome. *Chin Bull Bot* **56**, 433–442.

---

\* Author for correspondence. E-mail: zhdang\_1982@aliyun.com

(责任编辑: 白羽红)