西北植物学报，2018，38(7)：1244－1253
Acta Bot. Boreal.-Occident. Sin.

文章编号：1000-4025(2018)07-1244-10　　doi：10.7606/j.issn.1000-4025.2018.07.1244

山地虎耳草和棒腺虎耳草转录组 SSR 和 SNP 分析

李　彦[1,2]，焦秀洁[1,3]，更吉卓玛[1,2]，贾留坤[1,2]，王智华[1,2]，陈世龙[1]，高庆波[1,4*]

（1 中国科学院 西北高原生物研究所，高山植物适应与进化重点实验室，西宁 810001；2 中国科学院大学，北京 100039；3 中国科学院 西北高原生物研究所，信息与学报编辑部，西宁 810001；4 中国科学院 西北高原生物研究所，青海省作物分子育种重点实验室，西宁 810001）

摘　要：利用 Illumina HiSeq™ 2000 对山地虎耳草和棒腺虎耳草进行转录组测序，分析和比较其 SSR 和 SNP 特征。结果表明：山地虎耳草 63 763 条 Unigene 序列中含有 4 622 个 SSR，发生频率为 7.25%，有 110 种重复基元，平均每 10.00 kb 出现一个 SSR 位点；棒腺虎耳草 60 972 条 Unigene 序列中含有 4 542 个 SSR，发生频率为7.45%，有 85 种重复基元，平均每 10.40 kb 出现一个 SSR 位点，略低于山地虎耳草。山地虎耳草和棒腺虎耳草转录组序列的 SSR 优势基元均为三核苷酸重复。2 个物种的转录组 SSR 以 5～10 次的较低重复次数为主，长度主要集中于 12～30 bp，具有较高的多态性。山地虎耳草和棒腺虎耳草中分别获得 118 424 个和 112 006 个 SNP 位点，编码区的 SNP 位点分别占 30.40%和 28.59%，且在编码 SNP 中同义突变所占比例(30.27%、28.48%)远高于非同义突变(0.13%、0.11%)。比较发现，2 个物种的各项检索结果基本一致，推测与选取的组织部位、组织的发育阶段以及物种的亲缘关系有关。

关键词：山地虎耳草；棒腺虎耳草；转录组；简单重复序列(SSR)；单核苷酸多态性(SNP)

中图分类号：Q346^{+}.5；Q789　　**文献标志码**：A

Analysis of SSR and SNP in Transcriptome of *Saxifraga sinomontana* and *Saxifraga consanguinea*

LI Yan[1,2], JIAO Xiujie[1,3], GENGJI Zhuoma[1,2], JIA Liukun[1,2], WANG Zhihua[1,2], CHEN Shilong[1], GAO Qingbo[1,4*]

(1 Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China; 2 University of Chinese Academy of Sciences, Beijing 100039, China; 3 Editorial Department of Library and Journal, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China; 4 Key Laboratory of Crop Molecular Breeding of Qinghai Province, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001, China)

Abstract: Transcriptome analyses of *Saxifraga sinomontana* and *Saxifraga consanguinea* were carried out by Illumina HiSeq™ 2000, then the characteristics of SSR and SNP were generalized according to the sequences. The result indicated that: 4 622 SSR sites were identified among 63 763 Unigenes with the frequency of 7.25% in *S. sinomontana*, there were 110 repeat motifs and the density of SSRs was 1/10.00

收稿日期：2018-04-07；**修改稿收到日期**：2018-06-07

基金项目：中国科学院"西部之光"人才培养引进计划；中国科学院青年创新促进会项目(2016378)

作者简介：李　彦(1991－)，女，在读博士研究生，主要从事青藏高原植物适应与进化研究。E-mail：1772915413@qq.com

*通信作者：高庆波，副研究员，主要从事虎耳草属植物分类学及系统发育学研究。E-mail：qbgao@nwipb.cas.cn

kb; for *S. consanguinea*, 4 542 SSRs were distributed in 60 972 Unigenes which accounted for 7.45%, there were 85 kinds of repeat motifs and SSRs occurred every 10.40 kb in length, which was slightly lower than that of *S. sinomontana*. For the two species, the tri-nucleotide was dominant repeat motif. The repeat numbers of SSRs were mainly from 5 to 10 and their motif length mainly ranged from 12－30 bp, which suggested that these SSRs displayed high levels of polymorphism. Besides, there were 118 424 and 112 006 SNPs in *S. sinomontana* and *S. consanguinea*. The proportion of SNPs in the coding region were 30.40% and 28.59% respectively, and the proportion of the synonymous mutations in the coding region (30.27%, 28.48%) was significantly higher than that of the nonsynonymous mutations (0.13%, 0.11%). It was found that the relevant indicators of the two species showed no significant difference, which was probably related to the tissue and its development stage, as well as the phylogenetic relationships of the species.

Key words: *Saxifraga sinomontana*; *Saxifraga consanguinea*; transcriptome; simple sequence repeats (SSR); simple nucleotide polymorphism (SNP)

近年来，随着现代分子生物学技术的发展，第二代简单重复序列标记和第三代单核苷酸多态性标记已逐渐成为动植物遗传多样性分析、系统发育和进化研究中使用较为广泛的遗传标记，并成为了生命科学研究领域一个不可或缺的工具。简单重复序列(simple sequence repeats，SSR)，又称微卫星(microsatellite)，是一类由几个核苷酸(2～6 个)为重复单位组成的长达几十个核苷酸的重复序列，广泛分布于真核生物基因组中，约每 10～50 kb 的序列中就有 1 个微卫星位点，且具有共显性、长度短、多态性高、易于检测和相对保守等特点[1-3]。

单核苷酸多态性(single nucleotide polymorphism，SNP)指染色体基因组中单个核苷酸的变异而引起的 DNA 序列多态性，形式包括单碱基的缺失、插入、转换及颠换等[4]。SNP 是二等位基因，具有在基因组中分布广泛、多样性高及可高通量自动化检测等特点[5]。

虎耳草属(*Saxifraga* L.)是虎耳草科(Saxifragaceae)中最大的属，大约有 450～500 种，主要分布在欧洲和亚洲的高山地区[6-7]。中国产约 220 种虎耳草属植物，主要分布在青藏高原-喜马拉雅地区[8]。山地虎耳草(*Saxifraga sinomontana* J. T. Pan & Gornall)和棒腺虎耳草(*Saxifraga consanguinea* W. W. Smith)均隶属于虎耳草科虎耳草属，是多年生草本植物，在中国主要分布于青海、甘肃、四川、云南及西藏等地，其生境多为高海拔地区的高山草甸、灌丛和石隙[8]，是青藏高原地区高寒草甸生态系统的重要组成部分。此外，二者形态学差别较大，主要体现为山地虎耳草基生叶发达，边缘具有褐色卷曲长柔毛，而棒腺虎耳草茎基部叶腋处有丝状鞭匐枝，且基生叶密集聚成莲座状，边缘具腺睫毛(短棒状)[8]。目前 DNA 分子标记已广泛应用于虎耳草属植物的系统发育学和谱系地理学研究，结果表明对于该属内的不同物种，其遗传结构和进化历史不尽相同[6-7,9-12]。

本研究基于山地虎耳草和棒腺虎耳草的高通量测序结果，分析和比较 SSR 和 SNP 在 2 个物种内的分布规律和特点，为后期 SSR 标记的开发和系统发育学研究奠定理论基础。

1　材料和方法

1.1　样品采集和高通量测序

山地虎耳草(*S. sinomontana*)和棒腺虎耳草(*S. consanguinea*)分别采集于青海省玉树藏族自治州玉树县小苏莽乡(32°34′20.7″N，97°12′41.6″E，4 880 m)、青海省玉树藏族自治州玉树县勒巴沟(32°55′18.2″N，97°13′54.4″E，3 667 m)。将野外采集的活体材料置于室内种植 68 d，再采取二者同一丛植株上的叶片，放入冷冻管中，用液氮处理约 15 s 后放入－80 ℃冰箱保存。凭证标本保存于中国科学院西北高原生物研究所青藏高原生物标本馆(HNWP)。

分别从山地虎耳草和棒腺虎耳草的叶片材料中提取 100 μg 总 RNA；利用诺禾致源生物信息科技有限公司的 Illumina HiSeq™ 2000 高通量测序平台对其进行测序；对测得的原始序列(Raw reads)进行过滤：去除带接头(adapter)的、N 比例大于 10% 的以及低质量的 reads，得到干净的读序(Clean reads)；最终用 Trinity[13] 将其拼接成一个转录组，并取每条基因中最长的转录本作为 Unigene，以此进行后续分析。

1.2　SSR 和 SNP 的筛选和统计分析

用 MicroSatellite (MISA；http://pgrc. ipk-

gatersleben. de/misa/)对 Unigene 进行 SSR 检测、筛选和分析。检索标准同时包括精确型(perfect)和复合型(compound)SSR 重复单元[14],各重复微卫星类型重复次数设定如下:两碱基(di-nucleotide repeats,DNRs)至少重复 6 次,三碱基(tri-nucleotide repeats,TNRs)至少重复 5 次,四碱基(tetra-nucleotide repeats,TTNRs)至少重复 5 次,五碱基(penta-nucleotide repeats,PTNRs)至少重复 5 次,六碱基(hexanucleotide repeats,HXNRs)至少重复 5 次。最终对 SSR 出现频率、重复基元类型、重复次数及其多态性进行统计分析。

通过 samtools 和 picard-tools 等工具对比对结果进行染色体坐标排序并去掉重复的 reads 等处理,最后利用变异检测软件 CATK2[15] 以 Unigene 为参考序列对 reads 进行 SNP Calling,并对原始结果进行过滤:去除质量值小于 30,距离小于 5 的 SNP。最终对得到的 SNP 位点位于编码区或者非编码区,以及属于编码区中的同义转换或者非同义转换的 SNP 数量进行统计分析。

2 结果与分析

2.1 山地虎耳草和棒腺虎耳草转录组测序结果

对于 RNA-seq 技术,其测序错误率会随着测序序列长度的增加而升高,这是测序过程中化学试剂的消耗所导致的[16-17],其次,可能因为随机引物与 RNA 模板的不完全结合使得前 6 个碱基的位置也会发生较高的测序错误率[17],且单个碱基位置的测序错误率一般在 1%以下。本研究中,山地虎耳草和棒腺虎耳草分别获得 94 855 756 条和 93 118 446 条 Raw reads,过滤后分别获得 90 311 228 条、88 874 280条 Clean reads,分别占 Raw reads 的 95.21%和 95.44%。其中,两物种的单碱基错误率分别为0.035%和 0.04%,碱基 Q_{20} 分别为94.36%、94.00%,碱基 Q_{30} 分别为 88.98%、88.38%,碱基 G 和 C 的数量总和比例分别为 42.39%和 42.44%。

用 Trinity 软件对所得的 Clean reads 进行拼接,最终山地虎耳草获得 176 110 个 Transcripts 和 63 763 个 Unigene,棒腺虎耳草获得 150 308 个 Transcripts 和 60 972 个 Unigene(图 1);之后对 2 个物种的 Transcripts 和 Unigene 的长度统计结果(图 1)表明,在山地虎耳草中 Transcripts 和 Unigene 总的核苷酸数分别为 189 919 691 个、46 218 250个,棒腺虎耳草中二者总的核苷酸数分别为 180 129 302 个、47 241 106个。

2.2 山地虎耳草和棒腺虎耳草 SSR 的频率及其分布

采用 MISA 对 Unigene 进行 SSR 检测,结果显示:山地虎耳草中含有 SSR 的序列为 7 700 条,发生频率为 12.08%,其中 6 454 条序列含有单个 SSR,1 246条序列含有 1 个以上的 SSR。表 1 显示,山地虎耳草中共检测出 4 622 个 SSR,包括 4 098 个完全型 SSR 和 524 个复合型 SSR,其发生频率为 7.25%(检测出的 SSR 数量与总序列数目的比值)。在棒腺虎耳草中,共7 073条序列含有 SSR,发生频率为 11.60%,其中5 981条序列含有单个 SSR,1 092 条序列含有 1 个以上的 SSR,共检测出 4 542 个 SSR,包括 4 051 个完全型 SSR 和 491 个复合型 SSR,发生频率为7.45%(表 1)。从分布情况来看,山地虎耳草转录组序列中平均每 10.00 kb 出现一个 SSR,棒腺虎耳草中平均每 10.40 kb 出现一个 SSR,略低于前者(表 1)。

对 2 个物种的 SSR 类型进行统计发现,二至六核苷酸重复类型均有出现,但各类型出现的频率和分布的平均距离相差较大。表 1 显示,在山地虎耳

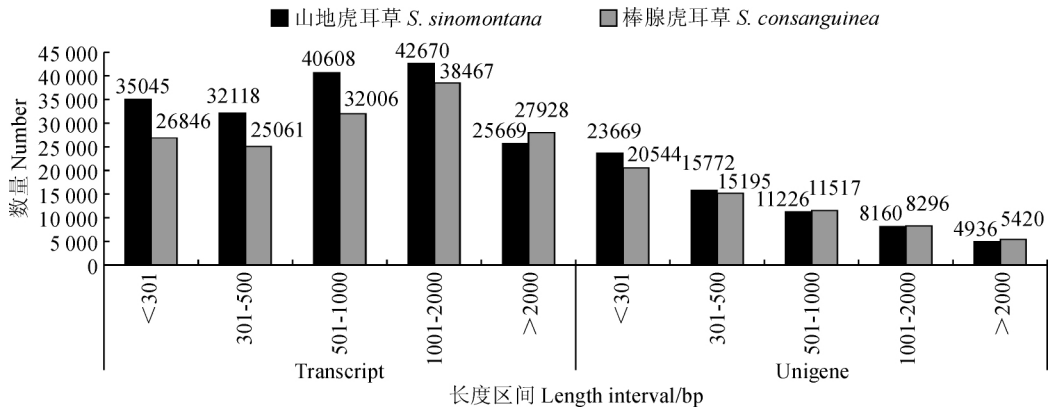

图 1 山地虎耳草和棒腺虎耳草拼接后的 Transcript 与 Unigene 长度分布

Fig. 1 The length distribution of Transcripts and Unigenes after assemblage of *S. sinomontana* and *S. consanguinea*

草和棒腺虎耳草中，三核苷酸重复类型的 SSR 含量均为最多，分别占总 SSR 的 55.50%和 56.36%；其次为二核苷酸重复类型，所占比例分别为30.23%和30.32%；其他类型（四核苷酸、五核苷酸、六核苷酸和复合核苷酸重复）所占比例较小，总和分别为14.28%和13.32%。从分布情况来看，两个物种不同重复基元 SSR 分布的平均距离均差别较大。其中，在山地虎耳草中，三核苷酸重复数量最多，出现频率为4.02%，每条 SSR 分布的平均距离为 18.02 kb，六核苷酸重复最少，出现频率为 0.02%，平均距离为3 355.25 kb；在棒腺虎耳草中，三核苷酸的出现频率为4.20%，每条 SSR 分布的平均距离为18.45 kb，与山地虎耳草不同的是，在该物种中五核苷酸重复最少，出现频率仅为 0.01%，平均距离为5 905.14 kb。此外，棒腺虎耳草各重复基元类型的平均距离均高于山地虎耳草（表 1）。

2.3　山地虎耳草和棒腺虎耳草 SSR 基元类型和比例

表 2 显示，在山地虎耳草转录组 4 098 个完全型 SSR 中共发现了 110 种重复基元，其中二至六核苷酸重复基元分别有 6 种、30 种、42 种、19 种和 13 种。在棒腺虎耳草转录组 4 051 个完全型 SSR 中则发现了 85 种重复基元，明显少于山地虎耳草，其中二至六核苷酸重复基元分别有 6 种、30 种、31 种、8 种和 10 种。

在 2 个物种的二核苷酸重复基元中，均属 AG/TC 出现次数最多，分别有 521 个和 548 个，为二碱基的优势重复单元，分别占二核苷酸重复基元 SSR 总数的 37.29%和 39.80%；同时 AG/TC 也是所有二至六核苷酸重复基元中数量最多的 SSR，而其余的二核苷酸重复类型在 2 个物种中所占比例也均有不同。此外，2 个物种的三碱基和四碱基的优势重复单元也有所不同，山地虎耳草中，AAG/TTC（233 个）和 AAGA/TTCT（8 个）出现频率最高，在棒腺虎耳草中出现频率最高的则是 CTT/GAA（203 个）和 AAAT/TTTA（14 个）。五核苷酸和六核苷酸在 2 个物种中出现频率普遍偏低（表 2）。

2.4　山地虎耳草和棒腺虎耳草 SSR 重复次数分布

研究表明，SSR 基元重复次数变异而引起的位点长度变化是产生位点多态性的主要原因[18-19]。通过对山地虎耳草 4 098 个和棒腺虎耳草 4 051 个完全型 SSR 进行分类统计，结果（图 2）发现，随着重复次数的增加，二者的 SSR 数量均逐渐减少。且 2 个物种的 SSR 均主要分布在 5～10 次的较低重复次数基元中，其中山地虎耳草有 4 036 个 SSRs，占总 SSR 的 98.49%；棒腺虎耳草中有 3 994 个 SSRs，占98.59%；11 次、12 次和 14 次为一般重复次数基元，在 2 个物种中分别有 61 个和 60 个 SSRs，分别占1.49%和 1.48%；20 次以上为较高重复次数基元，在山地虎耳草和棒腺虎耳草中分别只出现

表 1　山地虎耳草和棒腺虎耳草 SSR 序列的出现频率

Table 1　Frequency of SSR sequences in transcriptome *S. sinomontana* and *S. consanguinea*

重复基元类型 Repeat type	山地虎耳草 *S. sinomontana*				棒腺虎耳草 *S. consanguinea*			
	数量 Number	比例 Proportion/%	频率 Frequency/%	平均距离 Average distance /kb	数量 Number	比例 Proportion/%	频率 Frequency/%	平均距离 Average distance/kb
二核苷酸 Di-nucleotide	1 397	30.23	2.19	33.08	1 377	30.32	2.26	34.31
三核苷酸 Tri-nucleotide	2 565	55.50	4.02	18.02	2 560	56.36	4.20	18.45
四核苷酸 Tetra-nucleotide	104	2.25	0.16	444.41	96	2.11	0.16	492.09
五核苷酸 Penta-nucleotide	19	0.41	0.03	2 432.54	8	0.18	0.01	5 905.14
六核苷酸 Hexa-nucleotide	13	0.28	0.02	3 355.25	10	0.22	0.02	4 724.11
复合 Compound	524	11.34	0.82	88.20	491	10.81	0.81	96.21
合计 Total	4 622	100	7.25	10.00	4 542	100	7.45	10.40

注：比例：各核苷酸 SSR 在总 SSR 中所占比例；频率：含有 SSR 的序列数目与总序列数目的比值；平均分布距离：序列总长度与 SSR 数目的比值

Note: Proportion: Proportion in all SSRs; Frequency: The percentage of SSR number in all sequences; Average distance: Ratio of total sequence length and SSR number

表 2 山地虎耳草和棒腺虎耳草转录组中不同 SSR 序列的出现情况

Table 2 Occurrence of different SSR sequences in transcriptome of *S. sinomontana* and *S. consanguinea*

	山地虎耳草 *S. sinomontana*				棒腺虎耳草 *S. consanguinea*			
重复类型 Repeat type	重复基元 Repeat motif	数量 Number	比例 Proportion/%	频率 Frequency/%	重复基元 Repeat motif	数量 Number	比例 Proportion/%	频率 Frequency/%
二核苷酸 Di-nucleotide	AG/TC	521	11.27	0.817	AG/TC	548	13.53	0.899
	AT/TA	346	7.49	0.543	CT/GA	313	7.73	0.513
	CT/GA	292	6.32	0.458	AT/TA	299	7.38	0.490
	AC/TG	131	2.83	0.205	AC/TG	121	2.99	0.198
	CA/GT	101	2.19	0.158	CA/GT	90	2.22	0.148
	CG/GC	6	0.13	0.009	CG/GC	6	0.15	0.010
三核苷酸 Tri-nucleotide	AAG/TTC	233	5.04	0.365	CTT/GAA	203	5.01	0.333
	CTT/GAA	202	4.37	0.317	AAG /TTC	187	4.62	0.307
	AAC/TTG	180	3.89	0.282	AGA/TCT	181	4.47	0.297
	AGA/TCT	164	3.55	0.257	AAC/TTG	177	4.37	0.290
	ACT/TGA	157	3.40	0.246	ATC/TAG	155	3.83	0.254
	CTA/GAT	157	3.40	0.246	TCA/AGT	155	3.83	0.254
	AGT/TCA	144	3.12	0.226	CAA/GTT	153	3.78	0.251
	ACA/TGT	143	3.09	0.224	GAT/CTA	140	3.46	0.230
	CAA/GTT	129	2.79	0.202	TGA/ACT	127	3.14	0.208
	ATC/TAG	118	2.55	0.185	TGT/ACA	125	3.09	0.205
	CCA/GGT	118	2.55	0.185	CCA/GGT	125	3.09	0.205
	ACC/TGG	100	2.16	0.157	CAT/GTA	118	2.91	0.194
	其他 others	720	15.59	1.13	其他 others	714	17.63	1.173
四核苷酸 Tetra-nucleotide	AAGA/TTCT	8	0.17	0.013	AAAT/TTTA	14	0.35	0.023
	AGAA/TCTT	6	0.13	0.009	AAAC/TTTG	8	0.20	0.013
	CAAA/GTTT	6	0.13	0.009	AATA/TTAT	7	0.17	0.011
	AACA/TTGT	5	0.11	0.008	ACAA/TGTT	5	0.12	0.008
	—	—	—	—	AGAA/TCTT	5	0.12	0.008
	—	—	—	—	ATCA/TAGT	5	0.12	0.008
	—	—	—	—	CAAA/GTTT	5	0.12	0.008
	—	—	—	—	CTTT/GAAA	5	0.12	0.008
	其他 others	79	1.75	3.072	其他 others	42	0.96	0.074
五核苷酸 Penta-nucleotide	总计 total	19	0.38	0.038	总计 total	8	0.16	0.016
六核苷酸 Hexa-nucleotide	总计 total	13	0.26	0.026	总计 total	9	0.18	0.018
总计 Total	—	4 098	88.66	6.427	—	4 051	89.19	6.644

注：比例：各核苷酸 SSR 在总 SSR 中所占比例；频率：含有 SSR 序列数目与总序列数目的比值

Note: Proportion: Proportion in all SSRs; Frequency: The percentage of SSR number in all sequences

了 50 次和 25 次的重复，且均仅含 1 个 SSR(图 2)。

据统计，在山地虎耳草和棒腺虎耳草中，均属二核苷酸基元重复次数类型最多，跨度最大，分别有 8 种和 7 种，且主要类型均为 6 次重复，分别有 627 个、596 个(图 3)，占相应物种二核苷酸基元的 44.88%和 43.28%；位于二核苷酸之后的是三核苷

酸，分别为山地虎耳草中的 ATT/TAA(5 种)和棒腺虎耳草中的 CCA/GGT(5 种)出现的类型最多；2 物种的四核苷酸、五核苷酸和六核苷酸重复基元中多以 5 次、6 次重复类型出现。此外，在这 2 个物种的 5 种核苷酸基元中，随着重复次数的增加，SSR 数量所占比例都有逐渐减少的趋势(图 3，A、B)。

2.5　山地虎耳草和棒腺虎耳草 SSR 基元长度分布及其多态性

图 4 和表 3 显示，山地虎耳草和棒腺虎耳草的完全型 SSR 基元长度区间分别为 12～100 bp 和 12～75 bp，其中最大的片段长度分别为前者二核苷酸重复 50 次(100 bp)和后者三核苷酸重复 25 次(75 bp)的 SSR。从整体来看，二者 SSR 的分布范围较为集中，主要在 12～30 bp，且在所有 SSR 中，最多的为 15 bp 长度的 SSR，其中山地虎耳草有 1 529 个，占 37.31%，棒腺虎耳草有 1 581 个，占39.03%，并且均为 5 次重复的三核苷酸基元(图 4)。

研究表明，当 SSR 基序长度大于等于 20 bp 时其多态性较高，长度在 12～20 bp 时多态性中等，而长度在 12 bp 以下时多态性极低[20]。本研究筛选得到的山地虎耳草和棒腺虎耳草转录组 SSR 的长度均大于等于 12 bp，其中前者 12～19 bp 的 SSR 有 3 439 个(83.92%)，后者有 3 462 个(85.46%)(表 3)，这些 SSR 具有中等多态性；而 2 种虎耳草大

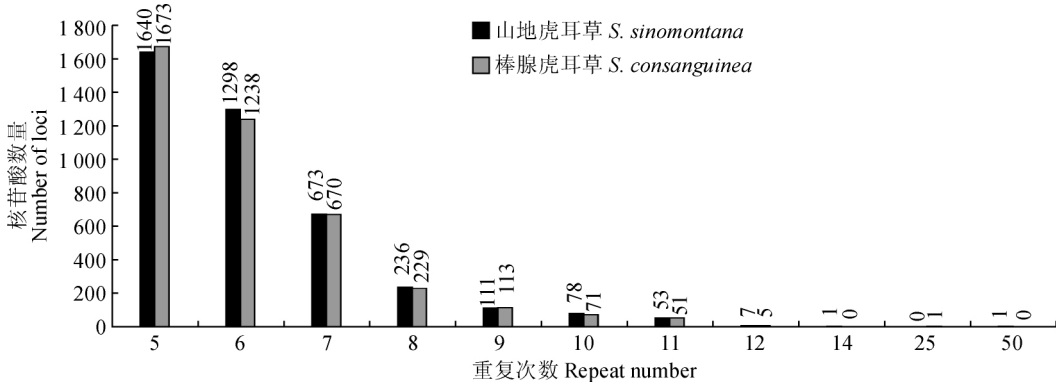

图 2　山地虎耳草和棒腺虎耳草转录组 SSR 重复次数分布

Fig. 2　The distribution of repeat number of SSRs in transcriptome of *S. sinomontana* and *S. consanguinea*

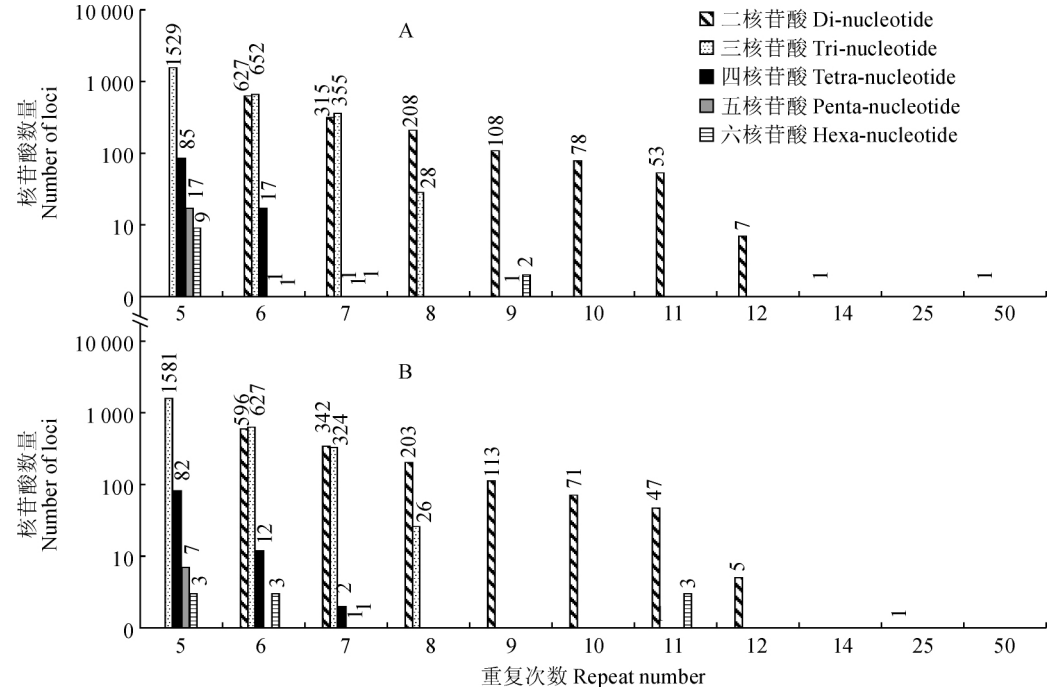

图 3　山地虎耳草(A)和棒腺虎耳草(B)转录组 SSR 不同重复类型的重复次数分布

Fig. 3　The distribution of repeat number of SSRs for different repeat types in transcriptome of *S. sinomontana* (A) and *S. consanguinea* (B)

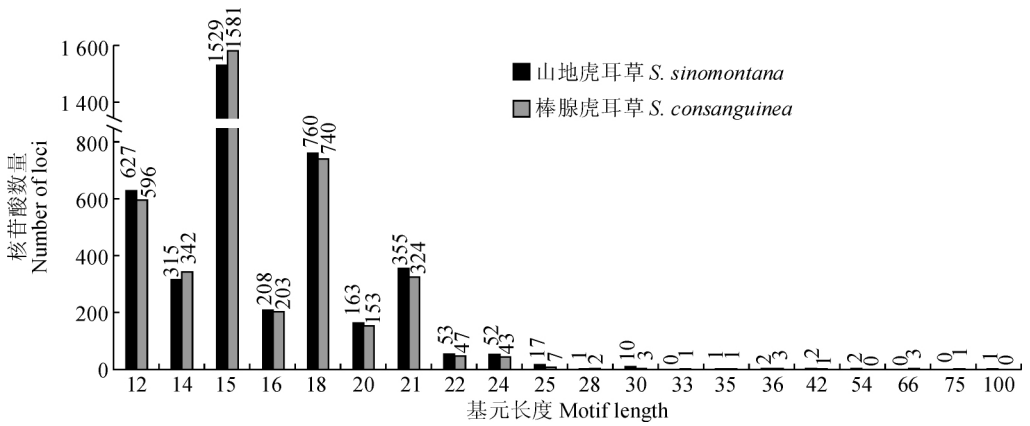

图 4 山地虎耳草和棒腺虎耳草转录组 SSR 基元长度分布

Fig. 4 The distribution of motif length of SSRs in transcriptome of *S. sinomontana* and *S. consanguinea*

表 3 山地虎耳草和棒腺虎耳草转录组 SSR 不同重复类型的基元长度分布

Table 3 The distribution of motif length of SSRs for different repeat types in transcriptome of *S. sinomontana* and *S. consanguinea*

重复类型 Repeat type	长度 Length/bp	山地虎耳草 *S. sinomontana*		棒腺虎耳草 *S. consanguinea*	
		SSR 数量 Number of SSRs	SSR 所占百分比 Percent of total SSRs/%	SSR 数量 Number of SSRs	SSR 所占百分比 Percent of total SSRs/%
二核苷酸 Di-nucleotide	12	627	44.88	596	43.28
	14	315	22.55	342	24.84
	16	208	14.89	203	14.74
	18	108	7.73	113	8.21
	20	78	5.58	71	5.16
	22	53	3.79	47	3.41
	24	7	0.50	5	0.36
	100	1	0.07	0	0.00
三核苷酸 Tri-nucleotide	15	1 529	59.61	1 581	61.76
	18	652	25.42	627	24.49
	21	355	13.84	324	12.66
	24	28	1.09	26	1.02
	33	0	0.00	1	0.04
	42	1	0.04	0	0.00
	75	0	0.00	1	0.04
四核苷酸 Tetra-nucleotide	20	85	81.73	82	85.42
	24	17	16.35	12	12.50
	28	1	0.96	2	2.08
	36	1	0.96	0	0.00
五核苷酸 Penta-nucleotide	25	17	89.47	7	87.50
	30	1	5.26	0	0.00
	35	1	5.26	1	12.50
六核苷酸 Hexa-nucleotide	30	9	69.23	3	30.00
	36	1	7.69	3	30.00
	42	1	7.69	1	10.00
	54	2	15.38	0	0.00
	66	0	0.00	3	30.00

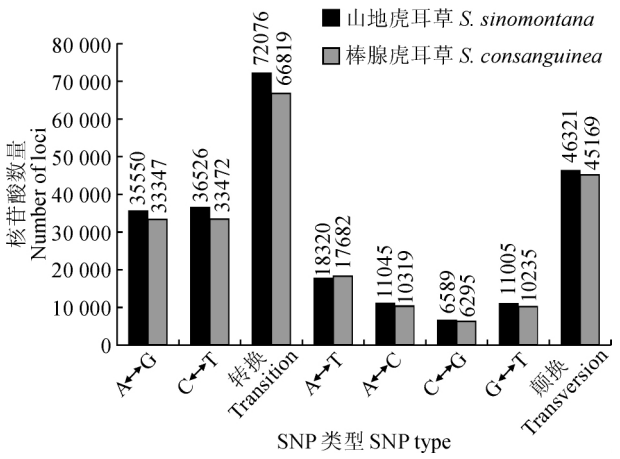

图 5　山地虎耳草和棒腺虎耳草 SNPs 类型分析

Fig. 5　The analysis of SNP types in transcriptome of *S. sinomontana* and *S. consanguinea*

于等于 20 bp 的 SSR 分别有 659 个(16.08%)、589 个(14.54%)(表 3)，这些 SSR 具有较高的多态性。由此推测本研究中山地虎耳草和棒腺虎耳草转录组 SSR 的多态性均在中等以上。此外，研究发现高级基元 SSR 的多态性普遍比低级基元的低[21]。本研究中山地虎耳草和棒腺虎耳草的二、三核苷酸基元总和分别占完全型 SSR 的 96.68%、97.19%，在长度大于等于 20 bp 的 SSR 中，两物种所包含的低级基元(二核苷酸和三核苷酸)总数分别为 523 条、475 条，占长度大于等于 20 bp 的所有 SSR 的 79.36%、80.65%，表明大部分山地虎耳草和棒腺虎耳草转录组 SSR 具有高多态性潜能(表 3)。

2.6　山地虎耳草和棒腺虎耳草的 SNP 位点特征分析

通过数据处理，最终在山地虎耳草和棒腺虎耳草中分别获得 118 424 个和 112 006 个 SNP 位点，其中非编码区的 SNP 位点分别为 82 420 个(69.60%)、79 986个(71.41%)，编码区的 SNP 位点分别为 36 004个(30.40%)、32 020 个(28.59%)。在山地虎耳草的编码 SNP 中，同义突变有 35 849 个(30.27%)，非同义突变 155 个(0.13%)；棒腺虎耳草的编码 SNP 中，同义突变有 31 899 个(28.48%)，非同义突变有 121 个(0.11%)。

对二者 SNP 进行类型分析结果(图 5)发现，在山地虎耳草中，转换类型有 72 076 个(60.86%)，颠换类型有 46 321 个(39.11%)；棒腺虎耳草中，转换类型有 66 819个(59.66%)，颠换类型有 45 169 个(40.33%)。

3　讨　论

转录组代表了特定物种的组织或细胞在不同发育阶段、不同生理状态下的全部 mRNA 总和[22]，且不同物种、同一物种的不同个体、同一个体的不同组织以及同一组织不同时期的转录本表达情况都有所不同[23]。本研究从山地虎耳草和棒腺虎耳草转录组中分别获得了 63 763 个和 60 972 个 Unigene，为后续分析奠定了坚实的基础。其中，山地虎耳草中，碱基 Q_{20} 和 Q_{30} 分别为 94.36%、88.98%，棒腺虎耳草中相应比例为 94.00%、88.38%，研究指出，当 Q_{30} 值在 80%以上就认为测序质量非常可靠[24]，而碱基 Q_{20} 与碱基识别的错误率呈对数相关，其表示每 100 个序列碱基中仅有 1 个出错的概率[25]。

从 2 物种的 Unigene 中分别检测出了 4 622 个和 4 542 个 SSR，平均分布距离为 1/10.00 kb、1/10.40 kb，二者差别不大，但与其他高山植物相比，其平均分布距离高于冷蒿(1/18.46 kb)[26]和川西獐牙菜(1/12.6 kb)[27]，与蓝玉簪龙胆(1/9.97 kb)[28]相差较小，但低于唐古特红景天(1/8.52 kb)[29]。由此表明，2 个物种转录组中 SSR 的数量较为丰富。此外，山地虎耳草和棒腺虎耳草转录组中 SSR 出现频率也相似，分别为 7.25%、7.45%，与唐古特红景天(7.1%)[29]的出现频率较为接近，高于冷蒿(2.61%)[26]和蓝玉簪龙胆(6.12%)[28]，但低于川西獐牙菜(8.16%)[27]。出现这种差异可能与物种选择、组装方法、SSR 搜索的标准及分析方法等有关。

研究表明，大多数植物的 SSR 主要以二核苷酸和三核苷酸重复为主要类型，但是主导重复基元的类型有所不同[27,30]。本研究发现，这 2 个物种转录组 SSR 的优势基元均是三核苷酸重复，这与冷蒿[26]、蓝玉簪龙胆[28]和唐古特红景天[29]和等植物的优势基元结果相一致；但在金银花[31]和芝麻[32]等植物中二核苷酸重复占主导地位，在川西獐牙菜[27]和灯盏花[30]等植物中主导类型为二核苷酸和三核苷酸重复基元。这种主导重复基元的不同可能与物种自身的差异有关。此外，有研究指出三核苷酸和六核苷酸 SSR 重复基序的突变情况，可能是一种有利于植物进化的突变[33]，在山地虎耳草和棒腺虎耳草中以三核苷酸 SSR 为主体的分布可能与长期以来自然选择所导致的适应性变化有关。

多项研究指出，作为碱基序列的重要特征之一，GC 含量反映了基因的结构、功能和进化信息，SSR 序列中 GC 含量的增加会使某些氨基酸序列的增加而获得某些特定功能，如胁迫抗性、转录调控、信号转导等[14,34-35]。并且在大多数植物中 GC 重复基元

很少出现，例如在唐古特红景天[29]、金银花[31]和小麦[36]等植物中均未发现该重复基元，但在山地虎耳草和棒腺虎耳草转录组 SSR 中均检测到了 6 个 GC 重复基元，这种现象在川西獐牙菜[27]和大豆[37]中也出现过。其次，山地虎耳草和棒腺虎耳草对青藏高原高寒干旱、土壤贫瘠等极端环境的适应机制是否与 GC 重复单元有关还需要进行更深入的探讨。

对山地虎耳草和棒腺虎耳草的 SNP 分析发现，两者 SNP 的转换类型比例均明显高于颠换类型，并且在转换类型中，C↔T 发生频率较高，与蓝玉簪龙胆的 SNPs 结果相一致[28]，这一现象可能是由于 SNP 在 CG 序列上出现最为频繁，并且 CG 中的 C 常为甲基化，在自发地脱氨后便成为 T 所导致[38]。此外，对这 2 个物种的编码 SNP 比较发现，物种间同义突变个数和非同义突变个数均为接近，但物种内同义突变个数均明显高于非同义突变，这可能是由于在自然选择的作用下，绝大多数引起氨基酸序列的突变因为降低了物种的适合度而遭到淘汰，进而导致了蛋白质编码区的非同义突变率低于同义突变率。

青藏高原作为虎耳草属植物多样性中心之一，虽然资源丰富独特，但相关物种的基因组学研究相对滞后，遗传信息较为缺乏。本研究通过分析和比较山地虎耳草和棒腺虎耳草转录组序列中 SSR 和 SNP 的分布情况，发现二者的结果差别较小，这可能与选取的组织部位相同、发育阶段相同以及物种间的系统发育学关系较近有关。本研究结果可为今后对山地虎耳草和棒腺虎耳草进行引物设计、生态适应和系统发育学研究提供基础，为保护生物学提供理论依据。

参考文献：

[1] LITT M, LUTY J A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene[J]. *American Journal of Human Genetics*, 1989, **44**(3): 397-401.

[2] WEBER J L. Informativeness of human (dC-dA)n·(dG-dT)n polymorphism[J]. *Genomics*, 1990, **7**(4): 524-530.

[3] SLATE J, COLTMAN D W, GOODMAN S J, *et al*. Bovine microsatellite loci are highly conserved in red deer (*Cervus elaphus*), sika deer (*Cervus nippon*) and Soay sheep (*Ovis aries*)[J]. *Animal Genetics*, 2015, **29**(4): 307-315.

[4] **杨昭庆，洪坤学. 单核苷酸多态性的研究进展[J]. 国外医学：遗传学分册**, 2000, **23**(1): 4-8.
YANG Z Q, HONG K X. Progress in single nucleotide polymorphisms[J]. *Foreign Medical Sciences* (Section of Genetics), 2000, **23**(1): 4-8.

[5] MORIN P A, LUIKART G, WAYNE R K, *et al*. SNPs in ecology, evolution and conservation[J]. *Trends in Ecology & Evolution*, 2004, **19**(4): 208-216.

[6] GAO Q B, LI Y H, GORNALL R J, *et al*. Phylogeny and speciation in *Saxifraga* sect. *Ciliatae* (Saxifragaceae): evidence from *psbA-trnH*, *trnL-F* and ITS sequences[J]. *Taxon*, 2015, **64**(4): 703-713.

[7] EBERSBACH J, MUELLNER-RIEHL A N, MICHALAK I, *et al*. In and out of the Qinghai-Tibet Plateau: divergence time estimation and historical biogeography of the large arctic-alpine genus *Saxifraga* L.[J]. *Journal of Biogeography*, 2017, **44**(4): 900-910.

[8] WU C Y, RAVEN P H. Flora of China [M]. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press, 2001, **8**: 208-344.

[9] ABBOTT R J, COMES H P. Evolution in the Arctic: a phylogeographic analysis of the circumarctic plant, *Saxifraga oppositifolia* (Purple saxifrage)[J]. *New Phytologist*, 2003, **161**(1): 211-224.

[10] DECHAINE E G, ANDERSON S A, MCNEW J M, *et al*. On the evolutionary and biogeographic history of *Saxifraga* sect. *Trachyphyllum* (Gaud.) Koch (Saxifragaceae Juss.)[J]. *PLoS One*, 2013, **8**(7): e69814.

[11] EBERSBACH J, SCHNITZLER J, FAVRE A, *et al*. Evolutionary radiations in the species-rich mountain genus *Saxifraga* L.[J]. *BMC Evolutionary Biology*, 2017, **17**(1): 119.

[12] GAO Q B, LI Y, GENGJI Z M, *et al*. Population genetic differentiation and taxonomy of three closely related species of *Saxifraga* (Saxifragaceae) from southern Tibet and the Hengduan Mountains[J]. *Frontiers in Plant Science*, 2017, **8**: 1 325.

[13] GRABHERR M G, HAAS B J, YASSOUR M, *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome[J]. *Nature Biotechnology*, 2011, **29**(7): 644-652.

[14] LI S X, YIN T M. Map and analysis of microsatellites in the genome of *Populus*: the first sequenced perennial plant[J]. *Science in China Series C*: Life Sciences, 2007, **50**(5): 690-699.

[15] MCKENNA A, HANNA M, BANKS E, *et al*. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. *Genome Research*, 2010, **20**(9): 1 297-1 303.

[16] ERLICH Y, MITRA P P, DELABASTIDE M, *et al*. Alta-Cyclic: a self-optimizing base caller for next-generation sequencing[J]. *Nature Methods*, 2008, **5**(8): 679-682.

[17] JIANG L, SCHLESINGER F, DAVIS C A, *et al*. Synthetic spike-in standards for RNA-seq experiments[J]. *Genome research*, 2011, **21**(9): 1 543-1 551.

[18] **张水仙. 基于蒙药冷蒿基因组低通量测序的 SSR 遗传多样性分析[D]. 北京：中央民族大学**, 2013.

[19] 刘　越，范增华，孙洪波，等. 裂叶牵牛获得表达序列标签资源的简单序列重复信息分析[J]. 中国药学杂志，2011，**46**(23)：1 790-1 794.
LIU Y，FAN Z H，SUN H B，*et al*. Analysis of SSR information in EST resource of *Pharbitis nil*[J]. *Chinese Pharmaceutical Journal*，2011，**46**(23)：1 790-1 794.

[20] TEMNYKH S，DECLERCK G，LUKASHOVA A，*et al*. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.)：Frequency，length variation，transposon associations，and genetic marker potential[J]. *Genome Research*，2001，**11**(8)：1 441-1 452.

[21] DREISIGACKER S，ZHANG P，WARBURTON M L，*et al*. SSR and pedigree analyses of genetic diversity among CIMMYT wheat lines targeted to different megaenvironments [J]. *Crop Science*，2004，**44**(2)：381-388.

[22] 李小白，向　林，罗　洁，等. 转录组测序(RNA-seq)策略及其数据在分子标记开发上的应用[J]. 中国细胞生物学学报，2013，**35**(5)：720-726.
LI X B，XIANG L，LUO J，*et al*. The strategy of RNA-seq，application and development of molecular marker derived from RNA-seq[J]. *Chinese Journal of Cell Biology*，2013，**35**(5)：720-726.

[23] 祁云霞，刘永斌，荣威恒. 转录组研究新技术：RNA-Seq 及其应用[J]. 遗传，2011，**33**(11)：1 191-1 202.
QI Y X，LIU Y B，RONG W H. RNA-Seq and its applications：a new technology for transcriptomics[J]. *Hereditas*，2011，**33**(11)：1 191-1 202.

[24] 贾新平，孙晓波，邓衍明，等. 鸟巢蕨转录组高通量测序及分析[J]. 园艺学报，2014，**41**(11)：2 329-2 341.
JIA X P，SUN X B，DENG Y M，*et al*. Sequencing and analysis of the transcriptome of *Asplenium nidus*[J]. *Acta Horticulturae Sinica*，2014，**41**(11)：2 329-2 341.

[25] EWING B，HILLIER L，WENDL M C，*et al*. Base-calling of automated sequencer traces using phred. I. Accuracy assessment[J]. *Genome Research*，1998，**8**(3)：175-185.

[26] 岳春江，陈川川，郭凤仙，等. 蒙药冷蒿转录组 SSR 信息分析[J]. 中国农业科技导报，2016，**18**(6)：31-43.
YUE C J，CHEN C C，GUO F X，*et al*. Data mining of simple sequence repeats in transcriptome sequences of Mongolia medicinal plant *Artemisia frigida* Willd. [J]. *Journal of Agriculture Science and Technology*，2016，**18**(6)：31-43.

[27] 刘　越，岳春江，王　翊，等. 藏茵陈川西獐牙菜转录组 SSR 信息分析[J]. 中国中药杂志，2015，**40**(11)：2 068-2 076.
LIU Y，YUE C J，WANG Y，*et al*. Data mining of simple sequence repeats in transcriptome sequences of Tibetan medicinal plant Zangyinchen *Swertia mussotii*[J]. *China Journal of Chinese Materia Medica*，2015，**40**(11)：2 068-2 076.

[28] 田尊哲，高庆波，陈世龙，等. 基于高通量测序分析青藏高原特有植物蓝玉簪龙胆的 SSR 和 SNP 特征[J]. 植物研究，2016，**36**(5)：747-752.
TIAN Z Z，GAO Q B，CHEN S L，*et al*. Characteristics of SSR and SNP in *Gentiana veitchiorum* in QingHai-Tibetan Plateau，by high-throughput sequencing[J]. *Bulletin of Botanical Research*，2016，**36**(5)：747-752.

[29] 雷淑芸，高庆波，付鹏程，等. 基于 Solexa 高通量测序的唐古特红景天微卫星信息分析[J]. 植物研究，2014，**34**(6)：829-834.
LEI S Y，GAO Q B，FU P C，*et al*. Analysis on microsatellites in *Rhodiola algida* based on Solexa Sequencing[J]. *Bulletin of Botanical Research*，2014，**34**(6)：829-834.

[30] 陈　茵，李翠婷，姜倪皓，等. 灯盏花转录组中 SSR 位点信息分析及其多态性研究[J]. 中国中药杂志，2014，**39**(7)：1 220-1 224.
CHEN Y，LI C T，JIANG N H，*et al*. SSR information in *Erigeron breviscapus* transcriptome and polymorphism analysis[J]. *China Journal of Chinese Materia Medica*，2014，**39**(7)：1 220-1 224.

[31] 蒋　超，袁　媛，刘贵明，等. 基于 EST-SSR 的金银花分子鉴别方法研究[J]. 药学学报，2012，**47**(6)：803-810.
JIANG C，YUAN Y，LIU G M，*et al*. EST-SSR identification of *Lonicera japonica* Thunb. [J]. *Acta Pharmaceutica Sinica*，2012，**47**(6)：803-810.

[32] WEI W L，QI X Q，WANG L H，*et al*. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using Illumina paired-end sequencing and development of EST-SSR markers[J]. *BMC Genomics*，2011，**12**(1)：451.

[33] 李炎林，杨星星，张家银，等. 南方红豆杉转录组 SSR 挖掘及分子标记的研究[J]. 园艺学报，2014，**41**(4)：735-745.
LI Y L，YANG X X，ZHANG J Y，*et al*. Studies on SSR molecular markers based on transcriptome of *Taxus chinensis* var. *mairei*[J]. *Acta Horticulturae Sinica*，2014，**41**(4)：735-745.

[34] 贾新平，叶晓青，梁丽建，等. 基于高通量测序的海滨雀稗转录组学研究[J]. 草业学报，2014，**23**(6)：242-252.
JIA X P，YE X Q，LIANG L J，*et al*. Transcriptome characteristics of *Paspalum vagiatum* analyzed with Illumina sequencing technology[J]. *Acta Prataculturae Sinica*，2014，**23**(6)：242-252.

[35] MORGANTE M，HANAFEY M，POWELL W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes[J]. *Nature Genetics*，2002，**30**(2)：194-200.

[36] 杨　会，杨在君，魏淑红，等. 基于转录组序列的小麦 ETS-SSR 标记筛选与染色体定位[J]. 西华师范大学学报(自然科学版)，2014，**35**(4)：315-321.
YANG H，YANG Z J，WEI S H，*et al*. Screening and chromosomal localization of EST-SSR molecular markers based on transcriptome sequencing of wheat[J]. *Journal of China West Normal University* (Natural Sciences)，2014，**35**(4)：315-321.

[37] 陈相艳，李　伟，戴海英，等. 大豆 EST 资源的 SSR 信息分析[J]. 大豆科学，2009，**28**(3)：394-399.
CHEN X Y，LI W，DAI H Y，*et al*. Analysis of SSR information in EST resource of soybean (*Glycine max*)[J]. *Soybean Science*，2009，**28**(3)：394-399.

[38] BROOKES A J. The essence of SNPs [J]. *Gene*，1999，**234**(2)：177-186.

（编辑：宋亚珍）