

## 基于 Solexa 高通量测序的唐古特红景天 (*Rhodiola algida*) 微卫星信息分析

雷淑芸<sup>1,2</sup> 高庆波<sup>1</sup> 付鹏程<sup>1</sup> 杨慧玲<sup>1\*</sup> 陈世龙<sup>1</sup> 张发起<sup>1,3</sup>

(1. 中国科学院高原生物适应与进化重点实验室, 中国科学院西北高原生物研究所, 西宁 810001; 2. 中国科学院大学, 北京 100039; 3. 西南大学生命科学学院, 重庆 400715)

**摘要** 利用所获得的 Solexa 高通量唐古特红景天转录组拼接 EST 序列进行微卫星位点的挖掘分析, 期望为红景天属 SSR 标记的开发提供生物信息学依据。在得到的 6552 条 EST 序列中, 三碱基最多, 占总 EST 序列的 41.50%; 单核苷酸和二核苷酸重复类型的 SSR 含量相似, 分别为 27.76% 和 24.76%; 二至六碱基微卫星分布密度与其对应的 SSR 含量成正比。在单核苷酸重复类型中, T 和 A 重复类型最多, 分别为总 SSR 的 14.91%、12.70%, 而 G 和 C 重复类型则很少; 在二核苷酸重复类型中, AG 重复类型最多, 占总 SSR 的 5.60%, GA 和 TC 重复类型次之, 分别为 4.75%、4.72%; 在三核苷酸重复类型中, GAA 重复类型最多, 为总 SSR 的 1.85%, GAT 次之, 为 1.79%, TTC、TCT、TCA、GGA、GCT、GAG 重复类型间的 SSR 数相差不大; 四、五、六核苷酸重复类型则很少。除五、六核苷酸重复类型外, 其长度变化与其对应的重复类型碱基长度成反比; 同种重复类型中, 微卫星的长度与其对应的 SSR 数成反比。

**关键词** 唐古特红景天; 高通量测序; 微卫星

**中图分类号:** Q949.751.1 **文献标志码:** A **doi:** 10.7525/j.issn.1673-5102.2014.06.019

## Analysis on Microsatellites in *Rhodiola algida* Based on Solexa Sequencing

LEI Shu-Yun<sup>1,2</sup> GAO Qing-Bo<sup>1</sup> FU Peng-Cheng<sup>1</sup> YANG Hui-Ling<sup>1\*</sup> CHEN Shi-Long<sup>1</sup> ZHANG Fa-Qi<sup>1,3</sup>

(1. Key laboratory of Adaption and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810001; 2. University of Chinese Academy of Sciences, Beijing 100039; 3. School of Life Sciences, Southwest University, Chongqing 400715)

**Abstract** Solexa sequencing was used to analyze microsatellites in *Rhodiola algida* based on EST-assembled sequences, expecting to find out the genetic diversity and to develop microsatellite markers for *Rhodiola*. A total of 6552 EST sequences were obtained in the current study, trinucleotide repeats were the most with 41.5%, mononucleotide and dinucleotide repeats were almost similar with 27.76% and 24.76%, respectively. The distribution densities of dinucleotides repeats to hexanucleotides were directly proportional to the number of SSR number. In mononucleotides repeats, thymine (T) and adenine (A) accounted for 14.91% and 12.70%, respectively, while with less G and C repeats. In dinucleotide repeats, AG repeats were the dominated with 5.60%, followed by GA (4.75%) and TC (4.72%). In trinucleotides repeats, GAA were the most frequent, accounted for 1.85%, followed by GAT (1.79%), while TTC, TCT, TCA, GGA, GCT, GAG with similar composition. The tetranucleotide, pentanucleotide and hexanucleotide repeats were very few. Except for pentanucleotide and hexanucleotide repeats, the variation of repeats are inversely proportional to the length of repeated nucleotide, while on one type, the variation of repeats is in negative correlation.

**Key words** *Rhodiola algida*; solexa sequencing; microsatellite

红景天是景天科 (Crassulaceae) 红景天属 (*Rhodiola* L.) 的多年生草本或亚灌木植物, 属于天然配方的珍贵药用植物<sup>[1]</sup>。红景天主要分布于北半球的喜马拉雅山区、亚洲西北部和北美洲, 是

基金项目: 国家自然科学基金 (31270270); 中国科学院知识创新工程重要方向项目 (KSCX2-EW-J-26)

第一作者简介: 雷淑芸 (1989—), 女, 硕士研究生, 主要从事植物遗传多样性学研究。

\* 通信作者: E-mail: yanghuiling@nwipb.cas.cn

收稿日期: 2014-03-10

青藏高原地区特殊且重要的植物类群<sup>[2]</sup>。唐古特红景天 (*Rhodiola algida*) 是青藏高原特有的多年生草本植物,是高山冻原植被带高山草甸上的主要建群种。对红景天遗传多样性及进化历史的研究有助于了解青藏高原隆升和第四纪冰期气候波动对青藏高原植物演化历史的了解<sup>[3]</sup>。研究发现红景天第四纪冰期时在高原台面和东南部边缘存在多个相互隔离的避难所,并发生不同程度的异域分化<sup>[4-5]</sup>。与青藏高原地区其他植物类群的研究类似,目前已开展的红景天遗传多样性研究大多仅利用叶绿体 DNA 标记,而较少涉及到双亲遗传的核 DNA<sup>[6-10]</sup>,对植物演化历史的描述是不全面的<sup>[11]</sup>。

微卫星 (Microsatellite or Simple Sequence Repeats, SSRs) 作为一种共显性表达分子标记,具有数量多,在基因组内分布均匀,多态性信息丰富,易于检测等优点,广泛用于动植物遗传多样性分析、分子育种研究等<sup>[12-14]</sup>。早期微卫星的分离技术主要是小片段 DNA 克隆法,通过酶切构建小片段 DNA 文库后进行测序筛选<sup>[15]</sup>。由于该方法耗时费力及筛选效率较低,又提出利用滤膜、磁珠等构建微卫星富集文库,为微卫星标记的开发提供了一种快速有效的技术手段<sup>[16-17]</sup>。近年来,随着高通量测序技术的发展及成本降低,通过高通量测序获得海量 EST 序列。基于 EST (Expressed Sequence Tag) 序列大量开发 SSR 引物标记的方法正越来越引起人们的关注。由于该方法获得的 SSR 标记来源于基因表达序列,与传统方法获得的 SSR 标记相比,其具有可能与基因功能相关、种间通用性强、开发周期短等优点<sup>[18]</sup>。而目前关于红景天 EST 序列来源的 SSR 研究还鲜有报道。

因此,本研究在唐古特红景天 Solexa 高通量转录组测序并拼接组装获得的大量 EST 序列的基础上,充分挖掘微卫星信息,通过分析唐古特红景天微卫星重复序列特征,为红景天属 SSR 标记的开发提供生物信息学依据。

## 1 材料方法

### 1.1 Solexa 高通量测序及拼接组装

本研究所用唐古特红景天采于海拔 3 770 m 的青海省大通县小达板山高山草甸 (地理坐标: 37°13' 11.8"N, 101°46' 59.4"E), 唐古特红景天全株采集后迅速投入液氮中,带回实验室后于 -80℃ 保存。凭证标本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)。从野外采集的唐古特红景天材料 (包括茎、叶及花等混合材料) 中提取 100 μg 总 RNA。转录组测序方案按照华大基因 Illumina Solexa HiSeq™ 2000 操作流程进行。获得的原始读序 (Raw reads) 经过滤后得到干净读序 (Clean reads)。通过短读序组装软件 Trinity 得到单基因簇 (Unigene)<sup>[19]</sup>。用 Tgicl 将其去冗余和进一步拼接,然后再对这些序列进行同源转录本聚类,得到最终的 Unigene。将 Unigene 序列与蛋白数据库 NR、Swiss-Prot、KEGG 和 COG 做 blastx 比对 (evalue < 0.000 01), 取比对结果最好的蛋白确定 Unigene 的序列方向。如果不同库之间的比对结果有矛盾,则按 NR、Swiss - Prot、KEGG 和 COG 的优先级确定 Unigene 的序列方向,跟以上四个库皆比不上的 Unigene 我们用软件 ESTScan 确定序列的方向。对于能确定序列方向的 Unigene 我们给出其从 5' 到 3' 方向的序列,对于无法确定序列方向的 Unigene 我们给出组装软件得到的序列。

### 1.2 微卫星检测、筛选及统计分析

基于转录组的 SSR 检测是以组装出来 Unigene 作为参考序列,使用 MicroSatellite (MISA; <http://pgrc.ipk-gatersleben.de/misa/>), 来找出所有的 SSR。对所有 SSR 重复单元在 Unigene 上前后序列的长度进行筛选,只保留前后序列均不小于 150 bp 的 SSR。检索标准同时包括精确型 (perfect) 及复合型 (compound) SSR 重复单元<sup>[20]</sup>; 各重复微卫星类型重复次数设定如下: 两碱基 (dinucleotide repeats, DNRs) 至少重复 6 次, 三碱基 (trinucleotide repeats, TNRs) 至少重复 5 次, 四碱基 (tetranucleotide repeats, TTNRs) 至少重复 5 次, 五碱基 (pentanucleotide repeats, PTNRs) 至少重复 4 次, 六碱基 (hexanucleotide repeats, HXNRs) 至少重复 4 次。此外,本研究中还对至少重复 12 次的单核苷酸重复类型 (mononucleotide repeats, MNNRs) 进行统计。利用 Excel 软件统计微卫星的数量及发生频率、不同核苷酸重复类型的 SSR 数比较、所占比例及分布密度及不同长度重复类型微卫星变异等数据。

## 2 结果与分析

### 2.1 唐古特红景天微卫星特征

由唐古特红景天 Solexa 高通量测序共获得 61 889 230 个 Raw reads, 过滤后获得 53 956 910 个

Clean reads。通过短读序组装软件 Trinity 等处理后最终获得 83 395 个 Unigene, 总核苷酸数为 55 822 353 bp, Unigene 平均长 669 bp, N50 1 001 bp。经 MISA 软件发掘、excel 软件处理后,得到唐古特红景天序列的微卫星基本特征(表 1)。共检测到 SSRs 的数量为 6 552, 其中复合型 SSR 的数量为 247, 含有 SSR 的序列有 5 947 条, 且有 537 条序列中含有 SSR 的数量超过 1, 其总的微卫星序列发生率为 0.071, 平均跨度为 8 519.90 bp。

表 1 唐古特红景天序列中微卫星的数量及发生频率  
Table 1 Number and distribution of SSRs in *R. algida* sequences

特征 Character	数量 Number
检测序列总数 Total number of sequences examined	83 395
检测序列总长度 Total size by examined sequences (bp)	5 582 233
检测到的 SSRs 总数 Total number of SSRs identified	6 552
复合型 SSR 总数 Number of compound microsatellites	247
包含有 SSR 的序列数 Number of SSR - containing sequences	5 947
包含有超过 1 个 SSR 的序列总数 Number of sequences with more than one SSR	537
总的微卫星序列发生频率 Total relative abundance	0.071
微卫星序列平均跨度 Average distance (bp)	8 519.90

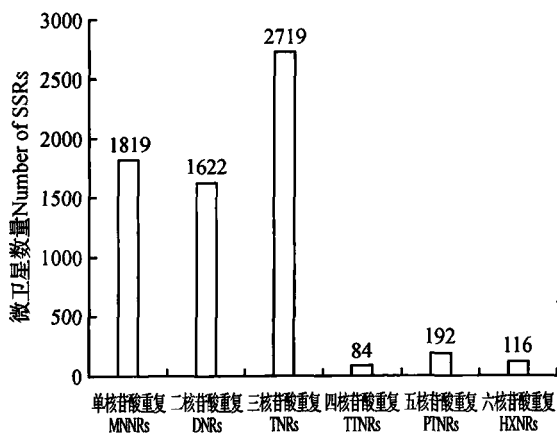


图 1 唐古特红景天不同核苷酸重复类型的 SSR 数比较  
Fig. 1 Number diversification of SSR in nucleotide repeats sequence of *R. algida*

2.2 唐古特红景天微卫星的丰度及分布密度分析  
在检测的唐古特红景天基因组 SSR 数据库中,以三核苷酸为重复类型的 SSR 含量最多,占总

SSR 的 41.50%; 单核苷酸和二核苷酸重复类型的 SSR 含量相差不大,分别为 27.76% 和 24.76%; 而四核苷酸、五核苷酸、六核苷酸重复类型的 SSR 含量很少,分别为 1.28%、2.93%、1.77% (图 1)。此 6 种核苷酸重复类型的分布密度与其对应的 SSR 含量成正比(表 2)。

表 2 不同长度重复单元微卫星所占比例及分布密度  
Table 2 Proportion and distribution density of nucleotide repeat types in SSR database

核苷酸重复类型 Nucleotide repeat type	SSR 数 No. of SSR	占总 SSR 的比例 Percent of total SSRs (%)	密度(个/Mb) Density of SSR
单核苷酸重复 Mononucleotide repeats	1 819	27.76	32.6
二核苷酸重复 Dinucleotide repeats	1 622	24.76	29.1
三核苷酸重复 Trinucleotide repeats	2 719	41.50	48.7
四核苷酸重复 Tetranucleotide repeats	84	1.28	1.5
五核苷酸重复 Pentanucleotide repeats	192	2.93	3.4
六核苷酸重复 Hexanucleotide repeats	116	1.77	2.1
总计 Total	6 552	100.00	117.4

唐古特红景天微卫星包含 12 种不同二核苷酸重复的微卫星,60 种不同三核苷酸重复的微卫星,51 种不同四核苷酸重复的微卫星,135 种不同五核苷酸重复的微卫星,101 种不同六核苷酸重复的微卫星。其中在单核苷酸重复类型中,T 和 A 重复类型所占比例较高,分别为总 SSR 的 14.91%、12.7%; 二核苷酸重复的微卫星中,AG 重复类型比例最高(5.6%); 三核苷酸重复的微卫星中,GAA(1.85%) 重复类型所占比例最高,而 GAT、TTC、TCT、TCA、GGA、GCT、GAG 重复类型相近; 四核苷酸、五核苷酸、六核苷酸重复类型的微卫星占总 SSR 的数量较少,因此未就其具体类型分析(表 3)。

2.3 唐古特红景天 EST 序列中微卫星长度分布及变异分析

微卫星的核心序列是 2~6 个核苷酸,通常微卫星核心序列的重复数与该座位的等位基因数有很强的正相关,序列重复次数越多,其变异性越大,则该微卫星座位的等位基因数就越多<sup>[12]</sup>。因此,了解序列中不同长度微卫星的分布及变化有利于多态性高的 SSR 标记的获得。

表 3 主要核苷酸重复类型在唐古特红景天序列中的百分比

Table 3 Relative percentage of SSRs in *R. algida* sequences

主要核苷酸重复类型 Nucleotide repeat composition	重复碱基类型 Nucleotide repeat type	SSR 数 No. of SSR	占总 SSR 的百分比 Percent of total SSRs(%)
单核苷酸重复 Mononucleotide repeats	T	977	14.91
	A	832	12.70
	G	7	0.11
	C	3	0.05
二核苷酸重复 Dinucleotide repeats	AG	367	5.60
	GA	311	4.75
	TC	309	4.72
	CT	242	3.69
	AT	127	1.94
	TA	88	1.34
	TG	72	1.10
	GT	40	0.61
	AC	33	0.50
	CA	31	0.47
三核苷酸重复 Trinucleotide repeats	GAA	121	1.85
	GAT	117	1.79
	TTC	107	1.63
	TCT	102	1.56
	TCA	101	1.54
	GGA	101	1.54
	GCT	98	1.50
	GAG	93	1.42
	TGA	82	1.25
	CTG	82	1.25
	CAG	76	1.16
	ATC	76	1.16
	AAG	71	1.08
	CCA	68	1.04
	TTG	66	1.01
	CCA	66	1.01
	CTT	66	1.01
TGC	65	0.99	
AGC	65	0.99	
CAT	62	0.95	
TCC	62	0.95	
CTC	62	0.95	

分析显示唐古特红景天序列微卫星的平均长度为 24.21 bp,最长的为 60 bp,最短的为 12 bp,长度大于 20 bp 的微卫星仅占总 SSR 的 15.42%;除五、六核苷酸重复类型外,其长度变化与其对应的重复类型的碱基长度成反比(图 2;每一扇形区对应不同长度的微卫星,而该长度的微卫星所占的比例用百分比标注在对应的扇形区内或用连线标注在对应的扇形区一侧,括号内为其对应的微卫星长度);同种重复类型中,微卫星的长度与其对应的 SSR 数成反比(表 4)。

### 3 讨论

通过对唐古特红景天转录组序列 SSR 的分析发现,其 SSR 的总发生频率为 0.071,平均跨度为 8519.90 bp,高于模式植物杨树基因组中的平均跨度(1 883 bp)<sup>[21]</sup>。多项研究指出植物的 SSR 以二、三核苷酸重复类型为主,其中占优势的重复类型在不同植物中存在差异。例如在 EST-SSR 引物开发研究中发现毛果杨、芝麻等双子叶植物二核苷

表 4 不同长度重复类型微卫星长度分布及变异

Table 4 Length diversification of the microsatellites in *R. algida* sequences of nucleotide repeat

核苷酸重复类型 Nucleotide repeat type	长度 Length(bp)	SSR 数量 No. of SSRs	SSR 所占百分比 Percent of total SSR(%)
单核苷酸重复 Mononucleotide repeats	12	634	34.85
	13	361	19.85
	14	230	12.64
	15	120	6.60
	16	90	4.95
	17	47	2.58
	18	35	1.92
	19	43	2.36
	20	75	4.12
	21	88	4.84
二核苷酸重复 Dinucleotide repeats	22	64	3.52
	23	32	1.76
	12	658	40.57
	14	384	23.67
	16	310	19.11
	18	153	9.43
	20	69	4.25
三核苷酸重复 Trinucleotide repeats	22	41	2.53
	24	5	0.31
	30	1	0.06
	42	1	0.06
	15	1 857	68.30
四核苷酸重复 Tetranucleotide repeats	18	620	22.80
	21	204	7.50
	24	33	1.21
	30	5	0.18
五核苷酸重复 Pentanucleotide repeats	20	71	84.52
	24	8	9.52
	28	3	3.57
六核苷酸重复 Hexanucleotide repeats	32	2	2.38
	20	167	86.98
	25	22	11.46
七核苷酸重复 Heptanucleotide repeats	30	2	1.04
	45	1	0.52
	24	94	81.03
	30	14	12.07
	36	6	5.17
	54	1	0.86
八核苷酸重复 Octanucleotide repeats	60	1	0.86

酸重复类型为主<sup>[22-24]</sup>,谷类作物中三核苷酸重复类型为主<sup>[25-26]</sup>,而白菜、油菜及黄瓜等双子叶植物显示二、三核苷酸重复类型所占百分比相近<sup>[27-29]</sup>。唐古特红景天转录组序列中微卫星重复类型以三核苷酸重复类型为主(48.7%),单核苷酸和二核苷酸重复类型相近,分别为 29.1%、32.6%。这与之前研究结果相一致,说明这种核苷酸的重复类型特征并非唐古特红景天所特有。此外,测序序列的数量、组装方法、注释途径均会对微卫星重复类型产生影响。因此,在对唐古特红景天 SSR 检测及分析过程中,因参数设定、物种差异等原因使其重复类型分布存在与前述研究不同之处。

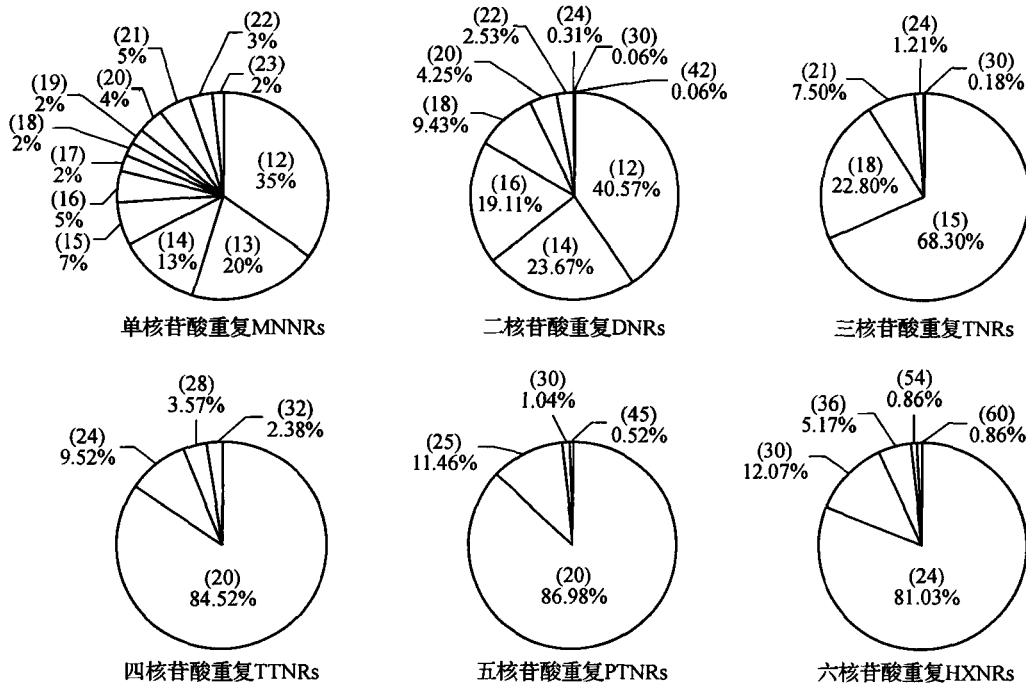


图 2 唐古特红景天序列不同长度重复类型微卫星变异

Fig. 2 Diversification of the microsatellites in *R. algida* sequences of nucleotide repeat

在唐古特红景天二核苷酸重复类型中,AG 碱基重复类型最多,占 5.60%,这与油茶、花旗松等 EST-SSR 序列的报道结果相一致<sup>[30-31]</sup>。三核苷酸重复类型中,各碱基重复类型所占百分比比较均匀,其中 GAA、GAT、TTC、TCT、TCA、GGA、GCT、GAG 等所占百分比相对较高,但相差不大。综观已报道研究结果,三碱基重复类型的优势分布密度在各物种中有较大的差异。在浙江山茶基因组序列 SSR 分布中,AAT 核苷酸重复类型分布密度最大<sup>[32]</sup>。通过比较分析发现水稻和大豆 SSR 的三核苷酸重复类型也不一致,CCC/GGC 在水稻中为最丰富类型,而大豆中则以 AAG 居多<sup>[25,33]</sup>。因此我们推断 GAA(或 GAT)这一重复单元可能是唐古特红景天或者红景天属 EST 序列的一个主要特征,但由于 SSR 位点的检测方法及设置标准的不同,即使在同一物种内部,也可能因为测序序列的数量,拼接方法及注释途径的不一致等,都可导致重复类型数量的差异。因此该特征是否成立还有待今后对红景天属植物的进一步研究。

此外,在多个研究中均提及 CG 重复单元可能与某些特定的功能相关,这可能是由于这些重复单元增加了 SSR 序列中的 GC 含量,从而导致某些氨基酸序列的增加,具有某些特定功能,如胁迫抗性、转录调控、信号传导等<sup>[21,34]</sup>。唐古特红景天

的转录组序列中也同样发现了少量的 AGC/GCT 重复单元,这些重复单元在红景天的生命活动中是否具有某些特定功能需要后期的进一步研究来证明。

综上所述,在唐古特红景天 Solexa 高通量转录组测序并拼接组装获得的大量 EST 序列的基础上,我们充分挖掘其微卫星信息,了解了唐古特红景天属微卫星重复序列特征。这一研究为今后红景天属 SSR 标记的开发提供生物信息学依据,可以更有针对性地去选择适合红景天属植物的 SSR 标记进行谱系地理等遗传多样性的研究,为阐述青藏高原隆升和第四纪冰期气候波动对青藏高原植物演化历史提供多方面的证据。

## 参 考 文 献

1. Wu Z Y, Raven P H, Hong D Y. Flora of China, Volume 8 [M]. Beijing: Science Press & Missouri: Missouri Botanical Garden Press, 2001.
2. 王强,阮晓,颜启传. 珍稀药用植物红景天[M]. 北京: 科学出版社, 2007.
3. Avise J C, Arnold J, Ball R M, et al. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics[J]. Annual review of ecology and systematics, 1987, 18: 489 - 522.
4. Gao Q B, Zhang D J, Chen S Y, et al. Chloroplast DNA phylogeography of *Rhodiola alsia* (Crassulaceae) in the Qinghai-Tibet Plateau[J]. Botany, 2009, 87(11): 1077 -

- 1088.
5. Gao Q B, Zhang D J, Duan Y Z, et al. Intraspecific divergences of *Rhodiola alsia* (Crassulaceae) based on plastid DNA and internal transcribed spacer fragments [J]. Botanical Journal of the Linnean Society, 2012, 168(2): 204 – 215.
  6. 段义忠, 张得钧, 高庆波, 等. 窄叶鲜卑花 (*Sibiraea angustata*) nrDNA ITS 和 cpDNA trnL-F 序列分子进化特点的分析 [J]. 植物研究, 2010, 30(2): 146 – 151.
  7. Yang F S, Li Y F, Ding X, et al. Extensive population expansion of *Pedicularis longiflora* (Orobanchaceae) on the Qinghai – Tibetan Plateau and its correlation with the Quaternary climate change [J]. Molecular Ecology, 2008, 17(23): 5135 – 5145.
  8. Zhang F Q, Gao Q B, Zhang D J, et al. Phylogeography of *Spiraea alpina* (Rosaceae) in the Qinghai-Tibetan Plateau inferred from chloroplast DNA sequence variations [J]. Journal of Systematics and Evolution, 2012, 50(4): 276 – 283.
  9. 于海彬, 张懿铨. 青藏高原及其周边地区高山植物谱系地理学研究进展 [J]. 西北植物学报, 2013, 33(6): 1268 – 1278.
  10. 张发起, 高庆波, 段义忠, 等. 横断山区高山绣线菊的谱系地理学研究 [J]. 广西植物, 2012, 32(5): 617 – 623.
  11. 王祎玲, 郭元涛, 钱增强, 等. 植物分子系统地理学及其研究进展 [J]. 西北植物学报, 2005, 25(6): 1250 – 1258.
  12. Wright J M, Bentzen P. Microsatellites: genetic markers for the future [M]. // Carvalho G R, Pitcher T J, et al. Molecular genetics in fisheries. London: Chapman & Hall Ltd, 1995: 117 – 121.
  13. Zane L, Bargelloni L, Patarnello T. Strategies for microsatellite isolation: a review [J]. Molecular ecology, 2002, 11(1): 1 – 16.
  14. Schrey A W, Heist E J. Microsatellite analysis of population structure in the shortfin mako (*Isurus oxyrinchus*) [J]. Canadian Journal of Fisheries and Aquatic Sciences, 2003, 60(6): 670 – 675.
  15. Rassmann K, Schlötterer C, Tautz D. Isolation of simple-sequence loci for use in polymerase chain reaction-based DNA fingerprinting [J]. Electrophoresis, 1991, 12(2/3): 113 – 118.
  16. Arranz S E, Avarre J C, Balasundaram C, et al. Permanent Genetic Resources added to Molecular Ecology Resources Database 1 December 2012 – 31 January 2013 [J]. Molecular ecology resources, 2013, 13: 546 – 549.
  17. Karagoyoz L, Kalcheva I D, Chapman V M. Construction of random small-insert genomic libraries highly enriched for simple sequence repeats [J]. Nucleic Acids Research, 1993, 21(16): 3911.
  18. Subramanian S, Mishra R K, Singh L. Genome-wide analysis of microsatellite repeats in humans: their abundance and density in specific genomic regions [J]. Genome Biol, 2003, 4(2): R13.
  19. 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.
  20. Weber J L. Informativeness of human (dC-dA)<sub>n</sub> (dG-dT)<sub>n</sub> polymorphisms [J]. Genomics, 1990, 7(4): 524 – 530.
  21. Li S, Yin T. 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.
  22. Tuskan G A, Gunter L E, Yang Z K, et al. Characterization of microsatellites revealed by genomic sequencing of *Populus trichocarpa* [J]. Canadian Journal of Forest Research, 2004, 34(1): 85 – 93.
  23. Kumpatla S P, Mukhopadhyay S. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species [J]. Genome, 2005, 48(6): 985 – 998.
  24. Wei W, Qi X, Wang L, 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.
  25. Cardle L, Ramsay L, Milbourne D, et al. Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. Genetics, 2000, 156(2): 847 – 854.
  26. Varshney R K, Thiel T, Stein N, et al. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species [J]. Cellular and Molecular Biology Letters, 2002, 7(2A): 537 – 546.
  27. 忻雅, 崔海瑞, 卢美贞, 等. 白菜 EST-SSR 信息分析与标记的建立 [J]. 园艺学报, 2006, 33(3): 549 – 554.
  28. 李小白, 张明龙, 崔海瑞. 油菜 EST 资源的 SSR 信息分析 [J]. 中国油料作物学报, 2007, 29(1): 20 – 25.
  29. 胡建斌, 李建吾. 黄瓜 EST-SSR 位点信息 [J]. 江西农业学报, 2008, 20(11): 58 – 60.
  30. 姜理平, 张磊, 鲍庆汉, 等. 浙江省在鼠类中检测到巴贝西原虫 DNA 片段 [J]. 中国媒介生物学及控制杂志, 2012, 23(4): 303 – 305.
  31. Amarasinghe V, Carlson J E. The development of microsatellite DNA markers for genetic analysis in Douglas-fir [J]. Canadian Journal of Forest Research, 2002, 32(11): 1904 – 1915.
  32. 史洁, 尹佟明, 管宏伟, 等. 油茶基因组微卫星特征分析 [J]. 南京林业大学学报: 自然科学版, 2012, 36(2): 47 – 51.
  33. Gao L, Tang J, Li H, et al. Analysis of microsatellites in major crops assessed by computational and experimental approaches [J]. Molecular Breeding, 2003, 12(3): 245 – 261.
  34. Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes [J]. Nature genetics, 2002, 30(2): 194 – 200.