

## 基于高通量测序分析青藏高原特有植物 蓝玉簪龙胆(*Gentiana veitchiorum*)的 SSR 和 SNP 特征

田尊哲<sup>1 2</sup> 高庆波<sup>1</sup> 陈世龙<sup>1</sup> 张发起<sup>1\*</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)

**摘要** 利用 Illumina HiSeq™2500 平台对青海省库泽县的蓝玉簪龙胆(*Gentiana veitchiorum*) 进行高通量测序, 共得到 SSR 序列 8 588 条, 对其 SSR 重复类型进行分析; 三核苷酸重复类型所占的比例最大占 54.7% (4696); 其次是二核苷酸重复类型占 41.3% (3543); 四核苷酸重复类型, 五核苷酸重复类型和六核苷酸重复类型所占的比例较少(共占 4%)。在二核苷酸重复类型中 AT/TA 重复类型所占的比例最大分别为 9.85% 和 9.5%。在微卫星中重复单元的长度大小和重复次数成负相关, 并且微卫星的总长度与重复单元的长度成正相关。在蓝玉簪龙胆的花(GP-F)和叶(GP-L)中分别得到 253 789 和 249 417 个 SNP 位点, 其中在非编码区上的比例为 51.29% 和 51.96%。分析发现蓝玉簪龙胆 SNP 位点在编码区中同义转换所占的比例(48.63% 和 47.96%) 要远远高于非同义转换的比例(0.08% 和 0.08%), 可能与功能基因序列相对稳定有关。

**关键词** 蓝玉簪龙胆; 高通量测序; 微卫星; 单核苷酸多态性; 青藏高原

中图分类号: Q949.776.4 文献标志码: A doi: 10.7525/j.issn.1673-5102.2016.05.016

## Characteristics of SSR and SNP in *Gentiana veitchiorum* in Qinghai-Tibetan Plateau by High-throughput Sequencing

TIAN Zun-Zhe<sup>1 2</sup> GAO Qing-Bo<sup>1</sup> CHEN Shi-Long<sup>1</sup> ZHANG Fa-Qi<sup>1\*</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)

**Abstract** We used the next generation sequencing (NGS) to capture the SSR and SNP marker in *Gentiana veitchiorum*, sampled in Zeku, Qinghai Province. A total of 8588 simple sequence repeats were generated through lumina HiSeq™ 2500. The dinucleotide repeats were the highest (54.7%), followed by the trinucleotide repeats (41.3%). A proportion of tetranucleotide, pentanucleotide and hexanucleotide repeats was less, which is 4% in total. In dinucleotide repeats, AT/TA repeats were the dominated ones, 9.85% and 9.5%, respectively. The length of nucleotide repeat and number of repeat were negative correlation, while there was a positive correlation between the total length of SSR sequences and the length of nucleotide repeat. A total of 253 789 and 249 417 SNPs were identified in GP-F and GP-L of *G. veitchiorum*. The proportion of the SNPs located in the noncoding region were 51.29% and 51.96%, respectively. The proportion of synonymous transition (48.63% and 47.96%) in coding region was significantly higher than that of nonsynonymous transition (0.08% and 0.08%), which might be caused by relatively conservative domains of functional genes.

**Key words** *Gentiana veitchiorum*; next generation sequencing; SSR; SNP; Qinghai-Tibetan Plateau

基金项目: 国家自然科学基金(31400322); 青海省国际科技合作项目(2014-HZ-812)

第一作者简介: 田尊哲(1989—) 男, 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: fqzhang@nwipb.cas.cn

收稿日期: 2016-03-31

Foundation item: National Natural Science Foundation of China(31400322); International S&T cooperation projects of Qinghai Province(2014-HZ-812)

First author introduction: TIAN Zun-Zhe(1989—) male, master degree, mainly engaged in the research of plant genetic diversity.

\* Corresponding author: E-mail: fqzhang@nwipb.cas.cn

Received date: 2016-03-31

青藏高原(Qinghai-Tibetan Plateau)是世界上海拔最高的高原,平均海拔达到4 500 m,面积达到 $2.5 \times 10^6 \text{ km}^2$ ,被誉为“世界屋脊”、“地球的第三极”<sup>[1]</sup>。蓝玉簪龙胆(*Gentiana veitchiorum* Hemsl.)为龙胆科(Gentianaceae)龙胆属(*Gentiana*)的多年生草本植物,在青藏高原地区广泛分布,高5~10 cm,营养叶和繁殖叶异型。龙胆花主要含有裂环烯醚类活性成分,如:落甘酸、獐芽菜苦苷、龙胆苦苷和獐牙菜苷等,其叶具有清湿热、泻肝胆湿火、镇咳利喉健胃的功能,主治感冒发烧目赤咽痛肺热咳嗽等<sup>[2]</sup>。对蓝玉簪龙胆的研究主要集中在化学成分以及药效等方面<sup>[3~5]</sup>,而在分子遗传等方面研究较少。青藏高原的温度从1960年就开始升高,这比中国其他地方的地区都要早<sup>[7~8]</sup>,并且在过去的50年中,青藏高原每十年温度就要增高0.3℃,比中国其他地区的温度都要上升的快<sup>[5]</sup>。青藏高原的环境变化对生物多样性有了很大的影响,为了更加深入的挖掘蓝玉簪龙胆资源和建立完善的种质资源评估和保护系统,因此很有必要开展蓝玉簪龙胆的遗传多样性研究。

近年来,随着分子技术的快速发展,尤其是第二代微卫星标记和第三代单核苷酸标记在探究生物群体内和群体间遗传变异及种间关系和遗传育种等研究中所凸显的优越性,使其运用范围越来越广。微卫星,又称简单重复序列标记(simple sequence repeats, SSRs),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,长度较短,且广泛分布于真核生物的基因组中,具有数量多,分布广且均匀,高度多样性,分析快速方便等优点<sup>[9~12]</sup>。此外,微卫星序列在群体中通常具有很高的多态性,而且一般为共显性,这些特点使得微卫星标记成为分子遗传研究中使用最为广泛的遗传标记之一。

单核苷酸多态性(Single Nucleotide Polymorphism, SNP)是指一物种不同个体在基因组水平上有单个核苷酸变异引起的DNA序列多态性,通常被认为是变异在频率大于1%时称为SNP<sup>[13]</sup>。SNP一般只涉及到碱基的转换(Transition)和颠换(Transversion),是许多真核生物中最丰富的遗传变异方式,具有数量多,分布广,突变率低,易实现自动化检测等优点<sup>[14]</sup>。

本研究以选取采集于青海泽库蓝玉簪龙胆的叶和花开展高通量测序,挖掘蓝玉簪龙胆的SSR和SNP信息,并分析其在该物种的特征,为蓝玉簪

龙胆的遗传多样性以及群体遗传学的研究奠定了分子基础。

## 1 样品的采集和方法

### 1.1 样品的采集和高通量测序、拼接

蓝玉簪龙胆(*G. veitchiorum* Hemsl.; Zhang 2014131)采集于青海省泽库县(35°4'N, 101°30'E, 3 681 m)。采取同一株上的叶片和花后放入液氮中保存,带回实验室于-80℃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)。

从蓝玉簪龙胆的叶(GP-L)和花(GP-F)中各提取100 μg总RNA;用Nanodrop检测RNA的纯度,用Qubit对RNA浓度进行精确定量,并用Agilent 2100精确检测RNA的完整性;然后用Oligo(dT)的磁珠富集蓝玉簪龙胆的mRNA;随后加入fragmentation buffer将mRNA打断成短片段,以mRNA为模板,用六碱基随机引物(random hexamers)合成一链cDNA,然后加入缓冲液、dNTPs和DNA polymerase I和RNase H合成双链cDNA,再用AMPure XP beads纯化双链cDNA。纯化的双链cDNA先进行末端修复,加A尾并连接测序接头,再用AMPure XP beads进行片段大小选择。最后进行PCR扩增,并用AMPure XP beads纯化PCR产物,得到最终的文库。文库合格后,把不同文库按照有效浓度及目标下机数据量的需求混池后在Illumina HiSeq™ 2500平台上测序。

对所得的数据进行质量评估,包括:测序错误率分布检查,A/T/G/C的含量分布检查,其中对单个碱基位置的错误率控制在1%以下,然后对测得的原始测序序列(raw reads)进行过滤:去除带接头(adapter)的和N的比例大于10%的reads,并去除低质量reads,得到干净的读序(clean reads)后,而后用Trinity<sup>[15]</sup>对得到的clean reads进行拼接以获取后续分析的参考序列,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的检测、筛选和统计分析

基于蓝玉簪龙胆转录组的SSR检测是以组装出来的Unigene作为参考序列,采用MicroSatellite(MISA; <http://pgra.ipk-gatersleben.de/misa/>)对Unigene进行SSR检测。SSR搜索标准有精确型(perfect)及复合型(compound)的SSR重复单元<sup>[16]</sup>,各重复微卫星类型重复次数设定为: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) 至少重复 4 次。并对蓝玉簪龙胆转录本的不同 SSR 类型以及 SSR 类型中密度进行统计分析。

通过 samtools 和 piacard-tools 等工具对比对结果进行染色体坐标排序, 去掉重复的 reads 等处理, 最后通过变异检测软件 GATK2<sup>[17]</sup> 以 Unigene 为参考序列对 reads 进行 SNP Calling, 并对原始结果进行过滤(过滤掉质量值小于 30, 距离小于 5 的 SNP)。并对得到的 SNP 位点位于编码区或者非编码区, 以及在编码区中的同义转换或者非同义转换的 SNP 数量进行统计。

## 2 结果与分析

### 2.1 测序产量, 质控和组装结果

随着高通量测序技术的发展及成本的降低, 通过高通量测序可以得到海量的序列, 但是在 RNA-seq 技术中, 测序错误率会随着测序序列长度的增加而升高<sup>[18-19]</sup>, 在单个碱基位置的测序错误率控制在 1% 以下, 在 GP-F 和 GP-L 的单碱基错误率都为 0.03%; 在由于测序得到的 reads 并不都是有效的, 里面含有带接头的, 重复的和测序质量低的 reads, 它们对组装及后续分析都会产生影响。在对 GP-F 和 GP-L 的转录本测得的 raw reads 去除含接头的、N 的比例大于 10% 的以及低质量的 reads 后, 分别得到 22 936 570 条和 22 355 444 条 clean reads, 分别占 Raw reads 的 92.80% 和 92.59%, 在两个 GP-F 和 GP-L 中质量值 > 20% 的碱基数占总碱基数的比例分别为 95.72% 和 95.96%, 质量值 > 30% 的碱基数占总碱基数的比例分别为 91.81% 和 92.14%, 碱基 G 和 C 含量总和分别为 43.23% 和 42.79% 以上。

在对 GP-F 和 GP-L 的 raw reads 进行处理后, 我们用组装软件 Trinity 对所得的 clean reads 进行组装, 对转录本和 Unigene 的长度进行统计(图 1), 其中在拼接得到的 Transcripts 中 N50 为 1 424, 而 N90 为 314; 在拼接得到的 Unigene 中 N50 为 910, 而 N90 为 256。其中拼接的 Transcripts 和 Unigene 总的核苷酸数分别为 177 300 983 个和 85 642 977 个。

### 2.2 SSR 的测序结果, 测序产量、质量及组装结果分析

在 Illumina HiSeq<sup>TM</sup> 2500 测序平台对蓝玉簪

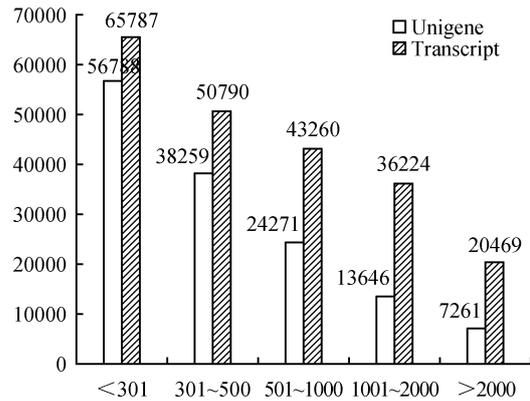


图 1 拼接后的 Unigene 与 Transcript 长度分布图  
Fig. 1 The length distribution of Unigene and Transcript after assembled

龙胆测序后, 去除 Raw reads 中含有接头的, N 含量比大于 10% 的和低质量的 reads 后, 用 Trinity 对 clean reads 进行拼接。在对 Trinity 拼接得到的转录序列为参考序列, 取每条基因中最长的转录本作为 Unigene。通过对 Unigene 进行 SSR 检测, 共检测的序列共 140 225 条, 总长度达到 85 642 977 bp, 共检测到 8 588 条 SSR 序。

通过对 SSR 类型进行分析发现: 双核苷酸重复类型 SSR 占 41.3% (3 543 个), 长度以 12 ~ 24 bp 为主; 三核苷酸重复类型 SSR 占 54.7% (4 696 个), 长度以 15 ~ 24 bp 为主; 四核苷酸重复类型 SSR 占 3.4% (296 个), 长度以 20 ~ 24 bp 为主; 五核苷酸重复类型和六核苷酸重复类型 SSR 较少(图 2)。对其进一步统计分析得到二核苷酸重复类型有 12 种, 三核苷酸重复类型有 60 种, 四核苷酸重复类型有 92 种, 五核苷酸重复类型 21 种, 六核苷酸重复类型有 32 种(图 3)。

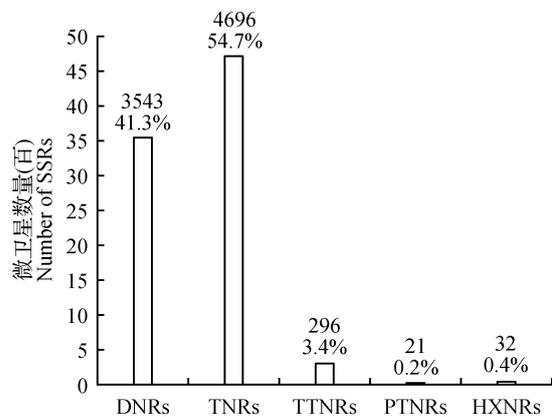


图 2 蓝玉簪龙胆不同核苷酸重复类型的 SSR 数量  
Fig. 2 Number diversification of SSR in nucleotide repeat sequence of *G. veitchiorum*

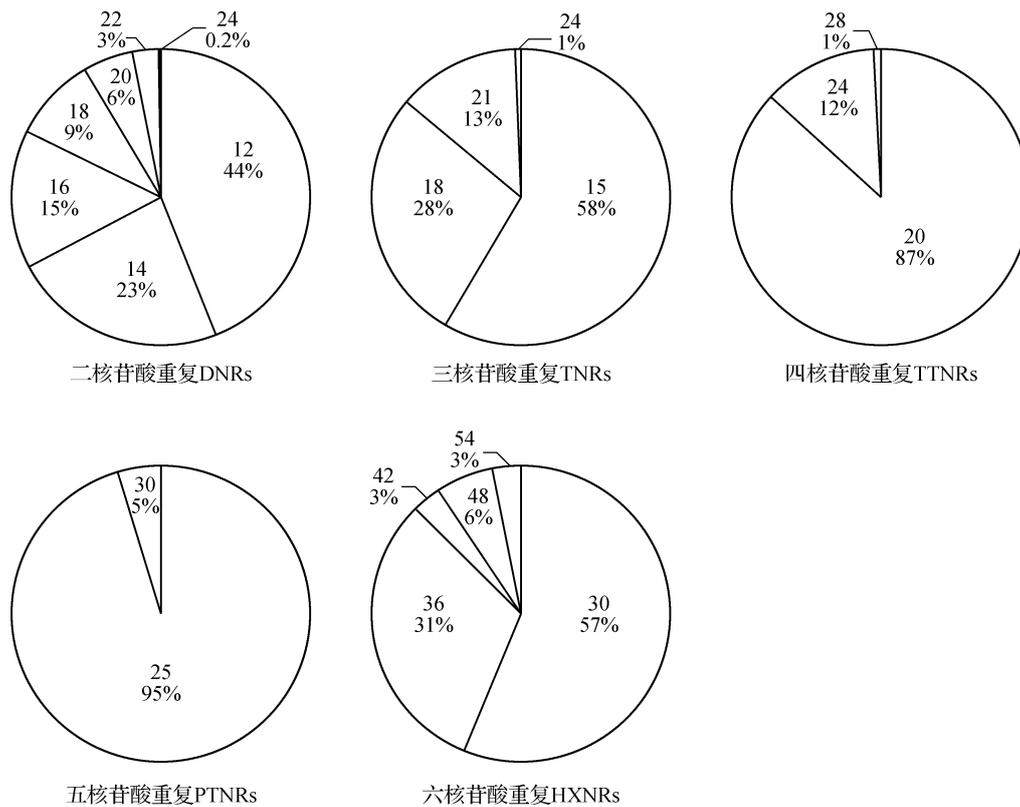


图 3 蓝玉簪龙胆序列不同长度重复类型微卫星长度分布及其变异 每一扇形区对应不同长度微卫星所占百分比,上面的值为对应的微卫星片段长度。

Fig. 3 Diversification of the microsatellites in *G. veitchiorum* sequence of nucleotide repeat Each sector represents the percent of different SSR, the number is the length of the SSR.

通过对主要的核苷酸重复类型的比较分析发现,在二核苷酸重复类型中,AT/TA 重复类型所占比例较高,AC/TG 次之;在三核苷酸重复类型中 GAA 重复类型所占比例较高,AGA/TCT 重复类型所占比例次之;四核苷酸重复类型、五核苷酸重复类型、六核苷酸重复类型的数量相对较少(表 1)。

### 2.3 蓝玉簪龙胆的 SNP 位点特征分析

通过 samtools 和 picard-tools 等工具比对 GP-F 和 GP-L 结果进行染色体坐标排序,去掉重复的 reads 等处理后,以 Unigene 序列为参考序列,通过变异检测软件 GATK2 进行 SNP Calling,并对原始结果进行过滤,在 GP-F 和 GP-L 中分别得到 253 789 和 249 417 个 SNP 位点,其中非编码区的 SNP 位点分别为 130 175 个和 129 593 个,编码区的 SNP 位点分别为 123 614 个和 119 824 个。其中在 GP-F 中编码 SNP 中,同义转换占 48.63% (123 431 个),非同义转换占 0.08% (201 个);在 GP-L 中的编码区的

SNP 中,同义转换占 47.96% (119 620),非同义转换占 0.08% (201 个;表 2)。对其进一步分析,GP-F 和 GP-L 的每个 SNP 的平均跨度分别为 699 和 711 bp。在 GP-F 和 GP-L 中的 SNP 的位点个数和跨度出现此差别表明在蓝玉簪龙胆在适应环境变化过程中,花的变异量强于叶的变异量,或者可能是由于花所受环境变化的影响更大。

对蓝玉簪龙胆 SNP 类型分析发现,在 GP-F 中转换类型占 58% (147 198 个),颠换类型占 42% (110 180 个);在 GP-L 中转换类型占 61% (52 144 个),颠换类型占 39% (97 272 个,表 3)。其中在 GP-F 和 GP-L 中的转换类型明显高于颠换类型,在 GP-F 和 GP-L 中的转换类型中,C <-> T 发生频率较高,这可能与 SNPs 在 CG 序列上出现的最为频繁,而 C(胞嘧啶)常以甲基化形式存在,在脱氨后即成为 T(胸腺嘧啶)有关<sup>[20]</sup>,在蓝玉簪龙胆的花和叶中的 SNPs 类型及其发生频率趋势基本一致。

表 1 主要核苷酸重复序列类型在蓝玉簪龙胆序列的百分比

Table 1 Relative percentage of SSRs in *G. veitchiorum* sequences

主要的重复单元 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( %)
二核苷酸重复 Dinucleotide repeats	AC	320	3.73
	AG	247	2.88
	AT	846	9.85
	CA	203	2.36
	CT	151	1.76
	GA	187	2.18
	GT	241	2.81
	TA	815	9.50
	TC	200	2.33
	TG	327	3.81
三核苷酸重复 Trinucleotide repeats	AAC	101	1.18
	AAG	161	1.87
	AAT	118	1.37
	AGA	154	1.79
	ATC	103	1.20
	CAA	100	1.16
	CCA	109	1.27
	GGA	107	1.25
	GGT	141	1.64
	TCA	121	1.41
	TCT	153	1.78
	TGA	108	1.26
	TGG	122	1.42
	TTC	161	1.87
TTG	129	1.50	

表 2 高通量测序鉴定的蓝玉簪龙胆的 SNPs 类型分类

Table 2 Identified the SNPs type of *G. veitchiorum* by high-throughput sequencing

样品 Sample	总的 SNP 位点数 Total no. of SNP	非编码区 SNP No-coding SNP	编码区 SNP Coding SNP	同义转换 Synonymous transition	非同义转换 Nonsynonymous transition
GP-F	253 789 (100%)	130 175 (51.29%)	123 614 (48.71%)	123 413 (48.63%)	201 (0.08%)
GP-L	249 417 (100%)	129 593 (51.96%)	119 824 (48.04%)	119 620 (47.96%)	201 (0.08%)

表 3 高通量测序鉴定的蓝玉簪龙胆 SNPs 类型分析

Table 3 Analysis the SNPs type of *G. veitchiorum* by high-throughput sequencing

样品 Sample	SNP 类型 SNP type	数量 Number	SNP 类型 SNP type	数量 Number
GP-F	转换 transition		颠换 transversion	
	A < - > G	73 590	A < - > T	37 596
	C < - > T	73 608	G < - > T	21 064
	—	—	C < - > G	14 974
	—	—	A < - > C	36 546
	合计	147 198(58%)	合计	110 180(42%)
GP-L	A < - > G	77 319	A < - > T	34 918
	C < - > T	74 825	G < - > T	22 448
	—	—	C < - > G	16 960
	—	—	A < - > C	22 946
		合计	152 144(61%)	合计

### 3 讨论

通过 Illumina HiSeq™ 2500 测序平台对蓝玉簪龙胆转录组测序共得到 8 588 条 SSR 序列,平均跨度为 9 973 bp,在蓝玉簪龙胆基因组中的 SSR 类型中,三核苷酸重复类型所占的比例最大(54.7%),其次是二核苷酸重复类型(41.3%),而在同属于龙胆亚族的川西獐牙菜(*Swertia mussotii*)的 SSR 的平均跨度为 12.6 kB,在 SSR 重复类型中三碱基重复所占的比例为 45.99%,二碱基重复类型所占的比例为 41.62%<sup>[21]</sup>。其中三碱基重复类型的比例相差比较大,此结果说明即使在同一亚族中不同属之间的 SSR 类型分布比例也有很大的区别。

在二、三碱基重复类型中占比例最高的碱基重复类型分别是 AT/TA 和 AAG/TTC,在川西獐牙菜中二、三碱基重复类型中比例最高的碱基重复类型分别为 AT/TA 和 AAT/TTA<sup>[21]</sup>,但是在唐古特红景天(*Rhodiola algida*)中的二碱基重复类型中比例较高的碱基重复类型为 AG/GA 和 TC/CT,三碱基重复类型中各个重复类型所占的比例较均匀<sup>[22]</sup>。蓝玉簪龙胆中二、三碱基重复类型比例较高的重复类型与红景天的差别比较大,而与川西獐牙菜的差别没那么大,这可能与不同植物之间的亲缘关系有关。对蓝玉簪龙胆的微卫星长度进行分析可以得出,在微卫星中重复单元的长度大小和重复次数成负相关,并且微卫星的总长度与重复单元的长度成正相关。

通过对蓝玉簪龙胆的 SNP 的分析发现,GP-F 含有 253 789 个 SNPs 位点,平均跨度 699 bp,GP-L 含有 SNPs 249 417 个,平均跨度 711 bp,这比青杨(*Populus cathayana*, 1/29 332 bp)的分布密度要高,这可能与物种、测序数量、SNP 分布的不均一性有关<sup>[23]</sup>。蓝玉簪龙胆的 SNP 位点非编码区所占的比例要比在编码区的比例高一些,和以往的研究结果一致<sup>[24~25]</sup>,可能是由于进化的原因和自然选择所致。蓝玉簪龙胆的 SNP 位点在编码区中同义变换所占的比例(48.63%和 47.96%)要远远大于非同义变换的比例(0.08%和 0.08%),可能是由于功能基因序列相对稳定所造成的。

综上所述,我们在高通量转录组测序的基础上,充分挖掘了蓝玉簪龙胆 SSR、SNP 信息,为构建蓝玉簪龙胆遗传学图谱以及谱系地理学研究提供了多方面的分子基础;该研究所获得 SSR 以及 SNP 分子标记,还为以后龙胆属植物的生物多样性研究以及保护生物学的研究提供了基础。

## 参 考 文 献

- Zheng D. The system of physico-geographical regions of the Qinghai-Xizang (Tibet) Plateau [J]. Science in China (Series D) ,1996 ,39( 4) : 410 - 417.
- 杨红澎 确生 吴锡冬 等. 蓝玉簪龙胆中苷类成分的研究 [J]. 中国中药杂志 2008 ,33( 21) : 2505 - 2507.  
Yang H P ,Que S ,Wu X D ,et al. Studies on glycosides from *Gentiana veitchiorum* [J]. China Journal of Chinese Materia Medica 2008 ,33( 21) : 2505 - 2507.
- Faver A ,Yuan Y M ,Küpfer P ,et al. Phylogeny of subtribe *Gentianinae*( Gentianaceae) : biogeographic inferences despite limitations in temporal calibration points [J]. Taxon , 2010 ,59( 6) : 1701 - 1711.
- Zhang Z F ,Yuan L ,Lu L Y ,et al. Hepatoprotective activity of *Gentiana veitchiorum* Hemsl. against carbon tetrachloride-induced hepatotoxicity in mice [J]. Chinese Journal of Natural Medicines 2014 ,12( 7) : 488 - 494.
- Mishiba K I ,Yamane K ,Nakatsuka T ,et al. Genetic relationships in the genus *Gentiana* based on chloroplast DNA sequence data and nuclear DNA content [J]. Breeding Science 2009 ,59( 2) : 119 - 127.
- Feng S ,Tang M C ,Wang D M. New evidence for the Qinghai-Xizang( Tibet) Plateau as a pilot region of climatic fluctuation in China [J]. Chinese Science Bulletin ,1998 , 43( 20) : 1745 - 1749.
- 蔡英 李栋梁 汤懋苍 等. 青藏高原近 50 年来气温的年代际变化 [J]. 高原气象 2003 ,22( 5) : 464 - 470.  
Cai Y ,Li D L ,Tang M C ,et al. Decadal temperature changes over Qinghai-Xizang Plateau in recent 50 years [J]. Plateau Meteorology 2003 ,22( 5) : 464 - 470.
- Piao S L ,Ciais P ,Huang Y ,et al. The impacts of climate change on water resources and agriculture in China [J]. Nature 2010 ,467( 7311) : 43 - 51.
- Ellegren H. Microsatellites: simple sequences with complex evolution [J]. Nature Reviews Genetics 2004 ,5( 6) : 435 - 445.
- Goldstein D B ,Schlotterer C. Microsatellites: evolution and applications [M]. Oxford ,England: Oxford University Press , 1999.
- Wright J M ,Bentzen P. Microsatellites: genetic markers for the future [M]. Carvalho G R ,Pitcher T J. Molecular genetics in fisheries. Netherlands: Springer ,1995: 117 - 121.
- Zane L ,Bargelloni L ,Paternello T. Strategies for microsatellite isolation: a review [J]. Molecular Ecology 2002 ,11( 1) : 1 - 16.
- Vignal A ,Milan D ,Sanristobal M ,et al. A review on SNP and other types of molecular markers and their use in animal genetics [J]. Genetics Selection Evolution ,2002 ,34( 3) : 275 - 305.
- 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.
- 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.
- 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.
- Mckenna A ,Hanna M ,Banks E ,et al. The genome analysis toolkit: a Map Reduce framework for analyzing next-generation DNA sequencing data [J]. Genome Research 2010 ,20( 9) : 1297 - 1303.
- Erich 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.
- Jiang L H ,Schlesinger F ,Davis C A ,et al. Synthetic spike-in standards for RNA-seq experiments [J]. Genome Research 2011 ,21( 9) : 1543 - 1551.
- Brookes A J. The essence of SNPs [J]. Gene ,1999 ,234( 2) : 177 - 186.
- 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) : 2068 - 2076.  
Liu Y ,Yue C J ,Wang Y ,et al. Data mining of simple sequence repeats in transcriptome sequences of Tibetan medicinal plant Zangyinchen *Swertia mussoitii* [J]. China Journal of Chinese Materia Medica ,2015 ,40( 11) : 2068 - 2076.
- 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( *Rhodiola algida*) 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.
- 雷淑云 张发起 ,Gulzar K 等. 利用高通量测序分析青藏高原地区青杨的 SSR 和 SNP 特征 [J]. 林业科学研究 2015 ,28( 1) : 37 - 43.  
Lei S Y ,Zhang F Q ,Khan G ,et al. Characteristic analysis of SSR and SNP in *Populus cathayana* on Qinghai-Tibetan Plateau by High-Throughput Sequencing [J]. Forest Research 2015 ,28( 1) : 37 - 43.
- Lijavetzky D ,Cabezas J ,Ibáñez A ,et al. High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology [J]. BMC Genomics 2007 ,8( 1) : 424 - 435.
- An C F ,Saha S ,Jenkins J N ,et al. Cotton (*Gossypium* spp.) R2R3-MYB transcription factors SNP identification , phylogenomic characterization , chromosome localization , and linkage mapping [J]. Theoretical and Applied Genetics 2008 ,116( 7) : 1015 - 1026.