

野三七转录组中 SSR 位点信息分析及其多态性研究

李翠婷, 张广辉, 马春花, 孟珍贵, 陈军文, 杨生超*

云南农业大学 云南省优势中药材规范化种植工程研究中心, 云南 昆明 650201

摘要: 目的 分析野三七 *Panax vietnamensis* var. *fuscidiscus* (PVF) 转录组中的 SSR 位点信息, 并设计简单重复序列 (SSR) 引物, 以期对野三七分子标记辅助育种提供有力工具。方法 利用 MISA 工具筛选了野三七转录组测序获得的 126 758 条 *unigenes*, 对其 SSR 位点信息进行了分析; 在此基础上利用 Primer3 设计 SSR 引物, 并随机选择 30 对 SSR 引物对 13 株不同来源的野三七进行多态性扩增分析。结果 在野三七的转录组中, 共搜索到 21 320 个 SSRs, 分布于 17 780 条 *unigenes*, SSR 位点出现频率为 16.82%。SSRs 位点中二核苷酸重复是主要类型, 占总 SSRs 的 52.52%; 其次是三核苷酸重复基序 (28.08%)。SSRs 所包含的重复基元中, 二核苷酸重复基元 AG/CT 和 AT/AT 是优势重复基元类型, 占总 SSRs 的 46.25%。利用 Primer3 共设计出 39 336 对 SSR 引物。随机选择 30 对引物进行 PCR 扩增, 其中 29 对 (96.67%) 扩增出清晰、可重复的条带, 15 对 (50.00%) 扩增条带表现出多态性。利用 UPGMA 作图, 将 13 个材料分为 2 类。结论 野三七转录组中 SSR 位点出现频率高, 类型丰富; 大量的 SSR 为其遗传多样性分析和遗传图谱构建提供了丰富的候选分子标记。

关键词: 野三七; 转录组; SSR 信息; 多态性; 遗传多样性

中图分类号: R282.12

文献标志码: A

文章编号: 0253-2670(2014)10-1468-05

DOI: 10.7501/j.issn.0253-2670.2014.10.021

Analysis on SSR loci information in transcriptome of *Panax vietnamensis* var. *fuscidiscus* and its polymorphism

LI Cui-ting, ZHANG Guang-hui, MA Chun-hua, MENG Zhen-gui, CHEN Jun-wen, YANG Sheng-chao

Yunnan Research Center on Good Agriculture Practice for Dominant Chinese Medicinal Materials, Yunnan Agriculture University, Kunming, 650201, China

Abstract: Objective The SSR loci information in the transcriptome of *Panax vietnamensis* var. *fuscidiscus* (PVF) was analyzed in this study to provide more powerful tools for molecular marker-assisted breeding in this plant. **Methods** Simple sequence repeats (SSR) loci were searched in all 126 758 *unigenes* by using MISA. SSR loci information was analyzed and SSR primers were designed by Primer3. Furthermore, 30 pairs of primers were randomly selected for the polymorphic analysis on 13 PVF plants collected from different habitats. **Results** A total of 21 320 SSRs were found in the transcriptome of PVF, distributed in 17 780 *unigenes* with the distribution frequency of 16.82%. Di-nucleotide repeat was the main type, accounted for as much as 52.52% of all SSRs, followed by tri-nucleotide repeat motif (28.08%). The dinucleotide repeat motifs of AG/CT and AT/AT were the predominant repeat types (46.25%). Using Primer3, a total of 39 336 pairs of SSR primers were designed. For validating the availability of those SSR primers, we randomly selected 30 pairs of primers for PCR amplification. Among them, 29 pairs of primers (96.67%) produced clear and reproductive bands, 15 pairs of primers (50.00%) showed polymorphism, and 13 PVF plants were divided into two groups by UPGMA. **Conclusion** There are numerous SSRs in PVF transcriptome with high frequency and various types, this will provide the abundant candidate molecular markers for genetic diversity study and genetic map for this plant.

Key words: *Panax vietnamensis* var. *fuscidiscus* K.Komatsu, S. Zhu & S.Q.Cai; transcriptome; SSR information; polymorphism; genetic diversity

野三七 *Panax vietnamensis* var. *fuscidiscus* K. Komatsu, S. Zhu & S. Q. Cai 是越南人参 *P. vietnamensis*

Ha et Grushv. 的一个变种, 为五加科人参属植物, 主要分布于我国云南省南部的红河州以及老挝等国家。

收稿日期: 2013-12-23

基金项目: 国家自然科学基金项目 (81303164); 国家“十二五”支撑计划课题 (2011BAI13B01)

作者简介: 李翠婷 (1988—), 女, 河南洛阳人, 在读硕士研究生, 从事药用植物资源评价与利用的研究。

Tel: 18213012327 E-mail: xiangfei8880118@163.com

*通信作者 杨生超 Tel: (0871)5227609 E-mail: shengchaoyang@163.com

最早由 Zhu 等^[1]在云南省金平县境内发现并命名,具有活血、化瘀和止血等功效,用于治疗肿痛、伤痛和月经不调等^[2]。野三七含有大量的奥克梯隆型皂苷珠子参皂苷 R_2 ^[3-4], 以及三七皂苷 R_1 , 人参皂苷 R_{g1} 、 R_{b1} 、 R_d 和 R_e 等^[5-10]。近年来研究发现,野三七也分布在红河州的元阳县、绿春县等地。民间也有从老挝等国家高价购买野三七销售,或者作为种源人工种植。一些野三七植株生长达 15 年以上,高温和多雨的环境下表现出明显的抗病性。因此,野三七是三七 *P. notoginseng* (Burk.) F. H. Chen 抗病育种的优良野生种质资源之一。

DNA 分子标记是鉴定种质资源遗传多样性的重要手段,也是一种重要的辅助育种方法^[11]。简单重复序列 (simple sequence repeats, SSR), 一般以 1~6 个碱基为核心序列,具有高度多态性。目前已利用 SSR 标记构建了许多物种的染色体遗传图谱,并被广泛应用于基因定位及亲缘关系分析、品种鉴定和动植物育种等领域^[12]。本研究分析了野三七转录组中的 SSR 位点,并随机验证了 SSR 引物的有效性,以期野三七分子标记辅助育种提供有力工具。

1 材料与方法

1.1 野三七转录组测序与 SSR 的筛选

野三七根和叶片 RNA 混合样品的转录组测序由北京博奥公司进行,共获得 126 758 条 *unigenes*。利用 *est_timmer.perl* 程序过滤 *unigenes*, 参数为: $-amb=2, 50$; $-tr5=T, 5, 50$; $-tr3=A, 5, 50$; $-cut=100, 700$; 再利用 *CD_HIT* 去除冗余序列, 参数为: $-c 1.00 -n 5 -M 2\ 000$ 。利用 *misa.pl* 搜索和定位 SSR, 搜索的标准为重复单元长度 2~6 bp, 二核苷酸至少重复 6 次, 三核苷酸至少重复 5 次, 四至六核苷酸至少重复 4 次; 同时筛选被间隔小于或等于 100 bp 碱基间隔的复合型 SSR。单核苷酸重复容易发生错配而测序失败, 所以没有选择。

1.2 SSR 引物设计

调用 Primer 3.0 程序对含有 SSR 位点的两端序列设计引物, 引物设计标准: 退火温度 (T_m) 在 55~64 °C, 上、下游引物的 T_m 值相差不大于 5 °C; 引物长度在 18~27 bp; GC 量在 40%~60%; 尽量避免出现引物二级结构如发卡结构、二聚体、错配和引物二聚体。为验证 SSR 引物是否可用, 随机选择合成 30 对引物 (昆明毓秀生物科技有限公司), 进行 PCR 扩增。

1.3 野三七 DNA 提取与 PCR 扩增

13 株采自不同产地的野三七 (表 1), 按照 Zhu

表 1 不同来源的野三七植株

Table 1 PVF plants collected from different habitats

编号	来源	编号	来源
MX11068	金平县铜厂乡勐谢村	Laos1203	老挝
MX11075	金平县铜厂乡勐谢村	Laos-1	老挝
MX11082	金平县铜厂乡勐谢村	ZM-1	金平县者米乡
YY1145	元阳县	ZM-2	金平县者米乡
YY-2	元阳县	TS1232	金平县铜厂乡炭山
YY-3	元阳县	TS1239	金平县铜厂乡炭山
Laos1201	老挝		

等^[1]的方法对其形态特征进行鉴定。野三七的典型特征是无托叶、无苞片, 成熟果实时为红色, 顶端有一黑色斑点。提取叶片 DNA (北京百泰克生物技术有限公司 DNA 提取试剂盒)。20 μ L PCR 扩增体系包含 2.5 mmol/L dNTPs 1 μ L, EasyTaq DNA polymerase 1 μ L, 10 $\times$ EasyTaq 缓冲液 2 μ L, 10 μ mol/L 的上游和下游引物各 1 μ L, 10 ng/ μ L DNA 1 μ L, ddH₂O 13 μ L。PCR 扩增程序为 94 °C 预变性 3 min 后 35 个循环, 每个循环包含 94 °C 变性 30 s, 最佳 T_m 退火 1 min, 72 °C 延伸 1 min。循环结束后再 72 °C 延伸 5 min。扩增产物用 8% 非变性聚丙烯酰胺凝胶检测。

1.4 数据处理

采用人工读带的方法, 将电泳图上可重复的、清晰的条带计为“1”, 同一位置无带或不易分辨的弱带计为“0”, 建立原始数据矩阵。利用软件 NTsys 2.10e 和 Popgene 3.2 计算每对引物扩增位点的等位位点数、多态性信息量、遗传距离, 按 UPGMA 进行聚类绘图。

2 结果与分析

2.1 野三七转录组 SSR 位点数量与分布频率

利用 MISA 工具搜索野三七转录组的 126 758 条 *unigenes* 序列, 共在 17 780 条 *unigene* 中找到 21 320 个符合条件的 SSRs, 发生频率 (含 SSR 的 *unigene* 数与总 *unigene* 数之比) 为 14.03%。14 862 条 *unigene* 含有单个 SSR 位点, 2 918 条 *unigene* (16.41%) 含有一个以上 SSR 位点, 1 207 条 *unigene* 含有复合 SSRs。野三七转录组中 SSR 位点的出现频率 (SSR 数目与总 *unigene* 的数目比值) 为 16.82%。平均每 10 kb 有 1.29 个 SSR 位点 (表 2)。

野三七转录组 SSR 种类较为丰富, 从重复次数

表2 野三七转录组中 SSR 搜索结果

Table 2 Searching results of SSR in transcriptome of PVF

项目	数量
搜索序列总数	126 758
搜索序列总长度 / bp	16 529, 1 103
SSR 总数	21 320
SSR 发生频率 / %	14.03
SSR 出现频率 / %	16.82
10 kb 含 SSR 数目	1.29
含 SSR 位点的序列数目	17 780
含超过 1 个 SSR 位点的序列数目	2 918
含复合型 SSR 的序列数目	1 207

表3 野三七转录组中不同基序长度和重复的 SSR 数量分布

Table 3 Distribution of SSRs with different motif types and repeat numbers in transcriptome of PVF

基序长度 / bp	重复次数									总计	比例 / %
	4	5	6	7	8	9	10	11	>11		
2	0	0	4 030	2 480	1 576	1 240	1 230	627	14	11 197	52.52
3	0	3 433	1 448	966	126	2	0	3	9	5 987	28.08
4	1 789	446	96	12	0	6	0	0	3	2 352	11.03
5	812	94	10	7	0	0	0	0	1	924	4.33
6	755	15	73	1	9	0	4	2	1	860	4.03
总计	3 356	3 988	5 657	3 466	1 711	1 248	1234	632	28	21 320	100
比例 / %	15.74	18.71	26.53	16.26	8.03	5.85	5.79	2.96	0.13	100	—

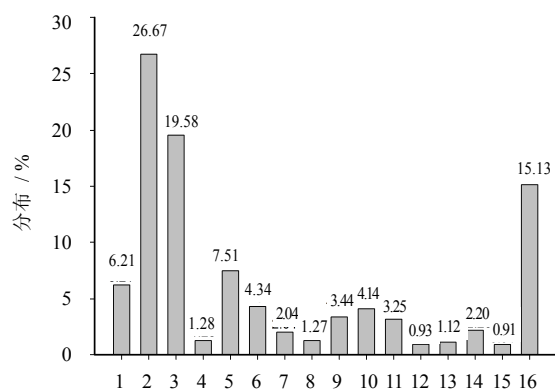


图1 野三七转录组中 SSR 基序类型分布

Fig. 1 Distribution of SSR type in transcriptome of PVF

2.2 野三七转录组 SSR 的可用性评价

SSR 标记多态性是判断其可用性的重要依据。SSR 长度是影响其多态性高低的重要因素,当 SSR 长度大于或等于 20 bp 时多态性较高,长度在 12~20 bp 的 SSR 多态性中等,而长度在 12 bp 以下时多

看,发现重复基元以重复 6 次出现的频率最高,有 5 657 个,占总 SSR 的 26.53%。其次为 4、5、7 次重复,分别占总 SSR 数量的 15.74%、18.71% 和 16.26% (表 3)。

从野三七 SSR 核苷酸基序类型来看 (图 1), SSR 以二核苷酸 AG/CT 和 AT/AT 重复基序为主要类型,占总 SSR 的 46.25%。其次是三、四核苷酸重复基序,其出现频率分别为 28.08% 和 11.03%,以 4~7 次重复为主。三核苷酸重复序列 AAG/CTT 最多,ACT/AGT 重复序列最少;四核苷酸重复序列 AAAT/ATTT 为最多,AAAG/CTTT 次之,ACAT/ATGT 为最少。

态性极低^[13]。本研究 SSR 筛选过程中已将多态潜能低于 12 bp 以下的 SSR 过滤去掉。野三七转录组中,长度为 12~20 bp 的 SSR 最多,共有 16 510 个,占非复合型 SSR 总数 (20 113 个) 的 82.09%,其次是长度为 21~30 bp,共 2 140 个 (11.98%) (图 2)。

2.3 SSR 引物筛选与多态性分析

利用 Primer3 共设计出 39 336 对 SSR 引物。为

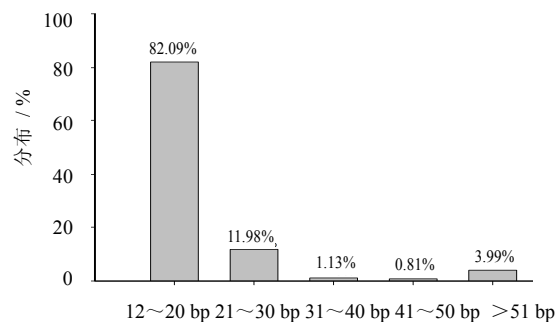


图2 野三七转录组中 SSR 长度分布

Fig. 2 Distribution of SSR length in transcriptome of PVF

了验证其引物的有效性, 随机挑选 30 对引物对 13 株不同来源的野三七基因组 DNA 进行 SSR-PCR 扩增。扩增结果表明, 除 1 对引物没有扩增产物外, 其余 29 对 (96.67%) 引物均能扩增出可重复的、清晰的条带。其中 24 对 (80.00%) PCR 扩增产物与预期大小复合, 有 4 对 (13.33%) 扩增产物长度超过预期, 1 对引物 (3.33%) 扩增产物小于预期。

15 对 (50.00%) 表现出多态性差异 (表 4), 多态性信息量 (PIC) 范围为 0.29~0.50, 平均值为 0.44。利用 SSR 多态性对 13 个野三七材料进行聚类分析, 13 个野三七材料被聚成 2 类 (图 3), 一类是购买自老挝, 另一类是采自云南省红河州的金平县、元阳县, 说明老挝的野三七种质和国内种质遗传距离较远。

表 4 SSR-PCR 筛选引物
Table 4 Filtering primers of SSR-PCR

编号	正向引物	反向引物	No	Ne	I	NP	PPB	PIC
P01	GTTATCATGTAGTCCCTCAAACACC	TTTAGTCATGTGCCCTCCT	0.33	0.27	0.21	1.6	17.78	0.47
P07	GCAATGTGTAGGAGACTTTGTAACC	GCTTGCCAACTGCAGTAGATTT	0.42	0.34	0.27	1.8	20.00	0.50
P09	GCTGAATATGAAACGGAAATAGAGC	AAAGTGCATAGACTGAAGGTTCTT	0.25	0.20	0.16	1.2	13.33	0.49
P12	CTGCACATACAAACAAGCAGCTAAT	TGGTTCTGTTTGCATTCTGT	0.47	0.38	0.30	2.6	28.89	0.32
P15	TGCTTATTCTGGACATGCAAAAC	ACAACAAACAAGTCTCCTGTAGAGC	0.38	0.31	0.24	2.2	24.44	0.29
P18	CAAACAAGACATACCATAACAGAGG	ACTCATCAATCACCCAAGTGTAGAT	0.49	0.39	0.31	3.6	40.00	0.41
P19	ATTCAATGCTATGAAGCACACCAC	AAGGAAAGCCAGGGATGTCT	0.39	0.31	0.25	2.2	24.44	0.46
P24	CAAACCAGGCTGTGTTGTTG	CATCATCATCAGATTCATCGCTAGT	0.49	0.39	0.31	3.2	35.55	0.47
P26	ATCCGGATGCTCCAAATC	GGGAAATAGATGGTCTCTAATAGC	0.48	0.39	0.31	3.4	37.78	0.48
P30	TAAACTTATGGTCTGGCGATAACC	GTGGAAGGCTTCAATCAGACTTTAT	0.47	0.38	0.21	2.6	28.89	0.46
P31	GCGTCTTTGTGTGAAAGAACT	GAGCAAGGCAATCTACATTCTACAC	0.33	0.27	0.21	1.4	15.55	0.50
P36	CACTGAGAAGGAAGAAGCAGTATAAGG	AGATAAGAAATGGCTAACGGCTTTG	0.49	0.39	0.31	3.0	33.33	0.49
P41	CAGAAGCTGTTAATGTGATTAGCTC	TTCTCTACTAGCTGATGGATCCTGT	0.27	0.21	0.17	0.8	8.89	0.38
P43	ATAAAGTCTGAACCGGACTATAACG	TTCGGTAGTCACGCTTAACAAT	0.50	0.40	0.32	4.4	48.89	0.50
P50	TTGATTATACCTCCCTTACGTTGTC	CCATGCAAATAACACATACATGGTC	0.52	0.41	0.33	4.2	46.67	0.46
平均值			0.42	0.34	0.27	2.5	28.15	0.44

No-等位基因观察值 Ne-等位基因预期值 I-Shannon 信息指数 NP-多态性位点数 PPB-多态性位点比例

No-observed number of alleles Ne-expected number of alleles I-Shannon's information index NP-number of polymorphic loci PPB-percentage of polymorphic loci

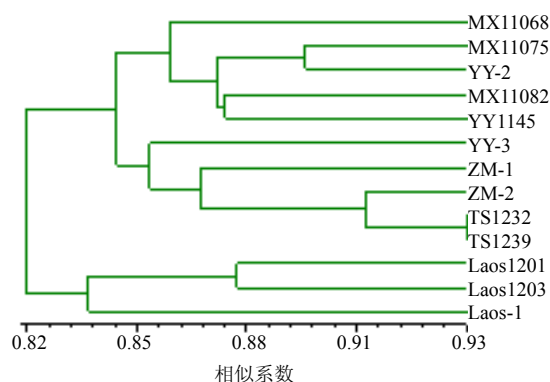


图 3 野三七的 UPGMA 聚类图

Fig. 3 Dendrogram for tested PVF plants by UPGMA

3 讨论

野三七转录组中存在丰富的 SSR 位点。本研究从野三七转录组中共搜索出 21 320 个 SSR, 分布于 17 780 条 *unigenes*。SSR 出现频率为 16.86%, 高于玉米 (1.5%)、水稻 (4.7%)、大麦 (3.4%)、高粱

(3.6%) 和小麦 (3.2%) 等作物^[14], 这种差异可能与 SSR 搜索标准、数据库大小和物种等有关。

多数植物 EST-SSR 以二核苷酸和三核苷酸重复为主, 但主要的重复单元不同。野三七转录组 SSR 重复单元以二核苷酸为主, 占有 SSRs 的 52.52%, 其次是三核苷酸重复单元。这与大豆、番茄、棉花、杨树、拟南芥、大麦等作物不同, 这些植物中 SSR 的主要类型是三核苷酸重复^[15-16]。GC 重复单元在多数植物中很难见到, 出现频率都很低^[17-19]。本研究也发现, 野三七转录组中, SSR 位点最多的是二核苷酸的重复单元 AT/AT 和 AG/CT, 仅有 1 个 GC 重复单元。

总的来说, 野三七转录组 SSR 不但出现频率高, 而且类型丰富; 从多态性潜能的角度考虑, 搜索到的这些 SSR 也具有较高的可用性。因此, 本研究的结果为进一步开发新的野三七功能基因 SSR 标记奠定了基础, 这种标记的建立对于加速野三七功能基

因资源的开发利用、丰富其分子标记类型、遗传资源评价、绘制遗传图谱、实现特定性状的辅助选择和进行比较基因组学研究都具有重要的意义。

参考文献

- [1] Zhu S, Fushimi H, Cai S Q, *et al.* A new variety of the genus *Panax* from Southern Yunnan, China and its nucleotide sequences of 18S ribosomal RNA gene and matK gene [J]. *Jap J Bot*, 2003, 78: 86-94.
- [2] 杨兴仓, 司民真, 刘仁明, 等. 野三七的红外光谱分析 [J]. *光散射学报*, 2011, 23(2): 162-167.
- [3] Takao K, Midori T, Harukuni T, *et al.* Anti-tumor-promoting activity of majonoside-R₂ from *Vietnamese Ginseng*, *Panax vietnamensis* Ha et Grushv. [J]. *Biol Pharm Bull*, 1998, 21(8): 834-838.
- [4] Takao K, Midori T, Eiichiro I, *et al.* Cancer chemopreventive activity of majonoside-R₂ from *Vietnamese Ginseng*, *Panax vietnamensis* Ha et Grushv. [J]. *Cancer Lett*, 1999, 147: 11-16.
- [5] Zhu S, Zou K, Fushimi H, *et al.* Comparative study on triterpene saponins of *ginseng* drugs [J]. *Planta Med*, 2004, 70: 666-677.
- [6] Zhu S, Zou K, Fushimi H, *et al.* Comparative study on triterpene saponins of *ginseng* drugs [J]. *Planta Med*, 2004, 70: 666-677.
- [7] Nguyen M D, Nguyen T N, Ryoji K, *et al.* Saponins from *Vietnamese ginseng*, *Panax vietnamensis* Ha et Grushv. collected in central Vietnam: I [J]. *Chem Pharm Bull*, 1993, 41(11): 2010-2014.
- [8] Nguyen M D, Ryoji K, Kazuhiro O, *et al.* Saponins from *Vietnamese ginseng*, *Panax vietnamensis* Ha et Grushv. collected in central Vietnam: II [J]. *Chem Pharm Bull*, 1994, 42(1): 115-122.
- [9] Nguyen M D, Ryoji K, Kazuhiro O, *et al.* Saponins from *Vietnamese ginseng*, *Panax vietnamensis* Ha et Grushv. collected in central Vietnam: III [J]. *Chem Pharm Bull*, 1994, 42(3): 634-640.
- [10] Quan L T, Ketut A, Yasuhiro T, *et al.* Triterpene saponins from *Vietnamese ginseng* (*Panax vietnamensis*) and their hepatocytoprotective activity [J]. *J Nat Prod*, 2001, 64(4): 456-461.
- [11] 黄映萍. DNA 分子标记研究进展 [J]. 中山大学研究生学刊: 自然科学医学版, 2010, 31(2): 27-36.
- [12] 闫秋良. 基于生物信息学方法从牛和绵羊表达序列标签中筛选 SSR 标记的初步研究 [D]. 西安: 西北农林科技大学, 2007.
- [13] Temnykh S, de Clerck G, Lukashova A, *et al.* Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.) [J]. *Genome Res*, 2001, 11: 1441-1452.
- [14] Kantety R V, Rota M L, Matthews D E, *et al.* Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat [J]. *Plant Mol Biol*, 2002, 48: 501-510.
- [15] Cardle L, Ramsay L, Mibourne D, *et al.* Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. *Genetics*, 2000, 156: 847-854.
- [16] Varshney R K, Graner A, Sorrells M. Genic microsatellite markers in plants: features and applications [J]. *Trends Biotechnol*, 2005, 23(1): 48-55.
- [17] Lidan S, Weiru Y, Qixiang Z, *et al.* Genome-wide characterization and linkage mapping of simple sequence repeats in *Mei* (*Prunus mume* Sieb. et Zucc.) [J]. *PLoS One*, 2013, 8(3): e59562.
- [18] Cardle L, Ramsay L, Milbourne D, *et al.* Computational and experimental characterization of physically clustered simple sequence repeats in plants [J]. *Genetics*, 2000, 156: 847-854.
- [19] Marshall H, Wayne W X, Brian W K, *et al.* Identification of differential gene expression in *Brassica rapa* Nectarines through expressed sequence tag analysis [J]. *PLoS One*, 2010, 5(1): e8782.