

番红花 EST 资源的 SSR 信息分析

陈国庆

(湖北师范学院 生命科学学院, 湖北 黄石 435002)

摘要: 从 NCBI 公共数据库下载得到 6745 条番红花 EST, 通过前处理得到全长为 612.01 kb 的无冗余 EST 1431 条。在这些序列中搜索出 108 个 SSR, 分布于 103 条 EST 中, 出现频率为 7.55%。这些 EST-SSR 的平均分布距离是 5.67 kb。二核苷酸重复和三核苷酸重复是番红花主要的重复类型, 分别占总 EST-SSR 的 30.56% 和 37.96%。AG/CT 和 AAG/CTT 是二、三核苷酸重复中的优势基元, 分别占二、三核苷酸重复的 66.67% 和 29.27%。本研究为番红花 EST-SSR 标记的建立和进一步应用奠定了基础。

关键词: 番红花; 药用植物; EST; EST-SSR; 微卫星

中图分类号: Q75 **文献标识码:** A **文章编号:** 1000-3142(2011)01-0043-04

Analysis of SSR information in EST resource of saffron crocus (*Crocus sativus*)

CHEN Guo-Qing

(College of Life Science, Hubei Normal University, Huangshi 435002, China)

Abstract: 6745 ESTs of saffron crocus in the public database of NCBI were downloaded and analysed, resulting in 1431 non-redundant ESTs with total length about 612.01 kb. A total of 108 SSRs distributed in 103 ESTs were detected, accounting for 7.55% of the non-redundant ESTs. The average distribution distance of these EST-SSRs was about 5.67 kb. Dinucleotide and trinucleotide repeats were the main types in *Crocus sativus*, accounting for 30.56% and 37.96% of all the SSRs, respectively. AG/CT and AAG/CTT are the most frequent motifs, accounting for 66.67% and 29.27% in dinucleotide and trinucleotide repeats, respectively. The present study provides a base for the development and further application of EST-SSR markers in saffron crocus.

Key words: saffron crocus; medicinal plants; EST; EST-SSR; microsatellite

简单序列重复 (Simple Sequence Repeats, SSRs), 又叫微卫星 (Microsatellite) 或短串联重复 (Short Tandem Repeats), 通常是指由 1~6 个碱基组成的核苷酸重复基元。SSRs 广泛分布于真核生物基因组中, 具有共显性表达、多态性高等优点, 因而得到了越来越多的应用 (Powell 等, 1996; Toth 等, 2000)。然而, SSR 标记的开发需时长、花费高, 大大限制了该分子标记的使用; 即使采用 SSR 富集方法可以提高产出、减低成本, 仍然需要投入大量的

人力和财力 (Kumapatla 等, 2004)。近年来, 随着功能基因组学的发展, 表达序列标签 (expressed sequence tags, ESTs) 已成为开发新型分子标记的重要资源。EST-SSR 的开发建立在生物信息学方法的基础上, 不但省时, 而且大大降低成本。相对于传统的基因组 SSR (genomic SSR), 基于 EST 的微卫星标记 (EST-SSR) 具有独特的优越性: 第一, 序列保守性高, 通用性强, 能在近缘属种间获得较好扩增; 第二, 由于其本身来自基因表达序列, 可能与功

收稿日期: 2010-04-19 修回日期: 2010-10-29

基金项目: 国家自然科学基金 (30900157); 湖北省教育厅青年项目 (Q20082201); 湖北师范学院人才引进项目 (2007F13) [Supported by the National Natural Science Foundation of China (30900157); the Youth Foundation of Hubei Educational Committee (Q20082201); the Talent Introduction Funds of Hubei Normal University (2007F13)]

作者简介: 陈国庆 (1979-), 女, 河北邯郸人, 博士, 研究方向为分子生态学, (E-mail) gqchen1001@yahoo.com.cn.

能基因表达有着直接或间接的关系,可作为一种功能性分子标记,因此也称基因 SSR (genic SSR) (Varshney 等,2005)。

番红花 (*Crocus sativus*) 又名藏红花、西红花,为鸢尾科番红花属植物,原产西班牙、希腊和伊朗等国,引入我国已有 600 多年历史。它是一种珍贵的药用植物,具有镇静、祛痰、刺激、解痉、活血化瘀等功效,用于胃痛、痛经,可治疗痢疾、麻疹、发热黄疸、肝脾肿大、泌尿道感染及糖尿病等多种疾病(陈书安等,2001)。由于仅以柱头入药,资源极其有限,故价格昂贵,被誉为“植物黄金”。为充分合理地利用这一珍贵药用植物,开展番红花资源多样性研究势在必行。目前尚未见有关番红花基因组 SSR 标记的报道,而公共数据库中已积累了大量的番红花 EST 序列,为 EST-SSR 标记的开发提供了契机。本研究对现有番红花 EST 中的 SSR 信息进行了全面分析,以明确其 EST-SSR 的发生频率和特点,为进一步筛选番红花 EST-SSR 标记和探索其在遗传育种、图谱构建中的应用奠定基础。

1 材料与方 法

1.1 番红花 EST 序列的来源

番红花 EST 来自美国国家生物技术信息中心 (National Center for Biotechnology Information, NCBI) (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) 的 GenBank 数据库,共计 6 745 条序列。

1.2 EST 的前处理与片段重叠群 (contig) 分析

用 EST-trimmer 软件 (<http://pgrc.ipk-gatersleben.de/misa/download/est-trimmer.pl>) 除去 5' 端或 3' 端 50 bp 内重复次数达 5 次和 5 次以上的 polyT 或 polyA; 除去长度小于 100 bp 的 EST 序列; 对于长度大于 700 bp 的 EST, 则保留其 5' 端的 700 bp。用 DNASTar 软件 (LASERGENE, DNASTAR, Inc.) 对每个物种的 EST 序列分别进行片段重叠群分析和聚类。装配参数设置为: 在至少 40 个核苷酸的重复区域中, 最小匹配百分比大于 95% 的 EST 序列可以进行拼接与延伸。

1.3 EST-SSR 的发掘

用 MISA 程序 (<http://pgrc.ipk-gatersleben.de/misa/>) 对经过前处理的无冗余 EST 进行 SSR 查找。筛选标准为单核苷酸重复次数不小于 12 次, 二核苷酸重复次数不小于 6 次, 三至六核苷酸重复

次数不小于 5 次。同时筛选和记录被少数碱基 (间隔小于 10 或等于 10 bp) 打断的不完全重复的 SSR。

2 结果与分析

2.1 番红花 EST-SSR 出现的频率

从 NCBI 公共数据库下载得到 6 745 条来自番红花的 EST, 总长度约为 2.64 Mb, 经过处理后得到全长为 612.01 kb 的无冗余 EST 序列 1 431 条, 可见数据库中存在大量的番红花冗余 EST 序列。对所得的无冗余 EST 进行搜索, 共检出 103 条含有 SSR 的 EST 序列, 发生频率 (含有 SSR 的 EST 数目与 EST 数目的比值) 为 7.20% (表 1)。其中只含有 1 个 SSR 的 EST 有 98 条, 含有 2 个或 2 个以上 SSR 的有 5 条, 共计 108 个 SSR, 占无冗余 EST 的 7.55%, 即为 EST-SSR 的出现频率。在 5 条含 2 个以上 SSR 的 EST 中, 相邻 SSR 之间的距离均大于 10 bp。从分布情况来看, 平均每 5.67 kb 就出现 1 个 SSR, 但不同重复类型间差异很大 (表 1)。

表 1 番红花不同重复类型的 EST-SSR 分布
Table 1 Percentage of EST-SSR of different repeat types in *Crocus sativus*

重复类型 Repeat type	SSR 数目 No. of SSRs	%	出现频率 Frequency of occurrence	
			%	kb
单核苷酸重复 Mononucleotide	20	18.52	1.40	30.6
二核苷酸重复 Dinucleotide	33	30.56	2.31	18.55
三核苷酸重复 Trinucleotide	41	37.96	2.87	14.93
四核苷酸重复 Tetranucleotide	8	7.41	0.56	76.5
五核苷酸重复 Pentanucleotide	4	3.70	0.28	153
六核苷酸重复 Hexanucleotide	2	1.85	0.14	306
总计 Total	108	100.00	7.55	5.67

在不同核苷酸重复类型中, 最常见的是二核苷酸和三核苷酸重复, 分别占总 EST-SSR 的 30.56% 和 37.96%; 单核苷酸 SSR 所占比例不足 20%; 四至六核苷酸 SSR 总计约占 13%, 所占比例随着重复基元核苷酸数目的增大而降低 (表 1)。

2.2 番红花 EST-SSR 的特性

在搜索出的番红花 EST-SSR 中, 共观察到 23 种不同的重复基元, 其中单核苷酸重复仅 1 种, 二核苷酸重复 3 种, 三、四、五、六核苷酸重复分别 10、4、3、2 种 (表 2)。单核苷酸重复只检测到 A/T 重复基元, 没有发现 C/G 重复类型。二核苷酸重复中, AG/CT 出现的频率最高, 占有二核苷酸重复的

表 2 非冗余番红花 EST 中各类型 SSR 的分布
Table 2 Distribution of different SSRs in the non-redundant ESTs of *Crocus sativus*

重复单元 SSR Motif	重复次数 Number of repetitions													合计 Total
	5	6	7	8	9	10	11	12	13	14	15	≥16		
A/T	—	—	—	—	—	—	—	8	4	5	1	2	20	
AC/GT	—	1	1	1	1	—	—	—	—	—	—	—	4	
AG/CT	—	5	3	1	1	—	2	—	2	—	1	7	22	
AT/AT	—	5	—	—	—	—	—	1	—	—	—	1	7	
AAC/GTT	3	—	—	—	—	—	—	—	—	—	—	—	3	
AAG/CTT	4	5	2	—	1	—	—	—	—	—	—	—	12	
AAT/ATT	3	—	—	—	—	—	—	—	—	—	—	—	3	
ACC/GGT	—	1	1	—	—	—	—	—	—	—	—	—	2	
ACG/CTG	2	—	—	—	—	1	—	—	—	—	—	—	3	
ACT/ATG	1	—	—	—	—	—	—	—	—	—	—	—	1	
AGC/CGT	1	3	—	—	—	—	1	—	—	—	—	—	5	
AGG/CCT	2	—	1	1	—	—	—	—	—	—	—	—	4	
AGT/ATC	1	—	—	—	—	—	1	—	—	1	—	—	3	
CCG/CGG	4	—	1	—	—	—	—	—	—	—	—	—	5	
AAAG/CTTT	1	1	—	—	—	—	—	—	—	—	—	—	2	
AACT/ATTG	—	1	—	—	—	—	—	—	—	—	—	—	1	
AATT/AATT	1	1	1	—	—	—	—	—	—	—	—	—	3	
AGAT/ATCT	1	1	—	—	—	—	—	—	—	—	—	—	2	
AAAAG/CTTTT	1	—	—	—	—	—	—	—	—	—	—	—	1	
AAAGG/CCTTT	—	1	—	—	—	—	—	—	—	—	—	—	1	
ACGAT/ATGCT	1	1	—	—	—	—	—	—	—	—	—	—	2	
AAAACC/GGTTTT	1	—	—	—	—	—	—	—	—	—	—	—	1	
AAGGAG/CCTCTT	1	—	—	—	—	—	—	—	—	—	—	—	1	

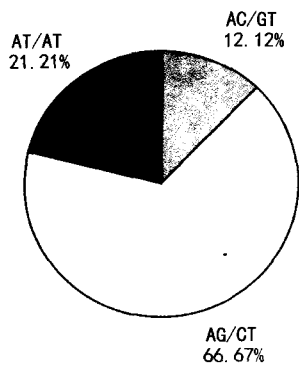


图 1 二核苷酸重复中不同重复基元的比例
Fig. 1 Percentage of different motifs in dinucleotide repeats

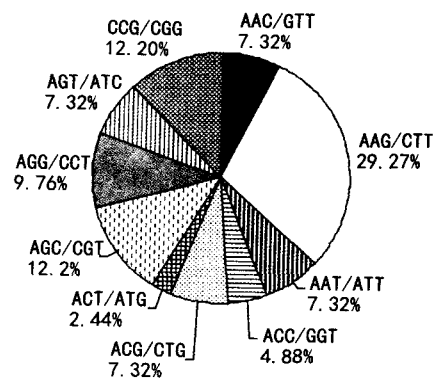


图 2 三核苷酸重复中不同重复基元的比例
Fig. 2 Percentage of different motifs in trinucleotide repeats

66.67%，其次是 AT 重复基元，但是没有检测到 CG 重复基元(图 1)。三核苷酸重复中，AAG/CTT 出现 12 次，所占比例达 29.27%；其次是 CCG/CGG 和 AGC/CGT，所占比例均为 12.20%，仅次于 AAG/CTT 重复基元(图 2)。四、五、六核苷酸重复的各种重复类型的出现次数均小于 5(表 2)。就所有重复基元来看，所占比例最高的是 AG/CT (20.37%)，其次是 A/T (18.52%) 和 AAG/CTT (11.11%)(图 3)。

3 讨论

本研究对 GenBank 中现有的番红花 EST 进行分析，结果表明 EST-SSR 的出现频率为 7.55%，SSR-EST 的发生频率为 7.20%。分析得到的番红花 SSR-EST 发生频率高于 Kumpatla & Mukhopadhyay(2005)对番红花 553 条 EST 的研究结果 (5.79%)。造成差别的原因可能有：(1)所分析的

EST 数量过少,可能导致结果的偏差。Kumpatla & Mukhopadhyay(2005)的分析仅针对当时 GenBank 中为数不多的 500 多条番红花 EST。(2)数据分析方法的不同造成结果的差异。Kumpatla & Mukhopadhyay(2005)首先在 ESTs 中查找 SSR,再对含有 SSR 的 EST 进行拼接聚类来排除冗余的 SSR-EST;而本研究则先进行 EST 拼接聚类、去冗余,再进行 SSR 筛选,这样在片段重叠群聚类分析中,可以将一些被打断的 SSR 拼接起来,从而增加 SSR 检出频率。葛佳等(2005)在对大白菜(*Brassica campestris* ssp. *pekinensis*)进行 EST-SSR 分析时发现,非冗余 EST 中 SSR 总数比排除冗余前还要高,证实了片段重叠群分析可以拼接一些被打断的 SSR。(3)筛选 SSR 的参数设置不同,也会造成结果的差异。Kumpatla & Mukhopadhyay(2005)筛选 SSR 的标准为单核苷酸重复次数不小于 15 次,二至四核苷酸重复次数不小于 5 次;而本分析采用标准为单核苷酸重复次数不小于 12 次,二核苷酸重复次数不小于 6 次,三至六核苷酸重复次数不小于 5 次。我们采用 Kumpatla & Mukhopadhyay(2005)的筛选标准进行重新分析,结果发现 133 个 SSR 分布于 125 条无冗余 EST 序列中,含有 SSR 的 EST 发生频率是 8.74%,SSR 的出现频率为 9.29%。这也同时表明,SSR 筛选标准改变时,EST-SSR 的分布频率和 SSR-EST 的发生频率都会随之改变。因此,将不同研究结果加以比较时,应当采用同一标准。

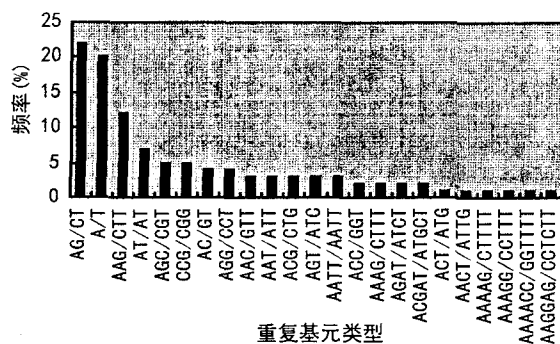


图 3 基于重复基元类型的番红花 EST-SSR 比例分布
Fig. 3 Percentage of different motifs in EST-SSRs of *Crocus sativus*

大多数植物的 EST-SSR 以三核苷酸重复占主导地位,其次是二核苷酸重复(李永强等,2004;

Varshney 等,2005)。番红花也表现出类似情况,筛选出的 EST-SSR 中三核苷酸重复所占比例最大(37.96%),二核苷酸重复也较多(30.56%)。番红花二核苷酸重复中,AG/CT 出现频率最高,与多数植物中已报道的情况相同(李永强等,2004);在三核苷酸重复中,则以 AAG/CTT 重复为主,与拟南芥(*Arabidopsis thaliana*)、柑橘(*Citrus*)、大豆(*Glycine max*)等植物类似(Cardle 等,2000;Gao 等,2003;Chen 等,2006)。CCG/CGG 重复基元在番红花三核苷酸重复类型中所占比例也较高(图 2),这种重复基元在单子叶植物大麦(*Hordeum vulgare*)、玉米(*Zea mays*)、水稻(*Oryza sativa*)和高粱(*Sorghum bicolor*)中是占据主导地位的三核苷酸重复类型。这些占优势的核苷酸重复基元可能与其编码相应蛋白质的使用频率较高有关,这一点在拟南芥中得到证实(范三红等,2003)。

番红花 EST 数据的不断积累为 EST-SSR 标记的开发提供了宝贵资源。当前分析表明番红花 EST-SSR 出现频率较高、类型较丰富,因此,本研究结果为进一步开发和筛选番红花 EST-SSR 标记奠定了基础,有助于该物种的遗传资源多样性评价、遗传图谱绘制、分子标记辅助育种和比较基因组学等相关研究。

参考文献:

- Cardle L, Ramsay L, Milbourne D, et al. 2000. Computational and experimental characterization of physically clustered simple sequence repeats in plants[J]. *Genetics*, 156:847-854
- Chen CX, Zhou P, Choi YA, et al. 2006. Mining and characterizing microsatellites from *Citrus* ESTs [J]. *Theor Appl Genet*, 112:1 248-1 257
- Chen SA(陈书安), Wang XD(王晓东), Zhao B(赵兵), et al. 2001. Advances in studies on *Crocus sativus*(藏红花的研究进展)[J]. *Chin Trad Herb Drugs*(中草药), 52(12):1 137-1 139
- Fan SH(范三红), Guo AG(郭嵩光), Shan LW(单丽伟), et al. 2003. Analysis of genetic code preference in *Arabidopsis thaliana*(拟南芥基因密码子偏爱性分析)[J]. *Prog Biochem Biophys*(生物化学与生物物理进展), 30(2):221-225
- Gao LF, Tang JF, Li HW, et al. 2003. Analysis of microsatellites in major crops assessed by computational and experimental approaches [J]. *Mol Breed*, 12:245-261
- Ge J(葛佳), Xie H(谢华), Cui CS(崔崇士), et al. 2005. Analysis of expressed sequences tags(ESTs) derived SSR markers in Chinese cabbage(*Brassica campestris* ssp. *pekinensis*)(大白菜表达序列标签 SSR 标记分析)[J]. *J Agric Biotech*(农业生物技术学报), 13(4):423-428
- Kumpatla SP, Manley MK, Horne EC, et al. 2004. An improved (下转第 96 页 Continue on page 96)

- Feng JC(冯建灿), Hu XL(胡秀丽), Mao XJ(毛训甲). 2002. Application of chlorophyll fluorescence dynamics to plant physiology in adverse circumstance(叶绿素荧光动力学在研究植物逆境生理中的应用)[J]. *Econom Fore Res*(经济林研究), **20**(4):14-18
- Hu XH(胡学华), Pu GL(蒲光兰), Xiao QW(肖千文), et al. 2007. Effects of water stress on chlorophyll fluorescence in leaves of plum(*Prunus salicina* Lindl.) (水分胁迫下李树叶绿素荧光动力学特性研究)[J]. *Chine J Eco-Agri*(中国生态农业学报), **15**(1):75-77
- Jiang CD, Gao HY, Zou Q. 2003. Changes of donor and acceptor side in photosystem 2 complex induced by iron deficiency in attached soybean and maize leaves[J]. *Photosynthetica*, **41**:267-271
- Krause GH, Weis E. 1988. Applications of chlorophyll fluorescence in photosynthesis research[J]. *Physiol Plant*, **74**:566-577
- Mommer L, Visser EJW. 2005. Underwater photosynthesis in flooded terrestrial plants; a matter of leaf plasticity[J]. *Ann Bot*, **96**:581-589
- Panda D, Rao DN, Sharma SG, et al. 2006. Submergence effect on rice genotypes during seedling stage; probing of submergence driven changes of photosystem II by chlorophyll a fluorescence induction O-J-I-P transients[J]. *Photosynthetica*, **44**:69-75
- Roháček K, Barták M. 1999. Technique of the modulated chlorophyll fluorescence; Basic concepts, useful parameters, and some applications[J]. *Photosynthetica*, **37**:339-363
- Vyal YA, Dyukova GR, Leonova NA, et al. 2007. Adaptation of the photosynthetic apparatus of the immature broadleaf trees to the floodplain conditions[J]. *Russ J Plant Physiol*, **54**:58-62
- Wang KF(王可芬), Xu CH(许春辉), Zhao FH(赵福洪), et al. 1997. The effects of water stress on some *in vivo* chlorophyll a fluorescence parameters of wheat flag leaves(水分胁迫对小麦旗叶某些体内叶绿素 a 荧光参数的影响)[J]. *Acta Biophysica Sin*(生物物理学报), **13**(2):273-278
- Wilbur KM, Anderson NG. 1948. Electrometric and colorimetric determination of carbonic anhydrase[J]. *J Biol Chem*, **176**:147-154
- Wei Y(魏媛), Yu LF(喻理飞), Zhang JC(张金池). 2007. Study on biomass construction of *Broussonetia papyrifera* sprouting population with different disturbed conditions in karst region(喀斯特地区不同干扰条件下构树萌株种群生物量构成)[J]. *J Nanjing Fore Univ, Nat Sci Edi*(南京林业大学学报·自然科学版), **31**:58-66
- Wu YY(吴沿友), Li XT(李西腾), Hao JC(郝建朝), et al. 2006. Study on the difference of the activities of carbonic anhydrase in different plants(不同植物的碳酸酐酶活力差异研究)[J]. *Guihaia*(广西植物), **26**(4):366-369
- Wu YY, Li PP, Zhao YG, et al. 2007. Study on photosynthetic characteristics of *Orychophragmus violaceus* related to shade-tolerance[J]. *Sci Horti*, **113**:173-176
- Wu YY, Liu CQ, Li PP, et al. 2009. Photosynthetic characteristics involved in adaptability to karst soil and alien invasion of paper mulberry (*Broussonetia papyrifera*) in comparison with mulberry (*Morus alba*) [J]. *Photosynthetica*, **47**(1):155-160
- Wu YY, Wu XM, Li PP, et al. 2005. Comparison of photosynthetic activity of *Orychophragmus violaceus* and oil-seed rape [J]. *Photosynthetica*, **43**(2):299-302
- Wu YY, Zhao XZ, Li PP, et al. 2006. A study on the activities of carbonic anhydrase of two species of bryophytes, *Tortula sinensis* and *Barbula convoluta*, *Cryptogamie Bryologie*, **27**(3):349-355
- Yang XQ(杨晓青), Zhang SQ(张岁岐), Liang ZS(梁宗锁), et al. 2004. Effects of water stress on chlorophyll fluorescence parameters of different drought resistance winter wheat cultivars seedlings(水分胁迫对不同抗旱类型冬小麦幼苗叶绿素荧光参数的影响)[J]. *Acta Bot Boreal-Occident Sin*(西北植物学报), **24**(5):812-816
- Yu LF(喻理飞), Zhu SQ(朱守谦), Ye JZ(叶镜中), et al. 2003. A study on evaluation of natural restoration for degraded karst forest(退化喀斯特森林自然恢复评价研究)[J]. *Sci Silv Sin*(林业科学), **36**(6):12-19
- Zhang SR(张守仁). 1999. A discussion on chlorophyll fluorescence kinetics parameters and their significance(叶绿素荧光动力学参数的意义及讨论)[J]. *Chin Bull Bot*(植物学通报), **16**(4):444-448

(上接第 46 页 Continue from page 46)

- enrichment procedure to develop multiple repeat classes of cotton microsatellite markers[J]. *Plant Mol Biol*, **22**:85a-85i
- Kumpatla SP, Mukhopadhyay S. 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous [J]. *Genome*, **48**:985-998
- Li YQ(李永强), Li HW(李宏伟), Gao LF(高立峰), et al. 2004. Progress of simple sequence repeats derived from expressed sequence tags(基于表达序列标签的微卫星标记(EST-SSRs)研究进展)[J]. *J Plant Genet Res*(植物遗传资源学报) **5**(1):91-95
- Powell W, Machray GC, Provan J. 1996. Polymorphism revealed by simple sequence repeats [J]. *Trends Plant Sci*, **1**:215-222
- Toth G, Gaspari Z, Jurka J. 2000. Microsatellites in different eukaryotic genomes: survey and analysis [J]. *Genome Res*, **10**:967-981
- Varshney RK, Graner A, Sorrells ME. 2005. Genic microsatellite markers in plants: features and applications[J]. *Trends Biotechnol*, **23**:48-55