

虞 杭,张得芳,樊光辉,等. 枸杞转录组 SSR 分布特征分析及其与基因组 SSR 分布特征的比较[J]. 江苏农业科学,2018,46(14):24-27.
doi:10.15889/j.issn.1002-1302.2018.14.006

枸杞转录组 SSR 分布特征分析及其与基因组 SSR 分布特征的比较

虞 杭¹,张得芳^{1,2},樊光辉^{1,2},王占林^{1,2}

(1. 青海大学,青海西宁 810016; 2. 青海省农林科学院/青海高原林木遗传育种实验室,青海西宁 810016)

摘要:通过 Illumina HiSeq 2500 高通量测序平台,利用 RNA-Seq 技术对青杞 1 号品种整体转录活动进行检测,并对得到的序列采用生物信息学手段进行拼接后查找 SSR,然后与基因组 SSR 进行比较。结果显示,转录组共获得 5 411 个重复单元长度为 2~6 碱基的微卫星重复序列,2 碱基重复的微卫星最为丰富,共 2 666 个,占 49.27%,其中 AG/CT 基序的重复数量最多;3 碱基重复类型有 2 609 个,占重复序列总数的 48.22%;4 碱基重复 118 个,5 碱基重复 10 个,6 碱基重复 8 个,分别占重复序列总数的 2.18%、0.18%、0.15%。而基因组共获得 14 733 个重复单元为 2~6 碱基的微卫星重复序列,3 碱基重复的微卫星最为丰富,共 9 799 个,占 66.51%,其中 GTT/CAA 基序的数量最多。青杞 1 号转录组 SSR 与基因 SSR 在重复序列分布上存在显著差异。

关键词:枸杞;RNA-Seq 技术;转录组;微卫星;遗传图谱

中图分类号: S567.1⁺9 **文献标志码:** A **文章编号:** 1002-1302(2018)14-0024-04

枸杞 (*Lycium barbarum*) 属于茄科 (Solanaceae) 枸杞属

(*Lycium* L.) 落叶灌木,是重要的药用资源,其果、叶、根等均能被利用,在中药配方中有重要地位^[1-2]。枸杞的营养含量非常丰富,富含蛋白质、维生素和多种氨基酸,还有类胡萝卜素以及钙、铁、锌、硒等对人体有益的元素,有很高的药用价值和保健效果。枸杞不仅有美容养生和抗衰老的功效,还具有抗癌、降血糖、降血脂以及护眼等功能^[3-7]。

收稿日期:2017-05-19

基金项目:青海省科技计划(编号:2014-NS-522);青海省自然科学基金青年科学基金(编号:2015-ZJ-926Q);青海省重大科技专项(编号:2015-NK-A2);青海高原林木遗传育种实验室(编号:2017-ZJ-Y29)。

作者简介:虞 杭(1994—),男,浙江义乌人,硕士研究生,主要从事森林培育方面的研究。E-mail:513859965@qq.com。

通信作者:王占林,研究员,主要从事森林培育和林木育种方面的研究。E-mail:1735105720@qq.com。

转录组是指某个物种或细胞在某一条件下所有转录产物的集合。RNA-Seq 技术能够在单核苷酸水平对特定物种的整体转录活动进行检测,以提供最全面的转录组信息^[8-12]。同样,RNA-Seq 技术的应用也非常普遍,利用此技术鉴定油

[11] Rengasamy P. World salinization with emphasis on Australia[J].

Journal of Experimental Botany,2006,57(5):1017-1023.

[12] Liu J Y, Zhu Y, Huang L L, et al. Combined effect of ethylene and salicylic acid - signaling insensitive mutation on *Arabidopsis* response to low temperature[J]. *Biologia Plantarum*,2016,60(3):523-531.

[13] Xia J C, Zhao H, Liu W Z, et al. Role of cytokinin and salicylic acid in plant growth at low temperatures[J]. *Plant Growth Regulation*,2008,57(3):211-221.

[14] Hao L, Zhao Y, Jin D D, et al. Salicylic acid - altering *Arabidopsis* mutants response to salt stress[J]. *Plant and Soil*,2012,354(7):81-95.

[15] Zhang J Y, Qu S C, Du X L, et al. Overexpression of the *Malus hupehensis* *MhTGA2* gene, a novel bZIP transcription factor for increased tolerance to salt and osmotic stress in transgenic tobacco [J]. *International Journal of Plant Sciences*,2012,59(5):441-453.

[16] Quilis J, Peñas G, Messeguer J, et al. The *Arabidopsis* *AtNPR1* inversely modulates defense responses against fungal, bacterial, or viral pathogens while conferring hypersensitivity to abiotic stresses in transgenic rice[J]. *Molecular Plant - Microbe Interactions*,2008,21

(9):1215-1231.

[17] 陈 宇. 油菜在多种逆境条件下 qRT-PCR 内参基因的稳定性评估及抗逆相关基因的表达分析[D]. 镇江:江苏大学,2014.

[18] Chou K C, Shen H B. Plant - mPLoc: a top - down strategy to augment the power for predicting plant protein subcellular localization[J]. *PLoS One*,2010,5(6):e11335.

[19] Wang Z, Chen Y, Fang H, et al. Selection of reference genes for quantitative reverse - transcription polymerase chain reaction normalization in *Brassica napus* under various stress conditions[J]. *Molecular Genetics and Genomics*,2014,289(5):1023-1035.

[20] 汪 雷,刘 瑶,丁一娟,等. 油菜菌核病研究进展[J]. 西北农林科技大学学报(自然科学版),2015,43(10):85-93.

[21] Kallenbach M, Alagna F, Baldwin I T, et al. *Nicotiana attenuata* *SIPK*, *WIPK*, *NPR1*, and fatty acid - amino acid conjugates participate in the induction of jasmonic acid biosynthesis by affecting early enzymatic steps in the pathway [J]. *Plant Signaling & Behavior*,2010,152(1):96-106.

[22] Yi S Y, Min S R, Kwon S Y. *NPR1* is instrumental in priming for the enhanced flg22 - induced MPK3 and MPK6 activation [J]. *Plant Pathology Journal*,2015,31(2):192-194.

菜(*Brassica campestris* L.)叶片干旱胁迫应答相关基因,从转录组水平揭示油菜适应干旱胁迫环境的分子机制^[13];在果树学中通过追踪柑橘(*Citrus reticulata* Blanco)、葡萄(*Vitis vinifera* L.)、香蕉(*Musa nana* Lour.)等10个常见果树的RNA-Seq实例,研究具体应用进展^[14];在地道药材形成机制研究以及改善牦牛基因结构信息上也起到了至关重要的作用^[15]。

简单重复序列(simple sequence repeat,简称SSR)又称微卫星,是核苷酸串联重复单元(1~6个核苷酸),在真核及原核生物基因组中都有分布,SSR标记可分为表达序列标签SSR(EST-SSR)和基因组SSR^[16-22]。此标记技术具有高多态性、高重复性和较广的覆盖面等特点,目前在构建植物的遗传图谱、分析遗传多态性上有较普遍的应用,对遗传多样性评价和种质鉴定也起到了很大的帮助^[17,23-25]。EST-SSR是在已有EST序列的基础上,用电子筛选鉴别SSR,再用PCR检测,避免了SSR引物开发过程中的克隆和测序步骤,很好地利用了现有数据,节约了开发成本^[26]。

本研究对枸杞转录组进行测序,通过对不同基因序列长度类型微卫星的统计与分析,了解枸杞转录组SSR的特征及组成,并将其与基因组SSR进行分析比较,从而进一步了解枸杞基因组SSR和转录组SSR在分布特征上的变化规律。

1 材料与方法

1.1 供试材料

2015年春天在青海省林业科学研究所枸杞种质资源圃,采集青杞1号的新梢顶端刚长出的幼叶作为试验材料,采集叶片后立即放入液氮罐冷冻并带回实验室,保存于-80℃冰箱中。

1.2 RNA提取

采用植物总RNA提取试剂盒(北京天根生化科技有限公司)提取总RNA,按照试剂盒要求进行操作。

1.3 构建文库及上机测序

用带有Oligo(dT)的磁珠富集mRNA,然后加入fragmentation buffer将mRNA打断成短片段,以mRNA为模板,用6碱基随机引物合成第1链cDNA,然后加入缓冲液、dNTPs、DNA polymerase I、RNase H合成第2链cDNA,再用AMPure XP beads纯化双链cDNA。将处理后的cDNA先进行末端修复、加A尾并连接测序接头,再用AMPure XP beads进行选择。最后进行PCR扩增,并用AMPure XP beads纯化产物,得到最终的文库。把不同文库按照有效浓度及目标下机数据量的需求pooling后进行Illumina HiSeq测序。

1.4 分析方法

利用Misa软件对测序结果进行SSR查找,从较短的表达序列标签中挖掘SSR标记位点,识别SSR的重复单元并找寻其侧翼序列,并对数据进行整理分析^[27]。其中参数设定为:1~6碱基重复最短的重复数,依次为10、6、5、5、5、5个重复。

2 结果与分析

2.1 枸杞转录组与基因组微卫星丰度及分布密度分析

在本次测序结果中,对枸杞转录组多碱基重复的微卫星进行统计。2碱基重复单元和3碱基重复单元的SSR含量最多,分别占SSR统计总数的49.27%、48.22%,之后依次是4

碱基重复(2.18%)、5碱基重复(0.18%)、6碱基重复(0.15%)。含有66种不同重复碱基组成的2碱基重复微卫星,还有由不同重复碱基组成的3碱基、4碱基、5碱基、6碱基重复微卫星,分别有188、59、9、8种。通过对测得的低覆盖度的枸杞转录组的序列进行微卫星查找,从总长为75 398 046 bp的111 921个重叠中共查找出5 411个SSR;而基因组中SSR含量最多的是3碱基重复单元,约占SSR统计总数的66.51%。对测得基因组的序列进行微卫星查找,从总长为260 163 757 bp的880 315个重叠中共查找出14 733个SSR^[28]。利用微卫星密度公式计算得,转录组平均每13 934.2 bp出现1个SSR,而基因组平均每17 658.6 bp出现1个SSR^[28]。

2.2 枸杞转录组与基因组微卫星优势重复单元类型碱基构成

由图1可知,转录组中在66种2碱基重复单元中,AG/CT基序的数量最多,共1 285个,占48.20%,其次是AT/AT共845个,占31.70%;AC/GT共531个,占19.92%;CG/CG共5个,占0.19%。

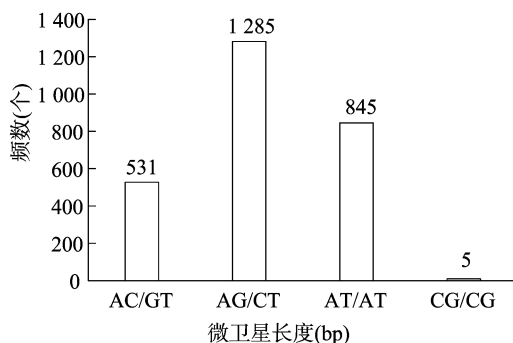


图1 2碱基重复微卫星

由图2可知,在188种3碱基重复单元中,AAC/GTT基序的数量最多,共690个,占26.44%,其次是AAG/CTT,共675个,占25.87%;ATC/ATG共345个,占13.22%;AAT/ATT共340个,占13.03%;ACC/GGT共185个,占7.10%;其余3碱基重复单元数量均较少。

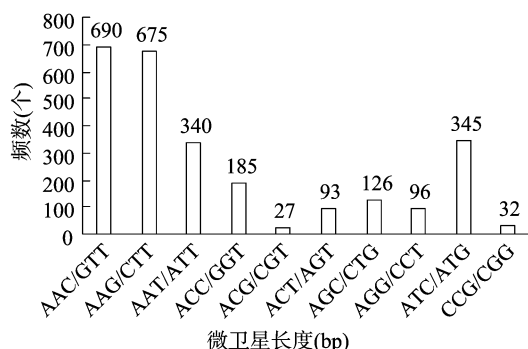


图2 3碱基重复微卫星

由图3可知,59种4碱基重复单元中,重复类型数量最大的基序为AAAT/ATTT,共35个,占29.66%,其次主要是AAAG/CTTT共34个,占28.81%;AAAC/GTTT共20个,占16.95%。

由图4可知,5碱基重复单元中,AAAAG/CTTTT共2个,占20%,其余均为1个,各占10%。

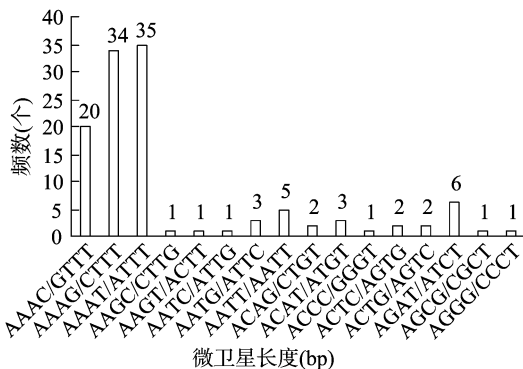


图3 4碱基重复微卫星

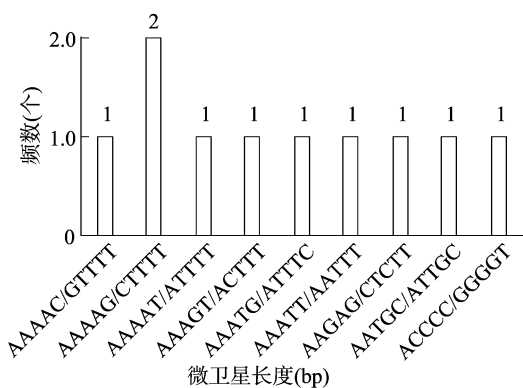


图4 5碱基重复微卫星

由图5可知,6碱基重复单元中的基序均为1个,各占12.50%。

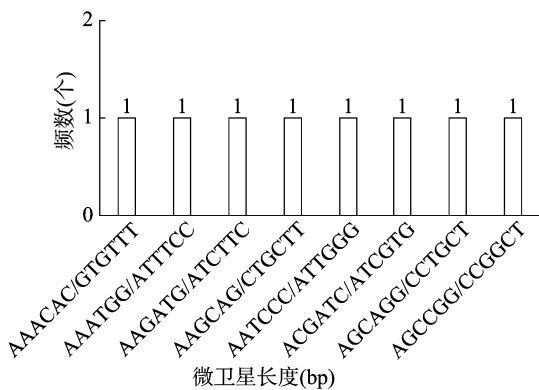


图5 6碱基重复微卫星

综合分析,转录组中2碱基重复数量最多,其次是3碱基重复,其中2碱基重复的优势序列为AG/CT、AT/AT,而6碱基重复丰度最低。

基因组中,在2碱基重复微卫星中AT/TA基序重复的数量最多,共1806个,占44.7%;在61种3碱基重复单元中,GTT/CAA基序的数量最多,共2744个,占28.0%,其次是ACA共738个,占7.5%;ATC共709个,占7.2%;AAC共483个,占4.9%;ATG共427个,占4.4%;其余3碱基重复单元数量均较少^[28]。

综合分析,基因组中3碱基重复数量最多,2碱基重复数量仅次于3碱基重复,其中3碱基重复的优势序列为GTT/CAA、ACA、ATC;同样6碱基重复丰度最低。

经计算得,枸杞转录组SSR的平均长度为16.19 bp,最

长为62 bp,最短为12 bp,其中主要以12~18 bp的微卫星为主,占总数的81.76%,长度>18 bp的微卫星仅占总数的18.24%。枸杞基因组SSR的平均长度为13.81 bp,最长为36 bp,最短为12 bp。其中主要以12~14 bp的微卫星为主,占总数的64.80%。长度>14 bp的微卫星仅占总数的35.20%^[28]。

3 结论与讨论

经测序分析,转录组中在总长为75 398 046 bp的有效读长中发现有5 411个微卫星分布,平均每13 934.2 bp出现1个微卫星;基因组中在总长为260 163 757 bp的有效读长中发现有14 733个微卫星分布,平均每17 658.6 bp出现1个微卫星^[28]。转录组SSR和基因组SSR的分布存在明显差异,转录组2碱基重复数量最多,而基因组中则是3碱基最丰富(表1)。并利用SSR密度计算公式 $D = N/L$ 算出密度。式中: L 代表重叠群总长(Mb); N 代表各重复微卫星数量(个); D 代表不同重复微卫星密度(个/Mb)。

表1 转录组与基因组不同长度重复单元微卫星所占比例及分布密度比较

重复类型	SSR 数量		所占比例(%)		密度(个/Mb)	
	转录组	基因组	转录组	基因组	转录组	基因组
2碱基重复	2 666	4 042	49.3	27.4	35.4	15.5
3碱基重复	2 609	9 799	48.2	66.5	34.6	37.7
4碱基重复	118	519	2.2	3.5	1.6	2.0
5碱基重复	10	281	0.2	1.9	1.3×10^{-7}	1.1
6碱基重复	8	92	0.1	0.6	1.1×10^{-7}	0.4
总数	5 411	14 733	100	100	71.6	56.6

枸杞转录组平均每13 934.2 bp出现1个微卫星,分布密度为71.6个/Mb。2碱基重复数量最多,占总数的49.3%,其次3碱基重复数量与2碱基相近,占48.2%,其余重复所占比例均较少。2碱基重复单元中,基序数量从多到少分别为AG/CT、AT/AT、AC/GT;3碱基重复单元中分别为AAC/GTT、AAG/CTT、ATC/ATG等。与转录组相比,基因组平均17 658.6 bp出现1个微卫星,分布密度为56.6个/Mb,小于转录组分布密度。基因组中3碱基重复最为普遍,所占比例较大,为66.5%,其次是2碱基重复,为27.4%,2、3碱基重复数量差距较大,其余重复所占比例均较小。2碱基重复单元中,基序数量较多的分别为AT/TA、GT/TA、AC/CA;3碱基重复单元中分别为GTT/CAA、ACA、ATC等。转录组SSR与基因组SSR分布特征差异较大,转录组中2碱基重复数量最多,而基因组中则是3碱基重复数量最多,基因组基序数量与种类也与转录组存在显著差异。

同为茄科植物的辣椒(*Capsicum annuum* L.),从转录组SSR上看,不考虑单碱基重复,2、3碱基重复分别占19.16%、23.18%,碱基重复数量相近,其余重复占很小比例。2碱基重复单元中以AG/CT基序数量最多,占58.02%;3碱基重复单元中以AAC/GTT基序数量最多,占27.8%^[29]。辣椒基因组SSR基序中2、3碱基重复分别占22.59%、29.01%,AT/AT、AAT/ATT分别是2碱基和3碱基重复单元中数量最多的基序。由此可见辣椒转录组SSR与基因组SSR分布特征较为类似,2、3碱基的重复数量都较为接近,且3碱基重复略多

于2碱基重复^[30]。

茄科的另一物种马铃薯 (*Solanum tuberosum* L.) 与上述2种植物的SSR分布特征又有不同。马铃薯转录组SSR,以宁夏薯4号为例,SSR重复类型从2核苷酸重复到9核苷酸重复均有,其中3碱基重复为主要的重复类型,占58.2%,其次是6碱基重复和2碱基重复,分别占12.8%、10.7%,3碱基重复单元中GAA/TTC基序出现的次数最多,占7%^[31]。在基因组SSR中,单碱基重复、2碱基重复、3碱基重复这3种重复类型占总SSR位点的94.16%,而4~6碱基重复占1.24%^[32]。可见马铃薯转录组SSR与基因组SSR无论是重复的碱基类型还是主要重复类型都存在较大差异。

有研究发现,单碱基重复和2碱基重复类型的SSR大多位于非编码区,而有部分3碱基重复类型位于编码区,在试验中发现,基因组SSR中,3碱基重复类型要明显多于2碱基重复类型。简化基因组测序是通过对基因组特定区域进行测序来反映部分基因组序列结构信息的测序技术,而表达序列标签(EST)中SSR结构及分布广,不仅可以存在于内含子,也存在于编码区、非编码区和调控区,数量庞大的SSR在基因组中分布均匀,可代表整个基因组。可见在转录组SSR中2、3碱基重复所占比例较大,而且重复数量相近。

根据本次试验结果分析表明,枸杞转录组SSR与基因组SSR分布在主要碱基重复类型上和主要的基序数量上都存在较显著差异,基因组序列能够帮助转录组注释数据,而转录组数据也可对校正基因组注释信息和发现新基因起到帮助,可为研究枸杞性状及多态性提供参考依据。

参考文献:

- [1] 如克亚·加帕尔,孙玉敬,钟烈州,等. 枸杞植物化学成分及其生物活性的研究进展[J]. 中国食品学报,2013,13(8):161-172.
- [2] 刘毓璟. CO₂浓度倍增处理下宁夏枸杞果实转录组测序及分析[D]. 银川:宁夏大学,2016.
- [3] 张惠玲. 枸杞的营养与功效[J]. 农业科学研究,2011,32(3):59-62.
- [4] 侯志恒. 青海省柴达木盆地枸杞品质地位与综合配套栽培技术[J]. 北方园艺,2010(12):190-192.
- [5] Zhao Z K, Yu H L, Liu B, et al. Antioxidative mechanism of *Lycium barbarum* polysaccharides promotes repair and regeneration following cavernous nerve injury[J]. Neural Regeneration Research, 2016, 11(8):1312-1321.
- [6] Mi X S, Chiu K, Van G, et al. Effect of *Lycium barbarum* polysaccharides on the expression of endothelin-1 and its receptors in an ocular hypertension model of rat glaucoma[J]. Neural Regeneration Research, 2012, 7(9):645-651.
- [7] Xing X W, Liu F Y, Xiao J, et al. Neuro-protective mechanisms of *Lycium barbarum* [J]. Neuro Molecular Medicine, 2016, 18(3):253-263.
- [8] 张春兰,秦孜娟,王桂芝,等. 转录组与RNA-Seq技术[J]. 生物技术通报,2012(12):51-56.
- [9] 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):1191-1202.
- [10] Lan D L, Xiong X R, Wei Y L, et al. RNA-Seq analysis of yak ovary: improving yak gene structure information and mining reproduction-related genes [J]. Science China Life Sciences, 2014, 57(9):925-935.
- [11] 张婷. 长雄野生水稻地下茎及耐冷性状功能基因组学及比较转录组学分析[D]. 武汉:武汉大学,2013.
- [12] Conesa A, Madrigal P, Tarazona S, et al. Erratum to: a survey of best practices for RNA-seq data analysis [J]. Genome Biology, 2016, 17:13.
- [13] 卢坤,张琳,曲存民,等. 利用RNA-Seq鉴定甘蓝型油菜叶片干旱胁迫应答基因[J]. 中国农业科学,2015,48(4):630-645.
- [14] 冯超,朱长青,徐昌杰,等. RNA-Seq在果树学研究中的应用[J]. 果树学报,2014,31(1):115-124.
- [15] 王晓明,宋经元,谢彩香,等. RNA-Seq与道地药材研究[J]. 药学学报,2014,49(12):1650-1657.
- [16] 鄢秀芹,鲁敏,安华明. 刺梨转录组SSR信息分析及其分子标记开发[J]. 园艺学报,2015,42(2):341-349.
- [17] 黄海燕,杜红岩,乌云塔娜,等. 基于杜仲转录组序列的SSR分子标记的开发[J]. 林业科学,2013,49(5):176-181.
- [18] 李满堂,张仕林,邓鹏,等. 洋葱转录组SSR信息分析及其多态性研究[J]. 园艺学报,2015,42(6):1103-1111.
- [19] Kalia R K, Rai M K, Kalia S, et al. Microsatellite markers: an overview of the recent progress in plants [J]. Euphytica, 2011, 177(3):309-334.
- [20] 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.
- [21] Yang Z J, Peng Z S, Yang H. Identification of novel and useful EST-SSR markers from de novo transcriptome sequence of wheat (*Triticum aestivum* L.) [J]. Genetics and Molecular Research, 2016, 15(1):1-13.
- [22] Zhou T, Li Z H, Bai G Q, et al. Transcriptome sequencing and development of genic SSR markers of an endangered Chinese endemic genus *Dipteronia* Oliver (Aceraceae) [J]. Molecules, 2016, 21(3):166.
- [23] 冯涛,刘娟,华夏雪. 利用SSR/SRAP分子标记鉴定桃早熟芽变[J]. 江苏农业科学,2017,45(6):42-44.
- [24] 赵雅楠,王颖,张东杰. 绿豆SSR-PCR反应体系的建立与优化[J]. 江苏农业科学,2017,45(8):23-26.
- [25] 王萍,母洪娜,耿春敏,等. 利用正交试验设计优化桂花的SSR-PCR体系[J]. 江苏农业科学,2017,45(2):44-46.
- [26] 黄心尧,卢茵. EST及EST-SSR分子标记应用[J]. 畜禽业, 2010(8):40-43.
- [27] 王希,陈丽,赵春雷. 利用MISA工具对不同类型序列进行SSR标记位点挖掘的探讨[J]. 中国农学通报,2016,32(10):150-156.
- [28] 党少飞,王占林,张得芳,等. 枸杞基因组微卫星特征分析[J]. 西北林学院学报,2016,31(1):97-102.
- [29] 刘峰,王运生,田雪亮,等. 辣椒转录组SSR挖掘及其多态性分析[J]. 园艺学报,2012,39(1):168-174.
- [30] 程蛟文. 辣椒全基因组SSR和SNP标记开发及应用[D]. 广州:华南农业大学,2016.
- [31] 巩福,程永芳,甘晓燕,等. 马铃薯转录组EST-SSR挖掘及其多样性分析[J]. 分子植物育种,2015,13(7):1535-1544.
- [32] 袁娟. 基于马铃薯全基因组序列SSR标记的开发及验证[D]. 雅安:四川农业大学,2013.