

邢冰,董诚明,魏硕,等. 怀菊转录组中 SSR 位点信息分析[J]. 江苏农业科学,2020,48(11):57-60.
doi:10.15889/j.issn.1002-1302.2020.11.011

怀菊转录组中 SSR 位点信息分析

邢冰¹,董诚明^{1,2},魏硕³,李曼¹,李询¹

(1. 河南中医药大学药学院,河南郑州 450046; 2. 呼吸疾病河南省协同创新中心,河南郑州 450046; 3. 四川大学,四川成都 610000)

摘要:以转录组测序数据为基础,利用 MISA 软件对简单重复序列(simple sequence repeat,简称 SSR)进行检测及分析,分析怀菊转录组中的 SSR 位点信息,并设计 SSR 引物,为后期 SSR 标记的开发和物种遗传多样性检测等提供参考。结果表明,共检测到 8 553 个 SSR 位点,它们分布在 7 369 条 Unigene 中,出现频率为 2.83%,SSR 位点的发生频率为 2.44%;其中单核苷酸重复基序最多、其次为三核苷酸重复基序。怀菊转录组基序长度主要集中于 12~19 bp,多具有中等多态性。对 SSR 序列设计引物,共得到 25 662 对引物可供今后使用。总之,怀菊 SSR 位点类型丰富,具有良好的多态性潜力及可开发性。

关键词:怀菊;SSR;位点信息;转录组

中图分类号: S567.23+9.01 **文献标志码:** A **文章编号:** 1002-1302(2020)11-0057-04

怀菊为菊科植物菊(*Chrysanthemum morifolium* Ramat.)的干燥头状花序^[1],栽培及药用历史悠久,研究表明,经过长期的自然环境改变和人工干预,

怀菊种质可能发生了变异与分化,品质下降^[2]。目前关于怀菊的研究主要集中于人工栽培、组织培养、质量评价等方面,对于其种质资源和分子标记辅助育种的研究尚少^[3-6]。因此,对于怀菊品种鉴别及遗传多样性的研究具有重要意义。

简单重复序列(simple sequence repeat,简称 SSR)是指由 1~6 个核苷酸为重复单位组成的简单串联重复序列。简单重复序列能反映出物种间高度的等位基因多样性,不同物种的重复次数、基

收稿日期:2019-09-20

基金项目:2016 年度中央引导地方科技发展专项资金“河南道地大宗药材种质评价及集约化种植与示范”。

作者简介:邢冰(1995—),女,河南郑州人,硕士研究生,主要从事药用植物研究。E-mail: xingbingywwz@163.com。

通信作者:董诚明,教授,主要从事药用植物研究。E-mail: dcm371@hactcm.edu.cn。

参考文献:

- [1] Liu W C, Kim I H. Effects of dietary xylanase supplementation on performance and functional digestive parameters in broilers fed wheat-based diets[J]. Poultry Science, 2017, 96(3): 566-573.
- [2] 聂国兴,王修启,明红,等. 黑曲霉产木聚糖酶稳定性的研究[J]. 华北农学报, 2004, 19(1): 112-115.
- [3] Stepan C M, Bailey S T, Bhat S, et al. The hormone resistin links obesity to diabetes[J]. Nature, 2001, 409(6818): 307-312.
- [4] Archibald A L, McClenaghan M, Hornsey V, et al. High-level expression of biologically active human alpha 1-antitrypsin in the milk of transgenic mice[J]. Proceedings of the National Academy of Sciences, 1990, 87(13): 5178-5182.
- [5] Niavarani A, Dehghanizadeh S, Zeinali S, et al. Development of transgenic mice expressing calcitonin as a β -lactoglobulin fusion protein in mammary gland[J]. Transgenic Research, 2005, 14(5): 719-727.
- [6] Barash I, Faerman A, Ratovitsky T, et al. Ectopic expression of β -lactoglobulin/human serum albumin fusion genes in transgenic mice;

hormonal regulation and in situ localization[J]. Transgenic Research, 1994, 3(3): 141-151.

- [7] Yin H F, Fan B L, Yang B, et al. Cloning of pig parotid secretory protein gene upstream promoter and the establishment of a transgenic mouse model expressing bacterial phytase for agricultural phosphorus pollution control[J]. Journal of Animal Science, 2006, 84(3): 513-519.
- [8] 张献伟,张冠冠,吴珍芳,等. 木聚糖酶-甘露聚糖酶融合酶基因 Linker 优化及其在猪肾 pK15 细胞中共表达[J]. 中国农业科学, 2013, 46(22): 4774-4783.
- [9] Powell B C. The keratin protein and genes of wool and hair[J]. Wool Technology and Sheep Breeding, 1996, 44(2): 100-118.
- [10] 王春生,安铁洙,白秀娟,等. 绵羊毛发角蛋白结合蛋白启动子的克隆与活性分析[J]. 中国兽医学报, 2008, 28(2): 137-139, 145.
- [11] 郭旭东,尹俊,杨东山,等. IGF-1 毛囊特异表达载体的构建以及转染绒毛羊胎儿成纤维细胞的研究[J]. 畜牧兽医学报, 2009, 40(10): 1460-1467.
- [12] 陈哲,雷明明,于建宁,等. 猪 *RELMB* 基因启动子区克隆及序列分析[J]. 江苏农业学报, 2015, 31(5): 1060-1064.

序类型、长度等差异很大,由此开发的 SSR 分子标记技术具有操作简便、稳定性好、多态性高等优点,已经在品种纯度鉴定、基因功能研究、遗传多样性分析等方面有广泛的应用^[7-8]。转录组测序技术可开发大量的 SSR 标记,且速度快、成本较低^[9],能够克服传统 SSR 分子标记技术的弊端。本研究基于高通量测序结果,获得大量怀菊转录组序列信息,并对转录组数据中的 SSR 位点进行检测,分析其分布及结构特点,以期对怀菊种质资源鉴定、遗传多样性分析及遗传图谱构建等提供基础。

1 材料与方法

1.1 材料

本试验中的转录组测序样本怀菊来源于河南省焦作市温县,种植于河南中医药大学药用植物园温室,采集后于河南中医药大学组织培养实验室进行离体培养得到怀菊愈伤组织、芽丛、幼苗。

1.2 方法

1.2.1 样品检测、文库构建与测序 将样品送至北京百迈客生物科技有限公司提取 mRNA,分别采用 Nanodrop、Qubit 2.0、Aglient 2100 方法检测 RNA 样品的纯度、浓度和完整性等,检测合格后,进行文库构建,分别使用 Qubit 2.0 和 Aglient 2100 方法对文库的浓度和插入片段大小进行检测,使用 Q-PCR 方法对文库的有效浓度进行准确定量,以保证文库质量。库检合格后,利用 Illumina HiSeq™ 2000 高通量测序平台对其进行转录组测序。

1.2.2 怀菊转录组 SSR 的筛选 采用 MISA 软件对组装出来的 Unigene 序列进行 SSR 检测、筛选。各重复次数搜索标准如下:单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸、六核苷酸重复次数至少分别为 10、6、5、5、5、5 次。对不同 SSR 类型在基因转录本的密度分布进行统计。同时筛选被间隔≤100 bp 碱基间隔的复合型 SSR。

1.2.3 怀菊 SSR 引物设计 使用 Primer 3.0 (2.3.5 版,默认参数)软件进行 SSR 引物设计,并针对预测到的每一个 SSR 位点分别设计 3 组引物供后期试验选择。主要的引物参数设置:退火温度(T_m)在 57~63℃,上、下游引物 T_m 值相差小于 5℃;扩增产物大小为 100~280 bp,引物长度为 10~27 bp;GC 含量小于 60%。

2 结果与分析

2.1 怀菊转录组测序结果

怀菊高通量测序后总计产出 83.67 GB 数据,干净读序经 Trinity 组装后最终获得 302 379 条 Unigene,过滤后发现,质量不低于 30 的碱基比例(Q30,测序错误率 < 1%)为 90.92%,表明测序结果良好,可对获得的数据进行进一步分析。

2.2 怀菊 SSR 数量与分布

采用 MISA 软件对 Unigene 进行 SSR 检测,共检出 8 553 个 SSR 位点,它们分布在 7 369 条 Unigene 中,其中有 4 477 条 Unigene 含单个 SSR 位点,有 2 553 条 Unigene 含 2 个及 2 个以上 SSR 位点。SSR 发生频率为 2.44%,出现频率为 2.83%,平均每 10 kb 有 1.27 个 SSR 位点(表 1)。

表 1 怀菊转录组中 SSR 搜索结果

项目	数量
搜索序列总数(条)	302 379
搜索序列总长度(bp)	164 773 301
SSR 总数(个)	8 553
SSR 发生频率(%)	2.44
SSR 出现频率(%)	2.83
10 kb 含 SSR 数目(条)	1.27
含 SSR 位点的序列数目(个)	7 369
含有超过 1 个 SSR 位点的序列数目(条)	2 553
含复合型 SSR 的序列数目(条)	339

2.3 怀菊转录组中 SSR 基元类型和比例

对 SSR 类型进行统计发现,二核苷酸至六核苷酸重复类型均有出现,不同核苷酸基序种类及重复次数差异明显。在考虑到碱基互补配对原则的情况下,单核苷酸重复单元有 4 种,占 60.77%,多数重复次数为 10~22 次,主要为 A/T 基序重复,而 C/G 基序出现频率较低;二核苷酸重复单元有 12 种,占 14.40%,多数重复次数为 6~10 次,重复次数最多的基序是 TG,其次是 TA;三核苷酸重复单元有 63 种,占 19.17%,多数重复次数为 5~7 次,重复次数最多的基序为 TGA,其次是 ATC;四核苷酸重复单元有 48 种,占 0.83%,重复次数最多的基序为 TAAA,其次是 TTAT;五、六核苷酸重复单元分别有 9、3 种,分别占 0.11%、0.04%,重复次数最多的基序分别是 GGGGA 与 TTGATA(表 2)。

2.4 怀菊 SSR 基元长度分布及其多态性

怀菊 SSR 基元长度区间为 10~191 bp。本研

表 2 怀菊转录组 SSR 重复基元序列特征

重复基元类型	数量 (个)	占比 (%)	重复基序种类数 (种)	数量最多重复 基元类型	数量最多基元的 重复次数(次)
单核苷酸	5 198	60.77	4	T	22
二核苷酸	1 232	14.40	12	TG	10
三核苷酸	1 640	19.17	63	TGA	7
四核苷酸	71	0.83	48	TAAA	5
五核苷酸	9	0.11	9	GGGGA	6
六核苷酸	3	0.04	3	TTGATA	6

究筛选出的怀菊 SSR 长度均 ≥ 10 bp,其中 12 ~ 19 bp 的 SSR 有 4 279 个(占比为 50.03%),具有中等多态性;而 ≥ 20 bp 的 SSR 有 621 个(占比为 7.26%),具有较高的多态性。此外,研究发现,高级基元 SSR 的多态性普遍比低级基元的低^[10]。本研究中,低级基元单核苷酸、二核苷酸和三核苷酸共占总 SSR 的 94.34%,而在长度 ≥ 20 bp 的 SSR 中共有 138 个,占长度 ≥ 20 bp 的所有 SSR 的 22.22%,表明大部分怀菊转录组 SSR 具有高多态性潜能(图 1)。

2.5 怀菊 SSR 引物设计

为了筛选出可在试验中应用的怀菊 SSR,使用 Primer 3.0 软件对含有 SSR 位点且 SSR 上下游序列

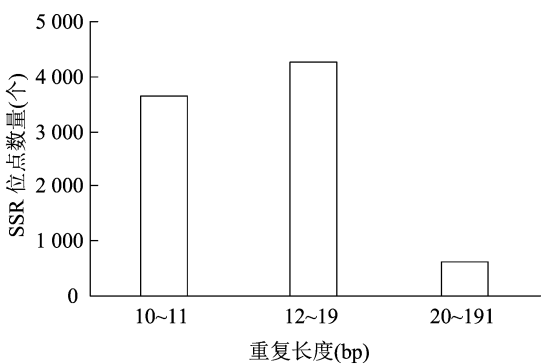


图1 SSR 重复长度分布

均不小于 150 bp 的 unigene 进行引物设计。结果根据 SSR 位点信息,批量设计了 25 662 对引物(部分引物见表 3)。

表 3 怀菊 SSR 引物序列

SSR 类型	SSR 序列	上游引物(5'→3')	下游引物(5'→3')	扩增产物长度 (bp)
单核苷酸重复	(T) ₁₃	CGCTCCAGATCTCAAGTCAA	GACAAAAGGAATGCAGGAAAA	268
	(A) ₁₂	TCAATCTTTTGCTTTTCGCCT	CACAATGAAATGCAGCAACA	208
	(A) ₁₀	TGATGAAAGCACAAAGTGGGA	AGGAGAAGTGCGATTCTCTGA	175
	(T) ₁₂	GAGTTCCAACGGCTGGTTTA	CTCCCGTGTGCTGATTTAAG	280
二核苷酸重复	(TG) ₈	CAAGGGCCACAATGTTTATA	CTTGGCTTTCTGCTTGAAGG	124
	(CT) ₆	ATGTTGGGGATGTTGATGCT	CAGGTAAGCAGGCCGTAGAG	272
	(GT) ₇	GCAAGACCTTGAGGATGAGG	AGTTGGTCTCTACCTCGCC	230
	(AT) ₆	TGTGCGACTGAATGTGTGAA	TCCCACCTAAACCGAATAACA	205
三核苷酸重复	(TGC) ₆	CGCCGACGTGTAGCAGAC	CCGCAGAGAAGCAGAAGC	125
	(TGT) ₆	GCCTCCTGTAGAGGACACCA	CTTCTGCAACCGCTACATCA	229
	(GCA) ₆	GAGCACACCAACGACCAAG	GTAGCGCGAGCTGTAGCC	273
	(TGA) ₆	TTGGGAAGAATCCGATGAAG	CTCTGGTTCCGAGTCTCCAG	268
四核苷酸重复	(TAAA) ₅	TTGGTGTTATAACTTTTCCTCAAA	CCATTGTGTGCTTGTTTTTTCA	271
	(TTAT) ₅	AAACTCCAAAAGGCAAATTTCA	CCTGCAAAACTCACAAGCAA	135
	(TGTT) ₅	TTTCTTGCCTTTGGTTGAGAA	AAGCGAACGCCACAACACTCT	130
五核苷酸重复	(GGCCG) ₅	GCCACGAGGACTGGGATAC	GGGGAATTCGAGGAGGTAAG	280
	(CTAAA) ₅	TGTCATCTTTTGTGTCATC	AAAGCAACAGCAGGGTCACT	168
六核苷酸重复	(TTGATA) ₆	GAAATTAATGGCTGGGAGCA	AGGCAAACAAACCATCCAAC	274
	(AACCTG) ₅	AAGGGGTCCAAAAGTGATCC	ATACTCCTGATGCCCCGTAC	270

注:表中括号内内容表示重复单元;括号外数字表示重复次数。

3 讨论与结论

近年来,通过转录组测序技术开发分子标记已成为一种主要的方式。Han 等基于转录组测序结果对野菊花 SSR 分子标记进行开发,并对其遗传分化等方面开展系统研究^[11]。Kobeissi 等使用 SSR 标记评估菊花遗传多样性,以利于种质资源保护和遗传育种^[12]。张运兴等利用 MISA 软件对 SSR 位点进行搜索,并开发系列引物,为菊花种质资源和遗传多样性分析提供依据^[13]。从本研究的结果来看,首先进行全基因组测序,再分析 SSR 信息,不仅可以得到更多的信息,准确性也大大提高。然而,由于不同品种的菊花之间基因组差异巨大,仅参考菊花基因组序列,并不能系统地对怀菊进行分析。本研究针对怀菊基因组进行初步组装,并进行 SSR 分析及引物设计,这有利于怀菊的种质鉴定,同时也对怀菊的遗传分析具有重要意义。

怀菊转录组测序后共获得 302 379 条 Unigene,利用 MISA 软件进行搜索,共发现 8 553 个 SSR 位点,其中以单核苷酸和三核苷酸重复为主,占 SSR 总数的 79.94%。张运兴等的研究表明,菊花中以三核苷酸重复类型出现频率最高^[13],而本研究发现,怀菊中的单核苷酸最多,可能是由于不同植物 SSR 的主导重复基元类型有所不同^[14-15],范三红等认为,可能是由于这种重复基元编码相应蛋白质的出现率较高^[16]。同时,样品、测序技术的不同,研究中检测及评价方法的不同,也可能会造成同类型物种的 SSR 信息出现一定的差异。本研究发现,A/T、TGA/ATC 分别是怀菊中单核苷酸和三核苷酸 SSR 类型的优势重复基元。有研究表明,单核苷酸中 A/T 重复的数量居多,而双子叶植物中三核苷酸主要重复类型是 AAG/CTT^[17]。张华丽等的研究表明,在万寿菊三核苷酸 SSR 类型中出现次数最多的重复基元是 ATC/ATG^[18]。初步分析这一差异可能与怀菊的 SSR 特性和物种的差异有关。SSR 长度是影响其多态性高低的重要因素^[19],本研究发现,怀菊中长度大于 12 bp 的 SSR 共有 4 900 个,占总数的 57.29%,具有中等以上的多态性,可以重点利用这类 SSR 进行怀菊分子标记的开发及遗传多样性等方面的研究。

综上所述,通过检测怀菊转录组序列中 SSR 信息,得到数量多、可用性强的 SSR 位点,对于加速怀菊 SSR 标记的开发、种质鉴定及遗传性状分析等研

究具有重要意义。

参考文献:

- [1] 国家药典委员会. 中华人民共和国药典(一部)[M]. 北京:化学工业出版社,2005:310-311.
- [2] 刘晓薇,张 飞,陈随清,等. 不同商品规格怀菊花的质量特征分析[J]. 中华中医药杂志,2018,33(6):2650-2655.
- [3] 张红瑞,黄 勇,周 艳,等. 河南 6 个栽培类型药菊内在质量的研究[J]. 中药材,2017,40(7):1507-1510.
- [4] 张红瑞,周 艳,黄 勇,等. 采收时间对 6 个栽培类型药菊产量品质的影响[J]. 山东农业科学,2016,48(7):82-85.
- [5] 周 洲,陈双臣,陈迪新,等. 怀菊组织培养与快速繁殖[J]. 时珍国医国药,2011,22(6):1452-1453.
- [6] 田 硕,苗明三. 菊花的研究及应用现状[J]. 中医学报,2014,29(3):378-380.
- [7] Senan S, Kizhakayil D, Sasikumar B, et al. Methods for development of microsatellite markers: an overview [J]. Notulae Scientiae Biologicae, 2014, 6(1): 1-13.
- [8] Sharma R, Maloo S R, Choudhary S, et al. Microsatellite markers: an important DNA fingerprinting tool for characterization of crop plants [J]. Journal of Plant Science Research, 2015, 31(1): 83.
- [9] 张 震,许彦明,陈永忠,等. 油茶转录组测序与 SSR 特征分析[J]. 西南林业大学学报,2018,38(6):63-68.
- [10] Zhang P, Dreisigacker S, Melchinger A E, et al. Quantifying novel sequence variation in CIMMYT synthetic hexaploid wheats and their backcross - derived lines using SSR markers [J]. Molecular Breeding, 2005, 15(1): 1-10.
- [11] Han Z Z, Ma X Y, Min W, et al. SSR marker development and intraspecific genetic divergence exploration of *Chrysanthemum indicum* based on transcriptome analysis [J]. BMC Genomics, 2018, 19(12): 291.
- [12] Kobeissi B, Saidi A, Kobeissi A, et al. Applicability of scot and SSR molecular markers for genetic diversity analysis in *Chrysanthemum morifolium* genotypes [J]. Biological Sciences, 2018, 89(3): 1-11.
- [13] 张运兴,李卫国,申亚琳,等. 菊花 EST-SSR 标记的开发与应用[J]. 武汉大学学报(理学版),2013,59(4):357-362.
- [14] 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.
- [15] 陈 茵,李翠婷,姜倪皓,等. 灯盏花转录组中 SSR 位点信息分析及其多态性研究[J]. 中国中药杂志,2014,39(7):1220-1224.
- [16] 范三红,郭嵩光,单丽伟,等. 拟南芥基因密码子偏爱性分析[J]. 生物化学与生物物理进展,2003,30(2):221-225.
- [17] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes [J]. Nature Genetics, 2002, 30(2): 194-200.
- [18] 张华丽,丛日晨,王茂良,等. 基于万寿菊转录组测序的 SSR 标记开发[J]. 园艺学报,2018,45(1):159-167.
- [19] 王 森,张 震,姜倪皓,等. 半夏转录组中的 SSR 位点信息分析[J]. 中药材,2014,37(9):1566-1569.