



引用格式:刘伟,蔡英丽,何培新.粗柄羊肚菌转录组的 SSR 分布和序列特征分析[J].轻工学报,2017,32(2):33-39.

中图分类号:Q75;S567.3 文献标识码:A

DOI:10.3969/j.issn.2096-1553.2017.2.006

文章编号:2096-1553(2017)02-0033-07

粗柄羊肚菌转录组的 SSR 分布和序列特征分析

Distribution and sequence characteristics of SSR in transcriptome of *Morchella crassipes*

刘伟^{1,2},蔡英丽²,何培新¹

LIU Wei^{1,2}, CAI Ying-li², HE Pei-xin¹

1. 郑州轻工业学院 食品与生物工程学院,河南 郑州 450001;

2. 华中农业大学 应用真菌研究所,湖北 武汉 430071

1. College of Food and Bioengineering, Zhengzhou University of Light Industry, Zhengzhou 450001, China;

2. Institute of Applied Mycology, Huazhong Agricultural University, Wuhan 430071, China

关键词:
简单重复序列;粗柄
羊肚菌;遗传标记;转
录组测序

Key words:
simple sequence repeat
(SSR);*Morchella*
crassipes; genetic
marker; transcriptome
sequencing

摘要:基于第二代转录组测序技术,对采集自河南省新郑市的粗柄羊肚菌(*Morchella crassipes*)M10 菌株进行了深度转录组测序,分析了全转录组简单重复序列 SSR (simple sequence repeat) 分布和序列特征.结果表明:1) 总计检测到 13 044 个 SSR 位点,分布在 8851 条 Trinity genes 中,SSR 在转录组序列中的平均密度为 2067.77 bp/个,远高于其他真菌,其中 2858 条 Trinity genes 中包含超过 1 个 SSR 位点,复合型 SSR 位点 1925 个. 2) 单碱基重复基元最多,为 8600 个,其中 A/T 重复类型远高于 C/G 重复类型,与多数大型真菌相同;其次是三碱基重复基元,为 2497 个,其中 ACC/GGT, AGC/CTG 和 AGG/CCT 重复类型占 60% 以上,且明显以 GC 的高频概率出现,在 SSR 分子标记或其他重复序列多态性标记的开发上,可优选富含 GC 的 SSR 设计引物.

收稿日期:2016-12-19

基金项目:河南省科技创新杰出青年项目(134100510017)

作者简介:刘伟(1984—),男,河南省镇平县人,华中农业大学助理研究员,主要研究方向为蕈菌遗传育种.

通信作者:何培新(1970—),男,河南省民权县人,郑州轻工业学院教授,博士,主要研究方向为生物工程.

Abstract: The transcriptome of *Morchella crassipes* M10 collected from Xinzheng of He'nan Province was sequenced by the next generation sequencing technology, and distribution characteristic of simple sequence repeats (SSR) were analyzed. The results showed that: 1) A total of 13 044 SSR sites were detected, which was distributed in 8851 Trinity genes, about 2858 Trinity genes had more than one SSR loci and the number of compound formation SSR loci was 1925. The average density of SSR loci was 2067.77 bp/piece in the transcriptome sequence of *Morchella crassipes*, which was much higher than that of other fungi. 2) The number of single base repeat motif was 8600, the largest one, among which the number of A/T repeat types was higher than C/G repeat types and the same as most macro-fungi; the number of tri-nucleotide (TGC) repeat motif was 2497, among which ACC/GGT, AGC/CTG and AGG/CCT accounted for 60% and obviously appeared with GC high probability. In the development of SSR molecular marker or other repeat polymorphic marker, SSR abundant with GC could be chosen as the best primers.

0 引言

简单重复序列 SSR (simple sequence repeat) 是一类由 1—6 个碱基组成的基本串联重复核苷酸序列, 也称为微卫星序列^[1]. SSR 广泛分布于真核生物基因组中. SSR 分子标记是基于重复序列两侧特定的保守序列设计引物, 来扩增检测重复序列长短差异的 DNA 分子标记. SSR 标记因其多态性高、重复性好、共显性、分布均匀、覆盖面广等优点, 广泛应用于生物遗传图谱绘制、遗传多样性分析、品种鉴定、基因定位和分子标记辅助育种等方面^[2]. SSR 标记按来源可分为基因组 SSR (genomic SSR, gSSR) 和表达序列标签 SSR (expressed sequence tag SSR, EST-SSR). 与 gSSR 相比, EST-SSR 源于基因的转录区域, 其多态性可能直接与基因的功能相关, 因此应用范围更广, 通用性更高^[3]. 基于转录组数据开发 SSR 标记在作物方面应用较多^[4]; 关于大型真菌, 仅对糙皮侧耳 (*Pleurotus ostreatus*)、灵芝 (*Ganoderma lucidum*)、黑木耳 (*Auricularia auricula*)、毛木耳 (*A. polytricha*)、茯苓 (*Poria cocos*) 等有所研究^[5-8].

新一代高通量测序技术的成熟和测序成本的急剧下降, 为开发功能基因的 SSR 标记创造了条件. 羊肚菌 (*Morchella* spp.) 是一类珍稀、名贵的食药兼用真菌, 具有较高的营养和保健

价值. 近年来, 梯棱羊肚菌 (*Morchella importuna*) 等黑色羊肚菌支系品种的人工种植在我国发展较快^[9]. 然而, 由于基础研究薄弱, 粗柄羊肚菌 (*Morchella crassipes*) 等黄色羊肚菌支系的种类至今没能成功驯化栽培. 分子标记开发是作物遗传育种的基本手段. 开发的羊肚菌分子标记主要集中于 RAPD, RFLP, AFLP, SCAR, SNP 等^[10-14], 关于羊肚菌 SSR 标记开发的研究鲜见报道. 鉴于此, 本研究拟基于粗柄羊肚菌的深度转录组 De Novo 组装数据, 分析全转录组的 SSR 分布和序列特征, 以期开发羊肚菌 SSR 分子标记提供数据支持, 为粗柄羊肚菌遗传育种研究和人工驯化栽培奠定基础.

1 材料与方法

1.1 供试菌株

野生粗柄羊肚菌标本, 采集自河南省新郑市, 通过组织分离获得供试菌株, 编号 M10. 标本的鉴定通过形态特征结合 rDNA 序列分析完成^[15]. M10 菌株保存储藏于郑州轻工业学院发酵工程研究室.

1.2 RNA 提取和质量控制

M10 菌株在 CYM 完全培养基 (葡萄糖 20 g, 蛋白胨 2 g, 酵母粉 2 g, 硫酸镁 0.5 g, 磷酸氢二钾 1 g, 磷酸二氢钾 0.46 g, 去离子水 1000 mL, pH 自然) 中于 25 °C 下静置培养 10 d,

过滤收集干净的菌丝体,置 $-80\text{ }^{\circ}\text{C}$ 下保存或直接使用. RNA 提取采用 STE 法^[16],简要步骤如下:取 0.5 g 左右菌丝体迅速用液氮研磨至粉状,装入预冷的 1.5 mL 离心管中,加 STE 裂解液裂解后,采用 PCA 液($V(\text{苯酚}):V(\text{氯仿}):V(\text{异戊醇})=25:24:1$)除去蛋白.获得的总 RNA 用 1.0% 的琼脂糖凝胶电泳和 Agilent 2100 生物分析仪进行 RNA 完整性检测,采用 Nanodrop 8000 测定 RNA 溶液的浓度.

1.3 转录组测序

按照 Illumina RNA 建库试剂盒的说明进行文库制备.本实验共计构建两个插入片段长度为 500 bp 的文库,通过 Illumina Nextseq 500 和 Miseq 测序平台(中国科学院水生生物研究所公共技术服务中心基因组信息平台)分别进行 PE 151 和 PE 175 链特异性双末端测序.

1.4 转录组 De Novo 拼接

首先对测序得到的原始数据进行去接头和初步过滤,得初始数据;再通过 FastQC 和 NGSQC Toolkit_v2.3.3 对获得的初始数据进行质量评价和控制,去除低质量的记录;使用 Trimmomatic-0.33 去除记录两端低质量碱基;最后使用 Trinity(v2.1.1)进行转录组 De Novo 从头组装.

1.5 SSR 搜索分析

使用 MISA 软件(<http://pgrc.ipk-gatersleben.de/misa/misa.html>)对转录组 Trinity gene 进行 SSR 位点搜索.设置单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸最少重复次数分别为 10,6,5,5,5,5.其两个复合 SSR 位点间最大间隔碱基数为 100 bp.将生成的文本文件导入 Excel 中,统计分析 SSR 的基本信息并作图.

2 结果与分析

2.1 粗柄羊肚菌转录组的 De Novo 组装

从粗柄羊肚菌培养菌丝体抽提 RNA,经生

物分析仪检测, RIN (RNA integrity number) 值 ≥ 6.5 ,浓度超过 $100\text{ ng}/\mu\text{L}$,达到 RNA 测序建库标准,即转入建库和测序步骤. RNA 测序在 Illumina 测序平台上进行,分别完成 PE 151 和 PE 175 的链特异性双末端测序.对测序获得的原始数据进行去接头、低质量记录、低质量碱基的去除和截断之后,共计获得高质量测序数据 7.6 G(包括 PE 151 4.2 G 和 PE 175 3.4 G),测序深度大于 150 倍(按基因组大小评估,粗柄羊肚菌的基因组预估在 50 M 左右).

使用 Trinity(v2.1.1)转录组组装软件,对粗柄羊肚菌的转录组进行 De Novo 从头组装,共计获得 43 925 条转录本,非冗余 Trinity genes 34 972 条,Contig N50 1219 bp,Contig 平均长度 770.13 bp,Contig 中值长度 440 bp,非冗余转录组碱基总长度 26 932 985 bp,转录组 GC 百分比为 48.29%.最短的 Trinity gene 长度为 224 bp,最长的 Trinity gene 长度为 11 778 bp,67.06%的 Trinity genes 长度小于 1000 bp,28.05%的 Trinity genes 长度为 1000~3000 bp(见表 1).

2.2 粗柄羊肚菌转录组 SSR 的分布特征

用 MISA 生物信息学软件,对粗柄羊肚菌的转录组进行 SSR 特征分析.在 34 972 条非冗余的 Trinity genes 中,总计搜索到 13 044 个 SSR 位点,分布在 8851 条 Trinity genes 中,转录组 SSR 出现的频率为 25.59%;SSR 在转录组序列中的平均密度为 2 067.77 bp/个,其中 2858 条 Trinity genes 中包含超过 1 个 SSR 位点;复合型 SSR 位点 1925 个,占总 SSR 总数的 14.8%.粗柄羊肚菌转录组中存在着各种 SSR 基元类型,不同碱基数量的 SSR 基元差异很大,其中,单碱基的重复个数最多,为 8600 个,占总 SSR 位点数量的 65.93%;其次是三碱基重复 2497 个,占比 19.14%;双碱基重复 1526 个,占比 11.70%;最后为四碱基、五碱基和六碱基重复,分别为 286 个、78 个和 57 个.

表1 粗柄羊肚菌转录组 De Novo 组装结果和转录组 SSR 分布特征信息

Table 1 Results of Trinity De Novo assembly and distribution characteristics of SSR in transcriptome of *Morchella crassipes*

转录组 De Novo 组装信息	数值	转录组 SSR 分布特征信息	数值
全转录本数量/条	43 925	总 SSR 位点数量/个	13 044
非冗余的 Trinity genes/条	34 972	包含 SSR 位点的 Trinity genes/条	8851
Contig N50/bp	1219	转录组 SSR 出现的频率/%	25.59
Contig 中值长度/bp	440	包含超过 1 个 SSR 位点的转录组数量/条	2858
Contig 平均长度/bp	770.13	多位点 SSR 发生的频率/%	8.26
非冗余转录组碱基总长度/bp	26 932 985	发生复合型 SSR 的转录组数量/条	1925
转录组 GC 百分比/%	48.29	复合型 SSR 发生的频率/%	5.56
最短的 Trinity genes 长度/bp	224	单核苷酸重复 SSR 数量/条	8600
最长的 Trinity genes 长度/bp	11 778	双核苷酸重复 SSR 数量/条	1526
长度 ≤ 1000 bp 的核苷酸/条	29 456	三核苷酸重复 SSR 数量/条	2497
长度 1000 ~ 2000 bp 的核苷酸/条	8451	四核苷酸重复 SSR 数量/条	286
长度 2000 ~ 3000 bp 的核苷酸/条	3870	五核苷酸重复 SSR 数量/条	78
长度 ≥ 3000 bp 的核苷酸/条	2148	六核苷酸重复 SSR 数量/条	57

2.3 粗柄羊肚菌转录组 SSR 的重复序列类型

图 1 为粗柄羊肚菌 SSR 基元特征和频率分布图,结合表 1 和图 1 可知,13 044 个转录组位点从属于 307 种碱基重复类型,其中单核苷酸至六核苷酸的重复类型分别有 4 种、12 种、60 种、118 种、60 种和 53 种. 单碱基 T 的重复次数最多(3595 次),占单碱基重复的 42%;其次是单碱基 A 的重复(3481 次),占单碱基重复数量的 40%. 双碱基重复中,CG/GC 重复的 SSR 位点最少,总计 31 个;AG/CT 和 AC/GT 重复类型最多,分别为 658 个和 506 个,分别占总双碱基重复类型的 43% 和 33%. 三碱基 SSR 重复序列中,ACC/GGT、AGC/CTG 和 AGG/CCT 是数量最多的 SSR 基元,分别有 540 个、513 个和 455 个,总计占三碱基重复类型的 60% 以上;其次是 CCG/CGG 和 AAG/CTT 三碱基重复. 41 种四碱基的 SSR 重复类型中,AAAAG/CTTTT 和 AAAAC/GTTTT 的数量较多,分别为 10 个和 6 个. 五碱基重复的 SSR 中,AACAGC/CTGTTG 出现的频率最高,有 7 次. 剩余的四碱基、五碱基和六碱基重复类型中,数量均比较平均,均在 1—4 个之间.

2.4 粗柄羊肚菌转录组 SSR 序列重复次数和序列长度特点

粗柄羊肚菌 6 种 SSR 重复基元的频率分布规律见图 2. 由图 2 可知,6 种碱基类型的 SSR 重复基元中,重复次数越低,SSR 的频率越高. 单碱基重复 10 次的 SSR 数量最多,占单碱基 SSR 总量的 36.24%,其次是 11 次的单碱基重复;单碱基(A)_n 重复的数量最高达 79 次,然后是 75 次和 71 次,远高于其他单碱基的重复长度. 双碱基重复 6 次的 SSR 有 688 个,占双碱基 SSR 数量的 45.08%;双碱基(TG)_n 和(TC)_n 最高重复次数分别为 23 次和 22 次,序列长度分别达 46 bp 和 44 bp. 三碱基(TGC)_n 重复次数最高达 25 次,序列长度达 75 bp;其次是五碱基的(TGGAT)_n,总计重复 14 次,序列长度达 70 bp.

3 结论与讨论

本研究按照通用的 SSR 基元特征参数,对粗柄羊肚菌的全转录组进行了 SSR 分布和序列特征分析,共计在 34 972 条非冗余的 Trinity genes 中搜索得到 13 044 个 SSR 位点,平均 2 067.77 bp 长度的转录序列包含 1 个 SSR 位点,

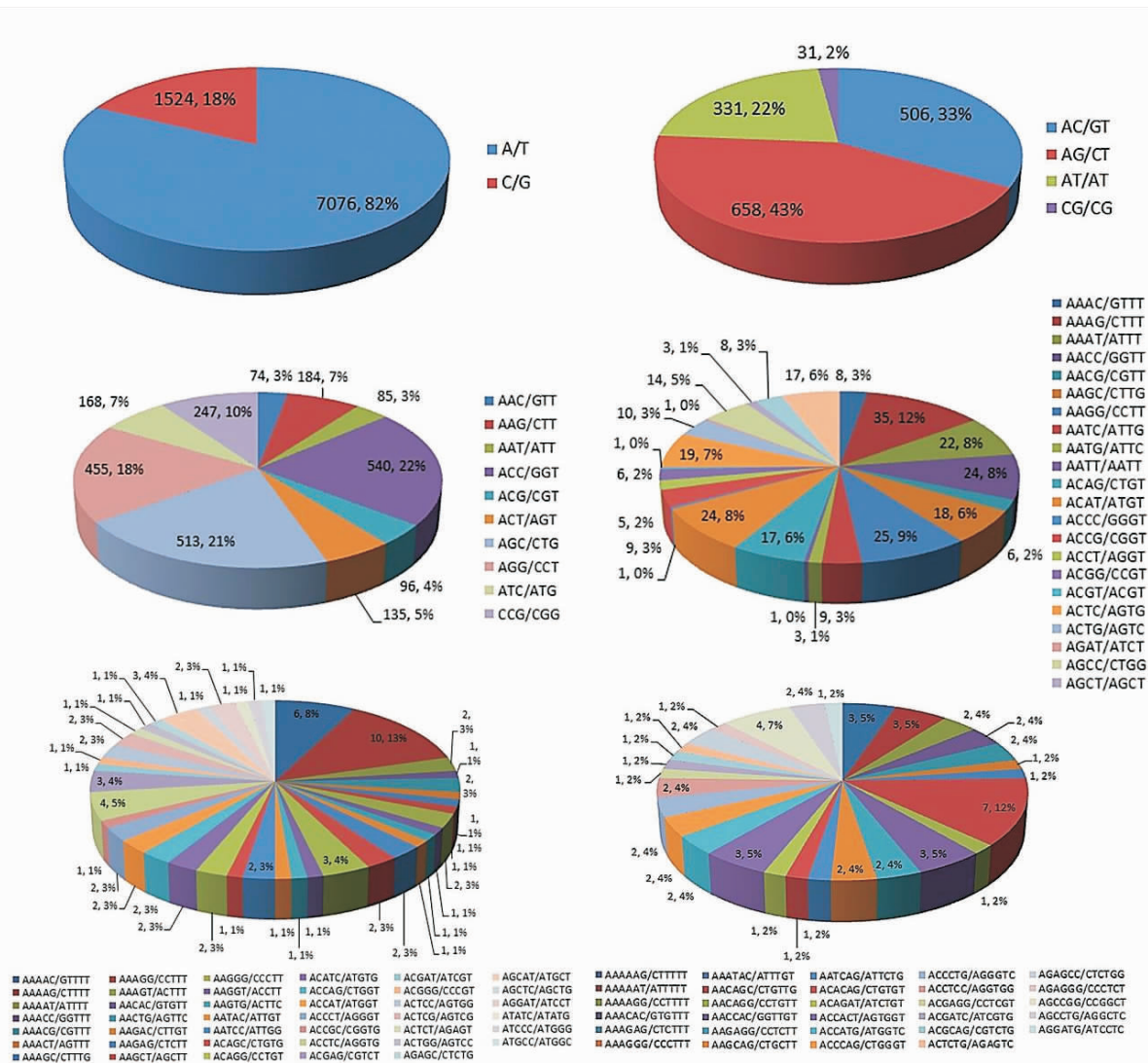


图 1 粗柄羊肚菌 SSR 基元特征和频率分布

Fig. 1 Motif characteristic and frequency distribution of SSR of *Morchella crassipes*

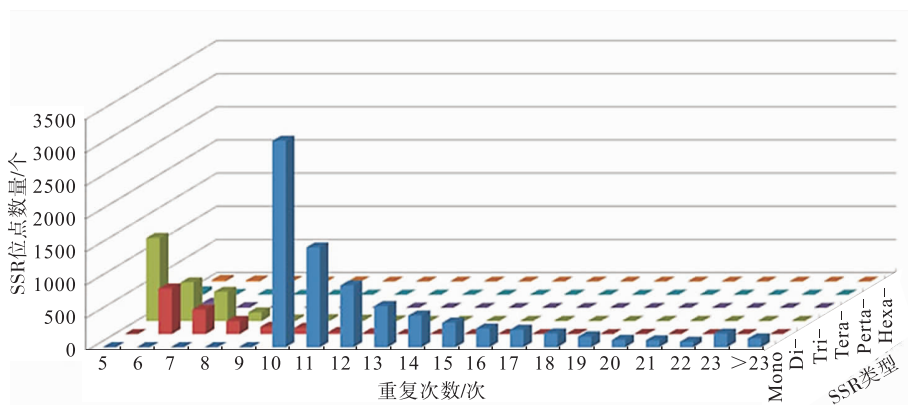


图 2 粗柄羊肚菌 6 种 SSR 重复基元的频率分布规律

Fig. 2 Law of frequency distribution of six SSR repeat motifs of *Morchella crassipes*

远高于每百万碱基转录组序列含有毛木耳 78 个和黑木耳 90 个的 SSR 序列密度^[7],也高于茯苓的每百万碱基转录组序列包含 58.78 个 SSR 序列的平均密度^[8].与植物相比,SSR 位点出现的频率高于洋葱(*Allium cepa*)的 14.1 kb/个^[17]和党参(*Codonopsis pilosula*)的 4.52 kb/个^[18],略低于刺梨(*Rosa roxburghii*)的 1.68 kb/个^[19].粗柄羊肚菌转录组 SSR 的密度也高于灵芝、灰盖鬼伞(*Coprinopsis cinerea*)、裂褶菌(*Schizophyllum commune*)、干朽菌(*Serpula lacrymans*)、双色蜡磨(*Laccaria bicolor*)、黄孢原毛平革菌(*Phanerochaete chrysosporium*)和绵腐卧孔菌(*Postia placenta*)基因组中的 SSR 密度^[6].这表明,粗柄羊肚菌的 SSR 分布频率较高,有助于后续相关分子标记的开发.

对粗柄羊肚菌转录组 SSR 基元序列特征分析发现,在单碱基重复基元中,A/T 重复远高于 C/G 重复序列类型,这与大多数的大型真菌相同,而与毛木耳和裂褶菌中 C/G 重复多于 A/T 重复类型相反^[7,20];三碱基重复基元是除单碱基重复之外最多的重复类型,尤其以 ACC/GGT,AGC/CTG 和 AGG/CCT 居多,占三碱基重复总量的 60% 以上,明显以 GC 的高频概率出现,表明富含 GC 的 SSR 在粗柄羊肚菌中占优势,类似于黑木耳和毛木耳的转录组 SSR 类型^[7].在 SSR 分子标记或其他重复序列多态性如 ISSR(inter-simple sequence repeat)标记的开发上,应优选富含 GC 的 SSR 设计引物.

粗柄羊肚菌转录组的 SSR 分布和序列特征分析,有助于后续开发羊肚菌 SSR 分子标记,为深入开展种质资源分析、遗传连锁图谱构建、系统发育分析和分子标记辅助育种等研究工作提供理论和技术支持.

参考文献:

- [1] 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.
- [2] POWELL W, MACHRAY G C, PROVAN J. Polymorphism revealed by simple sequence repeats [J]. *Trends in Plant Science*, 1996, 1 (7): 215.
- [3] 李炎林, 杨星星, 张家银, 等. 南方红豆杉转录组 SSR 挖掘及分子标记的研究 [J]. *园艺学报*, 2014, 41 (4): 735.
- [4] 王燕龙, 姜言生, 曲志才, 等. SSR 分子标记在作物种质资源鉴定中的应用 [J]. *山东农业科学*, 2012, 44 (10): 11.
- [5] QU J, HUANG C, ZHANG J. Genome-wide functional analysis of SSR for an edible mushroom *Pleurotus ostreatus* [J]. *Gene*, 2016, 575: 524.
- [6] QIAN J, XU H, SONG J, et al. Genome-wide analysis of simple sequence repeats in the model medicinal mushroom *Ganoderma lucidum* [J]. *Gene*, 2013, 512 (2): 331.
- [7] 周雁, 范秀芝, 陈连福, 等. SSR 在黑木耳和毛木耳转录组中的分布和序列特征 [J]. *菌物学报*, 2014, 33 (2): 280.
- [8] 何海, 郭继云, 马毅平, 等. 茯苓转录组 SSR 序列特征及其基因功能分析 [J]. *中草药*, 2015, 46 (23): 3558.
- [9] 何培新, 刘伟, 蔡英丽, 等. 我国人工栽培和野生黑色羊肚菌的菌种鉴定及系统发育分析 [J]. *郑州轻工业学院学报(自然科学版)*, 2015, 30 (3/4): 26.
- [10] DALGLEISH H J, JACOBSON K M. A first assessment of genetic variation among *Morchella esculenta* (Morel) populations [J]. *Journal of Heredity*, 2005, 96 (4): 396.
- [11] SINGH S K, KAMAL S, TIWARI M, et al. Arbitrary primer based RAPD—A useful genetic marker for species identification in morels [J]. *Journal of Plant Biochemistry and Biotechnology*, 2004, 13 (1): 7.

- [12] BUNYARD B A, NICHOLSON M S, ROYSE D J. A systematic assessment of *Morchella* using RFLP analysis of the 28S ribosomal RNA gene [J]. *Mycologia*, 1994, 88(6):762.
- [13] BUSCOT F, WIPF D, DI BATTISTA C, et al. DNA polymorphism in morels: PCR/RFLP analysis of the ribosomal DNA spacers and microsatellite-primed PCR [J]. *Mycological Research*, 1996, 100(1):63.
- [14] PAGLIACCIA D, DOUHAN G W, DOUHAN L A, et al. Development of molecular markers and preliminary investigation of the population structure and mating system in one lineage of black morel (*Morchella elata*) in the Pacific Northwestern USA [J]. *Mycologia*, 2011, 103(5):969.
- [15] 何培新, 刘伟. 粗柄羊肚菌分子鉴定及羊肚菌属真菌系统发育分析[J]. *江苏农业学报*, 2010, 26(2):395.
- [16] 税丕容, 郑晓冰, 林俊芳, 等. 简便高质量的食用菌总 RNA 提取方法[J]. *食用菌学报*, 2008, 15(1):32.
- [17] 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.
- [18] 王东, 曹玲亚, 高建平. 党参转录组中 SSR 位点信息分析[J]. *中草药*, 2014, 45(16):2390.
- [19] 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.
- [20] LI R, ZHU H, RUAN J, et al. De novo assembly of human genomes with massively parallel short read sequencing [J]. *Genome Research*, 2010, 20(2):265.