DOI:10.3724/SP.J.1238.2012.00162

# 辣椒多态性 EST-SSR 标记开发及连锁图谱构建

刘峰<sup>1a,3</sup>,谢玲玲<sup>1b</sup>,欧阳娴<sup>1a</sup>,王运生<sup>2,3</sup>,邹学校<sup>1a\*</sup>

(1.湖南省农业科学院 a.蔬菜研究所;b.西瓜甜瓜研究所,湖南 长沙 410125;2.湖南农业大学 生物安全科学技术 学院,湖南 长沙 410128;3.中国农业科学院 蔬菜花卉研究所,北京 100081)

摘 要: 采用 SSR 检索程序,从辣椒 221 037 条 unigenes 中筛选到 17 319 个 SSR 位点,设计了 10 468 对引物, 并对其进行 E-PCR 多态性检测。结果表明,1 538 对引物具有多态性,其中 554 条辣椒 EST 序列能匹配到番茄基 因组上,构成了 12 条辣椒 EST-SSR 连锁群,平均每条连锁群含 45.75 个 EST-SSR。对匹配到番茄基因组上的 EST-SSR 进行 GO(gene ontology)分类,结果有 481 条 EST 序列被分类,其中以初生代谢、细胞代谢、生物合成 过程为主,分别占 34.13%、28.35%、25.82%。在 KEGG map 中,91 条 EST 序列共涉及到 76 条代谢途径(KEGG), 约 67 条序列参与新陈代谢途径,32 条序列参与次生代谢产物合成途径。

关键词:辣椒;多态性;连锁图谱;表达序列标签技术;简单重复序列
 中图分类号:S641.3
 文献标志码:A
 文章编号:1007-1032(2012)02-0162-06

## Development of pepper EST-SSRs and an SSR-based linkage map

LIU Feng<sup>1a,3</sup>, XIE Ling-ling<sup>1b</sup>, OUYANG Xian<sup>1a</sup>, WANG Yun-sheng<sup>2,3</sup>, ZOU Xue-xiao<sup>1a\*</sup>

(1.a.Institute of Vegetables; b. Institute of Watermelon and Muskmelon, Hunan Academy of Agricultural Sciences, Changsha 410125, China; 2.College of Bio-Safety Science and Technology, Hunan Agricultural University, Changsha 410128, China; 3.Institute of Vegetables and Flowers Chinese Academy of Agricultural Sciences, Beijing 100081, China)

**Abstract**: In this study, 17 319 SSRs were identified from 221 037 ESTs of pepper using SSR finding soft. A total of 10 468 primer pairs were successfully designed and 1 538 primer pairs exhibited polymorphism by E–PCR. Among the polymorphic primer-containing ESTs, 554 ESTs containing SSR motifs were matched onto the tomato genome, forming 12 linkage groups with an average of 45.75 makers per group. Analysis of GO (gene ontology) showed that 481 ESTs were classified mainly to be involved in primary metabolic process, cellular metabolic process and biosynthetic process. In the KEGG map, 91 ESTs were involved in 76 pathways (KEGG), 67 ESTs took part in metabolic pathways and 32 participated in biosynthesis of second metabolites.

Key words: pepper; polymorphism; linkage map; expressed sequence tags (EST); simple sequence repeats (SSR)

目前,DNA分子标记(SSR、RAPD、AFLP、RFLP) 已成为遗传变异分析和作物分子辅助育种研究的重 要工具<sup>[1]</sup>。SSR(simple sequence repeats)分子标记<sup>[2]</sup>由 于具有信息量高、多等位性、共显性、结果稳定等特 点而被认为是非常理想的分子标记系统,在连锁图谱 的构建、基因定位、遗传多样性研究以及进化等方面 被广泛应用<sup>[3-4]</sup>。目前,公共序列数据库里有约200 个植物物种的超过3 000万条的EST。这些EST序列 源已被广泛应用于基因组和转录组分析<sup>[5]</sup>,特别是 含SSR基序的EST序列(EST–SSR)在玉米、大豆、水稻、小麦等作物中已被大量开发,并应用于遗传图 谱构建和遗传多样性分析<sup>[4,6]</sup>。

辣椒(*Capsicum annuum* L.)现有遗传图谱的 SSR 分子标记数量非常有限<sup>[5,7]</sup>,远不能满足辣椒高密度 作图的需要。笔者基于当前已公布的辣椒 EST 序列 (约 119 751 条)和辣椒转录组序列(约 230 000 条)进 行 SSR 挖掘、多态性分析和电子遗传图谱构建,现 将结果报道如下。

收稿日期: 2011-11-03

基金项目:国家自然科学基金项目(31101425)

作者简介:刘峰(1976—),男,湖南岳阳人,博士,助理研究员,主要从事作物遗传育种研究,liufengrich@126.com;\*通信作者, zou\_xuexiao@163.com

#### 1 材料与方法

#### 1.1 植物材料及其 RNA

供试辣椒品种为甜椒HDA149(法国INRA)和牛 角椒 9704B(取自湖南省农业科学院蔬菜研究所)。 每个辣椒材料选20粒种子,30℃恒温箱中育苗,用 TRIzol法分别提取HDA149幼根和9704B幼蕾的 RNA,-80℃保存备用。

#### 1.2 辣椒 EST-SSR 数据分析

对HDA149幼根组织和9704B花蕾组织进行 Illumina转录组测序。另外将NCBIAll Database数据 库(http://www.ncbi.nlm.nih.gov)中所有辣椒EST序列 下载到本地计算机,利用iAssembler (http://bioinfo. bti.cornell.edu/tool/iAssembler/)首先对所有序列进行 混合拼接,随后按来源分别拼接、建库。利用est\_ timmer.perl程序去除拼接序列中过短(<100 bp)的序 列以及mRNA5/端的"帽子"和3/端的"polyA尾巴" 结构。利用MISA程序(http://pgrc.ipk-gatersleben. de/misa/)对辣椒拼接序列进行SSR位点搜索。SSR搜 索条件为1~6 bp重复基序,SSR长度大于10 bp。

#### 1.3 辣椒 EST-SSR 标记开发

对含有SSRs位点的所有拼接序列建立fasta文本,在服务器上调用Primer 3程序(http:// frodo.wi.mit.edu/primer3/)对其进行批量引物设计。引物设计标准:

1) PCR产物大小为80~500 bp。

2) 引物长度为18~24 bp。

3) 退火温度为55~63 °C,上、下游引物的退火
 温度相差不大于5 °C。

 4) GC含量为40%~60%;尽量避免引物二级结构 Hairpin、Dimer、False Priming、Cross Dimer的出现。 对批量设计的SSR引物在3个拼接序列库中进行 E-PCR (http://www.ncbi.nlm.nih.gov/sutils/e-pcr/)。
 E-PCR条件为:允许引物1个碱基错配,1个碱基 Indel;PCR产物大小为100~500 bp。对E-PCR运行结 果进行多态性筛选,剔除没有扩增产物或在3个拼接 序列库中扩增产物均没有大小差异的SSR引物。

#### 1.4 辣椒连锁图谱的构建

将多态性SSR-EST拼接序列建立fasta文本,与 番茄基因组序列(http://solgenomics.net/)进行BLAT比 对,保留长度大于200 bp、序列连续匹配大于其长度 50%的EST序列。根据多态性EST序列,在番茄基因组上比对的位置,利用Joinmap软件构建连锁图谱。

#### 1.5 辣椒多态性 SSR-EST 序列功能分析

利用Blast2go程序(http://www.blast2go.org),对 匹配到番茄基因组上的多态性SSR-EST序列进行 功能注释(*E*<10<sup>-10</sup>)和GO分类(gene ontology, http://www.geneontology.org)及KEGG分析(kyoto encyclopedia of genes and genomes, http://www. genome.jp/kegg)。

#### 2 结果与分析

2.1 SSR 在辣椒 EST 序列中的分布

利用iAssembler拼接软件对所有辣椒转录组数 据进行unigene拼接,共获得221 037条unigenes,总 覆盖长度为97.3 Mbp,其中,HDA149、9704B分别 占总数据量的42.3%、40.4%,NCBI数据仅占17.3%。 对221 037条unigenes进行SSR位点搜索,共检测到 17 319个SSR位点,占7.83%,其发生频率为1/56 kb。 这些SSR分布在13 830条序列上,超过1个SSR位点 的共有2 343条,约占16.94%,其中复合SSRs约1 967 条,约占11.40%。17 319个SSR位点中,单核苷酸 重复占56.30%,二核苷酸重复占19.16%,三核苷酸 占23.18%。

EST-SSR重复基序以单核苷A/T、G/C为主,最 长的单核苷酸重复为146 bp;其次为三核苷酸重复基 序,其中以AAC/GTT最多(约占27.80%),AAG/CTT 次之(约占21.49%),ACG/CGT最少(约占0.62%);四、 五、六核苷酸仅占总SSR重复序列的0.13%。

#### 2.2 辣椒 EST-SSR 多样性分析及连锁图谱构建

对13 830条非冗余EST序列的SSR位点两侧进 行引物设计,共获得了10 468对特异引物,占总SSR 位点的60.44%。随后采用E-PCR对3个不同来源的 数据库进行引物多态性电子扩增,剔除来源于同一 数据库的非特异扩增引物和扩增产物大小一致的 引物,最后共产生1 538对多态性候选引物。

将1538对多态性候选引物所在EST序列与番茄 基因组进行比对,剔除连续匹配值小于50%的EST 序列,最后获得554条(36.02%)辣椒EST序列,构建 了12条辣椒EST-SSR连锁群(图1),平均每条连锁群 含45.75个EST-SSR。



2012 年 4 月

S ESSR9305 ESSR9305 ESSR8011 ESSR8011 ESSR80564 ESSR80501 ESSR8040 ESSR8040 ESSR8040 ESSR8040 ESSR8040 ESSR8040 ESSR2493 ESSR2493 ESSR2493 ESSR2493 ESSR2493	ESSR10388 ESSR15131 ESSR25131 ESSR4257 ESSR8064 ESSR8064 ESSR8064 ESSR8064	CESSR2198 ESSR3673 ESSR4970 ESSR6768 ESSR4084	ESSR9352 ESSR8117 ESSR8117 ESSR316 ESSR8316 ESSR687 ESSR4433	
690057 823978 1199947 1569871 1569871 1516401 1616401 1616445 1516445 1516445 1516445 1516445 1516445 1516445 1516465 15165865 2564536 2564536 256567 256567 257577777 2565877777777777777777777777777777777777	67763991 10403210 119964405 119964405 11996481 11996481 44067504 44067504 440575064 440575356 44875328 46358328	54321852 54322812 54322812 62417732 63188580 63287564	63323792 63512032 63512032 6395572036 6395572036 64615844 65077152 65000000 66000000	
SSR2519 ESSR2519 ESSR2519 ESSR2141 ESSR2141 ESSR2086 ESS	ESSR2381 ESSR2381 ESSR3289 ESSR3289	ESSR9455 ESSR9455 ESSR9455 ESSR9602 ESSR9602 ESSR9453 ESSR94938 ESSR4938 ESSR4938	ESSR75529 ESSR7348 ESSR5529 ESSR5529 ESSR5529 ESSR7574 ESSR7574	
263549 533701 809447 809447 809447 809447 809447 809447 809447 809447 809447 809447 809447 809447 809447 809446 876775 809447 1167281 117281 11	13722481 / 13722481 / 21504034	49371364 49371367 49738736 49738736 50281280 50281280 50435520 50435520	500001055 50995592 50995592 5115045012 51520480 51520480 54200000 542000000	
S ESSR1608 ESSR76992 ESSR7883 ESSR7897 ESSR7897 ESSR7897 ESSR7897 ESSR6897 ESSR6897 ESSR6897 ESSR6650	− ESSR5186 − ESSR3149 − ESSR3149	ESSR100 ESSR3509 ESSR1011 ESSR7481 ESSR2466 ESSR2466	ESSR743 ESSR5402 E	
461909 514338 792730 7955051 7955051 7955051 4518162 4518162 4518162 20899600	28141700 28141700 34608920 36008920	58060788 59296136 59296136 59545880 59638896 596398896 50341344 51197880	64039700 64519804 65000000 65000000	
SSR7313 ESSR7313 ESSR7313 ESSR7451 ESSR780 ESSR780 ESSR4875 ESSR4875 ESSR4875 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2357 ESSR2358	<ul> <li>ESSR2643</li> <li>ESSR2643</li> <li>ESSR8181</li> <li>ESSR818181</li> <li>ESSR5243</li> <li>ESSR3132</li> <li>ESSR3132</li> </ul>	CSSR9441 CSSR6782 CSSR6782 CSSR1379 CSSR5202 CSS CSSR5202		
760448 817229 2524550 33073376 3383555 3383555 3383555 3383555 3383555 3383555 3383555 3383555 3383555 5190825 519085 51	24424852 24424852 29841180 57502104 57705016 58770936 68175936 681022688 61022688	6120321446 62130356 621380400 622394400 622394400 6223948012 632498012 654310704 655682720 66730231296 66730231296	67000000 /	
<pre>     SSR6599     SSR6599     ESSR6599     ESSR7658     ESSR7658     ESSR7658     ESSR7659     ESSR7658     ESSR7652     ESSR78222     ESSR78222     ESSR78222     ESSR78222     ESSR78202     ESSR78002     ESSR78002     ESSR78002 </pre>	ESSR570 ESSR570 ESSR570 ESSR570 ESSR739 ESSR9328 ESSR9328 ESSR9328 ESSR9328 ESSR9348 ESSR9348 ESSR9348 ESSR9448 ESSR1736	ESSR1389 ESSR1389 ESSR1389		
51773 51773 2637795 2637795 2637795 2637795 2637795 50294188 50294188 50294188 50294188 50294188 50294188 5423496	57613650 57765972 582165972 58211496 58217712 59317712 59322096 60811424 60811176 61643088 62481328 62481328 62481328	63000000 (		
S ESSR8030 ESSR8030 ESSR8030 ESSR8989 ESSR6183 ESSR6183 ESSR5184 ESSR6183 ESSR5184 ESSR6115 ESSR6115 ESSR2115	ESSR5288 ESSR5288 ESSR5288 ESSR5288 ESSR7096 ESSR7096 ESSR79398 ESSR93398 ESSR29398	ESSR1745 ESSR5997 ESSR5997 ESSR7397 ESSR7397 ESSR73134 ESSR23134 ESSR23134 ESSR23134	ESSR3674 ESSR5211 ESSR5233 ESSR9133 ESSR9133 ESSR6545 ESSR6645 ESSR8087 ESSR8833	ESSR2100 ESSR1015 ESSR1030 ESSR6540 ESSR6527 ESSR9790 ESSR9790
309824 309824 494185 494270 1936214 1938214 22445486 22445486 22445486 22445486 22456865 5626956 5626956	7378424 7378429 9291222 53548322 55548332 55551304 55551304 5565452 5565452 5565452 5565452 55656452 55658320	58169988 59685680 60326504 60625504 615605322 62159872 62603568 62603568	62714884 63520112 63520112 63590192 64181100 64562400 64741672 64821760 64821760	64821784 64826632 64896632 64911696 64915548 65058616 65058616 65600000

第38卷第2期

Ch1 12

Ch1 11

Ch1 10

Ch1 09

图1 辣椒多态性EST-SSR在番茄基因组上的分布 Fig.1 Distribution of pepper polymorphic EST-SSRs on tomato genome

Ch1 08

Ch1 07

12条辣椒EST-SSR连锁群中,番茄第2染色体 上匹配的标记最多,为76个(约占13.71%),其次是1 号染色体,为74个(约占13.35%)。从图1可以看出, 辣椒多态性EST-SSR非均匀地分布在番茄染色体 上,且主要集中在每条染色体的两端,中间较少, 这可能是由于染色体末端更容易产生遗传交换而 形成多态性的EST-SSR。

2.3 辣椒多样性 SSR-EST 序列功能分析

对匹配到番茄基因组上的554个含有SSR motif 的EST序列进行BLAST比对(NCBI),共有547条EST 序列(98.7%, E<-10)被匹配 481条序列(约占86.82%) 被GO分类(图2),其中初生代谢、细胞代谢、生物 合成过程占多态性EST序列的主要部分,分别为 34.13%、28.35%、25.82%。另外一部分多态性EST 功能还涉及各种生物(非生物)胁迫(刺激)反应,以及 信号转导和次生代谢等。在KEGG 中,91条序列(约 占16.42%)共涉及到76条代谢途径,约67条序列(约 占73.62%)参与新陈代谢途径,32条序列(约占 35.16%)参与次生代谢产物合成途径。另外,还有19 条序列(约占20.87%)参与植物激素合成途径。





Fig.2 Gene Ontology classification of pepper polymorphic EST–SSRs and the metabolic pathways

### 3 结论与讨论

本研究中将辣椒1538对多态性SSR引物所在的 EST序列与番茄基因组进行匹配,其中554对(占 36.02%)多态性SSR-EST符合匹配标准,63.98%的 序列由于EST序列过短或低于匹配标准而被严格剔 除。根据554条SSR-EST在番茄基因组上的物理匹 配位点构建了辣椒12个EST-SSR连锁群。对这些分 布在番茄染色体上的多态性辣椒SSR-EST进行功 能分析,发现其中初生代谢(占34.13%)、细胞代谢 (占28.35%)、生物合成过程(占25.82%)占多态性EST 序列的主要部分,另外一部分EST涉及到胁迫反应 与次生代谢,表明这些含SSR基序的EST参与了辣 椒的大部分生理过程,因此,SSR-EST不仅是有用 分子标记,还可能是与目标性状关联的候选基因。

本研究中利用221 037条unigenes(97.3 Mbp)进行SSR位点搜索,共检测到17 319个SSR位点(约占7.83%),略低于前人仅基于公共数据库的辣椒EST-SSR频率(10.2%),但比小麦(3.2%),高粱(3.6%)、水稻(4.7%)等作物的EST-SSR频率高<sup>[8]</sup>,这种差异可能与SSR搜索标准、数据库大小及物种等有关。本研究结果"SSR重复主要以单核苷酸重复为主(占56.30%),三核苷酸重复次之(占23.18%)"与文献[9]中对49种双子叶植物EST-SSR的分析结果一致,即都是以单核苷酸重复占主导,二、三核苷酸重复次之。另有文献<sup>[8,10-11]</sup>报道EST-SSR重复

主要以三核苷酸重复为主。这种差异主要是由SSR 搜索参数设置不同所致。

对13 830条EST序列进行引物设计,共设计了 10 468对SSR位点特异引物,利用E-PCR扩增筛选 到多态性引物1 538对(占14.69%),开发的SSR量远 远高于目前关于辣椒EST-SSR分析的报道<sup>[5,12]</sup>,这 主要是因为本研究的EST数据量远远高于共公数据 库中的数据量。这些大量潜在的多态性SSR标记可 以应用于辣椒分子基因定位、遗传图谱构建等,也 可用作将来饱和图谱和转录图谱构建的补充。

生物基因组DNA之间存在广泛的共线性,尤其 是在近缘物种基因组间存在高度的序列同源性和 同线性<sup>[13-15]</sup>。对27 294条非冗余玉米ESTs与拟南芥 ESTs进行比较,发现68%的玉米ESTs与拟南芥编码 的ESTs同源;对1411条玉米unigenes与水稻ESTs进 行比对,发现74%的玉米基因与水稻基因同源,其 中47%为直系同源基因<sup>[16-17]</sup>;因此,物种间基因组 之间的共线性为未知基因组信息物种ESTs序列的 连锁分布提供了重要参考。

这些多态性EST-SSR在每条连锁群两端的 偏好性,基因在染色体上非绝对的共线性,以及染 色体小区内的缺失、插入、易位等分子重组,使得 辣椒EST-SSR连锁图谱在基因精细定位上产生局 限性<sup>[18-19]</sup>。辣椒EST-SSR连锁群能为基因初定位提 供参考位置区域,可以利用这一信息开发新的与目 标性状更紧密的分子标记。

#### 参考文献:

- Thiel T, Michalek W, Varshney R K, et al. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.)[J]. Theor Appl Genet, 2003, 106(3): 411–422.
- [2] Gupta P K, Rustgi S, Sharma S, et al. Transferable EST–SSR markers for the study of polymorphism and genetic diversity in bread wheat[J]. Mol Genet Genomics, 2003, 270(4): 315–323.
- [3] van Zijll de Jong E, Guthridge K M, Spangenberg G C, et al. Development and characterization of EST-derived simple sequence repeat (SSR) markers for pasture grass endophytes[J]. Genome, 2003, 46(2): 277–290.
- [4] Liang X , Chen X , Hong Y , et al .Utility of EST-derived SSR in cultivated peanut (*Arachis hypogaea* L.) and *Arachis* wild species[J]. BMC Plant Biol , 2009, 9:35.
- [5] Yi G , Lee J M , Lee S , et al . Exploitation of pepper EST–SSRs and an SSR–based linkage map[J]. Theoretical and Applied Genetics , 2006 , 114(1) : 113–130 .
- [6] Heesacker A, Kishore V K, Gao W, et al. SSRs and INDELs mined from the sunflower EST database :

Abundance , polymorphisms , and cross-taxa utility[J]. Theoretical and Applied Genetics , 2008 , 117(7) : 1021-1029 .

- [7] Djian-Caporalino C , Pijarowski L , Fazari A , et al. High-resolution genetic mapping of the pepper (*Capsicum annuum* L.) resistance loci Me3 and Me4 conferring heat-stable resistance to root-knot nematodes (*Meloidogyne* spp.) [J] .Theoretical and Applied Genetics ,2001 ,103(4) : 592–600 .
- [8] Kantety R V La Rota M Matthews D E et al Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat[J]. Plant Molecular Biology, 2002, 48(5): 501–510.
- [9] Kumpatla S P , Mukhopadhyay S . Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species[J] . Genome , 2005 , 48(6) : 985–998 .
- [10] Song Q, Marek L, Shoemaker R, et al. A new integrated genetic linkage map of the soybean[J]. Theoretical and Applied Genetics, 2004, 109(1): 122–128.
- [11] Durand J, Bodenes C, Chancerel E, et al. A fast and cost-effective approach to develop and map EST–SSR markers :Oak as a case study[J].BMC Genomics ,2010, 11: 570.
- [12] Min W K, Han J H, Kang W H, et al. Reverse random amplified microsatellite polymorphism reveals enhanced polymorphisms in the 3<sub>i</sub>ä end of simple sequence repeats in the pepper genome[J]. Mol Cells ,2008 ,26 :250–257.
- [13] Ku H M, Vision T, Liu J, et al. Comparing sequenced segments of the tomato and *Arabidopsis genomes*: Large-scale duplication followed by selective gene loss creates a network of synteny[J]. Proceedings of the National Academy of Sciences, 2000, 97(16): 9121.
- [14] Grant D ,Cregan P ,Shoemaker R C .Genome organization in dicots : Genome duplication in *Arabidopsis* and synteny between soybean and *Arabidopsis*[J] . Proceedings of the National Academy of Sciences , 2000, 97(8) : 4168.
- [15] Tang H, Bowers J E, Wang X, et al. Synteny and collinearity in plant genomes[J]. Science, 2008, 320 (5875): 486.
- Brendel V, Kurtz S, Walbot V. Comparative genomics of *Arabidopsis* and maize : Prospects and limitations[J]. Genome Biol, 2002, 3(3) : 1005–1006.
- [17] Salse J, Piegu B, Cooke R, et al. New in silico insight into the synteny between rice (*Oryza sativa* L.) and maize (*Zea mays* L.) highlights reshuffling and identifies new duplications in the rice genome[J]. The Plant Journal, 2004, 38(3): 396–409.
- [18] Varshney R K, Sigmund R, Borner A, et al. Interspecific transferability and comparative mapping of barley EST–SSR markers in wheat, rye and rice[J]. Plant Science, 2005, 168(1): 195–202.
- [19] 张西露,刘峰,戴雄泽,等.黄瓜性别控制基因 CsACSIG的分子标记及其快速检测[J].湖南农业大学 学报:自然科学版,2010,36(5):512-515.

责任编辑: 王赛群