

SSR 标记开发软件 SSR MINING 1.0 的编制

孙佳莹¹, 辛大伟¹, 单彩云¹, 刘春燕², 胡国华², 陈庆山¹

(¹东北农业大学大豆研究所, 哈尔滨 150030; ²黑龙江省农垦科研育种中心, 哈尔滨 150090)

摘要:【目的】研制利用生物信息学方法通过分析数量迅速增加的 EST 序列开发 SSR 标记的有力工具。【方法】通过对 SSR 标记本身的特点以及排列组合原理的分析, 提出快速有效的 SSR 标记开发算法, 并应用 Visual Basic 6.0 程序设计语言进行相关软件的开发。对 2004 年 8 月 2 日至 2005 年 6 月 12 日之间公布在 NCBI 上的 22 161 条大豆 EST 序列, 进行了 SSR 标记的检索。【结果】通过对现存 SSR 标记开发软件中所存在的问题的总结与分析, 编制完成 SSR 标记开发软件 SSR MINING 1.0。经过检索共发现 2 068 个 SSR 标记, 应用 DNAMAN 对其中 30 个标记进行了引物设计, 并合成进行试验验证, 最终获得了 14 个具有多态性的 SSR 标记。【结论】SSR MINING 1.0 是一个高效、灵活、界面友好、使用简便的 SSR 标记开发软件。

关键词: SSR 标记; EST 序列; Visual Basic 6.0; 遗传图谱; 大豆

Programming for SSR Marker Developing Software SSR MINING 1.0

SUN Jia-ying¹, XIN Da-wei¹, SHAN Cai-yun¹, LIU Chun-yan², HU Guo-hua², CHEN Qing-shan¹

(¹Northeast Agricultural University, Soybean Institute, Harbin 150030; ²Land Reclamation Research & Breeding Centre of Heilongjiang, Harbin 150090)

Abstract:【Objective】As the number of EST sequences increasing, which supplies a valuable resource for the development of SSR markers, and bioinformatics has become a powerful tool for mining SSR. 【Method】By analysis of the characteristic of SSR marker itself and permutation and combination theory of SSR as well, an efficient arithmetic was put forward. The relative software also was programmed by Visual Basic 6.0. 22 161 soybean EST sequences published between August of 2004 and June of 2005, as the resource to search SSR sites. 【Result】After summarizing and analyzing the problems of existed software for SSR markers finding, a new software named SSR MINING 1.0 was accomplished. After searching for SSR markers, 2 068 SSR markers were got, and with the help of DNAMAN, 30 pairs of primer were designed from the found SSR markers. Then synthesized them and did the experiment, 14 polymorphic SSR markers were found. 【Conclusion】SSR MINING 1.0 is an efficient, flexible, interface friendly, and convenient SSR markers developing software.

Key words: SSR marker; EST sequence; Visual Basic 6.0; Genetic map; Soybean

0 引言

【研究意义】简单重复序列 (simple sequence repeat, SSR), 也称微卫星 DNA (microsatellite DNA), 是指 DNA 分子中 1~5 个核苷酸的串联重复^[1,2]。SSR 以其在动植物基因组随机分布、高信息量和多态性、共显性和孟德尔遗传等优点^[3], 在遗传图谱的构建、遗传多样性分析、亲缘关系鉴定、DNA 指纹图谱构建、

功能基因标记等方面具有公认的优越性和应用前景^[4]。由于 SSR 位点的两端是相对保守的单拷贝序列, 因此可以根据其两端序列设计引物, 然后对 PCR 结果进行分析^[5,6]。然而传统的 SSR 标记引物设计过程较慢, 需要进行基因组文库的构建、重复序列克隆的识别和筛选、测序、引物设计等环节。SSR 引物的获得, 花费的时间长, 费用较大, 效率低^[7]。【前人研究进展】EST (expression sequence tag) 和 cDNA 作为表

收稿日期: 2007-04-17; 接受日期: 2007-08-27

基金项目: 国家科技攻关项目 (东北区, 2004BA907A26), 黑龙江省教育厅科技研究项目 (10551029), 国家“863”项目 (2006AA100104), 黑龙江省博士后科学研究基金 (LHK-04014)

作者简介: 孙佳莹 (1983-), 女, 黑龙江哈尔滨人, 硕士研究生, 研究方向为大豆生物技术。E-mail: sunjiaying_1983@163.com; Tel: 0451-5519194。通讯作者陈庆山 (1973-), 男, 黑龙江林甸人, 教授, 博士, 研究方向为大豆遗传育种与生物技术。E-mail: qshchen@sohu.com。胡国华 (1951-), 男, 上海人, 研究员, 博士, 研究方向为大豆遗传育种。E-mail: Hugh757@163.com

达基因部分序列，随着功能基因组学的发展而高速增加。GenBank、DDBJ 和 EMBL 等公共数据库提供的大量 EST 数据为 SSR 标记的开发提供了丰富的资源。通过对大量 EST 数据进行 SSR 分析，可以估测 SSR 分子标记出现的频率和分布特征，并据此提出相应的核心算法，以及相关的软件开发。已有一些学者进行了高等植物如拟南芥、玉米、大豆、水稻、小麦的 SSR 研究^[8-10]，开发了一些发掘 SSR 标记的程序，如 GRAMENE 网站提供的用 perl 脚本编写的 SSRIT. PI (<http://www.gramene.org/gramene/searchs/ssrtool>)^[11,12] 和华盛顿大学 Abajian 博士用 C 语言开发的 Sputnik 软件 (<http://abajian.net/sputnik/index.html>) 等^[8]。【本研究切入点】这些程序都立足将找到的标记列表给出重复类型和位置，但不能满足对 SSR 分子标记进行设计引物等更详细研究的需要，也不能用于对大量 EST 数据进行 SSR 分子标记统计分析研究^[13]。为了满足从 EST 数据中大规模进行 SSR 标记开发的需求，后续又有很多学者进一步开发出了其它可在本地计算机上使用的软件，如用 java 语言开发的 SSRFinder^[14] 以及应用 perl 脚本写成的软件 MISA (<http://pgrc.ipk-gatersleben.de/misa/>)^[15]，但由于其均为基于命令行的软件，缺少友好的界面，而限制了其使用。【拟解决的关键问题】本试验应用 Visual Basic6.0 程序设计语言尝试开发基于本地平台的、界面友好的、使用简单的 SSR 标记开发软件。

1 算法的提出与实现

1.1 SSR MINING 1.0 软件的核心算法

根据排列组合原理，结合 SSR 序列的自身特征，建立 SSR 标记的核心单元库，如不考虑重复的核心单元，则由 i 个碱基组成的核心单元应有 4^i 个，但由于其中存在有重复的核心单元序列，实际需要查找比对的核心单元序列远小于 4^i 。因此在建立核心单元库时，去除以下几种冗余的重复：

移码重复：当从大量的序列中发掘重复单元时，发现由于碱基数量很大，一些移码单元可看成同一条序列。例如：(CT)₅=CTCTCTCTCT=TCTCTCTCTC=(TC)₅，移位的一个碱基由于其相对于总碱基数来说可被忽略不计。

约数重复：对于像 4,6 这样的有约数的核心单元，要去除其约数重复。例如：4 碱基重复 (ATAT)_n=(AT)_{2n} 等同于 2 碱基重复。

单碱基重复：由于真核生物基因特有的 PolyA 和

PolyT 结构的存在，在建立核心单元库的时候应当将这一情况排除。因为这一结构上的单碱基重复不能算作 SSR 重复序列。

由此可推知，对于 i 个碱基组成的核心单元，其核心单元总数为^[13]：

$$\alpha_i = \frac{4^i - \sum_{l \in \alpha} \left(4^i - \sum_{k \in \beta} 4^k \right)}{i}$$

式中， i 为组成核心单元的碱基个数； α 为 i 除去本身外的所有约数； β 为 l 除去本身外的所有约数。

核心单元库的建立：通过总结各种碱基数目的核心单元组合的特点，其中单碱基重复单元、两碱基重复单元、三碱基重复单元、四碱基重复单元、五碱基重复单元以及六碱基重复单元的数目依次为：4、6、20、60、204 和 670，总计 964^[13]。

1.1.1 基于重复序列核心单元库的算法 依次建立各个移码、约数等重复单元库，再从总的重复单元库中去除，耗费了大量的资源，因此本软件在开发的过程中采取了由上一级最优库递推建立下一级最优库存的手法，对此过程进行了简化。具体方法如下：

首先，通过在 N-1 级最优库中的重复单元后依次加上 A、T、G、C 4 种碱基来建立 N 级初级库；然后，在 N-1 级最优库中的重复单元中插入与开头碱基不重复的碱基来建立 N 级插入库；再将这两个库合并，并去除其中的移码、约数重复以得到 N 级最优库。

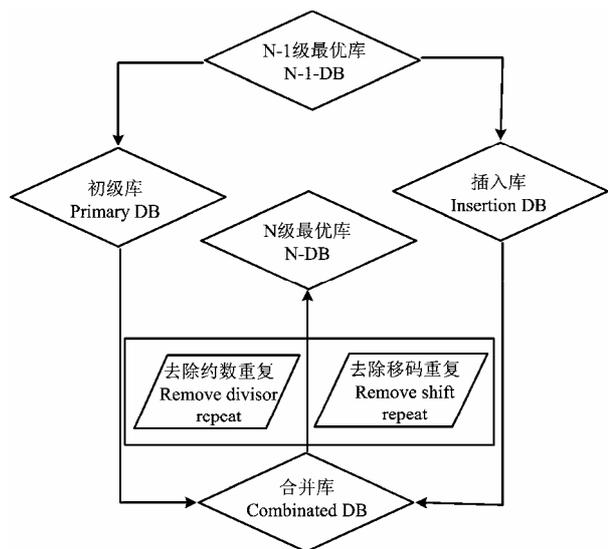


图 1 建立 SSR 最优重复单元库的流程
Fig. 1 Flow chart of the process building best SSR unit database

建立完成 SSR 重复单元库之后,从输入的序列中读入一条序列,检索其中是否包含重复单元库中的重复单元,若包含则输出超过设定的阈值的 SSR 重复单元、重复次数、起始位置、终止位置。

1.1.2 基于待测序列的算法 基于待测序列的算法可以通过对输入的待测序列的扫描,动态地获得 SSR 重复单元,省去建立重复单元库的时间和空间。在待测序列中,若遇到重复出现某特定次数(可人为设定)的序列后,将其作为待检测的核心单元进行检索。若在待测序列中发现大于某长度或重复次数多于某特定次数(可人为设定)的区域,即可取其两侧保守序列,设计特异引物,得到新的 SSR 标记位点。其具体算法描述如下:

定义字符串 S 的长度为 $|S|$, $S(i, j)$ 为 S 中从第 i 个碱基到第 j 个碱基的子串, m 为重复次数, k 代表 $S(i, j)$ 子串的长度。若满足: $S(i, j) = S(i+m*k, j+m*k)$, 其中 $[i, j \in [1, |S|)]$, 且 $i \leq j$, $m \in (1, |S|/k)$, $k \in (1, |S|/2)$, m, k 均为整数, 则 m 的累进值必须为 1, 如果 $S(i, j) \neq S(i+(m+1)*k, j+(m+1)*k)$, 即使 $S(i, j) = S(i+(m+2)*k, j+(m+2)*k)$, 则只计为 m 值, 而不计为 $m+2$ 。

如果上式中, $m \geq$ 阈值, 则输出重复单元、重复次数、起始位置、终止位置。

注意: 在 m 的累进值必须为 1, 如果 $S(i, j) \neq S(i+(m+1)*k, j+(m+1)*k)$, 即使 $S(i, j) = S(i+(m+2)*k, j+(m+2)*k)$, 则只计为 m 值, 而不计为 $m+2$ 。

1.2 SSR MINING 1.0 软件的编制和运行环境

SSR MINING 1.0 软件是应用 Microsoft 公司的 Visual Basic 6.0 中文版开发的。有关 Visual Basic 语言的介绍请参阅有关书籍。SSR MINING 1.0 软件的运行对于计算机硬件环境要求不高, 奔腾以上机型, Windows 9X/me/2000/XP 等系统均能正常运行, 但要达到理想的运算效率, 计算机的硬件配置不应过低。

1.3 SSR MINING 1.0 软件的使用

1.3.1 基于重复单元库的检索 第一步, 建立重复单元库。输入要建立的重复单元库的碱基数, 可以是固定数值, 或者是一个范围。完成后运行程序, 生成的重复单元库便以文本文件 (.txt) 的形式输出。

第二步, 对待测序列进行检索。在此仍要输入需检索的重复单元的碱基数, 同样既可以是定值或者是数值范围。除此之外, 还要给出最小的重复次数, 默认值为 5。待测序列即可直接粘贴到文本框中, 也可

以通过文件打开。注意这个检索模式仅支持 flatfile 格式(无格式文本)的序列, 并不支持其它格式的序列格式。运行程序, 所得结果即以文本文件的形式输出。

1.3.2 基于待测序列的检索 本部分支持两种序列格式: 一种是 flatfile 格式; 另一种是 fasta 格式。待测序列即可以直接粘贴到文本框中, 也可以通过文件打开。然后应给出相关的两个参数, 组成的重复单元碱基数, 既可以是固定值, 也可以是一个范围, 缺省值为 2~10; 以及最小重复单元出现次数, 缺省为 5。之后即可运行程序, 进行检索。本软件给出了 3 种不同的输出方式, 根据用户的不同需要输出检索结果:

①完全输出。此方式不排除约数重复, 可以更加全面的了解已知序列中 SSR 序列分布的情况; ②部分输出。按照用户的要求有针对性地输出检索结果, 部分排除某些序列; ③最简输出。完全排除约数重复等冗余, 输出最简结果, 由此可以获得已知序列的最简信息。

本部分亦可对序列进行各种转换, 如大小写转换、互补以及反向互补序列的转换等等。

1.4 实验验证

从 NCBI 上下载了 2004 年 8 月 2 日至 2005 年 6 月 12 日期间共 22 161 条大豆 EST 序列, 以 fasta 格式保存, 用于 SSR 标记位点的检索。共发现了 2 070 个 SSR 标记位点, 从中选取了 30 个应用 DNAMAN 软件对检索出的 SSR 标记位点进行引物设计, 并送交上海生工进行合成。

选取具有代表性的 8 个远缘品种^[16]: Clark、Harosoy、Williams、Amsoy、Chalston、东农 594、合丰 25 对设计获得的 SSR 标记位点的多态性等进行检验。采用 soybase 网站上公布的 PCR 循环 ($T_m=47^\circ\text{C}$) 以及体系^[17], 应用 6% 的聚丙烯酰胺凝胶进行电泳检测。

2 结果与分析

2.1 不同软件的检索结果

用 SSR MINING 1.0、SSRIT 和 Sputnik 对大豆的 22 161 条 EST 序列分别进行了处理, 所得结果如表 1 所示, SSR MINING 发现 SSR 位点总数略高于其余两个软件, 然而 Sputnik 由于其具有容错性, 多找出了几个含插入、缺失和错配 SSR 标记, 但其所发现的 SSR 位点总数在此 3 种软件中则为最少, 其它结果这 3 个程序的几乎相同。

经过具体操作发现, Sputnik 由于其本身软件的设计, 只能检索到 2~5 碱基的重复单元的 SSR 位点,

因此其得到的五、六碱基的 SSR 位点最少, 然而其对三、四碱基重复单元的 SSR 位点发现能力最强, 所得数目最多; 而其它类型的 SSR 位点则以 SSR MINING 所发现的为多, 具体情况如图 2 所示。

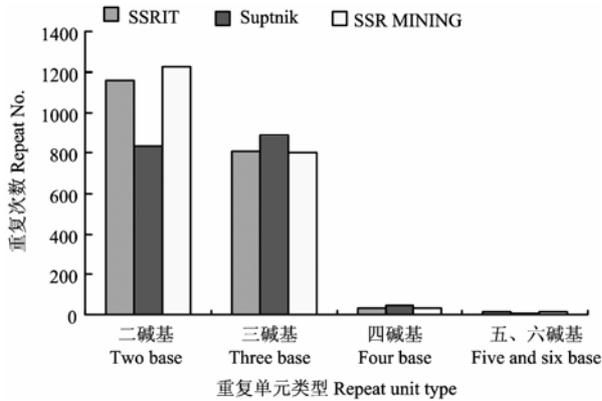


图 2 不同软件检索结果 SSR 位点的分布
Fig. 2 Distribution of the SSR sites searched by different software

2.2 实验结果分析

在 2004.08.02 至 2005.06.12 期间, NCBI 上共公布了 22 161 条大豆 EST 序列。本研究应用软件 SSR MINING 1.0 从中发现了 2 068 个 SSR 标记位点。其中二核苷酸重复单元的 EST-SSR 最多, 共有 1 227 个, 占总数的 59.28% , 其余类型的出现次数及频率

见表 1。

表 1 SSR 中不同重复单元出现的频率
Table 1 Frequency of different type SSR units

重复单元类型 Repeat unit type	数量 No.	频率 Frequency(%)
二核苷酸 Di-unit	1227	59.28
三核苷酸 Tri-unit	801	38.70
四核苷酸 Tetra-unit	30	1.45
五、六核苷酸 Penta, hexa-unit	10	0.48

在 1 227 条二核苷酸重复单元的 ESR-SSRs 中, AG/TC 重复单元的数量最多, 有 535 条, 占总数的 25.85%, 而仅有一条含 GC/CG 的重复单元。本研究中共检索到 801 条三核苷酸重复单元的 EST-SSRs, 其中 AAG/TTC 重复单元出现频率最高, 为 106 条, 占 20.01%; 其次为 AAT/TTA 重复单元的 EST-SSRs, 为 132 条, 占总数的 16.48%。具体分布见图 3。

从中选取 30 个 SSR 标记位点, 应用 DNAMAN 设计引物, 并送交上海生工合成, 经聚丙烯酰胺凝胶电泳检测发现 14 个位点具有多态性, 占总数的 46.67%, 初步证明所检测到的 SSR 位点, 可以在大豆上作为 SSR 标记使用; 11 个位点有扩增, 但无多态, 占总数的 36.67%; 5 个位点无扩增, 占总数的 16.67%。由此可见本软件已达到应用的基本要求, 可以作为 SSR 标记的开发软件在实际中应用 (图 4)。

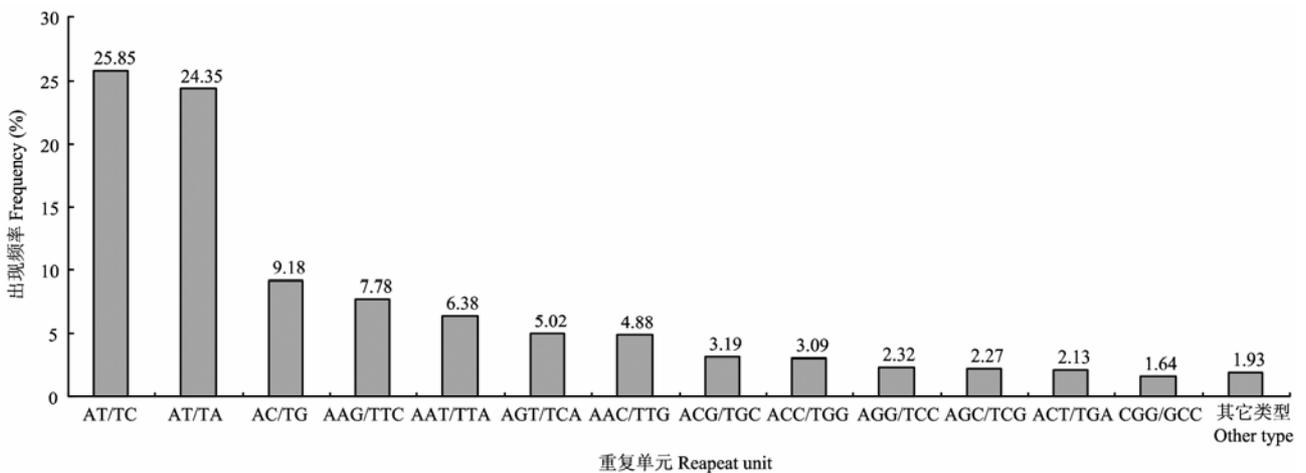
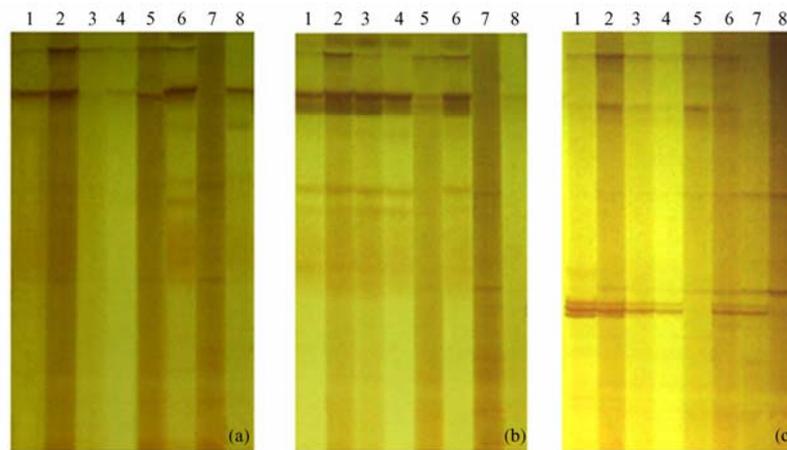


图 3 基于重复单元类型的 EST-SSRs 分布
Fig. 3 Distribution of different type unit of EST-SSRs



(a) Si23; (b) Si24; (c) Si25

1: Clark; 2: Harsoy; 3: Williams; 4: Amsoy; 5: Noir1; 6: 合丰 25; 7: 东农 594; 8: Charlition
1: Clark; 2: Harsoy; 3: Williams; 4: Amsoy; 5: Noir1; 6: Hefeng 25; 7: Dongnong 594; 8: Charlition

图 4 部分引物对大豆基因组 DNA 中的 SSR 标记的扩增产物

Fig. 4 Partial amplification of SSR production in soybean genome

3 讨论

目前普遍认为 SSR 标记开发软件应具备 4 大特性^[18]: ①较高的计算效率。随着生物信息学的发展, 用于 SSR 标记开发的序列越来越大, 因此用于序列分析的工具的效率就显得尤为重要; ②灵活性和显著性。在处理大序列的时候, 需要考虑一系列相关的变异如: 插入、缺失和错配。因此软件的灵活性也就变得至关重要, 而与此同时, 为了评估结果的质量, 显著性的定义也成为了关键; ③交互显示系统。待测序列的输入和检测结果的输出皆要通过交互显示系统, 因此基因研究者不仅需要获得基于整个基因组以及染色体的可视结果, 而且对于特定区段的详尽的可视化数据的作用也不可忽视; ④可扩展性。在长期的研究过程中, 重复序列分析仅仅是序列分析的一个基本的步骤。基于重复序列的分析可进行更深入的研究, 因此重复序列检索程序应该能够提供一种简易的界面来与其它高级的分析软件进行接口。

国内外在与大豆相关的 SSR 分子标记开发方面开展了一些研究, 但是这些研究并不很完善, 大多都集中在实验室的开发方法上。实验室开发方法消耗大量的人力、物力, 并非一般科研单位所承担得起的, 因此限制了 SSR 标记的应用。然而生物信息学开发方法的核心即为提出合理有效的核心算法, 开发出相关软件。但现在已提出的 SSR 标记开发的核心算法无论在

功能或效率等方面上均有很多不完善的地方, 有待于进一步提高。而且并未完全具备以上提出的 4 点特性, 有些已开发的软件虽然具有较好的计算效率, 但由于其基于命令行的开发平台, 使其使用变得尤为困难, 且缺乏友好的界面, 因此限制了其使用。

通过对已开发成功并为研究人员们普遍使用的软件之间的对比来验证本试验所提出算法。这里主要采用较为常用的软件如 SSRIT 和 Sputnik, 从而获得图 2。

由此可见, 新算法在结果输出以及检索方式上均较 SSRIT 和 Sputnik 这两个网络软件灵活多样 (如表 2 所示)。而且处于本地计算机上, 不受网络的限制, 可以灵活便捷地进行使用, 节省了大量的时间和网络资源, 而且在相同的参数设定下, 本软件较之其它软件检索出了更多的 SSR 位点, 而且输出结果灵活多样, 便于下一步研究的分析处理, 使得 SSR 标记的开发更加高效。

虽然本试验所提出的新算法虽不具有容错性, 且尚无法检测出复合 SSR 标记位点, 并且没有实现 SSR 标记开发的全自动化过程; 但依据其所编制出来的软件 SSR MINING 1.0 可与其它引物设计软件如 DNAMAN、Primer3.0 等组合, 而且使用灵活简便, 界面友好。

4 结论

基于 SSR 标记本身的特点以及排列组合原理而开

表 2 不同 SSR 标记开发软件的比较

Table 2 Comparison of different SSR marker developing soft wares

	SSR MINING 1.0	SSRIT	Sputnik
容量 Content	[1, 2000]	<800	无限 Infinite
重复单元长度 Length of repeat unit	不限 Infinite	1~10	2~5
形式 Install form	本地计算机 Local	网络 Internet	网络、本地计算机 Internet, local
速度的稳定性 Stability of speed	稳定 Stable	受网络限制 Up to internet	受网络限制 Up to internet
容错性 Flexibility	无 No	有 Yes	有 Yes
文件输入格式 Format of input file	Flatfile: FASTA	FASTA	FASTA
检索的形式 Searching form	最简无冗余检索 None-redundancy search 按人为指定的重复检索 Specified search 全重复检索 All-repeat search	最简无冗余检索 None-redundancy search	最简无冗余检索 None-redundancy search
结果输出形式 Form of output	文本文档输出.txt file 文本框直接输出 Text	网页输出 Internet 表格 Table	网页输出 Internet 命令行输出 Command 电子邮件输出 E-mail

发出的 SSR 标记开发软件 SSR MIN-ING 1.0, 在 EST-SSRs 的开发中具有较高的效率, 与相关的引物设计软件结合, 对于已经公布 EST 数据的生物来说, 是一个比较理想的 SSR 开发软件。

References

- [1] Tautz D, Renz M. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. *Nucleic Acids Research*, 1984, 12: 4127-4138.
- [2] Tautz D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research*, 1989, 7: 6463-6471.
- [3] Powell W, Machray G C, Provan J. Polymorphism revealed by simple sequence repeats. *Trends Plant Science*, 1996, 7: 215-222.
- [4] 陈海梅, 李林志, 卫宪云, 李斯深, 雷天东, 胡海洲, 张宪省. 小麦 EST-SSR 标记的开发、染色体定位和遗传作图. *科学通报*, 2005, 50(20): 2208-2216.
Chen H M, Li L Z, Wei X Y, Li S S, Hu H Z, Zhang X X. Development chromosome location and genetic mapping of EST-SSRs in wheat. *Chinese Science Bulletin*, 2005, 50(20): 2208-2216. (in Chinese)
- [5] Wang Z, Weber J L, Zhong G, Tanksley S D. Survey of plant short tandem DNA repeats. *Theoretical and Applied Genetics*, 1994, 8:1-6.
- [6] Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A. The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 1996, 2: 225-238.
- [7] 谢皓, 陈学珍, 杨柳, 王建立. EST-SSR 标记的发展和在植物遗传研究中的应用. *北京农学院学报*, 2005, 20(4): 73-76.
Xie H, Chen X Z, Yang L, Wang J L. Development and application of EST-SSR marker in plant genetic. *Journal of Beijing Agricultural College*, 2005, 20(4), 73-76. (in Chinese)
- [8] Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R. Computational and experimental characterization of physically clustered simple sequence repeats in plants. *Genetics*, 2000, 156: 847-854.
- [9] Tóth G, Gáspári Z, Jurka J. Microsatellites in different eukaryotic genomes: Survey and analysis. *Genome Research*, 2000, 10: 967-981.
- [10] Morgante M, Hanafey M, Powell W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics*, 2002, 30: 194-200.
- [11] Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S. Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon association, and genetic marker potential. *Genome Research*, 2001, 11: 1441-1452.
- [12] Kantety R V, La Rota M, Matthews D E, Sorrells M E. Data mining for simple sequence repeats in expressed sequence tags from barley,

- maize, rice, sorghum and wheat. *Plant Molecular Biology*, 2002, 48: 501-510.
- [13] 汤继凤, 曹永生, 高丽锋, 贾继增. 用生物信息技术构建 cSSR 分子标记开发体系. *中国农业科学*, 2004, 37(3): 328-332.
Tang J F, Cao Y S, Gao L F, Jia J Z. Mining cSSRs from EST data using bioinformatics approaches. *Scientia Agricultura Sinica*, 2004, 37(3): 328-332. (in Chinese)
- [14] Gao L F, Tang J F, Li H W, Jia J Z. Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding*, 2003, 12: 245-261.
- [15] Thiel T, Michalek W, Varshney R K, Thiel A G. Exploiting EST databases for the development of cDNA derived microsatellite markers in barley (*Hordeum vulgare* L.). *Theoretical and Applied Genetics*, 2003, 106: 411-422.
- [16] Song Q J, Marek L F, Shoemaker R C, Lark K G, Concibido V G, Delannay X, Specht J E, Cregan P B. A new integrated genetic linkage map of soybean. *Theoretical and Applied Genetics*, 2004, 109: 122-128.
- [17] Soybase: <http://soybase.org/resources/ssr.php>.
- [18] Kurtz S, Benson G, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. Computation and visualization of degenerate repeats in complete genomes. *Molecular Biology*, 2000(8): 228-238.

(责任编辑 毕京翠)