

技术主题

Technology Feature

玉米 DNA 指纹鉴定中两种 SSR 数据分析软件 GeneMapper 和 GeneMarker 的应用比较

王蕊^{1,2*}, 田红丽^{2*}, 王凤格², 赵久然², 易红梅², 王璐², 李楠³, 霍永学³

1. 首都师范大学, 北京, 100089
2. 北京市农林科学院玉米研究中心, 北京, 100097
3. 北京华生恒业科技有限公司, 北京, 100083

*第一共同作者

✉ 通讯作者: maizezhao@126.com ✉ 作者

分子植物育种, 2012 年, 第 10 卷, 第 24 篇 doi: 10.5376/mpb.cn.2012.10.0024

收稿日期: 2012 年 04 月 24 日

接受日期: 2012 年 05 月 16 日

发表日期: 2012 年 06 月 05 日

这是一篇采用 Creative Commons Attribution License 进行授权的开放取阅论文。只要对本原作有恰当的引用, 版权所有人允许并同意第三方无条件的使用与传播。

引用格式(中文):

王蕊等, 2012, 玉米 DNA 指纹鉴定中两种 SSR 数据分析软件 GeneMapper 和 GeneMarker 的应用比较, 分子植物育种(online) Vol.10 No.24 pp.1171-1178 (doi: 10.5376/mpb.cn.2012.10.0024)

引用格式(英文):

Wang et al., 2012, The Application Comparison of Two SSR Data Analysis Softwares GeneMapper and GeneMarker Used in the Identification of Maize Variety, Fenzi Zhiwu Yuzhong (online) (Molecular Plant Breeding) Vol.10 No.24 pp.1171-1178 (doi: 10.5376/mpb.cn.2012.10.0024)

摘要 GeneMapper 与 GeneMarker 均是 SSR 数据分析的主要软件。本文使用 40 对引物对 11 套亲子组合分析可发现, GeneMapper 与 GeneMarker 分析结果吻合度达 99%, 说明两种软件可靠性均很高。其中 GeneMarker 在二倍体筛选、去除 Pull-up 峰、平移 Panel 等功能上进一步优化, 减少了人工分析的时间, 同时降低了人工操作出现的误差。GeneMarker 软件的升级为构建玉米 DNA 指纹数据库的标准化与自动化提供重要辅助作用。

关键词 玉米; SSR; GeneMarker; DNA 指纹鉴定

The Application Comparison of Two SSR Data Analysis Softwares GeneMapper and GeneMarker Used in the Identification of Maize Variety

Wang Rui^{1,2*}, Tian Hongli^{2*}, Wang Fengge², Zhao Jiuran², Yi Hongmei², Wang Lu², Li Nan³, Huo Yongxue³

1. Capital Normal University, Beijing, 100089
2. Maize Research Center, Beijing Academy Agriculture and Forestry Sciences, Beijing, 100097
3. Beijing Todaysoft Limited, Beijing, 100083

*These authors contributed equally

✉ Corresponding author, maizezhao@126.com; ✉ Authors

Abstract GeneMapper and GeneMarker are main softwares in the analysis of SSR data. 40 pairs of primers were used to analysis 11 groups of hybrids and their parents, and the parallel degree reached 99% between GeneMapper and GeneMarker, therefore this could conclude that the reliability of two softwares were reliability. Some functions were updated in GeneMarker, such as amphiploid filter, removing Pull-up peaks, shifting Panels. These changes reduced the manual work, as well as decreased the manual errors. The updating of GeneMarker made an important effort for the standardization and automatic construction of maize DNA fingerprint database.

Keywords Maize; SSR; GeneMarker; DNA fingerprint identification

研究背景

SSR 标记作为近年来发展较快的分子标记技术, 具有多态性高、重复性高、共显性遗传、操作简单快速等优点, 并在品种 DNA 指纹鉴定中被广泛应用。基于毛细管电泳的 SSR 检测平台被各研究单位、检测机构所青睐。北京市农林科学院玉米研究中心从 2002 年起开展基于 SSR 标记的玉米品种

DNA 指纹技术研究, 目前已经建立简单快速、准确高效的毛细管荧光电泳检测技术。

随着毛细管电泳检测技术日益成熟, 检测样品量急剧增加的情况下, 对数据分析也提出了更高的要求, 急需操作简单、自动化程度高、准确性高的数据分析软件来支持 DNA 指纹鉴定技术。目前用于分析 SSR 扩增片段的主流分析软件为 GeneMapper

和 GeneMarker, 这两个软件算法不同、具有各自的优缺点。

本文使用我国推广面积较大的 11 个杂交种及其父母本作为研究材料,对 GeneMapper 和 GeneMarker 两个软件从数据分析准确性、操作简便性、自动化分析程度等全方面进行比较分析,以期从事 DNA 指纹鉴定研究的人员提供软件选择的依据。

1 结果与分析

根据本实验材料 40 对 SSR 引物的数据分析结果,从以下 6 个方面对 GeneMapper 和 GeneMarker 两个软件进行应用比较,分析比较两软件在解决方案上的差异。

1.1 去除 Pull-up 峰

在毛细管电泳中,引物标记的颜色有 4 种,当多重组合电泳时,某一色荧光引物某一位置的特异峰较高引起同一位置的其它颜色荧光峰值升高,这种峰称之为 Pull-up 峰。Pull-up 峰是影响数据分析的重要因素,进行数据分析时易将峰高较高的 Pull-up 峰判为特异峰,同时也影响了分析效率。针对于此 GeneMarker 软件设计了便利的矩阵算法,可批量完成去除 Pull-up 峰的功能。如图 1 所示,使用 GeneMarker 分析,可根据每次电泳结果,根据图像设定不同荧光峰高比值,可直接将大部分 Pull-up 峰去除,设置的比值越精确, Pull-up 峰消除效果越好。如图 2 所示,在 GeneMapper 软件分析中,黑色对红色、绿色对黑色等均有 Pull-up 峰,其中最高的 Pull-up 峰高接近 1 000,易与特异峰混淆。而 GeneMarker 软件中去除得较为彻底。

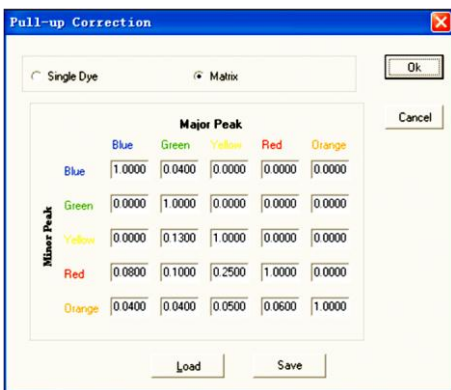


图 1 GeneMarker-BMSTC v1.0 去除 Pull-up 峰设置界面
Figure 1 The setup interface of removing Pull-up peaks of GeneMarker-BMSTC v1.0

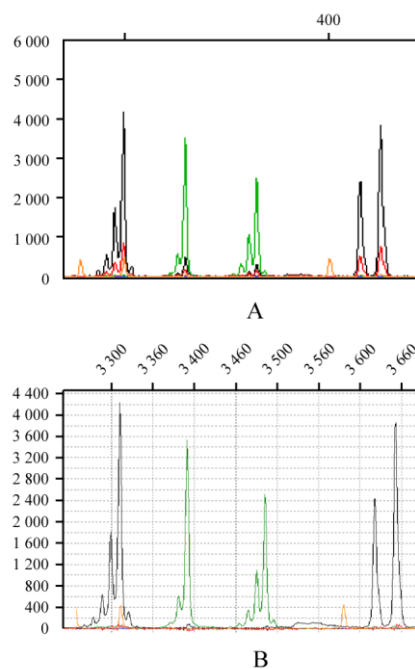


图 2 毛细管电泳中 Pull-up 峰去除前后图
注: A: 去除 Pull-up 峰前(GeneMapper v3.7); B: 去除 Pull-up 峰后(GeneMarker-BMSTC v1.0)

Figure 2 Before and after removing Pull-up peaks in the capillary electrophoresis

Note: A: Before removing Pull-up peaks (viewed by GeneMapper v3.7); B: After removing Pull-up peaks (viewed by GeneMarker-BMSTC v1.0)

1.2 Panel 平移功能

不同批次的电泳可能会出现峰的微小滑移现象(差异在 0.5 bp 以内)。数据库为了保证不同批次数据能够有效整合,需要使用固定 Panel 进行数据分析,当数据出现滑移时,可能出现大量数据不能进入 BIN 的范围,只有重新设定 Panel 或加宽 Panel 来解决。GeneMarker 软件具备 Panel 整体平移的功能,如图 3A 所示,当峰大多不在 BIN 内,可将此引物的 Panel 作为一个整体平移,使峰最大化进入 BIN 设置范围内,调整后如图 3B 所示。

1.3 连续多峰识别功能

连续多峰一般在扩增片段大于 300 bp 时易出现,主要是由于扩增片段长度过长, Taq 酶在模板链上滑动引起的,多是由一系列相差 2 bp 的连续峰峰组成,峰高普遍偏低。连续峰是荧光检测大片段位点时的常见峰形,而且容易引起软件识别的问题,容易将峰高较低连续峰去除或将连续峰计算为多峰,这两种分析结果都要尽力避免,因此 GeneMarker-BMSTC v1.0

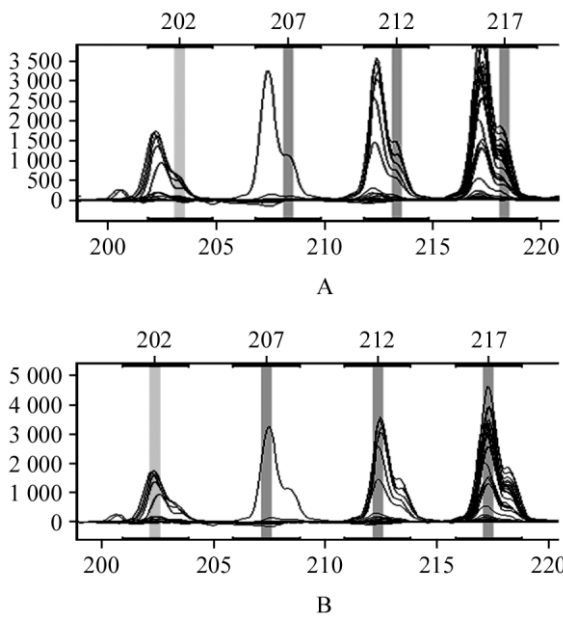


图 3 GeneMarker-BMSTC v1.0 软件 Panel 平移界面(引物: phi080k15)

注: A: Panel 平移前; B: Panel 平移后

Figure 3 Panel moving interface of GeneMarker-BMSTC v1.0 (Primer: phi080k15)

Note: A: Before moving Panel; B: After moving Panel

软件改进了荧光电泳中连续多峰的识别方法。如图 4 所示, 杂交种及父本在 403 位点上为连续峰, 使用 GeneMapper 软件分析时将父本连续峰位点计算成两个峰, 使用 GeneMarker 软件分析时, 因为能自动识别连续峰, 因此亲本和杂交种在连续峰位点均只显示同一等位基因位点, 不需额外删减位点, 提高了数据分析的准确度与效率。

1.4 高低峰选择功能

在毛细管电泳图中, 有时会出现两等位基因不对称扩增的情况, 导致出现高低峰。高低峰出现的主要原因包括以下两点: 一是 PCR 扩增时出现了不对称扩增, 这在玉米杂交种出现亲本互补带型时出现的可能性较大, 当杂交种某个位点上的条带也为单带或者是自交系时, 出现这种情况的可能性较小; 二是采用混株法提取 DNA 时, 样品一致性出现问题, 导致最终混入杂带, 这种情况的低峰峰高通常较低, 可通过人工看板去除掉。为了保证软件分析时特异峰能够被软件保留, 而非特异峰以及杂带可被软件去掉, 在 GeneMarker 软件中优化了高低峰选择功能, 可通过不同参数设置, 自动识别峰高偏低的特异峰。如图 5 所示, Local Region 中可

设定峰高比值, 在面对不同农作物中以及不同检测对象时, 应采取不同的比例。如图 6 所示, 虽然峰高有所差异, 但峰高较低的连续峰在 GeneMapper 与 GeneMarker 中均可入选。

1.5 N+1 峰自动过滤功能

N+1 峰在毛细管电泳中比较常见, 主要特征是在同一位置上出现了相差 1 bp 的两个峰(易红梅等, 2006)。一般认为与 PCR 反应过程中普通 Taq 酶在产物 3'端自动加入一个碱基有关。使用 GeneMapper 分析时, 为了避免 N+1 峰作为两个峰, 需要不断加宽 BIN 设置中此等位基因的范围, 致使相邻等位基因的范围可能出现交叉, 降低了 SSR 的多态性。而 GeneMarker 软件可自动识别 N+1 峰, 若两峰在同一个等位基因范围内, 则计算为同一位点; 若不在同一个等位基因范围内, 则滤掉峰分值(score 值)较小的峰, 将峰分值较大的峰作为主峰。如图 7 所示, 勾选“Use Low-High Filter”可使用此过滤算法。如图 8 所示, 两软件均可自动识别 N+1 峰。

1.6 二倍体筛选功能

玉米为二倍体植物, SSR 标记是共显性标记, 如果采用单株提取 DNA, 则在同一引物只能出现两个特异峰(杂合位点)或一个特异峰(纯合位点)。但在混株提取 DNA 进行指纹分析时, 由于各种原因有时会出现三峰或多峰的情况(赵久然等, 2003)。使用 GeneMapper 分析时需要人工手动删除杂带, 极大影响分析速度, 无法实现自动化, 而 GeneMarker-BMSTC v1.0 版增加了二倍体筛选(即三峰, 多峰过滤)功能。应用该算法后, 同一引物范围内最多保留两个峰。保留原则为在同一引物范围内, 有多峰时仅保留 Score 值最好的 2 个, 其余都删除; 当 Score 值相同时, 取强度较大的峰。具体方法为: 运行 GeneMarker, 导入原始数据后, 点击运行, 出现如图 9 所示的对话框, 勾选“Use Amphiploid Filter”即可。如图 10 所示, PCR 扩增时出现少量杂带, GeneMapper 需要人工删除杂带, 而 GeneMarker 可以自动识别并删除多余杂带, 适合大规模的自动化样品分析。

2 讨论

经过多年的研究、应用, SSR 指纹鉴定技术在玉米品种检测中的应用已趋向成熟, 尤其是核心引物、实验流程、判断标准等方面均已在各检测机构被广泛应用, 并且各步工作基本达到准确、适

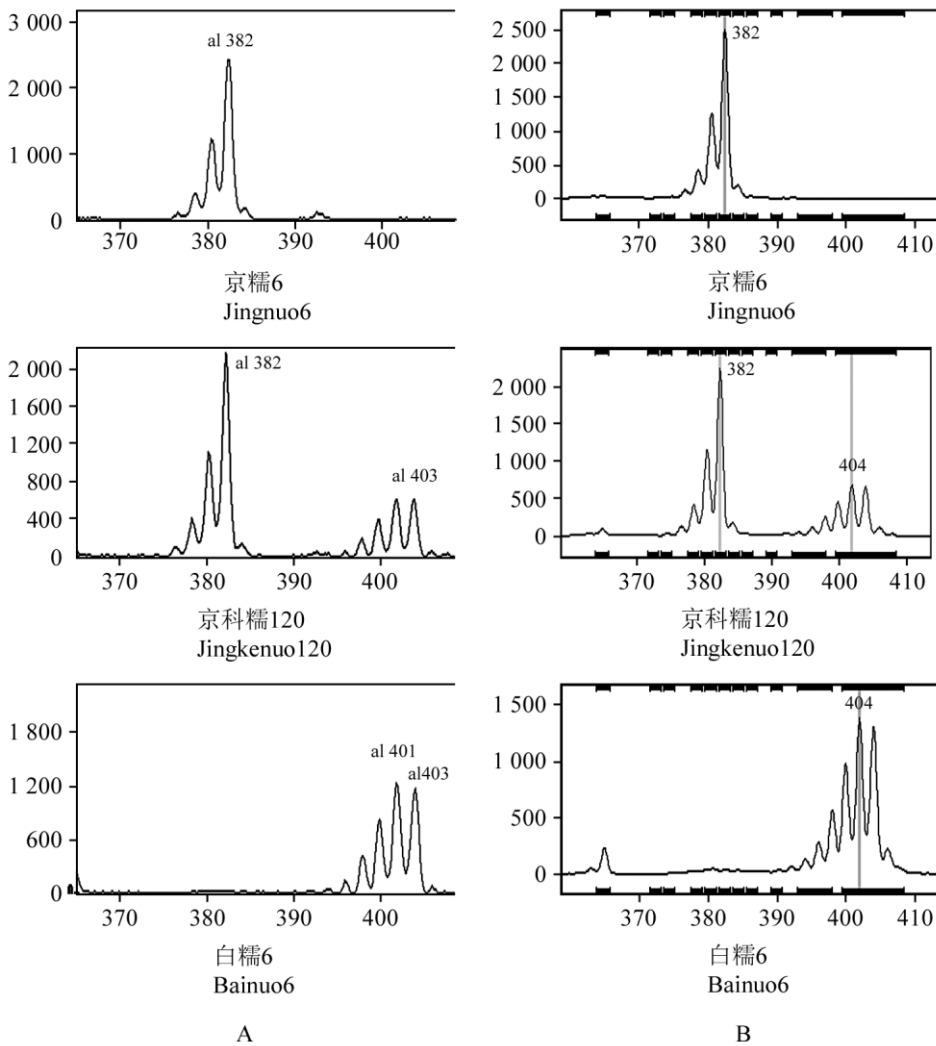


图 4 京科糯 120 其亲本的荧光检测结果(引物: bnlg2291k4)

注: A: GeneMapper v3.7 分析结果; B: GeneMarker-BMSTC v1.0 分析结果

Figure 4 PCR products of Jingkenuo120 and their parents detected with fluorescence detection (Primer: bnlg2291k4)

Note: A: The analysis result of GeneMapper v3.7; B: The analysis result of GeneMarker-BMSTC v1.0

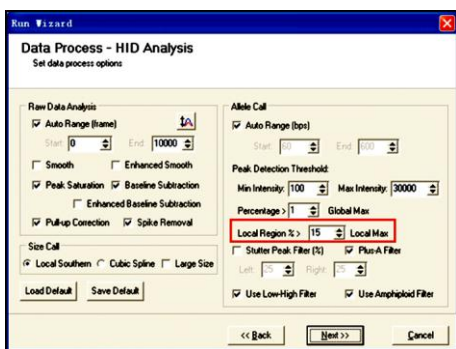


图 5 GeneMarker-BMSTC v1.0 的低峰筛选功能

Figure 5 The function of low peak filter of GeneMarker-BMSTC v1.0

用、快速、高效的要求(Wang et al., 2011; 王风格和赵久然, 2011, 中国农业科学技术出版社, pp.74-78)。但是目前面对各种检测需求的出现, 检测样品量的急剧增加, 对数据分析提出了更高的要求。所选择的 SSR 指纹数据分析软件不但需要满足急剧增长的数据量和数据质量要求, 而且要在易操作性、通用性、自动化等方面提出了更高的要求。同时为了不同实验室间的数据整合与共享, 构建的指纹数据库具有更高的通用性, 分析工具得到的数据格式、存储方式也应满足上述要求(UPOV, 2007)。

通过本文一系列的比较分析可以看出, 在分析相同样品数据时, GeneMapper 与 GeneMarker 软件

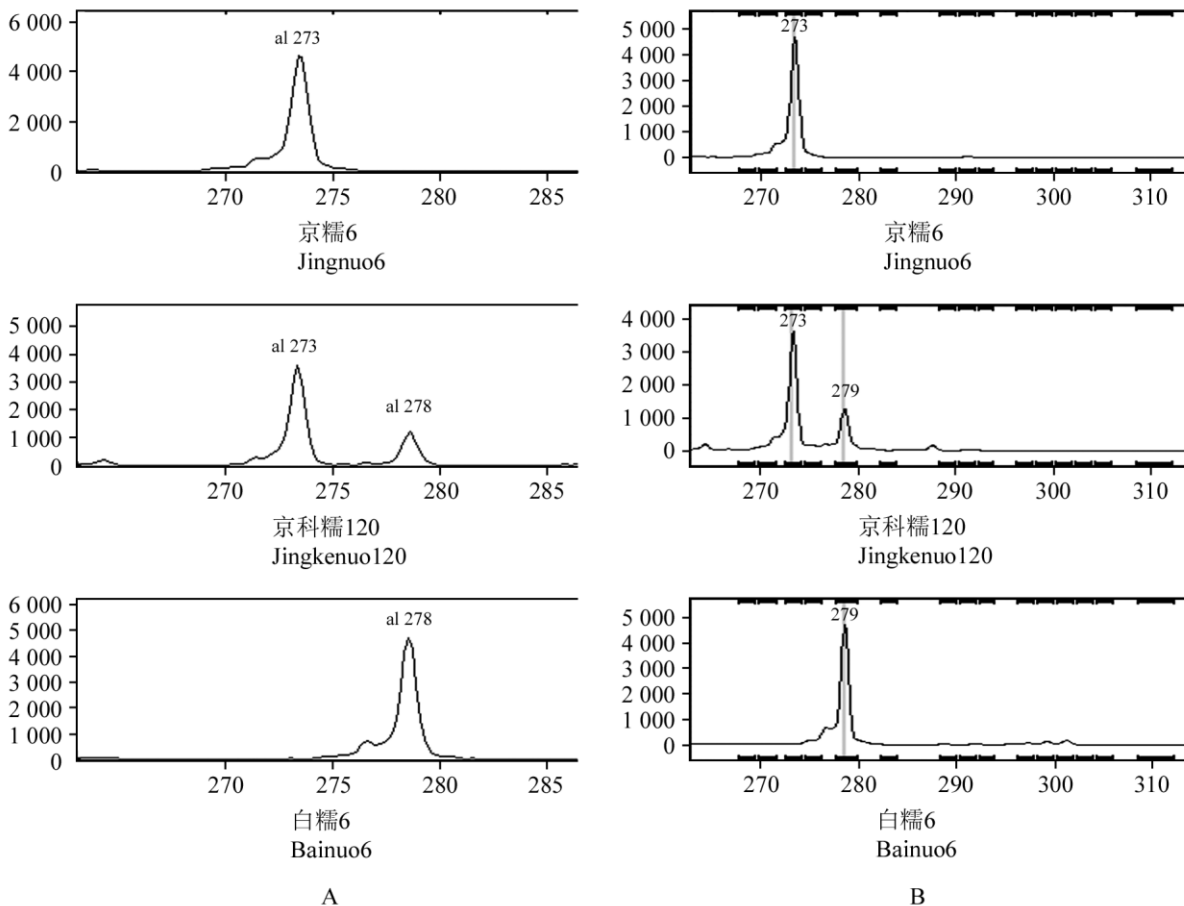


图 6 京科糯 120 及其亲本的荧光检测结果(引物: umc1705w1)
注: A: GeneMapper v3.7 分析结果; B: GeneMarker-BMSTC v1.0 分析结果
Figure 6 PCR products of Jingkenuo120 and their parents detected with fluorescence detection (Primer: umc1705w1)
Note: A: The analysis result of GeneMapper v3.7; B: The analysis result of GeneMarker-BMSTC v1.0

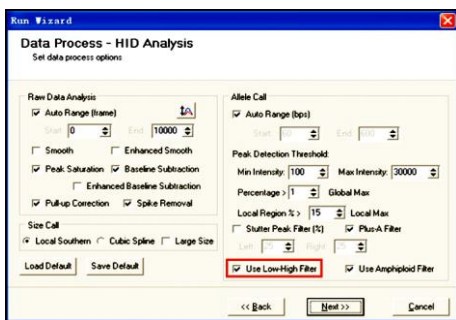


图 7 GeneMarker-BMSTC v1.0 的过滤算法
Figure 7 The function of filter arithmetic of GeneMarker-BMSTC v1.0

的分析数据吻合度达 99%，说明两种软件可靠性都非常高。同时 GeneMarker 相对于 GeneMapper 做出了很多升级，比如二倍体筛选、自动去除 Pull-up 峰、识别连续峰、平移 Panel 等一系列功能，免除

了人工看板增删条带的繁杂工序，大幅度提高了数据分析过程的自动化。GeneMarker 软件因提供功能订制，所以能够实现与数据库对接，即电泳后的 SSR 片段信息作为源数据，在软件自动分析及人工检查后，数据可直接上传至指纹数据库系统，与数据库的对接进一步完成了数据分析自动化的过程。

SSR 指纹数据分析不仅需要开发完善一个功能强大的数据分析软件，还需要一些后续研究工作，比如引物峰型特征描述、BIN 的确定(王风格等, 2006)。SSR 标记具有种属特异性，不同农作物建立不同的引物 Panel，每对引物根据其等位基因特征建立 BIN。每对引物 BIN 的设置包括规定引物范围，各个等位基因的名称、范围，以及记录等位基因的频率，标注稀有等位基因等等。这些工作需要经过对大量样品的分析，并且在 DNA 指纹数据库构建的过程中不断累积完善，作为重要的备注信息添加到 SSR 分析

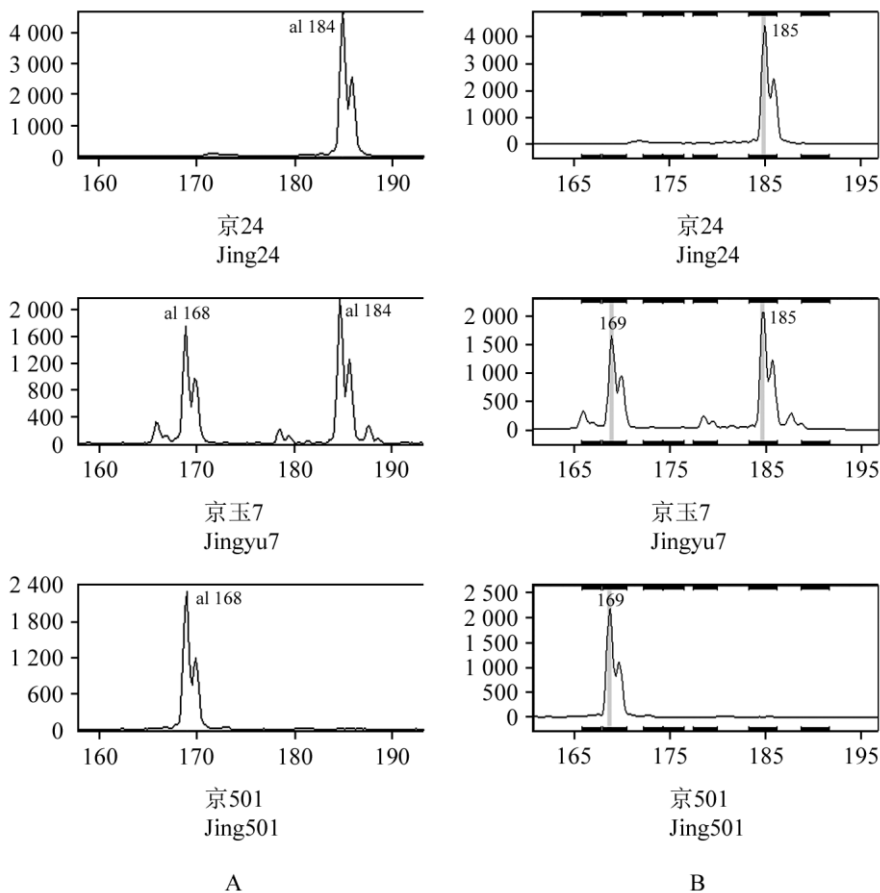


图 8 京玉 7 及其亲本的荧光检测结果(引物: umc1506K12)

注: A: GeneMapper v3.7 分析结果; B: GeneMarker-BMSTC v1.0 分析结果

Figure 8 PCR products of Jingyu7 and their parents detected with fluorescence detection (Primer: umc1506K12)

Note: A: The analysis result of GeneMapper v3.7; B: The analysis result of GeneMarker-BMSTC v1.0

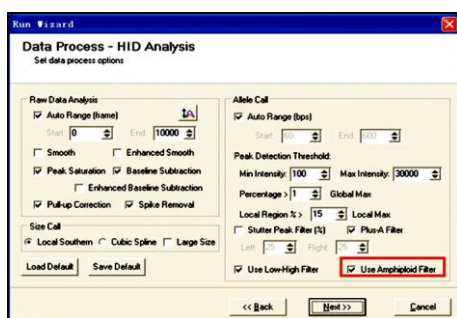


图 9 GeneMarker-BMSTC v1.0 的二倍体筛选功能设置

Figure 9 The setup of the function of the amphiploid filter of GeneMarker-BMSTC v1.0

软件中, 为 DNA 指纹数据分析提供方便和标尺。综上所述, 对于大量数据的分析工作离不开一款功能齐全, 可靠性强, 定制灵活的 SSR 数据分析软件对以上过程进行简化, 同时能在进一步的数据库构建中实现无缝连接, 为整个数据分析平台搭建的

可靠性、便捷性、合理性、扩展性提供重要保障。

3 材料与方法

3.1 材料

本文使用 11 个杂交种及其父母本为研究对象, 材料信息见表 1, 杂交种均为人工组配。

3.2 实验方法

3.2.1 DNA 提取

本文采用 CTAB 小量提法提取 DNA (王风格和赵久然, 2011, 中国农业科学技术出版社, pp.155)。DNA 质量和浓度用琼脂糖电泳和 NanoDrop 2000 (Thermo Scientific)紫外分光光度计进行测定, 根据实际测量值调节 DNA 工作液浓度。

3.2.2 PCR 扩增

PCR 扩增体系: 模板 DNA 3 μ L, 0.25 μ mol/L 引物, 1 \times PCR Buffer, 0.15 μ mol/L dNTP, 2.5 μ mol/L

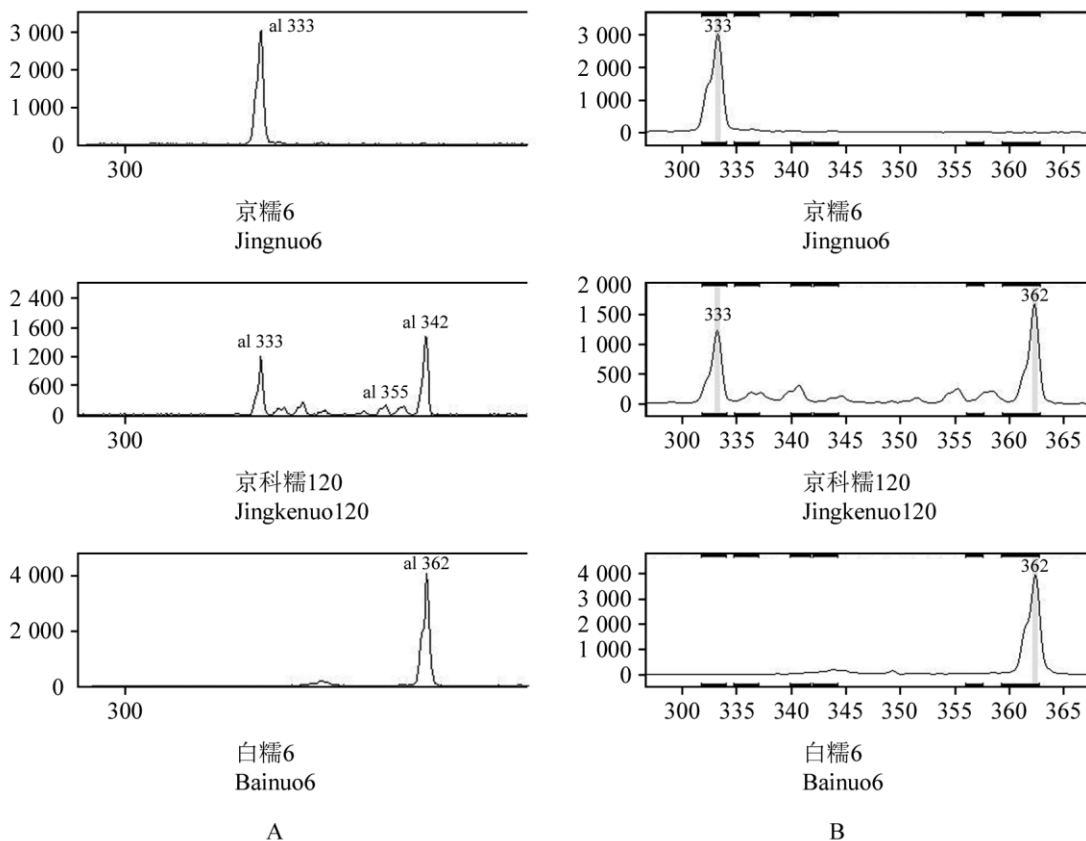


图 10 京科糯 120 其亲本的荧光检测结果(引物:phi053k2)

注: A 为 GeneMapper v3.7 分析结果; B 为 GeneMarker-BMSTC v1.0 分析结果

Figure 10 PCR products of Jingkenuo120 and their parents detected with fluorescence detection (Primer: phi053k2)

Note: A: The analysis result of GeneMapper v3.7; B: The analysis result of GeneMarker-BMSTC v1.0

MgCl₂, 1 单位 Taq 酶, 反应总体积为 20 μL。每对引物中的一条 5'端用荧光染料标记, 选用 PET、NED、VIC、FAM 四种荧光染料(Applied Biosystems, USA 公司合成)。PCR 反应程序: 94℃预变性 5 min, 1 个循环; 94℃变性 40 s, 60℃退火 35 s, 72℃延伸 45 s, 共 35 个循环; 72℃延伸 10 min; 4℃保存。引物选用适于玉米 DNA 指纹分析的 40 核心 SSR 引物(即 20 对基本核心引物+20 扩展核心引物)(王凤格和赵久然, 2011, 中国农业科学技术出版社, pp.189-191)。

3.2.3 检测扩增产物

毛细管电泳: PCR 产物在毛细管荧光电泳系统 AB 3730XL DNA 分析仪(Applied Biosystems, USA)上进行 10 重电泳, 即按照扩增片段大小和荧光颜色差异将 10 对引物的扩增产物混合在一起电泳。在 96 孔电泳板的单个孔中分别加入 1.5 μL 10 对引物 PCR 产物的混合物, 9 μL 甲酰胺, 0.12 μL 分子量内标(GeneScan™-500 LIZ, Applied Biosystems, USA)。95℃变性 5 min, 4℃保存 10 min, 1 000 rpm 离心 1 min

后, 于 AB 3730XL DNA 分析仪上进行电泳。预电泳时间为 15 kV, 2 min; 电泳时间为 15 kV, 20 min。用 Date Collection v1.0 软件收集原始数据。

3.2.4 数据分析

用 GeneMapper v3.7 和 GeneMarker-BMSTC v1.0 软件(由北京市农林科学院玉米研究中心与北京华生恒业科技有限公司基于玉米共同研发)分别对 11 组供试材料的 40 对 SSR 引物的原始数据进行分析 and 校正。

作者贡献

王蕊是本研究的实验设计和实验研究的执行人; 田红丽参与实验设计, 试验结果分析, 指导论文的写作与修改; 王凤格, 赵久然是项目的构思者及负责人, 指导实验设计与论文修改; 易红梅, 王璐参与实验研究; 李楠, 霍永学提供软件相关指导。全体作者都阅读并同意最终的文本。

致谢

感谢北京华生恒业科技有限公司的李欣, 江彬在软件

表1 11组供试材料杂交种及亲本名称

Table 1 Names of 11 groups of hybrids and their parent used in the study

| 编号 No. | 杂交种 Hybrids | 母本 Female parents | 父本 Male parents |
|-----------|-------------------------|----------------------|--------------------|
| 1 | 先玉 335 Xianyu335 | PH6WC | PH4CV |
| 2 | 郑单 958 Zhengdan958 | 郑 58 Zheng58 | 昌 7-2 Chang7-2 |
| 3 | 农大 108 Nongda108 | X178 | 黄 C HuangC |
| 4 | 鲁单 981 Ludan981 | 齐 319 Qi319 | Lx9801 |
| 5 | 浚单 20 Xundan20 | 浚 9058 Xun9058 | 浚 92-8 Xun92-8 |
| 6 | 蠡玉 6 Liyu6 | 连 87 Lian87 | 543 |
| 7 | 沈单 16 Shendan16 | K12 | 沈 137 Shen137 |
| 8 | 东单 60 Dongdan60 | A801 | 丹 598 Dan598 |
| 9 | 京玉 7 Jingyu7 | 京 24 Jing24 | 京 501 Jing501 |
| 10 | 掖单 13 Yedan13 | 478 | 丹 340 Dan340 |
| 11 | 京科糯 120 Jingkenuo120 | 京糯 6 Jingnuo6 | 白糯 6 Bainuo6 |

方面提供重要指导与耐心帮助!本研究由十二五农村领域国家科技技术课题(2011BAD35B09), 国家国际科技合作项目(2010DFB33740)联合资助。

参考文献

UPOV (International Union for the Protection of New Varieties of Plants), 2007, Guidelines for DNA-profiling: molecular marker selection and database construction,

BMT Guidelines (proj.9). http://www.upov.int/en/documents/c/41/bmt_guidelines_proj_9.pdf

Wang F.G., Zhao J.R., Dai J.R., Guo J.L., Yuan Y.P., Wang L., Yi H.M., Sun S.X., and Lu B., 2006, Criteria for the construction of maize DNA fingerprint database, *Yumi Kexue (Journal of Maize Sciences)*, 14(6): 66-68 (王凤格, 赵久然, 戴景瑞, 郭景伦, 原亚萍, 王璐, 易红梅, 孙世贤, 吕波, 2006, 玉米 DNA 指纹数据库建库标准规范的建立, *玉米科学*, 14(6): 66-68)

Wang F.G., Tian H.L., Zhao J.R., Yi H.M., Wang L., and Song W., 2011, Development and characterization of a core set of SSR markers for fingerprinting analysis of Chinese maize varieties, *Maydica*, 5 (1):7-18

Yi H.M., Wang F.G., Zhao J.R., Wang L., Guo J.L., and Yuan Y.P., 2006, Comparison of two maize SSR detection methods: capillary electrophoresis with fluorescence detection method and denaturing PAGE silver-staining detection method, *Huabei Nongxuebao (Acta Agriculturae Boreali-Sinica)*, 21(5): 64-67 (易红梅, 王凤格, 赵久然, 王璐, 郭景伦, 原亚萍, 2006, 玉米品种 SSR 标记毛细管电泳荧光检测法与变性 PAGE 银染检测法的比较研究, *华北农学报*, 21(5): 64-67)

Zhao J.R., Wang F.G., Guo J.L., Chen G., Liao Q., Sun S.X., Chen R.M., and Liu L.Z., 2003, Series of research on establishing DNA fingerprinting pool of Chinese new maize cultivars II. Confirmation of a set of SSR core primer pairs, *Yumi Kexue (Journal of Maize Sciences)*, 11(2): 3-5, 8 (赵久然, 王凤格, 郭景伦, 陈刚, 廖琴, 孙世贤, 陈如明, 刘龙洲, 2003, 中国玉米新品种 DNA 指纹库建立系列研究 II. 适于玉米自交系和杂交种指纹图谱绘制的 SSR 核心引物的确定, *玉米科学*, 11(2): 3-5, 8)