



## 技术主题

## Technology Feature

# SSR 数据格式转换软件 DataFormater

樊文强<sup>1,2</sup>, 盖红梅<sup>3</sup>, 孙鑫<sup>1,2</sup>, 杨爱国<sup>1</sup>, 张忠锋<sup>1</sup>, 任民<sup>1</sup>

1. 中国农业科学院烟草研究所, 烟草遗传改良与生物技术重点开放实验室, 青岛, 266101
2. 中国农业科学院研究生院, 北京, 100081
3. 青岛市农业科学研究院, 青岛, 266100

✉ 通讯作者, [renmin@caas.cn](mailto:renmin@caas.cn); ✉ 作者

分子植物育种, 2016 年, 第 14 卷, 第 5 篇 doi: 10.5376/mpb.cn.2016.14.0005

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):

樊文强等, 2016, SSR 数据格式转换软件 DataFormater, 分子植物育种(online), 14(5): 1029-1034 (doi: 10.5376/mpb.cn.2016.14.0005)

引用格式(英文):

Fan et al., 2016, DataFormater, A Software for SSR Data Formatting to Develop Population Genetics Analysis, Fenzi Zhiwu Yuzhong (online) (Molecular Plant Breeding), 14(5): 1029-1034 (doi: 10.5376/mpb.cn.2016.14.0005)

**摘要** 随着分子群体遗传学的快速发展和 SSR 标记的广泛应用, 一些分子群体遗传学分析软件相继问世。但是不同软件的 SSR 数据输入文件格式并不一致, 人工转换费时费力, 且易出错。因此, 本研究利用 Python 高级计算机语言开发了 SSR 分子标记数据格式转换软件 DataFormater。该软件具有友好的图形用户界面, 能高效准确的把 SSR 原始数据快速转换为 Popgene、Ntsys、PowerMarker、Structure、Tassel、SPAGeDi 等常用分子群体遗传学分析软件的输入文件, 同时还具过滤稀有等位、过滤无多态位点、用户数据校验等功能。与人工转换相比, DataFormater 可大大提高转换的效率和准确性。因此, DataFormater 将是分子群体遗传学研究中不可或缺的工具软件。现在用户可以通过 DataFormater 软件的主页进行下载, 网址为: [www.ycsjk.com.cn/dataformater/home/](http://www.ycsjk.com.cn/dataformater/home/), 或通过电子邮件 [renmin@caas.cn](mailto:renmin@caas.cn) 索取该软件。

**关键词** 分析分子群体遗传学, 分子标记, SSR, 数据格式, 软件, Python 计算机语言

## DataFormater, A Software for SSR Data Formatting to Develop Population Genetics Analysis

Fan Wenqiang<sup>1,2</sup>, Ge Hongmei<sup>3</sup>, Sun Xin<sup>1,2</sup>, Yang Aiguo<sup>1</sup>, Zhang Zhongfeng<sup>1</sup>, Ren Min<sup>1</sup>

1 Key Laboratory of Tobacco Genetic Improvement and Biotechnology, Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao 266101

2 Post Graduate School, Chinese Academy of Agricultural Sciences, Beijing 100081

3 Qingdao Academy of Agricultural Sciences, Qingdao, 266100

✉ Corresponding author, [renmin@caas.cn](mailto:renmin@caas.cn); ✉ Authors

**Abstract** With the rapid development of molecular population genetics and the widespread application of SSR marker, much molecular population genetics software appears, consecutively. However, the format of SSR data for different genetics software is not the same. Manual data transform for different genetic software would take long time and would always make mistakes. Therefore, the software DataFormater was developed with Python language for SSR data formatting to different population genetics software. The software has a friendly graphical user interface, and can transform original SSR data into the input file for Popgene, Ntsys, PowerMarker, Structure, Tassel and SPAGeDi. Meanwhile, it can filter rare alleles and monomorphic loci, and validate users' original data. Compared with artificial way, the software DataFormater can improve the efficiency and accuracy of data formater significantly. So that it's a necessary tool in the study of molecular population genetics. Now, the user can download DataFormater conveniently through [www.ycsjk.com.cn/dataformater/home/](http://www.ycsjk.com.cn/dataformater/home/) or email [renmin@caas.cn](mailto:renmin@caas.cn).

**Keywords** Molecular population genetics; Molecular marker; Simple sequence repeats; Data format; Software; Python language

收稿日期: 2016 年 02 月 29 日

接受日期: 2016 年 03 月 02 日

发表日期: 2016 年 03 月 07 日

基金项目: 本研究由烟草基因组计划重大专项(110201301008)、国家自然科学基金青年基金(31301320)和作物生物学国家重点实验室开放课题(2013KF12)共同资助

## 研究背景

分子群体遗传学是指动植物进化生物学、种质资源、遗传育种和关联作图等基础理论学科, 它由 Kimura 于 1971 年首次提出, 此后 Kreitman (1983) 在黑腹果蝇中开展了基于 DNA 序列变异检测手段的实验分子群体遗传学研究。微卫星 DNA 又叫简



单重复序列(simple sequence repeats, SSR), 由于其共显性遗传、多态性高、重复性好等特点, 在分子群体遗传学研究中得到广泛应用, 如遗传多样性分析(Vigoroux et al., 2005; 盖红梅等, 2005; 任民等, 2005)、核心种质指纹图谱构建(徐军等, 2011)、遗传图谱构建(Röder et al., 1998; Somers et al., 2004; Bindler et al., 2011)、关联分析(Remington et al., 2001; Breseghello and Sorrells, 2006; 王兰芬等, 2007; 任民等, 2014)、育种亲本评估(盖红梅等, 2009)等, 并由此发展出了大量的分析模型、算法以及分子群体遗传学分析软件。

目前, 常用的以 SSR 数据为基础的分子群体遗传学软件主要有 Popgene、Ntsys、PowerMarker、Structure、Tassel、SPAGeDi 等。但是这些软件之间数据输入格式互不兼容, 给前期数据处理带来了很大的困难。而且, 随着高通量 SSR 分型技术的发展, 分子遗传学的数据量日益庞大, 单纯依靠手工处理各种软件的数据格式, 已无法满足研究的需求。迫切需要一种使用灵活、简洁方便、准确无误的数据格式转换软件, 以提高工作效率, 降低科研成本, 避免各种人工操作失误。虽有部分软件具有分子标记数据格式转换的功能, 但均无法满足当前研究的需要(盖红梅和任民, 2011)。

针对这些问题, 盖红梅等(盖红梅和任民, 2011)于 2011 年使用 Microsoft VBA 语言开发了 DataTrans1.0 软件, 实现了将 SSR 原始 bp 数据转换为 Popgene、Ntsys、PowerMarker、Structure 和 Tassel 软件需要的输入格式。随着 DataTrans 的应用(许鲲等, 2014; 方珊茹等, 2012; 郑永胜等, 2014; 郭阳等, 2013), 日益对软件的升级提出了迫切的需求。

基于 Microsoft Office 办公套件的 Excel 运行, 随着 Office 版本的升级以及用户安装方式的不同, 导致了大量的兼容性问题, 给软件的维护和更新带来了诸多困难。加之转化后的数据还需人工保存成相应软件的输入文件, 存在操作上较复杂等原因。

本研究所 DataTrans1.0 的基础上, 又利用 Python 高级计算机语言开发了具有图形用户界面的独立运行软件 DataFormater。该软件全面支持了“01 带型”和“bp 值型”两种常见 SSR 标记原始数据格式向 Popgene、Ntsys、PowerMarker、Structure、Tassel、SPAGeDi 等 6 种分子群体遗传学软件输入文件的转换, 同时还增加了过滤稀有等位、过滤无

多态位点和数据检验的功能。不但提供了更加友好的用户体验, 并进一步提高了工作效率和数据转换的准确性, 也给软件的后续维护升级带来了便利。

## 1 结果与分析

### 1.1 需求分析

在基于 SSR 标记的遗传多样性分析、关联分析等研究中主要用到的软件包括: Ntsys、Popgene、Powermarker、Structure、SPAGeDi 和 Tassel 等 6 种, 用户的原始数据根据电泳检测方式的不同有“01 带型”和“bp 值型”两种, 因此本软件需要将两种用户原始数据直接生成上述 6 种软件的输入文件。通过分析 6 种目标输入文件格式可知, Ntsys 和 Popgene 的输入文件格式基于“01 带型”数据, 而其他 4 种软件的输入文件格式则基于“bp 值型”。由于基于“bp 值型”的输入文件格式应用更加广泛, 因此本研究基于“bp 值型”定义了一个中间数据结构, 所有的目标输入文件均通过该中间数据结构生成。用户的“bp 值型”原始数据加载到内存后直接解析为中间数据结构, 而用户的“01 带型”原始数据加载到内存后还需要增加一步“01 转 bp”的过程, 首先模拟成“bp 值型”原始数据, 进而解析为中间数据结构, 以实现目标输入文件格式的生成。除此之外, 在用户数据分析的过程中, 通常还要按照一定比例过滤稀有等位, 以及去掉无多态的标记位点, 因此在软件开发过程中均需实现相应的功能。

### 1.2 软件的功能实现

本研究所开发软件的数据处理流程如图 1 所示。输入系统的用户原始数据类型为“01 带型”和“bp 值型”两种类型。软件的核心功能分为“数据处理”和“格式转换”两个模块。“数据处理”模块负责实现过滤稀有等位和过滤无多态的位点等功能, 同时还包括了将“01 带型”数据转换成“bp 值型”的功能; “格式转换”模块负责实现从中间数据结构到各种目标文件格式的转换, 并将转换结果保存成磁盘文件。此外, 依据需求分析软件还实现了其他功能模块, 例如读取用 Excel 文件保存的用户原始数据, 在读取的同时进行数据检验, 对有错误的数据给出明确的提示; 将读取后的用户数据, 随时保存成新的 Excel 文件等。

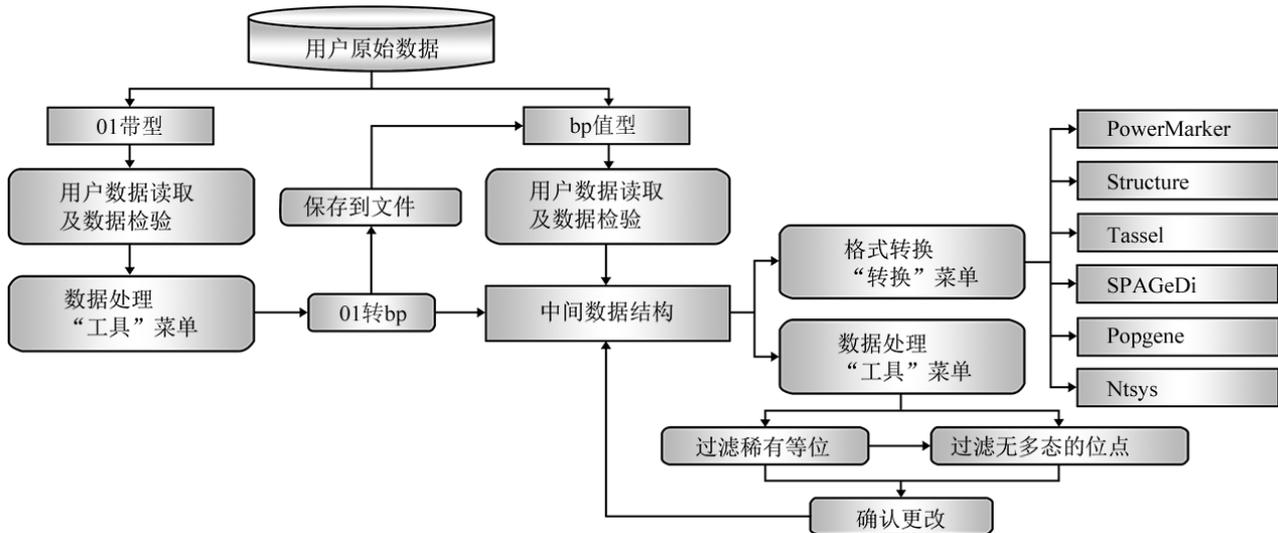


图 1 DataFormatter 软件的数据处理流程图  
 Figure 1 Flow chart of data processing for DataFormatter

### 1.3 用户界面

该软件提供了友好的图形用户界面, 其主界面为典型的窗口程序布置(图 2; 图 3)。最上方为标题栏, 其下为菜单栏和工具栏, 再下方为显示用户数据的窗体表格, 最下方是状态栏。通过“文件”菜单可以从磁盘读取数据, 或保存数据到磁盘; 通过“工具”菜单可以调用软件提供的各种数据处理功能; 通过“转换”菜单下的各命令, 可以直接从用户数据生成目标软件的输入文件。因此, 在未加载用户数据时, “转换”菜单下的命令呈不可用状态。软件的工具栏则提供了常用数据处理和数据转换命令的快捷访问。



图 2 DataFormatter 软件的用户界面  
 Figure 2 User interface of DataFormatter



图 3 DataFormatter 软件的命令菜单  
 Figure 3 Command menu of DataFormatter

### 1.4 DataFormatter 软件的使用方法

#### 1.4.1 读取“01 带型”格式的用户数据

“01 带型”格式的用户数据必须首先转换成“bp 值型”格式的用户原始数据后才能进行下一步的数据处理或数据转换。因此, 打开 DataFormatter 软件, 点击用户界面工具栏中“01 转 bp”按钮, 选择需进行“01 转 bp”的用户原始数据, 如果原始数据中无格式错误, 软件将自动把用户数据转换为“bp 值型”格式。如果用户数据中存在格式错误, 软件将给出相应的提示, 用户可根据提示修改相应的错误, 并重复“01 转 bp”直至成功。

#### 1.4.2 过滤稀有等位

如果需要过滤稀有等位点, 则在打开“bp 值型”用户原始数据后, 点菜单栏“工具”菜单中的“过滤稀有等位”菜单项, 在弹出的对话框中输入要过滤的稀有等位频率, 接着点击“确认变更”菜单项完成稀有等位过滤。过滤后的数据还可进一步过滤无多态的位点或者保存数据到磁盘文件。

#### 1.4.3 用户数据到输入文件的生成

当用户数据正确读取后, 该软件就可以根据用户的原始数据直接生成各种分析软件的输入文件。例如, 若生成 PowerMarker 软件的数据输入文件, 就点击菜单栏“转换”菜单中的“PowerMarker”菜单项, 或工具栏中“PowerMarker”快捷按钮, 在弹出的“保存”对话框中选择保存文件的目录和文件名即可。生成其他软件输入文件的操作步骤以此类推。



## 2 讨论

SSR 标记与其它分子标记相比, 具有基因组覆盖度广, 多态性高、重复性好、呈共显性遗传等特点, 在分子群体遗传学研究中得到了广泛应用。而且随着目前高通量 SSR 分型技术的快速发展, 分子遗传学研究的数据量日益增多, 单纯用手工进行数据转换过于耗时耗力, 且易出错。因此, 本研究为了满足 SSR 原始数据快速转换为不同软件的输入文件, 开发出了能即刻把 SSR 原始数据转换为目标软件输入文件的 DataFormater 软件, 该软件运行速度快, 为 SSR 数据的深入分析提供了高效快捷的途径。

DataFormater 软件是对 DataTrans 1.0 的更新换代, 不仅继承了 DataTrans 1.0 无需安装, 占用内存小, 不产生系统垃圾的所有优点, 而且数据转换速度比 DataTrans 1.0 更快, 同时还增加了过滤稀有等位和无多态性位点等功能。此外, DataFormater 软件是用 Python 语言编写的, 具有独立的操作界面, 能在 Windows、Linux、MacOS 等主流操作系统上运行, 具备很好的兼容性。

DataFormater 软件目前包含了分子群体遗传学常用软件 Popgene、Ntsys、PowerMarker、Structure、SPAGeDi 和 Tassel 的数据转换功能, 基本上满足了研究人员数据转换的需求。但是随着科技技术的飞速发展, 日后必定会有新的分子群体遗传学分析软件问世, 因此 DataFormater 软件也会根据发展不断完善, 为广大使用者提供更多的功能支持。

## 3 材料与方 法

### 3.1 原始数据格式

当前 SSR 检测方法主要有聚丙烯酰胺凝胶电泳和毛细管凝胶电泳(王立新等, 2012)。前者的 SSR 分型数据一般为“01 带型”格式, 而后的 SSR 分型数据为“bp 值型”格式(盖红梅和任民, 2011)。为此, 本研究所用的原始数据包括了“bp 值型”和“01 带型”2 种格式。上述 2 种原始数据格式在 Excel 软件中的组织和保存方式如下:

#### 3.1.1 “01 带型”原始数据的输入格式

用户的原始数据可以直接用 Excel 表格录入和保存(2007 及以上版本), DataFormater 支持直接读取表格内的数据。从 Excel 表格中 C1 单元格开始, 沿第 1 行输入引物名称。在第 1 列输入每个材料的编号, 在第 2 列输入代表居群或其他分类的编号。其他单元格内输入相应的带型信息, 缺失用“9”表示(图 4)。

#### 3.1.2 “bp 值型”原始数据的输入格式

同样将原始数据保存 Excel 2007 及其以上版

本中。从 Excel 表格中 C1 单元格开始, 沿第 1 行输入引物名称, 每个引物的数据占 2 列。第 1 列和第 2 列的内容与“01 带型”数据格式相同。其他单元格内输入带型的 bp 值, 空带用英文状态的“?”表示(图 5)。

|   | A   | B | C    | D | E | F | G | H    | I | J | K |
|---|-----|---|------|---|---|---|---|------|---|---|---|
| 1 |     |   | M-01 |   |   |   |   | M-02 |   |   |   |
| 2 | V01 | 1 | 0    | 0 | 1 | 0 | 0 | 0    | 0 | 1 | 0 |
| 3 | V02 | 1 | 0    | 0 | 1 | 0 | 1 | 1    | 0 | 0 | 0 |
| 4 | V03 | 1 | 0    | 0 | 0 | 0 | 0 | 9    | 9 | 9 | 9 |
| 5 | V04 | 2 | 0    | 1 | 1 | 0 | 0 | 0    | 1 | 0 | 0 |
| 6 | V05 | 2 | 1    | 0 | 0 | 0 | 0 | 0    | 1 | 0 | 0 |
| 7 | V06 | 2 | 0    | 0 | 0 | 1 | 0 | 0    | 0 | 0 | 1 |

图 4 “01 带型”数据在 Excel 表格中的组织形式  
 Figure 4 Organizational format of “01” data in Excel

|   | A   | B | C    | D   | E    | F   |
|---|-----|---|------|-----|------|-----|
| 1 |     |   | M-01 |     | M-02 |     |
| 2 | V01 | 1 | 103  | 103 | 103  | 103 |
| 3 | V02 | 1 | 103  | 105 | 101  | 101 |
| 4 | V03 | 1 | 102  | 102 | ?    | ?   |
| 5 | V04 | 2 | 103  | 103 | 102  | 102 |
| 6 | V05 | 2 | 101  | 101 | 104  | 104 |
| 7 | V06 | 2 | 104  | 104 | 102  | 102 |

图 5 “bp 值型”数据在 Excel 表格中的组织形式  
 Figure 5 Organizational format of “bp” data in Excel

### 3.2 开发语言

该软件基于 Python 高级计算机语言开发, 图形用户界面(GUI)采用了 wxPython 图形库, 利用 openpyxl 库实现了对 Excel (2007-2013 版)文件读写的功能。软件的打包采用 pyinstaller 2.1。

### 3.3 运行环境

由于 Python 计算机语言具有良好的跨平台能力, 因此本研究所开发的软件 DataFormater 亦可跨平台运行在 Windows、Linux、MacOS 等多种操作系统平台之上。不论何种操作系统平台, 为了保证程序的正确运行, 应具备(表 1)所列的运行环境。

表 1 DataFormater 软件的运行环境要求  
 Table 1 Runtime environment of DataFormater

| 运行环境                | 版本            |
|---------------------|---------------|
| Runtime environment | Version       |
| Python              | 2.X, 推荐 2.7 版 |
| wxPython            | 2.8           |
| openpyxl            | 2.2.5         |



## 作者贡献

樊文强、盖红梅、孙鑫、杨爱国、张忠锋、任民是本研究的实验设计和实验研究的执行人;樊文强、孙鑫参与运算处理和论文初稿的写作;盖红梅、杨爱国、张忠锋参与软件运行环境的测试;任民是项目的构思者及负责人,进行全面指导。

## 致谢

本研究由烟草基因组计划重大专项(110201301008)、国家自然科学基金青年基金(31301320)和作物生物学国家重点实验室开放课题(2013K-F12)共同资助。

## 参考文献

- Bindler G., Plieske J., Bakaher N., Gunduz I., Ivanov N., Van der Hoeven R., Ganai M., and Donini P., 2011, A high density genetic map of tobacco (*Nicotiana tabacum* L.) obtained from large scale microsatellite marker development, *Theor. Appl. Genet.*, 123(2): 219-230
- Breseghele F., and Sorrells M.E., 2006, Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars, *Genetics*, 172(2): 1165-1177  
<http://dx.doi.org/10.1534/genetics.105.044586>
- Fang S.R., Zheng Y.M., Wu C.Z., Liu Y.Q., and Xiong X.J., 2012, Genetic diversity analysis of main male sterile lines for indica hybrid rice based on SSR markers, *Fujian Nongye Xuebao (Fujian Journal of Agricultural Sciences)*, 27(11): 1173-1177 (方珊茹, 郑燕梅, 吴春珠, 刘玉芹, 熊雪娇. 基于 SSR 标记的杂交水稻主要不育系遗传多样性分析. 福建农业学报, 2012, 27(11): 1173-1177)
- Ge H.M., Chen C.B., Shen F.F., Zhang W.X., Ren M., Wang Y.W., and Yang Q.W., 2005, Genetic diversity and conservation strategy of *Oryza rufipogon* along the Haojiang river in Guangxi Zhuang Autonomous Region, *Zhiwu Yichuan Ziyuan Xuebao (Journal of Plant Genetic Resources)*, 6(2): 156-162 (盖红梅, 陈成斌, 沈法富, 张万霞, 任民, 王玉薇, 杨庆文, 2005, 广西武宣濠江流域普通野生稻居群遗传多样性及保护研究, 植物遗传资源学报, 6(2):156-162)
- Ge H.M., and Ren M., 2011, DataTrans1.0, a software for microsatellite data processing based on Excel macro, *Fenzi Zhiwu Yuzhong (Online) (Molecular Plant Breeding (Online))*, 9: 1359-136 (盖红梅, 任民, SSR 数据处理宏程序 DataTrans 1.0, 分子植物育种(网络版), 9: 1359-1365)
- Ge H.M., Wang L.F., You G.X., Hao C.Y., Dong Y.C., and Zhang X.Y., 2009, Fundamental roles of cornerstone breeding lines in wheat reflected by SSR random

- scanning, *Zhongguo Nongye Kexue (Scientia Agricultura Sinica)*, 42(5): 1503-1511 (盖红梅, 王兰芬, 游光霞, 郝晨阳, 董玉琛, 张学勇, 2009, 基于 SSR 标记的小麦骨干亲本育种重要性研究, 中国农业科学, 42(5): 1503-1511)
- Guo Y., Xiong H.P., Chen P., Wang Y.Z., Chen J.K., Tan L.T., Zheng J.S., and Yu C.M., 2013, Analysis on genetic diversity of S1 progeny of ramie variety Zhongzhu No. 2 by SSR markers, *Zhongguo Maye Kexue (Plant Fiber Sciences in China)*, 35(2): 75-80 (郭阳, 熊和平, 陈平, 王延周, 陈继康, 谭龙涛, 郑建树, 喻春明, 2013, 中苧 2 号苧麻自交 S1 代遗传多样性的 SSR 标记分析, 中国麻业科学, 35(2): 75-80)
- Kimura M., 1971, Theoretical foundation of population genetics at the molecular level, *Theor. Popul. Biol.*, 2(2): 174-208  
[http://dx.doi.org/10.1016/0040-5809\(71\)90014-1](http://dx.doi.org/10.1016/0040-5809(71)90014-1)
- Kreitman M., 1983, Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster*, *Nature*, 304: 412-417  
<http://dx.doi.org/10.1038/304412a0>
- Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M., and Buckler E.S., 2001, Structure of linkage disequilibrium and phenotypic associations in the maize genome, *Proc. Natl. Acad. Sci. USA*, 98(20): 11479-11484  
<http://dx.doi.org/10.1073/pnas.201394398>
- Ren M., Chen C.B., Rong T.Z., Zhang W.X., Ge H.M., and Yang Q.W., 2005, Genetic diversity of *Oryza rufipogon* Griff. in southeast region of Guangxi in China, *Zhiwu Yichuan Ziyuan Xuebao (Journal of Plant Genetic Resources)*, 6(1): 31-36 (任民, 陈成斌, 荣廷昭, 张万霞, 盖红梅, 杨庆文, 2005, 桂东南地区普通野生稻遗传多样性研究, 植物遗传资源学报, 6(1): 31-36)
- Ren M., Zhang C.J., Jiang C.H., Cheng L.R., Jia X.H., and Yang A.G., 2014, Association analysis of tobacco aroma constituents based on high density SSR linkage group, *Zhongguo Yancao Xuebao (Acta Tabacaria Sinica)*, 20(4): 88-93 (任民, 张长静, 蒋彩虹, 程立锐, 贾兴华, 杨爱国, 2014, 基于高密度 SSR 连锁群的烟草致香物质关联分析, 中国烟草学报, 20(4): 88-93)
- Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P., and Ganai M.W., 1998, A microsatellite map of wheat, *Genetics*, 149(4): 2007-2023
- Somers D.J., Isaac P., and Edwards K., 2004, A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.), *Theor. Appl. Genet.*, 109(6): 1105-1114  
<http://dx.doi.org/10.1007/s00122-004-1740-7>



- Vigoroux Y., Mitchell S., Matsuoka Y., Hamblin M., Kresovich S., Smith J.S., Jaqueth J., Smith O.S., and Doebley J., 2005, An analysis of genetic diversity across the maize genome using microsatellites, *Genetics*, 169(3): 1617-1630  
<http://dx.doi.org/10.1534/genetics.104.032086>
- Wang L.F., Balfourier F., Hao C.Y., Exbrayat-vinson F., Dong Y.C., Ge H.M., and Zhang X.Y., 2007, Comparison of genetic diversity level between European and East-Asian wheat collections using SSR markers, *Zhongguo Nongye Kexue (Scientia Agricultura Sinica)*, 40(12): 2667-2678 (王兰芬, Balfourier F., 郝晨阳, Exbrayat-vinson F., 董玉琛, 盖红梅, 张学勇, 2007, 欧洲与东亚小麦品种遗传多样性的比较分析, *中国农业科学*, 40(12): 2667-2678)
- Wang L.X., Zhang X.J., Shi X.Y., Gao H., and Zhao Z.Y., 2012, Establishment of SSR fingerprinting database on major apple (*Malus × domestica*) cultivars, *Guoshu Xuebao (Journal of Fruit Science)*, 29(6): 971-977 (王立新, 张小军, 史星云, 高华, 赵政阳, 2012, 苹果栽培品种 SSR 指纹图谱的构建, *果树学报*, 29(6): 971-977)
- Xu J., Liu Y.H., Ren M., Mou J.M., Zhang X.W., Chen Y.C., and Wang Z.D., 2011, SSR fingerprint map analysis of tobacco germplasms, 2011, *Zhongguo Yancao Kexue (Chinese Tobacco Science)*, 32(2): 62-65 (徐军, 刘艳华, 任民, 牟建民, 张兴伟, 陈雅琼, 王志德, 2011, 普通烟草种质资源的 SSR 标记与指纹图谱分析, *中国烟草科学*, 32(2): 62-65)
- Xu K., Li F., Wu J.F., Gu T.C., Chen B.Y., Gao G.Z., Yan G.X., Li J., Qiao J.W., Wang N., and Wu X.M., 2014, Fingerprint identification platform of capillary electrophoresis detection with fluorescent SSR markers on national winter rapeseed varieties (lines) field trials, *Zhongguo Youliao Zuowu Xuebao (Chinese Journal of Oil Crop Sciences)*, 36(2): 150 -159 (许鲲, 李锋, 吴金锋, 谷铁城, 陈碧云, 高桂珍, 闫贵欣, 李俊, 乔江伟, 汪念, 伍晓明, 2014, SSR 荧光标记毛细管电泳法与国家冬油菜区试指纹鉴定平台的构建, *中国油料作物学报*, 36(2): 150 -159)
- Zheng Y.S., Zhang H., Wang D.J., Sun J.M., Wang X.M., Duan L.L., Li H., Wang W., and Li R.Y., 2014, Development of a wheat variety identification system based on fluorescently labeled SSR markers, *Zhongguo Nongye Kexue (Scientia Agricultura Sinica)*, 47(19): 3725-3735 (郑永胜, 张晗, 王东建, 孙加梅, 王雪梅, 段丽丽, 李华, 王玮, 李汝玉, 2014, 基于荧光检测技术的小麦品种 SSR 鉴定体系的建立, *中国农业科学*, 47(19): 3725-3735)