MapGene2Chrom 基于 Perl 和 SVG 语言绘制基因物理图谱

晁江涛，孔英珍，王倩，孙玉合，龚达平，吕婧，刘贯山

中国农业科学院烟草研究所，烟草行业烟草基因资源重点实验室，青岛 266101

摘要: 遗传图谱表现形式简洁明了，为分析遗传规律、克隆基因提供了便利。Gbrowse、MapViewer 等工具虽然能够协助研究人员绘制相似形式的物理图谱，但有很大的局限性：(1)数据需提前布置好；(2)输出结果无法灵活修改。鉴于此，文章基于 Perl 和 SVG 语言，开发了一款生物辅助作图软件 MapGene2Chrom 的本地版与网页版，该软件能够依据输入数据快速绘制相应的物理图谱。该软件输入数据格式简单，输出结果易于修改，图片格式为 SVG 矢量图，具有很好的移植性，以期为研究人员绘制物理图谱提供便利。

关键词: 物理图谱；基因分布图；SVG 矢量图；作图；生物软件

MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages

Jiangtao Chao, Yingzhen Kong, Qian Wang, Yuhe Sun, Daping Gong, Jing Lv, Guanshan Liu

Key Laboratory for Tobacco Gene Resources, Tobacco Research Institute of Chinese Academy of Agricultural Sciences, Qingdao 266101, China

Abstract: Genetic linkage map is helpful for analysis on heredity of some gene families and map-based gene cloning because of its simple and elegant manifestation. One software is in need to draw a gene physical map, which shows a manner similar to the genetic linkage map, based on the relative physical distance between genes. Although some tools like GBrowse and MapViewer etc. are available to draw gene physical map, there are obvious limitations for them: (1) the data need to be decorated in advance; (2) users can't modify results. Therefore, we developed a bio-assisted mapping software——MapGene2Chrom with PC and web versions, which is based on Perl and SVG languages. The software can be used to draw the corresponding physical map quickly in SVG format based on the input data. It will become a useful tool for drawing gene physical map with the advantages of simple input data format, easily modified output and very good portability.

Keywords: physical map; gene distribution map; SVG vector graph; draw map; bio-software

收稿日期: 2014−06−24; 修回日期: 2014−08−06
基金项目: 中国烟草总公司科技重大专项(编号：110201301005[JY-05])项目资助
作者简介: 晁江涛，硕士，助理研究员，研究方向：生物信息学。E-mail: chaojiangtao@caas.cn
通讯作者: 刘贯山，研究员，研究方向：烟草突变体鉴定与利用。E-mail: liuguanshan@caas.cn
DOI: 10.16288/j.yczz.2015.01.013
网络出版时间: 2014-11-19 16:47:08
URL: http://www.cnki.net/kcms/detail/11.1913.R.20141119.1647.002.html

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

依据分子标记之间的遗传距离绘制一张简洁、美观、大方的遗传图谱，为研究人员分析某一性状的遗传规律及图位克隆基因提供了便利。目前能够绘制遗传图谱的软件有 4 种：MAPMAKER[1]，JoinMap[2~4]，Mapplotter[5]和 MapDraw[6]。前两种软件均具备计算遗传距离和绘制连锁图谱的功能，其中 MAPMAKER 侧重于 Mac OS 平台；JoinMap 更侧重于 Windows 平台。后两种软件具备输出遗传图谱的功能，是 MAPMAKER 在 Windows 平台图形化功能的补充。

当研究人员绘制一张物理图谱时，虽然可以借助于现有公共数据库，如 NCBI(www.ncbi.nlm.nih.gov)、JGI(www.phytozome.net)、TAIR(www.arabidopsis.org)中的 GBrowse[7]、MapViewer(www.ncbi.nlm.nih.gov/mapview)等工具来实现，但这些工具存在较大的局限性：(1)数据需事先布置在这些工具中；(2)输出结果样式无法灵活定制。这大大限制了这些工具在绘制美观物理图谱方面的应用。鉴于此，本文基于 Perl(Practical Extraction and Report Language) 和 SVG(Scalable Vector Graphics)矢量图语言，开发了一款生物辅助软件 MapGene2Chrom 的本地版及网页版。输入指定格式的数据，并做简单的参数设置，即可快速绘制一张简洁、美观、大方的基因物理图谱，如果输出效果不够美观，只需修改参数重新运行软件即可。网页版软件访问地址：http://www.tobaccomdb.com/tools/index_mapGene2Chrom.html，源码可从该网页下载，或直接与作者联系索取。

1 MapGene2Chrom 的原理与特点

MapGene2Chrom 基于 Perl 和 SVG 语言，根据染色体上基因的相对物理距离，可快速画出不同染色体上基因的分布矢量图。基因的位置信息通过常用的记事本、写字板或 Microsoft Excel 办公软件进行简单处理即可作为输入文件；之后，运行软件，用户将得到一张基因分布 SVG 矢量图，该文件可通过谷歌 chrome、FireFox、IE9+或者其他支持 SVG 的浏览器查看。由于用 MapGene2Chrom 软件画图时，采用直接输出 SVG 标记语句，所以用户只要有 Perl 语言基本环境即可使用。下文将分别介绍本地版与网页版软件的使用方法。

2 MapGene2Chrom 本地版的使用方法

从 JGI(phytozome v9.0)数据库下载二穗短柄草(*Brochypodium distachyon, Bdistachyon*)的基因注释数据 (ftp://ftp.jgi-psf.org/pub/compgen/phytozome/v9.0/*Bdistachyon*/)，随机选取基因数据。以 Windows 操作系统为例，分别从运行环境、数据准备、运行软件、相关参数等方面介绍其使用方法。

2.1 运行环境

第一步：打开 DOS 命令窗口。点击桌面左下角的“开始”—>“运行” 输入“cmd”即可打开 DOS 命令窗口。

第二步：在 DOS 命令窗口，输入命令“perl-v”，如果能看到 Perl 软件的版本号，表示电脑可以使用本软件。如果看到反馈信息为：不是内部或外部命令，也不是可运行的程序或批处理文件，则说明电脑需要安装 Perl 程序，下载页面为：http://www.perl.org/get.html。

2.2 数据准备

绘制染色体上的基因分布图，需要收集基因相关信息：(1)基因名称；(2)基因开始位置；(3)基因结束位置；(4)基因所在染色体名称；(5)基因所在染色体的长度。之后将(1)~(4)信息依次汇总至文本文件 input_gene_info.txt 中；将(4)、(5)信息汇总至文本 input_chrom_info.txt 中。

本文通过自编写的 Perl 程序 randomSelect-Info.pl(其源码可软件包中找到)分别从每条染色体中选取 60、80、100 个基因数据样本，依次存入文本文件 60_bd_origin_gene_info.txt、80_bd_origin_gene_info.txt、100_bd_origin_gene_info.txt 中。

2.3 运行软件

如需了解软件用法，打开 DOS 窗口输入“perl mapGene2Chrom.pl”即可获得相关信息。其输入参数有：(1)-i <input> 输入文件 1：包含基因及染色体的信息；(2)-chrom <chrom_info> 输入文件 2：包含染色体名称及长度信息；(3)-o <output> 输出文件前缀：后缀为“.svg”；(4)-setup <setup_files> 配置文件：绘制分布图所需的参数值，用文本打开修改即可。

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

将 2.2 中准备好的数据输入软件，即可获得结果文件 output_60.svg，output_80.svg，output_100.svg 中。本文选取不同密度的分布图做对比(图 1) ;同时，也抽查了不同染色体、等量基因的效果(图 2)。这些均为矢量图，随意放大或缩小，可直接用于发表文章。

2.4　参数说明

在设计之初，将各个元素的绘图参数逐一放入配置文件 setup.txt 中，以便于用户个性化定制显示效果。在配置文件中：(1)注释行以“ # ”开头，相关说明信息会存入注释行中。(2)含有“ = ”的行，为参数设置行，等号前的字符为参数名称；等号后的字符为参数值。涉及绘图效果的参数共有 30 个，这些参数涉及标题的字体、字体大小及颜色；基因名称的字体、字体大小及颜色；染色体边框线宽度、颜色；连接线的宽度、颜色等，详情请参考表 1。

3　MapGene2Chrom 网页版的使用方法

MapGene2Chrom 网页版与本地版相比，相同点：核心算法与参数设置相同；不同点：网页版操作更为简单、直观，但受网络因素影响较大，数据计算量限制为 100 kb(约为 400 个基因)。绘图参数共 30 个，详情参考表 1，界面如图 3 所示。

在使用 MapGene2Chrom 网页版时，只需将指定格式的基因信息和染色体信息粘贴至文本框，点击绘图即可得到输出结果(图 4)。由于绘图方式采用的是 SVG 语言，建议使用谷歌 Chrome 或 FireFox 浏览器。

4　MapGene2Chrom 核心算法

软件绘制物理图谱的大致流程(图 5)为：(1)读取输入文件，分析共有几条染色体、每条染色体上有几个基因；(2)对每条染色体上的基因信息，依据起始位置升序排列；(3)统一单位，以染色体最长的为参考，计算每像素代表的序列长度，进而计算每条染色体的尺寸，并分配其具体位置；(4)依据计算结

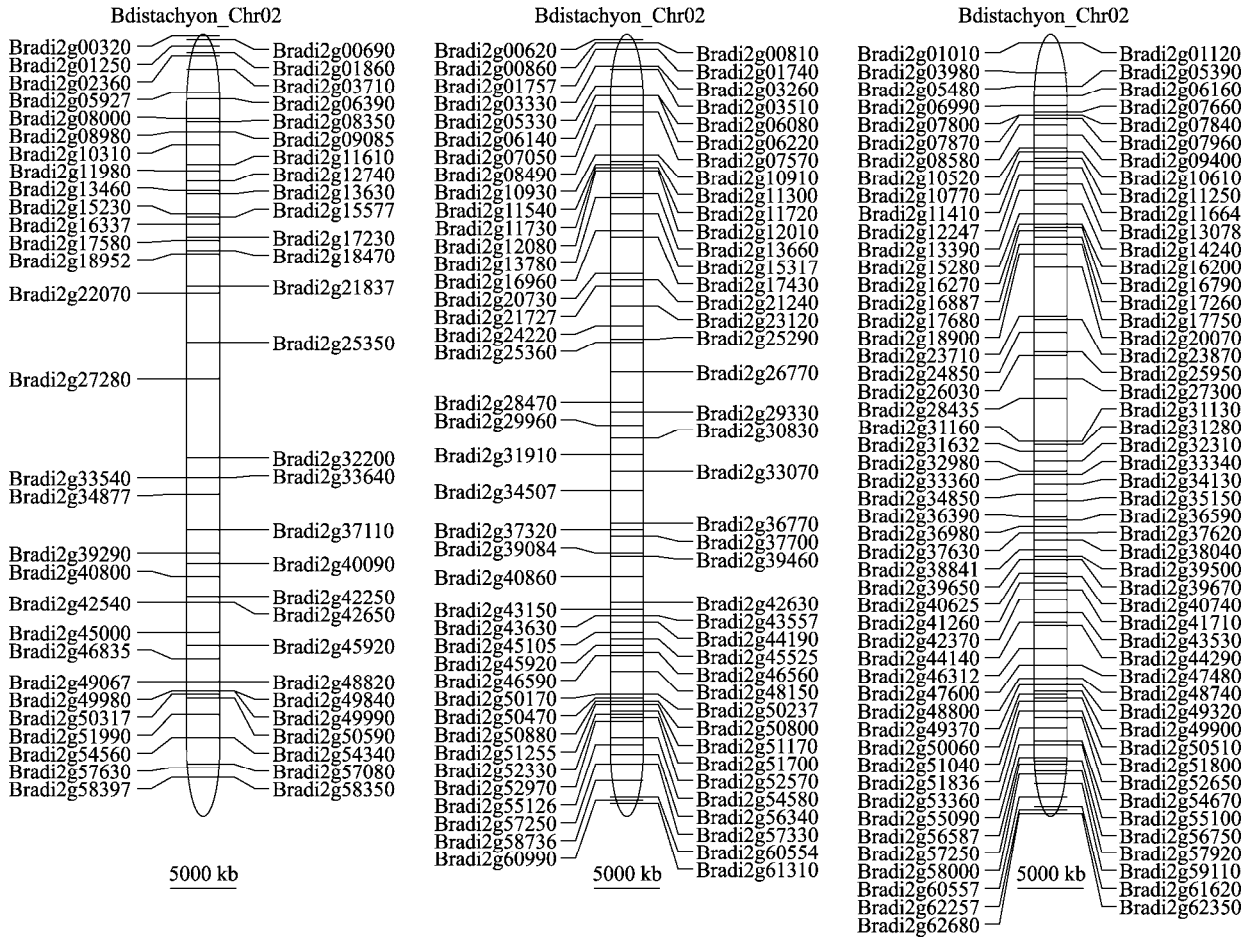

图 **1**　从染色体 **Chr02** 上分别随机选取 **60、80、100** 个基因绘制不同密度分布图的效果

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

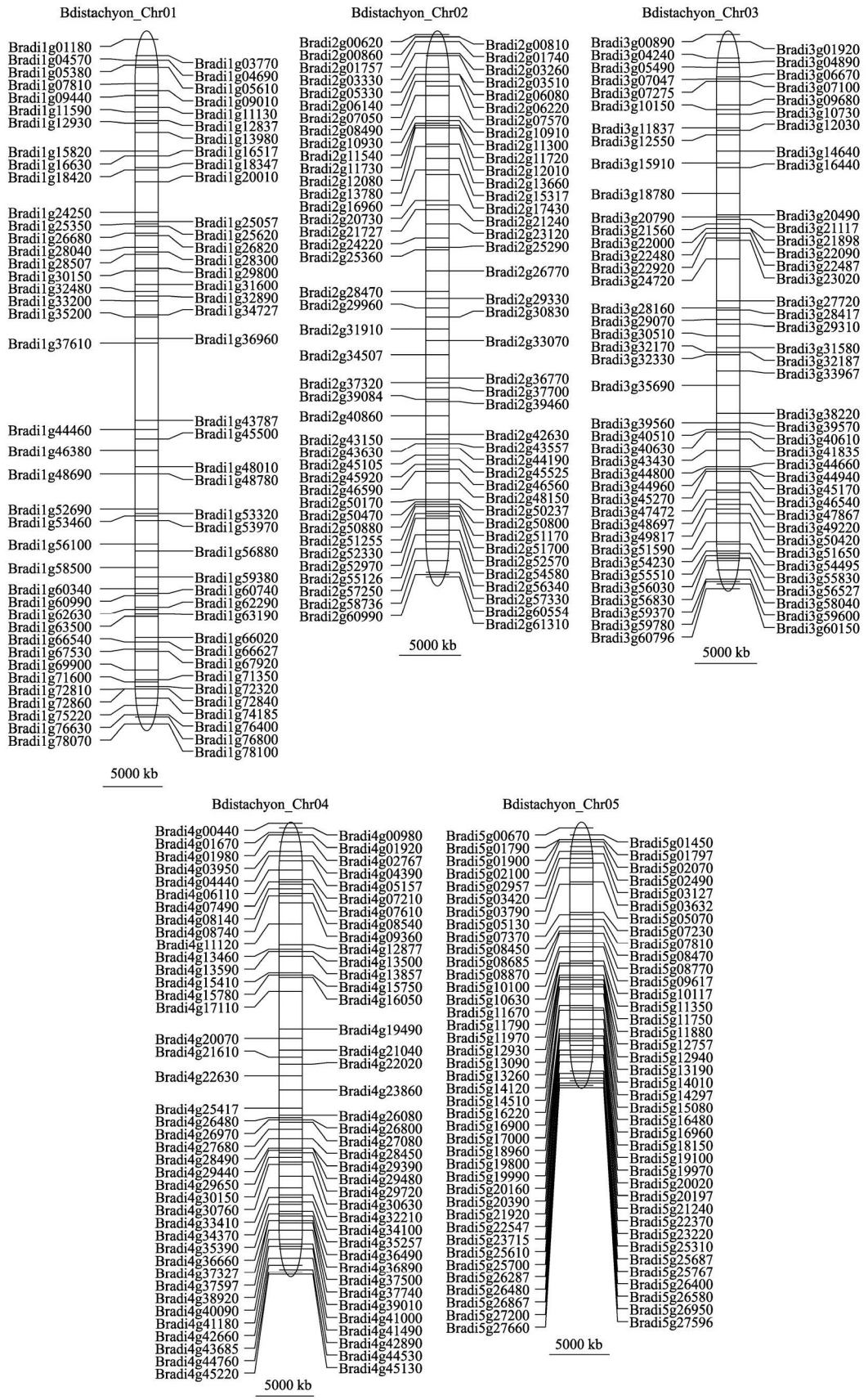

图 2 从二穗短柄草的 5 条染色体上随机选取 80 个基因的分布图效果

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

表 **1**　配置文件参数说明

参数名称	默认参数	参数说明
title_font_family	Arial	标题字体
title_font_size	10	标题字体大小
title_font_color	black	标题字体颜色
svg_width	1000	SVG 图片所占的最大宽度
svg_height	900	SVG 图片初始高度
svg_chrom_height	350	绘制染色体区域的高度
svg_chrom_width	300	绘制染色体区域的宽度
svg_chrom_fill_color	white	绘制染色体区域的填充颜色
svg_chrom_border_color	gray	染色体区域边界线的颜色
svg_chrom_border_width	1	染色体区域边界线宽度
chrom_init_len	300	染色体高度
chrom_init_width	15	染色体宽度
chrom_rx	21	染色体圆角半径
chrom_ry	21	染色体圆角半径
chrom_fill_color	white	染色体边框内的填充色
chrom_border_color	black	染色体边框线颜色
chrom_border_width	1	染色体边框线宽度
gene_line_type	1	基因位置显示方式，1 显示中间线；2 显示上下边界
gene_line_color	gray	基因线的颜色
gene_line_width	0.5	基因线的宽度
gene_name_font_color	black	基因名称字体颜色
gene_name_font_family	Arial	基因名称字体
gene_name_font_size	8	基因名称字体大小
geneName2chrom_margin	20	基因名称与染色体边框线的间距
link_polyline_color	gray	基因名称与基因线连线的颜色
link_polyline_width	0.5	基因名称与基因线连线的宽度
scale_chrom_margin_y	20	刻度尺与染色体的垂直间距
scale_len	20	刻度尺显示的长度
scale_unit	bp	相对距离的单位,只可取 bp 或 cM
scale_unit_float	3	单位距离值保留的有效数字位数。

注：表中涉及宽度、高度、位置信息的单位均为像素(pixel)。关于颜色值可为 none, gray, white, black, red, yellow, pink, green, blue 等。

果绘制染色体名称及染色体边框；(5)绘制基因名称及连接线；(6)在染色体正下方绘制刻度尺，并将所有绘图信息输出至 SVG 文件，供用户查看。

5　讨　论

MapGene2Chrom 是基于 Perl 语言开发的一种生物辅助作图软件，目前有本地版和网页版软件供用户选择使用。本地版软件的优点是数据量不受限制；缺点是操作较为繁琐，用户需具备一定计算机基础。网页版软件的优点是操作界面简单、直观，所有用户均可使用；缺点是数据量受限制(限制为 100 kb，约 400 条基因数据)。软件通过分析不同染色体上的基因相对距离来描绘其分布图，以直观、简洁的方式展示给用户。除此之外，软件的用途还可做以下延伸：(1)绘图功能亦适用于 Scaffold；(2)如果将基因的物理距离变为分子标记的遗传距离，物理图谱就变成了遗传连锁图谱，两者之间的换算需要用户自行完成；(3)相对位置的单位除了碱基对(Base pair, bp)

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

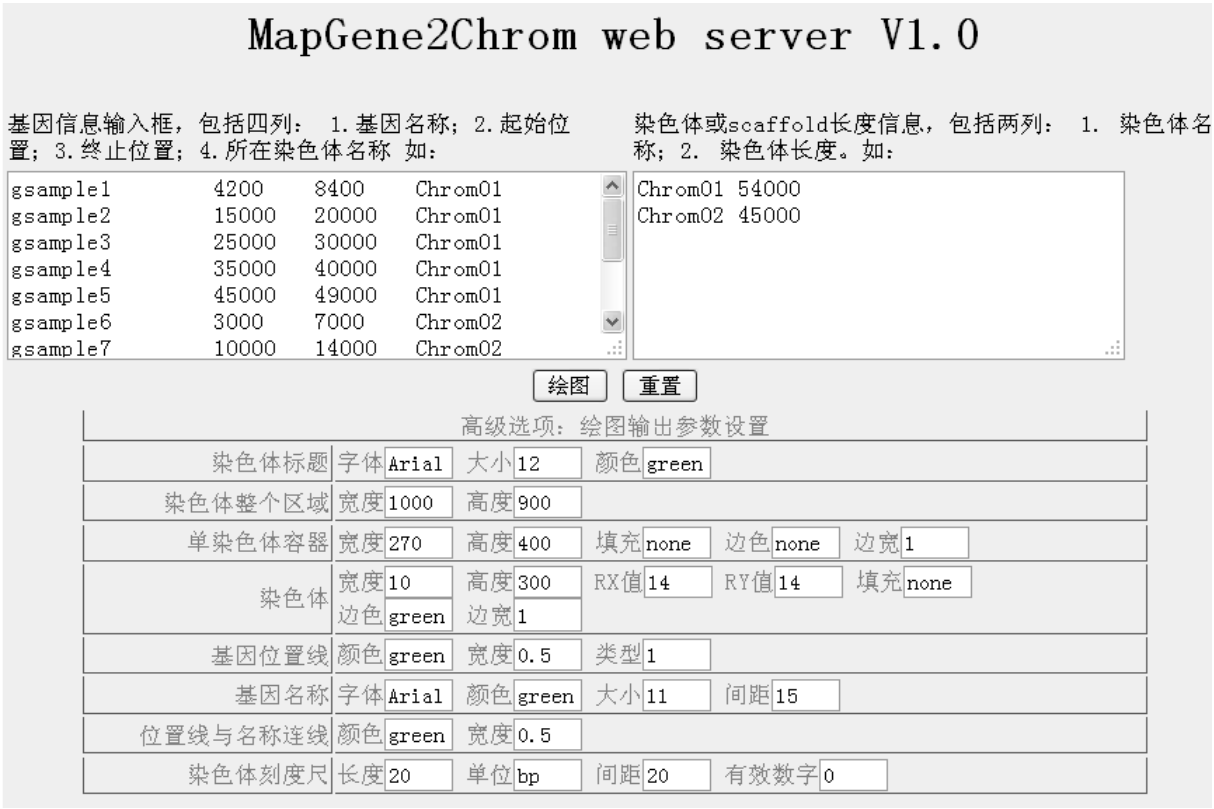

图 3 MapGene2Chrom 网页版操作界面

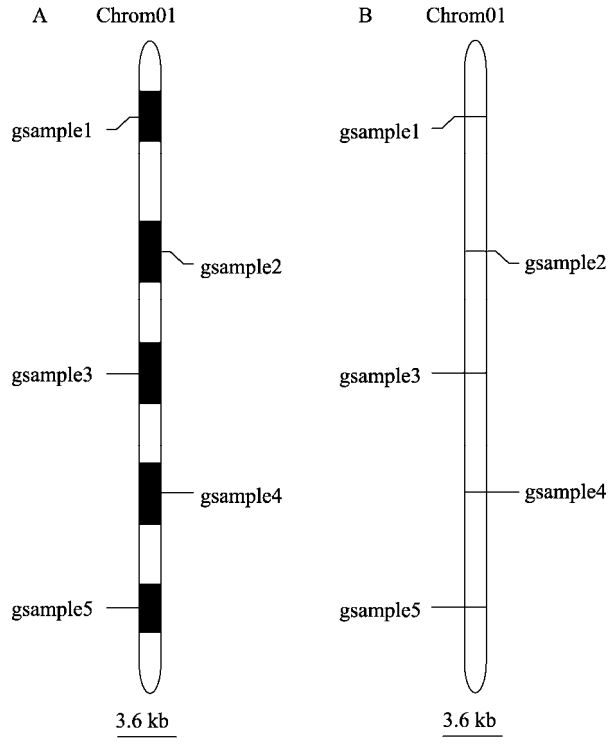

图 4 绘图采用不同基因线类型值的显示效果
A：上下边界；B：边界中间线。

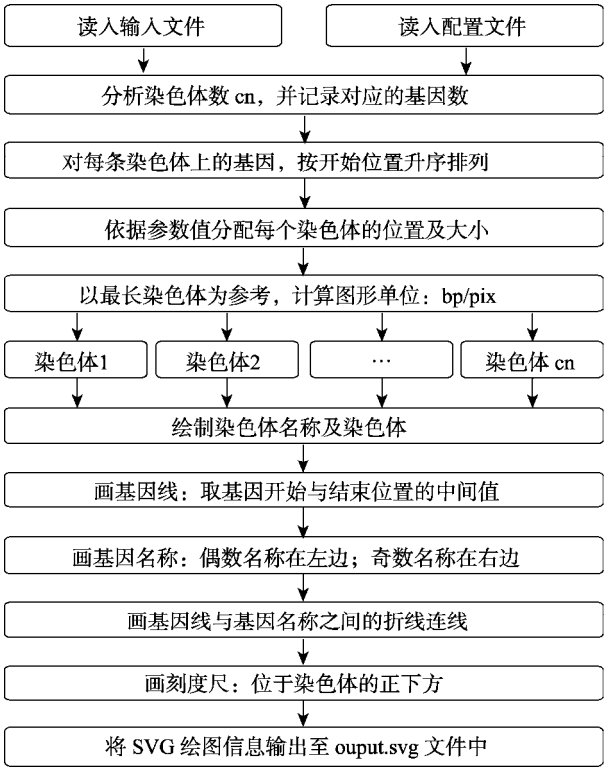

图 5 MapGene2Chrom 执行流程图

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

外，也可用遗传距离里摩(Centi-Morgan, cM)，此项参数是依据输入数据的单位来定的；(4)本文展示的分布图最多为 100 个基因，如需显示更多，本地版软件需修改配置文件中的参数 svg_chrom_height 和 chrom_init_len；网页版软件，需调整单染色体容器及染色体的高度。

另外，用户需要注意的是：在对同一物种构建遗传图谱时，不同的试验家系和试验群体规模，其遗传图谱并不完全一致；再加上基因组中重组热点的存在，会导致遗传图谱与物理图谱并非严格的一一对应关系。所以，软件不会直接将物理距离与遗传距离直接变换，具体选择物理距离(bp)还是遗传距离(cM)绘图，是由用户的数据来决定的，两者之间无准确的换算关系，取决于用户的经验。

MapGene2Chrom 输出结果为 SVG 矢量图，推荐用谷歌 Chrome、Firefox 等常用浏览器中打开查看，能够随意放大或缩小、简单灵活、易于修改，便于研究人员使用，为生物信息学分析提供辅助。

参考文献

[1] Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, 1987, 1(2): 174–181.

[2] VAN Ooijen JW. Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res*, 2011, 93(5): 343–349.

[3] Stam P. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J*, 1993, 3(5): 739–744.

[4] Stam P. JoinMap 2. 0 deals with all types of plant mapping populations. In: Plant Genome III Abstracts. San Diego, USA, 1995.

[5] 刘仁虎, 孟金陵. MapDraw, 在 Excel 中绘制遗传连锁图的宏. 遗传, 2003, 25(3): 317–321.

[6] 沈利爽, 郑先武, 朱立煌. Mapplotter——个输出遗传图谱、图示基因型和 QTL 曲线图形的软件. 遗传, 2000, 22(3): 172–174.

[7] Stein LD, Mungall C, Shu SQ, Caudy M, Mangone M, Day A, Nickerson E, Stajich JE, Harris TW, Arva A, Lewis S. The generic genome browser: a building block for a model organism system database. *Genome Res*, 2002, 12(10): 1599–1610.

(责任编委: 吴为人)

?1994-2015 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net