Chromosome Res (2016) 24:217–223
DOI 10.1007/s10577-015-9513-5

chromDraw: an R package for visualization of linear and circular karyotypes

Jan Janečka · Martin A. Lysak

Received: 23 September 2015 / Revised: 9 December 2015 / Accepted: 11 December 2015 / Published online: 20 January 2016

Abstract Species-specific sets of chromosomes—karyotypes—are traditionally depicted as linear ideograms with individual chromosomes represented by vertical bars. However, linear visualization has its limitations when the shared collinearity and/or chromosomal rearrangements differentiating two or more karyotypes need to be demonstrated. In these instances, circular visualization might provide easier comprehension and interpretation of inter-species chromosomal collinearity. The chromDraw graphical tool was developed as a user-friendly graphical tool for visualizing both linear and circular karyotypes based on the same input data matrix. The output graphics, saved in two different formats (EPS and SVG), can be easily imported to and modified in presentation and image-editing computer programs. The tool is freely distributed under GNU General Public License (GPL) and can be installed from Bioconductor or from the chromDraw home page.

Keywords Visualization · Chromosome · Karyotype · R

Responsible Editor: Hans de Jong

J. Janečka · M. A. Lysak (✉)
CEITEC-Central European Institute of Technology, Masaryk University, Kamenice 5, Brno 62500, Czech Republic
e-mail: martin.lysak@ceitec.muni.cz

J. Janečka
e-mail: jan.janecka@ceitec.muni.cz

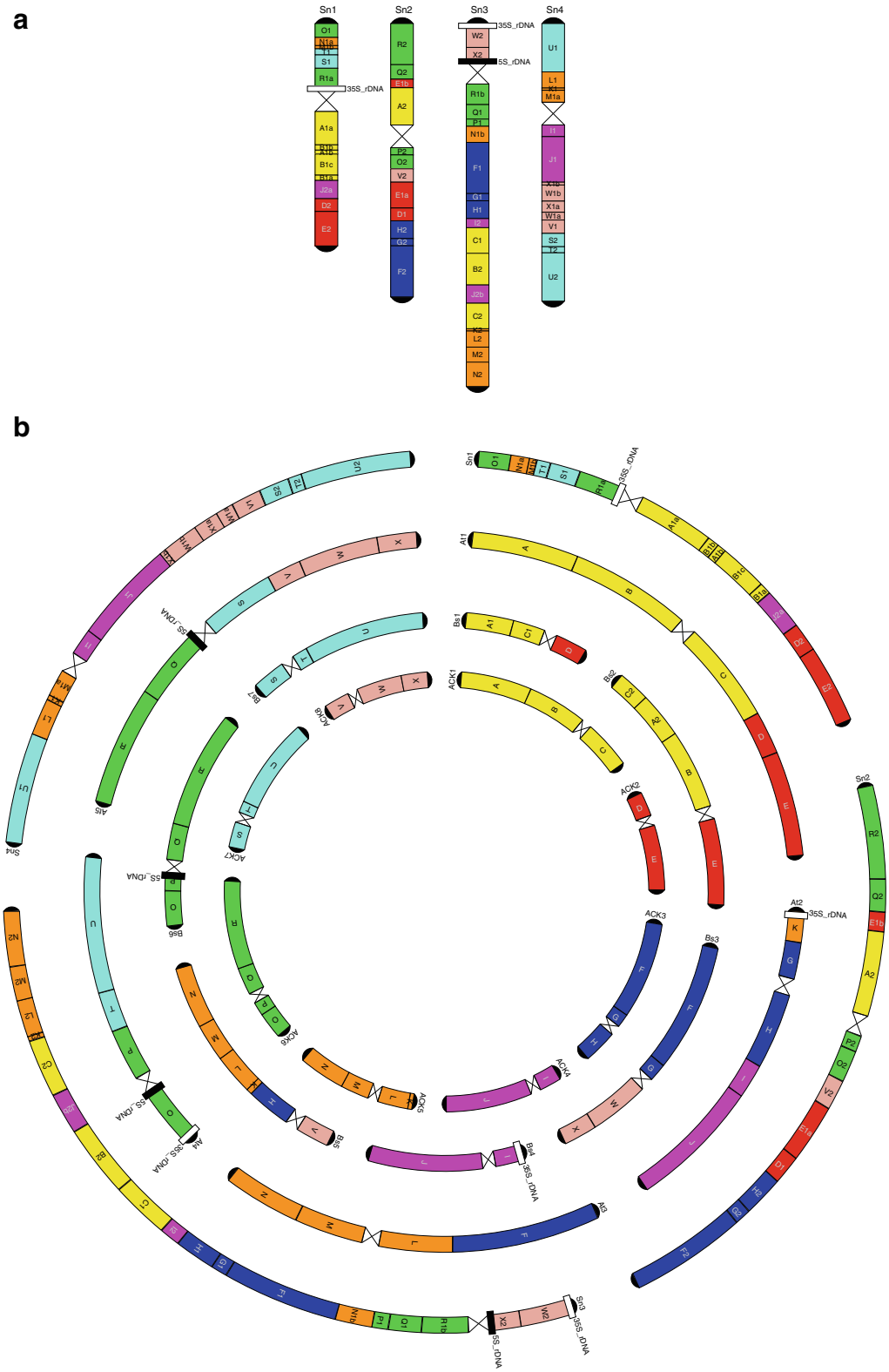
Abbreviations

BED	Browser Extensible Display
C++	Programming language
EPS	Encapsulated PostScript
GPL	General Public License
R	Programming language and software environment
rDNA	Ribosomal DNA
RGB	Red, green, blue
SVG	Scalable Vector Graphic

Introduction

Karyotypes represent chromosome sets that are usually constant for an individual, population, species, or species group. Karyotypes and important chromosomal features were traditionally described by simplified ideograms or idiograms (both terms can be found in the relevant literature and are being used reciprocally) representing chromosomes as linearized vertical bars. Until the onset of computerization and the development of specialized graphical software and hardware, ideograms were drawn by hand. Both the accelerated development of computer-aided graphical tools together with the rapid growth of the Internet and web environments like Java or Flash paved the way for computer-based drawing of ideograms.

Computer-aided karyotype visualizations started to be more common during the late 1980s. A chromosome image analyzing system (CHIAS) has been developed


a
Sn1
Sn2
Sn3
Sn4
35S_rDNA
5S_rDNA
b

particularly to analyze plant chromosomes (Fukui 1986; Fukui and Iijima 1991). CHIAS is primarily an analytical karyotyping tool designed to describe quantitative chromosomal features, such as chromosome length, arm ratio, and chromatin condensation patterns, on Giemsa- or fluorescently stained metaphase chromosomes. Another development in visualization of chromosomes was marked by first published libraries for graphical visualization of chromosomes (Dohi et al. 1993). Later, computerization and the ever-decreasing prices of computers fostered the development of several graphical tools and software. With the Internet becoming more accessible, online graphical tools not requiring installation of special libraries have been employed for karyotype visualization; one of the first client–server applications for visualization of data from public databases was Zomit (Pook et al. 1998). This trend was further underlined by the introduction of Java runtime environment and Flash. Today, several freeware or shareware software for chromosomal and karyotype visualizations are available, for example CrusView (Chen and Wang 2013), ggbio (Yin et al. 2012), GenoPlotR (Guy et al. 2010), Circos (Krzywinski et al. 2009) or DNAnexus (www.dnanexus.com).

Except for the ggbio R package (which however was not designed to visualize ideograms of more than one species), all existing graphical tools were designed for karyotype visualization either in a linear or circular fashion. Whereas linear visualization is usually used to display chromosomal structures and chromosomes within a given karyotype, circular ideograms are preferred for pictorial comparison of two or more karyotypes. Here, we describe principal features of a newly developed graphical tool chromDraw, designed to visualize both linear and circular karyotypes using the same data matrix. As the open Bioconductor software platform (Gentleman et al. 2004), containing a broad range of free packages for bioinformatic analyses and visualization, is widely used by the bioinformatics community, chromDraw was designed as one of the Bioconductor packages.

Results

Design and implementation

BiocCheck (Bioconductor 2014) and BiocInstaller (Tenenbaum and Team 2014) R packages were used during development of the package. The main functionality of chromDraw was written in the C++ language. The library Board (Fourey 2014) was used for drawing graphic primitives. The integration of R and the C++ core of chromDraw was made by the Rcpp package (Eddelbuettel et al. 2011) and allows complete hiding of the C++ implementation for an R user. Because R is a multi-platform software, users can run chromDraw on several platforms such as Linux/UNIX, Windows, or Mac OS. ChromDraw has no requirements for special hardware or software to successfully run the package; R software and required packages are needed (Rcpp, GenomicRanges). A library board has to be installed to compile and run the command line version of this tool in the operating system. ChromDraw can also be run as an online tool accessed from the web page www.plantcytogenomics.org/chromDraw. The online version is limited to simultaneous circular visualization of two karyotypes, based on sample data matrices available from the chromDraw home page. Both the online and locally installed versions have no limitations as to the number of circularly visualized karyotypes when using the user's own matrices; however, there are practical limits to the number of compared genomes (see below). The tool is freely distributed under GNU General Public License (GPL) version 3.

Data representation

In chromDraw, one input file has to be used for visualization of both linear and circular karyotypes. Data describing the structure of individual chromosomes are presented in an open text file format and information on karyotypes is saved in the main data file. Each karyotype contains chromosomal specifications such as acronyms, aliases, and ranges. A chromosome is represented as one or more blocks filling the whole chromosome's

◀ **Fig. 1** **a** chromDraw linear karyotype visualization in *Stenopetalum nutans* (Sn, $n = 4$). The genome contains 24 genomic blocks (A to X; some blocks split into two or three parts) duplicated by a whole-genome duplication event. Data matrix based on Mandáková et al. (2010). **b** chromDraw circular visualization of four ancestral or extant karyotypes from the mustard family (Brassicaceae): *S. nutans* (Sn, $n = 4$), *Arabidopsis thaliana* (At, $n = 5$), *Boechera stricta* (Bs, $n = 7$) and Ancestral Crucifer Karyotype (ACK, $n = 8$). Data matrices are based on Mandáková et al. (2010, 2015) and Schranz et al. (2006), and are available at the chromDraw home page. 5S rDNA and 35S rDNA loci are visualized as *black* and *white rectangles*, respectively

space except for its centromere and telomeres. If the position of the centromere is unknown, or if holokinetic chromosomes should be visualized, this feature can remain undefined. A chromosome block is defined, among other attributes, by its acronym and range. Both chromosomal and block ranges must be represented by integers and must be in the same units. The second file is a simple text file containing definitions of colors, where each color should be described by a unique name and RGB value. However, implemented default colors can be used to successfully run the application. Specific chromosomal features (e.g., ribosomal RNA gene loci, heterochromatic bands or knobs), and landmarks (e.g., localized DNA probes or constructs) can be visualized within both linear and circular karyotypes. Any chromosomal landmark is defined by its position on a particular chromosome as well as by its size, color, and label.

The number of simultaneously visualized karyotypes is not restricted. However, to ensure sufficient legibility of small fonts in circular karyotype visualizations, it is recommended to display no more than six karyotypes with an average of ten chromosomes per presentation slide or A4-sized figure. Input text data file (s) are accepted by the function *chromDraw*, which may be used in R or on the command prompt in locally installed versions.

Data manipulation and functionalities

As chromDraw is one of the Bioconductor R packages, we allowed for the Bioconductor package GenomicRanges (Lawrence et al. 2013) to be used as an alternative input data structure in chromDraw. GenomicRanges provides data structures for representation of annotated genomic ranges, whereby one structure contains information about one karyotype. Processing of this type of data by chromDraw is possible using the function *convertInputData*, converting GenomicRanges structures into the main data file format. In R, colors differentiating individual blocks can be represented as a data frame. The conversion of colors from the data frame to a color data text file is made by the function *convertInputColors*. The function *chromDrawGR* automatically makes both data conversions described above (*convertInputData* and *convertInputColors*). All conversion outputs are saved in the working directory in R.

chromDraw can also visualize data saved in the Browser Extensible Display (BED) file format. The use of the BED data file is possible through the parameter "format", enabling to set the input data format. To store data in the BED format, at least the first nine fields and color definitions for each record have to be used. However, when using the BED files and not violating the principles of this format, only a single karyotype can be visualized.

The output graphics are saved in two formats: Encapsulated PostScript (EPS) and Scalable Vector Graphic (SVG). The first one may be used to create postscript documents or can be imported into slide show presentation programs (e.g., Libre Office Impress or Microsoft Powerpoint), whereas the SVG files can be modified by other graphical tools or converted to other graphical formats.

The chromDraw package is included in Bioconductor R packages (www.bioconductor.org/packages) and may be easily installed. Once installed, chromDraw may be run with the following command:

```
library(chromDraw)
```

The package contains an example that can be executed by the following commands:

```
OUTPATH=file.path(getwd());
INPATH=system.file(
    'extdata',
    'Ack_and_Stenopetalum_nutans.txt',
    package='chromDraw')
COLORPATH=system.file('extdata',
          'default_colors.txt',
          package='chromDraw')
chromDraw(argc=7,
     argv=c("chromDraw",
          "-c",COLORPATH,
          "-d",INPATH,
          "-o",OUTPATH));
```

This example illustrates the use of the package based on text data files. To use the GenomicRanges structures, the following function is used:

```
chromDrawGR(list(karyotype1),
colors);
```

A more detailed description of how to use these functions is available from the chromDraw web page and the documentation file.

Application examples

An example of a linear ChromDraw visualization output for the karyotype of an Australian plant species, *Stenopetalum nutans* ($n=4$), is shown in Fig. 1a. A

circular visualization of the same karyotype, together with three other karyotypes, is shown in Fig. 1b. In these examples, all chromosomal arms are built of so-called genomic blocks (A to X) facilitating inter-genomic comparisons of chromosomal collinearity. An optional functionality to display chromosomal landmarks is demonstrated here by placing rDNA loci on the respective chromosomes. Figure 2 shows a chromDraw circular comparison of human and mouse karyotypes. This example, with the mouse karyotype ($n = 19 + XY$) shown inside the human karyotype ($n = 22 + XY$), illustrates different chromosome numbers and rearrangements of genomic blocks differentiating both species.

Fig. 2 chromDraw circular visualization of mouse (Mm, $n = 19 + XY$) and human (Hs, $n = 22 + XY$) karyotypes. Colors correspond to 21 mouse chromosomes; chromosomal regions that are collinear among the two genomes are labeled with the same color. Data matrices are based on data obtained from CoGe (www.genomevolution.org/coge/, Lyons and Freeling 2008)

Discussion

Here we described the principal functionalities of an R package, chromDraw, developed for linear and circular visualizations of eukaryotic karyotypes using one input data matrix. ChromDraw's features enable drawing of linear ideograms that can then be aligned and compared with each other in the form of circular diagrams displaying inter-species chromosomal collinearity and structural differences. Additionally, the developed graphical tool enables to plot various chromosome-specific landmarks and DNA probes on selected chromosomes. Output vector graphics are versatile, can be easily modified in a number of graphical software, and the postscript files can be readily inserted into slide show presentation programs such as Libre Office Impress or Microsoft PowerPoint. As R is a free multi-platform software, the chromDraw package can be run on several platforms and its core can also be run separately like a command line application, or through the online web-based version. There are a number of directions in which the package might be extended in the future. The online available database of karyotypes will be constantly updated and useful functionalities and settings of karyotype visualization will be implemented. Special requirements and features can be implemented by any user after downloading the source code. However, all changes to the published version can be made only by the chromDraw developer upon request.

There are several other freely available software for linear or circular visualization of karyotypes. However, all the software and tools are limited to either linear or circular types of visualization. GenoPlotR (Guy et al. 2010) is an R package plotting linear gene and genome maps. This package is suited for comparison of two chromosomal regions or chromosomes in a linear fashion. If multi-karyotype comparisons are required, circular visualizations drawn by chromDraw or Circos are more informative. The widely used Circos software (Krzywinski et al. 2009) was developed for visualizing data in a circular layout. Although Circos provides publication-quality circular visualizations, linear karyotypes cannot be drawn using the same data matrix. ChromDraw, in contrast, generates both types of visualization using one input data matrix. Both types of visualization can be also produced using an R package ggbio (Yin et al. 2012). However, this package was primarily designed to visualize the detailed structure of genomic regions (e.g., splicing patterns, exons vs. introns), whereas chromDraw is visualizing large-scale chromosomal regions, chromosomes, key chromosomal features, and multiple complete karyotypes.

Acknowledgments We thank Dr Terezie Mandáková and Petra Hloušková for helpful suggestions and comments on the functionality of the chromDraw package. Drs Matej Lexa and Jiří Hon are acknowledged for ideas on how to expand the package and for introduction to the R world, respectively. This research was supported by a grant from the Czech Science Foundation (P501/12/G090) and by European Social Fund (CZ.1.07/2.3.00/20.0189).

References

Bioconductor PM (2014) BiocCheck: Bioconductor-specific package checks. [Software]. R package version 1.1.9

Chen H, Wang X (2013) CrusView: a java-based visualization platform for comparative genomics analyses in brassicaceae species. Plant Physiol 163:354–362

Dohi H, Ishizuka M, Minoshima S, Shimizu N (1993) GeneView: multi-language human gene mapping library with a graphical user interface. CABIOS 9:459–464

Eddelbuettel D, François R, Allaire J, Chambers J, Bates D, Ushey K (2011) Rcpp: Seamless R and C++ integration. J Stat Softw 40:1–18

Fourey S (2014) LibBoard: a vector graphics C++ library (Version 0.9.0). [Software]. GREYC Laboratory

Fukui K (1986) Standardization of karyotyping plant chromosomes by a newly developed chromosome image analyzing system (CHIAS). Theor Appl Genet 72:27–32

Fukui K, Iijima K (1991) Somatic chromosome map of rice by imaging methods. Theor Appl Genet 81:589–596

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80

Guy L, Kultima JR, Andersson SG (2010) genoPlotR: comparative gene and genome visualization in R. Bioinformatics 26: 2334–2335

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Care VJ (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9, e1003118

Lyons E, Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J 53:661–673

Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA (2010) Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. Plant Cell Online 22:2277–2290

Mandáková T, Schranz ME, Sharbel TF, Jong H, Lysak MA (2015) Karyotype evolution in apomictic Boechera and the origin of the aberrant chromosomes. Plant J 82: 785–793

Pook S, Vaysseix G, Barillot E (1998) Zomit: biological data visualization and browsing. Bioinformatics 14: 807–814

Schranz ME, Lysak MA, Mitchell-Olds T (2006) The ABC's of comparative genomics in the Brassicaceae: building blocks of crucifer genomes. Trends Plant Sci 11:535–542

Tenenbaum D, Team B (2014) BiocInstaller: Install/Update Bioconductor and CRAN Packages. [Software]. R package version 1.15.5

Yin T, Cook D, Lawrence M (2012) ggbio: an R package for extending the grammar of graphics for genomic data. Genome Biology 13: R77