

Package ‘nplplot’

February 20, 2015

Version 4.5

Date 2014-04-30

Title Plotting linkage and association results

Author Robert V Baron <rvb5@pitt.edu>, Nandita Mukhopadhyay,
Xinyu Tang, Daniel E. Weeks <weeks@pitt.edu>

Maintainer Daniel E. Weeks <weeks@pitt.edu>

Description This package provides routines for plotting linkage and association results along a chromosome, with marker names displayed along the top border. There are also routines for generating BED and BedGraph custom tracks for viewing in the UCSC genome browser. The data reformatting program Mega2 uses this package to plot output from a variety of programs.

License GPL (>= 3)

URL <http://watson.hgen.pitt.edu/register>

Repository CRAN

Date/Publication 2014-05-01 09:32:57

NeedsCompilation no

R topics documented:

| | |
|-------------------------|----|
| bedplot | 2 |
| genomeplot | 3 |
| lods1 | 4 |
| lods2 | 4 |
| nplplot | 5 |
| nplplot.multi | 7 |
| nplplot.old | 9 |
| prepareplot | 12 |

| | |
|--------------|-----------|
| Index | 14 |
|--------------|-----------|

`bedplot`*Creation of BED and BedGraph custom tracks*

Description

Generates matched sets of files for linkage or association statistics along a chromosome for viewing in the UCSC genome browser from an input file containing a table of marker names, physical positions and one or more statistical scores.

Usage

```
bedplot(bed.data)
```

Arguments

`bed.data` File containing a table of marker names, physical position and scores.

Details

bed.data example:

| Marker | Position | TRAIT_ALL |
|--------|----------|-----------|
| M1 | 144255 | 0.670 |
| - | 144305 | 0.640 |
| M3 | 144355 | 0.590 |
| - | 144378 | 0.600 |
| M2 | 144400 | 0.610 |

Bedplot creates two types of files: a BED.* file containing a custom BED annotation track and a BedGraph.* file custom BedGraph annotation track. These files have the same suffix as the input bed.data file. When there are multiple scores in bed.data file, a matched pair of BedGraph track BED track files is created for each score, labelled with the score names, as well as the chromosome numbers, e.g. BedGraph.score1.* and BED.score1.*, BedGraph.score2.* and BED.score2.*, etc.

Value

TRUE or FALSE depending on whether runs successfully.

Examples

```
## Not run: bedplot("bed.data.05")
```

`genomeplot`*Creation of Genome Graph files*

Description

The `genomeplot` function generates two formatted files, one containing “chromosome base” formatted genome data and the other containing marker-specific results with dbSNP SNP IDs for displaying genome-wide data sets in the UCSC genome browser.

Usage

```
genomeplot(gg.data)
```

Arguments

`gg.data` a file containing chromosome, marker, physical position and scores.

Details

`gg.data` example:

| Chromosome | Marker | Position | TRAIT_ALL |
|------------|--------|----------|-----------|
| 5 | M1 | 0.000 | 0.670 |
| 5 | - | 2.500 | 0.640 |
| 5 | M3 | 5.000 | 0.590 |
| 5 | - | 6.500 | 0.600 |
| 5 | M2 | 8.000 | 0.610 |
| 8 | M4 | 0.000 | 0.670 |
| 8 | - | 2.500 | 0.640 |
| 8 | M6 | 5.000 | 0.590 |
| 8 | - | 6.500 | 0.600 |
| 8 | M5 | 8.000 | 0.610 |

Two files are created, “GG.positions.all” for the “chromosome base” format, and “GG.markers.all” for the marker-names based format. When there are multiple scores in `gg.data` file, this results in matched pairs of files, one for each score, labelled with the score names, e.g. `GG.positions.score1.all`, and `GG.markers.score1.all`, `GG.positions.score2.all` and `GG.markers.score2.all`, and so on.

Value

TRUE or FALSE depending on whether runs successfully.

Examples

```
## Not run: genomeplot("GG.data.all")
```

| | |
|-------|---|
| lods1 | <i>LOD score table for chromosome 1</i> |
|-------|---|

Description

This is a data frame with the first two columns containing marker names and positions, followed by three columns of LOD scores.

Usage

```
data(lods1)
```

Format

There are 100 markers in the table.

| | |
|-------|---|
| lods2 | <i>LOD score table for chromosome 2</i> |
|-------|---|

Description

This is a data frame with the first two columns containing marker names and positions, followed by three columns of LOD scores.

Usage

```
data(lods2)
```

Format

There are 87 markers in the table.

Description

Plots linkage or association statistics along a chromosome, contained within a data frame or a file. Marker names are displayed along the top border.

Usage

```
npplot(plotdata=NULL, filename=NULL, yline=2.0, ymin=0, ymax=3.0,
       header=TRUE, yfix=FALSE, title=NULL, draw.lgnd=TRUE,
       xlabel="", ylabel="", lgndx=NULL, lgndy=NULL, lgndtxt=NULL,
       cex.legend = 0.7, cex.axis=0.7, tcl=1,
       bw=TRUE, my.colors=NULL, ltypes=NULL, ptypes=NULL,
       na.rm=TRUE, plot.width=0.0, ...)
```

Arguments

| | |
|------------|---|
| plotdata | A data frame containing marker names in the first column, marker map positions in the second column, and statistical scores in column 3 onwards. |
| filename | A table format file containing the plot data as described above. |
| header | TRUE or FALSE depending on whether the plotdata or file has a header line. |
| ylines | Y-value for displaying a horizontal cut-off line. If 'yfix' is set to TRUE and Y-line falls outside of [ymin, ymax], then the cut-off line is omitted. |
| ymin, ymax | Y-axis minimum and maximum values. If non-NULL values are provided, and yfix is set to TRUE, then the plot area will be cropped to these values. If yfix is set to FALSE, then ymin and ymax values are ignored. |
| yfix | TRUE or FALSE to denote whether plot area should be cropped to the ymin, ymax values. This has no effect if ymin, ymax values are NULL. |
| title | Used as the subtitle of the plot. |
| xlabel | X-axis label. May interfere with the display of the subtitle provided as the title argument. |
| ylabel | Y-axis label. |
| draw.lgnd | TRUE or FALSE denoting whether a plot legend should be displayed. |
| lgndx | X coordinate for the legend box, passed to the legend command. Ignored if draw.legend is set to FALSE. If set to NULL with draw.legend set to TRUE, the X-coordinate is automatically calculated. |
| lgndy | Y coordinate for the legend box, passed to the legend command. Ignored if draw.legend is set to FALSE. If set to NULL with draw.legend set to TRUE, the Y-coordinate is automatically calculated. |
| lgndtxt | Vector of strings to use in the legend. |

| | |
|-------------------------|--|
| <code>cex.legend</code> | Character scaling for legend, passed as the <code>cex</code> argument to the <code>legend</code> command. |
| <code>cex.axis</code> | Character scaling for the axis, passed to the <code>axis</code> command for drawing the top border. |
| <code>tcl</code> | Length of ticks for the top border, passed to the <code>axis</code> command. |
| <code>bw</code> | TRUE or FALSE depending on whether plots should be drawn in color. If set to FALSE, then the colors defined by <code>my.colors</code> are used. |
| <code>my.colors</code> | Vector of color specifications as described in the <code>par</code> command. Ignored if <code>bw</code> above is set to FALSE. If <code>bw</code> is to TRUE and <code>my.colors</code> is set to NULL, the <code>rainbow</code> palette will be used. |
| <code>ltypes</code> | Vector of line types for the plots. Each non-zero line type is passed on to a <code>lines</code> command. Use 0 or 'none' if a line is to be skipped. If NULL, no lines will be drawn. For line types see the <code>par</code> command. If set to "default", line-types 1 through the number of plots is used. |
| <code>ptypes</code> | Vector of characters giving the point types, to be passed onto the <code>points</code> command. Use 'none' if no points are to be drawn for a score column. If NULL, no points will be displayed. If both the line-type and point-type specification for a results column is set to 'none', that column will not be plotted. |
| <code>na.rm</code> | TRUE or FALSE depending on whether points with Y-coordinates set to NAs should be skipped. Setting <code>na.rm</code> to TRUE eliminates discontinuities in the plots. |
| <code>plot.width</code> | A number giving the width of the plot in inches. This is used to decide whether some marker names should be dynamically hidden, if they are too close to each other along the top border. If set to 0, the default page-size is used to set the width. |
| <code>...</code> | Further graphical parameters to be passed onto the 'plot', 'lines' and 'points' commands. |

Details

The `nplplot` function draws multiple curves within a single plot by automatically calling 'plot', 'lines', and 'points' multiple times, thus making it easy for the user to plot many columns of results using a single plot command. It is intended for the display of linkage and association analysis results such as LOD scores and P-values. It allows the marker names to be displayed along the top border of the plot, as well as a significance threshold line.

The input plot data has to be in a specific tabular format with each column separated by white-space:

Here is an example:

| Marker | Position | score1 | score2 | score3 |
|--------|----------|--------|--------|--------|
| d1s228 | 0.00 | 0.546 | 0.345 | 0.142 |
| d1s429 | 1.00 | 0.346 | 0.335 | 0.252 |
| d1s347 | 2.00 | 0.446 | 0.245 | 0.342 |

This example file contains a header, therefore the header argument should be set to TRUE.

Lines 2-4 contain scores at various marker positions. **Missing scores can be denoted with either "." or "NA"**. The position column cannot have missing data. There can be any number of score columns within a file and will be plotted as separate curves within the same plot. Each file is plotted as a separate plot.

Value

TRUE or FALSE depending on whether the plot data was successfully plotted.

See Also

[nplplot.multi](#), [nplplot.old](#)

Examples

```
# plot with legend
par(omi=c(0.05, 0.05, 0.5, 0.05))
data(lods1, package="nplplot")
nplplot(plotdata=lods1, draw.lgnd=TRUE)

# plot without legend
data(lods2, package="nplplot")
nplplot(plotdata=lods2, draw.lgnd=FALSE)

# plotting from a data file
datadir <- paste(system.file("data", package="nplplot"), .Platform$file.sep, sep="")
nplplot(filename=paste(datadir, "lods2.txt.gz", sep=""))
```

nplplot.multi

Plotting linkage or association statistics for multiple results files

Description

Wrapper function for the 'nplplot' function. Creates mutiple plots from a list of plot files, with custom graphical parameters set by header files.

Usage

```
nplplot.multi(filenamees, plotdata = NULL, col=2, row=2, mode="1",
              output="screen", headerfiles=NULL, lgnd="page",
              customtracks=FALSE, mega2mapfile=NULL,
              pagewidth=NULL, pageheight=NULL, topmargin=0.25,
              ...)
```

Arguments

| | |
|--------------|---|
| filenames | Vector of strings giving file names containing tables of linkage analysis results. See nplplot for a description of the file format. |
| plotdata | List of dataframes by chromosome containing tables of linkage analysis results. See nplplot for a description of the format. |
| col | Integer indicating number of columns of plots to be drawn on a page. |
| row | Integer signifying number of rows of plots will be drawn on a page. |
| mode | 'p' or 'l' to denote 'portrait' or 'landscape' mode. |
| output | String giving file name to save plots in. If set to 'screen', plots will be displayed and not saved. The file format is determined by the filename extension: '.pdf' for PDF, or '.ps' for postscript. If no extension is provided, or is not recognized, a PDF file will be produced with '.pdf' appended to the file name. |
| headerfiles | Files containing R language commands to set various plot parameters, which are passed onto the nplplot command. The recommended use is to have one headerfile per plot file. For a list of parameters, consult the nplplot documentation. If the number of headerfiles is fewer than plot files, the last header file will be reused as many times as needed. If more headerfiles are provided than necessary, the last ones will be ignored. |
| lgnd | TRUE, FALSE, 'page' or a list consisting of plot numbers. If a single value is given, TRUE causes legends to be drawn inside every plot, FALSE omits legends altogether, and 'page' causes a legend to be drawn inside the first plot on every page. If a list of numbers is provided, only plots corresponding to these numbers will have legends. |
| customtracks | TRUE or FALSE. If set to TRUE, data files are created to draw custom tracks within the UCSC genome browser in BED format, as well as a combined data file to add a genome-wide track over all chromosomes present in the data. If set to TRUE, a mega2mapfile also needs to be supplied (see below). |
| mega2mapfile | Mega2 annotated format map file containing physical positions for all the markers present in the nplplot input data files. Rather than a file name, the name of a <i>data.frame</i> containing what would have been read from the file, may be given. |
| pagewidth | A number denoting width of the plot page in inches. If set to NULL, a width of 7.0 is used for the plot area. Assumes that a margin of 0.5 will be available around the plot area for axis annotations. |
| pageheight | A number denoting height of the plot page in inches. If set to NULL, a height of 10.0 is used for the plot area. Assumes that a margin of 0.5 will be available around the plot area for axis annotations. |
| topmargin | A number denoting the width of the outside top margin of each plot. Since this contains marker names, it may need to be increased to accommodate long names. |
| ... | Further graphical parameters to be passed onto the 'plot', 'lines' and 'points' commands within nplplot. |

Details

This function is designed for use within the Mega2 software to generate graphical output for some of the target analysis options, namely Merlin, SimWalk2 and Allegro. It calls `nplplot` repeatedly to create plots corresponding to each input file. The input arguments control characteristics of all plots together, whereas the header files allow customization within each plot. Thus, it is expected that there should be as many header files as there are plot data files.

This function can also be used to create custom tracks within the UCSC genome browser, as well as a genome-wide plot. To use this feature, make sure that the names of the `nplplot` input data files each have a "Mega2-style" chromosome extension (01 through 09 for chromosomes 1 through 9, otherwise the chromosome number, or X for the human X-chromosome, 23).

To make this function more useful to other **R** programs, you may directly supply a `data.frame` for the `mapfile` argument and a list of `data.frames` for the `plotdata` argument and `NULL` for the `filename` argument. (The name of each list element is the corresponding chromosome.)

Value

TRUE or FALSE depending on whether all plot commands were successful.

See Also

`nplplot`, `nplplot.old`

Examples

```
datadir <- paste(system.file("data", package="nplplot"),
                .Platform$file.sep, sep="")
f1 <- paste(datadir, "lods1.txt.gz", sep="")
f2 <- paste(datadir, "lods2.txt.gz", sep="")
h1 <- system.file("extdata", "lods1header.R", package="nplplot")
h2 <- system.file("extdata", "lods2header.R", package="nplplot")
nplplot.multi(c(f1, f2), col=1, row=2, output="screen",
             headerfiles=c(h1, h2), topmargin=0.5)
```

`nplplot.old`

LOD score plotting (old version of nplplot)

Description

Plots score curves contained within one or more specified results files.

Usage

```
nplplot.old(files, col=2, row=2, mode="p", output="screen",
            yline=2.0, ymin=NULL, ymax=NULL, yfix=FALSE, batch=FALSE,
            headerfiles=NULL, titles=NULL, xlabel="", ylabel="",
            lgnd="page", lgndx=NULL, lgndy=NULL, bw=TRUE, na.rm=TRUE)
```

Arguments

| | |
|-------------|---|
| files | List of files names (strings). Each file produces a separate plot. |
| col | For multiple plots on a single page of pdf or postscript output, this item defines the number of columns of plots, and should be an integer greater than or equal to 1. Default is set to 2. |
| row | For multiple plots on a page of pdf or postscript output, this defines the number of rows of plots, (value should be 1 or greater). Default value is set to 2. |
| mode | Orientation for pdf or postscript output, "p" for portrait "l" for landscape. |
| output | File name for saving plots; "screen", the default causes the plots to be displayed on the screen. To produce a pdf file use the extension .pdf. To produce a postscript file, use the .ps file name extension. If no extension is given a pdf file is produced. |
| ylines | Y-value for displaying a horizontal cut-off line. |
| ymin, ymax | Y-axis minimum and maximum values with default values NULL. If non-NULL values are provided, and yfix is set to TRUE, then the plot area will be cropped to these values. |
| yfix | Set to TRUE or FALSE depending on whether ymin and ymax should be enforced across all plots irrespective of whether the plot data lie within these bounds. Ignored if ymin or ymax are set to NULL. |
| batch | TRUE or FALSE, to determine whether the display screen should be closed. If nplplot is called within R, this should be set to FALSE. |
| headerfiles | List of file names, one for each data file specified above. Each header-file contains a string with column names corresponding to the columns in the data file. These column names are used in the plot legend. If set to NULL (the default), nplplot uses the first item in each column of a data file as plot legend. If a headerfile is provided, then nplplot will attempt to read in the first line of the datafile as data, so the user should be careful not to put in a headerline as well as a headerfile. |
| titles | Array of strings denoting titles for each plot. If there are not enough titles, the last string is recycled for the remaining plots. Default is an empty string. |
| xlabl | Array of strings, to use as the x-axis label on each plot. |
| ylabl | Array of strings to use as the y-axis label on each plot. |
| lgnd | TRUE, FALSE, "page" or a list of plot numbers denoting whether the legend should be drawn in all plots, none, first plot on a page, or specific plot numbers. Default "page". |
| lgndx | NULL or a real value if a specific x-coordinate should be used to position the legend. Default NULL. |
| lgndy | NULL or a real value if a specific y-coordinate should be used to position the legend. Default NULL. |
| bw | TRUE or FALSE depending on whether plots should be drawn in color. A list of six colors are defined within nplplot, which are successively used to draw each curve, and reused as necessary. The order in which these colors are used is: magenta, lightblue, grey, navyblue, lightcyan and pink. The 7th color, reserved for black and white plots is black. |

na.rm TRUE or FALSE depending on whether NAs should be removed prior to plotting the data. Including NAs will produce broken plots, when lines are drawn. This may be desirable in some cases, if missing data needs to be reported.

Details

Usually these results would be LOD scores, p-values, or $\log_{10}(\text{p-values})$. This is targeted towards p-values or LOD scores obtained at various marker positions from statistical analysis of genetic data. A results file has to be in a specific tabular format with each column separated by white-space :

- A) First line = header line
- B) Next set of lines = any number of data lines
- C) Final two lines = line type & point type definition.

Here is an example:

| marker | location | score1 | score2 | score3 |
|--------|----------|--------|--------|--------|
| d1s228 | 0.00 | 0.546 | 0.345 | 0.142 |
| d1s429 | 1.00 | 0.346 | 0.335 | 0.252 |
| d1s347 | 2.00 | 0.446 | 0.245 | 0.342 |
| ltype | -99.99 | 1 | 2 | 3 |
| ptype | -99.99 | 15 | 16 | 17 |

In this example, line 1 column headers for the score columns may be used as labels within the legend, as described in the usage of the "headerfile" argument. **The first two headers are ignored.**

Lines 2-4 contain scores at various marker positions. **Missing scores can be denoted with either "." or "NA"**. The position column cannot have missing data. There can be any number of score columns within a file and will be plotted as separate curves within the same plot. Each file is plotted as a separate plot.

The last two lines give line types and point types for each curve. A zero line or point type will not plot lines or points for that score column respectively. For allowable ptype values, consult the R documentation for "points". For line types, consult the documentation on "par".

The names in the first column are used as axis labels on the top of the plot border. Setting a name in the **marker** column to "-" will result in no label at that position.

Value

TRUE or FALSE depending on whether the input files were read in successfully.

See Also

[nplplot](#), [nplplot.multi](#)

Examples

```
## Not run: nplplot.old("lod.1", output="lod.1.ps", batch=T, headerfiles="hdr.1")
## Not run: nplplot.old(c("lod.1", "lod.2"), col=1, row=2, headerfiles=c("hdr.1", "hdr.2"))
```

```
prepareplot          Prepare input data files for bedplot and genomeplot
```

Description

The prepareplot function prepares input data files for [bedplot](#) and [genomeplot](#) functions from nplplot-formatted score files and a Mega2 annotated format map file with physical positions.

Usage

```
prepareplot(prefix, chrlist=c(1:23,25), mapfile, output="both")
```

Arguments

| | |
|---------|---|
| prefix | Prefix of the names of R table files, e.g. "RMERLINDATA" for R table files "RMERLINDATA.01", "RMERLINDATA.02", etc. Using chrlist below, it automatically finds R table files with the specified prefix and chromosome-specific extensions to convert. Alternatively, prefix may be a list of <i>data.frames</i> named by the chromosomes supplied in chrlist. |
| chrlist | List of chromosome numbers to create plots for, default 1 through 23. Chromosomes 23 and 25 produces files for the X chromosome X, 25 denoting pseudo-autosomal markers on chromosome X. |
| mapfile | Mega2 annotated format map file, containing marker names and exactly one set of physical positions. mapfile may instead be a <i>data.frame</i> containing the same information as the map file, viz. the marker names and physical positions. |
| output | Which plotting function to generate data for, "both" for both bedplot and genomeplot functions, "bed" for generating input files for bedplot function, "GG" for generating input file for genomeplot function. output is set to default "both". |

Details

mapfile example:

| Chromosome | Map.h.a | Name | Map.h.m | Map.h.f | Build52.p |
|------------|---------|------|---------|---------|-----------|
| 5 | 0.0 | M1 | 0.0 | 0.0 | 144255 |
| 5 | 5.0 | M3 | 2.0 | 7.0 | 144355 |
| 5 | 8.0 | M2 | 4.0 | 12.0 | 144400 |
| 8 | 0.0 | M4 | 0.0 | 0.0 | 144255 |
| 8 | 5.0 | M6 | 2.0 | 7.0 | 144355 |
| 8 | 8.0 | M5 | 4.0 | 12.0 | 144400 |

The names of R table files should be linkage or association analysis score files in nplplot-format with Mega2-style file names, i.e., having a common specified prefix and 01-09, 11-24, X, or XY as suffixes. The list of suffixes are determined by the chromosome list. If this list includes 23 or X, R

table files with either the “23” suffix or “X” suffix are accepted. If both files exist, the one with the “X” suffix is read in and the user warned. If the XY chromosome is chosen, R table files can have either “24” or “XY” as a suffix, with “XY” suffixed file having precedence.

The prepareplot function generates chromosome-specific formatted score files “bed.data.\#” for use by [bedplot](#) with the same suffix as the R table file. If X chromosome is chosen, the output file is named “bed.data.23”. If XY chromosome is chosen, those records on XY chromosome are included in “bed.data.23” file. The output file “bed.data.\#” contains marker names and physical positions followed by one or more score columns. The header is taken from the input score file(s).

Prepareplot generates a combined file over all chromosomes “GG.data.all” for [genomeplot](#). For pseudo-autosomal markers denoted by chromosome XY or 24, these scores are assigned the X chromosome. The output file “GG.data.all” contains four or more columns with headings. The first, second and third columns contain chromosomes, marker names and physical positions respectively, followed by one or more score columns with score names as headers.

Value

TRUE or FALSE depending on whether runs successfully.

Examples

```
## Not run: prepareplot("RMERLINDATA", c(5,8), "map.all", "GG")
```

Index

*Topic **datasets**

lods1, [4](#)

lods2, [4](#)

*Topic **plotting**

nplplot, [5](#)

nplplot.multi, [7](#)

nplplot.old, [9](#)

*Topic **utilities**

bedplot, [2](#)

genomeplot, [3](#)

prepareplot, [12](#)

axis, [6](#)

bedplot, [2](#), [12](#), [13](#)

genomeplot, [3](#), [12](#), [13](#)

legend, [5](#), [6](#)

lods1, [4](#)

lods2, [4](#)

nplplot, [5](#), [8](#), [9](#), [11](#)

nplplot.multi, [7](#), [7](#), [11](#)

nplplot.old, [7](#), [9](#), [9](#)

par, [6](#)

prepareplot, [12](#)

rainbow, [6](#)