genomePlots

October 15, 2018

Version: 2v1

Title: genomePlots

Description: Plots genomic coordinate-based data including bigBed, bed, bigWig and bedGraph for defined genomic loci.

Author: Benjamin Barwick [cre,aut], Christopher Scharer [aut, ctb], Tian Mi [ctb]

Depends R (v3.4.3)

Imports: stats4, parallel, stats, graphics, grDevices, utils, datasets, methods, base, BSgenome, rtracklayer, GenomicRanges, GenomeInfoDb, Biostrings, XVector, IRanges, S4Vectors, BiocGenerics

Suggests: XML, Rsamtools, GenomicAlignments, bitops, futile.options, zlibbioc, futile.logger, lambda.r, BiocParallel, tools, Biobase, RCurl, SummarizedExperiment

Git url: https://github.com/cdschar/genomePlots

genomePlots Usage:

        All four files are required to run the script and generate a plot. Additionally, the BSgenome package for the genome of interest must be installed and will be loaded during plotting.

Files:

refFlat2bed.R – Converts a UCSC refFlat file containing gene annotation data into a format that can be used as a gene track in genomePlot.R

genomePlot.R – Driver script that generates the plots.

Arguments

1. **output format** – options include: pdf, screen, svg
2. **species** – set BSgenome name for species you are plotting data for. Human and mouse options are included.
3. **plot.pattern** – TRUE/FALSE if the specified genomic sequence should be annotated
4. **pattern** – Genomic sequence to find and annotate. For example, this can be CG for DNA methylation or a restriction site.
5. **plot options** – control the size of some of the plot features

regions.txt – Each line is a list of regions to plot with customizable options to skip regions or invert the direction of regions so that transcription is always left to right for example .

Column description and functions

1. **name** – Name for the region tha will be used as the output file.
2. **plot** – TRUE/FALSE to determine if region should be plotted.
3. **chr** – chromosome name for region to plot.
4. **start** – Starting chromosome coordinate for region to plot.
5. **end** – End chromosome coordinate for region to plot.
6. **plot.fwd** – TRUE/FALSE if the regions should be plotted in reverse.
7. **notes** – Not used but can be used to store information about each region.

tracks.txt – List of data tracks to plot. The columns control the order of plotting and the color of different plots. Examples of how to overlay plots are shown. There are options to plot bed objects with different pch points. Big files can be locally on your computer or hosted on a server as the examples show here.

Column description and functions

1. **track** – General name for a track. If two tracks share the same name it will overlap them with the order column which one is plot first.
2. **name** – Track name

3. **type** – Type of data to plot: bed, bed15, bigWig, bedGraph
4. **height** – Vertical height of track.
5. **show** – TRUE/FALSE to determine if data is plotted.
6. **order** – Number that is unique for each track and determines the order from top to bottom.
7. **min** – For bigWig and bedGraph data the minimum value on y-axis. NA allowed.
8. **max** – For bigWig and bedGraph data the maximum value on y-axis. NA allowed.
9. **bed.height** – Height of bed rectangular objects.
10. **block.height** – Secondary, bed height value for exons or other bed15 data.
11. **bed.name** – TRUE/FALSE if bed object names should be plotted. NA allowed.
12. **bed.arrow** – TRUE/FALSE if bed object strand should be annotated with an arrow.
13. **contains.color** – TRUE/FALSE for if track should be plotted in color.
14. **plot.type** – Type of plot: polygon (traditional for bigWig), line, bar, points.
15. **smooth** – TRUE/FALSE if lowess smoothing should be used on data.
16. **ptsPch** – For point plots the pch code controlling the type of character to plot.
17. **ptsCex** – For point plots the cex numeric code controlling the size of character to plot.
18. **color** – Fill color for track. RGB color code supported.
19. **border** – Outline color for track. RGB color code supported.
20. **line.lwd** - For line plots controls the thickness of line to plot.
21. **line.lty** - For line plots controls the type of line to plot.
22. **legend** – TRUE/FALSE if legend should be annotated for each track to the left of each track outside the scale.
23. **label.track** - TRUE/FALSE if legend should be annotated for each track inside the scale above the data.
24. **label.size** – Cex numerical value to control the size of label text.
25. **label.orientation** – Controls if the label to the left of the track is vertical (1) or horizontal (0).
26. **label.line** – Controls if a line should be included after the label name for the above track label.
27. **label.adj** – Numerical value to adjust the label.
28. **label.name** – Name that will be used to label the track.
29. **file** – The location on the local computer of the file to be plotted. For 'big' files the data can be hosted on a website.

Plot.region.lib.R – R library containing the plotting functions.