Package 'RWebLogo'

August 29, 2016

Type Package

Title plotting custom sequence logos

Version 1.0.3

Date 2014-04-14

Author Omar Wagih

Maintainer Omar Wagih <wagih@ebi.ac.uk>

Description RWebLogo is a wrapper for the WebLogo python package
that allows generating of customised sequence logos. Sequence logos are
graphical representations of the sequence conservation of nucleotides (in a
strand of DNA/RNA) or amino acids (in protein sequences). Each logo
consists of stacks of symbols, one stack for each position in the sequence.
The overall height of the stack indicates the sequence conservation at that
position, while the height of symbols within the stack indicates the
relative frequency of each amino or nucleic acid at that position. In
general, a sequence logo provides a richer and more precise description of,
for example, a binding site, than would a consensus sequence.

Imports findpython

License BSD_3_clause + file LICENSE

SystemRequirements Python (>=2.6) in path and NumPy module installed,
ghostscript (>=9.0)

NeedsCompilation no

Repository CRAN

Date/Publication 2014-08-10 23:50:59

R topics documented:

| weblogo | *Plot a sequence logo* |

Description

This function will plot a sequence logo given aligned sequences. For more details on the parameters, see the WebLogo user manual at <http://weblogo.threeplusone.com/manual.html>

Usage

```
weblogo(seqs, file.in, open = TRUE, verbose = TRUE, return.cmd = F,
  datatype = "plain", file.out, format = "pdf", sequence.type = "protein",
  alphabet, units = "bits", composition = "auto", weight, first.index,
  lower, upper, ignore.lower.case = TRUE, reverse, complement,
  size = "large", stacks.per.line = 40, title, label, show.xaxis = TRUE,
  xlabel, annotate, yaxis, show.yaxis = TRUE, ylabel, show.ends = FALSE,
  fineprint = "", ticmarks = 1, errorbars = FALSE,
  reverse.stacks = TRUE, color.scheme = "auto", stack.width = 10.8,
  aspect.ratio = 5, box = FALSE, resolution = 300, scale.width = TRUE,
  rotate.numbers = FALSE, hide.tics = FALSE)
```

Arguments

seqs	Aligned sequences as an R character vector. Sequences must all have the same length. Alternatively, you can provide a file containing your sequence alignment using file.in.
file.in	A file containing your sequence alignment in one of the following formats: clustal, fasta, plain, msf, genbank, nbrf, nexus, phylip, stockholm, intelligenetics, table, array, transfac. This file is only to be provided if you are not inputting data with 'seqs'. To set your data format, see the datatype option.
open	Open the generated logo file? (default: TRUE).
verbose	Write status messages to the R console? (default: TRUE).
return.cmd	Logical indicating if RWebLogo should return the WebLogo command generated (default: FALSE).
datatype	Type of multiple sequence alignment or position weight matrix file: ('clustal', 'fasta','plain', 'msf', 'genbank', 'nbrf', 'nexus', 'phylip', 'stockholm', 'intelligenetics', 'table', 'array', 'transfac'). You usually don't need to specify this, as weblogo will try figure out the format of your file.
file.out	Output file. For example, /path/to/dir/mylogo.pdf. By default this is your working directory + RWebLogo. + selected format.
format	Format of output: 'pdf' (default) 'eps', 'png', 'jpeg', 'svg'.
sequence.type	The type of sequence data: 'protein', 'rna' or 'dna'.
alphabet	The set of symbols to count, e.g. 'AGTC'. All characters not in the alphabet are ignored. If neither the alphabet nor sequence-type are specified then weblogo will examine the input data and make an educated guess.

units	A unit of entropy ('bits' (default), 'nats', 'digits'), or a unit of free energy ('kT', 'kJ/mol', 'kcal/mol'), or 'probability' for probabilities.
composition	The expected composition of the sequences: 'auto' (default), 'equiprobable', 'none' (do not perform any compositional adjustment), a CG percentage, a species name ('H. sapiens', 'E. coli', 'S. cerevisiae', 'C. elegans', 'D. melanogaster', 'M. musculus', 'T. thermophilus'), or an explicit distribution as a named numerical vector (e.g. c(A=10, C=40, G=40, T=10)). The automatic option uses a typical distribution for proteins and equiprobable distribution for everything else.
weight	The weight of prior data. Default depends on alphabet length.
first.index	Index of first position in sequence data (default: 1).
lower	Lower bound index of sequence to display.
upper	Upper bound index of sequence to display.
ignore.lower.case	
	Disregard lower case letters and only count upper case letters in sequences.
reverse	reverse sequences.
complement	complement DNA sequences.
size	Specify a standard logo size: 'small', 'medium', 'large' (default).
stacks.per.line	
	Maximum number of logo stacks per logo line (default: 40).
title	Logo title text.
label	A figure label, e.g. '2a'.
show.xaxis	Display sequence numbers along x-axis? (default: TRUE).
xlabel	X-axis label.
annotate	A comma separated list or vector of custom stack annotations, e.g. '1,3,4,5,6,7' or c(1,3,4,5,6,7). Annotation list must be same length as sequences.
yaxis	Height of yaxis in units (default: maximum value with uninformative prior).
show.yaxis	Display entropy scale along y-axis? (default: TRUE).
ylabel	Y-axis label (default: depends on plot type and units).
show.ends	Label the ends of the sequence? (default: FALSE).
fineprint	The fine print text at the bottom right corner (default: blank).
ticmarks	Distance between ticmarks (default: 1.0).
errorbars	Display error bars? (default: FALSE).
reverse.stacks	Draw stacks with largest letters on top? (default: TRUE).
color.scheme	Specify a standard color scheme ('auto', 'base pairing', 'charge', 'chemistry', 'classic', 'hydrophobicity', 'monochrome').
stack.width	Width of a logo stack (default: 10.8).
aspect.ratio	Ratio of stack height to width (default: 5).
box	Draw boxes around symbols? (default: FALSE).

resolution	Bitmap resolution in dots per inch (DPI). Low resolution bitmaps with DPI<300 are antialiased (default: 300 DPI).
scale.width	Scale the visible stack width by the fraction of symbols in the column? i.e. columns with many gaps of unknowns are narrow (default: TRUE).
rotate.numbers	Rotate values of x-axis? (default: FALSE).
hide.tics	Hide tic marks? (default: FALSE).

Examples

```
# Make a sequence logo using an external alignment file format
# In this example we'll use the EMBOSS alignment format or msf
# However, you can use any format supported by WebLogo e.g. fasta
fpath = system.file("extdata", "example_data.msf", package="RWebLogo")
weblogo(file.in=fpath)
# Now for an example using an alignment as an R character vector
aln <- c('CCAACCCAA', 'CCAACCCTA', 'AAAGCCTGA', 'TGAACCGGA')
# Simple WebLogo
weblogo(seqs=aln)
# Lets get rid of those ugly error bars and add some text!
weblogo(seqs=aln, errorbars=FALSE, title='Yay, No error bars!',
        fineprint='RWebLogo 1.0', label='1a')
# We can also change the format of the output like this
weblogo(seqs=aln, format='png', resolution=500)
# You can change the axis labels like this
weblogo(seqs=aln, xlabel='My x-axis', ylabel='Awesome bits')
# You get the idea! See ?weblogo for more awesome options!
```

Index

weblogo,