

Introduction to motif manipulation with the “universalmotif” package

Benjamin Tremblay^{*1}

¹University of Waterloo, Waterloo, Canada

^{*}b2tremblay@uwaterloo.ca

12 juillet 2018

Abstract

Import, manipulate, and export motifs with R.

Contents

1	Introduction to sequence motifs.	4
2	Importing and exporting motifs	4
2.1	Importing.	4
2.2	Exporting.	4
3	Converting motif class	4
3.1	Starting from another package	4
3.2	Starting as a universalmotif motif	4
4	Modifying motifs and related functions	4
4.1	Accessing and changing universalmotif slots	4
4.2	Converting motif type	4
4.3	Comparing and merging motifs	4
4.4	Multi-letter information	4
5	Motif creation	4
5.1	From a PCM/PPM/PWM/ICM matrix.	4
5.2	From sequences or character strings	4
5.3	Generating random motifs	4
6	Sequence-related functions	4
6.1	Sequence shuffling and creation	4
6.2	Searching for motifs in sequences.	4

Motif manipulation with “universalmotif”

- 7 Motif visualization. 4
 - 7.1 Motif logos 4
 - 7.2 Motif trees with ggtree and motifStack 4
- Session info 4

Motif manipulation with “universalmotif”

1 Introduction to sequence motifs

2 Importing and exporting motifs

2.1 Importing

2.2 Exporting

3 Converting motif class

3.1 Starting from another package

3.2 Starting as a universalmotif motif

4 Modifying motifs and related functions

4.1 Accessing and changing universalmotif slots

4.2 Converting motif type

4.3 Comparing and merging motifs

4.4 Multi-letter information

5 Motif creation

5.1 From a PCM/PPM/PWM/ICM matrix

5.2 From sequences or character strings

5.3 Generating random motifs

6 Sequence-related functions

6.1 Sequence shuffling and creation

6.2 Searching for motifs in sequences

7 Motif visualization

7.1 Motif logos

7.2 Motif trees with ggtree and motifStack

Motif manipulation with “universalmotif”

```
#> R version 3.5.1 (2018-07-02)
#> Platform: x86_64-apple-darwin15.6.0 (64-bit)
#> Running under: macOS High Sierra 10.13.4
#>
#> Matrix products: default
#> BLAS: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRblas.0.dylib
#> LAPACK: /Library/Frameworks/R.framework/Versions/3.5/Resources/lib/libRlapack.dylib
#>
#> locale:
#> [1] fr_CA.UTF-8/fr_CA.UTF-8/fr_CA.UTF-8/C/fr_CA.UTF-8/fr_CA.UTF-8
#>
#> attached base packages:
#> [1] stats      graphics  grDevices datasets  utils      methods    base
#>
#> other attached packages:
#> [1] BiocStyle_2.8.2      rmarkdown_1.10      universalmotif_0.98.0
#> [4] testthat_2.0.0      BiocCheck_1.16.0    nvimcom_0.9-71
#> [7] pacman_0.4.6        usethis_1.3.0       devtools_1.13.5.9000
#> [10] colorout_1.2-0      forcats_0.3.0       stringr_1.3.1
#> [13] dplyr_0.7.5         purrr_0.2.5         readr_1.1.1
#> [16] tidyr_0.8.1         tibble_1.4.2        ggplot2_2.2.1
#> [19] tidyverse_1.2.1
#>
#> loaded via a namespace (and not attached):
#> [1] readxl_1.1.0          backports_1.1.2
#> [3] VGAM_1.0-5            plyr_1.8.4
#> [5] lazyeval_0.2.1        splines_3.5.1
#> [7] BiocParallel_1.14.1   GenomeInfoDb_1.16.0
#> [9] TFBSTools_1.18.0      digest_0.6.15
#> [11] htmltools_0.3.6       BiocInstaller_1.30.0
#> [13] GO.db_3.6.0           gdata_2.18.0
#> [15] magrittr_1.5          memoise_1.1.0
#> [17] BSgenome_1.48.0       Biostrings_2.48.0
#> [19] annotate_1.58.0       modelr_0.1.2
#> [21] matrixStats_0.53.1    R.utils_2.6.0
#> [23] colorspace_1.3-2      blob_1.1.1
#> [25] rvest_0.3.2           xfun_0.2
#> [27] haven_1.1.1           callr_2.0.4
#> [29] crayon_1.3.4          RCurl_1.95-4.10
#> [31] jsonlite_1.5          roxygen2_6.0.1
#> [33] graph_1.58.0          bindr_0.1.1
#> [35] TFMPvalue_0.0.8       ape_5.1
#> [37] glue_1.2.0            gtable_0.2.0
#> [39] zlibbioc_1.26.0       XVector_0.20.0
#> [41] DelayedArray_0.6.1    pkgbuild_1.0.0
#> [43] evd_2.3-3            BiocGenerics_0.26.0
#> [45] scales_0.5.0         DBI_1.0.0
#> [47] bibtex_0.4.2         Rcpp_0.12.17
#> [49] xtable_1.8-2         tidytree_0.1.9
#> [51] foreign_0.8-70       bit_1.1-14
#> [53] stats4_3.5.1         htmlwidgets_1.2
```

Motif manipulation with “universalmotif”

```
#> [55] httr_1.3.1          getopt_1.20.2
#> [57] R.methodsS3_1.7.1    pkgconfig_2.0.1
#> [59] XML_3.98-1.11       PWMEnrich_4.16.0
#> [61] ggseqlogo_0.1        AnnotationDbi_1.42.1
#> [63] tidyselect_0.2.4     rlang_0.2.1
#> [65] reshape2_1.4.3       munsell_0.5.0
#> [67] biocViews_1.48.2     cellranger_1.1.0
#> [69] tools_3.5.1          cli_1.0.0
#> [71] DirichletMultinomial_1.22.0 RSQLite_2.1.1
#> [73] ade4_1.7-11          broom_0.4.4
#> [75] evaluate_0.10.1      rGADEM_2.28.0
#> [77] yaml_2.1.19          ggtree_1.12.0
#> [79] knitr_1.20           processx_3.1.0
#> [81] bit64_0.9-7          caTools_1.17.1
#> [83] splitstackshape_1.4.4 KEGGREST_1.20.0
#> [85] bindrcpp_0.2.2        RBGL_1.56.0
#> [87] nlme_3.1-137         R.oo_1.22.0
#> [89] powerLaw_0.70.1      xml2_1.2.0
#> [91] debugme_1.1.0        compiler_3.5.1
#> [93] rstudioapi_0.7       curl_3.2
#> [95] png_0.1-7            MotIV_1.36.0
#> [97] msa_1.12.0           treeio_1.4.1
#> [99] stringi_1.2.3        desc_1.2.0
#> [101] lattice_0.20-35      CNEr_1.16.1
#> [103] Matrix_1.2-14        commonmark_1.5
#> [105] psych_1.8.4          stringdist_0.9.5.1
#> [107] pillar_1.2.3         RUnit_0.4.32
#> [109] optparse_1.6.0       Rdpack_0.8-0
#> [111] data.table_1.11.4     grImport_0.9-1
#> [113] bitops_1.0-6         gbRd_0.4-11
#> [115] rtracklayer_1.40.3    GenomicRanges_1.32.3
#> [117] R6_2.2.2             bookdown_0.7
#> [119] MotifDb_1.22.0       motifStack_1.24.0
#> [121] IRanges_2.14.10      codetools_0.2-15
#> [123] MASS_7.3-50          gtools_3.5.0
#> [125] assertthat_0.2.0     seqLogo_1.46.0
#> [127] pkgload_1.0.0        SummarizedExperiment_1.10.1
#> [129] rprojroot_1.3-2      withr_2.1.2
#> [131] GenomicAlignments_1.16.0 Rsamtools_1.32.0
#> [133] mnormt_1.5-5         S4Vectors_0.18.3
#> [135] GenomeInfoDbData_1.1.0 parallel_3.5.1
#> [137] hms_0.4.2            grid_3.5.1
#> [139] rvcheck_0.1.0        Biobase_2.40.0
#> [141] lubridate_1.7.4      tinytex_0.5
```