The pepDat User Guide

Renan Sauteraud*

April 17, 2025

Contents

1 Introduction

`pepDat` is an R package that stores sample files, data for vignettes and peptide collections. It is intended to be used in conjunction with other packages for peptide analysis and visualisation: `pepStat` anf `Pviz`.

As with any R package, it should first be loaded in the session

```
library(pepDat)
```

2 Peptide collections

2.1 Information

For each peptide, the following collections display information about the position relative to the reference sequence, the alignment, the trimmed alignment, the zSums for the physicochemical properties as well as the clades where they can be found.

*rsautera@fhcrc.org

2.2 Structure

The datasets in this package are `GRanges` objects. For more information about the class, its accessors and setters, please refer to `GenomicRanges` documentation.

2.3 pep_hxb2

This collection is based on the alignment of the reference HIV sequence hxb2 and seven subtypes (clades) A, B, C, D, M, CRF01 and CRF02.

2.4 pep_hxb2JPT

This collection adds a few more clades to the previous one: CM244, CON_01_AE, LAI_A04321 and MN_DD328842.

2.5 pep_mac239

This collection is for SIV, with the clades mac239 and E660.

2.6 pep_m239smE543

This collection is for SIV, with the clades mac239 and E543.

3 Loading the collections

The peptide collections can be loaded like any other R dataset.

```
data(pep_hxb2)
head(pep_hxb2)

## GRanges object with 6 ranges and 10 metadata columns:
##                    seqnames    ranges strand |           names
##                       <Rle> <IRanges>  <Rle> |     <character>
##   MRVKETQMNWPNLWK     gp160      1-16      * | MRVKETQMNWPNLWK
##   MRVMGIQKNYPLLWR     gp160      1-16      * | MRVMGIQKNYPLLWR
##   MRVMGIQRNCQHLWR     gp160      1-16      * | MRVMGIQRNCQHLWR
##   MRVKGIRKNYQHLWR     gp160      1-16      * | MRVKGIRKNYQHLWR
##   MRVRGILRNWQQWWI     gp160      1-16      * | MRVRGILRNWQQWWI
##   MRVRGIERNYQHLWR     gp160      1-16      * | MRVRGIERNYQHLWR
##                              aligned          trimmed     seqNb       clade
##                             <factor>         <factor> <integer> <character>
##   MRVKETQMNWPNLWK MRVKETQMNWPNL----WK MRVKenWPNL----WK         1       CRF01
##   MRVMGIQKNYPLLWR MRVMGIQKNYPLL----WR MRVMgnYPLL----WR         1       CRF02
##   MRVMGIQRNCQHLWR MRVMGIQRNCQHL----WR MRVMgnCQHL----WR         1           A
##   MRVKGIRKNYQHLWR MRVKGIRKNYQHL----WR MRVKgnYQHL----WR         1           B
##   MRVRGILRNWQQWWI MRVRGILRNWQQW----WI MRVRgnWQQW----WI         1           C
##   MRVRGIERNYQHLWR MRVRGIERNYQHL----WR MRVRgnYQHL----WR         1           D
```

```
##                          z1        z2        z3        z4        z5
##                   <numeric> <numeric> <numeric> <numeric> <numeric>
##   MRVKETQMNWPNLWK     -3.14      9.87     -8.14      6.72      1.36
##   MRVMGIQKNYPLLWR    -13.12      3.19    -11.96      7.09      0.76
##   MRVMGIQRNCQHLWR      1.65      4.17    -11.22     10.31     -2.99
##   MRVKGIRKNYQHLWR      3.95      9.78    -18.51     12.55     -0.87
##   MRVRGILRNWQQWWI    -11.42     10.26    -16.19     10.33     -2.37
##   MRVRGIERNYQHLWR      6.00     10.76    -17.14      8.52     -1.91
##   -------
##   seqinfo: 1 sequence from an unspecified genome; no seqlengths
```

4 Create collections

While the package comes with datasets for HIV and SIV. It is possible to create new collections for different organisms or proteins using the `create_db` function in `pepStat`. Please refer to `pepStat`'s user guide and ?create_db for more information.