

The PEP.db User Guide

Renan Sauteraud*

July 1, 2013

Contents

1	Introduction	1
2	Peptide collections	1
2.1	Information	1
2.2	pep_hxb2	1
2.3	pep_hxb2JPT	1
2.4	pep_mac239	1
2.5	pep_mac239	2
3	Loading the collections	2

1 Introduction

PEP.db is an R package that stores peptide collections and provide functions for their analysis. It is intended to be used in conjunction with other packages for peptide analysis and visualisation: HIV.db, pepStat and Pviz.

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

```
> library(PEP.db)
```

2 Peptide collections

2.1 Information

For each peptide, the following collection 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.

2.2 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.

*rsautera@fhcrc.org

2.3 pep_hxb2JPT

This collection adds a few more clades to the previous one: CM244, CON_01_AE, LALA04321 and MN_DD328842.

2.4 pep_mac239

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

2.5 pep_mac239

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 data.

```
> data(pep_hxb2)
> pep_hxb2
```

RangedData with 1423 rows and 10 value columns across 1 space

	space	ranges		names	aligned
	<factor>	<IRanges>		<character>	<factor>
MRVKETQMNWPNLWK	gp160	[1, 16]		MRVKETQMNWPNLWK	MRVKETQMNWPNL----WK
MRVMGIQKNYPPLLWR	gp160	[1, 16]		MRVMGIQKNYPPLLWR	MRVMGIQKNYPPLL----WR
MRVMGIQRNCQHLWR	gp160	[1, 16]		MRVMGIQRNCQHLWR	MRVMGIQRNCQHL----WR
MRVKGIRKNYQHLWR	gp160	[1, 16]		MRVKGIRKNYQHLWR	MRVKGIRKNYQHL----WR
MRVRGILRNWQQWWI	gp160	[1, 16]		MRVRGILRNWQQWWI	MRVRGILRNWQQW----WI
MRVRGIERNYQHLWR	gp160	[1, 16]		MRVRGIERNYQHLWR	MRVRGIERNYQHL----WR
MRVMGIQRNWQHLWR	gp160	[1, 16]		MRVMGIQRNWQHLWR	MRVMGIQRNWQHL----WR
KETQMNWPNLWKWGT	gp160	[4, 19]		KETQMNWPNLWKWGT	KETQMNWPNL----WKWGT
MGIQKNYPPLLWRWGT	gp160	[4, 19]		MGIQKNYPPLLWRWGT	MGIQKNYPPLL----WRWGT
...
VRAILNIPRRIRQGL	gp160	[837, 851]		VRAILNIPRRIRQGL	VRAILNIPRRIRQGL
GRAILNIPRRIRQGL	gp160	[837, 851]		GRAILNIPRRIRQGL	GRAILNIPRRIRQGL
ILHIPRRIRQGFERA	gp160	[840, 854]		ILHIPRRIRQGFERA	ILHIPRRIRQGFERA
ILHIPRRIRQGLERA	gp160	[840, 854]		ILHIPRRIRQGLERA	ILHIPRRIRQGLERA
IRNIPRRIRQGFEEA	gp160	[840, 854]		IRNIPRRIRQGFEEA	IRNIPRRIRQGFEEA
ILNIPRRIRQGLERA	gp160	[840, 854]		ILNIPRRIRQGLERA	ILNIPRRIRQGLERA
IPRRIRQGFERALL	gp160	[843, 856]		IPRRIRQGFERALL	IPRRIRQGFERALL
IPRRIRQGFEEALL	gp160	[843, 856]		IPRRIRQGFEEALL	IPRRIRQGFEEALL
IPRRIRQGLERALL	gp160	[843, 856]		IPRRIRQGLERALL	IPRRIRQGLERALL
	trimmed	seqNb		clade	z1
	<factor>	<integer>		<character>	<numeric>
MRVKETQMNWPNLWK	MRVKenWPNL----WK	1		CRF01	-3.14
MRVMGIQKNYPPLLWR	MRVMgnYPLL----WR	1		CRF02	-13.12
MRVMGIQRNCQHLWR	MRVMgnCQHL----WR	1		A	1.65

MRVKGIRKQNYQHLWR	MRVKgnYQHL----	WR	1	B	3.95
MRVRGILRNWQQWWI	MRVRgnWQQW----	WI	1	C	-11.42
MRVRGIERNYQHLWR	MRVRgnYQHL----	WR	1	D	6.00
MRVMGIQRNWQHLWR	MRVMgnWQHL----	WR	1	M	-3.55
KETQMNPWLWKWGT	KenWPNL----	WKWGT	2	CRF01	-2.78
MGIQKNYPLLWRWGT	MgnYPLL----	WRWGT	2	CRF02	-12.76
...		
VRAILNIPRRIRQGL	VRAILNIPRRIRQGL		276	D	-3.31
GRAILNIPRRIRQGL	GRAILNIPRRIRQGL		276	M,CRF02	1.33
ILHIPRRIRQGFERA	ILHIPRRIRQGFERA		277	A	1.87
ILHIPRRIRQGLERA	ILHIPRRIRQGLERA		277	B,CRF01	1.81
IRNIPRRIRQGFEEA	IRNIPRRIRQGFEEA		277	C	6.97
ILNIPRRIRQGLERA	ILNIPRRIRQGLERA		277	M,D,CRF02	2.39
IPRRIRQGFERALL	IPRRIRQGFERALL		278	A	-0.99
IPRRIRQGFEEALL	IPRRIRQGFEEALL		278	C	-4.27
IPRRIRQGLERALL	IPRRIRQGLERALL		278	M,B,D,CRF01,CRF02	-1.05
	z2	z3	z4	z5	
	<numeric>	<numeric>	<numeric>	<numeric>	
MRVKETQMNPWLWK	9.87	-8.14	6.72	1.36	
MRVMGIQKNYPLLWR	3.19	-11.96	7.09	0.76	
MRVMGIQRNCQHLWR	4.17	-11.22	10.31	-2.99	
MRVKGIRKQNYQHLWR	9.78	-18.51	12.55	-0.87	
MRVRGILRNWQQWWI	10.26	-16.19	10.33	-2.37	
MRVRGIERNYQHLWR	10.76	-17.14	8.52	-1.91	
MRVMGIQRNWQHLWR	9.78	-14.34	13.57	-1.93	
KETQMNPWLWKWGT	7.93	-3.74	4.80	0.16	
MGIQKNYPLLWRWGT	1.25	-7.56	5.17	-0.44	
...	
VRAILNIPRRIRQGL	-4.42	-21.25	0.40	6.95	
GRAILNIPRRIRQGL	-5.84	-19.35	0.43	6.59	
ILHIPRRIRQGFERA	2.05	-18.05	4.52	3.74	
ILHIPRRIRQGLERA	-1.19	-20.60	3.26	5.20	
IRNIPRRIRQGFEEA	0.70	-15.18	0.05	5.72	
ILNIPRRIRQGLERA	-1.52	-19.82	-1.79	6.72	
IPRRIRQGFERALL	0.53	-18.09	0.74	4.23	
IPRRIRQGFEEALL	-4.29	-13.99	-1.39	5.70	
IPRRIRQGLERALL	-2.71	-20.64	-0.52	5.69	