

# Package ‘rSWeeP’

January 16, 2022

**Title** Functions to creation of low dimensional comparative matrices of Amino Acid Sequence occurrences

**Version** 1.6.0

**Description** The SWeeP method was developed to favor the analyzes between amino acids sequences and to assist alignment free phylogenetic studies. This method is based on the concept of sparse words, which is applied in the scan of biological sequences and its the conversion in a matrix of occurrences. Aiming the generation of low dimensional matrices of Amino Acid Sequence occurrences.

**biocViews** Software,StatisticalMethod,SupportVectorMachine,Technology,Sequencing,Genetics,Alignment

**Depends** R (>= 4.0)

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.0.2

**Imports** pracma, stats

**Suggests** Biostrings, methods, knitr, rmarkdown, BiocStyle

**VignetteBuilder** knitr

**git\_url** <https://git.bioconductor.org/packages/rSWeeP>

**git\_branch** RELEASE\_3\_14

**git\_last\_commit** 5b84f77

**git\_last\_commit\_date** 2021-10-26

**Date/Publication** 2022-01-16

**Author** Danrley R. Fernandes [com, cre, aut]

**Maintainer** Danrley R. Fernandes <DanrleyRF@gmail.com>

## R topics documented:

orthBase . . . . .	2
sWeeP . . . . .	3
<b>Index</b>	<b>5</b>

---

orthBase	<i>Generate a orthonormal matrix (lin, col)</i>
----------	-------------------------------------------------

---

**Description**

Generate a orthonormal matrix in a specified size, lin by col.

**Usage**

```
orthBase(lin, col)
```

**Arguments**

lin	Number of rows in the desired matrix
col	Number of columns in the desired matrix

**Value**

A orthonormal matrix in a specified size, lin by col.

**Author(s)**

Danrley R. Fernandes.

**See Also**

[sWeeP](#), [orth](#)

**Examples**

```
orthBase(160000, 10)

lin <- 160000
col <- 10
orthBase(lin = lin, col = col)
```

---

`sWeeP`*A vectorial comparative method to amino acid sequence.*

---

**Description**

The "Spaced Words Projection (SWeeP)" is a method for representing biological sequences using compact vectors. SWeeP uses the spacedwords concept by scanning sequences and generating indices to create a higherdimensional matrix of occurrences that is later projected into a smaller randomly oriented orthonormal base (PIERRI, 2019). This way the resulting matrix will conserve the comparational data but will have a selectable size

**Usage**

```
sWeeP(xfas, baseMatrix)

## S4 method for signature 'character'
sWeeP(xfas, baseMatrix)

## S4 method for signature 'AAStringSet'
sWeeP(xfas, baseMatrix)
```

**Arguments**

<code>xfas</code>	A AAStringSet or a FASTA format file
<code>baseMatrix</code>	A orthonormal matrix with 160.000 coordinates

**Details**

The SWeeP method was developed to favor the comparison between complete proteomic sequences and to assist in machine learning analyzes. This method is based on the concept of spaced words, which are used to scan biological sequences and project them into matrix of occurrences, favoring the manipulation of the data. The `sWeeP` function can project a matrix  $n$  by  $m$ , where  $n$  is the number of sequences in the analyzed `xfas` and  $m$  is the number of columns in `baseMatrix` matrix.

**Value**

A matrix resulted by the projection of the sequences in `xfas` in the `baseMatrix` matrix

**Author(s)**

Danrley R. Fernandes.

**References**

Pierri, C. R. et al. SWeeP: Representing large biological sequences data sets in compact vectors. Scientific Reports, accepted in December 2019. doi: 10.1038/s41598-019-55627-4.

**Examples**

```
baseMatrix <- orthBase(160000,10)
path <- system.file(package = "rSWeeP", "extdata", "exdna.fas")
return <- sWeeP(path,baseMatrix)
distancia <- dist(return, method = "euclidean")
tree <- hclust(distancia, method="ward.D")
plot(tree, hang = -1, cex = 1)
```

# Index

[orth](#), [2](#)

[orthBase](#), [2](#)

[sWeeP](#), [2](#), [3](#)

[sWeeP](#), [AAStringSet-method \(sWeeP\)](#), [3](#)

[sWeeP](#), [character-method \(sWeeP\)](#), [3](#)