

# Package ‘FMIndex’

August 25, 2023

**Title** A package for computing FM index

**Version** 0.99.0

**Description** The package computes the FM index of a DNA string contained in a fasta file.  
The user can specify the path where he wants to save the FM index data structures.  
The results will be also printed on the console.

**Imports** Biostrings, utils

**Suggests** knitr, rmarkdown, BiocStyle, testthat

**License** GPL-2

**Encoding** UTF-8

**Roxygen** list(markdown = TRUE)

**RoxygenNote** 7.2.3

**VignetteBuilder** knitr

**biocViews** Normalization, Preprocessing

**Depends** R (>= 3.6)

**NeedsCompilation** no

**Author** Sara Resta [aut, cre] (YOUR-ORCID-ID)

**Maintainer** Sara Resta <sara.resta@mail.polimi.it>

## R topics documented:

FMIndex-package . . . . .	2
FMindex . . . . .	2
getBWTcharacter . . . . .	3
getFcolumn . . . . .	4
getLcolumn . . . . .	4
getSA . . . . .	5
getSuffix . . . . .	5
getTallyTable . . . . .	6
getTallyTableLine . . . . .	6
<b>Index</b>	<b>7</b>

---

FMIndex-package	<i>FMIndex - Computer the FM index for a DNA sequence contained in a fasta file The FMIndex package implements all the functions necessary to compute the data structures of the FM index</i>
-----------------	---

---

## Description

Package: FMIndex  
 Type: Package  
 Date: 2023-06-20  
 License: GPL (>=2)

## Author(s)

Sara Resta [aut, cre]  
 Politecnico di Milano  
 Maintainer: Sara Resta  
 E-mail: [sara.resta@mail.polimi.it](mailto:sara.resta@mail.polimi.it)

## References

<http://some.url/here/maybe/>  
 Resta et al.

---

FMIndex	<i>Writes the FM index</i>
---------	----------------------------

---

## Description

This function writes the FM index of the DNA sequence contained in the .fasta file given in input

## Usage

```

FMindex(
  fastafilename,
  usrp,
  tallywidth = 1,
  includeTerminationCharacter = TRUE,
  includeSA = FALSE
)
  
```

**Arguments**

fastafile	The .fasta file containing one single DNA sequence, files containing multiple sequences are not allowed
usrpath	The directory path where the user wants to save the files: it must end with '/' (you can leave it empty if you want to save the files in your working directory).
tallywidth	How you want to compress the tally table. Default value=1
includeTerminationCharacter	default value: true
includeSA	set it to TRUE if you want to print also the suffix array.

**Value**

A list containing all the data structures necessary for the FM index

**Examples**

```
FMIndex(system.file("extdata/examples/seq5.fasta", package="FMIndex"), "")
FMIndex(system.file(
  "extdata/examples/seq5.fasta", package="FMIndex"), "", tallywidth=1)
```

---

getBWTcharacter	<i>Compute the i-th character of the BWT</i>
-----------------	--

---

**Description**

This function computes the i-th character of the BWT of a given sequence which corresponds to the i-th character in the L column.

**Usage**

```
getBWTcharacter(index, sequence, suffixarray)
```

**Arguments**

index	The position which you want to find the character
sequence	The string you want to know the i-th L column character
suffixarray	An integer vector containing the suffix array of the input string

**Value**

A character containing the i-th character of the input sequence

---

getFcolumn	<i>Compute the F column</i>
------------	-----------------------------

---

### Description

This function gets a DNASTring and gives the F column of the FM index

### Usage

```
getFcolumn(seq, includeTerminationCharacter = TRUE)
```

### Arguments

seq	the DNASTring to be analysed
includeTerminationCharacter	True if you want to include the ".", false if you want to exclude it. Default value:TRUE

### Value

An integer vector with all the frequencies in the dna sequence given in input

### Examples

```
getFcolumn(Biostrings::DNASTring("ACCT"))
```

---

getLcolumn	<i>Compute the L column of a sequence</i>
------------	---

---

### Description

This function gets a character and gives the F column of the FM index

### Usage

```
getLcolumn(sequence)
```

### Arguments

sequence	the DNASTring to be analysed
----------	------------------------------

### Value

A DNASTring containing the L column of the FM index

**Examples**

```
getLcolumn(Biostrings::DNASTring("ACGT"))  
getLcolumn(Biostrings::DNASTring("AACCGT"))
```

---

`getSA`*Compute the suffix array of a string*

---

**Description**

This function computes the suffix array of a given string

**Usage**

```
getSA(sequence)
```

**Arguments**

`sequence`            a DNASTring representing the string sequence

**Value**

An integer array containing the suffix array

**Examples**

```
getSA(Biostrings::DNASTring("ACGT"))  
getSA(Biostrings::DNASTring("AACCGT"))
```

---

`getSuffix`*Get the suffix*

---

**Description**

This function gets the suffix of the string starting from a position given in input

**Usage**

```
getSuffix(index, sequence)
```

**Arguments**

`index`            An integer representing the starting position of the string  
`sequence`        The sequence you want to find the suffix

**Value**

The substring starting at the given position

---

getTallyTable	<i>Create the Tally Table</i>
---------------	-------------------------------

---

**Description**

Creates the Tally Table of a given BWT of a string

**Usage**

```
getTallyTable(l.col, rowwidth = 1)
```

**Arguments**

l.col	A DNASTring containing the L column of the FM index
rowwidth	Step for saving rows of the tally table. The default value is 1 which corresponds to save all the rows of the tally table

**Value**

The tally table of the input sequence

**Examples**

```
getTallyTable(Biostrings::DNASTring("T.AACCG"))
getTallyTable(Biostrings::DNASTring("T.AACCG"), 2)
```

---

getTallyTableLine	<i>Create the i-th row of the Tally Table</i>
-------------------	---

---

**Description**

This function creates the i-th row of the Tally Table

**Usage**

```
getTallyTableLine(index, l.col, alphabet)
```

**Arguments**

index	An integer representing the index of the i-th row you want to build
l.col	A DNASTring object containing the L column of the FM index
alphabet	A character vector containing the alphabet to consider in counting the elements

**Value**

the i-th row of the Tally Table

# Index

FMIndex (FMIndex-package), [2](#)

FMIndex, [2](#)

FMIndex-package, [2](#)

getBWTcharacter, [3](#)

getFcolumn, [4](#)

getLcolumn, [4](#)

getSA, [5](#)

getSuffix, [5](#)

getTallyTable, [6](#)

getTallyTableLine, [6](#)