

# Package ‘FMIndex’

August 28, 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 (>= 4.1.0)

**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 . . . . .	3
getBWTcharacter . . . . .	4
getFcolumn . . . . .	4
getLcolumn . . . . .	5
getSA . . . . .	6
getSuffix . . . . .	6
getTallyTable . . . . .	7
getTallyTableLine . . . . .	8
<b>Index</b>	<b>9</b>

---

FMIndex-package	<i>FMIndex - Compute the FM index for a DNA sequence contained in a fasta file</i>
-----------------	--

---

## Description

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

## Details

The FMIndex package implements all the functions necessary to compute the data structures of the FM index. The package has one main function which computes all the FM index data structures and saves them into different text files (the path is specified by the user):

- F\_col.txt for the F column (cont the occurrences of the unique characters)
- L\_col.txt for the L column (BWT)
- suffix\_array.txt for the suffix array
- tally\_table.txt for the tally table

The functions that compute the FM index data structures are available: they all take a DNASTring as input and the return values are printed on the console.

NOTE: to make the strings compatible with DNASTring objects the termination character used to build the BWT is "."

## 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)

---

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, uspath, tallywidth = 1, includeEndChar = TRUE)
```

## Arguments

fastafilename	The .fasta file containing one single DNA sequence, files containing multiple sequences are not allowed
uspath	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. It states how often (in terms of rows) you save one checkpoint. Default value=1
includeEndChar	True if you want to include the "." in the F column, false if you want to exclude it. Default value:TRUE

## Details

This function may raise an error if:

- The FASTA file provided contains more than one sequence
- The FASTA file provided contains no sequences

## Value

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

## Examples

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

---

getBWTcharacter	<i>Helper function: compute the i-th character of the BWT</i>
-----------------	---

---

### Description

This is a helper function that performs a specific task for the main function [getLcolumn](#). It 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

### See Also

[getLcolumn](#)

---

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, includeEndChar = TRUE)
```

### Arguments

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

**Details**

This function may raise a warning if

- The provided sequence is empty

**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

**Details**

- The provided sequence is empty

**Value**

A list containing the L column (BWT) and the suffix array

**Examples**

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

---

getSA	<i>Compute the suffix array of a string</i>
-------	---

---

**Description**

This function computes the suffix array of a given string

**Usage**

```
getSA(sequence)
```

**Arguments**

sequence            a DNASTring representing the string sequence

**Details**

This function may raise a warning if

- The provided sequence is empty

**Value**

An integer array containing the suffix array

**Examples**

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

---

getSuffix	<i>Helper function: get the suffix</i>
-----------	--

---

**Description**

This is a helper function that performs a specific task for the main function [getSA](#). It 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

**See Also**

[getSA](#)

---

getTallyTable

*Create the Tally Table*

---

**Description**

Creates the Tally Table of a given BWT of a string

**Usage**

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

**Arguments**

<code>l.col</code>	A DNASTring containing the L column of the FM index
<code>rowwidth</code>	Step (in term of number of rows) for saving checkpoints of the tally table. The default value is 1 which corresponds to save all the rows of the tally table

**Details**

This function may raise an error if

- The argument rowwidth is higher than one
- The argument rowwidth is lower than one
- The input sequence (argument l.col) doesn't contain the termination character "."
- The input sequence (argument l.col) contains more than one termination character "."

This function may raise a warning if

- The input sequence contains only the termination character

**Value**

The tally table of the input sequence

**Examples**

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

---

getTallyTableLine	<i>Helper function: create the i-th row of the Tally Table</i>
-------------------	--

---

**Description**

This is a helper function that performs a specific task for the main function [getTallyTable](#). It 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

**See Also**

[getTallyTable](#)



# Index

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

FMIndex, [3](#)

FMIndex-package, [2](#)

getBWTcharacter, [4](#)

getFcolumn, [4](#)

getLcolumn, [4](#), [5](#)

getSA, [6](#), [6](#), [7](#)

getSuffix, [6](#)

getTallyTable, [7](#), [8](#)

getTallyTableLine, [8](#)