

Package ‘Rqc’

April 10, 2015

Type Package

Title Quality Control Tool for High-Throughput Sequencing Data

Version 1.0.4

Author Welliton Souza, Benilton Carvalho <beniltoncarvalho@gmail.com>

Maintainer Welliton Souza <well309@gmail.com>

Description Rqc is an optimised tool designed for quality control and assessment of high-throughput sequencing data. It performs parallel processing of entire files and produces a report which contains a set of high-resolution graphics.

License GPL (>= 2)

Depends BiocParallel, ShortRead, ggplot2

Imports BiocGenerics, Biostrings, IRanges, methods, reshape2, S4Vectors, knitr (>= 1.7), BiocStyle

Suggests rmarkdown

VignetteBuilder knitr

biocViews Sequencing, QualityControl

URL <https://github.com/labbc/b/Rqc>

R topics documented:

Rqc-package	2
rqc	2
rqcCycleAverageQualityCalc	3
rqcCycleAverageQualityPlot	4
rqcCycleBaseCallsCalc	5
rqcCycleBaseCallsLinePlot	6
rqcCycleBaseCallsPlot	6
rqcCycleGCCalc	7
rqcCycleGCPlot	8
rqcCycleQualityBoxCalc	9
rqcCycleQualityBoxPlot	10

rqcCycleQualityCalc	10
rqcCycleQualityPlot	11
rqcQA	12
rqcReadQualityCalc	13
rqcReadQualityPlot	14
rqcReadWidthCalc	14
rqcReadWidthPlot	15
rqcReport	16
RqcResultSet-class	17

Index	18
--------------	-----------

Rqc-package	<i>Quality Control Tool for High-Throughput Sequencing Data</i>
-------------	---

Description

Rqc is an optimized tool designed for quality assessment of high-throughput sequencing data. It performs parallel processing of entire files and produces a report, which contains a set of high-resolution images that can be directly used on publications.

Author(s)

Welliton Souza, Benilton Carvalho

Maintainer: Welliton Souza <well309@gmail.com>

Examples

```
options(device.ask.default = FALSE)
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
rqc(folder, ".fastq.gz")
```

rqc	<i>Main Rqc function</i>
-----	--------------------------

Description

Rqc is an optimized tool designed for quality assessment of high-throughput sequencing data. It performs parallel processing of entire files and produces an HTML report, which contains a set of high-resolution images that can be directly used on publications.

Usage

```
rqc(path=".", pattern,
     sample=TRUE, n=1e6,
     outdir=tempdir(), file="rqc_report",
     openBrowser=TRUE)
```

Arguments

path	directory path that contains input files.
pattern	a regex expression that matches to input file names
sample	it reads a random sample from files if this parameter is TRUE.
n	number of sequences to read from each input file. This represents sample size if 'sample' parameter is TRUE, if not represents the chunk size to read on each iteration. By default, it reads a sample of one million sequences from each input file.
outdir	output directory path. Is created a temporary directory by default.
file	output file name.
openBrowser	if TRUE opens report file on default Internet Browser.

Value

A invisible named list of RqcResultSet objects, each one represents a file.

Author(s)

Welliton Souza

See Also

[rqcQA](#)

Examples

```
options(device.ask.default = FALSE)
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
rqc(folder, ".fastq.gz", openBrowser=FALSE)
```

rqcCycleAverageQualityCalc

Per cycle average quality calculation

Description

Calculates average quality per cycle.

Usage

```
rqcCycleAverageQualityCalc(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Data frame ready for plot.

Author(s)

Welliton Souza

See Also

[rqcCycleAverageQualityPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
df <- rqcCycleAverageQualityCalc(rqcResultSet)
qplot(x=cycle, y=quality, colour=filename, data=df, geom="line")
```

`rqcCycleAverageQualityPlot`

Per cycle average quality plot

Description

Creates a line graph of per cycle average quality.

Usage

```
rqcCycleAverageQualityPlot(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by [rqc](#) and [rqcQA](#) functions.

Value

Plot object from [ggplot](#) function.

Author(s)

Welliton Souza

See Also

[rqcCycleAverageQualityCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcCycleAverageQualityPlot(rqcResultSet)
```

`rqcCycleBaseCallsCalc` *Per cycle base call calculation*

Description

Calculates average base call per cycle.

Usage

```
rqcCycleBaseCallsCalc(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

List of data frames ready for plot.

Author(s)

Welliton Souza

See Also

[rqcCycleBaseCallsPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
dfList <- rqcCycleBaseCallsCalc(rqcResultSet)
df <- do.call(rbind, dfList)
qplot(x=cycle, y=value, colour=variable, group=variable, data=df,
      geom="line", facets=~filename)
```

`rqcCycleBaseCallsLinePlot`

Per cycle base calls line plot

Description

Creates a line graph of per cycle base calls.

Usage

```
rqcCycleBaseCallsLinePlot(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

Plot object from `ggplot` function.

Author(s)

Welliton Souza

See Also

[rqcCycleBaseCallsCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcCycleBaseCallsLinePlot(rqcResultSet)
```

`rqcCycleBaseCallsPlot` *Per cycle base calls plot*

Description

Creates a bar graph of per cycle base calls.

Usage

```
rqcCycleBaseCallsPlot(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

Plot object from `ggplot` function.

Author(s)

Welliton Souza

See Also

[rqcCycleBaseCallsCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcCycleBaseCallsPlot(rqcResultSet)
```

`rqcCycleGCCalc` *Per cycle percentual GC.*

Description

Calculates per cycle percentual GC.

Usage

```
rqcCycleGCCalc(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

Data frame ready for plot.

Author(s)

Welliton Souza

See Also

[rqcCycleGCPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
df <- rqcCycleGCCalc(rqcResultSet)
qplot(x=cycle, y=gc, colour=filename, data=df, geom="line")
```

rqcCycleGCPlot	<i>Per cycle percentual GC plot</i>
----------------	-------------------------------------

Description

Creates a line graph of per cycle percentual GC.

Usage

```
rqcCycleGCPlot(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

Plot object from `ggplot` function.

Author(s)

Welliton Souza

See Also

[rqcCycleGCCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcCycleGCPlot(rqcResultSet)
```

`rqcCycleQualityBoxCalc`*Per cycle quality percentiles calculation for boxplot*

Description

Calculates per cycle quality percentiles to create boxplot.

Usage

```
rqcCycleQualityBoxCalc(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

List of data frames ready for plot.

Author(s)

Welliton Souza

See Also

[rqcCycleQualityBoxPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
dfList <- rqcCycleQualityBoxCalc(rqcResultSet)
df <- do.call(rbind, dfList)
qplot(x=cycle, ymin=ymin, lower=lower, middle=middle, upper=upper, ymax=ymax,
      data=df, geom="boxplot", stat="identity", facets=~filename)
```

rqcCycleQualityBoxPlot

Per cycle quality boxplot

Description

Plots per cycle quality boxplot.

Usage

```
rqcCycleQualityBoxPlot(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Plot object from [ggplot](#) function.

Author(s)

Welliton Souza

See Also

[rqcCycleQualityBoxCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcCycleQualityBoxPlot(rqcResultSet)
```

rqcCycleQualityCalc *Per cycle quality calculation*

Description

Calculates per cycle quality percentiles.

Usage

```
rqcCycleQualityCalc(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Data frame ready for plot.

Author(s)

Welliton Souza

See Also

[rqcCycleQualityPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
dfList <- rqcCycleQualityCalc(rqcResultSet)
df <- do.call(rbind, dfList)
qplot(x=cycle, y=percentiles, fill=value, data=df, geom="bar",
      stat="identity", facets=~filename)
```

`rqcCycleQualityPlot` *Per cycle quality plot*

Description

Creates a graph of per cycle quality.

Usage

```
rqcCycleQualityPlot(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Plot object from [ggplot](#) function.

Author(s)

Welliton Souza

See Also[rqcCycleQualityCalc](#)**Examples**

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcCycleQualityPlot(rqcResultSet)
```

rqcQA*Quality Assessment Rqc function*

Description

Process a set of files and returns a list of quality control data. Files must be FASTQ format, compressed or not.

Usage

```
rqcQA(files, sample = TRUE, n = 1e6)
```

Arguments

files	A vector of file paths.
sample	It reads a random sample from files if this parameter is TRUE.
n	Number of sequences to read from each input file. This represents sample size if 'sample' parameter is TRUE, if not represents the chunk size to read on each iteration. Default is read a sample of one million sequences from each input file.

Details

Input files are read using `FastStreamer` and `FastSampler` classes of [ShortRead](#) package. Process multiple files in parallel using `bplapply` function of [BiocParallel](#) package.

Value

A named list of `RqcResultSet` objects, each one represents a file.

Author(s)

Welliton Souza

See Also[rqc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(full.names=TRUE, path=folder)
rqcResultSet <- rqcQA(files)
rqcReadQualityPlot(rqcResultSet)
```

rqcReadQualityCalc	<i>Per read quality calculation</i>
--------------------	-------------------------------------

Description

Calculates per read quality

Usage

```
rqcReadQualityCalc(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Data frame ready for plot.

Author(s)

Welliton Souza

See Also

[rqcReadQualityPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
df <- rqcReadQualityCalc(rqcResultSet)
qplot(x=quantile, y=value, colour=filename, data=df, geom="line")
```

`rqcReadQualityPlot` *Per read quality plot*

Description

Plots the quality of all the files by read.

Usage

```
rqcReadQualityPlot(rqcResultSet)
```

Arguments

`rqcResultSet` list of `RqcResultSet` objects created by `rqc` and `rqcQA` functions.

Value

Plot object from `ggplot` function.

Author(s)

Welliton Souza

See Also

[rqcReadQualityCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcReadQualityPlot(rqcResultSet)
```

`rqcReadWidthCalc` *Per read width calculation*

Description

Calculates amount of per read width

Usage

```
rqcReadWidthCalc(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Data frame ready for plot.

Author(s)

Welliton Souza

See Also

[rqcReadWidthPlot](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
df <- rqcReadWidthCalc(rqcResultSet)
qplot(x=width, y=count, data=df, geom="bar", stat="identity", facets=~filename)
```

rqcReadWidthPlot *Per read width plot*

Description

Creates bar graph of per read width from all elements of input list.

Usage

```
rqcReadWidthPlot(rqcResultSet)
```

Arguments

rqcResultSet list of RqcResultSet objects created by [rqc](#) and [rqcQA](#) functions.

Value

Plot object from [ggplot](#) function.

Author(s)

Welliton Souza

See Also

[rqcReadWidthCalc](#)

Examples

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(folder, "fastq.gz", full.names=TRUE)
rqcResultSet <- rqcQA(files)
rqcReadWidthPlot(rqcResultSet)
```

rqcReport

Quality Control HTML Report

Description

Generates an HTML report file.

Also creates a directory called "figure" in outdir path.

Usage

```
rqcReport(rqcResultSet, outdir=tempdir(), file="rqc_report")
```

Arguments

rqcResultSet	list of RqcResultSet objects created by rqc and rqcQA functions.
outdir	output directory path. Is created a temporary directory by default.
file	output file name.

Value

Report file path.

Author(s)

Welliton Souza

See Also

[rqc](#)

[rqcQA](#)

Examples

```
options(device.ask.default = FALSE)
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(full.names=TRUE, path=folder)
rqcResultSet <- rqcQA(files)
reportFile <- rqcReport(rqcResultSet)
browseURL(reportFile)
```

RqcResultSet-class *Quality control data*

Description

Represents quality control data used by Rqc functions. Extends .QA class from ShortRead package.

Index

*Topic **graphics**

- [rqc](#), [2](#)
- [rqcCycleAverageQualityPlot](#), [4](#)
- [rqcCycleBaseCallsLinePlot](#), [6](#)
- [rqcCycleBaseCallsPlot](#), [6](#)
- [rqcCycleGCPlot](#), [8](#)
- [rqcCycleQualityBoxPlot](#), [10](#)
- [rqcCycleQualityPlot](#), [11](#)
- [rqcReadQualityPlot](#), [14](#)
- [rqcReadWidthPlot](#), [15](#)

*Topic **package**

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

*Topic **qc**

- [rqc](#), [2](#)
- [rqcCycleAverageQualityCalc](#), [3](#)
- [rqcCycleAverageQualityPlot](#), [4](#)
- [rqcCycleBaseCallsCalc](#), [5](#)
- [rqcCycleBaseCallsLinePlot](#), [6](#)
- [rqcCycleBaseCallsPlot](#), [6](#)
- [rqcCycleGCCalc](#), [7](#)
- [rqcCycleGCPlot](#), [8](#)
- [rqcCycleQualityBoxCalc](#), [9](#)
- [rqcCycleQualityBoxPlot](#), [10](#)
- [rqcCycleQualityCalc](#), [10](#)
- [rqcCycleQualityPlot](#), [11](#)
- [rqcReadQualityCalc](#), [13](#)
- [rqcReadQualityPlot](#), [14](#)
- [rqcReadWidthCalc](#), [14](#)
- [rqcReadWidthPlot](#), [15](#)

- [rqcCycleBaseCallsCalc](#), [5](#), [6](#), [7](#)
- [rqcCycleBaseCallsLinePlot](#), [6](#)
- [rqcCycleBaseCallsPlot](#), [5](#), [6](#)
- [rqcCycleGCCalc](#), [7](#), [8](#)
- [rqcCycleGCPlot](#), [7](#), [8](#)
- [rqcCycleQualityBoxCalc](#), [9](#), [10](#)
- [rqcCycleQualityBoxPlot](#), [9](#), [10](#)
- [rqcCycleQualityCalc](#), [10](#), [12](#)
- [rqcCycleQualityPlot](#), [11](#), [11](#)
- [rqcQA](#), [3–11](#), [12](#), [13–16](#)
- [rqcReadQualityCalc](#), [13](#), [14](#)
- [rqcReadQualityPlot](#), [13](#), [14](#)
- [rqcReadWidthCalc](#), [14](#), [15](#)
- [rqcReadWidthPlot](#), [15](#), [15](#)
- [rqcReport](#), [16](#)
- [RqcResultSet-class](#), [17](#)

- [ShortRead](#), [12](#)

- [BiocParallel](#), [12](#)

- [bplapply](#), [12](#)

- [ggplot](#), [4](#), [6–8](#), [10](#), [11](#), [14](#), [15](#)

- [Rqc \(Rqc-package\)](#), [2](#)

- [rqc](#), [2](#), [3–16](#)

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

- [rqcCycleAverageQualityCalc](#), [3](#), [4](#)

- [rqcCycleAverageQualityPlot](#), [4](#), [4](#)