

# Package ‘Rqc’

May 10, 2024

**Type** Package

**Title** Quality Control Tool for High-Throughput Sequencing Data

**Version** 1.38.0

**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 (>= 0.25.1), Biostrings, IRanges, methods, S4Vectors, knitr (>= 1.7), BiocStyle, plyr, markdown, grid, reshape2, Rcpp (>= 0.11.6), biovizBase, shiny, Rsamtools, GenomicAlignments, GenomicFiles

**LinkingTo** Rcpp

**Suggests** rmarkdown, testthat

**VignetteBuilder** knitr

**biocViews** Sequencing, QualityControl, DataImport

**URL** <https://github.com/labbcB/Rqc>

**BugReports** <https://github.com/labbcB/Rqc/issues>

**RoxygenNote** 7.1.1

**Encoding** UTF-8

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

**git\_branch** RELEASE\_3\_19

**git\_last\_commit** 68a2092

**git\_last\_commit\_date** 2024-04-30

**Repository** Bioconductor 3.19

**Date/Publication** 2024-05-09

## Contents

|   |           |
|---|-----------|
| Rqc-package . . . . .                     | 2         |
| checkpoint . . . . .                      | 3         |
| detectFileFormat . . . . .                | 4         |
| fromRRDNA . . . . .                       | 4         |
| matdist . . . . .                         | 5         |
| rqc . . . . .                             | 6         |
| rqcCycleAverageQualityCalc . . . . .      | 8         |
| rqcCycleAverageQualityPcaCalc . . . . .   | 9         |
| rqcCycleBaseCallsCalc . . . . .           | 10        |
| rqcCycleGCCalc . . . . .                  | 11        |
| rqcCycleQualityBoxCalc . . . . .          | 12        |
| rqcCycleQualityCalc . . . . .             | 13        |
| rqcFileHeatmap . . . . .                  | 14        |
| rqcGroupCycleAverageQualityCalc . . . . . | 15        |
| rqcQA . . . . .                           | 16        |
| rqcReadFrequencyCalc . . . . .            | 18        |
| rqcReadQualityBoxCalc . . . . .           | 19        |
| rqcReadQualityCalc . . . . .              | 20        |
| rqcReadWidthCalc . . . . .                | 21        |
| rqcReport . . . . .                       | 22        |
| RqcResultSet-class . . . . .              | 23        |
| rqcShinyReport . . . . .                  | 25        |
| stats4trim . . . . .                      | 26        |
| subsetByGroup . . . . .                   | 27        |
| subsetByPair . . . . .                    | 27        |
| toRRDNA . . . . .                         | 28        |
| <b>Index</b>                              | <b>30</b> |

---

Rqc-package

*Quality Control Tool for High-Throughput Sequencing Data*

---

### 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", pair=c(1,1), workers=1)
```

---

checkpoint

*Save time storing longer analysis step on disk*

---

**Description**

This utility function can be used to save time on task that takes long time to complete. A Rda file are written on disk containing only objects setted to keep. If checkpoint function find related Rda file then this Rda will be loaded.

**Usage**

```
checkpoint(
  label,
  CODE,
  path = ".",
  overwrite = FALSE,
  verbose = FALSE,
  keep = NULL
)
```

**Arguments**

|           |   |
|-----------|---|
| label     | name of this code, will create a Rda file with the same name. |
| CODE      | R code.   |
| path      | directory to write/load Rda file.                             |
| overwrite | Rerun CODE and replace Rda file.                              |
| verbose   | argument passed to load function                              |
| keep      | vector of object/variable name to keep. NULL means error.     |

**Value**

Nothing.

**Note**

Experimental function.

**Author(s)**

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
```

---

detectFileFormat      *Detect file format*

---

**Description**

Detect file format

**Usage**

```
detectFileFormat(file)
```

**Arguments**

file                    file name

**Value**

FastqFile or BamFiles objects

**Examples**

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(full.names=TRUE, path=folder)
input <- lapply(files, detectFileFormat)
sapply(input, class)
```

---

fromRRDNA              *Revert codified DNA sequences to original DNA sequences.*

---

**Description**

This function receives a vector of strings containing codified DNA and returns a vector of string containing original DNA sequences.

**Usage**

```
fromRRDNA(rrdnas)
```

**Arguments**

rrdnas            Vector of codified DNA (character vector).

**Value**

Vector of original DNA sequences (character vector).

**Note**

This function is used internally to restore original DNA sequences stored in RqcResultSet objects (per file top reads).

**Author(s)**

Welliton Souza

**See Also**

[perFileTopReads](#)

**Examples**

```
dna <- "ATCG"  
dna.converted <- toRRDNA(dna)  
dna.reverted <- fromRRDNA(dna.converted)  
all.equal(dna, dna.reverted)
```

---

matdist

*Distance matrix of the similarity between the DNA sequences.*

---

**Description**

This function receives a vector of strings representing codified DNA sequences and returns a integer matrix representing the similarities between all sequences from input vectors.

**Usage**

```
matdist(rrdnas)
```

**Arguments**

rrdnas            Vector of codified DNA sequences (character vector).

**Value**

Matrix  $n \times n$ , where  $n$  is the length of the largest original DNA sequence.

**Note**

This function is used internally to compute data for `rqcFileHeatmap` function.

**Author(s)**

Welliton Souza

**See Also**

[rqcFileHeatmap](#)

**Examples**

```
dna1 <- toRRDNA("atcgn")
dna2 <- toRRDNA("atcga")
matdist(c(dna1, dna2))
```

---

rqc

*Main Rqc function*

---

**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 = 1e+06,
  group = NULL,
  top = 10,
  pair = NULL,
  outdir = tempdir(),
  file = "rqc_report",
  openBrowser = TRUE,
  workers = multicoreWorkers()
)
```

**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.                |
| group       | group name for each input file.  |
| top         | number of top over-represented reads. Default is 10 reads.   |
| pair        | combination of files for paired-end reads. By default, all input files are treated as single-end. For paired-end, please define a vector of numbers where two index with the same value represent a pair. Examples, single-end c(1,2,3,4) and paired-end c(1,1,2,2). |
| 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.   |
| workers     | Number of parallel workers. Set 1 to serial. Default value from <a href="#">multicoreWorkers</a> .   |

**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", pair=c(1,1), workers=1, openBrowser=FALSE)
```

---

`rqcCycleAverageQualityCalc`*Per cycle average quality by files*

---

**Description**

This function plots line graph of per cycle average quality.

**Usage**

```
rqcCycleAverageQualityCalc(rqcResultSet)
```

```
rqcCycleAverageQualityPlot(rqcResultSet)
```

**Arguments**

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

**Value**

ggplot2 object

**Functions**

- `rqcCycleAverageQualityCalc`: calculates necessary statistics

**Author(s)**

Welliton Souza

**See Also**

[rqcGroupCycleAverageQualityPlot](#) plots cycle-specific quality by groups

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcCycleAverageQualityPlot(rqcResultSet)
```



---

`rqcCycleAverageQualityPcaCalc`*Biplot of PCA of per cycle read average quality*

---

**Description**

This function creates a Biplot of PCA of per cycle read average quality

**Usage**

```
rqcCycleAverageQualityPcaCalc(rqcResultSet)
```

```
rqcCycleAverageQualityPcaPlot(rqcResultSet)
```

**Arguments**

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

**Value**

Plot object from `ggplot` function.

**Functions**

- `rqcCycleAverageQualityPcaCalc`: calculates necessary statistics

**Author(s)**

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcCycleAverageQualityPcaPlot(rqcResultSet)
```

---

`rqcCycleBaseCallsCalc` *Per cycle base calls plot*

---

### Description

Creates a bar graph of per cycle base calls.

### Usage

```
rqcCycleBaseCallsCalc(rqcResultSet)

rqcCycleBaseCallsLinePlot(rqcResultSet)

rqcCycleBaseCallsPlot(rqcResultSet)
```

### Arguments

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

### Value

Plot object from `ggplot` function.

### Functions

- `rqcCycleBaseCallsCalc`: calculates necessary statistics
- `rqcCycleBaseCallsLinePlot`: creates a line graph

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
rqcCycleBaseCallsPlot(rqcResultSet)
```

---

|                |                                     |
|----------------|-------------------------------------|
| rqcCycleGCCalc | <i>Per cycle percentual GC plot</i> |
|----------------|-------------------------------------|

---

### Description

Creates a line graph of per cycle percentual GC.

### Usage

```
rqcCycleGCCalc(rqcResultSet)
```

```
rqcCycleGCPlot(rqcResultSet)
```

### Arguments

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

### Value

Plot object from `ggplot` function.

### Functions

- `rqcCycleGCCalc`: calculates necessary statistics

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
rqcCycleGCPlot(rqcResultSet)
```

rqcCycleQualityBoxCalc

*Per cycle quality box plot*

---

### Description

Plots per cycle quality box plot.

### Usage

```
rqcCycleQualityBoxCalc(rqcResultSet)
```

```
rqcCycleQualityBoxPlot(rqcResultSet)
```

### Arguments

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

### Value

Plot object from [ggplot](#) function.

### Functions

- [rqcCycleQualityBoxCalc](#): calculates necessary statistics

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcCycleQualityBoxPlot(rqcResultSet)
```

---

`rqcCycleQualityCalc`    *Per cycle quality plot*

---

**Description**

Creates a graph of per cycle quality.

**Usage**

```
rqcCycleQualityCalc(rqcResultSet)
```

```
rqcCycleQualityPlot(rqcResultSet)
```

**Arguments**

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

**Value**

Plot object from `ggplot` function.

**Functions**

- `rqcCycleQualityCalc`: calculates necessary statistics

**Author(s)**

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, workers=1)  
}, keep="rqcResultSet")  
rqcCycleQualityPlot(rqcResultSet)
```

---

|                |   |
|----------------|---|
| rqcFileHeatmap | <i>Heatmap of distance matrix of top over-represented reads</i> |
|----------------|---|

---

### Description

This function plots a heatmap of distance matrix of top over-represented reads. This function does not work with list of RqcResultSet objects, only with one RqcResultSet object.

### Usage

```
rqcFileHeatmap(  
  rqcResultSet,  
  dist.method = "euclidean",  
  hclust.method = "ward.D"  
)
```

### Arguments

`rqcResultSet` RqcResultSet object created by `rqc` and `rqcQA` functions.  
`dist.method` the distance measure to be used by `dist` function.  
`hclust.method` the agglomeration method to be used by `hclust` function.

### Value

Plot object from `ggplot` function.

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcFileHeatmap(rqcResultSet[[1]])
```

---

`rqcGroupCycleAverageQualityCalc`*Per group average quality across cycles*

---

**Description**

This function plots cycle-specific quality by groups

**Usage**

```
rqcGroupCycleAverageQualityCalc(rqcResultSet)
```

```
rqcGroupCycleAverageQualityPlot(rqcResultSet)
```

**Arguments**

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

**Value**

ggplot2 object

**Functions**

- `rqcGroupCycleAverageQualityCalc`: calculates necessary statistics

**Author(s)**

Welliton Souza

**See Also**

[rqcCycleAverageQualityPlot](#) plots cycle-specific quality by files

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
rqcGroupCycleAverageQualityPlot(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(
  x,
  sample = TRUE,
  n = 1e+06,
  group = rep("None", length(x)),
  top = 10,
  pair = seq_along(x),
  ...
)

## S4 method for signature 'list'
rqcQA(x, sample, n, group, top, pair, workers = multicoreWorkers())

## S4 method for signature 'character'
rqcQA(
  x,
  sample = TRUE,
  n = 1e+06,
  group = rep("None", length(x)),
  top = 10,
  pair = seq_along(x),
  workers = multicoreWorkers()
)

## S4 method for signature 'BamFile'
rqcQA(x, sample, n, group, top, pair)

## S4 method for signature 'FastqFile'
rqcQA(x, sample, n, group, top, pair)

```

### Arguments

|        |  |
|--------|--|
| x      | input file(s)  |
| 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. |



|         |  |
|---------|--|
| group   | group name for each input file.  |
| top     | number of top over-represented reads. Default is 10 reads.   |
| pair    | combination of files for paired-end reads. By default, all input files are treated as single-end. For paired-end, please define a vector of numbers where two index with the same value represent a pair. Examples, single-end c(1,2,3,4) and paired-end c(1,1,2,2). |
| ...     | other parameters   |
| workers | number of parallel workers   |

### 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.

### Methods (by class)

- `list`: process a list of `FastqFile` and `BamFile` objects.
- `character`: automatically detects file format (using `detectFileFormat` function) of input files then process.
- `BamFile`: process only one BAM file.
- `FastqFile`: process only one FASTQ file.

### Author(s)

Welliton Souza

### See Also

[rqc](#)

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
rqcReadQualityPlot(rqcResultSet)
```

---

rqcReadFrequencyCalc *Read frequency plot*

---

### Description

This function creates a bar graph of read frequency (in percentage).

### Usage

```
rqcReadFrequencyCalc(rqcResultSet)
```

```
rqcReadFrequencyPlot(rqcResultSet)
```

### Arguments

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

### Value

Plot object from `ggplot` function.

### Functions

- `rqcReadFrequencyCalc`: calculates necessary statistics

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcReadFrequencyPlot(rqcResultSet)
```

---

`rqcReadQualityBoxCalc` *Per read mean quality box plot*

---

### Description

This function creates a graphic chart with box plots describing per read mean quality distribution for each input file

### Usage

```
rqcReadQualityBoxCalc(rqcResultSet)
```

```
rqcReadQualityBoxPlot(rqcResultSet)
```

### Arguments

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

### Value

Plot object from `ggplot` function.

### Functions

- `rqcReadQualityBoxCalc`: calculates necessary statistics

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcReadQualityBoxPlot(rqcResultSet)
```

---

rqcReadQualityCalc     *Per read quality plot*

---

### Description

Plots the quality of all the files by read.

### Usage

```
rqcReadQualityCalc(rqcResultSet)
```

```
rqcReadQualityPlot(rqcResultSet)
```

### Arguments

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

### Value

Plot object from [ggplot](#) function.

### Functions

- [rqcReadQualityCalc](#): calculates necessary statistics

### Author(s)

Welliton Souza

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
rqcReadQualityPlot(rqcResultSet)
```

---

|                  |                            |
|------------------|----------------------------|
| rqcReadWidthCalc | <i>Per read width plot</i> |
|------------------|----------------------------|

---

**Description**

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

**Usage**

```
rqcReadWidthCalc(rqcResultSet)
```

```
rqcReadWidthPlot(rqcResultSet)
```

**Arguments**

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

**Value**

Plot object from `ggplot` function.

**Functions**

- `rqcReadWidthCalc`: calculates necessary statistics

**Author(s)**

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
rqcReadWidthPlot(rqcResultSet)
```

---

`rqcReport`*Quality Control HTML Report*

---

### Description

Generates an HTML report file.

### Usage

```
rqcReport(  
  rqcResultSet,  
  outdir = tempdir(),  
  file = "rqc_report",  
  keepMD = FALSE,  
  templateFile = system.file("templates", package = "Rqc", "rqc_report.Rmd")  
)
```

### Arguments

|                           |  |
|---------------------------|--|
| <code>rqcResultSet</code> | list of <code>RqcResultSet</code> objects created by <code>rqc</code> and <code>rqcQA</code> functions.  |
| <code>outdir</code>       | output directory path. It is created a temporary directory by default.   |
| <code>file</code>         | output file name.  |
| <code>keepMD</code>       | If true <code>Rqc</code> does not delete markdown file. <code>knit</code> function takes RMarkdown template file (within package) and generates a temporary Markdown file. Next <code>markdownToHTML</code> function takes this markdown file and creates final HTML file. |
| <code>templateFile</code> | Path of Rmarkdown file as <code>Rqc</code> web report template.  |

### Details

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

### Value

Report file path.

### Author(s)

Welliton Souza

### See Also

[rqc](#)  
[rqcQA](#)

**Examples**

```

options(device.ask.default = FALSE)
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
reportFile <- rqcReport(rqcResultSet)
browseURL(reportFile)

```

---

RqcResultSet-class      *Class RqcResultSet*

---

**Description**

Class RqcResultSet  
 Frequency distribution of cycle-specific base call  
 Frequency distribution of cycle-specific quality  
 File information  
 Top over-represented sequencing reads  
 Read frequency table  
 Frequency distribution of per read mean quality  
 Frequency distribution of read width

**Usage**

```

perCycleBasecall(x)

## S4 method for signature 'RqcResultSet'
perCycleBasecall(x)

## S4 method for signature 'list'
perCycleBasecall(x)

perCycleQuality(x)

## S4 method for signature 'RqcResultSet'
perCycleQuality(x)

## S4 method for signature 'list'
perCycleQuality(x)

perFileInformation(x)

```

```
## S4 method for signature 'RqcResultSet'  
perFileInformation(x)  
  
## S4 method for signature 'list'  
perFileInformation(x)  
  
perFileTopReads(x)  
  
## S4 method for signature 'RqcResultSet'  
perFileTopReads(x)  
  
## S4 method for signature 'list'  
perFileTopReads(x)  
  
perReadFrequency(x)  
  
## S4 method for signature 'RqcResultSet'  
perReadFrequency(x)  
  
## S4 method for signature 'list'  
perReadFrequency(x)  
  
perReadQuality(x)  
  
## S4 method for signature 'RqcResultSet'  
perReadQuality(x)  
  
## S4 method for signature 'list'  
perReadQuality(x)  
  
perReadWidth(x)  
  
## S4 method for signature 'RqcResultSet'  
perReadWidth(x)  
  
## S4 method for signature 'list'  
perReadWidth(x)
```

**Arguments**

x                    RqcResultSet object or list of RqcResultSet objects

**Value**

data frame  
data frame  
data frame  
data frame



data frame  
data frame  
data frame

### Examples

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {  
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")  
  files <- list.files(full.names=TRUE, path=folder)  
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)  
}, keep="rqcResultSet")  
head(perCycleBasecall(rqcResultSet))  
head(perCycleQuality(rqcResultSet))  
head(perReadFrequency(rqcResultSet))  
head(perReadQuality(rqcResultSet))  
head(perReadWidth(rqcResultSet))  
perFileInformation(rqcResultSet)  
perFileTopReads(rqcResultSet)
```

---

rqcShinyReport

*Interactive Quality Control Report*

---

### Description

This function runs a Shiny web application of interactive Rqc report. This is useful for large amount of files and sample groups.

### Usage

```
rqcShinyReport(rqcResultSet)
```

### Arguments

rqcResultSet list of [RqcResultSet-class](#) objects

### Value

function

### Author(s)

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
# rqcShinyReport(rqcResultSet)
```

---

stats4trim

*Minimum read mean quality and maximum percentage loss of reads estimations for trimming step.*


---

**Description**

This function estimates how many reads would be lost if the sequences are filtered by a minimum read mean quality value. Also this function estimates what is the minimum read mean quality value for filtering and lose max percentage defined.

**Usage**

```
stats4trim(rqcResultSet, qmin, pmax)
```

**Arguments**

**rqcResultSet** list of RqcResultSet objects created by `rqc` and `rqcQA` functions.  
**qmin** Minimum read mean quality value (between 0 and 41).  
**pmax** Maximum percentage of reads permitted been lost during trimming step.

**Value**

A data frame containing estimated minimum quality and maximum percentage for each input file.

**Author(s)**

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
stats4trim(rqcResultSet, qmin=20)
stats4trim(rqcResultSet, pmax=10)
```

---

|               |  |
|---------------|--|
| subsetByGroup | <i>Subset RqcResultSet object by group name.</i> |
|---------------|--|

---

**Description**

This function subsets RqcResultSet object function by group name.

**Usage**

```
subsetByGroup(rqcResultSet, group)
```

**Arguments**

|              |  |
|--------------|--|
| rqcResultSet | list of RqcResultSet objects created by <a href="#">rqc</a> and <a href="#">rqcQA</a> functions. |
| group        | Name of the group to subset  |

**Value**

list of RqcResultSet objects from only one group.

**Author(s)**

Welliton Souza

**Examples**

```
folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
files <- list.files(full.names=TRUE, path=folder)
rqcResultSet <- rqcQA(files, workers=1, group=c("a", "b"))
perFileInformation(subsetByGroup(rqcResultSet, "a"))
```

---

|              |  |
|--------------|--|
| subsetByPair | <i>Subset RqcResultSet object by pair files.</i> |
|--------------|--|

---

**Description**

This function subsets RqcResultSet object function by pair files.

**Usage**

```
subsetByPair(rqcResultSet, pair)
```

**Arguments**

|              |  |
|--------------|--|
| rqcResultSet | list of RqcResultSet objects created by <a href="#">rqc</a> and <a href="#">rqcQA</a> functions. |
| pair         | index of the pair  |

**Value**

list of RqcResultSet objects from only one pair.

**Author(s)**

Welliton Souza

**Examples**

```
checkpoint("Rqc", path=system.file(package="Rqc", "extdata"), {
  folder <- system.file(package="ShortRead", "extdata/E-MTAB-1147")
  files <- list.files(full.names=TRUE, path=folder)
  rqcResultSet <- rqcQA(files, pair=c(1,1), workers=1)
}, keep="rqcResultSet")
perFileInformation(subsetByPair(rqcResultSet, 1))
```

---

toRRDNA

*Title: Convert DNA sequences to Reduced Representation format*

---

**Description**

This function receives a vector of strings (character vector) containing DNA sequences and returns a vector of strings containing codified DNA.

**Usage**

```
toRRDNA(dnas)
```

**Arguments**

dnas                    Vector of DNA sequences (character vector).

**Value**

Vector of DNA converted to reduced representation format (character vector).

**Note**

This function is used internally to compute top over-represented reads and to store in RqcResultSet objects (per file top reads).

**Author(s)**

Welliton Souza

**See Also**

[perFileTopReads](#)

**Examples**

```
dna <- "ATCGNATCGTA"  
dna.converted <- toRRDNA(dna)  
nchar(dna)  
nchar(dna.converted)
```

# Index

## \* **graphics**

- [rqc](#), [6](#)
- [rqcCycleAverageQualityCalc](#), [8](#)
- [rqcCycleAverageQualityPcaCalc](#), [9](#)
- [rqcCycleBaseCallsCalc](#), [10](#)
- [rqcCycleGCCalc](#), [11](#)
- [rqcCycleQualityBoxCalc](#), [12](#)
- [rqcCycleQualityCalc](#), [13](#)
- [rqcFileHeatmap](#), [14](#)
- [rqcGroupCycleAverageQualityCalc](#),  
[15](#)
- [rqcReadFrequencyCalc](#), [18](#)
- [rqcReadQualityBoxCalc](#), [19](#)
- [rqcReadQualityCalc](#), [20](#)
- [rqcReadWidthCalc](#), [21](#)

## \* **package**

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

## \* **qc**

- [rqc](#), [6](#)
- [rqcCycleAverageQualityCalc](#), [8](#)
- [rqcCycleAverageQualityPcaCalc](#), [9](#)
- [rqcCycleBaseCallsCalc](#), [10](#)
- [rqcCycleGCCalc](#), [11](#)
- [rqcCycleQualityBoxCalc](#), [12](#)
- [rqcCycleQualityCalc](#), [13](#)
- [rqcFileHeatmap](#), [14](#)
- [rqcGroupCycleAverageQualityCalc](#),  
[15](#)
- [rqcReadFrequencyCalc](#), [18](#)
- [rqcReadQualityBoxCalc](#), [19](#)
- [rqcReadQualityCalc](#), [20](#)
- [rqcReadWidthCalc](#), [21](#)
- [stats4trim](#), [26](#)
- [subsetByGroup](#), [27](#)
- [subsetByPair](#), [27](#)

## \* **trimming**

- [stats4trim](#), [26](#)

[BamFile](#), [17](#)

[BiocParallel](#), [17](#)

[bplapply](#), [17](#)

[checkpoint](#), [3](#)

[detectFileFormat](#), [4](#), [17](#)

[dist](#), [14](#)

[FastqFile](#), [17](#)

[fromRRDNA](#), [4](#)

[ggplot](#), [9–14](#), [18–21](#)

[hclust](#), [14](#)

[knit](#), [22](#)

[markdownToHTML](#), [22](#)

[matdist](#), [5](#)

[multicoreWorkers](#), [7](#)

[perCycleBasecall](#) ([RqcResultSet-class](#)),  
[23](#)

[perCycleBasecall](#), list-method  
([RqcResultSet-class](#)), [23](#)

[perCycleBasecall](#), [RqcResultSet-method](#)  
([RqcResultSet-class](#)), [23](#)

[perCycleQuality](#) ([RqcResultSet-class](#)), [23](#)

[perCycleQuality](#), list-method  
([RqcResultSet-class](#)), [23](#)

[perCycleQuality](#), [RqcResultSet-method](#)  
([RqcResultSet-class](#)), [23](#)

[perFileInformation](#)  
([RqcResultSet-class](#)), [23](#)

[perFileInformation](#), list-method  
([RqcResultSet-class](#)), [23](#)

[perFileInformation](#), [RqcResultSet-method](#)  
([RqcResultSet-class](#)), [23](#)

[perFileTopReads](#), [5](#), [28](#)

[perFileTopReads](#) ([RqcResultSet-class](#)), [23](#)

[perFileTopReads](#), list-method  
([RqcResultSet-class](#)), [23](#)

- perFileTopReads, RqcResultSet-method (RqcResultSet-class), 23
- perReadFrequency (RqcResultSet-class), 23
- perReadFrequency, list-method (RqcResultSet-class), 23
- perReadFrequency, RqcResultSet-method (RqcResultSet-class), 23
- perReadQuality (RqcResultSet-class), 23
- perReadQuality, list-method (RqcResultSet-class), 23
- perReadQuality, RqcResultSet-method (RqcResultSet-class), 23
- perReadWidth (RqcResultSet-class), 23
- perReadWidth, list-method (RqcResultSet-class), 23
- perReadWidth, RqcResultSet-method (RqcResultSet-class), 23
  
- Rqc (Rqc-package), 2
- rqc, 6, 8–15, 17–22, 26, 27
- Rqc-package, 2
- rqcCycleAverageQualityCalc, 8
- rqcCycleAverageQualityPcaCalc, 9
- rqcCycleAverageQualityPcaPlot (rqcCycleAverageQualityPcaCalc), 9
- rqcCycleAverageQualityPlot, 15
- rqcCycleAverageQualityPlot (rqcCycleAverageQualityCalc), 8
- rqcCycleBaseCallsCalc, 10
- rqcCycleBaseCallsLinePlot (rqcCycleBaseCallsCalc), 10
- rqcCycleBaseCallsPlot (rqcCycleBaseCallsCalc), 10
- rqcCycleGCCalc, 11
- rqcCycleGCPlot (rqcCycleGCCalc), 11
- rqcCycleQualityBoxCalc, 12
- rqcCycleQualityBoxPlot (rqcCycleQualityBoxCalc), 12
- rqcCycleQualityCalc, 13
- rqcCycleQualityPlot (rqcCycleQualityCalc), 13
- rqcFileHeatmap, 6, 14
- rqcGroupCycleAverageQualityCalc, 15
- rqcGroupCycleAverageQualityPlot, 8
- rqcGroupCycleAverageQualityPlot (rqcGroupCycleAverageQualityCalc), 15
  
- rqcQA, 7–15, 16, 18–22, 26, 27
- rqcQA, BamFile-method (rqcQA), 16
- rqcQA, character-method (rqcQA), 16
- rqcQA, FastqFile-method (rqcQA), 16
- rqcQA, list-method (rqcQA), 16
- rqcReadFrequencyCalc, 18
- rqcReadFrequencyPlot (rqcReadFrequencyCalc), 18
- rqcReadQualityBoxCalc, 19
- rqcReadQualityBoxPlot (rqcReadQualityBoxCalc), 19
- rqcReadQualityCalc, 20
- rqcReadQualityPlot (rqcReadQualityCalc), 20
- rqcReadWidthCalc, 21
- rqcReadWidthPlot (rqcReadWidthCalc), 21
- rqcReport, 22
- RqcResultSet-class, 23
- rqcShinyReport, 25
  
- ShortRead, 17
- stats4trim, 26
- subsetByGroup, 27
- subsetByPair, 27
  
- toRRDNA, 28