

Package ‘systemPipeRdata’

April 1, 2025

Title systemPipeRdata: Workflow templates and sample data

Version 2.10.0

Date 2024-10-19

Author Thomas Girke

Maintainer Thomas Girke <thomas.girke@ucr.edu>

biocViews Genetics, Infrastructure, DataImport, Sequencing, RNASeq, ChIPSeq, MethylSeq, SNP, GeneExpression, Coverage, GeneSetEnrichment, Alignment, QualityControl, ImmunoOncology, RiboSeq, WorkflowStep

Description systemPipeRdata complements the systemPipeR workflow management system (WMS) by offering a collection of pre-designed data analysis workflow templates. These templates are easily accessible and can be readily loaded onto a user's system with a single command. Once loaded, the WMS can immediately utilize these templates for efficient end-to-end analysis, serving a wide range of data analysis needs.

Depends R (>= 3.6.0)

Imports methods, Biostrings, BiocGenerics, jsonlite, remotes

Suggests GenomicFeatures, GenomicRanges, IRanges, Rsamtools, ShortRead, rtracklayer, RUnit, BiocStyle, knitr, rmarkdown, systemPipeR, kableExtra, magrittr, dplyr

VignetteBuilder knitr

License Artistic-2.0

NeedsCompilation no

URL <https://github.com/tgirke/systemPipeRdata>, <https://systempipe.org/>

git_url <https://git.bioconductor.org/packages/systemPipeRdata>

git_branch RELEASE_3_20

git_last_commit 04d9bb0

git_last_commit_date 2024-10-29

Repository Bioconductor 3.20

Date/Publication 2025-04-01

Contents

| | |
|--------------------------|----------|
| availableWF | 2 |
| genWorkenvir | 3 |
| getSubsetReads | 4 |
| pathList | 5 |
| Index | 7 |

| | |
|-------------|--|
| availableWF | <i>List Available Workflows Templates at systemPipeRdata package</i> |
|-------------|--|

Description

This function checks the workflow templates availability from systemPipeRdata package and also from [systemPipeR Organization](#) on GitHub.

Usage

```
availableWF(github = FALSE)
```

Arguments

| | |
|--------|---|
| github | logical. If TRUE, it will return current experimental workflow templates available on systemPipeR Organization. |
|--------|---|

Details

Internally, this function uses the GitHub API, and there is an access limit per hour. For more details, please check: `system("curl -i https://api.github.com/users/<username>")`.

Value

Workflow templates are printed on the console, and also return an invisible list with the names of the workflows templates available at systemPipeRdata package. If `github = TRUE`, it will return an additional data.frame with current workflow templates available on systemPipeR Organization in the list.

Note

We are assuming that workflow templates repositories under [systemPipeR Organization](#) content the keyword "Workflow Template" on the Description section and "Topics" section, we expected "systempiper" and "release" or "development" words.

Author(s)

Daniela Cassol

See Also

[genWorkenvir](#).

Examples

```
availableWF()
## Not run:
## List Workflow Templates from \code{systemPipeR} Organization
availableWF(github = TRUE)

## End(Not run)
```

genWorkenvir

Generate workflow templates

Description

Generates workflow templates for systemPipeR package. The template environments contain a predefined directory structure along with run parameter files and sample data. The structure of the workflow templates and the sample data are described in all details in the Overview Vignette of the [systemPipeR package](#).

Usage

```
genWorkenvir(workflow, mydirname=NULL, bam=FALSE)
```

Arguments

| | |
|-----------|--|
| workflow | character string of workflow templates to be generated. Supported values can be checked with the <code>\link{availableWF}()</code> function. |
| mydirname | Specifies the name of the workflow directory. The default NULL uses the name of the chosen workflow. An error is issued if a directory of the same name and path exists already. |
| bam | If bam=TRUE pregenerated short read alignment (BAM) files will be included in the results directory of the workflow environment. Note, these BAM files have been generated with the HISAT2 aligner using the FASTQ files provided in the data directory. The default bam=FALSE omits this step meaning no BAM files will be copied into the results directory. |

Details

Check the output of `\link{availableWF}()` to the current workflow templates available on systemPipeR Organization.

For an `interactive()` session, the `readline()` function provides the option choose between proceeding or not, through options: yes or no. For non-interactive use, if there is no package install, the option yes will be selected.

Value

Workflow directory containing sample data and parameter files along with the following subdirectories:

| | |
|--------|------------------------|
| param/ | stores parameter files |
| data/ | stores input data |

results/ stores output results

For more details, please consult the Overview Vignette (HTML) of the systemPipeR package (<http://bioconductor.org/packages/systemPipeR>).

Author(s)

Thomas Girke and Daniela Cassol

Examples

```
## Return location of sample data
samplepaths <- pathList()
## Not run:
## Generate varseq workflow environment
genWorkenvir(workflow="varseq", mydirname=NULL)
setwd("varseq")

## End(Not run)
```

getSubsetReads *Subsetting fastq data*

Description

Returns subsets of fastq files data based on specific mapping regions or list of genes or GRanges object.

Usage

```
getSubsetReads(args,
  geneList = NULL,
  gr = NULL,
  MappingRegion = 1:1e+05,
  sample_range = 90000:1e+05,
  truncate_refs = TRUE,
  id_read_number = TRUE,
  annotation = "data/tair10.gff",
  reference = "data/tair10.fasta",
  annot_outname = "tair10_sub.gff",
  ref_outname = "tair10_sub.fasta",
  outdir = "data/subset/",
  silent = FALSE
)
```

Arguments

| | |
|---------------|--|
| args | object of class SYSargs2. |
| geneList | selected genes list to retrieve the reads from the fastq file. |
| gr | an object containing genomic ranges to retrieve the reads from the fastq file. |
| MappingRegion | integers ranges of start and end of chromosome position to retrieve the reads from the fastq file. |

| | |
|----------------|--|
| sample_range | random range to subsetted the fastq file. |
| truncate_refs | logical. If TRUE it will generate reference genome and annotation subset file. |
| id_read_number | if fastq file contains sequence name with read number (<code>\$ri --define-seq '@\$sn[_\$rn]/\$ri '</code>). |
| annotation | path to annotation file. |
| reference | path to reference genome. |
| annot_outname | character name of the annotation output file. |
| ref_outname | character name of the reference genome output file. |
| outdir | path to output directory. |
| silent | if set to TRUE, all messages returned by the function will be suppressed. |

Value

Workflow directory containing sample data and parameter files along with the following subdirectories:

| | |
|----------|------------------------|
| param/ | stores parameter files |
| data/ | stores input data |
| results/ | stores output results |

For more details, please consult the Overview Vignette (HTML) of the systemPipeR package (<http://bioconductor.org/packages/systemPipeR>).

Author(s)

Thomas Girke, Shiyuan Guo and Daniela Cassol

Examples

```
## Not run:
getSubsetReads(args, MappingRegion = 1:900, sample_range = 800:900, outdir = "data/subset/", silent = FALSE)
getSubsetReads(args, MappingRegion = 1:900, sample_range = NULL, outdir = "data/subset/", silent = FALSE)

## End(Not run)
```

| | |
|----------|---------------------------------------|
| pathList | <i>Return location of sample data</i> |
|----------|---------------------------------------|

Description

Function to return paths to sample data provided by sytemPipeRdata package.

Usage

```
pathList()
```

Value

list

Author(s)

Thomas Girke

Examples

```
samplepaths <- pathList()
```

Index

* **utilities**

- availableWF, [2](#)
- genWorkenvir, [3](#)
- getSubsetReads, [4](#)
- pathList, [5](#)

availableWF, [2](#)

genWorkenvir, [2, 3](#)
getSubsetReads, [4](#)

pathList, [5](#)