

qrqc

October 25, 2011

FASTASummary-class *'FASTASummary'* class representing the summaries of a FASTA file

Description

This class contains the same slots as the [SequenceSummary](#), but it is used to indicate the data originated from a FASTA file.

Slots

[FASTASummary](#) has the slots inherited from [SequenceSummary](#).

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

See Also

[FASTQSummary](#) is the counterpart of this class for FASTQ data.

[readSeqFile](#) is the function that takes a FASTA file and returns a [FASTASummary](#) object.

[plotBases](#) is a function that plots the distribution of bases over sequence length for a particular [FASTASummary](#) object. [plotGC](#) combines and plots the GC proportion.

[plotSeqLengths](#) is a function that plots a histogram of sequence lengths for a particular [FASTASummary](#) object.

Examples

```
showClass("FASTASummary")
```

```
FASTQSummary-class 'FASTQSummary' class representing the summaries of a FASTQ file
```

Description

This class contains the same slots as the [SequenceSummary](#), as well as additional slots for quality information.

Slots

In addition to the slots inherited from [SequenceSummary](#), [FASTQSummary](#) contains:

`quality` a string indicating the type of quality (used to convert ASCII characters to quality integers). Either "phred", "solexa", or "illumina".

`qual.freqs` a dataframe of quality frequencies by position, if the file was a FASTQ file.

`mean.qual` a numeric that is the mean quality across all positions, weighted by the number of reads that extended to that position.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

See Also

[FASTASummary](#) is the counterpart of this class for FASTA data.

[readSeqFile](#) is the function that takes a FASTQ file and returns a [FASTQSummary](#) object.

[plotBases](#) is a function that plots the distribution of bases over sequence length for a particular [FASTQSummary](#) object. [plotGC](#) combines and plots the GC proportion.

[plotQuals](#) is a function that plots the distribution of qualities over sequence length for a particular [FASTQSummary](#) object.

[plotSeqLengths](#) is a function that plots a histogram of sequence lengths for a particular [FASTQSummary](#) object.

Examples

```
showClass("FASTQSummary")
```

```
SequenceSummary-class
      'SequenceSummary' class representing the summaries of a sequence
      file
```

Description

A sequence file read in with [readSeqFile](#) is summarized by a C call. This is a base class with slots common to both [FASTQSummary](#) and [FASTASummary](#). This is not usually instantiated directly.

Slots

`filename` the filename processed by `readSeqFile`.

`base.freqs` a dataframe of base frequencies by position. Each column is a nucleotide (there is a column for position too), and each row contains the count frequencies of bases for that position.

`seq.lengths` a numeric vector of the number of sequences of a particular length (the length is the position in the vector).

`hash` a numeric vector of the count frequencies of sequences (the sequences are in the name attribute).

`hashed` a logical indicating whether the sequences were hashed in `readSeqFile`.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

See Also

[FASTQSummary](#) and [FASTASummary](#) are the classes that inherit from `SequenceSummary`.

[readSeqFile](#) is the function that takes a FASTQ or FASTA file and returns a `FASTQSummary` object or `FASTASummary` object.

Examples

```
showClass("SequenceSummary")
```

`makeReport-methods` *Make an HTML report from a FASTASummary of FASTQSummary object*

Description

`makeReport` takes a `FASTQSummary` or `FASTASummary` object, creates an HTML report, and writes it to a file within a directory. The directory naming is incremental so past reports will not be overwritten.

Usage

```
makeReport(obj, outputDir=".")
```

Arguments

`obj` an object that is either `FASTQSummary` or `FASTASummary`.

`outputDir` an optional character argument to indicate the report output directory. By default, the current directory.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

Examples

```
## Load a FASTQ file, with sequence hashing.
s.fastq <- readSeqFile(system.file('extdata', 'test.fastq', package='qrqc'))

## Make and save a report
makeReport(s.fastq)
```

plotBases-methods *Plot Bases by Position*

Description

plotBases plots the frequency or proportion of bases by position in the read.

plotBases uses the Sanger base color scheme: blue is Cytosine, green is Adenine, black is Guanine, red is Thymine, and purple in N (any base). Other IUPAC nucleotides are colored using **RColorBrewer**.

Usage

```
plotBases(obj, type="freq", bases=NULL, legend=TRUE)
```

Arguments

obj	an S4 object of class that inherits from SequenceSummary (either FASTASummary or FASTQSummary) from readSeqFile.
type	a character string that is either "freq" or "prop" indicating whether to plot frequencies or proportions on the y-axis.
bases	a vector of characters indicating which bases to include. The default value NULL indicates to plot <code>_all_</code> bases.
legend	a logical value indicating whether to include a legend on the top right.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

Examples

```
## Load a FASTQ file, with sequence hashing.
s.fastq <- readSeqFile(system.file('extdata', 'test.fastq', package='qrqc'))

## Plot base frequencies
plotBases(s.fastq, type="freq")

## Plot base proportions
plotBases(s.fastq, type="prop")
```

plotGC-methods *Plot per Base GC Content by Position*

Description

plotGC plots the GC proportion by position.

Usage

```
plotGC(obj)
```

Arguments

obj an S4 object of class that inherits from `SequenceSummary` (either `FASTASummary` or `FASTQSummary`) from `readSeqFile`.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

Examples

```
## Load a FASTQ file, with sequence hashing.
s.fastq <- readSeqFile(system.file('extdata', 'test.fastq', package='qrgc'))

## Plot Qualities
plotGC(s.fastq)
```

plotQuals-methods *Plot a Base Quality Boxplot by Position*

Description

plotQuals plots quality statistics by position. Optionally, it adds a lowess curve through the qualities, which is fit with data randomly drawn from the distribution of qualities at each position. A histogram of the sequence length distribution is plotted above the quality plot when `histogram` is `TRUE`.

A legend is plotted on the bottom left if `legend` is `TRUE` (this location is used because this where the bases are likely to be of highest quality, and thus not overlap the legend). The grey lines indicate the range of the 10% and 90% quantiles, the orange lines indicate the range of the 25% and 75% quantiles, the blue point is the median, the green dash is the mean, and the purple line is the lowess curve if `lowess` is `TRUE`.

Usage

```
plotQuals(obj, ylim='relative', lowess=TRUE, histogram=TRUE, legend=TRUE)
```

Arguments

obj	an S4 object of class FASTQSummary from readSeqFile.
ylim	either 'relative' or 'fixed', which will scale the y axis to either the relative range (from the data) or absolute range of qualities.
lowess	a logical value indicating whether to fit a lowess curve through the quality plot.
histogram	a logical value indicating whether to add a histogram of the sequence length distribution above the quality plot.
legend	a logical value indicating whether a legend is to be included.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

Examples

```
## Load a FASTQ file, with sequence hashing.
s.fastq <- readSeqFile(system.file('extdata', 'test.fastq', package='qrqc'))

## Plot Qualities
plotQuals(s.fastq)
```

plotSeqLengths-methods

Plot Histogram of Sequence Lengths

Description

plotSeqLengths plots histogram of sequence lengths.

Usage

```
plotSeqLengths(obj)
```

Arguments

obj	an S4 object of class that inherits from SequenceSummary (either FASTASummary or FASTQSummary) from readSeqFile.
-----	-------------------------------------------------------------------------------------------------------------------------------------------------------------------

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

Examples

```
## Load a FASTQ file, with sequence hashing.
s.fastq <- readSeqFile(system.file('extdata', 'test.fastq', package='qrqc'))

## Plot Qualities
plotSeqLengths(s.fastq)
```

`readSeqFile`*Read and Summarize a Sequence (FASTA or FASTQ) File*

Description

`readSeqFile` reads a FASTQ or FASTA file, summarizing the nucleotide distribution across position (cycles) and the sequence length distributions. If `type` is 'fastq', the distribution of qualities across position will also be recorded. If `hash` is TRUE, the unique sequences will be hashed with counts of their frequency.

Usage

```
readSeqFile(filename, type='fastq', max.length=1000,  
            quality='illumina', hash=TRUE, verbose=FALSE)
```

Arguments

<code>filename</code>	the name of the file which the sequences are to be read from.
<code>type</code>	either 'fastq' or 'fasta', representing the type of the file. FASTQ files will have the quality distribution by position summarized.
<code>max.length</code>	the largest sequence length likely to be encountered. For efficiency, a matrix larger than the largest sequence is allocated to *this* size in C, populated, and then trimmed in R. Specifying a value too small will lead to an error and the function will need to be re-run.
<code>quality</code>	either 'illumina', 'phred', or 'solexa', this determines the quality offsets and range. See the values of <code>QUALITY.CONSTANTS</code> for more information.
<code>hash</code>	a logical value indicating whether to hash sequences
<code>verbose</code>	a logical value indicating whether be verbose (in the C backend).

Value

An S4 object of [FASTQSummary](#) or [FASTASummary](#) containing the summary statistics.

Author(s)

Vince Buffalo <vsbuffalo@ucdavis.edu>

See Also

[FASTQSummary](#) and [FASTASummary](#) are the classes of the objects returned by `readSeqFile`.

[plotBases](#) is a function that plots the distribution of bases over sequence length for a particular [FASTASummary](#) or [FASTQSummary](#) object. [plotGC](#) combines and plots the GC proportion.

[plotQuals](#) is a function that plots the distribution of qualities over sequence length for a particular [FASTASummary](#) or [FASTQSummary](#) object.

[plotSeqLengths](#) is a function that plots a histogram of sequence lengths for a particular [FASTASummary](#) or [FASTQSummary](#) object.

Examples

```
## Load a FASTQ file, with sequence hashing.
s.fastq <- readSeqFile(system.file('extdata', 'test.fastq', package='qrqc'))

## Load a FASTA file, without sequence hashing.
s.fasta <- readSeqFile(system.file('extdata', 'test.fasta', package='qrqc'),
                      type='fasta', hash=FALSE)
```

Index

*Topic **classes**

FASTASummary-class, [1](#)
FASTQSummary-class, [2](#)
SequenceSummary-class, [2](#)

*Topic **file**

makeReport-methods, [3](#)
readSeqFile, [7](#)

*Topic **graphics**

plotBases-methods, [4](#)
plotGC-methods, [5](#)
plotQuals-methods, [5](#)
plotSeqLengths-methods, [6](#)

*Topic **methods**

plotQuals-methods, [5](#)

FASTASummary, [1-7](#)

FASTASummary-class, [1](#)

FASTQSummary, [1-7](#)

FASTQSummary-class, [2](#)

makeReport (*makeReport-methods*), [3](#)

makeReport, FASTASummary-method
(*FASTASummary-class*), [1](#)

makeReport, FASTQSummary-method
(*FASTQSummary-class*), [2](#)

makeReport-methods, [3](#)

plotBases, [1, 2, 7](#)

plotBases (*plotBases-methods*), [4](#)

plotBases, SequenceSummary-method
(*SequenceSummary-class*), [2](#)

plotBases-methods, [4](#)

plotGC, [1, 2, 7](#)

plotGC (*plotGC-methods*), [5](#)

plotGC, SequenceSummary-method
(*SequenceSummary-class*), [2](#)

plotGC-methods, [5](#)

plotQuals, [2, 7](#)

plotQuals (*plotQuals-methods*), [5](#)

plotQuals, FASTQSummary-method
(*FASTQSummary-class*), [2](#)

plotQuals-methods, [5](#)

plotSeqLengths, [1, 2, 7](#)

plotSeqLengths

(*plotSeqLengths-methods*), [6](#)

plotSeqLengths, SequenceSummary-method
(*SequenceSummary-class*), [2](#)

plotSeqLengths-methods, [6](#)

readSeqFile, [1-3, 7](#)

SequenceSummary, [1, 2, 4-6](#)

SequenceSummary-class, [2](#)