

Package ‘ebimetagenomics’

September 7, 2017

Type Package

Title EBI Metagenomics Portal

Version 0.3

Date 2017-08-22

Author Darren Wilkinson

Maintainer Darren Wilkinson <darren.wilkinson@nc1.ac.uk>

Description Functions for querying the EBI Metagenomics Portal <<https://www.ebi.ac.uk/metagenomics/>>. The current main focus is on taxa abundance data, but the intention is that this package should evolve into a general purpose package for working with EBI Metagenomics data using R.

License LGPL-3

Depends R (>= 3.0.0), sads, vegan, breakaway, methods

Suggests

Repository CRAN

Repository/R-Forge/Project ebimetagenomics

Repository/R-Forge/Revision 29

Repository/R-Forge/DateTimeStamp 2017-09-07 17:00:52

Date/Publication 2017-09-07 17:26:25 UTC

NeedsCompilation no

R topics documented:

analyseOtu	2
convertOtuTad	3
getProjectsList	3
getProjectSummary	4
getRunOtu	5
getSampleOtu	5
mergeOtu	6
plotOtu	7

projectRuns	8
projectSamples	9
read.otu.tsv	9
read.project.csv	10
runsBySample	11

Index	12
--------------	-----------

analyseOtu	<i>Analysis of OTU count data</i>
------------	-----------------------------------

Description

This function takes as input a data frame representing OTU counts, such as returned by `getRunOtu`, and conducts a range of analyses relating to taxa diversity and coverage estimation.

Usage

```
analyseOtu(otu, plot = TRUE)
```

Arguments

<code>otu</code>	An OTU data frame.
<code>plot</code>	If TRUE, calls <code>plotOtu</code> to generate a plot.

Value

The function returns a numeric vector with named elements representing the results of various analyses.

TODO: Explain output properly...

See Also

[getSampleOtu](#), [getRunOtu](#), [plotOtu](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runs=runsBySample(ps,samp[2])
otu=getRunOtu(ps,runs[1])
analyseOtu(otu,plot=FALSE)
```

convertOtuTad	<i>Converts an OTU count data frame to a TAD data frame</i>
---------------	---

Description

This function takes as input a data frame representing OTU counts, such as returned by `getRunOtu`, and tabulates it in the form of a taxa abundance distribution (TAD).

Usage

```
convertOtuTad(otu)
```

Arguments

`otu` An OTU data frame.

Value

An R data frame representing a taxa abundance distribution (TAD). The first column is species abundance and the second column represents the number of species with that level of abundance in the sample.

See Also

[getSampleOtu](#), [getRunOtu](#), [mergeOtu](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runs=runsBySample(ps, samp[2])
otu=getRunOtu(ps, runs[1])
tad=convertOtuTad(otu)
head(tad)
```

getProjectsList	<i>Download the EMG projects list into a data frame</i>
-----------------	---

Description

This downloads an EMG projects list (detailed info) into an R data frame for subsequent analysis.

Usage

```
getProjectsList()
```

Value

An R data frame containing a list of all public EMG projects, one row per project.

See Also

[getProjectSummary](#), [projectSamples](#)

Examples

```
p1 = getProjectsList()
str(p1)

# Find big projects
biggies = p1$Study.ID[p1$Number.Of.Samples >= 100]
ps = getProjectSummary(biggies[3])

# Find some 16S projects
p1$Study.ID[agrep("16S", p1$Study.Name)]
```

`getProjectSummary` *Download an EMG project summary into a data frame*

Description

This downloads an EMG project summary file into an R data frame for subsequent analysis.

Usage

```
getProjectSummary(projectID)
```

Arguments

`projectID` A character string representing an EMG project ID (eg. SRP047083).

Value

An R data frame containing a list of all samples and runs associated with the project, one row per run. The `projectID` is associated with the data frame via the attribute `project.id`.

See Also

[getProjectsList](#), [projectSamples](#), [projectRuns](#), [runsBySample](#), [read.project.csv](#)

Examples

```
ps=getProjectSummary("SRP047083")
# Tabulate number of runs per by sample
table(ps$Sample.ID)
```

getRunOtu	<i>Downloads OTU count data associated with a EMG run</i>
-----------	---

Description

This function downloads an OTU count data file associated with a particular run from the EMG portal. The OTU data is returned as a data frame.

Usage

```
getRunOtu(summ, runID, verb=FALSE, plot.preston=FALSE)
```

Arguments

summ	An EMG project summary file, as returned by <code>getProjectSummary</code> .
runID	A string representing a valid EMG run ID within the project summ.
verb	If true, outputs runID to console.
plot.preston	If true, produces a Preston plot for the taxa abundance distribution of the run.

Value

An R data frame containing OTU count data. The first column is the OTU ID, the second is the count, and the third contains a taxonomic classification.

See Also

[getSampleOtu](#), [mergeOtu](#), [projectRuns](#), [convertOtuTad](#), [plotOtu](#), [read.otu.tsv](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runs=runsBySample(ps, samp[2])
getRunOtu(ps, runs[1])
```

getSampleOtu	<i>Downloads and merges all of the OTU count data associated with a EMG sample</i>
--------------	--

Description

This function downloads the OTU count data files from all runs associated with a particular sample from the EMG portal, and merges it using `mergeOtu`. The merged OTU data is returned as a data frame.

Usage

```
getSampleOtu(summ, sampleID, verb=TRUE, plot.preston=FALSE)
```

Arguments

summ	An EMG project summary file, as returned by <code>getProjectSummary</code> .
sampleID	A string representing a valid EMG sample ID within the project summ.
verb	If true, outputs each runID to console.
plot.preston	If true, produces a Preston plot for the taxa abundance distribution of each run as it is downloaded.

Value

An R data frame containing OTU count data. The first column is the OTU ID, the second is the count, and the third contains a taxonomic classification.

See Also

[getRunOtu](#), [mergeOtu](#), [projectSamples](#), [convertOtuTad](#), [plotOtu](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runs=runsBySample(ps, samp[2])
getRunOtu(ps, runs[1])
```

mergeOtu	<i>Merges two OTU count data data frames to produce a combined OTU count data data frame</i>
----------	--

Description

This function merges two or more OTU count data data frames, such as returned by `getRunOtu`. The merged OTU data is returned as a data frame. The ordering of the two data frames to be merged should not be significant. The operation of merging should also be associative.

Usage

```
mergeOtu(...)
```

Arguments

... A sequence of OTU data frames to be merged.

Value

An R data frame containing OTU count data. The first column is the OTU ID, the second is the count, and the third contains a taxonomic classification.

See Also

[getSampleOtu](#), [getRunOtu](#), [projectRuns](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runs=runsBySample(ps,samp[2])

# merge two:
otu1=getRunOtu(ps,runs[1])
otu2=getRunOtu(ps,runs[2])
mergeOtu(otu1,otu2)

# merge many:
otuList=lapply(as.list(runs[1:10]),function(rid){getRunOtu(ps,rid)})
Reduce(mergeOtu,otuList)
```

plotOtu

Plots OTU count data

Description

This function takes as input a data frame representing OTU counts, such as returned by `getRunOtu`, plots it on the graphics device.

Usage

```
plotOtu(otu)
```

Arguments

otu An OTU data frame.

Value

The function returns silently, and is called purely for its side-effect of rendering plots on the graphics device.

See Also

[getSampleOtu](#), [getRunOtu](#), [convertOtuTad](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runs=runsBySample(ps,samp[2])
otu=getRunOtu(ps,runs[1])
plotOtu(otu)
```

projectRuns

Find runs associated with an EMG project

Description

This function queries an already-downloaded project summary file for vector of EMG run IDs associated with the project.

Usage

```
projectRuns(summ)
```

Arguments

summ An EMG project summary file, as returned by `getProjectSummary`.

Value

An R vector containing a list of strings representing EMG run IDs.

See Also

[getProjectSummary](#), [projectSamples](#), [runsBySample](#)

Examples

```
ps=getProjectSummary("SRP047083")
projectRuns(ps)
```

projectSamples	<i>Find samples associated with an EMG project</i>
----------------	--

Description

This function queries an already-downloaded project summary file for vector of EMG sample IDs associated with the project.

Usage

```
projectSamples(summ)
```

Arguments

summ An EMG project summary file, as returned by `getProjectSummary`.

Value

An R vector containing a list of strings representing EMG sample IDs.

See Also

[getProjectSummary](#), [projectRuns](#), [runsBySample](#)

Examples

```
ps=getProjectSummary("SRP047083")
projectSamples(ps)
```

read.otu.tsv	<i>Reads OTU count data associated with a EMG run</i>
--------------	---

Description

This function reads an OTU count data file associated with a particular run from the EMG portal. The OTU data is returned as a data frame.

Usage

```
read.otu.tsv(fileName,...)
```

Arguments

fileName A string representing a filename (or URL) corresponding to a valid EMG OTU TSV file.

... Extra arguments to be passed to `read.delim`.

Value

An R data frame containing OTU count data. The first column is the OTU ID, the second is the count, and the third contains a taxonomic classification.

See Also

[getSampleOtu](#), [mergeOtu](#), [getRunOtu](#)

<code>read.project.csv</code>	<i>Read an EMG project summary into a data frame</i>
-------------------------------	--

Description

This reads an EMG project summary file into an R data frame for subsequent analysis.

Usage

```
read.project.csv(fileName,projectID,...)
```

Arguments

<code>fileName</code>	Filename (or URL) corresponding to a EMG project summary CSV file.
<code>projectID</code>	String corresponding to the EMG project ID (sometimes known as study ID) associated with the file.
<code>...</code>	Extra parameters to be passed on to <code>read.csv</code> .

Value

An R data frame containing a list of all samples and runs associated with the project, one row per run. The `projectID` is associated with the data frame via the attribute `project.id`.

See Also

[getProjectSummary](#), [projectSamples](#), [projectRuns](#), [runsBySample](#)

runsBySample	<i>Find runs associated with a particular sample ID within an EMG project</i>
--------------	---

Description

This function queries an already-downloaded project summary file for vector of EMG run IDs associated with a particular sample ID within the project.

Usage

```
runsBySample(summ, sampleID)
```

Arguments

summ	An EMG project summary file, as returned by <code>getProjectSummary</code> .
sampleID	A string representing a valid EMG sample ID within the project summ.

Value

An R vector containing a list of strings representing EMG run IDs.

See Also

[getProjectSummary](#), [projectSamples](#), [projectRuns](#)

Examples

```
ps=getProjectSummary("SRP047083")
samp=projectSamples(ps)
runsBySample(ps, samp[2])
```

Index

*Topic **EMG**

- analyseOtu, [2](#)
- convertOtuTad, [3](#)
- getProjectsList, [3](#)
- getProjectSummary, [4](#)
- getRunOtu, [5](#)
- getSampleOtu, [5](#)
- mergeOtu, [6](#)
- plotOtu, [7](#)
- projectRuns, [8](#)
- projectSamples, [9](#)
- read.otu.tsv, [9](#)
- read.project.csv, [10](#)
- runsBySample, [11](#)

analyseOtu, [2](#)

convertOtuTad, [3](#), [5–7](#)

getProjectsList, [3](#), [4](#)

getProjectSummary, [4](#), [4](#), [8–11](#)

getRunOtu, [2](#), [3](#), [5](#), [6](#), [7](#), [10](#)

getSampleOtu, [2](#), [3](#), [5](#), [5](#), [7](#), [10](#)

mergeOtu, [3](#), [5](#), [6](#), [6](#), [10](#)

plotOtu, [2](#), [5](#), [6](#), [7](#)

projectRuns, [4](#), [5](#), [7](#), [8](#), [9–11](#)

projectSamples, [4](#), [6](#), [8](#), [9](#), [10](#), [11](#)

read.otu.tsv, [5](#), [9](#)

read.project.csv, [4](#), [10](#)

runsBySample, [4](#), [8–10](#), [11](#)