Package ‘stampr’

January 19, 2021

Type Package

Title Perform the STAMP analysis

Version 1.0

Date 2020-12-23

Author Ashton Trey Belew

Maintainer Ashton Trey Belew <`abelew@gmail.com`>

Description A reimplementation of the STAMP process.

License GPL-2 | file LICENSE

Depends dplyr, tidyr

VignetteBuilder knitr

ByteCompile true

Encoding UTF-8

RoxygenNote 7.1.1

Collate '01_stampr.r'
'calculate_nb.r'
'filter_given_reference.r'
'filter_tags.r'
'plot_calibration.r'
'plot_read_density.r'
'predict_cfu.r'
'preprocess_reads.r'
'read_idx.r'
'read_qiime_otus.r'

R topics documented:

`calculate_nb` *Given the metadata and tag data, calculate Nb values.*

Description

This is minor cleanup of the original implementation I received from Dr. Lee's collaborators. I do have the text of the original paper, but it is difficult to interpret; so I fell back on basically retranscribing the original code.

Usage

```
calculate_nb(tags, generations = 1, metadata_column = "Nb", write = NULL)
```

Arguments

`tags`	Input tags provided by count_tags.pl and hopefully count_otus.pl
`generations`	Set the g parameter from the published paper.

`filter_given_reference` *Filter samples using the tags observed in the reference samples.*

Description

Take the result from read_tags() and use it to find which tags are in the non-reference samples. One caveat about this process, it splits the data by replicate because tags are not global. Thus we will need to consider what is the best way to handle the filtered data. The simplest is to just merge the pieces back together.

Usage

```
filter_given_reference(tag_list, reference_cutoff = 1, remove_nas = FALSE)
```

Arguments

`reference_cutoff`	Minimum number of reads in the reference samples to be considered 'real'
`remove_nas`	Replace NA with 0?
`index_list`	Result from count_tags

`filter_given_reference_list` *This does the real work for filter_given_reference().*

Description

Create a list of matrices containing only the tags observed in each replicates' reference sample.

Usage

```
filter_given_reference_list(index_list, reference_cutoff = 1)
```

Arguments

`index_list`	Result from count_tags
`reference_cutoff`	Minimum number of reads in the reference samples to be considered 'real'

`filter_tags` *Try out some heuristics for filtering the tag data.*

Description

Given our focus on RNAseq, it is not a stretch to guess that we would assume some of those methods would be useful. Here is a cpm implementation of some simple filtering.

Usage

```
filter_tags(tag_list, multiplier = 1)
```

Arguments

`tag_list`	Result from one form of read_tags(). This contains the metadata and the matrix of reads/tag.
`multiplier`	Arbitrary multiplier to make the filter more stringent.

`filter_topn_tags` *Keep only the top-n most abundant tags*

Description

In Dr. Lee's experiment, they explicitly know the number of input tags. So we can explicitly keep only those n most observed tags.

Usage

```
filter_topn_tags(tag_list, replicate_column = "replicate", topn = 600)
```

Arguments

`tag_list`	Existing tag data
`replicate_column`	Tags should be shared across replicates.
`topn`	How many tags to keep.

`gather_tags_per_replicate` *Extract the tags observed in each replicate's reference sample.*

Description

This ought to be extended to multiple samples/reference.

Usage

```
gather_tags_per_replicate(meta, reads_per_tag, reference_cutoff = 1)
```

Arguments

`meta`	Metadata matrix.
`reads_per_tag`	The set of tags from the index list.

make_frequency_df — *The f(i,s) term in the Nb estimation equation requires frequency estimates.*

Description

Thus, take the sum of each column and divide every value by it.

Usage

```
make_frequency_df(mtrx)
```

Arguments

`mtrx`	Matrix to transform.

plot_calibration — *Plot the calibration curve of the tags observed vs. CFU*

Description

Currently this is just a simple ggplot scatterplot. I want to make it more configurable and fun.

Usage

```
plot_calibration(
  tags,
  x_column = "cfu",
  y_column = NULL,
  color_column = "replicate",
  transform_x = "log2",
  transform_y = "log2"
)
```

Arguments

`tags`	Tag data.
`x_column`	Metadata column for the x axis.
`y_column`	Metadata column for the y axis.
`color_column`	Metadata column for colors.
`transform_x`	Perform a transformation on the x axis?
`transform_y`	Perform a transformation on the y axis?

plot_read_density *Make a boxplot of read density*

Description

Make a boxplot of read density

Usage

```
plot_read_density(retlist)
```

Arguments

retlist	from read_idx or read_qiime

predict_cfu *Try out some methods to go from tags/Nb back to CFU.*

Description

This function has not been completed, mostly because I am not sure if it will ever be used, but also because I am not sure of the appropriate model. I have never had an opportunity to use modelling to estimate data, so this is mostly a place for me to play around with something that I should know but don't.

Usage

```
predict_cfu(meta, from = "Nb", to = "cfu", provided = NULL)
```

Arguments

meta	Metadata.
from	Factor from which to extrapolate.
to	Factor to which to return.
provided	Dataframe of incomplete data.

`preprocess_stamp` *Preprocess raw reads using cutadapt and my little perl script.*

Description

Preprocess raw reads using cutadapt and my little perl script.

Usage

```
preprocess_stamp(
  metadata,
  raw_column = "raw_fastq",
  trimmed_column = "trimmed_fastq",
  index_column = "index_table",
  output = NULL
)
```

Arguments

`metadata`	Sample metadata.
`column`	Metadata column containing the locations of the raw reads.

`read_qiime_otus` *Create data structures similar to read_idx using qiime otus.*

Description

There are some problems with this implementation still, primarily because the version of qiime does not helpfully give the tag sequences. I have found some ways around this, but since I don't really like that implementation I haven't finished it yet.

Usage

```
read_qiime_otus(
  metadata,
  otu_column = "qiime_otus",
  xref_sequence = FALSE,
  trimmed_column = "trimmed_reads",
  output = NULL
)
```

Arguments

`metadata`	Sample metadata.
`output`	Write the matrix to this file, if provided.
`column`	Metadata column containing the qiime output.

read_tags *Interpret the files produced by my count_idx.pl script.*

Description

This uses the index_table column in the metadata, reads this files associated with it, and creates some data structures with the results. These include a modified version of the metadata, containing some summary information, a table of the reads/tag observed, and a long table for ggplot.

Usage

```
read_tags(
  metadata,
  id_column = "sampleid",
  index_column = "index_table",
  output = NULL,
  cutoff = 3
)
```

Arguments

`metadata`	Sample metadata.
`id_column`	Column in the metadata containing the sample names.
`index_column`	Column in the metadata containing the tag tables.
`output`	Write out the matrix to this file (if provided).
`cutoff`	Initial reads/tag filter.

stampr *stampr: Some functions for analyzing sequencing data intended to quantify RT error rates.*

Index