

ETEC 16S dataset

November 1, 2018

This data package contains the data used in the analyses found in "Individual-specific changes in the human gut microbiota after challenge with enterotoxigenic *Escherichia coli* and subsequent ciprofloxacin treatment". DNA was amplified using 'universal' primers targeting the V1-V2 region of the 16S rRNA gene (small subunit of the ribosome) in bacteria - 338R (5'- CATGCTGCCTCCCGTAGGAGT -3') and 27F (5'- AGAGTTTGATCCTGGCTCAG -3'). Both forward and reverse primers had a 5 prime portion specific for use with 454 GS-FLX Titanium sequencing technology and the forward primers contained a barcode between the Titanium and gene specific region, so that samples could be pooled to a multiplex level of 132 samples per instrument run.

16S rRNA gene sequencing was performed for all available stool samples. After sequencing, 124 samples passed quality controls, corresponding to data from 5 volunteers with the most unambiguous cases of diarrhea during the study (54 samples) and 7 volunteers without diarrheal symptoms who had the most stool samples (78 samples). The raw data have been submitted to NCBI under project ID: PRJNA298336.

Data is stored as an `MRExperiment`-class object. The count matrix was generated using DNAClust (<http://dnaclust.sourceforge.net/>). For more details please refer to the paper.

The help file `?etec16s` describes the example dataset.

1 The Data

We start by loading the library and the data.

```
> suppressMessages(library(metagenomeSeq))
> library(etec16s)
> data(etec16s)
```

This will load the `etec16s` object of class `MRExperiment`. As described in the `metagenomeSeq` vignette, `print` (or `show`) will display summary information.

```
> etec16s
```

```

MRexperiment (storageMode: environment)
assayData: 6423 features, 124 samples
  element names: counts
protocolData: none
phenoData
  sampleNames: 2 3 ... 333 (124 total)
  varLabels: SubjectID Dose ... AntibPrev (7 total)
  varMetadata: labelDescription
featureData
  featureNames: 3 5 ... 148089 (6423 total)
  fvarLabels: OTU.ID Center ... Species (8 total)
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:

```

The data in `etec16s` is the substrate for the analysis described in "Individual-specific changes in the human gut microbiota after challenge with enterotoxigenic *Escherichia coli* and subsequent ciprofloxacin treatment". Included in the `MRexperiment` object are the counts, phenotype and feature information.

The phenotype information can be accessed with the `phenoData` and `pData` methods:

```
> phenoData(etec16s)
```

```

An object of class 'AnnotatedDataFrame'
  sampleNames: 2 3 ... 333 (124 total)
  varLabels: SubjectID Dose ... AntibPrev (7 total)
  varMetadata: labelDescription

```

```
> head(pData(etec16s))
```

| | SubjectID | Dose | Day | AnyDayDiarrhea | Diarrhea | AntiGiven | AntibPrev |
|----|-----------|-------------------------|-----|----------------|----------|-----------|-----------|
| 2 | E01JH0003 | 2 x 10 ⁶ cfu | -1 | 0 | 0 | 0 | 0 |
| 3 | E01JH0004 | 2 x 10 ⁶ cfu | -1 | 1 | 0 | 0 | 0 |
| 8 | E01JH0011 | 2 x 10 ⁶ cfu | -1 | 1 | 0 | 0 | 0 |
| 10 | E01JH0013 | 2 x 10 ⁶ cfu | -1 | 0 | 0 | 0 | 0 |
| 12 | E01JH0016 | 2 x 10 ⁶ cfu | -1 | 1 | 0 | 0 | 0 |
| 13 | E01JH0017 | 2 x 10 ⁵ cfu | -1 | 1 | 0 | 0 | 0 |

The feature information including cluster representative sequence can be accessed with the `featureData` and `fData` methods:

```
> featureData(etec16s)
```

```
An object of class 'AnnotatedDataFrame'
  featureNames: 3 5 ... 148089 (6423 total)
  varLabels: OTU.ID Center ... Species (8 total)
  varMetadata: labelDescription
```

```
> head(fData(etec16s))
```

| | OTU.ID | Center | Phylum | Class | Order |
|----|--------|-----------------|-----------------|-------------------------|--------------------|
| 3 | 3 | 130_6133 | Bacteroidetes | Bacteroidia | Bacteroidales |
| 5 | 5 | 14_1391 | Bacteroidetes | Bacteroidia | Bacteroidales |
| 10 | 10 | 111_7283 | Bacteroidetes | Bacteroidia | Bacteroidales |
| 11 | 11 | 130_1817 | Bacteroidetes | Bacteroidia | Bacteroidales |
| 13 | 13 | 82_8248 | Bacteroidetes | Bacteroidia | Bacteroidales |
| 14 | 14 | 131_3769 | Verrucomicrobia | Verrucomicrobiae | Verrucomicrobiales |
| | | Family | Genus | Species | |
| 3 | | Bacteroidaceae | Bacteroides | Bacteroides vulgatus | |
| 5 | | Bacteroidaceae | Bacteroides | Bacteroides uniformis | |
| 10 | | Bacteroidaceae | Bacteroides | Bacteroides uniformis | |
| 11 | | Bacteroidaceae | Bacteroides | Bacteroides vulgatus | |
| 13 | | Bacteroidaceae | Bacteroides | Bacteroides ovatus | |
| 14 | | Akkermansiaceae | Akkermansia | Akkermansia muciniphila | |

The raw or normalized counts matrix can be accessed with the MRcounts function:

```
> head(MRcounts(etec16s[,1:10]))
```

| | 2 | 3 | 8 | 10 | 12 | 13 | 16 | 20 | 21 | 24 |
|----|---|---|---|----|----|----|----|----|----|----|
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 11 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Using this class, the object can be easily subsetted, for example:

```
> etec16s_day84 = etec16s[,which(pData(etec16s)$Day == 84)]
> etec16s_day84
```

```
MRexperiment (storageMode: environment)
assayData: 6423 features, 12 samples
  element names: counts
protocolData: none
phenoData
```

```
sampleNames: 306 307 ... 333 (12 total)
varLabels: SubjectID Dose ... AntibPrev (7 total)
varMetadata: labelDescription
featureData
  featureNames: 3 5 ... 148089 (6423 total)
  fvarLabels: OTU.ID Center ... Species (8 total)
  fvarMetadata: labelDescription
experimentData: use 'experimentData(object)'
Annotation:
```