Package 'curatedMetagenomicData'

June 30, 2016

Type Package

Title Curated Metagenomic Data of the Human Microbiome

Version 0.99.1

Author Levi Waldron <levi.waldron@sph.cuny.edu>,
Faizan Malik <faizankmalik2012@yahoo.com>,
Valerie Obenchain <Valerie.Obenchain@roswellpark.org>,
Edoardo Pasolli <edoardo.pasolli@unitn.it>,
Nicola Segata <nicola.segata@unitn.it>,
Lucas Schiffer <schiffer.lucas@gmail.com>,
MorganMartin <martin.morgan@roswellpark.org>

Maintainer Levi Waldron <levi.waldron@sph.cuny.edu>

Description The curatedMetagenomicData package provides taxonomic, functional, and gene marker abundance abundance in samples collected from different body sites.

Depends R (>= 3.3.0),
ExperimentHubData (>= 0.101.1),
Biobase

Imports BiocGenerics

Suggests BiocStyle,
testthat,
rmarkdown,
gplots,
ExperimentHub,
phyloseq

VignetteBuilder rmarkdown

License Artistic-2.0

biocViews ExperimentData, CancerData, ExpressionData, Homo_sapiens_Data, Metagenomics

R topics documented:

`curatedMetagenomicData-package`

Taxonomic, functional, and gene marker abundance in samples collected from different body sites available as `ExpressionSet` *objects.*

Description

Shotgun metagenomics sequencing can simultaneously profile bacteria, archeae, fungi, and viruses to the strain level, enabling us to better understand relationships between microbiota to human health. In this package we present a collection of relative abundance of taxa, marker datasets, and gene function across 3000 samples collected from various human body sites. The package contains two types of marker datasets:

- '_pres': presence/absence of taxonomic markers
- '_ab': abundance level of markers across samples

The data have been parsed into `ExpressionSet` objects and are available in the Bioconductor `ExperimentHub`.

Details

See the vignette for examples of using these data in differential gene expression analysis.

browseVignettes("curatedMetagenomicData")

Details of how these data were created are in the scripts/ directory of the source package.

Examples

```
library(ExperimentHub)
hub <- ExperimentHub()
x <- query(hub, c("Metagenomics", "CancerData"))
x
## Not run:
## download resource
x[[1]]

## End(Not run)

## Get the package name associated with a resource:
packageName <- package(x[1])
packageName

## Using the package name as a search term returns all records from
## the same package.
query(hub, "curatedMetagenomicData")
```

Candela_Africa.stool.eset

Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota

---

Description

Through human microbiome sequencing, we can better understand how host evolutionary and ontogenetic history is reflected in the microbial function. However, there has been no information on the gut metagenome configuration in hunter-gatherer populations, posing a gap in our knowledge of gut microbiota (GM)-host mutualism arising from a lifestyle that describes over 90% of human evolutionary history. Here, we present the first metagenomic analysis of GM from Hadza hunter-gatherers of Tanzania, showing a unique enrichment in metabolic pathways that aligns with the dietary and environmental factors characteristic of their foraging lifestyle. We found that the Hadza GM is adapted for broad-spectrum carbohydrate metabolism, reflecting the complex polysaccharides in their diet. Furthermore, the Hadza GM is equipped for branched-chain amino acid degradation and aromatic amino acid biosynthesis. Resistome functionality demonstrates the existence of antibiotic resistance genes in a population with little antibiotic exposure, indicating the ubiquitous presence of environmentally derived resistances. Our results demonstrate how the functional specificity of the GM correlates with certain environment and lifestyle factors and how complexity from the exogenous environment can be balanced by endogenous homeostasis. The Hadza gut metagenome structure allows us to appreciate the co-adaptive functional role of the GM in complementing the human physiology, providing a better understanding of the versatility of human life and subsistence.

Usage

```
data( Candela_Africa.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Rampelli S, Schnorr SL, Consolandi C, Turroni S, Severgnini M, Peano C, Brigid
```

```
 Laboratory: Department of Pharmacy and Biotechnology, University of Bologna, Bologna 40126, Italy
 Contact information:
 Title: Metagenome Sequencing of the Hadza Hunter-Gatherer Gut Microbiota
 URL:
 PMIDs: 25981789

 Abstract: A 213 word abstract is available. Use 'abstract' method.
 notes:
  platform_summary:
     NA
  platform_accession:
     NA
  platform_title:
     GAIIx
  platform_technology:
     Illumina
  platform_distribution:
     NA
  platform_manufacturer:
     Illumina
  processor:
    MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 38 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
Candela_Africa
            38

sampleID:
   Length     Class      Mode
       38 character character

subjectID:
   Length     Class      Mode
       38 character character

bodysite:
stool
   38

disease:
 n
```

```
38

age:
   Length     Class      Mode
       38 character character

gender:
female   male
    16     22

country:
   italy tanzania
      11       27

sequencing_technology:
Illumina
      38

pubmedid:
25981789
      38

camp:
bologna dedauko sengele
     11      20       7
```

`Chatelier_gut_obesity.stool.eset`

Richness of human gut microbiome correlates with metabolic markers

---

Description

We are facing a global metabolic health crisis provoked by an obesity epidemic. Here we report the human gut microbial composition in a population sample of 123 non-obese and 169 obese Danish individuals. We find two groups of individuals that differ by the number of gut microbial genes and thus gut bacterial richness. They contain known and previously unknown bacterial species at different proportions; individuals with a low bacterial richness (23% of the population) are characterized by more marked overall adiposity, insulin resistance and dyslipidaemia and a more pronounced inflammatory phenotype when compared with high bacterial richness individuals. The obese individuals among the lower bacterial richness group also gain more weight over time. Only a few bacterial species are sufficient to distinguish between individuals with high and low bacterial richness, and even between lean and obese participants. Our classifications based on variation in the gut microbiome identify subsets of individuals in the general white adult population who may be at increased risk of progressing to adiposity-associated co-morbidities.

Usage

```
data( Chatelier_gut_obesity.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, Falony G, Almeida M,
  Laboratory: INRA, Institut National de la Recherche Agronomique, US1367 Metagenopolis, 78350 Jouy
   Contact information:
   Title: Richness of human gut microbiome correlates with metabolic markers
   URL:
   PMIDs: 23985870

   Abstract: A 167 word abstract is available. Use 'abstract' method.
   notes:
    platform_summary:
       NA
    platform_accession:
       NA
    platform_title:
        Genome Analyzer II
    platform_technology:
       Illumina
    platform_distribution:
       NA
    platform_manufacturer:
       Illumina
    processor:
      MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 278 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
Chatelier_gut_obesity
                  278

sampleID:
   Length     Class      Mode
      278 character character

subjectID:
   Length     Class      Mode
      278 character character

bodysite:
```

```
stool
  278

disease:
leaness       n obesity
     89      25     164

country:
denmark
    278

sequencing_technology:
Illumina
     278

pubmedid:
23985870
     278

paired_end_insert_size:
   Length     Class      Mode
      278 character character

read_length:
44/75    75    90
    1   162   115

total_reads:
   Length     Class      Mode
      278 character character

matched_reads:
   Length     Class      Mode
      278 character character

uniquely_matching_reads:
   Length     Class      Mode
      278 character character

uniquely_matched_reads:
   Length     Class      Mode
      278 character character

gene_number:
   Length     Class      Mode
      278 character character

gene_number_for_11m_uniquely_matched_reads:
   Length     Class      Mode
      278 character character

hitchip_probe_number:
```

```
   Length     Class      Mode
      278 character character

bmi:
   Length     Class      Mode
      278 character character

gene_count_class:
hgc lgc
212  66

hitchip_probe_class:
hpc lpc
221  57

X.SampleID:
Metaphlan2_Analysis
                278
```

`hmp.anterior_nares.eset` *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.anterior_nares.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
```

```
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 94 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
hmp
 94

sampleID:
   Length     Class      Mode
       94 character character

subjectID:
   Length     Class      Mode
       94 character character

bodysite:
anterior_nares
            94

disease:
 n
94
```

```
gender:
female   male
    36     58

country:
usa
 94

sequencing_technology:
Illumina
      94

pubmedid:
22699609
      94

X.SampleID:
Metaphlan2_Analysis
                 94

visit_number:
 1  2  3
61 32  1

snprnt:
   Length     Class      Mode
       94 character character

wmsphase:
 1
94
```

`hmp.buccal_mucosa.eset` *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and

ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.buccal_mucosa.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 123 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
hmp
123
```

```
sampleID:
   Length     Class      Mode
      123 character character

subjectID:
   Length     Class      Mode
      123 character character

bodysite:
buccal_mucosa
          123

disease:
  n
123

gender:
female   male
    57     66

country:
usa
123

sequencing_technology:
Illumina
     123

pubmedid:
22699609
     123

X.SampleID:
Metaphlan2_Analysis
                123

visit_number:
 1  2  3
75 45  3

snprnt:
   Length     Class      Mode
      123 character character

wmsphase:
  1
123
```

---

hmp.hard_palate.eset *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.hard_palate.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 1 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
  1

sampleID:
SRS062878
        1

subjectID:
765560005
        1

bodysite:
hard_palate
          1

disease:
n
1

gender:
male
   1

country:
usa
  1

sequencing_technology:
Illumina
       1

pubmedid:
22699609
       1

X.SampleID:
Metaphlan2_Analysis
                  1

visit_number:
2
1

snprnt:
```

```
700114333
        1

wmsphase:
1
1
```

---

hmp.keratinized_gingiva.eset

Structure, function and diversity of the healthy human microbiome

---

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.keratinized_gingiva.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
```

```
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 6 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
hmp
  6

sampleID:
   Length     Class      Mode
        6 character character

subjectID:
763496533 763577454 763961826
        2         2         2

bodysite:
keratinized_gingiva
                  6

disease:
n
6

gender:
female   male
     2      4

country:
usa
  6

sequencing_technology:
Illumina
       6
```

```
pubmedid:
22699609
       6

X.SampleID:
Metaphlan2_Analysis
                  6

visit_number:
1 2
3 3

snprnt:
   Length     Class      Mode
        6 character character

wmsphase:
1
6
```

`hmp.1_retroauricular_crease.eset`

Structure, function and diversity of the healthy human microbiome

---

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.1_retroauricular_crease.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
```

```
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 9 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
  9

sampleID:
   Length     Class      Mode
        9 character character

subjectID:
   Length     Class      Mode
        9 character character

bodysite:
l_retroauricular_crease
                      9

disease:
```

```
n
9

gender:
male
   9

country:
usa
  9

sequencing_technology:
Illumina
       9

pubmedid:
22699609
       9

X.SampleID:
Metaphlan2_Analysis
                  9

visit_number:
1 2
4 5

snprnt:
   Length     Class      Mode
        9 character character

wmsphase:
1
9
```

`hmp.mid_vagina.eset` *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and

ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.mid_vagina.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 2 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
hmp
  2
```

```
sampleID:
SRS014466 SRS015072
        1         1

subjectID:
763577454
        2

bodysite:
mid_vagina
         2

disease:
n
2

gender:
female
     2

country:
usa
  2

sequencing_technology:
Illumina
       2

pubmedid:
22699609
       2

X.SampleID:
Metaphlan2_Analysis
                  2

visit_number:
1 2
1 1

snprnt:
700023120 700023727
        1         1

wmsphase:
1
2
```

`hmp.palatine_tonsils.eset`
Structure, function and diversity of the healthy human microbiome

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.palatine_tonsils.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 6 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
  6

sampleID:
   Length     Class      Mode
        6 character character

subjectID:
763496533 763577454 763961826 764042746
        2         2         1         1

bodysite:
palatine_tonsils
               6

disease:
n
6

gender:
female   male
     3      3

country:
usa
  6

sequencing_technology:
Illumina
       6

pubmedid:
22699609
       6

X.SampleID:
Metaphlan2_Analysis
                  6
```

```
visit_number:
1 2
2 4

snprnt:
   Length     Class      Mode
        6 character character

wmsphase:
1
6
```

`hmp.posterior_fornix.eset` *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.posterior_fornix.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
```

```
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 63 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
 63

sampleID:
   Length     Class      Mode
       63 character character

subjectID:
   Length     Class      Mode
       63 character character

bodysite:
posterior_fornix
              63

disease:
 n
63

gender:
female
    63

country:
usa
```

```
 63

sequencing_technology:
Illumina
      63

pubmedid:
22699609
      63

X.SampleID:
Metaphlan2_Analysis
                 63

visit_number:
 1 1t  2  3
37  1 22  3

snprnt:
   Length     Class      Mode
       63 character character

wmsphase:
 1
63
```

`hmp.r_retroauricular_crease.eset`

Structure, function and diversity of the healthy human microbiome

---

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.r_retroauricular_crease.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 18 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
 18

sampleID:
   Length     Class      Mode
       18 character character

subjectID:
   Length     Class      Mode
       18 character character

bodysite:
```

```
r_retroauricular_crease
                     18

disease:
 n
18

gender:
female   male
     2     16

country:
usa
 18

sequencing_technology:
Illumina
      18

pubmedid:
22699609
      18

X.SampleID:
Metaphlan2_Analysis
                 18

visit_number:
 1  2
 6 12

snprnt:
   Length     Class      Mode
       18 character character

wmsphase:
 1
18
```

---

hmp.saliva.eset *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to

vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.saliva.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 5 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------
```

```
dataset_name:
hmp
  5

sampleID:
SRS013942 SRS014468 SRS014692 SRS015055 SRS019120
        1         1         1         1         1

subjectID:
763496533 763577454 763961826
        2         2         1

bodysite:
saliva
     5

disease:
n
5

gender:
female   male
     2      3

country:
usa
  5

sequencing_technology:
Illumina
       5

pubmedid:
22699609
       5

X.SampleID:
Metaphlan2_Analysis
                  5

visit_number:
1 2
3 2

snprnt:
700021297 700023122 700023346 700023710 700037591
        1         1         1         1         1

wmsphase:
1
5
```

hmp.stool.eset *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 152 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
152

sampleID:
   Length     Class      Mode
      152 character character

subjectID:
   Length     Class      Mode
      152 character character

bodysite:
stool
  152

disease:
  n
152

gender:
female   male
    66     86

country:
usa
152

sequencing_technology:
Illumina
     152

pubmedid:
22699609
     152

X.SampleID:
Metaphlan2_Analysis
                152
```

```
visit_number:
 1  2  3
88 60  4

snprnt:
   Length     Class      Mode
      152 character character

wmsphase:
  1
152
```

---

hmp.subgingival_plaque.eset

Structure, function and diversity of the healthy human microbiome

---

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.subgingival_plaque.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
```

```
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 8 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
hmp
  8

sampleID:
   Length     Class      Mode
        8 character character

subjectID:
763435843 763496533 763577454 763961826 764042746
        1         2         2         2         1

bodysite:
subgingival_plaque
                 8

disease:
n
8

gender:
female   male
     3      5

country:
usa
```

```
  8

sequencing_technology:
Illumina
       8

pubmedid:
22699609
       8

X.SampleID:
Metaphlan2_Analysis
                  8

visit_number:
1 2
4 4

snprnt:
   Length     Class      Mode
        8 character character

wmsphase:
1
8
```

`hmp.supragingival_plaque.eset`

Structure, function and diversity of the healthy human microbiome

---

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.supragingival_plaque.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 128 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
128

sampleID:
   Length     Class      Mode
      128 character character

subjectID:
   Length     Class      Mode
      128 character character

bodysite:
```

```
supragingival_plaque
                128

disease:
  n
128

gender:
female   male
    56     72

country:
usa
128

sequencing_technology:
Illumina
     128

pubmedid:
22699609
     128

X.SampleID:
Metaphlan2_Analysis
                128

visit_number:
 1  2  3
76 49  3

snprnt:
   Length     Class      Mode
      128 character character

wmsphase:
  1
128
```

`hmp.throat.eset` *Structure, function and diversity of the healthy human microbiome*

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to

vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.throat.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 7 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------
```

```
dataset_name:
hmp
  7

sampleID:
   Length     Class      Mode
        7 character character

subjectID:
763496533 763577454 763961826 765560005
        2         2         2         1

bodysite:
throat
     7

disease:
n
7

gender:
female   male
     2      5

country:
usa
  7

sequencing_technology:
Illumina
       7

pubmedid:
22699609
       7

X.SampleID:
Metaphlan2_Analysis
                  7

visit_number:
1 2
3 4

snprnt:
   Length     Class      Mode
        7 character character

wmsphase:
1
7
```

hmp.tongue_dorsum.eset

Structure, function and diversity of the healthy human microbiome

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.tongue_dorsum.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
```

```
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 137 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
137

sampleID:
   Length     Class      Mode
      137 character character

subjectID:
   Length     Class      Mode
      137 character character

bodysite:
tongue_dorsum
          137

disease:
  n
137

gender:
female   male
    61     76

country:
usa
137

sequencing_technology:
Illumina
     137

pubmedid:
22699609
     137

X.SampleID:
Metaphlan2_Analysis
                137
```

```
visit_number:
 1  2  3
79 53  5

snprnt:
   Length     Class      Mode
      137 character character

wmsphase:
  1
137
```

`hmp.vaginal_introitus.eset`

Structure, function and diversity of the healthy human microbiome

---

Description

Studies of the human microbiome have revealed that even healthy individuals differ remarkably in the microbes that occupy habitats such as the gut, skin and vagina. Much of this diversity remains unexplained, although diet, environment, host genetics and early microbial exposure have all been implicated. Accordingly, to characterize the ecology of human-associated microbial communities, the Human Microbiome Project has analysed the largest cohort and set of distinct, clinically relevant body habitats so far. We found the diversity and abundance of each habitat's signature microbes to vary widely even among healthy subjects, with strong niche specialization both within and among individuals. The project encountered an estimated 81-99% of the genera, enzyme families and community configurations occupied by the healthy Western microbiome. Metagenomic carriage of metabolic pathways was stable among individuals despite variation in community structure, and ethnic/racial background proved to be one of the strongest associations of both pathways and microbes with clinical metadata. These results thus delineate the range of structural and functional configurations normal in the microbial communities of a healthy population, enabling future characterization of the epidemiology, ecology and translational applications of the human microbiome.

Usage

```
data( hmp.vaginal_introitus.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Huttenhower C, Gevers D, Knight R et al.
  Laboratory: Human Microbiome Project Consortium
  Contact information:
  Title: Structure, function and diversity of the healthy human microbiome
  URL:
  PMIDs: 22699609

  Abstract: A 187 word abstract is available. Use 'abstract' method.
```

```
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 3 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
hmp
  3

sampleID:
SRS014465 SRS015071 SRS062752
        1         1         1

subjectID:
763577454 764042746
        2         1

bodysite:
vaginal_introitus
                3

disease:
n
3

gender:
female
     3

country:
```

```
usa
  3

sequencing_technology:
Illumina
       3

pubmedid:
22699609
       3

X.SampleID:
Metaphlan2_Analysis
                  3

visit_number:
1 2
1 2

snprnt:
700023119 700023726 700114429
        1         1         1

wmsphase:
1
3
```

`Neilsen_genome_assembly.stool.eset`
Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes

---

Description

Most current approaches for analyzing metagenomic data rely on comparisons to reference genomes, but the microbial diversity of many environments extends far beyond what is covered by reference databases. De novo segregation of complex metagenomic data into specific biological entities, such as particular bacterial strains or viruses, remains a largely unsolved problem. Here we present a method, based on binning co-abundant genes across a series of metagenomic samples, that enables comprehensive discovery of new microbial organisms, viruses and co-inherited genetic entities and aids assembly of microbial genomes without the need for reference sequences. We demonstrate the method on data from 396 human gut microbiome samples and identify 7,381 co-abundance gene groups (CAGs), including 741 metagenomic species (MGS). We use these to assemble 238 high-quality microbial genomes and identify affiliations between MGS and hundreds of viruses or genetic entities. Our method provides the means for comprehensive profiling of the diversity within complex metagenomic samples.

Usage

```
data( Neilsen_genome_assembly.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Nielsen HB, Almeida M, Juncker AS, Rasmussen S, Li J, Sunagawa S, Plichta DR, C
  Laboratory: 1] Center for Biological Sequence Analysis, Technical University of Denmark, Kongens
   Contact information:
  Title: Identification and assembly of genomes and genetic elements in complex metagenomic samples
   URL:
   PMIDs: 24997787

   Abstract: A 153 word abstract is available. Use 'abstract' method.
   notes:
    platform_summary:
       NA
    platform_accession:
       NA
    platform_title:
       Illumina
    platform_technology:
       Illumina
    platform_distribution:
       NA
    platform_manufacturer:
       Illumina
    processor:
      MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 382 samples
Platform type: NA
----------------------------
Available sample meta-data:
----------------------------

dataset_name:
Neilsen_genome_assembly
                    382

sampleID:
   Length     Class      Mode
      382 character character

subjectID:
   Length     Class      Mode
      382 character character

bodysite:
```

```
stool
  382

disease:
    ibd_crohn_disease ibd_ulcerative_colitis                  n          n_relative
                   21                   127                187                  47

age:
   Length     Class      Mode
      382 character character

gender:
female   male
   222    160

country:
denmark   spain
    163     219

sequencing_technology:
Illumina
     382

pubmedid:
24997787
     382

bmi:
   Length     Class      Mode
      382 character character

X.SampleID:
Metaphlan2_Analysis
                382

sampling_day:
   Length     Class      Mode
      382 character character

known_consumers_of_a_defined_fermented_milk_product_.dfmp.:
dfmp   na
  19  363

mgs_richness:
   Length     Class      Mode
      382 character character

mgs_profile_matched_sample_pairs:
   Length     Class      Mode
      382 character character

ethnicity:
```

```
white
  382
```

Quin_gut_liver_cirrhosis.stool.eset
Alterations of the human gut microbiome in liver cirrhosis

Description

Liver cirrhosis occurs as a consequence of many chronic liver diseases that are prevalent worldwide. Here we characterize the gut microbiome in liver cirrhosis by comparing 98 patients and 83 healthy control individuals. We build a reference gene set for the cohort containing 2.69??million genes, 36.1% of which are novel. Quantitative metagenomics reveals 75,245 genes that differ in abundance between the patients and healthy individuals (false discovery rate??<??0.0001) and can be grouped into 66 clusters representing cognate bacterial species; 28 are enriched in patients and 38 in control individuals. Most (54%) of the patient-enriched, taxonomically assigned species are of buccal origin, suggesting an invasion of the gut from the mouth in liver cirrhosis. Biomarkers specific to liver cirrhosis at gene and function levels are revealed by a comparison with those for type 2 diabetes and inflammatory bowel disease. On the basis of only 15 biomarkers, a highly accurate patient discrimination index is created and validated on an independent cohort. Thus microbiota-targeted biomarkers may be a powerful tool for diagnosis of different diseases.

Usage

```
data( Quin_gut_liver_cirrhosis.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Qin N, Yang F, Li A, Prifti E, Chen Y, Shao L, Guo J, Le Chatelier E, Yao J, Wu
  Laboratory: 1] State Key Laboratory for Diagnosis and Treatment of Infectious Disease, The First
   Contact information:
   Title: Alterations of the human gut microbiome in liver cirrhosis
   URL:
   PMIDs: 25079328

   Abstract: A 172 word abstract is available. Use 'abstract' method.
   notes:
    platform_summary:
       NA
    platform_accession:
       NA
    platform_title:
       HiSeq 2000
    platform_technology:
       Illumina
    platform_distribution:
       NA
    platform_manufacturer:
```

```
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 232 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
Quin_gut_liver_cirrhosis
                     232

sampleID:
   Length     Class      Mode
      232 character character

subjectID:
   Length     Class      Mode
      232 character character

bodysite:
stool
  232

disease:
cirrhosis         n
      118       114

age:
   Length     Class      Mode
      232 character character

gender:
female   male
    80    152

country:
china
  232

sequencing_technology:
Illumina
     232

pubmedid:
```

```
25079328
     232

bmi:
   Length     Class      Mode
      232 character character

X.SampleID:
Metaphlan2_Analysis
                232

stage:
 discovery validation
       178         54

cirrhotic:
  n   y
114 118

hbv_related:
  n   y
135  97

alcohol_related:
  n   y
200  32

other_causes_related:
   Length     Class      Mode
      232 character character

inr:
   Length     Class      Mode
      232 character character

crea:
   Length     Class      Mode
      232 character character

alb:
   Length     Class      Mode
      232 character character

tb:
   Length     Class      Mode
      232 character character

pt:
   Length     Class      Mode
      232 character character

ascites:
```

```
absent   mild  sever   NA's
    54     36     28    114

he:
grade_1    none    NA's
      2     116     114

ctp:
   Length     Class      Mode
      232 character character

meld:
   Length     Class      Mode
      232 character character

antivirus:
   Length     Class      Mode
      232 character character

X..blocker:
       none propranolol        NA's
        106          12         114
```

`t2dmeta_long.stool.eset` *A metagenome-wide association study of gut microbiota in type 2 diabetes*

Description

Assessment and characterization of gut microbiota has become a major research area in human disease, including type 2 diabetes, the most prevalent endocrine disease worldwide. To carry out analysis on gut microbial content in patients with type 2 diabetes, we developed a protocol for a metagenome-wide association study (MGWAS) and undertook a two-stage MGWAS based on deep shotgun sequencing of the gut microbial DNA from 345 Chinese individuals. We identified and validated approximately 60,000 type-2-diabetes-associated markers and established the concept of a metagenomic linkage group, enabling taxonomic species-level analyses. MGWAS analysis showed that patients with type 2 diabetes were characterized by a moderate degree of gut microbial dysbiosis, a decrease in the abundance of some universal butyrate-producing bacteria and an increase in various opportunistic pathogens, as well as an enrichment of other microbial functions conferring sulphate reduction and oxidative stress resistance. An analysis of 23 additional individuals demonstrated that these gut microbial markers might be useful for classifying type 2 diabetes.

Usage

```
data( t2dmeta_long.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng
  Laboratory: BGI-Shenzhen, Shenzhen 518083, China.
  Contact information:
  Title: A metagenome-wide association study of gut microbiota in type 2 diabetes
  URL:
  PMIDs: 23023125

  Abstract: A 161 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx and HiSeq 2000
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 290 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
t2dmeta_long
         290

sampleID:
   Length     Class      Mode
      290 character character

subjectID:
   Length     Class      Mode
      290 character character

bodysite:
```

```
stool
  290

disease:
   n  t2d NA's
 136  135   19

age:
   Length     Class      Mode
      290 character character

gender:
female   male   NA's
   119    152     19

country:
china  NA's
  271    19

sequencing_technology:
Illumina
     290

pubmedid:
23023125
     290

bmi:
   Length     Class      Mode
      290 character character

X.SampleID:
Metaphlan2_Analysis
                290

stage:
 stage_i stage_ii     NA's
      72      199       19

height:
   Length     Class      Mode
      290 character character

weight:
   Length     Class      Mode
      290 character character

diabetic:
   n    y NA's
 136  135   19

fbg:
```

```
   Length     Class      Mode
      290 character character

sbp:
   Length     Class      Mode
      290 character character

dbp:
   Length     Class      Mode
      290 character character

fins:
   Length     Class      Mode
      290 character character

fcp:
   Length     Class      Mode
      290 character character

hbalc:
   Length     Class      Mode
      290 character character

tg:
   Length     Class      Mode
      290 character character

tcho:
   Length     Class      Mode
      290 character character

hdl:
   Length     Class      Mode
      290 character character

ldl:
   Length     Class      Mode
      290 character character
```

`t2dmeta_short.stool.eset` *A metagenome-wide association study of gut microbiota in type 2 diabetes*

Description

Assessment and characterization of gut microbiota has become a major research area in human disease, including type 2 diabetes, the most prevalent endocrine disease worldwide. To carry out analysis on gut microbial content in patients with type 2 diabetes, we developed a protocol for a metagenome-wide association study (MGWAS) and undertook a two-stage MGWAS based on deep

shotgun sequencing of the gut microbial DNA from 345 Chinese individuals. We identified and validated approximately 60,000 type-2-diabetes-associated markers and established the concept of a metagenomic linkage group, enabling taxonomic species-level analyses. MGWAS analysis showed that patients with type 2 diabetes were characterized by a moderate degree of gut microbial dysbiosis, a decrease in the abundance of some universal butyrate-producing bacteria and an increase in various opportunistic pathogens, as well as an enrichment of other microbial functions conferring sulphate reduction and oxidative stress resistance. An analysis of 23 additional individuals demonstrated that these gut microbial markers might be useful for classifying type 2 diabetes.

Usage

```
data( t2dmeta_short.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Qin J, Li Y, Cai Z, Li S, Zhu J, Zhang F, Liang S, Zhang W, Guan Y, Shen D, Peng
  Laboratory: BGI-Shenzhen, Shenzhen 518083, China.
  Contact information:
  Title: A metagenome-wide association study of gut microbiota in type 2 diabetes
  URL:
  PMIDs: 23023125

  Abstract: A 161 word abstract is available. Use 'abstract' method.
  notes:
   platform_summary:
      NA
   platform_accession:
      NA
   platform_title:
      GAIIx and HiSeq 2000
   platform_technology:
      Illumina
   platform_distribution:
      NA
   platform_manufacturer:
      Illumina
   processor:
     MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 73 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------
```

```
dataset_name:
t2dmeta_short
          73

sampleID:
   Length     Class      Mode
       73 character character

subjectID:
   Length     Class      Mode
       73 character character

bodysite:
stool
   73

disease:
  n t2d
 38  35

age:
   Length     Class      Mode
       73 character character

gender:
female   male
    35     38

country:
china
   73

sequencing_technology:
Illumina
      73

pubmedid:
23023125
      73

bmi:
   Length     Class      Mode
       73 character character

X.SampleID:
Metaphlan2_Analysis
                 73

stage:
stage_i
     73
```

```
height:
   Length     Class      Mode
       73 character character

weight:
   Length     Class      Mode
       73 character character

diabetic:
 n  y
38 35

fbg:
   Length     Class      Mode
       73 character character

sbp:
   Length     Class      Mode
       73 character character

dbp:
   Length     Class      Mode
       73 character character

fins:
   Length     Class      Mode
       73 character character

hbalc:
   Length     Class      Mode
       73 character character

tg:
   Length     Class      Mode
       73 character character

tcho:
   Length     Class      Mode
       73 character character

hdl:
   Length     Class      Mode
       73 character character

ldl:
   Length     Class      Mode
       73 character character
```

`Tito_subsistence_gut.stool.eset` *Subsistence strategies in traditional societies distinguish gut microbiomes*

Description

Recent studies suggest that gut microbiomes of urban-industrialized societies are different from those of traditional peoples. Here we examine the relationship between lifeways and gut microbiota through taxonomic and functional potential characterization of faecal samples from hunter-gatherer and traditional agriculturalist communities in Peru and an urban-industrialized community from the US. We find that in addition to taxonomic and metabolic differences between urban and traditional lifestyles, hunter-gatherers form a distinct sub-group among traditional peoples. As observed in previous studies, we find that Treponema are characteristic of traditional gut microbiomes. Moreover, through genome reconstruction (2.2-2.5???MB, coverage depth ?? 26-513) and functional potential characterization, we discover these Treponema are diverse, fall outside of pathogenic clades and are similar to Treponema succinifaciens, a known carbohydrate metabolizer in swine. Gut Treponema are found in non-human primates and all traditional peoples studied to date, suggesting they are symbionts lost in urban-industrialized societies.

Usage

```
data( Tito_subsistence_gut.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Obregon-Tito AJ, Tito RY, Metcalf J, Sankaranarayanan K, Clemente JC, Ursell L
  Laboratory: 1] Department of Anthropology, University of Oklahoma, Dale Hall Tower, 521  Norman,
   Contact information:
  Title: Subsistence strategies in traditional societies distinguish gut microbiomes
   URL:
   PMIDs: 25807110

   Abstract: A 146 word abstract is available. Use 'abstract' method.
   notes:
    platform_summary:
       NA
    platform_accession:
       NA
    platform_title:
       Illumina
    platform_technology:
       Illumina
    platform_distribution:
       NA
    platform_manufacturer:
       Illumina
    processor:
      MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 58 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
Tito_subsistence_gut
                  58

sampleID:
   Length     Class      Mode
       58 character character

subjectID:
   Length     Class      Mode
       58 character character

bodysite:
stool
   58

disease:
          n       obese  overweight underweight        NA's
         35           5          10           1           7

age:
   Length     Class      Mode
       58 character character

gender:
female   male     na
    30     27      1

country:
peru  usa
  36   22

sequencing_technology:
Illumina
      58

pubmedid:
25807110
      58

bmi:
```

```
   Length     Class      Mode
       58 character character

X.SampleID:
Metaphlan2_Analysis
                 58

ethnicity:
unknown   white
     36      22

population:
  matses   norman tunapuco
      24       22       12

bmi_class:
    healthy          na       obese  overweight underweight
         35           7           5          10           1

X16s_rrna:
yes
 58

shotgun_metagenome:
yes
 58

read_depth_16s:
   Length     Class      Mode
       58 character character

paired:
   Length     Class      Mode
       58 character character

single:
   Length     Class      Mode
       58 character character
```

`WT2D.stool.eset` *Gut metagenome in European women with normal, impaired and diabetic glucose control*

Description

Type 2 diabetes (T2D) is a result of complex gene-environment interactions, and several risk factors have been identified, including age, family history, diet, sedentary lifestyle and obesity. Statistical models that combine known risk factors for T2D can partly identify individuals at high risk of developing the disease. However, these studies have so far indicated that human genetics contributes little to the models, whereas socio-demographic and environmental factors have greater influence.

Recent evidence suggests the importance of the gut microbiota as an environmental factor, and an altered gut microbiota has been linked to metabolic diseases including obesity, diabetes and cardiovascular disease. Here we use shotgun sequencing to characterize the faecal metagenome of 145 European women with normal, impaired or diabetic glucose control. We observe compositional and functional alterations in the metagenomes of women with T2D, and develop a mathematical model based on metagenomic profiles that identified T2D with high accuracy. We applied this model to women with impaired glucose tolerance, and show that it can identify women who have a diabetes-like metabolism. Furthermore, glucose control and medication were unlikely to have major confounding effects. We also applied our model to a recently described Chinese cohort and show that the discriminant metagenomic markers for T2D differ between the European and Chinese cohorts. Therefore, metagenomic predictive tools for T2D should be specific for the age and geographical location of the populations studied.

Usage

```
data( WT2D.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Karlsson FH, Tremaroli V, Nookaew I, Bergstr??m G, Behre CJ, Fagerberg B, Niel
  Laboratory: Department of Chemical and Biological Engineering, Chalmers University of Technology
   Contact information:
  Title: Gut metagenome in European women with normal, impaired and diabetic glucose control
   URL:
   PMIDs: 23719380

   Abstract: A 229 word abstract is available. Use 'abstract' method.
   notes:
    platform_summary:
       NA
    platform_accession:
       NA
    platform_title:
       HiSeq 2000
    platform_technology:
       Illumina
    platform_distribution:
       NA
    platform_manufacturer:
       Illumina
    processor:
      MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)

featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 145 samples
Platform type: NA
```

```
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
WT2D
 145

sampleID:
   Length     Class      Mode
      145 character character

subjectID:
   Length     Class      Mode
      145 character character

bodysite:
stool
  145

disease:
impaired_glucose_tolerance                          n                       t2d
                        49                         43                        53

age:
   Length     Class      Mode
      145 character character

country:
   Length     Class      Mode
      145 character character

sequencing_technology:
Illumina
     145

pubmedid:
23719380
     145

bmi:
   Length     Class      Mode
      145 character character

X.SampleID:
Metaphlan2_Analysis
                145

hdl:
   Length     Class      Mode
      145 character character
```

```
ldl:
   Length     Class      Mode
      145 character character

classification:
igt ngt t2d
 49  43  53

gad.antibodies:
   Length     Class      Mode
      145 character character

whr:
   Length     Class      Mode
      145 character character

wc:
   Length     Class      Mode
      145 character character

cholesterol:
   Length     Class      Mode
      145 character character

triglycerides:
   Length     Class      Mode
      145 character character

creatinine:
   Length     Class      Mode
      145 character character

y.gt:
   Length     Class      Mode
      145 character character

fasting_glucose:
   Length     Class      Mode
      145 character character

fasting_insulin:
   Length     Class      Mode
      145 character character

hba1c:
   Length     Class      Mode
      145 character character

adiponectin:
   Length     Class      Mode
      145 character character
```

```
leptin:
   Length     Class      Mode
      145 character character

glp.1:
   Length     Class      Mode
      145 character character

fgf.19:
   Length     Class      Mode
      145 character character

hscrp:
   Length     Class      Mode
      145 character character

c.peptide:
   Length     Class      Mode
      145 character character

tnfa:
   Length     Class      Mode
      145 character character

il.1:
   Length     Class      Mode
      145 character character

cd163:
   Length     Class      Mode
      145 character character

statins:
 n  y
93 52

insulin:
  n   y
139   6

oral_anti.diabetic_medication:
  met sulph  NA's
   20     2   123

years_in_sweden:
   Length     Class      Mode
      145 character character
```

`Zeller_fecal_colorectal_cancer.stool.eset`
Potential of fecal microbiota for early-stage detection of colorectal cancer

---

Description

Several bacterial species have been implicated in the development of colorectal carcinoma (CRC), but CRC-associated changes of fecal microbiota and their potential for cancer screening remain to be explored. Here, we used metagenomic sequencing of fecal samples to identify taxonomic markers that distinguished CRC patients from tumor-free controls in a study population of 156 participants. Accuracy of metagenomic CRC detection was similar to the standard fecal occult blood test (FOBT) and when both approaches were combined, sensitivity improved > 45% relative to the FOBT, while maintaining its specificity. Accuracy of metagenomic CRC detection did not differ significantly between early- and late-stage cancer and could be validated in independent patient and control populations (N = 335) from different countries. CRC-associated changes in the fecal microbiome at least partially reflected microbial community composition at the tumor itself, indicating that observed gene pool differences may reveal tumor-related host-microbe interactions. Indeed, we deduced a metabolic shift from fiber degradation in controls to utilization of host carbohydrates and amino acids in CRC patients, accompanied by an increase of lipopolysaccharide metabolism.

Usage

```
data( Zeller_fecal_colorectal_cancer.stool.marker_ab.eset )
```

Format

```
experimentData(eset):
Experiment data
  Experimenter name: Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, Costea PI, Amiot A, B??hm J,
  Laboratory: Structural and Computational Biology Unit, European Molecular Biology Laboratory, He
   Contact information:
  Title: Potential of fecal microbiota for early-stage detection of colorectal cancer
   URL:
   PMIDs: 25432777

   Abstract: A 175 word abstract is available. Use 'abstract' method.
   notes:
    platform_summary:
       NA
    platform_accession:
       NA
    platform_title:
       HiSeq 2000/2500
    platform_technology:
       Illumina
    platform_distribution:
       NA
    platform_manufacturer:
       Illumina
    processor:
      MetaPhlAn2 (marker & abundance) / HUMAnN2 (geneFam., pathAbun., & pathCov.)
```

```
featureData(eset):
An object of class 'AnnotatedDataFrame': none
```

Details

```
assayData: 302989 features, 134 samples
Platform type: NA
---------------------------
Available sample meta-data:
---------------------------

dataset_name:
Zeller_fecal_colorectal_cancer
                          134

sampleID:
   Length     Class      Mode
      134 character character

subjectID:
   Length     Class      Mode
      134 character character

bodysite:
stool
  134

disease:
       cancer large_adenoma             n small_adenoma
           48            13            47            26

age:
   Length     Class      Mode
      134 character character

gender:
female   male
    59     75

country:
france
   134

sequencing_technology:
Illumina
     134

pubmedid:
25432777
     134
```

```
bmi:
   Length     Class      Mode
      134 character character

X.SampleID:
Metaphlan2_Analysis
                134

tnm_stage:
   Length     Class      Mode
      134 character character

ajcc_stage:
  i  ii iii  iv  na
 15   6   8  19  86

localization:
   Length     Class      Mode
      134 character character

fobt:
negative positive
     102       32

wif.1_gene_methylation_test:
      na negative positive
       4      104       26

group:
control     crc      na
     73      48      13
```

Index