Package ‘AssessORFData’

March 7, 2024

Type Package
Title Data and Files for the AssessORF Package
Version 1.20.0
Date 2019-04-07
Description This package provides access to mapping and results objects generated by the AssessORF package, as well as the genome sequences for the strains corresponding to those objects.
Depends R (>= 3.5.0)
Imports DECIPHER, utils
Suggests AssessORF, BiocStyle, knitr, rmarkdown
biocViews OrganismData, Bacillus_subtilis_Data, Escherichia_coli_Data, Pseudomonas_aeruginosa_Data, Staphylococcus_aureus_Data, Genome, Proteome, SequencingData
License GPL-3
Encoding UTF-8
LazyData FALSE
NeedsCompilation no
RoxygenNote 6.1.1
VignetteBuilder knitr
git_url https://git.bioconductor.org/packages/AssessORFData
git_branch RELEASE_3_18
git_last_commit ce79d8f
git_last_commit_date 2023-10-24
Repository Bioconductor 3.18
Date/Publication 2024-03-07
Author Deepank Korandla [aut, cre]
Maintainer Deepank Korandla <dkorandl@alumni.cmu.edu>
R topics documented:

- AP1 ................................................................. 2
- AssessORF_StrainIDs ................................. 4
- ATCC11842 .................................................. 4
- ATCC13032 .................................................. 6
- ATCC17978 .................................................. 8
- ATCC700084 ............................................... 9
- BW25113 .................................................... 11
- CCMP1375 .................................................. 13
- CECT5344 ................................................... 14
- CNRZ327 .................................................... 16
- COH1 .......................................................... 18
- D_UW_3_CX ............................................... 19
- EGD_e .......................................................... 21
- GetDataMapObj ........................................... 23
- GetGeneSources .......................................... 23
- GetResultsObj ............................................ 24
- GetStrainIDs ............................................... 25
- H37Rv ........................................................ 25
- HG001 ........................................................ 27
- Houston_1 .................................................. 28
- II1403 ........................................................ 30
- K_12_MG1655 ............................................. 32
- LAL14_1 ..................................................... 33
- MG1363 ..................................................... 35
- MGAS5005 .................................................. 36
- NCIB_3610 .................................................. 38
- PAO1 .......................................................... 40
- SaveGenomeToPath ...................................... 41
- SL1344 ....................................................... 42
- Strain10403S ................................ ............. 44
- Strain168 ................................................... 45
- TCH1516 ..................................................... 47

Index

| AP1 | Assessment Objects for Streptococcus pyogenes strain AP1 |

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Streptococcus pyogenes strain AP1
Usage

data(AP1_PreSaved_DataMapObj)
data(AP1_PreSaved_ResultsObj_GenBank)
data(AP1_PreSaved_ResultsObj_GeneMarkS2)
data(AP1_PreSaved_ResultsObj_Glimmer)
data(AP1_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'AP1_PreSaved_DataMapObj' is an object of subclass DataMap.
'AP1_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'AP1_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'AP1_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'AP1_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

• StrainID: AP1
• Species: S. pyogenes

Mapping object

The mapping object, 'AP1_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD006345. The related genomes used to determine evolutionary conservation all came from the family, Streptococcaceae excluding anomalous genomes and non-complete *Streptococcus pneumoniae* genomes. Links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'AP1_PreSaved_ResultsObj_GenBank', 'AP1_PreSaved_ResultsObj_GeneMarkS2', 'AP1_PreSaved_ResultsObj_Glimmer', and 'AP1_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GeneMarkS-2 (web server), Prodigal (2.6.3), and Glimmer...
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object. The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object. The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

**Getting the strain’s genome**

Use `SaveGenomeToPath("AP1", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

**Source**

Proteomcis data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006345](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006345)


---

**AssessORF_StrainIDs**  
*Vector of strain IDs used in the package*

**Description**

Vector of strain IDs, which describes the strains for which the package has data

**Usage**

data(AssessORF_StrainIDs)

**Format**

Character vector of length 26

---

**ATCC11842**  
*Assessment Objects for Lactobacillus delbrueckii subsp. bulgaricus strain ATCC 11842*

**Description**

Objects of class Assessment and either subclass DataMap or subclass Results for *Lactobacillus delbrueckii subsp. bulgaricus* strain ATCC 11842
Usage

data(ATCC11842_PreSaved_DataMapObj)
data(ATCC11842_PreSaved_ResultsObj_GenBank)
data(ATCC11842_PreSaved_ResultsObj_GeneMarkS2)
data(ATCC11842_PreSaved_ResultsObj_Glimmer)
data(ATCC11842_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'ATCC11842_PreSaved_DataMapObj' is an object of subclass DataMap.

'ATCC11842_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'ATCC11842_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'ATCC11842_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'ATCC11842_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

• StrainID: ATCC11842
• Species: L. delbrueckii bulgaricus

Mapping object

The mapping object, 'ATCC11842_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD006551. The related genomes used to determine evolutionary conservation all came from the genus Lactobacillus, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'ATCC11842_PreSaved_ResultsObj_GenBank', 'ATCC11842_PreSaved_ResultsObj_GeneMarkS2', 'ATCC11842_PreSaved_ResultsObj_Glimmer', and 'ATCC11842_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

**Getting the strain’s genome**

Use `Save GenomeToPath("ATCC11842", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

**Source**

Proteomcis data: [http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXD006551](http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXD006551)


---

<table>
<thead>
<tr>
<th>ATCC13032</th>
<th>Assessment Objects for Corynebacterium glutamicum strain ATCC 13032</th>
</tr>
</thead>
</table>

**Description**

Objects of class Assessment and either subclass DataMap or subclass Results for *Corynebacterium glutamicum* strain ATCC 13032

**Usage**

```r
data(ATCC13032_PreSaved_DataMapObj)
data(ATCC13032_PreSaved_ResultsObj_GenBank)
data(ATCC13032_PreSaved_ResultsObj_GeneMarkS2)
data(ATCC13032_PreSaved_ResultsObj_Glimmer)
data(ATCC13032_PreSaved_ResultsObj_Prodigal)
```

**Format**

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.
Details

'ATCC13032_PreSaved_DataMapObj' is an object of subclass DataMap.

'ATCC13032_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'ATCC13032_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'ATCC13032_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'ATCC13032_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: ATCC13032
- Species: C. glutamicum

Mapping object

The mapping object, 'ATCC13032_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD005812. The related genomes used to determine evolutionary conservation came from the genera Corynebacterium, Dietzia, and Tsukamurella, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, ‘ATCC13032_PreSaved_ResultsObj_GenBank’, ‘ATCC13032_PreSaved_ResultsObj_GeneMarkS2’, ‘ATCC13032_PreSaved_ResultsObj_Glimmer’, and ‘ATCC13032_PreSaved_ResultsObj_Prodigal’, store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("ATCC13032", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomcis data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD005812
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

ATCC17978 Assessment Objects for Acinetobacter baumannii strain ATCC 17978

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Acinetobacter baumannii strain ATCC 17978

Usage

data(ATCC17978_PreSaved_DataMapObj)
data(ATCC17978_PreSaved_ResultsObj_GenBank)
data(ATCC17978_PreSaved_ResultsObj_GeneMarkS2)
data(ATCC17978_PreSaved_ResultsObj_Glimmer)
data(ATCC17978_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'ATCC17978_PreSaved_DataMapObj' is an object of subclass DataMap.
'ATCC17978_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'ATCC17978_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'ATCC17978_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'ATCC17978_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

• StrainID: ATCC17978
• Species: A. baumannii
Mapping object

The mapping object, 'ATCC17978_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The proteomics data was generated in-house and collected as part of the work for the AssessORF paper. The raw mass spectra data can be found at ProteomeXchange dataset PXD012539. The related genomes used to determine evolutionary conservation all came from the genus Acinetobacter, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the mapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'ATCC17978_PreSaved_ResultsObj_GenBank', 'ATCC17978_PreSaved_ResultsObj_GeneMarkS2', 'ATCC17978_PreSaved_ResultsObj_Glimmer', and 'ATCC17978_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("ATCC17978", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD012539
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

ATCC700084 Assessment Objects for Mycobacterium smegmatis strain ATCC 700084

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Mycobacterium smegmatis strain ATCC 700084
Usage

```r
data(ATCC700084_PreSaved_DataMapObj)
data(ATCC700084_PreSaved_ResultsObj_GenBank)
data(ATCC700084_PreSaved_ResultsObj_GeneMarkS2)
data(ATCC700084_PreSaved_ResultsObj_Glimmer)
data(ATCC700084_PreSaved_ResultsObj_Prodigal)
```

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'ATCC700084_PreSaved_DataMapObj' is an object of subclass DataMap.

'ATCC700084_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'ATCC700084_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'ATCC700084_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'ATCC700084_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: ATCC700084
- Species: M. smegmatis

Mapping object

The mapping object, 'ATCC700084_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD003500. The related genomes used to determine evolutionary conservation all came from the genus Mycobacterium, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'ATCC700084_PreSaved_ResultsObj_GenBank', 'ATCC700084_PreSaved_ResultsObj_GeneMarkS2', 'ATCC700084_PreSaved_ResultsObj_Glimmer', and 'ATCC700084_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

**Getting the strain’s genome**

Use `SaveGenomeToPath("ATCC700084", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine. Please note that there is no genome sequence available for strain ATCC 700084 so the reference genome from strain MC2 155 was used instead.

**Source**

Proteomcis data: [http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXD003500](http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXD003500)


---

**BW25113 Assessment Objects for Escherichia coli strain BW25113**

**Description**

Objects of class Assessment and either subclass DataMap or subclass Results for *Escherichia coli* strain BW25113

**Usage**

```r
data(BW25113_PreSaved_DataMapObj)
data(BW25113_PreSaved_ResultsObj_GenBank)
data(BW25113_PreSaved_ResultsObj_GeneMarkS2)
data(BW25113_PreSaved_ResultsObj_Glimmer)
data(BW25113_PreSaved_ResultsObj_Prodigal)
```

**Format**

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.
Details

'BW25113_PreSaved_DataMapObj' is an object of subclass DataMap.

'BW25113_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'BW25113_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'BW25113_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'BW25113_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: BW25113
- Species: E. coli

Mapping object

The mapping object, 'BW25113_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD000498. The related genomes used to determine evolutionary conservation all came from the genus Escherichia, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'BW25113_PreSaved_ResultsObj_GenBank', 'BW25113_PreSaved_ResultsObj_GeneMarkS2', 'BW25113_PreSaved_ResultsObj_Glimmer', and 'BW25113_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("BW25113", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
CCMP1375

Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000498
NCBI's Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/
GenBank record: https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP009273.1

CCMP1375 Assessment Objects for Prochlorococcus marinus subsp. marinus strain CCMP1375

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Prochlorococcus marinus subsp. marinus strain CCMP1375

Usage

data(CCMP1375_PreSaved_DataMapObj)
data(CCMP1375_PreSaved_ResultsObj_GenBank)
data(CCMP1375_PreSaved_ResultsObj_GeneMarkS2)
data(CCMP1375_PreSaved_ResultsObj_Glimmer)
data(CCMP1375_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'CCMP1375_PreSaved_DataMapObj' is an object of subclass DataMap.
'CCMP1375_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'CCMP1375_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'CCMP1375_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'CCMP1375_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: CCMP1375
- Species: P. marinus
Mapping object

The mapping object, 'CCMP1375_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD005745. The related genomes used to determine evolutionary conservation all came from the order Synechococcales, and links to their sequences were downloaded from NCBI's Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'CCMP1375_PreSaved_ResultsObj_GenBank', 'CCMP1375_PreSaved_ResultsObj_GeneMarkS2', 'CCMP1375_PreSaved_ResultsObj_Glimmer', and 'CCMP1375_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain's genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain's genome

Use SaveGenomeToPath("CCMP1375", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomic data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX005745
NCBI's Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

Objects of class Assessment and either subclass DataMap or subclass Results for Pseudomonas pseudoalcaligenes strain CECT 5344
Usage

data(CECT5344_PreSaved_DataMapObj)
data(CECT5344_PreSaved_ResultsObj_GenBank)
data(CECT5344_PreSaved_ResultsObj_GeneMarkS2)
data(CECT5344_PreSaved_ResultsObj_Glimmer)
data(CECT5344_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'CECT5344_PreSaved_DataMapObj' is an object of subclass DataMap.
'CECT5344_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'CECT5344_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'CECT5344_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'CECT5344_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: CECT5344
- Species: P. pseudoalcaligenes

Mapping object

The mapping object, 'CECT5344_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD005745. The related genomes used to determine evolutionary conservation all came from the genus Pseudomonas, and links to their sequences were downloaded from NCBI's Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'CECT5344_PreSaved_ResultsObj_GenBank', 'CECT5344_PreSaved_ResultsObj_GeneMarkS2', 'CECT5344_PreSaved_ResultsObj_Glimmer', and 'CECT5344_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain's genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GeneMarkS-2 (web server), Prodigal (2.6.3), and Glimmer
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

**Getting the strain’s genome**

Use SaveGenomeToPath("CECT5344", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

**Source**

Proteomcis data: [http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXDO05745](http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXDO05745)


---

**Description**

Objects of class Assessment and either subclass DataMap or subclass Results for *Lactobacillus delbrueckii* subsp. *lactis* strain LBCNRZ327_V11

**Usage**

data(CNRZ327_PreSaved_DataMapObj)
data(CNRZ327_PreSaved_ResultsObj_GenBank)
data(CNRZ327_PreSaved_ResultsObj_GeneMarkS2)
data(CNRZ327_PreSaved_ResultsObj_Glimmer)
data(CNRZ327_PreSaved_ResultsObj_Prodigal)

**Format**

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.
Details

'CNRZ327_PreSaved_DataMapObj' is an object of subclass DataMap.

'CNRZ327_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'CNRZ327_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'CNRZ327_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'CNRZ327_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: CNRZ327
- Species: L. delbrueckii lactis

Mapping object

The mapping object, 'CNRZ327_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD006551. The related genomes used to determine evolutionary conservation all came from the genus Lactobacillus, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'CNRZ327_PreSaved_ResultsObj_GenBank', 'CNRZ327_PreSaved_ResultsObj_GeneMarkS2', 'CNRZ327_PreSaved_ResultsObj_Glimmer', and 'CNRZ327_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GeneMarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("CNRZ327", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomcis data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006551](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006551)

---

**COH1**  
*Assessment Objects for Streptococcus agalactiae strain COH1*

Description

Objects of class Assessment and either subclass DataMap or subclass Results for *Streptococcus agalactiae* strain COH1

Usage

data(COH1_PreSaved_DataMapObj)  
data(COH1_PreSaved_ResultsObj_GenBank)  
data(COH1_PreSaved_ResultsObj_GeneMarkS2)  
data(COH1_PreSaved_ResultsObj_Glimmer)  
data(COH1_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'COH1_PreSaved_DataMapObj' is an object of subclass DataMap.
'COH1_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'COH1_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'COH1_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'COH1_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: COH1
- Species: *S. agalactiae*
Mapping object

The mapping object, 'COH1_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The proteomics data was generated in-house and collected as part of the work for the AssessORF paper. The raw mass spectra data can be found at ProteomeXchange dataset PXD012567. The related genomes used to determine evolutionary conservation all came from the family Streptococcaceae, excluding anomalous genomes and non-complete Streptococcus pneumoniae genomes. Links to their sequences were downloaded from NCBI's Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'COH1_PreSaved_ResultsObj_GenBank', 'COH1_PreSaved_ResultsObj_GeneMarkS2', 'COH1_PreSaved_ResultsObj_Glimmer', and 'COH1_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain's genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("COH1", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomics data: http://proteomecentral.proteomeexchange.org/cgi/GetDataset?ID=PXD012567
NCBI's Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/
Usage

data(D_UW_3_CX_PreSaved_DataMapObj)
data(D_UW_3_CX_PreSaved_ResultsObj_GenBank)
data(D_UW_3_CX_PreSaved_ResultsObj_GeneMarkS2)
data(D_UW_3_CX_PreSaved_ResultsObj_Glimmer)
data(D_UW_3_CX_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'D_UW_3_CX_PreSaved_DataMapObj' is an object of subclass DataMap.
'D_UW_3_CX_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'D_UW_3_CX_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'D_UW_3_CX_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'D_UW_3_CX_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: D_UW_3_CX
- Species: C. trachomatis

Mapping object

The mapping object, 'D_UW_3_CX_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD003883. The related genomes used to determine evolutionary conservation all came from the phylum Chlamydiae, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'D_UW_3_CX_PreSaved_ResultsObj_GenBank', 'D_UW_3_CX_PreSaved_ResultsObj_GeneMarkS2', 'D_UW_3_CX_PreSaved_ResultsObj_Glimmer', and 'D_UW_3_CX_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use `SaveGenomeToPath("D_UW_3_CX", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

Source

Proteomics data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD003883](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD003883)


---

**EGD_e**

**Assessment Objects for Listeria monocytogenes strain EGD-e**

**Description**

Objects of class "Assessment" and either subclass "DataMap" or subclass "Results" for *Listeria monocytogenes* strain EGD-e

**Usage**

- `data(EGD_e_PreSaved_DataMapObj)`
- `data(EGD_e_PreSaved_ResultsObj_GenBank)`
- `data(EGD_e_PreSaved_ResultsObj_GeneMarkS2)`
- `data(EGD_e_PreSaved_ResultsObj_Glimmer)`
- `data(EGD_e_PreSaved_ResultsObj_Prodigal)`

**Format**

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the [AssessORF](https://www.ncbi.nlm.nih.gov/nuccore/AE001273.1) package.

**Details**

- 'EGD_e_PreSaved_DataMapObj’ is an object of subclass DataMap.
- 'EGD_e_PreSaved_ResultsObj_GenBank’ is an object of subclass Results with predicted genes from the GenBank database.
- 'EGD_e_PreSaved_ResultsObj_GeneMarkS2’ is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'EGD_e_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'EGD_e_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- **StrainID**: EGD_e
- **Species**: L. monocytogenes

Mapping object

The mapping object, 'EGD_e_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD000890. The related genomes used to determine evolutionary conservation all came from the genus Listeria, and links to their sequences were downloaded from NCBI's Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'EGD_e_PreSaved_ResultsObj_GenBank', 'EGD_e_PreSaved_ResultsObj_GeneMarkS2', 'EGD_e_PreSaved_ResultsObj_Glimmer', and 'EGD_e_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain's genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain's genome

Use SaveGenomeToPath("EGD_e", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomics data: http://proteomcentral.proteomexchange.org/cgi/GetDataset?ID=PXD000890
NCBI's Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/
**GetDataMapObj**

*Get a Data Map Object*

**Description**

Gets and returns the data map object for a specific strain

**Usage**

`GetDataMapObj(strainID)`

**Arguments**

- `strainID` Character string corresponding to the strain identifier.

**Details**

`GetDataMapObj` returns an object of class `Assessment` and subclass `DataMap` corresponding to the given strain ID. The given strain ID must be a part of the AssessORF set, and there is no partial matching. Otherwise, the function will error.

**Value**

An object of class `Assessment` and subclass `DataMap`

**Examples**

```r
mapObj <- GetDataMapObj("MGAS5005")
```

---

**GetGeneSources**

*Get the Gene Sources*

**Description**

Returns the list of gene sources used in making the results objects

**Usage**

`GetGeneSources()`

**Details**

`GetGeneSources` returns the list of gene sources (programs and databases) used in making the results objects in the AssessORF set.
GetResultsObj

Value
A character vector where each element corresponds to a single gene source

Examples
```
geneSourceSet <- GetGeneSources()
```

Description
Gets and returns the results object for a specific strain-gene source combo

Usage
```
GetResultsObj(strainID, geneSource = "Prodigal")
```

Arguments
- **strainID**: Character string corresponding to the strain identifier.
- **geneSource**: Character string corresponding to the gene source.

Details
GetDataMapObj returns an object of class Assessment and subclass Results corresponding to the given strain ID and the given gene source. The given strain ID must be a part of the AssessORF set, and there is no partial matching. Otherwise, the function will error. The given gene source must also be a part of the AssessORF set, but the function ignores case when checking if the given gene source is a part of the set.

Value
An object of class Assessment and subclass Results

Examples
```
resObj1 <- GetResultsObj("MGAS5005", "Prodigal")
resObj2 <- GetResultsObj("MGAS5005", "GenBank")
resObj3 <- GetResultsObj("MGAS5005", "GeneMarkS2")
resObj4 <- GetResultsObj("MGAS5005", "Glimmer")
```
GetStrainIDs

Get the Strain Identifiers

Description

Returns the list of strain identifiers for which the package has data

Usage

GetStrainIDs()

Details

GetStrainIDs returns the list of strains in the AssessORF set. This function is a shorter alternative to data("AssessORF_StrainIDs").

Value

A character vector where each element corresponds to a single strain identifier

Examples

allStrainIDs <- GetStrainIDs()

H37Rv Assessment Objects for Mycobacterium tuberculosis strain H37Rv

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Mycobacterium tuberculosis strain H37Rv

Usage

data(H37Rv_PreSaved_DataMapObj)
data(H37Rv_PreSaved_ResultsObj_GenBank)
data(H37Rv_PreSaved_ResultsObj_GeneMarkS2)
data(H37Rv_PreSaved_ResultsObj_Glimmer)
data(H37Rv_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.
Details

'H37Rv_PreSaved_DataMapObj' is an object of subclass DataMap.
'H37Rv_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'H37Rv_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'H37Rv_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'H37Rv_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- **StrainID**: H37Rv
- **Species**: M. tuberculosis

Mapping object

The mapping object, 'H37Rv_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD006117. The related genomes used to determine evolutionary conservation all came from the genus Mycobacterium, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'H37Rv_PreSaved_ResultsObj_GenBank', 'H37Rv_PreSaved_ResultsObj_GeneMarkS2', 'H37Rv_PreSaved_ResultsObj_Glimmer', and 'H37Rv_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("H37Rv", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006117
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

Assessment Objects for Staphylococcus aureus subsp. aureus strain HG001

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Staphylococcus aureus subsp. aureus strain HG001

Usage

data(HG001_PreSaved_DataMapObj)
data(HG001_PreSaved_ResultsObj_GenBank)
data(HG001_PreSaved_ResultsObj_GeneMarkS2)
data(HG001_PreSaved_ResultsObj_Glimmer)
data(HG001_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'HG001_PreSaved_DataMapObj’ is an object of subclass DataMap.
'HG001_PreSaved_ResultsObj_GenBank’ is an object of subclass Results with predicted genes from the GenBank database.
'HG001_PreSaved_ResultsObj_GeneMarkS2’ is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'HG001_PreSaved_ResultsObj_Glimmer’ is an object of subclass Results with predicted genes from the program Glimmer.
'HG001_PreSaved_ResultsObj_Prodigal’ is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: HG001
- Species: S. aureus
Mapping object

The mapping object, 'HG001_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD000702. The related genomes used to determine evolutionary conservation all came from the genus Staphylococcus, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'HG001_PreSaved_ResultsObj_GenBank', 'HG001_PreSaved_ResultsObj_GeneMarkS2', 'HG001_PreSaved_ResultsObj_Glimmer', and 'HG001_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("HG001", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine. Note that this genome comes from accession CP018205.1, which has the same genome sequence as accession NZ_CP018205.1. Unlike NZ_CP018205.1 however, CP018205.1 does not have any associated GenBank genes.

Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000702
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/
GenBank record: https://www.ncbi.nlm.nih.gov/nuccore/NZ_CP018205.1

Houston_1 Assessment Objects for Bartonella henselae strain Houston-1

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Bartonella henselae strain Houston-1
Usage

data(Houston_1_PreSaved_DataMapObj)
data(Houston_1_PreSaved_ResultsObj_GenBank)
data(Houston_1_PreSaved_ResultsObj_GeneMarkS2)
data(Houston_1_PreSaved_ResultsObj_Glimmer)
data(Houston_1_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'Houston_1_PreSaved_DataMapObj' is an object of subclass DataMap.
'Houston_1_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'Houston_1_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS2.
'Houston_1_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'Houston_1_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: Houston_1
- Species: B. henselae

Mapping object

The mapping object, 'Houston_1_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD000153. The related genomes used to determine evolutionary conservation came from the families Bartonellaceae, Brucellaceae, Phyllobacteriaceae, Rhizobiaceae. Links to their sequences were downloaded from NCBI's Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'Houston_1_PreSaved_ResultsObj_GenBank', 'Houston_1_PreSaved_ResultsObj_GeneMarkS2', 'Houston_1_PreSaved_ResultsObj_Glimmer', and 'Houston_1_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain's genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object. The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object. The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

**Getting the strain’s genome**

Use `SaveGenomeToPath("Houston_1", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

**Source**

Proteomcis data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000153](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000153)


---

**II1403**

Assessment *Objects for Lactococcus lactis subsp. lactis strain II1403*

**Description**

Objects of class `Assessment` and either subclass `DataMap` or subclass `Results` for *Lactococcus lactis subsp. lactis* strain II1403

**Usage**

```r
data(II1403_PreSaved_DataMapObj)
data(II1403_PreSaved_ResultsObj_GenBank)
data(II1403_PreSaved_ResultsObj_GeneMarkS2)
data(II1403_PreSaved_ResultsObj_Glimmer)
data(II1403_PreSaved_ResultsObj_Prodigal)
```

**Format**

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the `AssessORF` package.

**Details**

'II1403_PreSaved_DataMapObj' is an object of subclass `DataMap`.

'II1403_PreSaved_ResultsObj_GenBank' is an object of subclass `Results` with predicted genes from the GenBank database.

'II1403_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass `Results` with predicted genes from the program GeneMarkS-2.
'Il1403_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'Il1403_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- **StrainID**: Il1403
- **Species**: L. lactis

### Mapping object

The mapping object, 'Il1403_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD000494. The related genomes used to determine evolutionary conservation all came from the family, Streptococcaceae excluding anomalous genomes and non-complete *Streptococcus pneumoniae* genomes. Links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

### Results objects

The 4 results objects, 'Il1403_PreSaved_ResultsObj_GenBank', 'Il1403_PreSaved_ResultsObj_GeneMarkS2', 'Il1403_PreSaved_ResultsObj_Glimmer', and 'Il1403_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GeneMarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The_geneLeftPos, _GeneRightPos, and the_GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The_GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

### Getting the strain’s genome

Use SaveGenomeToPath("Il1403", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

### Source

- Proteomics data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000494](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD000494)
K_12_MG1655 Assessment Objects for Escherichia coli strain K-12 substrain MG1655

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Escherichia coli strain K-12 substrain MG1655

Usage

data(K_12_MG1655_PreSaved_DataMapObj)
data(K_12_MG1655_PreSaved_ResultsObj_GenBank)
data(K_12_MG1655_PreSaved_ResultsObj_GeneMarkS2)
data(K_12_MG1655_PreSaved_ResultsObj_Glimmer)
data(K_12_MG1655_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'K_12_MG1655_PreSaved_DataMapObj' is an object of subclass DataMap.

'K_12_MG1655_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'K_12_MG1655_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'K_12_MG1655_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'K_12_MG1655_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: K_12_MG1655
- Species: E. coli

Mapping object

The mapping object, 'K_12_MG1655_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD005901. The related genomes used to determine evolutionary conservation all came from the genus Escherichia, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.
Results objects

The 4 results objects, ‘K_12_MG1655_PreSaved_ResultsObj_GenBank’, ‘K_12_MG1655_PreSaved_ResultsObj_GeneMarkS2’, ‘K_12_MG1655_PreSaved_ResultsObj_Glimmer’, and ‘K_12_MG1655_PreSaved_ResultsObj_Prodigal’, store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("K_12_MG1655", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD005901
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

LAL14_1 Assessment Objects for Sulfolobus islandicus strain LAL14/1

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Sulfolobus islandicus strain LAL14/1

Usage

data(LAL14_1_PreSaved_DataMapObj)
data(LAL14_1_PreSaved_ResultsObj_GenBank)
data(LAL14_1_PreSaved_ResultsObj_GeneMarkS2)
data(LAL14_1_PreSaved_ResultsObj_Glimmer)
data(LAL14_1_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.
Details

'LAL14_1_PreSaved_DataMapObj' is an object of subclass DataMap.

'LAL14_1_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'LAL14_1_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'LAL14_1_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'LAL14_1_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- **StrainID**: LAL14_1
- **Species**: S. islandicus

Mapping object

The mapping object, 'LAL14_1_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD003074. The related genomes used to determine evolutionary conservation all came from the phylum Crenarchaeota, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'LAL14_1_PreSaved_ResultsObj_GenBank', 'LAL14_1_PreSaved_ResultsObj_GeneMarkS2', 'LAL14_1_PreSaved_ResultsObj_Glimmer', and 'LAL14_1_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("LAL14_1", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD003074
NCBI's Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

MG1363 Assessment Objects for Lactococcus lactis subsp. cremoris strain MG1363

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Lactococcus lactis subsp. cremoris strain MG1363

Usage

data(MG1363_PreSaved_DataMapObj)
data(MG1363_PreSaved_ResultsObj_GenBank)
data(MG1363_PreSaved_ResultsObj_GeneMarkS2)
data(MG1363_PreSaved_ResultsObj_Glimmer)
data(MG1363_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'MG1363_PreSaved_DataMapObj' is an object of subclass DataMap.
'MG1363_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'MG1363_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'MG1363_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'MG1363_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: MG1363
- Species: L. lactis
Mapping object

The mapping object, 'MG1363_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD011263. The related genomes used to determine evolutionary conservation all came from the family, Streptococcaceae excluding anomalous genomes and non-complete Streptococcus pneumoniae genomes. Links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'MG1363_PreSaved_ResultsObj_GenBank', 'MG1363_PreSaved_ResultsObj_GeneMarkS2', 'MG1363_PreSaved_ResultsObj_Glimmer', and 'MG1363_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("MG1363", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomic data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD011263
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

MGAS5005 Assessment Objects for Streptococcus pyogenes strain MGAS5005

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Streptococcus pyogenes strain MGAS5005
Usage

data(MGAS5005_PreSaved_DataMapObj)
data(MGAS5005_PreSaved_ResultsObj_GenBank)
data(MGAS5005_PreSaved_ResultsObj_GeneMarkS2)
data(MGAS5005_PreSaved_ResultsObj_Glimmer)
data(MGAS5005_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'MGAS5005_PreSaved_DataMapObj' is an object of subclass DataMap.

'MGAS5005_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'MGAS5005_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'MGAS5005_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'MGAS5005_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: MGAS5005
- Species: S. pyogenes

Mapping object

The mapping object, 'MGAS5005_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The proteomics data was generated in-house and collected as part of the work for the AssessORF paper. The raw mass spectra data can be found at ProteomeXchange dataset PXD012568. The related genomes used to determine evolutionary conservation all came from the family, Streptococcaceae excluding anomalous genomes and non-complete Streptococcus pneumoniae genomes. Links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'MGAS5005_PreSaved_ResultsObj_GenBank', 'MGAS5005_PreSaved_ResultsObj_GeneMarkS2', 'MGAS5005_PreSaved_ResultsObj_Glimmer', and 'MGAS5005_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.
For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

### Getting the strain’s genome

Use `SaveGenomeToPath("MGAS5005", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

### Source

Proteomcis data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX0012568](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX0012568)


---

### NCIB_3610 Assessment Objects for Bacillus subtilis subsp. subtilis strain NCIB 3610

### Description

Objects of class `Assessment` and either subclass `DataMap` or subclass `Results` for *Bacillus subtilis* subsp. *subtilis* strain NCIB 3610

### Usage

```r
data(NCIB_3610_PreSaved_DataMapObj)
data(NCIB_3610_PreSaved_ResultsObj_GenBank)
data(NCIB_3610_PreSaved_ResultsObj_GeneMarkS2)
data(NCIB_3610_PreSaved_ResultsObj_Glimmer)
data(NCIB_3610_PreSaved_ResultsObj_Prodigal)
```

### Format

All 5 objects have a list structure. For specifics on what the two types of `Assessment` objects contain, please see the `AssessORF` package.
Details

'NCIB_3610_PreSaved_DataMapObj' is an object of subclass DataMap.

'NCIB_3610_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'NCIB_3610_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'NCIB_3610_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'NCIB_3610_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: NCIB_3610
- Species: B. subtilis

Mapping object

The mapping object, 'NCIB_3610_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD006444. The related genomes used to determine evolutionary conservation all came from the genus Bacillus, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'NCIB_3610_PreSaved_ResultsObj_GenBank', 'NCIB_3610_PreSaved_ResultsObj_GeneMarkS2', 'NCIB_3610_PreSaved_ResultsObj_Glimmer', and 'NCIB_3610_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("NCIB_3610", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomic data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD006444
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

---

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Pseudomonas aeruginosa strain PAO1

Usage

data(PAO1_PreSaved_DataMapObj)
data(PAO1_PreSaved_ResultsObj_GenBank)
data(PAO1_PreSaved_ResultsObj_GeneMarkS2)
data(PAO1_PreSaved_ResultsObj_Glimmer)
data(PAO1_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'PAO1_PreSaved_DataMapObj' is an object of subclass DataMap.

'PAO1_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.

'PAO1_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.

'PAO1_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.

'PAO1_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: PAO1
- Species: P. aeruginosa
Mapping object

The mapping object, 'PAO1_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD004560. The related genomes used to determine evolutionary conservation all came from the genus Pseudomonas, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the `MapAssessmentData` function from the `AssessORF` package.

Results objects

The 4 results objects, 'PAO1_PreSaved_ResultsObj_GenBank', 'PAO1_PreSaved_ResultsObj_GeneMarkS2', 'PAO1_PreSaved_ResultsObj_Glimmer', and 'PAO1_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the `AssessGenes` function from the `AssessORF` package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The `GeneLeftPos`, `GeneRightPos`, and the `GeneStrand` within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The `GeneSource` list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use `SaveGenomeToPath("PAO1", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD004560
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

---

**SaveGenomeToPath**

**Save a Strain’s Genome to a Directory**

**Description**

Saves the genome for a specified strain to a given directory

**Usage**

`SaveGenomeToPath(strainID, filePath)`
Arguments

strainID    Character string corresponding to the strain identifier.
filePath    Character string corresponding to the path to the file path. Must end in `.fasta`.

Details

SaveGenomeToPath saves the genome for the specified strain ID to the given file path. If the file specified by the path already exists, it will be overwritten (with a warning). The given strain ID must be a part of the AssessORF set, and there is no partial matching. Otherwise, the function will error.

Note: there is no genome for strain ATCC700084, so a reference genome for the species (strain MC2155) is used instead.

Value

Invisibly returns `filePath`

Examples

```r
tmpFile <- paste0(tempfile(), "\".fasta\")
SaveGenomeToPath("MGASS005", tmpFile)
unlink(tmpFile)
```

SL1344 Assessment Objects for Salmonella enterica subsp. enterica serovar Typhimurium strain SL1344

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Salmonella enterica subsp. enterica serovar Typhimurium strain SL1344

Usage

```r
data(SL1344_PreSaved_DataMapObj)
data(SL1344_PreSaved_ResultsObj_GenBank)
data(SL1344_PreSaved_ResultsObj_GeneMarkS2)
data(SL1344_PreSaved_ResultsObj_Glimmer)
data(SL1344_PreSaved_ResultsObj_Prodigal)
```

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.
Details

'SL1344_PreSaved_DataMapObj' is an object of subclass DataMap.
'SL1344_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'SL1344_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'SL1344_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'SL1344_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: SL1344
- Species: S. typhimurium

Mapping object

The mapping object, 'SL1344_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD005579. The related genomes used to determine evolutionary conservation all came from the genus Salmonella, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'SL1344_PreSaved_ResultsObj_GenBank', 'SL1344_PreSaved_ResultsObj_GeneMarkS2', 'SL1344_PreSaved_ResultsObj_Glimmer', and 'SL1344_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("SL1344", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomcis data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD005579
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

---

<table>
<thead>
<tr>
<th>Strain10403S</th>
<th>Assessment Objects for Listeria monocytogenes strain 10403S</th>
</tr>
</thead>
</table>

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Listeria monocytogenes strain 10403S

Usage

data(Strain10403S_PreSaved_DataMapObj)
data(Strain10403S_PreSaved_ResultsObj_GenBank)
data(Strain10403S_PreSaved_ResultsObj_GeneMarkS2)
data(Strain10403S_PreSaved_ResultsObj_Glimmer)
data(Strain10403S_PreSaved_ResultsObj_Prodigal)

Format

All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details

'Strain10403S_PreSaved_DataMapObj' is an object of subclass DataMap.
'Strain10403S_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'Strain10403S_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'Strain10403S_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'Strain10403S_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.
For all 5 objects, values of two key identifying list items within each object are listed below.

• StrainID: Strain10403S
• Species: L. monocytogenes
Mapping object

The mapping object, 'Strain10403S_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD010000. The related genomes used to determine evolutionary conservation all came from the genus Listeria, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects

The 4 results objects, 'Strain10403S_PreSaved_ResultsObj_GenBank', 'Strain10403S_PreSaved_ResultsObj_GeneMarkS2', 'Strain10403S_PreSaved_ResultsObj_Glimmer', and 'Strain10403S_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkerS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain’s genome

Use SaveGenomeToPath("Strain10403S", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.

Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD010000
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/

Description

Objects of class Assessment and either subclass DataMap or subclass Results for Bacillus subtilis subsp. subtilis strain 168
Usage

data(Strain168_PreSaved_DataMapObj)
data(Strain168_PreSaved_ResultsObj_GenBank)
data(Strain168_PreSaved_ResultsObj_GeneMarkS2)
data(Strain168_PreSaved_ResultsObj_Glimmer)
data(Strain168_PreSaved_ResultsObj_Prodigal)

Format
All 5 objects have a list structure. For specifics on what the two types of Assessment objects contain, please see the AssessORF package.

Details
'Strain168_PreSaved_DataMapObj’ is an object of subclass DataMap.
'Strain168_PreSaved_ResultsObj_GenBank’ is an object of subclass Results with predicted genes from the GenBank database.
'Strain168_PreSaved_ResultsObj_GeneMarkS2’ is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'Strain168_PreSaved_ResultsObj_Glimmer’ is an object of subclass Results with predicted genes from the program Glimmer.
'Strain168_PreSaved_ResultsObj_Prodigal’ is an object of subclass Results with predicted genes from the program Prodigal.

For all 5 objects, values of two key identifying list items within each object are listed below.

- StrainID: Strain168
- Species: B. subtilis

Mapping object
The mapping object, ‘Strain168_PreSaved_DataMapObj’, stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain’s genome. The raw proteomics mass spectra data came from ProteomeXchange dataset PXD004565. The related genomes used to determine evolutionary conservation all came from the genus Bacillus, and links to their sequences were downloaded from NCBI’s Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects
The 4 results objects, ‘Strain168_PreSaved_ResultsObj_GenBank’, ‘Strain168_PreSaved_ResultsObj_GeneMarkS2’, ‘Strain168_PreSaved_ResultsObj_Glimmer’, and ‘Strain168_PreSaved_ResultsObj_Prodigal’, store how much evidence there is supporting or against each gene in a set of predicted genes for the strain’s genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.

For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer
(3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.

The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.

The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

**Getting the strain’s genome**

Use `SaveGenomeToPath("Strain168", <INSERT FILE PATH HERE>)` to save the genome sequence for the strain to your local machine.

**Source**

Proteomcis data: [http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX004565](http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX004565)


Details
'TCH1516_PreSaved_DataMapObj' is an object of subclass DataMap.
'TCH1516_PreSaved_ResultsObj_GenBank' is an object of subclass Results with predicted genes from the GenBank database.
'TCH1516_PreSaved_ResultsObj_GeneMarkS2' is an object of subclass Results with predicted genes from the program GeneMarkS-2.
'TCH1516_PreSaved_ResultsObj_Glimmer' is an object of subclass Results with predicted genes from the program Glimmer.
'TCH1516_PreSaved_ResultsObj_Prodigal' is an object of subclass Results with predicted genes from the program Prodigal.
For all 5 objects, values of two key identifying list items within each object are listed below.
- StrainID: TCH1516
- Species: S. aureus

Mapping object
The mapping object, 'TCH1516_PreSaved_DataMapObj', stores the mapping of proteomics evidence and evolutionary conservation evidence to the strain's genome. The proteomics data was generated in-house and collected as part of the work for the AssessORF paper. The raw mass spectra data can be found at ProteomeXchange dataset PXD012538. The related genomes used to determine evolutionary conservation all came from the genus Staphylococcus, and links to their sequences were downloaded from NCBI's Genome Browser. The object was built using the MapAssessmentData function from the AssessORF package.

Results objects
The 4 results objects, 'TCH1516_PreSaved_ResultsObj_GenBank', 'TCH1516_PreSaved_ResultsObj_GeneMarkS2', 'TCH1516_PreSaved_ResultsObj_Glimmer', and 'TCH1516_PreSaved_ResultsObj_Prodigal', store how much evidence there is supporting or against each gene in a set of predicted genes for the strain's genome. The four objects were built using the AssessGenes function from the AssessORF package, a set of genes, and the mapping object described here.
For the GenBank object, the genes were downloaded from the corresponding record in the GenBank database. For the other three objects, GenemarkS-2 (web server), Prodigal (2.6.3), and Glimmer (3.02) were run at default settings in order to generate the set of predicted genes for the corresponding results object.
The GeneLeftPos, GeneRightPos, and the GeneStrand within each of the four results objects provide positional information on the set of predicted genes used to generate that object.
The GeneSource list item within each of the four results objects describes where the predicted genes came from (either "GenBank", "GeneMarkS2", "Glimmer", or "Prodigal" respectively).

Getting the strain's genome
Use SaveGenomeToPath("TCH1516", <INSERT FILE PATH HERE>) to save the genome sequence for the strain to your local machine.
Source

Proteomics data: http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD012538
NCBI’s Genome Browser: https://www.ncbi.nlm.nih.gov/genome/browse/#!/prokaryotes/
Index

* datasets
  AssessORF_StrainIDs, 4
  AP1, 2
  AP1_PreSaved_DataMapObj (AP1), 2
  AP1_PreSaved_ResultsObj_GenBank (AP1), 2
  AP1_PreSaved_ResultsObj_GeneMarkS2 (AP1), 2
  AP1_PreSaved_ResultsObj_Glimmer (AP1), 2
  AP1_PreSaved_ResultsObj_Prodigal (AP1), 2
  AssessORF_StrainIDs, 4
  ATCC11842, 4
  ATCC11842_PreSaved_DataMapObj (ATCC11842), 4
  ATCC11842_PreSaved_ResultsObj_GenBank (ATCC11842), 4
  ATCC11842_PreSaved_ResultsObj_GeneMarkS2 (ATCC11842), 4
  ATCC11842_PreSaved_ResultsObj_Glimmer (ATCC11842), 4
  ATCC11842_PreSaved_ResultsObj_Prodigal (ATCC11842), 4
  ATCC13032, 6
  ATCC13032_PreSaved_DataMapObj (ATCC13032), 6
  ATCC13032_PreSaved_ResultsObj_GenBank (ATCC13032), 6
  ATCC13032_PreSaved_ResultsObj_GeneMarkS2 (ATCC13032), 6
  ATCC13032_PreSaved_ResultsObj_Glimmer (ATCC13032), 6
  ATCC13032_PreSaved_ResultsObj_Prodigal (ATCC13032), 6
  ATCC17978, 8
  ATCC17978_PreSaved_DataMapObj (ATCC17978), 8
  ATCC17978_PreSaved_ResultsObj_GenBank (ATCC17978), 8
  ATCC17978_PreSaved_ResultsObj_GeneMarkS2 (ATCC17978), 8
  ATCC17978_PreSaved_ResultsObj_Glimmer (ATCC17978), 8
  ATCC700084, 9
  ATCC700084_PreSaved_DataMapObj (ATCC700084), 9
  ATCC700084_PreSaved_ResultsObj_GenBank (ATCC700084), 9
  ATCC700084_PreSaved_ResultsObj_GeneMarkS2 (ATCC700084), 9
  ATCC700084_PreSaved_ResultsObj_Glimmer (ATCC700084), 9
  ATCC700084_PreSaved_ResultsObj_Prodigal (ATCC700084), 9
  CCMP1375, 13
  CCMP1375_PreSaved_DataMapObj (CCMP1375), 13
  CCMP1375_PreSaved_ResultsObj_GenBank (CCMP1375), 13
  CCMP1375_PreSaved_ResultsObj_GeneMarkS2 (CCMP1375), 13
  CCMP1375_PreSaved_ResultsObj_Glimmer (CCMP1375), 13
INDEX

Strain10403S_PreSaved_ResultsObj_Glimmer (Strain10403S), 44
Strain10403S_PreSaved_ResultsObj_Prodigal (Strain10403S), 44
Strain168, 45
Strain168_PreSaved_DataMapObj (Strain168), 45
Strain168_PreSaved_ResultsObj_GenBank (Strain168), 45
Strain168_PreSaved_ResultsObj_GeneMarkS2 (Strain168), 45
Strain168_PreSaved_ResultsObj_Glimmer (Strain168), 45
Strain168_PreSaved_ResultsObj_Prodigal (Strain168), 45

TCH1516, 47
TCH1516_PreSaved_DataMapObj (TCH1516), 47
TCH1516_PreSaved_ResultsObj_GenBank (TCH1516), 47
TCH1516_PreSaved_ResultsObj_GeneMarkS2 (TCH1516), 47
TCH1516_PreSaved_ResultsObj_Glimmer (TCH1516), 47
TCH1516_PreSaved_ResultsObj_Prodigal (TCH1516), 47