Package 'specmine.datasets'

July 14, 2026

Type Package

Title Data Sets for 'specmine'

Version 0.0.3

Date 2026-07-03

Depends R (>= 4.0.0)

LazyData true

RoxygenNote 8.0.0

Description Provides the data sets used to exemplify 'specmine'. These data sets were formerly distributed with 'specmine', but they exceed current CRAN policy for package size.

Encoding UTF-8

License GPL (>= 2)

URL https://github.com/PedroFontao/specmine.datasets

BugReports https://github.com/PedroFontao/specmine.datasets/issues

NeedsCompilation no

Author Christopher Costa [aut],
Marcelo Maraschin [aut],
Miguel Rocha [aut],
Pedro Fontão [aut, cre],
Sara Cardoso [aut],
Telma Afonso [aut],
C. Beleites [cph],
Jie Hao [cph],
Bruno Pereira [aut]

Maintainer Pedro Fontão <pedrofontao812004@gmail.com>

Repository CRAN

Date/Publication 2026-07-14 16:40:02 UTC

Contents

`cachexia` *Human cachexia data*

Description

Cachexia is a complex metabolic syndrome associated with an underlying illness (such as cancer) and characterized by loss of muscle with or without loss of fat mass (Evans et al., 2008). A total of 77 urine samples were collected, with 47 from patients with cachexia and 30 from control patients.

Usage

```
cachexia
```

Format

An object of class `"list"`

Source

Data originally made available through MetaboAnalyst.

References

Eisner et al. (2010) Learning to predict cancer-associated skeletal muscle wasting from 1h-nmr profiles of urinary metabolites. Metabolomics 7:25-34.

cassavaPPD — *Cassava postharvest physiological deterioration*

Description

Cassava is a root widely cultivated in tropical and subtropical regions for its starchy tuberous root, which is an important source of carbohydrates. It has a wide variety of applications, including animal feeding, culinary uses and alcoholic beverages. In some countries, cassava has also been tested as an ethanol biofuel feedstock.

Usage

```
cassavaPPD
```

Format

An object of class `"list"`

References

Uarrota et al. (2014) Metabolomics combined with chemometric tools (pca, hca, pls-da and svm) for screening cassava (manihot esculenta crantz) roots during postharvest physiological deterioration. Food Chemistry 161:67-78.

propolis — *Brazilian propolis from different harvest seasons and agroecological regions (dataset)*

Description

Propolis, or bee glue, is a sticky dark-colored substance produced from collected buds or plant exudates (resin) by bees (Apis mellifera L.). The resin is masticated, salivary enzymes are added, and the partially digested material is mixed with beeswax and used in the hive to seal walls, strengthen comb borders and embalm dead invaders (Wollenweber et al., 1990). The propolis samples are from NMR data and were collected in autumn (AU), winter (WI), spring (SP) and summer (SM) of 2010 from Apis mellifera hives located in southern Brazil (Santa Catarina State). A total of 59 samples were collected, distributed by season as follows: SM - 16 samples, AU and SP - 15 samples each, and WI - 13 samples. Three agroecological regions were defined for the different apiaries and distributed as follows: Highlands - 12 samples, Plain - 11 samples and Plateau - 36 samples.

Usage

```
propolis
```

Format

An object of class "list"

References

E. Wollenweber, B. M. Hausen, and W. Greenaway. Phenolic constituents and sensitizing properties of propolis, poplar balsam and balsam of Peru. Bulletin de Groupe Polyphenol, 15:112-120, 1990.

M. Maraschin, A. Somensi-Zeggio, S. K. Oliveira, S. Kuhnen, M. M. Tomazzoli, A. C. M. Zeri, R. Carreira, and M. Rocha. A machine learning and chemometrics assisted interpretation of spectroscopic data - a NMR-based metabolomics platform for the assessment of Brazilian propolis. 2012.

---

propolisSampleList *Brazilian propolis from different harvest seasons and agroecological regions (sample list)*

Description

Propolis, or bee glue, is a sticky dark-colored substance produced from collected buds or plant exudates (resin) by bees (Apis mellifera L.). The resin is masticated, salivary enzymes are added, and the partially digested material is mixed with beeswax and used in the hive to seal walls, strengthen comb borders and embalm dead invaders (Wollenweber et al., 1990). The propolis samples are from NMR data and were collected in autumn (AU), winter (WI), spring (SP) and summer (SM) of 2010 from Apis mellifera hives located in southern Brazil (Santa Catarina State). A total of 59 samples were collected, distributed by season as follows: SM - 16 samples, AU and SP - 15 samples each, and WI - 13 samples. Three agroecological regions were defined for the different apiaries and distributed as follows: Highlands - 12 samples, Plain - 11 samples and Plateau - 36 samples.

Usage

```
propolisSampleList
```

Format

An object of class "list"

References

E. Wollenweber, B. M. Hausen, and W. Greenaway. Phenolic constituents and sensitizing properties of propolis, poplar balsam and balsam of Peru. Bulletin de Groupe Polyphenol, 15:112-120, 1990.

M. Maraschin, A. Somensi-Zeggio, S. K. Oliveira, S. Kuhnen, M. M. Tomazzoli, A. C. M. Zeri, R. Carreira, and M. Rocha. A machine learning and chemometrics assisted interpretation of spectroscopic data - a NMR-based metabolomics platform for the assessment of Brazilian propolis. 2012.

spectra_options *Information on the library of NMR reference spectra in the specmine package*

Description

This dataset provides information on the library of NMR spectra used as references in NMR metabolite identification.

Usage

```
spectra_options
```

Format

A data frame with 1816 observations on the following 9 variables. Each observation corresponds to a spectrum in the library.

`SPCMNS` A character vector with the spectra IDs.

`SPCMNM` A character vector with the metabolite IDs of the corresponding spectra.

`FREQUENCY` A character vector with the frequencies at which the spectra were obtained.

`NUCLEUS` A character vector indicating the examined nucleus. All observations are '1H'.

`PH` A character vector with the pH of the samples from which the spectra were obtained. May contain missing values.

`TEMPERATURE` A character vector with the temperature at which the spectra were obtained. May contain missing values.

`SOLVENT` A character vector with the solvent of the samples from which the spectra were obtained.

`ORIGINAL_DATABASE_ID` Whenever available, a character vector with the ID of the corresponding spectrum from the database where it was originally acquired.

`DATABASE` A character vector specifying the database from which the spectra were taken.

References

The spectra were taken from the HMDB, BMRB and SDBS databases. Some spectra were internally acquired and are mentioned as OUR in the `DATABASE` variable.

`spinalCord` *Mouse spinal cord LC-MS data*

Description

This dataset consists of 12 LC-MS spectra samples in netCDF format from mouse spinal cord, divided into two groups: wild type and knockout. The data was obtained from MetaboAnalyst and originates from a study describing a general strategy for identifying endogenous substrates of enzymes by untargeted LC-MS analysis of tissue metabolomes from wild-type and enzyme-inactivated organisms.

Usage

```
spinalCord
```

Format

An object of class `"list"`

References

A. Saghatelian, S.A. Trauger, E.J. Want, E.G. Hawkins, G. Siuzdak, and B.F. Cravatt. Assignment of endogenous substrates to enzymes by global metabolite profiling. Biochemistry, 43:14332-14339, 2004.

Index