Package 'WallomicsData'

October 12, 2022

Title Datasets for Multi-Omics Integration in a Plant Abiotic Stress Context

Version 1.0

Description
Datasets from the WallOmics project. Contains phenomics, metabolomics, proteomics and transcriptomics data collected from two organs of five ecotypes of the model plant Arabidopsis thaliana exposed to two temperature growth conditions. Exploratory and integrative analyses of these data are presented in Durufle et al (2020) <doi:10.1093/bib/bbaa166> and Durufle et al (2020) <doi:10.3390/cells9102249>.

License GPL-3

Encoding UTF-8

RoxygenNote 7.1.2

Depends R (>= 2.10)

LazyData true

NeedsCompilation no

Author Sébastien Déjean [aut, cre],
Harold Duruflé [aut],
Aurélie Mercadié [aut]

Maintainer Sébastien Déjean <sebastien.dejean@math.univ-toulouse.fr>

Repository CRAN

Date/Publication 2022-04-15 10:32:29 UTC

R topics documented:

`Altitude_Cluster` *Altitude Cluster*

Description

The **Altitude Cluster factor** identifies the environment height from which is originated a given plant from the sample under study, either high altitude (denoted *High*), moderate altitude (*Low*) or the reference group's environment height (*Col*, the lowest of all).

Usage

```
data("Altitude_Cluster")
```

Format

A factor with 3 levels.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the data
data("Altitude_Cluster")

# Count how many samples are in each group
table(Altitude_Cluster)
```

`Ecotype` *Ecotype*

Description

The **Ecotype factor** identifies the genotype specifically designed for a given ecosystem of the *A. thaliana* from which the studied sample comes from. We have a population of reference as well as 4 newly-described Pyrenean populations, namely:

- Columbia, denoted *Col* (originating from Poland, acts as the reference ecotype)
- Grip, denoted *Grip*
- Herran, denoted *Hern*
- L'Hospitalet-près-l'Andorre, denoted *Hosp*
- Chapelle Saint Roch, denoted *Roch*

Usage

```
data("Ecotype")
```

Format

A factor with 5 levels of *A. thaliana* genotypes.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the data
data("Ecotype")

# Count how many samples are in each group
table(Ecotype)
```

`Genetic_Cluster` *Genetic Cluster*

Description

The **Genetic Cluster factor** identifies the genetic group from which comes from the studied sample, either **Genetics Cluster 1** (constitued of *Grip* and *Roch* genotypes), **Genetics Cluster 2** (*Hern* and *Hosp* genotypes) or **Genetics Cluster 3** (*Col* genotype). See Ecotype for more information on genotypes.

Usage

```
data("Genetic_Cluster")
```

Format

A factor with 3 levels.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the data
data("Genetic_Cluster")

# Count how many samples are in each group
table(Genetic_Cluster)
```

Metabolomics_Rosettes *Metabolomics Rosettes*

Description

A dataset containing metabolomics variables measured on rosettes of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Metabolomics_Rosettes")
```

Format

A data frame with 30 rows and 6 variables:

- **Pectin_RGI**: Rhamnogalacturonan I (µg/100mg)
- **Pectin_HG**: Homogalacturonan (µg/100mg)
- **XG**: Xyloglucan (µg/100mg)
- **Pectin_linearity**: Linearity of pectin (Ratio)
- **Contribution_RG**: Contribution of rhamnogalacturonan to pectin population (Ratio)
- **RGI_branching**: Branching of Rhamnogalacturonan I (Ratio)

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Metabolomics_Rosettes")

# Look at simple statistics
summary(Metabolomics_Rosettes)

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# A graphical representation
plot(x = as.factor(substr(row.names(Metabolomics_Rosettes), 1, 7)),
     y = Metabolomics_Rosettes$Pectin_linearity, col = "white", lty = 0,
     xlab = "Genotype x Temperature groups",
     ylab = "Pectin linearity (Ratio)",
     main = "Pectin linearity distribution by genotype and growth temperature")
grid()
abline(h = 1, lty = 2)
points(x = as.factor(substr(row.names(Metabolomics_Rosettes), 1, 7)),
       y = Metabolomics_Rosettes$Pectin_linearity, type = "p", pch = 19, lwd = 5,
       col = colors)
```

Metabolomics_Stems *Metabolomics Stems*

Description

A dataset containing metabolomics variables measured on floral stems of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Metabolomics_Stems")
```

Format

A data frame with 30 rows and 6 variables:

- **Pectin_RGI**: Rhamnogalacturonan I (µg/100mg)
- **Pectin_HG**: Homogalacturonan (µg/100mg)
- **XG**: Xyloglucan (µg/100mg)
- **Pectin_linearity**: Linearity of pectin (Ratio)
- **Contribution_RG**: Contribution of rhamnogalacturonan to pectin population (Ratio)
- **RGI_branching**: Branching of Rhamnogalacturonan I (Ratio)

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Metabolomics_Stems")

# Look at simple statistics
summary(Metabolomics_Stems)

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# A graphical representation
plot(x = as.factor(substr(row.names(Metabolomics_Stems), 1, 7)),
     y = Metabolomics_Stems$Pectin_linearity, col = "white", lty = 0,
     xlab = "Genotype x Temperature groups",
     ylab = "Pectin linearity (Ratio)",
     main = "Pectin linearity distribution by genotype and growth temperature")
grid()
abline(h = 1, lty = 2)
points(x = as.factor(substr(row.names(Metabolomics_Stems), 1, 7)),
       y = Metabolomics_Stems$Pectin_linearity, type = "p", pch = 19, lwd = 5,
       col = colors)
```

`Metadata` *Metadata*

Description

Bioinformatics Annotation and description, using the WallProtDB database, of all the Cell Wall Proteins (CWPs) identified on rosettes and floral stems of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for additionnal information.

Usage

```
data("Metadata")
```

Format

A data frame with 474 rows and 4 variables:

- **Acc_number**: GenBank accession number (gene name)
- **Functional_classes**: Functional classes of the CWPs
- **Protein_families**: Protein families of the CWPs
- **Putative_functions**: Putative functions of the CWPs

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Metadata")

# Look at the dataset's dimensions
dim(Metadata)
head(Metadata)

# How many functional classes ?
table(Metadata$Functional_classes)

# How many protein families ?
table(Metadata$Protein_families)
```

`Phenomics_Rosettes` *Phenomics Rosettes*

Description

A dataset containing phenotypic variables measured on rosettes of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Phenomics_Rosettes")
```

Format

A data frame with 30 rows and 5 variables:

- **Mass**: rosette mass (g)
- **Diameter**: rosette diameter (cm)
- **Leaves_number**: total number of leaves
- **Density**: rosette density (g/cm²)
- **Area**: projected rosette area (cm²)

Source

doi: 10.3390/cells9102249

Examples

```
# Load the data
data("Phenomics_Rosettes")

# Look at simple statistics
dim(Phenomics_Rosettes)
summary(Phenomics_Rosettes)

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# A graphical representation: Leaves number distribution
plot(x = as.factor(substr(row.names(Phenomics_Rosettes), 1, 7)),
     y = Phenomics_Rosettes$Leaves_number, col = "white", lty = 0,
     xlab = "Genotype x Temperature groups",
     ylab = "Number of rosette leaves",
     main = "Rosette leaves' distribution by genotype and growth temperature"
     )
grid()
points(x = as.factor(substr(row.names(Phenomics_Rosettes), 1, 7)),
       y = Phenomics_Rosettes$Leaves_number, type = "p", pch = 19, lwd = 5,
       col = colors)
```

Phenomics_Stems *Phenomics Stems*

Description

A dataset containing phenotypic variables measured on floral stems of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Phenomics_Stems")
```

Format

A data frame with 30 rows and 4 variables:

- **Mass**: floral stems mass (g)
- **Diameter**: floral stems diameter (mm)
- **Length**: length of the floral stems (cm)
- **Number_lateral_stems**: number of lateral stems)

Source

doi: 10.3390/cells9102249

Examples

```
# Load the data
data("Phenomics_Stems")

# Look at simple statistics
dim(Phenomics_Stems)
summary(Phenomics_Stems)

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# A graphical representation: Lateral stems distribution
plot(x = as.factor(substr(row.names(Phenomics_Stems), 1, 7)),
     y = Phenomics_Stems$Number_lateral_stems, col = "white", lty = 0,
     xlab = "Genotype x Temperature groups",
     ylab = "Number of lateral stems",
     main = "Lateral stems' distribution by genotype and growth temperature"
     )
grid()
points(x = as.factor(substr(row.names(Phenomics_Stems), 1, 7)),
       y = Phenomics_Stems$Number_lateral_stems, type = "p", pch = 19, lwd = 5,
       col = colors)
```

`Proteomics_Rosettes_CW` *Proteomics Rosettes Cell Wall*

Description

A dataset containing the identification and quantification of Cell Wall Proteins (CWPs) performed using LC-MS/MS analysis on rosettes of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for additional information.

Usage

```
data("Proteomics_Rosettes_CW")
```

Format

A data frame with 30 rows and 364 variables.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Proteomics_Rosettes_CW")

# Look at data frame dimensions
dim(Proteomics_Rosettes_CW)

# Look at the first rows and columns
head(Proteomics_Rosettes_CW[,c(1:10)])

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# PCA on proteomics
res.pca <- prcomp(Proteomics_Rosettes_CW, center = TRUE, scale. = TRUE)
plot(res.pca$x[,"PC1"], res.pca$x[,"PC2"], pch = 19, xlab = "PC1", ylab = "PC2", lwd = 5,
     main = "Individuals' plot (1 x 2) - PCA on Rosettes Cell Wall Proteomics",
     col = colors)
text(res.pca$x[,"PC1"], res.pca$x[,"PC2"], labels = row.names(res.pca$x), cex = 0.8, pos = 3)
```

Proteomics_Stems_CW *Proteomics Stems Cell Wall*

Description

A dataset containing the identification and quantification of Cell Wall Proteins (CWPs) performed using LC-MS/MS analysis on floral stems of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for additionnal information.

Usage

```
data("Proteomics_Stems_CW")
```

Format

A data frame with 30 rows and 414 variables.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Proteomics_Stems_CW")

# Look at data frame dimensions
dim(Proteomics_Stems_CW)

# Look at the first rows and columns
head(Proteomics_Stems_CW[,c(1:10)])

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# PCA on proteomics
res.pca <- prcomp(Proteomics_Stems_CW, center = TRUE, scale. = TRUE)
plot(res.pca$x[,"PC1"], res.pca$x[,"PC2"], pch = 19, xlab = "PC1", ylab = "PC2", lwd = 5,
     main = "Individuals' plot (1 x 2) - PCA on Stems Cell Wall Proteomics",
     col = colors)
text(res.pca$x[,"PC1"], res.pca$x[,"PC2"], labels = row.names(res.pca$x), cex = 0.8, pos = 3)
```

`Temperature` *Temperature*

Description

The **Temperature factor** identifies the temperature at which the studied sample was exposed all along its growth, either 22°C (optimal condition) or 15°C (high altitude condition).

Usage

```
data("Temperature")
```

Format

A factor with 2 levels.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the data
data("Temperature")

# Count how many samples are in each group
```

```
table(Temperature)
```

`Transcriptomics_Rosettes`
Transcriptomics Rosettes

Description

A dataset containing all the transcripts obtained by RNA-seq performed, according to the standard Illumina protocols, on rosettes of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Transcriptomics_Rosettes")
```

Format

A data frame with 30 rows and 19763 variables.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Transcriptomics_Rosettes")

# Look at data frame dimensions
dim(Transcriptomics_Rosettes)

# Look at the first rows and columns
head(Transcriptomics_Rosettes[,c(1:10)])

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# PCA on transcriptomics
res.pca <- prcomp(Transcriptomics_Rosettes, center = TRUE, scale. = TRUE)
plot(res.pca$x[,"PC1"], res.pca$x[,"PC2"], pch = 19, xlab = "PC1", ylab = "PC2", lwd = 5,
     main = "Individuals' plot (1 x 2) - PCA on Rosettes' Transcriptomics",
     col = colors)
text(res.pca$x[,"PC1"], res.pca$x[,"PC2"], labels = row.names(res.pca$x), cex = 0.8, pos = 3)
```

`Transcriptomics_Rosettes_CW` *Transcriptomics Rosettes Cell Wall*

Description

A dataset containing the transcripts encoding Cell Wall Proteins (CWPs) sorted from the 19 763 transcripts (see Transcriptomics_Rosettes) obtained by RNA-seq performed, according to the standard Illumina protocols, on rosettes of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Transcriptomics_Rosettes_CW")
```

Format

A data frame with 30 rows and 364 variables.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Transcriptomics_Rosettes_CW")

# Look at data frame dimensions
dim(Transcriptomics_Rosettes_CW)

# Look at the first rows and columns
head(Transcriptomics_Rosettes_CW[,c(1:10)])

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# PCA on transcriptomics
res.pca <- prcomp(Transcriptomics_Rosettes_CW, center = TRUE, scale. = TRUE)
plot(res.pca$x[,"PC1"], res.pca$x[,"PC2"], pch = 19, xlab = "PC1", ylab = "PC2", lwd = 5,
     main = "Individuals' plot (1 x 2) - PCA on Rosettes Cell Wall Transcriptomics",
     col = colors)
text(res.pca$x[,"PC1"], res.pca$x[,"PC2"], labels = row.names(res.pca$x), cex = 0.8, pos = 3)
```

`Transcriptomics_Stems` *Transcriptomics Stems*

Description

A dataset containing all the transcripts obtained by RNA-seq performed, according to the standard Illumina protocols, on floral stems of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Transcriptomics_Stems")
```

Format

A data frame with 30 rows and 22570 variables.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Transcriptomics_Stems")

# Look at data frame dimensions
dim(Transcriptomics_Stems)

# Look at the first rows and columns
head(Transcriptomics_Stems[,c(1:10)])

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# PCA on transcriptomics
res.pca <- prcomp(Transcriptomics_Stems, center = TRUE, scale. = TRUE)
plot(res.pca$x[,"PC1"], res.pca$x[,"PC2"], pch = 19, xlab = "PC1", ylab = "PC2", lwd = 5,
     main = "Individuals' plot (1 x 2) - PCA on Stems' Transcriptomics",
     col = colors)
text(res.pca$x[,"PC1"], res.pca$x[,"PC2"], labels = row.names(res.pca$x), cex = 0.8, pos = 3)
```

Transcriptomics_Stems_CW

Transcriptomics Stems Cell Wall

Description

A dataset containing the transcripts encoding Cell Wall Proteins (CWPs) sorted from the 22 570 transcripts (see Transcriptomics_Stems) obtained by RNA-seq performed, according to the standard Illumina protocols, on floral stems of five *A. thaliana* genotypes at two growth temperatures. See Ecotype and Temperature for more information.

Usage

```
data("Transcriptomics_Stems_CW")
```

Format

A data frame with 30 rows and 414 variables.

Source

doi: 10.3390/cells9102249

Examples

```
# Load the dataset
data("Transcriptomics_Stems_CW")

# Look at data frame dimensions
dim(Transcriptomics_Stems_CW)

# Look at the first rows and columns
head(Transcriptomics_Stems_CW[,c(1:10)])

# Create a colors' vector
colors <- c(rep("#A6CEE3",3), rep("#1F78B4",3), rep("#B2DF8A",3), rep("#33A02C",3),
            rep("#FB9A99",3), rep("#E31A1C",3), rep("#FDBF6F",3), rep("#FF7F00",3),
            rep("#CAB2D6",3), rep("#6A3D9A",3))

# PCA on transcriptomics
res.pca <- prcomp(Transcriptomics_Stems_CW, center = TRUE, scale. = TRUE)
plot(res.pca$x[,"PC1"], res.pca$x[,"PC2"], pch = 19, xlab = "PC1", ylab = "PC2", lwd = 5,
     main = "Individuals' plot (1 x 2) - PCA on Stems Cell Wall Transcriptomics",
     col = colors)
text(res.pca$x[,"PC1"], res.pca$x[,"PC2"], labels = row.names(res.pca$x), cex = 0.8, pos = 3)
```

Index