Package 'ChiliCross'

July 21, 2021

Type Package

Title Gene expression inheritance in two crosses of Capsicum

Version 1.0

Date 2021-07-19

Author Christian Escoto-Sandoval and Octavio Martinez

Maintainer <octavio.martinez@cinvestav.mx>

Description Contains standardized gene expression for two crosses of chili pepper (Capsicum annuum L.) during fruit development and results for Gene Ontology enrichment analyses for sets of genes. There is also a function to plot cross results.

License GPL-3

LazyData TRUE

R topics documented:

ChiliCross-package *Gene expression inheritance in two crosses of Capsicum*

Description

Contains standardized gene expression for two crosses of chili pepper (Capsicum annuum L.) during fruit development and results for Gene Ontology enrichment analyses for sets of genes. There is also a function to plot cross results.

Details

Gene expression data were obtained from RNA-Seq libraries constructed from accessions CM (criollo de Morelos 334, domesticated), QU (piquin Queretaro, wild) and the two F1s: CQ resulting from CM female x QU male, and QC resulting from QU female x CM male. Estimation of gene expression was performed for each one of the participants in the crosses at 0, 10, 20, 30, 40, 50 and 60 Days After Anthesis (DAA). You can explore the way in which gene expression of 22374 genes behaved in those two crosses during fruit development. Gene expression was standardized, thus the vector of gene expression of the three participants in a cross has a mean of 0 and standard deviation of 1. Gene expression was classified into 4 classes and 19 exhaustive sub-models. You can also instal our R package Salsa from https://zenodo.org/record/4767445#.YKJmFGauJn5 to complement the study of gene expression for a larger set of accessions.

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.

Author(s)

Christian Escoto-Sandoval and Octavio Martinez

Maintainer: <octavio.martinez@cinvestav.mx>

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
# Assume that you are interested in MYB transcription factors (TF)
# and want to see how some of those genes were expressed in the crosses
# Let's isolate the identifiers of some MYB:
temp.ids <- gene$id[grep('MYB', gene$Prot.Desc)]
length(temp.ids) # How many genes are there?
# Let's see the Protein Description (gene$Prot.Desc) of some of them
head(gene$Prot.Desc[is.element(gene$id, temp.ids)])
temp.ids[1] # The first of these genes
gene[gene$id==temp.ids[1],]
# Let's plot the expression pattern of this gene in cross C1,
cross.plot(x = C1[C1$id == temp.ids[1], ], cross = "C1")
# and in the second cross, C2,
cross.plot(x = C2[C1$id == temp.ids[1], ], cross = "C2")
# Now let's tabulate the models of all those 107 MYB TF at time
# 00 DAA (the mature flower at 0 Days After Anthesis)
table(C1$sub.model[(is.element(C1$id, temp.ids))&(C1$time=="00")])
# We see that there a 2 genes with sub-model P2<P1<F1 at time 00 in C1
# and we want to know which genes are those
temp.ids2 <- C1$id[(is.element(C1$id, temp.ids))&(C1$time=="00")&(C1$sub.model=="P2<P1<F1")]
temp.ids2
gene[is.element(gene$id, temp.ids2),]
```

```
# Now let's see the inheritance pattern of the first one in C1:
cross.plot(x = C1[C1$id == temp.ids2[1], ], cross = "C1")
# and in the second cross
cross.plot(x = C2[C2$id == temp.ids2[1], ], cross = "C2")
# Remove the temporal vectors
rm(temp.ids, temp.ids2)
```

---

all.models   *All models in both crosses (C1 and C2)*

Description

A simple data frame with auxiliar variables to describe characteristics of classes, models and sub-models.

Usage

```
data("all.models")
```

Format

A data frame with 19 observations on the following 5 variables.

sub.model a character vector with sub-model.

mi a numeric vector with numerical identifier of the model.

model a character vector with model (does not take into account significance).

class a character vector with the class corresponding to the sub-model.

is.main a logical vector denoting if the sub-model has only inequalities.

Details

This data frame is a guide to some of the variables in data.frames C1 and C2.

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(all.models)
# Show first rows of the data.frame
head(all.models)
# Shows only the sub-models that are equal to the models
all.models[all.models$is.main==TRUE,]
```

C1	*Data for cross 1 (C1).*

Description

All results for cross C1, including standardized expressions for each cross participant.

Usage

```
data("C1")
```

Format

A data frame with 156618 observations on the following 13 variables.

id a numeric vector identifying the gene (as gene$id).

time a character vector with the time of chili pepper development stage (00, 10, up to 60 DAA).

mi a numeric vector with the identifier of the model.

model a character vector with the model.

class a character vector with the class to which the model belongs.

C1.left a character vector with the two left hand side participants of the cross (only for not-null sub-models).

C1.right a character vector with the two right hand side participants of the cross (only for not-null sub-models).

sig.left a logical vector to signal if the difference of gene expression between the left hand side participants is significant.

sig.right a logical vector to signal if the difference of gene expression between the right hand side participants is significant

SE.P1.CM a numeric vector with the Standardized Expression (SE) for P1.CM

SE.F1.CQ a numeric vector with the Standardized Expression (SE) for F1.CQ

SE.P2.QU a numeric vector with the Standardized Expression (SE) for P2.QU

sub.model a character vector with the sub-model.

Details

This data frame gives all relevant information for the inheritance of each gene at each time of development in cross C1.

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(C1)
# How many of each one of the models?
table(C1$model)
# How many of each one of the combinations time x model?
table(C1$time, C1$model)
```

C2   *Data for cross 2 (C2).*

Description

All results for cross C1, including standardized expressions for each cross participant.

Usage

```
data("C2")
```

Format

A data frame with 156618 observations on the following 13 variables.

id a numeric vector identifying the gene (as gene$id).

time a character vector with the time of chili pepper development stage (00, 10, up to 60 DAA).

mi a numeric vector with the identifier of the model.

model a character vector with the model.

class a character vector with the class to which the model belongs.

C2.left a character vector with the two left hand side participants of the cross (only for not-null sub-models).

C2.right a character vector with the two right hand side participants of the cross (only for not-null sub-models).

sig.left a logical vector to signal if the difference of gene expression between the left hand side participants is significant.

sig.right a logical vector to signal if the difference of gene expression between the right hand side participants is significant.

SE.P1.CM a numeric vector with the Standardized Expression (SE) for P1.CM

SE.F1.QC a numeric vector with the Standardized Expression (SE) for F1.QC

SE.P2.QU a numeric vector with the Standardized Expression (SE) for P2.QU

sub.model a character vector with the sub-model.

Details

This data frame gives all relevant information for the inheritance of each gene at each time of development in cross C2.

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(C2)
# How many of each one of the models?
table(C2$model)
# How many of each one of the combinations time x model?
table(C2$time, C2$model)
```

---

cross.plot   *Plots standardized expression in one of the crosses*

Description

This function plots standardized gene expression of any subset of exactly 7 rows of the datasets C1 (cross 1) or C2 (cross 2). The plot will be annotated with all details needed to appreciate the behavior of the cases plot. See details for a full description of the annotations.

Usage

```
cross.plot(x = C1[C1$id == 32, ], cross = "C1")
```

Arguments

x   A subset of exactly seven rows of datasets C1 or C2, depending of parameter
                cross.

cross   Either, "C1" or "C2", depending of which cross is used.

Details

The plot will show standardized gene expression in the X-axis and the seven cases in the Y-axis. It will always have a legend with the symbols used for each one of the participants in the cross. The left margin presents the gene identifier, while the right margin presents the time of development. Each row is annotated below each one of the 7 lines with the cross that was used and the sub-model. The sub-model is based in statistical evaluation of the differences with an approximate False Discovery Rate of 5 percent. If there is a single gene identifier (id), then the plot will have a title and subtitle describing the gene that was plot, but that will not happen in the cases where there is more than one different gene in the 7 rows.

Value

The function is used by its side effects, but invisibly returns the argument x.

Author(s)

Octavio Martinez

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

See Also

C1, C2.

Examples

```
# The default (plot the gene expression for gene with id=32 in C1)
cross.plot()
# The same gene but in C2
cross.plot(x = C2[C2$id == 32, ], cross = "C2")
# The fist seven rows of C1
cross.plot(C1[1:7,], "C1")
# The fist seven rows of C2
cross.plot(C2[1:7,], "C2")
# Now, assume that you are interested in "DNA polymerase"
head(gene[grep("DNA polymerase", gene$Prot.Desc),], 1)
# Thus you see that id=813 is annotated one of the subunits.
# Let's see the plots in C1 and C2
cross.plot(C1[C1$id == 813, ], cross = "C1")
cross.plot(C2[C2$id == 813, ], cross = "C2")
```

---

gene *Identifiers for genes*

Description

For each one of the expressed genes this data frame contains protein identifiers and descriptions.

Usage

```
data("gene")
```

Format

A data frame with 29946 observations on the following 4 variables.

id a numeric vector; numeric gene identifier.

ProtId a character vector; protein identifier.

Prot.Desc a character vector; protein description

isTF a logical vector; Is the coded protein a Transcription Factor?

Details

Descriptors are based in the reference genome CM334 v1.6.

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(gene)
# A random row of the data.frame
gene[sample(c(1:nrow(gene)), size=1), ]

# Show data for some Transcription Factors.
head(gene[gene$isTF==TRUE,])
```

GO.enr.BP   *Gene Ontology Enrichment results for Biological Process*

Description

All 133 sets of genes with a specific combination of sub-model and time of expression were subjected to Gene Ontology (GO) enrichment analysis for Biological Process (BP). Original results were filtered to let only cases that fulfilled a False Discovery Rate of 10 percent and an estimated odd value larger than one.

Usage

```
data("GO.enr.BP")
```

Format

A data frame with 3096 observations on the following 17 variables.

fullName a character vector with the full name of the combination cross.sub-model.time

cross a character vector with the cross from which the genes were selected; either C1 or C2.

sub.model a character vector with the sub-model of genes in the target set (1 of 19 possible values).

time a character vector with the time of chili pepper development stage (00, 10, up to 60 DAA).

s.class a character vector with the class of the sub-model (one of 4 possible values).

group a character vector with the group at which the GO belongs (exclusive to only one sub-model or Shared by to or more sub-models).

aspect.id a numeric vector with the identifier of the BP.

GO a character vector with the GO identifier.

desc a character vector with the description of the GO BP.

odds a numeric vector with the estimated of odd values.

enrich a logical vector with TRUE only when the estimated odd is equal or larger than 2.

P a numeric vector with p-values.

Q a numeric vector with q-values (for FDR calculation).

AnnTarg a numeric vector with the number of genes Annotated with aspect and in Target set.

NotAnnTarg a numeric vector with the number of genes Not Annotated with aspect and in Target set.

AnnNotTarg a numeric vector with the number of genes Annotated with aspect and Not in Target set.

NotAnnNotTarg a numeric vector with the number of genes Not Annotated with aspect and Not in Target set

Details

Data were obtained in our Salsa package (see references).

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(GO.enr.BP)
# Table the number of enriched BPs by cross and group
table(GO.enr.BP$cross, GO.enr.BP$group)
# Of those how many with odds equal or larger than 2
table(GO.enr.BP$cross[GO.enr.BP$odds>=2], GO.enr.BP$group[GO.enr.BP$odds>=2])
# Looking for BP which contains 'cell cycle'
head(GO.enr.BP[grep('cell cycle', GO.enr.BP$desc),], 1)
# And specifically with enrichment in 'regulation of cell cycle'
GO.enr.BP[GO.enr.BP$aspect.id==2125,]
```

GO.enr.CC   *Gene Ontology Enrichment results for Cell Component*

Description

All 133 sets of genes with a specific combination of sub-model and time of expression were subjected to Gene Ontology (GO) enrichment analysis for Cell Component (CC). Original results were filtered to let only cases that fulfilled a False Discovery Rate of 10 percent and an estimated odd value larger than one.

Usage

```
data("GO.enr.CC")
```

Format

A data frame with 374 observations on the following 17 variables.

fullName a character vector with the full name of the combination cross.sub-model.time

cross a character vector with the cross from which the genes were selected; either C1 or C2.

sub.model a character vector with the sub-model of genes in the target set (1 of 19 possible values).

time a character vector with the time of chili pepper development stage (00, 10, up to 60 DAA).

s.class a character vector with the class of the sub-model (one of 4 possible values).

group a character vector with the group at which the GO belongs (exclusive to only one sub-model or Shared by to or more sub-models).

aspect.id a numeric vector with the identifier of the CC.

GO a character vector with the GO identifier.

desc a character vector with the description of the GO CC.

odds a numeric vector with the estimated of odd values.

enrich a logical vector with TRUE only when the estimated odd is equal or larger than 2.

P a numeric vector with p-values.

Q a numeric vector with q-values (for FDR calculation).

AnnTarg a numeric vector with the number of genes Annotated with aspect and in Target set.

NotAnnTarg a numeric vector with the number of genes Not Annotated with aspect and in Target set.

AnnNotTarg a numeric vector with the number of genes Annotated with aspect and Not in Target set.

NotAnnNotTarg a numeric vector with the number of genes Not Annotated with aspect and Not in Target set

Details

Data were obtained in our Salsa package (see references).

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(GO.enr.CC)
# Table of cross x class in this data.frame
table(GO.enr.CC$cross, GO.enr.CC$s.class)
# See the 4 rows with cross="C1" and class="High F1"
GO.enr.CC[(GO.enr.CC$cross=="C1")&(GO.enr.CC$s.class=="High F1"),]
# How many cases of apoplast enrichment?
nrow(GO.enr.CC[GO.enr.CC$desc=="apoplast",])
# Note that apoplast is aspect.id=488
# See some tables for cases of enrichment in apoplast
table(GO.enr.CC$sub.model[GO.enr.CC$aspect.id==488])
table(GO.enr.CC$cross[GO.enr.CC$aspect.id==488])
table(GO.enr.CC$time[GO.enr.CC$aspect.id==488])
# Looking specific cases
GO.enr.CC[(GO.enr.CC$aspect.id==488)&(GO.enr.CC$time=="10"),]
```

GO.enr.MF   *Gene Ontology Enrichment results for Molecular Function*

Description

All 133 sets of genes with a specific combination of sub-model and time of expression were subjected to Gene Ontology (GO) enrichment analysis for Biological Process (BP). Original results were filtered to let only cases that fulfilled a False Discovery Rate of 10 percent and an estimated odd value larger than one.

Usage

```
data("GO.enr.MF")
```

Format

A data frame with 1772 observations on the following 17 variables.

fullName a character vector with the full name of the combination cross.sub-model.time

cross a character vector with the cross from which the genes were selected; either C1 or C2.

sub.model a character vector with the sub-model of genes in the target set (1 of 19 possible values).

time a character vector with the time of chili pepper development stage (00, 10, up to 60 DAA).

s.class a character vector with the class of the sub-model (one of 4 possible values).

group a character vector with the group at which the GO belongs (exclusive to only one sub-model or Shared by to or more sub-models).

aspect.id a numeric vector with the identifier of the MF.

GO a character vector with the GO identifier.

desc a character vector with the description of the GO MF.

odds a numeric vector with the estimated of odd values.

enrich a logical vector with TRUE only when the estimated odd is equal or larger than 2.

P a numeric vector with p-values.

Q a numeric vector with q-values (for FDR calculation).

AnnTarg a numeric vector with the number of genes Annotated with aspect and in Target set.

NotAnnTarg a numeric vector with the number of genes Not Annotated with aspect and in Target set.

AnnNotTarg a numeric vector with the number of genes Annotated with aspect and Not in Target set.

NotAnnNotTarg a numeric vector with the number of genes Not Annotated with aspect and Not in Target set

Details

Data were obtained in our Salsa package (see references).

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(GO.enr.MF)
# Which is the MF more frequently enriched?
temp.t <- table(GO.enr.MF$aspect.id)
temp.t <- temp.t[order(temp.t, decreasing=TRUE)]
head(temp.t)
head(GO.enr.MF[GO.enr.MF$aspect.id==778,])
# Obtain a table of those cases
temp.t <- GO.enr.MF[GO.enr.MF$aspect.id==778,] # 22 rows
table(temp.t$cross, temp.t$s.class)
rm(temp.t)
```

---

results.key   *Key for results*

Description

A remainder of participants of the crosses useful for users.

Usage

```
data("results.key")
```

Format

A data frame with 2 observations on the following 4 variables.

cross a character vector with definitions of the crosses.

P1 a character vector with definition of who is P1

P2 a character vector with definition of who is P2

F1 a character vector with definition of who is F1

Source

GEO Series accession number GSE165448. Link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE165448

References

Christian Escoto-Sandoval, Neftali Ochoa-Alejo and Octavio Martinez (In preparation, 2021). Inheritance of gene expression during fruit development in crosses between a domesticated and a wild accession of chili pepper; Escoto-Sandoval, C., Flores-Diaz, A., Reyes-Valdes, M.H., Neftali Ochoa-Alejo and Octavio Martinez. A method to analyze time expression profiles demonstrated in a database of chili pepper fruit development. Sci Rep 11, 13181 (2021). https://doi.org/10.1038/s41598-021-92672-4

Examples

```
data(results.key)
# To remind you how the crosses were performed
results.key
```

Index