

# Package ‘QuESTr’

October 12, 2022

**Type** Package

**Title** Constructing Quantitative Environment Sensor using  
Transcriptomes

**Version** 0.1.1

**Author** Takahiko Koizumi, Kenta Suzuki, Yasunori Ichihashi

**Maintainer** Takahiko Koizumi <takahiko.koizumi@riken.jp>

**Description** A method for prediction of environmental conditions based on transcriptome data linked with the environmental gradients. This package provides functions to overview gene-environment relationships, to construct the prediction model, and to predict environmental conditions where the transcriptomes were generated. This package can quest for candidate genes for the model construction even in non-model organisms' transcriptomes without any genetic information.

**Depends** R (>= 3.5.0)

**License** MIT + file LICENSE

**Language** en-US

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.1.2

**Suggests** knitr, rmarkdown, testthat (>= 3.0.0)

**URL** <https://github.com/takakoizumi/QuESTr>

**VignetteBuilder** knitr

**Config/testthat/edition** 3

**Imports** ggplot2, kernlab

**NeedsCompilation** no

**Repository** CRAN

**Date/Publication** 2022-02-15 20:00:04 UTC

**R topics documented:**

|                   |          |
|-------------------|----------|
| Pinus . . . . .   | 2        |
| q.clean . . . . . | 3        |
| q.opt . . . . .   | 3        |
| q.pca . . . . .   | 4        |
| q.rank . . . . .  | 5        |
| q.sort . . . . .  | 6        |
| quest . . . . .   | 7        |
| <b>Index</b>      | <b>8</b> |

Pinus

*Transcriptomes of Pinus roots under a Temperature Gradient***Description**

This dataset gives the TPM values of 200 selected genes obtained from 60 Pinus root samples (30 samples each for training and test data) under a temperature gradient, generated by RNA-seq.

**Usage**

Pinus

**Details**

A gene expression data matrix of 30 root samples of *P. thunbergii* under five temperature conditions (8, 13, 18, 23, 28 °C) with six biological replicates is in the first element of the list.

A gene expression data matrix of another 30 root samples of *P. thunbergii* under the same condition is in the second one.

Temperature conditions where 30 root samples in each data matrix were generated are in the third one.

Gene expressions are normalized in the TPM value.

**Source**

original (not published)

**References**

original (not published)

---

|         |   |
|---------|---|
| q.clean | <i>Clean data by eliminating genes with many missing values</i> |
|---------|---|

---

**Description**

Clean data by eliminating genes with many missing values

**Usage**

```
q.clean(x, missing = 0.1, lowest = 10)
```

**Arguments**

|         |  |
|---------|--|
| x       | A data matrix (raw: samples, col: genes).                                      |
| missing | A ratio of missing values in each column allowed to be remained in the data.   |
| lowest  | The lowest value recognized in the data (e.g., TPM, FPKM, or raw read counts). |

**Value**

A data matrix (raw: samples, col: qualified genes)

**Author(s)**

Takahiko Koizumi

**Examples**

```
data(Pinus)
train.raw <- Pinus$train
ncol(train.raw)

train <- q.clean(train.raw)
ncol(train)
```

---

|       |  |
|-------|--|
| q.opt | <i>Estimate the optimal number of genes to construct QuEST model</i> |
|-------|--|

---

**Description**

Estimate the optimal number of genes to construct QuEST model

**Usage**

```
q.opt(x, y, range = 5:50, method = "linear", rep = 1)
```

**Arguments**

|        |  |
|--------|--|
| x      | A data matrix (row: samples, col: genes).  |
| y      | A vector of an environment in which the samples were collected.  |
| range  | A sequence of numbers of genes to be tested for MAE calculation (default: 5:50).   |
| method | A string to specify the method of regression for calculating R-squared values. "linear" (default), "quadratic" or "cubic" regression model can be specified. |
| rep    | The number of replications for each case set by range (default: 1).  |

**Value**

A sample-MAE curve

**Author(s)**

Takahiko Koizumi

**Examples**

```
data(Pinus)
train <- q.clean(Pinus$train)
target <- Pinus$target
q.opt(train[1:10, ], target[1:10], range = 5:15)
```

---

|       |   |
|-------|---|
| q.pca | <i>Visualize gene expression similarity using principal coordinate analysis</i> |
|-------|---|

---

**Description**

Visualize gene expression similarity using principal coordinate analysis

**Usage**

```
q.pca(x, y, method = "linear", lower.thr = 0, n.gene = ncol(x), size = 1)
```

**Arguments**

|           |  |
|-----------|--|
| x         | A data matrix (row: samples, col: genes).  |
| y         | A vector of an environment in which the samples were collected.  |
| method    | A string to specify the method of regression for calculating R-squared values. "linear" (default), "quadratic" or "cubic" regression model can be specified. |
| lower.thr | The lower threshold of R-squared value to be indicated in a PCA plot (default: 0).   |
| n.gene    | The number of candidate genes for QuEST model to be indicated in a PCA plot (default: ncol(x)).  |
| size      | The size of symbols in a PCA plot (default: 1).  |

**Value**

A PCA plot

**Author(s)**

Takahiko Koizumi

**Examples**

```
data(Pinus)
train <- q.clean(Pinus$train)
target <- Pinus$target
q.pca(train, target)
```

---

|        |   |
|--------|---|
| q.rank | <i>Visualize R-squared value distribution in gene-environment interaction</i> |
|--------|---|

---

**Description**

Visualize R-squared value distribution in gene-environment interaction

**Usage**

```
q.rank(
  x,
  y,
  method = "linear",
  lower.thr = 0,
  n.gene = ncol(x),
  upper.xlim = ncol(x)
)
```

**Arguments**

|            |  |
|------------|--|
| x          | A data matrix (row: samples, col: genes).  |
| y          | A vector of an environment in which the samples were collected.  |
| method     | A string to specify the method of regression for calculating R-squared values. "linear" (default), "quadratic" or "cubic" regression model can be specified. |
| lower.thr  | The lower threshold of R-squared value to be included in QuEST model (default: 0).   |
| n.gene     | The number of genes to be included in QuEST model (default: ncol(x)).  |
| upper.xlim | The upper limitation of x axis (i.e., the number of genes) in the resulted figure (default: ncol(x)).  |

**Value**

A rank order plot

**Author(s)**

Takahiko Koizumi

**Examples**

```
data(Pinus)
train <- q.clean(Pinus$train)
target <- Pinus$target
train <- q.sort(train, target)
q.rank(train, target)
```

---

|        |  |
|--------|--|
| q.sort | <i>Sort and truncate genes according to the strength of gene-environment interaction</i> |
|--------|--|

---

**Description**

Sort and truncate genes according to the strength of gene-environment interaction

**Usage**

```
q.sort(x, y, method = "linear", n.gene = ncol(x), trunc = 1)
```

**Arguments**

|        |  |
|--------|--|
| x      | A data matrix (raw: samples, col: genes).  |
| y      | A vector of an environment in which the samples were collected.  |
| method | A string to specify the method of regression for calculating R-squared values. "linear" (default), "quadratic" or "cubic" regression model can be specified. |
| n.gene | The number of genes to be included in QuEST model (default: ncol(x)).  |
| trunc  | a threshold to be truncated (default: 1).  |

**Value**

A data matrix (raw: samples, col: sorted genes)

**Author(s)**

Takahiko Koizumi

**Examples**

```

data(Pinus)
train <- q.clean(Pinus$train)
target <- Pinus$target
cor(target, train[, 1])

train <- q.sort(train, target, trunc = 0.5)
cor(target, train[, 1])

```

quest

*Construct and apply the QuEST model with your own data***Description**

Construct and apply the QuEST model with your own data

**Usage**

```
quest(x, y, newx = x, method = "linear", lower.thr = 0, n.gene = 0)
```

**Arguments**

|           |  |
|-----------|--|
| x         | A data matrix (row: samples, col: genes).  |
| y         | A vector of an environment in which the samples were collected.  |
| newx      | A data matrix (row: samples, col: genes).  |
| method    | A string to specify the method of regression for calculating R-squared values. "linear" (default), "quadratic" or "cubic" regression model can be specified. |
| lower.thr | The lower threshold of R-squared value to be used in QuEST model (default: 0).   |
| n.gene    | The number of candidate genes to be used in QuEST model (default: 30).   |

**Value**

A vector of the environment in which the samples of newx were collected

**Author(s)**

Takahiko Koizumi

**Examples**

```

data(Pinus)
train <- q.clean(Pinus$train)
test <- Pinus$test
test <- test[, colnames(train)]
target <- Pinus$target
cor(target, quest(train, target, newx = test, method = "cubic"))

```

# Index

\* **dataset**  
Pinus, 2

Pinus, 2

q.clean, 3  
q.opt, 3  
q.pca, 4  
q.rank, 5  
q.sort, 6  
quest, 7