Package 'AdvISERMPYRO'

August 6, 2014

Type Package

Title Identification using SparsE Representation of Multiplex PYROsequencing signal

Version 1.0

Date 2014-04-11

Author Jerome Ambroise

Maintainer <jerome.ambroise@uclouvain.be>

Depends penalized

Description Pyrosequencing is a cost-effective DNA sequencing technology that has many applications including rapid genotyping of a broad spectrum of bacteria. If molecular typing requires to genotype multiple DNA stretches, several pyrosequencing primers could be used simultaneously, hence creating overlapping primer-specific signals which are not visually uninterpretable. AdvISER-M-PYRO is an efficient algorithm that can reliably be used to translate the global multiplex pyrosequencing signal in sequences corresponding to each genomic region. the new SENATOR method was developed to select and optimize the nucleotide dispensation order to be used all along the multiplex pyrosequencing experiment. Irrespective of the location of the pyrosequencing primer, the dispensation order is a crucial feature avoiding similarities between uniplex pyrosequencing signals.

License GPL-2

R topics documented:

1

AdvISERMPYRO	*Amplicon Identification using SparsE Representation of Mulitplex PY-ROsequencing signal*

Usage

```
AdvISERMPYRO(signaltest, dictionary, L1, SCT)
```

Arguments

signaltest	Testing Multiplex pyrosequencing signal of length n. Each element of this signal corresponds to a peak height. The length of the signal (n) is therefore equal to the number of dispensated nucleotides.
dictionary	Representative dictionary that include at least one signal (of length n) for each potential Unique Nucleotide Sequence in each genomic region.
L1	L1-norm penalty of the regression model used within AdvISERMPYRO. Default value = 0.05.
SCT	Significant Contribution Threshold. Any Unique Nucleotide Sequence having a contribution to the testing signal lower than the SCT is removed. Default value = 1.

Value

A vector with that includes for each genomic region, the amplicon that was identified by AdvIS-ERMPYRO and its contribution to the signal. This vector also includes a reliability index of the global identification (r). This index corresponds to the correlation between the n predicted values of the penalized regression model and the n values of the pyrosequenced signal. An r index close to 1 indicates therefore a reliable identification.

Author(s)

Ambroise J.

References

Ambroise J., Deccache Y., Irenge L., Encho S., Robert A., Gala J.-L.: Amplicon Identification using SparsE Representation of Multiplex PYROsequencing signal (AdvISER-M-PYRO): Application to bacterial resistance genotyping. Bioinformatics, 2014

Examples

```
data(dictionary)
data(BS035)
AdvISERMPYRO(signaltest=BS035,dictionary=dictionary,L1=0.05,SCT=1)
```

ATCC35218	*Multiplex pyrosequencing signal generated from ATCC35218 bacterial strain*

Usage

```
data(ATCC35218)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain ATCC35218.

Examples

```
data(ATCC35218)
```

ATCC700603	*Multiplex pyrosequencing signal generated from ATCC700603 bacterial strain*

Usage

```
data(ATCC700603)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain ATCC700603.

Examples

```
data(ATCC700603)
```

BS031	*Multiplex pyrosequencing signal generated from BS031 bacterial strain*

Usage

```
data(BS031)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain BS031.

Examples

```
data(BS031)
```

BS035	*Multiplex pyrosequencing signal generated from BS035 bacterial strain*

Usage

```
data(BS035)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain BS035.

Examples

```
data(BS035)
```

---

dictionary	*Dictionary that includes theoritical and experimental simplex pyrosequencing signals for each amplicon that can be found in each genomic region.*

Usage

```
data(dictionary)
```

Format

the data frame includes 1 column for each signal in the dictionary. Several signals can correspond to the same UNS. The number of rows corresponds to the length of the dispensation order.

Details

The first 6 letters must correspond to the genomic region label while the next letters must correspond to the UNS label.

Examples

```
data(dictionary)
```

| DSM22313 | *Multiplex pyrosequencing signal generated from DSM22313 bacterial strain* |

Usage

```
data(DSM22313)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain DSM22313.

Examples

```
data(DSM22313)
```

| DSM22314 | *Multiplex pyrosequencing signal generated from DSM22314 bacterial strain* |

Usage

```
data(DSM22314)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain DSM22314.

Examples

```
data(DSM22314)
```

| evaluator | *evaluator* |

Description

Evaluate the quality of 1 dispensation order

Usage

```
evaluator(seqlist, dispensation)
```

Arguments

```
seqlist
dispensation
```

MMA55   *Multiplex pyrosequencing signal generated from MMA55 bacterial strain*

Usage

```
data(MMA55)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain ATCC35218.

Examples

```
data(MMA55)
```

---

pyrogenerator   *Generation of a theoritical pyrosequencing signal*

Usage

```
pyrogenerator(sequence, dispensation)
```

Arguments

```
sequence
dispensation
```

R021   *Multiplex pyrosequencing signal generated from R021 bacterial strain*

Usage

```
data(R021)
```

Format

This vector contains the peak heights of the muliplex pyrosequencing signal observed for strain R021.

Examples

```
data(R021)
```

---

senator   *SElecting the Nucleotide dispensATion Order*

Description

the new SENATOR method was developed to select and optimize the nucleotide dispensation order to be used all along the multiplex pyrosequencing experiment. Irrespective of the location of the pyrosequencing primer, the dispensation order is a crucial feature avoiding similarities between uniplex pyrosequencing signals.

Usage

```
senator(seqlist, length.disp, Ntest, Nresult)
```

Arguments

seqlist	A list with all Unique Nucleotide Sequences expected to be found within each genomic region.
length.disp	The length of the dispensation order candidates.
Ntest	The number of dispensation order candidates.
Nresult	The number of best dispensation orders which are displayed.

Value

A data.frame with the best dispensation order candidates at the first lines. Each dispensation order is associated to the maximum value of all pairwise correlation coefficients between uniplex pyrosequencing signals.

Author(s)

Ambroise J.

References

Ambroise J., Deccache Y., Irenge L., Encho S., Robert A., Gala J.-L.: Amplicon Identification using SparsE Representation of Multiplex PYROsequencing signal (AdvISER-M-PYRO): Application to bacterial resistance genotyping. Bioinformatics, 2014

Examples

```
data(seqlist)

senator(seqlist,length.disp=14,Ntest=100,Nresult=5)
```

seqlist	*List of all UNS expected to be found in each genomic region*

Usage

```
data(seqlist)
```

Examples

```
data(seqlist)
```

Index