

# Package ‘tAI’

July 22, 2025

**Type** Package

**Title** The tRNA Adaptation Index

**Version** 0.2.2

**Date** 2025-03-19

**Description** Functions and example files to calculate the tRNA adaptation index, a measure of the level of co-adaptation between the set of tRNA genes and the codon usage bias of protein-coding genes in a given genome. The methodology is described in dos Reis, Wernisch and Savva (2003) <[doi:10.1093/nar/gkg897](https://doi.org/10.1093/nar/gkg897)>, and dos Reis, Savva and Wernisch (2004) <[doi:10.1093/nar/gkh834](https://doi.org/10.1093/nar/gkh834)>.

**License** GPL (>= 2)

**LazyData** TRUE

**RoxygenNote** 7.3.2

**Encoding** UTF-8

**NeedsCompilation** no

**Author** Mario dos Reis [aut, cre]

**Maintainer** Mario dos Reis <[mariodosreis@gmail.com](mailto:mariodosreis@gmail.com)>

**Repository** CRAN

**Date/Publication** 2025-04-16 14:50:02 UTC

## Contents

ecolik12 . . . . .	2
get.s . . . . .	2
get.tai . . . . .	3
get.ws . . . . .	4
nc.adj . . . . .	5
nc.f . . . . .	6
ts.test . . . . .	7

<b>Index</b>	<b>9</b>
--------------	----------

---

 ecolik12

*E. coli K-12 codon bias and tRNA numbers*


---

### Description

A list with elements `trna`, a vector of length 64 of tRNA gene copy numbers in the Escherichia coli K-12 genome, `w`, a data frame with some codon bias statistics for 49 E. coli K-12 coding genes, and `m`, a 49 by 61 matrix of codon frequencies for the 49 genes in question.

### Usage

```
ecolik12
```

### Format

An object of class `list` of length 3.

### Author(s)

Mario dos Reis

### Examples

```
# 87 tRNA genes in the E. coli K-12 genome:
sum(ecolik12$trna)

# Two copies are isoacceptors for Phe, with anticodon GAA (codon TTC)
ecolik12$trna[2]

# ecolik12$w, a data frame with codon bias statistics
names(ecolik12$w)

# Effective number of codons vs. gene length (in codons)
plot(ecolik12$w$Nc, ecolik12$w$L_aa, xlab="Nc", ylab="Gene length")
```

---

 get.s

*Correlation between tAI and Nc adjusted*


---

### Description

Calculates the correlation between tAI and Nc (adjusted for GC content at third codon positions).

### Usage

```
get.s(tAI, nc, gc3)
```

**Arguments**

tAI                    a vector of length n with tAI values for genes  
nc                     a vector of length n with Nc values for genes  
gc3                    a vector of length n with GC content at third codon positions for genes

**Value**

Numeric of length one with the correlation between tAI and Nc adjusted

**Author(s)**

Mario dos Reis

---

*get.tai*                    *The tRNA adaptation index*

---

**Description**

Calculates the tRNA adaptation index (tAI) of dos Reis et al. (2003, 2004).

**Usage**

`get.tai(x, w)`

**Arguments**

x                      an n by 60 matrix of codon frequencies for n open reading frames.  
w                      a vector of length 60 of relative adaptiveness values for codons.

**Details**

The tRNA adaptation index (tAI) is a measure of the level of co-adaptation between the set of tRNA genes and the codon usage bias of protein-coding genes in a given genome. STOP and methionine codons are ignored. The standard genetic code is assumed.

**Value**

A vector of length n of tAI values.

**Author(s)**

Mario dos Reis

## References

dos Reis M., Wernisch L., and Savva R. (2003) Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.*, **31**: 6976–85.

dos Reis M., Savva R., and Wernisch L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**: 5036–44.

## See Also

[get.ws](#)

## Examples

```
# Calculate relative adaptiveness values (ws) for E. coli K-12
eco.ws <- get.ws(trNA=ecolik12$trna, sking=1)

# Calculate tAI for a set of 49 E. coli K-12 coding genes
eco.tai <- get.tai(ecolik12$m[,-33], eco.ws)

# Plot tAI vs. effective number of codons (Nc)
plot(eco.tai, ecolik12$w$Nc, xlab="tAI", ylab="Nc")
```

---

<code>get.ws</code>	<i>Relative adaptiveness values</i>
---------------------	-------------------------------------

---

## Description

Calculates the relative adaptiveness values of codons based on the number of tRNA genes.

## Usage

```
get.ws(trNA, s = NULL, sking)
```

## Arguments

<code>trNA</code>	a vector of length 64 with tRNA gene copy numbers
<code>s</code>	a vector of length 9 with selection penalties for codons
<code>sking</code>	a vector of length 1 indicating the superkingdom

## Details

The relative adaptiveness values are calculated as described in dos Reis et al. (2003, 2004). If `s = NULL`, the `s` values are set to the optimised values of dos Reis et al. (2004). `sking` indicates the superkingdom, with 0 indicating Eukaryota, and 1 Prokaryota.

**Value**

A vector of length 60 of relative adaptiveness values.

**Author(s)**

Mario dos Reis

**References**

dos Reis M., Wernisch L., and Savva R. (2003) Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.*, **31**: 6976–85.

dos Reis M., Savva R., and Wernisch L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**: 5036–44.

**Examples**

```
eco.ws <- get.ws(trNA=ecolik12$trna, sking=1)
```

---

 nc.adj

*Adjusted effective number of codons (Nc)*

---

**Description**

The adjusted Nc is  $f(gc3s) \cdot Nc$

**Usage**

```
nc.adj(nc, gc3)
```

**Arguments**

nc a vector of length n with the effective number of codons for genes

gc3 a vector of length n with corresponding GC composition at third codon positions

**Details**

The adjusted Nc is calculated as described in dos Reis et al. (2004).

**Value**

A vector of length n with adjusted Nc values

**Author(s)**

Mario dos Reis

## References

dos Reis M., Savva R., and Wernisch L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**: 5036–44.

## See Also

[nc.f](#) for the function used to calculate f(gc3s)

## Examples

```
eco.ncadj <- nc.adj(ecolik12$w$Nc, ecolik12$w$GC3s)
plot(eco.ncadj ~ ecolik12$w$Nc, xlab="Nc", ylab="Nc adjusted")
```

---

nc.f

*Nc vs. GC3s*

---

## Description

Calculates the expected Nc value of a gene for a given GC content at the third codon positions.

## Usage

```
nc.f(x)
```

## Arguments

x a vector of GC contents at third codon positions

## Details

Without selection on codon bias, the expected value of Nc as a function of GC content at third positions, x, is given by

$$f(x) = -6 + x + 34/(x^2 + (1.025 - x)^2).$$

This equation follows dos Reis et al. (2004, see also Wright 1990 for the original).

## Value

A vector of Nc values for the given GC contents.

## Author(s)

Mario dos Reis

## References

- Wright F. (1990) The 'effective number of codons' used in a gene. *Gene*, **87**: 23–9.
- dos Reis M., Savva R., and Wernisch L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**: 5036–44.

## Examples

```
curve(nc.f(x), xlab="GC3s content", ylab="Nc")
points(ecolik12$w$GC3s, ecolik12$w$Nc, pch=19)
```

---

ts.test

*Monte Carlo test of correlation between tAI and Nc adjusted*

---

## Description

Calculates the p-value (using a Monte Carlo or randomisation test) that the correlation (the S value) between tAI and the adjusted Nc for a set of genes is different from zero.

## Usage

```
ts.test(m, ws, nc, gc3s, ts.obs, samp.size, n = 1000)
```

## Arguments

m	a k by 60 matrix of codon frequencies for k genes
ws	vector of length 60 of relative adaptiveness values of codons
nc	vector of length k of Nc values for genes
gc3s	vector of length k of GC content at third codon position for genes
ts.obs	vector of length 1 with observed correlation between tAI and Nc adjusted for the k genes
samp.size	a vector of length 1 with the number of genes to be sampled from m (see details)
n	the number of permutations of ws in the randomisation test

## Details

The Monte Carlo test is described in dos Reis et al. (2004). When working with complete genomes, matrix m can have a very large number of rows (large k). In this case it may be advisable to choose samp.size < k to speed up the computation.

## Value

A list with elements p.value, the p-value for the test, and ts.simulated, a vector of length n with the simulated correlations between tAI and adjusted Nc.

**Author(s)**

Mario dos Reis

**References**

dos Reis M., Savva R., and Wernisch L. (2004) Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Res.*, **32**: 5036–44.

**Examples**

```
eco.ws <- get.ws(trNA=ecolik12$trna, sking=1)
eco.tai <- get.tai(ecolik12$m[,-33], eco.ws)
ts.obs <- get.s(eco.tai, ecolik12$w$Nc, ecolik12$w$GC3s)

# The S-value (dos Reis et al. 2004):
ts.obs # [1] 0.9065442

# There seems to be a high correlation between tAI and Nc adjusted for
# the 49 genes in ecolik12$m. Is the correlation statistically significant?
ts.mc <- ts.test(ecolik12$m[,-33], eco.ws, ecolik12$w$Nc, ecolik12$w$GC3s,
                ts.obs, samp.size=dim(ecolik12$m)[1])
# The p-value is zero:
ts.mc$p.value # [1] 0

# Histogram of simulated S-values:
hist(ts.mc$ts.simulated, n=50, xlab = "Simulated S values",
     xlim=c(min(ts.mc$ts.simulated), ts.obs))
# Add the observed S-value as a red vertical line:
abline(v=ts.obs, col="red")
```



# Index

## \* datasets

ecolik12, [2](#)

ecolik12, [2](#)

get.s, [2](#)

get.tai, [3](#)

get.ws, [4](#), [4](#)

nc.adj, [5](#)

nc.f, [6](#), [6](#)

ts.test, [7](#)