

GROWTHPRED

Main references:

The background on the method can be found in:

Vieira-Silva S, Rocha EPC (2010) The Systemic Imprint of Growth and its Uses in Ecological (Meta)genomics. PLoS genetics doi: 10.1371/journal.pgen.1000808.

The manuscript describing this server is in preparation:

Vieira-Silva S *et al.* Growthpred: Sequence based prediction of minimum generation times for Prokaryotes. Nucleic Acids Research (Web Server issue). *In preparation.*

General description:

This application predicts the minimum generation time for a bacterial or archaeal organism based on its codon usage bias intensity (CUB). The CUB index is calculated given two input sets of sequences: 1) **input_1**: highly expressed genes 2) **input_2**: coding sequences from the complete genome or subset of non-HEG genes.

Growthpred calculates the codon bias index F [1]. F is adapted from $\Delta ENC'$ [2] and S [3]. The predicted minimum generation time (d in hours) is calculated from a linear model built on a dataset of 214 bacteria and archaea, which correlates F and the box-cox transformation of d ($R^2=0.58$). The optimal growth temperature (OGT in °C) can be taken into account if the organism is not mesophilic. If there is no prior knowledge about OGT, growthpred can estimate it from amino acid frequency following the equations in [4].

$$d = \left[1 - 0.1664 \cdot (1.743 - 0.7372 \cdot F - 0.0226 \cdot OGT) \right]^{-1/0.1664}$$

The application outputs:

1. **outfile.results** with average and the standard deviation of the predicted minimum generation time, the quartet nucleotide frequencies, the number of sequences in the two inputs and the optimal growth temperature.
2. **outfile.errors** with the identifiers of ignored sequences (i.e. which have internal stop codons or length not multiple of 3).
3. optional: **outfile.ribs** with ribosomal protein genes retrieved by similarity with a database.
4. optional: **outfile.cub** with codon usage bias indexes (ENC [5], ENC' [6], P [1], CAI [7]) for each gene in `input_2`.

Growthpred options:

1. Input:

- a. Input_1 and input_2 sequences can be pasted or uploaded in the entry fields. **(-f -g)**
- b. Input_1 is retrieved from input_2 by similarity with a database of HEG genes. **(-b)**
- c. Input_1 and input_2 are replaced by *E. coli* K-12 example files: ribosomal coding sequences and complete genome coding sequences. **(-e)**

2. Input type: (cf. [1] for details)

- a. Inputs sequences are from one organism (sequences can be concatenated to increase the accuracy of the generation time prediction)
- b. Inputs sequences are from several organisms (e.g. metagenome) (sequences can not be concatenated) **(-m)**

3. Prediction options:

- a. Select the genetic code **(-c)**
 - 0 : Universal
 - 1 : Yeast mitochondrial
 - 2 : Vertebrate mitochondrial
 - 3 : Mold/Protozoan/Mycoplasma/Spiroplasma
 - 4 : Invertebrate mitochondrial
 - 5 : Candida cylindracea
 - 6: Ciliate
 - 7 : Euplotes
 - 8 : Echinoderm mitochondrial
- b. Remove start **(-s)** and stop codons **(-S)** from each sequence
- c. Optimal growth temperature (OGT):
 - default = 36
 - user defined in entry field **(-T)**
 - estimated from amino acid frequency in input_2 **(-t)**

4. Output:

- a. Select a name for your output **(-o)**
- b. If **(-b)** recover the file of retrieved ribosomal coding sequences (FASTA) **(-r)**
- c. Recover the file of codon usage bias indexes for each gene in input_2 **(-i)**
 - ENC [5]
 - ENC'[6]
 - P [1]
 - CAI[7]

References:

1. Vieira-Silva S, Rocha EPC (2010) The Systemic Imprint of Growth and its Uses in Ecological (Meta)genomics. PLoS genetics doi: 10.1371/journal.pgen.1000808.
2. Rocha EP (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res 14: 2279-2286.
3. Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE (2005) Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res 33: 1141-1153.
4. Zeldovich KB, Berezovsky IN, Shakhnovich EI (2007) Protein and DNA sequence determinants of thermophilic adaptation. PLoS Comput Biol 3: e5.
5. Wright F (1990) The 'effective number of codons' used in a gene. Gene 87: 23-29.
6. Novembre JA (2002) Accounting for Background Nucleotide Composition When Measuring Codon Usage Bias. Mol Biol Evol 19: 1390-1394.
7. Sharp PM, Li WH (1987) The Codon Adaptation Index - a Measure of Directional Synonymous Codon Usage Bias, and Its Potential Applications. Nucleic Acids Research 15: 1281-1295.