
Gene enumeration

X48860_en

Recall that in prokaryotic genomes, the sequence coding for a protein occurs as one contiguous open reading frame, and that an open reading frame begins with the start codon ATG (methionine) in most species and ends with a stop codon TAA, TAG, or TGA.

For example, the DNA sequence of Bacteriophage ϕ -X174, which was the first genome to be sequenced, has 117 open reading frames (11 of which are protein coding genes) within a circular single strand of 5,386 nucleotides.

Write pseudocode, Python code, and C++ code for the gene enumeration problem. Make two submissions, including the pseudocode as a comment to both the Python and the C++ code.

Input

The input is a string s (a genomic sequence) over the alphabet $\Sigma = \{A, C, G, T\}$.

Output

The output is all minimal substrings of s (all open reading frames) from a start codon to a stop codon.

Sample input

```
GAGTTTTATCGCTTCCATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTG
```

Sample output

```
ATGACGCAGAAGTTAACACTTTCGGATATTTCTGATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTGTTACGAATTAAATCG
ATGAGTCGAAAAATTATCTTGATAAAGCAGGAATTACTACTGCTTGTTACGAATTAAATCGAAGTGGACTGCTGGCGGAAAATGAGAAAATTCGA
ATGAGAAAATTCGACCTATCCTTGCGCAGCTCGAGAAGCTTTACTTTGCGACCTTCGCCATCAACTAACGATTCTGTCAAAAACGACGCGTTG
ATGAGGAGAAGTGGCTTAATATGCTTGGCACGTTTCGTCAGGACTGGTTAGATATGAGTCACATTTTGTTCATGGTAGAGATTCTCTTGTGACA
ATGCTTGGCACGTTTCGTCAGGACTGGTTAGATATGAGTCACATTTTGTTCATGGTAGAGATTCTCTTGTGACATTTTAAAGAGCGTGGATTA
ATGAGTCACATTTTGTTCATGGTAGAGATTCTCTTGTGACATTTTAA
ATGGTAGAGATTCTCTTGTGACATTTTAA
ATGCTGTCAACCACTAA
```

Problem information

Author : Gabriel Valiente

Generation : 2021-09-30 17:01:07

© Jutge.org, 2006–2021.

<https://jutge.org>