

Gene Coding Regions (2)

V62955_en

In a DNA sequence, start codons and stop codons delimit a coding region of a gene –a.k.a. coding DNA sequence (CDS)–, which is the part of a gene that is translated into a protein. Any CDS starts with the universal start codon ATG and ends with the first occurrence of a stop codon (either TAA, TAG, or TGA).

Write a program that, given a DNA sequence, counts codons inside each CDS found in the sequence, and prints the count of each codon in the CDS.

Input

A sequence of codons (triplets of A, T, C, G), which may contain zero or more CDS delimited by start/stop codons.

The input sequence may appear in several lines, with one or more whitespaces or newlines between one codon and the next.

Assume that the CDS in the sequence –if any– are well-formed. That is, if a start codon appears, an end codon will appear later before the sequence ends or the start codon appears again. No end codon will appear if no start codon has previously appeared.

Output

For each CDS (fragment comprised between a start and an end codon), the codons belonging to the CDS are printed in alphabetical order, along with their frequencies.

If no CDS is found in the sequence, the output is "No CDS found".

Sample input 1

```
TGC ATG GCT CCG GCT AAG
TAA TGC CGT ATG
AAT CTC AAT GAG AAT CCG TAG AAG
```

Sample output 1

```
CDS 1
AAG 1
CCG 1
GCT 2
CDS 2
AAT 3
CCG 1
CTC 1
GAG 1
```

Sample input 2

```
CGT TGC TAC TGC CGT TGC GCT CCG
GCT AAG AAT CTC AAT GAG GAT AAG
```

Sample output 2

```
No CDS found
```

Sample input 3

```
CCA CGG ATA GAA ACA GGA CCC CGA AAG CTG CAC ACC GAC CAA
TAC ATG GTG GGT AGA TCT GAG GCA CTT TTT TTA ACA TAT TAT
GGC TCT TCA TGC CCT GGT CGA AGG CAG TAG CGT TTA CGT CAG
TGT CGT ACT TAT GGA ACG TCC GAT TGG TTT TTG ATC ATG GTC
CAC CTC AAG TAT CAT AAC AGA TTT GTT GGT GCC CCA CTA CTA
GTA GAT TCG TGG GGA TCG GTG AAA ACC CGG ACC TGA AAG CAG
GTC GTG CTC TAT CGG GAC CGG GAA AAG AAT TGG AGT CCC TTC CTC
GAC CTG CTT CCT CCA CCT CTA GGC AGC AAC TCA ACA AAA ACC
```

```
AGA GTG ATA CCC GCT ATG TCA ACC CGA ACT AGG CCG GCA TCG
GCA TTT GCC GTT CAG AAC CGC AGC TTG CAT CGT ACA GGG ACC
GGT CCC AAA CTA AAA CTC TAA CTC ACC ACG CGA AAA CTA ATC
TCA ACA TAT TAT
CAG CAT AGT ATT GCT CTA CAG ACT AAC TAC
```

Sample output 3

CDS 1	GGT 1
ACA 1	GTA 1
AGA 1	GTC 1
AGG 1	GTG 1
CAG 1	GTT 1
CCT 1	TAT 1
CGA 1	TCG 2
CTT 1	TGG 1
GAG 1	TTT 1
GCA 1	CDS 3
GGC 1	AAA 2
GGT 2	AAC 1
GTG 1	ACA 1
TAT 2	ACC 2
TCA 1	ACT 1
TCT 2	AGC 1
TGC 1	AGG 1
TTA 1	CAG 1
TTT 1	CAT 1
CDS 2	CCC 1
AAA 1	CCG 1
AAC 1	CGA 1
AAG 1	CGC 1
ACC 2	CGT 1
AGA 1	CTA 1
CAC 1	CTC 1
CAT 1	GCA 2
CCA 1	GCC 1
CGG 1	GGG 1
CTA 2	GGT 1
CTC 1	GTT 1
GAT 1	TCA 1
GCC 1	TCG 1
GGA 1	TTG 1
	TTT 1

Problem information

Author: Lluís Padró

Generation: 2026-01-25T13:11:04.932Z

© *Jutge.org*, 2006–2026.

<https://jutge.org>