

TD Méthodologie de la Programmation

29 janvier 2021

Recommandations

Pour votre convenance, l'énoncé est disponible au format .ipynb et .pdf. **Il s'agit du même exercice.** Vous pouvez réaliser les exercices directement sur jupyter et rendre le fichier .ipynb modifié par vos soins. Alternativement, vous pouvez rendre l'exercice directement en python (fichier .py).

Dans tous les cas, pensez bien à indiquer votre nom, prénom et n. d'étudiant.

Vous devrez envoyer votre travail (exercice et fichier de sortie) sur le mail `josue.melka02@univ-paris8.fr`. Tout travail rendu après le délai fixé ne sera pas pris en compte.

Important: inutile de s'arrêter sur les notions de biologie exposées pour réaliser les exercices, elles servent ici simplement d'illustration. Au besoin, reportez vous aux exemples pour mieux comprendre les questions.

1 Exercices

L'ADN est constitué de deux brins formant une double hélice. Chaque brin est formé par une chaîne de *nucléotides*, qui contiennent l'une des bases suivantes:

- A pour *Adénine*
- T pour *Thymine*
- G pour *Guanine*
- C pour *Cytosine*

L'ordre dans lequel se succèdent les nucléotides le long d'un brin d'ADN constitue la *séquence* de ce brin. C'est elle qui porte l'information génétique.

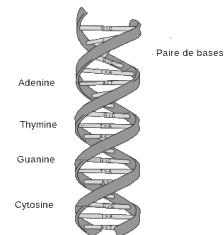


Illustration (Wikipedia)

1.1 Lecture d'une séquence

Nous pouvons stocker une séquence d'ADN dans un fichier simplement à l'aide d'une chaîne de caractères de cette forme: 'AATGTCCAGTCAGTTG... '.

Écrire une fonction `lire_adn(filename)` qui lit et retourne le contenu d'un tel fichier.
Par exemple, `lire_adn("ncov-s.txt")` retournera la séquence "ATGTTGTTTTTCTT... "

1.2 Transcription

La *transcription* est un mécanisme qui permet de copier l'ADN dans le noyau de la cellule pour former un ARN.

Lors de la transcription, la Thymine (T) est remplacée par l'Uracile (U).

Écrire une fonction `transcrire_adn(seq)` qui transcrit une séquence d'ADN en ARN.
Par exemple, `transcrire_adn('AATGTCCAGTCAG')` doit renvoyer 'AAUGUCCAGUCAG'.

1.3 Découpage de l'ARN

L'ARN est découpé en *codons*, constitués de 3 bases successives, qui correspondent chacun à un acide aminé. Ils sont lus les uns à la suite des autres.

Écrire une fonction `arn_vers_codons(seq)` permettant de découper une séquence d'ADN en une liste de codons.

Par exemple, `arn_vers_codons('AAUGUCCAGUCA')` renverra: `['AAU', 'GUC', 'CAG', 'UCA']`

1.4 Traduction

La *traduction* de l'ADN consiste à lire l'ARN issu de la transcription pour synthétiser une protéine sous forme d'une chaîne d'acides aminés.

Le dictionnaire ci-dessous donne la correspondance entre un codon et un acide aminé. Par exemple, le codon GUC code pour l'acide aminé Val.

```
codage = {'GUC': 'Val', 'GUG': 'Val', 'UCA': 'Ser', 'CUC': 'Leu', 'AUU': 'Ile', 'CCU': 'Pro', 'CGG': 'Arg',
          'GGU': 'Gly', 'CAU': 'His', 'CGU': 'Arg', 'GCG': 'Ala', 'AGA': 'Arg', 'CUG': 'Leu', 'GUU': 'Val',
          'AUA': 'Ile', 'UGU': 'Cys', 'GCC': 'Ala', 'UAU': 'Tyr', 'ACU': 'Thr', 'UGG': 'Trp', 'CAG': 'Gln',
          'CAC': 'His', 'AGC': 'Ser', 'GGG': 'Gly', 'CUU': 'Leu', 'ACA': 'Thr', 'CUA': 'Leu', 'GAA': 'Glu',
          'AUC': 'Ile', 'CGC': 'Arg', 'UGC': 'Cys', 'UCC': 'Ser', 'AAU': 'Asn', 'UUC': 'Phe', 'CCC': 'Pro',
          'AGG': 'Arg', 'UGA': None, 'UAG': None, 'UAA': None, 'AUG': 'Met', 'UUG': 'Leu', 'UUA': 'Leu',
          'AGU': 'Ser', 'GGC': 'Gly', 'GAU': 'Asp', 'CCA': 'Pro', 'CGA': 'Arg', 'GUA': 'Val', 'CAA': 'Gln',
          'UCU': 'Ser', 'UCG': 'Ser', 'ACC': 'Thr', 'GAG': 'Glu', 'GGA': 'Gly', 'GCU': 'Ala', 'GAC': 'Asp',
          'AAC': 'Asn', 'AAG': 'Lys', 'GCA': 'Ala', 'CCG': 'Pro', 'AAA': 'Lys', 'ACG': 'Thr', 'UAC': 'Tyr',
          'UUU': 'Phe'}
```

On remarquera que plusieurs codons différents peuvent coder pour le même acide aminé. Notez également que certains codent pour la valeur None. Ces derniers indiquent la fin de la partie "codante" et **stoppent** la synthèse de la protéine.

Écrire une fonction `traduire_arn(seq)` qui traduit l'ARN et renvoie la liste d'acides aminés correspondante. **Attention**, la chaîne doit s'arrêter au codon STOP.

Par exemple, `traduire_arn('AAUGUCCAGUAGUCA')` retournera `['Asn', 'Val', 'Gln']`.

1.5 Acide aminés

On souhaite avoir un tableau permettant de savoir quels sont les codons qui correspondent à chaque acide aminé, sous cette forme:

```
Ala : GCA, GCC, GCG, GCU
Arg : AGA, AGG, CGA, CGC, CGG, CGU
Asn : AAC, AAU
...
```

Écrire le code permettant de générer le tableau complet, et l'enregistrer dans le fichier `amines.txt`. Ce fichier sera rendu avec le devoir.