Master de Bioinformatique **Perl**
Universités Bordeaux 1 & 2 **Mai 2006**

Perl - Evaluation
Durée : 2h

Modalités

Au début de l'épreuve, téléchargez l'archive contenant les fichiers fournis pour l'évaluation sur le site du cours : `http://dept-info.labri.fr/~dutour/ENS/Perl_2005-06`.

A l'issu de l'épreuve, vous enverrez par mail à `dutour@labri.fr` une archive dont le nom sera <votre_nom>.tgz (ou `.zip`), et qui contiendra l'ensemble des scripts demandés :
`exo1.pl`
`exo2.pl`
`exo3.pl`
`exo4.1.pl`, `exo4.2.pl`, `exo4.3.pl`

Exercices

Exercice 1. (3 points) Script mystère.
Fourni(s) : `exo1.pl`, `seq.fa`.
Consultez et exécutez le script fourni. Répondez alors aux questions posées directement à la fin du script lui-même :

1. Décrivez brièvement ce que fait ce script.
2. Justifiez en quelques mots la structure de l'algorithme mis en œuvre dans ce script.
3. Que doit-on changer (et pourquoi) si l'on veut rajouter, par exemple, le calcul du nombre de codons START et STOP dans ce script ?
   (On ne demande pas ici d'écrire l'algorithme, mais d'expliquer le concept qui doit changer pour l'écrire)

Exercice 2. (4 points) Codons Start et Stop.
Fourni(s) : `exo2.pl`, `multiSeqs.fa`.

1. Modifiez le script fourni pour qu'il compte le nombre de codons Start (ATG) et de codons Stop (TAA, TGA, TAG) dans chacune des séquences, et affiche les résultats sous le format suivant (fichier `multiSeqs.fa`) :
```
AB031069 :
        43 codons START
        64 codons STOP
HSP53G :
        239 codons START
        852 codons STOP
```
2. Justifiez, directement à la fin du script, les choix algorithmiques que vous avez faits pour ce traitement.

Exercice 3. (6,5 points) Traduction en séquence protéique.
Fourni(s) : `Code.txt`, `multiSeqs.fa`, `multiSeqstest.fa`.
Ecrivez le script `exo3.pl` en trois étapes successives (il n'est pas interdit de s'inspirer des scripts précédents) :

1. Gérer les arguments du script de sorte que l'on puisse l'utiliser avec deux fichiers en argument : `exo3.pl Code.txt multiSeqs.fa`.
2. Lire le fichier contenant le code et remplir une table de hachage associant chaque codon à l'acide aminé qu'il code.
   Remarque : les codons Stop coderont un acide aminé X comme suggéré dans le fichier `Code.txt` fourni.
3. Traduire chaque séquence du fichier multifasta passé en argument en séquence protéique et l'afficher à l'écran.
   Remarque : même si c'est faux biologiquement, traduire "mécaniquement" la séquence en entier, y compris les Start et les Stop.
   Exemple : résultat attendu pour le fichier `multiSeqstest.fa` :
```
SEQ1TEST :
RXIN
SEQ2TEST :
FPSSPRAPRGCWELXSERFYLGEMAPIMP
```

Exercice 4. (6,5 points) Lecture d'un fichier Genbank.
Fourni(s) : `AB031069.gb`, `AB042770.gb`.

1. Ecrivez le script `exo4.1.pl` qui permet de lire un fichier GenBank, et de stocker dans 2 variables string différentes :
   - l'ensemble des annotations (telles quelles, sans modification),
   - la séquence (filtrée correctement pour ne garder que les atgc).
2. Faites une copie de `exo4.1.pl` en `exo4.2.pl`.
   Puis modifiez ce script pour qu'il récupère les positions du CDS et affiche la séquence correspondante (le fichier `AB031069.gb` contient un seul CDS).
3. Faites une copie de `exo4.2.pl` en `exo4.3.pl`.
   Puis modifiez ce script pour qu'il traite et affiche TOUS les CDS présents (le fichier `AB042770.gb` contient plusieurs CDS).