

24/09/2008 – Durée : 2 heures

Consignes

Vous avez deux heures pour écrire les programmes Perl correspondant à chacun des trois exercices ci-dessous. Vous avez accès à vos notes de cours, ainsi qu'aux ressources disponibles en ligne.

A l'issue du contrôle, merci de m'envoyer un email avec vos programmes Perl en pièces jointes. SVP n'oubliez pas de mentionner vos nom et prénom sous forme d'un commentaire en entête de chacun de vos programmes.

NB : la note tiendra compte de la qualité des commentaires et de la présentation des résultats.

Exercice 1 : Manipulation de références

Soit les variables radicaux et terminaisons telles que :

```
@radicaux = ("aimer", "manger", "donner");
```

```
@terminaisons = ("ai", "as", "a", "ons", "ez", "ont");
```

Ecrire un programme Perl (fichier nommé `exo1.pl`) permettant de conjuguer les radicaux ci-dessus au futur avec affichage à l'écran. Votre programme utilisera une fonction nommée `conjuguer`.

Exercice 2 : Traitement de séquences

Sur la base de données PDB (<http://www.rcsb.org/pdb/home/home.do>), vous allez récupérer le fichier au format PDB de la protéine ayant le numéro d'accès 1AJQ.

1. Ecrire un programme (fichier nommé `exo2.pl`) qui affiche le NOM de la protéine et la date de soumission à PDB.
2. Ajouter à ce programme, la possibilité de générer un fichier de sortie au format FASTA.

NB :

– Dans le fichier PDB, la séquence protéique est codée en 3 lettres alors que le format FASTA code la séquence protéique en 1 lettre (utilisation d'un tableau associatif).

– La protéine est composée de 2 chaînes protéiques différentes, la chaîne A et B :

```
SEQRES 16 A 209 ASN TYR PRO LEU LYS PHE ASN GLN GLN ASN SER
SEQRES 1 B 557 SER ASN MET TRP VAL ILE GLY LYS SER LYS ALA
```

Le fichier FASTA sera de la forme :

```
>ANTIBIOTIC RESISTANCE A | 1AJQ
EQSSSEILIVRDEYGMPHIYANDTWHLFYGYGYV...
>ANTIBIOTIC RESISTANCE B | 1AJQ
SNMWVIGLSLAQDALAIMVNGPQFGWYAPAYTY...
```

Exercice 3 : BioPerl

1. Ecrire un programme BioPerl (fichier nommé `exo3a.pl`) réalisant les tâches suivantes :
 - (a) demande d'un identifiant de séquence à l'utilisateur
 - (b) récupération de la séquence au format genbank
 - (c) affichages des 20 premières bases de la séquence
 - (d) calcul du complément de cette sous-séquence
 - (e) écriture du résultat dans un fichier de sortie au format fasta (fichier nommé `identifiant.fasta`)
2. Ecrire un programme BioPerl (fichier nommé `exo3b.pl`) prenant en entrée un nom de répertoire contenant des fichiers au format genbank, et qui, pour chacun de ces fichiers, affiche les informations suivantes :
 - (a) l'identifiant de la séquence
 - (b) la longueur de la séquence
 - (c) le complément de la séquence
 - (d) la caractéristique 'exon' (si elle existe)

NB : vous testerez votre programme en créant un répertoire de test contenant 3 fichiers au format genbank.