

ANT3814 Examen final partie 1: Unix et R

L'examen final pour ANT3814 se déroulera en deux parties. La première partie (lundi) couvrira la ligne de commande Unix et R (leçons 1-5), et la deuxième partie (mercredi) se concentrera sur les sujets de bioinformatique que nous avons appris en classe (leçons 6-10).

C'est un examen "à livre ouvert", vous êtes donc libre d'utiliser le matériel dans les diapositives et les recherches Google pour répondre à ces questions.

Les deux parties de l'examen doivent être remises avant la fin de cette semaine (dimanche 23 juin à 23:59). Veuillez télécharger vos feuilles de réponses dans les dossiers correspondants sur StudiUM.

Section 1: Bases de l'Opération en Ligne de Commande

Question 1 :

Si vous travaillez dans votre répertoire home (~), comment listez-vous tous les fichiers et répertoires qu'il contient, y compris les fichiers cachés ?

Question 2 :

Quelle commande utiliseriez-vous pour copier un fichier nommé `data.txt` dans un répertoire nommé `backup` ?

Question 3 :

Comment pouvez-vous afficher le contenu d'un fichier nommé `results.txt` page par page ?

Question 4 :

Écrivez la commande pour supprimer un répertoire nommé `old_files` et tout son contenu.

Question 5 :

Comment recherchez-vous un mot-clé spécifique, "bioinformatique", dans un fichier nommé `research.txt` ?

Question 6 :

Expliquez la différence entre `>` et `>>` lors de la redirection de la sortie vers un fichier.

Question 7 :

Comment afficheriez-vous les 15 premières lignes d'un fichier nommé `sample.txt` ?

Question 8 :

Quel est l'objectif de la commande `man` dans Unix ?

Question 9 :

Écrivez la commande pour déplacer un fichier nommé `results.txt` du répertoire courant vers un répertoire nommé `output` situé dans le même répertoire.

Question 10 :

Quelle commande utiliseriez-vous pour trouver le répertoire de travail actuel ?

Section 2: Variables Bash et `awk`

Question 11 :

Comment déclarez-vous une variable nommée `count` et lui assignez-vous la valeur `10` en Bash ?

Question 12 :

Écrivez une commande utilisant `awk` pour imprimer la deuxième colonne d'un fichier nommé `data.csv`.

Question 13 :

Comment ajoutez-vous 5 à une variable nommée `num` en Bash ?

Question 14 :

En bash, que signifient les variables environnementales `$0`, `$1` et `$2` ?

Question 15 :

Quelle est la différence entre les guillemets simples et les guillemets doubles en bash ?

Question 16 :

Si vous avez deux fichiers contenant de nombreuses séquences d'ADN sur des lignes séparées, quelle commande utiliseriez-vous pour imprimer uniquement les lignes uniques ?

Question 17 :

Comment concaténez-vous les valeurs de deux variables, `var1` et `var2`, et stockez-vous le résultat dans une nouvelle variable `var3` ?

Section 3: Expressions Régulières

Question 18 :

Écrivez une expression régulière pour correspondre à une adresse e-mail.

Question 19 :

Comment utiliseriez-vous `sed` pour remplacer toutes les occurrences du mot "apple" par "orange" dans un fichier nommé `fruits.txt` ?

Question 20 :

Que fait la commande `grep` suivante ?

```
grep -v 'ATG' sequences.txt
```

Question 21 :

Décrivez la fonction du `.` (point) dans les expressions régulières.

Question 22 :

Que représente le symbole `^` dans une expression régulière ?

Question 23 :

Écrivez une commande Perl pour substituer toutes les instances de ATG par TAC dans la ligne de séquence suivante ATGCGAAGGATGAAG.

Question 24 :

Si vous avez un fichier d'appel de variantes appelé `sample.vcf`, comment n'imprimeriez-vous que les noms des colonnes et les 10 premières lignes des SNPs ?

Section 4: Boucles `for`, Boucles `while`

Question 25 :

Écrivez une boucle `for` pour itérer sur tous les fichiers `.txt` dans le répertoire courant et imprimer leurs noms.

Question 26 :

Écrivez une boucle `for` qui imprime les nombres de 1 à 5.

Question 27 :

Comment utiliseriez-vous une boucle `while` pour lire les lignes d'un fichier nommé `input.txt` et les imprimer ?

Question 28 :

Comment utiliseriez-vous une boucle `while` pour imprimer chaque ligne d'un fichier et mettre le numéro de ligne et un deux-points devant elle ? Par exemple, si les 5 premières lignes d'un fichier appelé `sequences.txt` sont :

```
GGGAGAGATG
GCTGATGCGCAGAG
AATTAGGATGGAAT
AATATAGGAG
ATAGGAGAAT
```

imprimez le texte suivant :

```
1: GGGAGAGATG
2: GCTGATGCGCAGAG
3: AATTAGGATGGAAT
4: AATATAGGAG
5: ATAGGAGAAT
```

Question 29 :

Quel est le résultat du script suivant ? Les 5 premières lignes du fichier `metadata.txt` sont :

```
primates
carnivores
cétacés
félidés
canidés
```

```
#!/bin/bash

filename='metadata.txt'
while read line; do
    alignment=$(echo $line | awk '{print $1}')
    echo "iqtree2 -s ${alignment}.aligned.gb.fasta -m MFP -b 1000"
done < $filename
```

Question 30 :

Utilisez une boucle `for` pour imprimer "est un étudiant dans ANT3814" après chaque nom dans la liste de noms suivante : Charlotte Jimmy Mikerlange Sarah

Question 31 :

Écrivez une boucle `for` pour itérer sur une liste de répertoires (`dir1`, `dir2`, `dir3`) et créer un fichier nommé `readme.txt` dans chacun d'eux. Astuce : lisez à propos de la commande bash `touch` pour créer un nouveau fichier vide.

Question 32 :

Si vous avez un répertoire avec 50 fichiers où vous avez accidentellement enregistré des noms comme `sample1.txx`, `sample2.txx`, `sample50.txx`, écrivez une boucle pour renommer tous les fichiers afin qu'ils se terminent par `.txt` au lieu de `.txx`.

Section 5: R et ggplot2

Question 33 :

Comment installez-vous le package `ggplot2` en R ? Écrivez la commande.

Question 34 :

Expliquez comment vous pouvez charger le package `ggplot2` après l'avoir installé.

Question 35 :

Écrivez la commande R pour créer un simple vecteur contenant les nombres de 1 à 5.

Question 36 :

Comment créez-vous un tableau de données en R avec deux colonnes : `gene` (contenant les noms des gènes "gene1", "gene2", "gene3") et `expression` (contenant les valeurs 10, 20, 30) ?

Question 37 :

Quelle est la commande pour extraire la 6ème ligne d'un tableau de données nommé `data` ?

Question 38 :

Écrivez la commande `ggplot2` pour créer un diagramme de dispersion avec `ggplot(data, aes(x = variable1, y = variable2))`. Incluez la fonction `geom` nécessaire pour tracer les points.

Question 39 :

Comment pouvez-vous ajouter un titre "Gene Expression Plot" à un diagramme de dispersion `ggplot` ?

Question 40 :

Décrivez comment vous pouvez changer l'étiquette de l'axe x en "Gene Name" et l'étiquette de l'axe y en "Expression Level" dans un diagramme `ggplot`.

Question 41 :

Écrivez la commande R pour filtrer les lignes dans un tableau de données `df` où la colonne `expression` est supérieure à 15.