

Feuille d'exercices n° 4

Scripts BASH / Python

Exercice 4.1

On souhaite afficher à l'écran les chaînes de bases de fichiers [FASTQ](#) en colorisant chaque base en fonction de son [score phred](#) Q :

$Q < 30$	$30 \leq Q < 60$	$60 \leq Q$
----------	------------------	-------------

1. En utilisant le module Python [textcolors.py](#) disponible sur madoc, écrire le module `colorize_fastq.py` prenant en entrée un nom de fichier au format FASTQ et affichant les séquences de bases qui s'y trouvent avec les couleurs correspondant à leurs *scores phred*. On pourra tester le module avec le fichier [example.fq](#) disponible sur madoc ;
2. Écrire un script BASH `fastq_color.sh` prenant en entrée un nom de répertoire et appelant le module `colorize_fastq.py` sur les dix fichiers FASTQ de tailles les plus petites se trouvant dans ce répertoire ou ses sous-répertoires. Pour vérifier qu'un fichier est au format FASTQ, on regardera si la 3^e ligne du fichier commence par un caractère « + ». On affichera un message d'erreur et l'on retournera le code d'erreur 1 si le script est appelé avec le mauvais nombre de paramètres. On pourra tester le script sur l'arborescence contenue dans l'archive [paeruginosa-reads.tar.gz](#) sur madoc.

Exercice 4.2

1. Écrire un programme `trim_fastq.py` prenant en entrée un nom de fichier au format FASTQ, nettoyant l'extrémité droite de chaque séquence de la plus longue séquence dont les bases ont toutes un score phred strictement inférieur à 30 et affichant sur la sortie les séquences nettoyées au format FASTA ;
2. Écrire le programme `trim_fastq.sh` prenant en entrée une liste de noms de fichiers au format FASTQ et sauvegardant chaque fichier au format FASTA sous le même nom que le fichier FASTQ de départ, en remplaçant l'extension initiale par `.fasta` (exemple : `SRR396636.fq` devient `SRR396636.fasta`). Au passage, on aura nettoyé les séquences avec le programme `trim_fastq.py`. Le programme devra afficher un message d'aide s'il est appelé sans paramètre ; il devra aussi s'assurer que le nom du fichier FASTA créé n'existe pas déjà et demandera au besoin un nouveau nom de fichier à l'utilisateur. Enfin, on s'assurera que chaque fichier traité est bien au format FASTQ ; si ce n'est pas le cas, on affichera un message d'avertissement et on passera au fichier suivant.