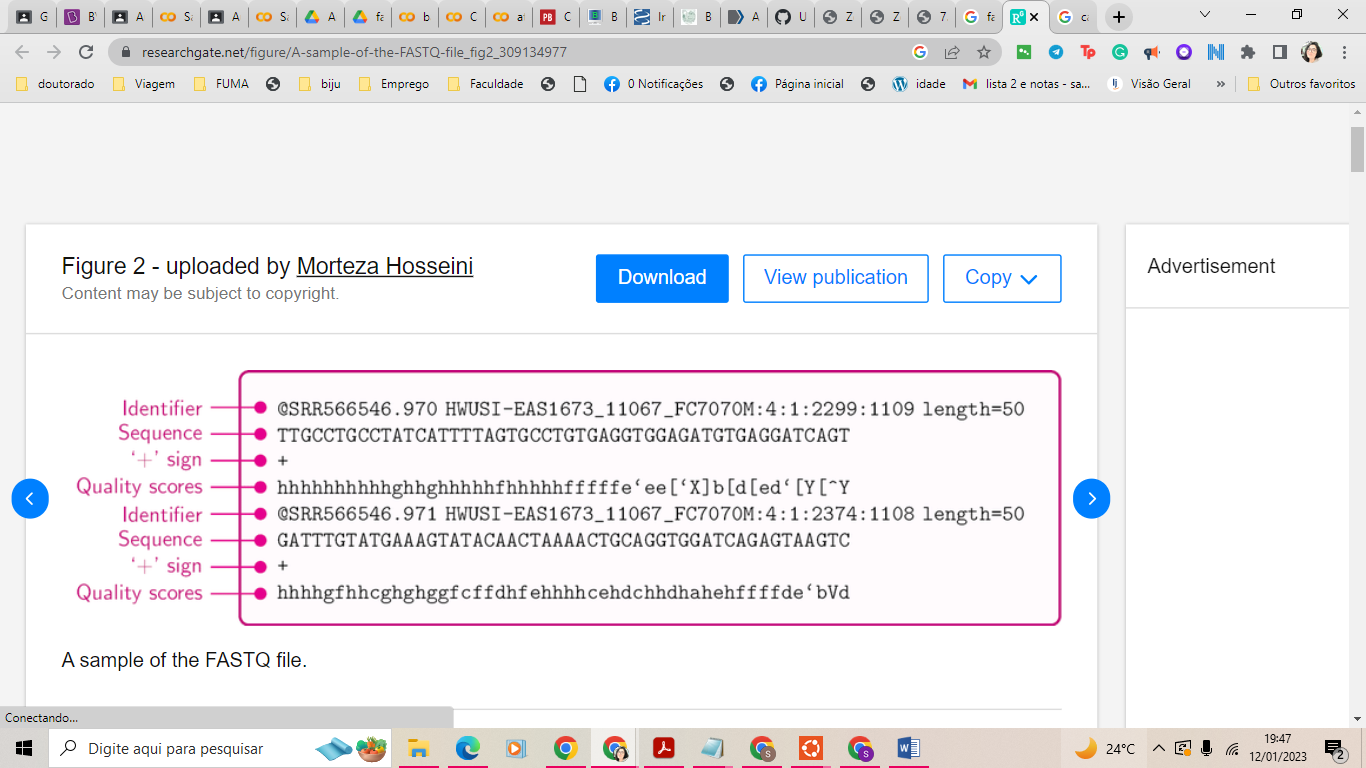
1. O que é um arquivo do tipo FASTQ e como posso verificar se um FASTQ é válido?

FASTQ é um arquivo projetado para lidar com a saída de métricas de qualidade de máquinas de sequenciamento. Nesse formato, as pontuações de sequência e qualidade são representadas como caracteres ASCII simples (escrita baseada no alfabeto romano cuja função é padronizar a forma como os computadores representam letras, números, acentos, sinais diversos e alguns códigos de controle).

O formato usa quatro linhas para cada sequência, e essas quatro linhas são empilhadas umas sobre as outras em arquivos de texto gerados por fluxos de trabalho de sequenciamento. Cada uma das 4 linhas representará uma leitura.

Segue a imagem de um arquivo fastq.

A linha 1 começa com o caractere ‘@’ e é seguida por um identificador de sequência e uma descrição opcional. Essa linha é utilizada pela tecnologia de sequenciamento e geralmente contém informações específicas para a tecnologia. Ele pode conter IDs de lâmina de fluxo, números de pista e informações sobre pares de leitura.

A linha 2 é a sequência de letras (nucleotídeo)

A linha 3 começa com um caractere ‘+’; ele marca o fim da sequência e é opcionalmente seguido pelo mesmo identificador de sequência novamente na linha 1.

A linha 4 codifica os valores de qualidade para a sequência na linha 2 e deve conter o mesmo número de símbolos como letras na sequência. Cada letra corresponde a um índice de qualidade. Embora possa haver diferentes definições dos índices de qualidade, um padrão de fato no campo é usar “Phred Quality Scores”. Essas pontuações representam a probabilidade de a base ser considerada errada.

Para se verificar se um arquivo fastq é válido podemos utilizar o software fastq

Instalação

$ sudo apt install fastqc

Em seguida utilizar o código abaixo:

$ fastqc <fastq\_file> --extract -o <output\_file>

Segue abaixo a imagem de um arquivo de validação de um arquivo fastq

