**Teste Admissional**

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

O FASTQ é um formato de arquivo de texto. Nesse formato, os dados (geralmente sequências de nucleotídeos) são organizados em:

@cabeçalho (@ seguido de identificadores)

ATTCATCTATCATCA (Sequência)

**+** (Separador. Obs.: Pode vir acompanhado da repetição do cabeçalho)

Dados de qualidade de cada base (codificados em tabela ASCII)

Para a avaliação da validade de um arquivo FastQ é possível utilizar várias ferramentas como FastQValidator ou ValidateFastQ.

**2. Quais são as etapas mais comuns de um pipeline de bioinformática de NGS para DNAseq e quais ferramentas podem ser usadas em cada etapa?**

Verificação da validade dos dados FASTQ: **(FastQValidator; ValidateFastQ)**

Verificação da qualidade dos dados FASTQ: **(Fastqc; MultiQC)**

Trimagem dos reads e remoção dos adaptadores: **(Trimmomatic)**

Mapeamento dos reads em genoma referência (no caso de montagem com genoma referência): **(BWA, Bowtie)**

Montagem dos reads: **(SPAdes, Velvet)**

Validação da montagem: **(QUAST)**

Predição de regiões codantes: **(ORF Finder)**

Anotação dos dados: **(Maker2)**

**3. O que é o dogma central da biologia e como a bioinformática pode ser aplicada em cada um dos processos?**

DNA é transcrito em RNA que é traduzido em proteínas. A bioinformática tem sido essencial em várias áreas de genômica, transcriptômica e proteômica. Em dados genômicos, a bioinformática pode ser utilizada para avaliar a presença de genes em organismos/cepas específicos, variações (SNV, inserções e deleções) em regiões conhecidas em genes de interesse, predição de regiões codantes, análise filogenética de organismos/cepas, predição do impacto funcional de variantes e comparação de genes homólogos. A partir de dados transcriptômicos, a bioinformática pode ser aplicada para avaliar diferentes níveis de expressão de genes de interesse em determinadas condições e realizar análises de enriquecimento de vias metabólicas. Já com dados proteicos, por meio de bioinformática estrutural é possível predizer estruturas de proteínas e predizer sua função. Além disso, pode-se estudar estruturalmente o impacto de modificações de resíduos de aminoácido, realizar o desenho de fármacos para proteínas específicas por meio de *docking* molecular, estudar a comportamento de proteínas em ambientes miméticos.

**4. Considerando que o alinhamento de sequências biológicas é uma das atividades mais recorrentes e importantes na área de bioinformática. Comente sobre os algoritmos computacionais de alinhamentos mais utilizados e em que cenários uns são mais indicados do que outros.**

Algoritmos de alinhamento de sequência têm sido essenciais na bioinformática desde seus primórdios. Nesse sentido, ferramentas como Blast e ClustalW revolucionaram a análise de dados biológicos. Nesse cenário, ferramentas de alinhamento local, como o Blast e o Fasta, são aplicados buscando regiões com maior semelhança entre pares de sequências. Já ferramentas de alinhamento global, como ClustalW, Muscle e MAFFT, alinham sequências inteiras, de ponta a ponta. Alinhamentos locais são indicados para encontrar padrões conservados dentro de sequências, enquanto alinhamentos globais são usados para comparar sequências de genes ou proteínas homólogos ou de função similar.