

AULA PRÁTICA GENÔMICA – Explorando a WEB 1- BMC

Parte 1 (40 minutos)- Em sala de aula (em grupo de duas pessoas- Não precisa entregar):

1- Com o objetivo de simular um sequenciamento e montagem de sequências de DNA, você receberá uma série de sequências de uma porção de genoma e deve alinhar e montar um contig. Notar que, para facilitar o trabalho, todas as sequências estão representadas apenas no mesmo sentido (5'-3', em relação ao gene). Use fitas adesivas para facilitar o seu trabalho.

2- Após a montagem, identificar a janela aberta de leitura mais provável (“open read frame”- ORF). Quantas possíveis fases de janelas abertas de leitura nós temos nessa sequência? Por que?

3- responda: porque em uma sequência de DNA dupla fita esperamos ter 6 janelas abertas de leitura?

Parte 2 (1 hora)- RESPONDA com a ajuda do computador (em grupo de duas pessoas) (devem entregar apenas uma folha de respostas, colocando o nome dos dois nos exercícios, que devem ser entregues no final da aula).

Baseado na sequencia abaixo no formato FASTA:

```
>sequência a ser analisada
ATGACCGCGCGGGATTTCCCTGGAATTCCTGAGCGACTGGCTCCAGCCAGCGGTTTCCAGTCGGCCCAAT
TGCGTGAAATCGAGATCCTGTTGGGCCTGGAGGACACGCAGCGTATTTTCGATCGGCAACGGGTGCTCGTA
CCGGGATGCACTCAAATTGCCCGGTGGCGCGCCGTCATCGTCGGCGCAACCGCTCGAAGCACGTGCGGCC
GATGGGCAAGCTTCAAGCATTGTCTGTACGACTGGCTGTCACGCGTGCCCATTTGACGGCTCGAATGCGC
CAGCGGATATCACGCGGTTTCTTGAGGCGTACCTGGCCTCGCTGCGTGCGGAGAATGATCGCCGACTCAA
GCTGGCAAGCAGTGGCCTGGCTGCGGCAGAGATCGAGCAGTTGAGGGCGCGTTATGCGGCAGAAGACGCT
GGCGCACAGGCATTTCTATTGGCCGAAGACGATCCGCAGGCCAGCGATGCGGAGCGTGAAACGCGGCGTG
CGGGACGTGCGGCGATGCTCTTTATAGAGAGCTACCGCGAATTGCCGCAGCTTGCATGGCCGCATGCGTT
GCTCGAAAGCGTGCTTTGAGCTGGAGCAGGCCATGCTGATCTGGCGACAGCGGCATGCCCGCATGGTGGAG
CGCTTCATCGGGCGCCGCATAGGCACAGGGGGTTCCAGCGGCGTCTGACTATCTGGATCAGACCGCGCGCT
ACCGCATTTTACCGAGCTGTGGACAGTGCCTTCCCTGTTGCTGCGCAAGGCTGCCGTGCCCCAGATCCG
ATCGAGTGCGGGCTACGGATTTCGTGATGGAGGGCTTTGCTTGA
```

1. Abra o computador no site BLAST: <https://blast.ncbi.nlm.nih.gov/Blast.cgi> .
Teste a sequência acima no BLASTN (nucleotídeo-nucleotídeo) e BLASTX (nucleotídeo- proteína) e responda:

- a) O que é formato FASTA?
- b) O que você encontrou com BLASTN e com BLASTX? Qual a razão da diferença?
- c) Defina também BLASTP, tBLASTN, tBLASTX
- d) Refaça o BLASTX mas antes de pedir o BLAST coloque no DATABASE apenas MODEL organisms. Qual o resultado agora?
 - i) Veja também o resultado de gráfico. Por que as cores são diferentes, o que elas representam?
 - ii) Qual o organismo modelo principal do alinhamento dessa proteína?
 - iii) Como um gene bacteriano alinhou com um gene de eucarionte?
- e) O que indica a ordem de alinhamento do resultado do BLAST?
- f) Verifique nesse resultado o valor de e-value. Como ele varia em relação aos dois BLASTX efetuados. O que significa Evalue?
- g) Qual é o provável organismo de cujo genoma foi obtida essa sequência?
- h) Qual é a provável anotação desse gene? (ou seja qual o nome do gene?)

2. Agora vá para a página do ORF FINDER-
<https://www.ncbi.nlm.nih.gov/orffinder/>

faça um copiar-colar de toda a sequência e aplique, no formato FASTA, e veja o resultado.

RESPONDA

- a) Quantas ORFs o programa encontra? Por que temos 6 fases abertas de leitura?
- b) Quais é a provável ORF correta, qual frame? Por que?
- c) Para essa ORF faça o BLASTP e SmartBLAST (já está oferecido ao usuário no final da página). Avalie os resultados e o que você encontra de interessante (evolutivamente) nesse alinhamento encontrado.