

Tarefa 1 Bioinformática 2011/2

Entregue o relatório para rafaelmguedes@yahoo.com.br utilizando como assunto TAREFA_1.

1. Logar na sua conta na maracatu.icb.ufmg.br
(Caso não lembre a senha, escreva para o e-mail acima)
2. Criar diretório tarefa1 e dentro de tarefa1 criar diretórios chromat_dir e edit_dir

```
mkdir tarefa1  
cd tarefa1  
mkdir chromat_dir  
mkdir edit_dir
```

3. Entrar em chromat_dir e copiar os cromatogramas correspondentes ao número que lhe foi atribuído em <http://cromatina.icb.ufmg.br/tarefa/> (**substitua <1> no comando abaixo pelo seu número**)

```
cp /home/bacharelado/ROTEIRO/1/* .
```

4. “Da um ls dá dá um ls” para confirmar a transferência, depois suba um diretório e entre em edit_dir

```
cd ..  
cd edit_dir
```

5. Rode o programa PHRED trimando as seqüências de baixa qualidade

```
phred -id ../chromat_dir -trim_alt "" -trim_cutoff 0.032 -trim_out -sa saida -qa saida.qual
```

6. Copie para o relatório uma das seqüências FASTA e seu correspondente arquivo de QUALIDADE e explique o significado dos números no cabeçalho e no arquivo .qual.
7. Elimine as seqüências de vetor com seqclean. Uma biblioteca de vetores denominada univec já está formatada na maracatu. Escolha e mostre no relatório uma seqüência, antes e depois de retirar o vetor.

```
seqclean saida -v /home/bacharelado/phred_aula/vetor/univec -o semvetor
```

8. Conte quantas seqüências foram obtidas no arquivo semvetor e reporte no relatório

```
cat semvetor | grep ">" -c
```

9. Forme consensos com cap3. Pesquise e relate brevemente o papel do programa cap3.

cap3 semvetor

10. Analise o resultado contando, como acima, quantos contigs (semvetor.cap.contigs) foram formados e quantos singlets (semvetor.cap.singlets) e reporte no relatório

11. Descubra qual o BLAST correto para anotar seqüências de nucleotídeos utilizando um database proteico (UniProt) e anote os contigs localmente. (substitua <?> pelo BLAST correto, Ex: blastn)

**blastall -p ? -i semvetor.cap.contigs -d
/home/bacharelado/blast_aula/uniprot/uniprotdb -e 1e-15 -a 4 -F F -o blast_contigs &**

Acompanhe o andamento do processo com o comando top. Pode levar alguns minutos

top

Para sair do top, aperte a tecla q

12. Reporte no relatório (não é necessário colocar todos os hits de cada sequencia) o resultado do BLAST e COMENTE (por exemplo, descrição do hit, e-value, se foi significativo)

13. Usando copiar e colar, repita a análise com BLASTx online no NCBI. Compare e discuta os resultados no relatório.

14. Por fim, responda as seguintes perguntas:

a. Ao rodar o programa PHRED, você utilizou o parâmetro `-trim_cutoff 0.032`. O que você esperaria obter em relação ao : (1) tamanho das sequencias; (2) quantidade de contigs; caso utilizasse `-trim_cutoff 0.01`?
Quais valores de PHRED correspondem as porcentagens de erro 0.032 e 0.01 citadas acima? Pesquise a fórmula que converte % de erro a valor de qualidade.

b. Rode o item 5 novamente sem o comando `-trim_out` e explique a diferença nos arquivos de saída.

**phred -id ../chromat_dir -trim_alt "" -trim_cutoff 0.032 -sa saida2 -qa
saida2.qual**

c. Discuta o que seria observado no item 11 caso o valor de e-value fosse alterado para `-e 1e-3`.

- d. Você possui uma seqüência nucleotídica desconhecida. Para tentar aferir alguma similaridade com seqüências já conhecidas faz um BLASTn no NCBI. Fica então muito surpreso ao verificar que não ha hits significativos. Tenta então fazer um BLASTx e vê uma lista de possíveis homólogos! Com base em seus conhecimentos biológicos, discuta a diferença de resultados.