

Trabalho prático 1 - Bioinformática 2011/Primeiro semestre

Entregue o relatório para rafaelmguedes@yahoo.com.br até 10:00h de 18/06 utilizando como “assunto” do e-mail “TAREFA1”

CUIDADO se copiar e colar os comandos abaixo no terminal, pois alguns caracteres podem ser alterados.

- Logar na sua conta na maracatu.icb.ufmg.br
(Caso não lembre a senha, escreva para o e-mail acima)
- Criar diretório tarefa1 e dentro de tarefa1 criar diretórios chromat_dir e edit_dir
mkdir tarefa1
cd tarefa1
mkdir chromat_dir
mkdir edit_dir
- Entrar em chromat_dir e copiar os cromatogramas correspondentes ao número que lhe foi no atribuído em <http://cromatina.icb.ufmg.br/tarefa1/chamada2011.htm>.
(substitua <X> no comando abaixo pelo seu número)
cp /home/bacharelado/ROTEIRO/X/* . (inclua o ponto separado por espaço)
- “Dê um ls” para confirmar a transferência, suba um diretório e entre em edit_dir
cd ..
cd edit_dir
- 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
- Copie para o **relatório** uma das seqüências FASTA e seu correspondente arquivo de QUALIDADE e comente o resultado obtido. Para referência, veja http://cromatina.icb.ufmg.br/bacharelado/phred_bioinfo.htm.
- Elimine as seqüências de vetor com seqclean. Uma biblioteca de vetores denominada univec já está formatada na maracatu.
seqclean saida -v /home/bacharelado/phred_aula/vetor/univec -o semvetor
- Conte quantas seqüências foram obtidas no arquivo semvetor e reporte no **relatório**
cat semvetor | grep ">" -c
- Forme consensos com cap3
cap3 semvetor

- Analise o resultado contando, como acima, quantos contigs (semvetor.cap.contigs) foram formados e quantos singlets (semvetor.cap.singlets) e reporte no **relatório**
- Pesquise qual programa BLAST usar para anotar seqüências de nucleotídeos utilizando uma database protéica (UniProt) e anote os contigs (substitua <?> pelo BLAST correto, ex: blastn)

```
blastall -p ? -i semvetor.cap.contigs
-d /home/bacharelado/blast_aula/uniprot_5_2011/uniprot_17_5_11
-e 1e-10 -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)

Repita a análise com os singlets.

- Reporte no **relatório** (não é necessário colocar todos os hits, somente a anotação mais informativa a julgar pela análise dos melhores hits) e COMENTE (e-value obtido, etc.)
- Usando copiar do terminal e colar no site do BLAST na web, repita a análise com BLASTx online no NCBI. Compare e discuta os resultados no **relatório**.
- **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 e (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?
 - 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 BLAST local contra o UniProt 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, discuta a diferença de resultados.