# **CEN0336 - Introdução a Programação de Computadores Aplicada a Ciências Biológicas**

# **Prova II**

Nome: Miguel Ângelo Cyrillo Narbot

No USP: 11243516

## **Instruções Gerais**

Vocês terão desde o dia 3 de dezembro as 15:00 e até o dia 15 de dezembro as 23:59 para resolver as atividades da prova.

São 4 perguntas/atividades, cada uma precisa de uma resposta neste arquivo de word, que tem que ser carregado no sistema e-disciplinas até o dia e horário limite. Além disso algumas perguntas precisam que vocês desenvolvam scripts e carregam num repositório github especifico para a prova (indicado na pergunta 1). No dia 16 de dezembro as 00:01 eu descarregarei todos os repositórios para avaliar seus scripts. Os critérios de avaliação dos scripts são:

Eles têm que rodar sem erros de sintaxe (30%)

Eles têm que ser logicamente corretos (40%)

Eles têm que ter comentários explicando o que é feito em cada passo (20%)

Eles têm que ter controle da entrada dos usuários (10%)

## **Perguntas/Atividades**

1. Crie um repositório no github para esta prova. Escreva o endereço completo do seu repositório github como resposta desta pergunta.

Lembre-se que o repositório tem que ter o nome segundo o padrão:

P2\_NNNNNNN-CEN036

Onde NNNNNNN é seu número USP.

A resposta tem que ter o endereço completo, iniciando com https://"

O professor irá clonar seu repositório, com as respostas para os outros pontos da prova.

**R: https://github.com/MiguelNarbot/P2\_11243516-CEN036.git**

1. Crie um script de python com o nome script\_getORF.py que realize as seguintes operações:
2. Ler um arquivo multifasta da linha de comandos, usando o modulo sys.
3. Para cada registro fasta identificar o Open Reading Frame (Fase de Leitura Aberta: <https://pt.wikipedia.org/wiki/Fase_de_leitura_aberta> ou <https://en.wikipedia.org/wiki/Open_reading_frame>) mais comprido dentre as 6 fases de leitura possíveis. Os ORFs neste caso não têm introns. Repare que uma fase de leitura aberta tem que iniciar com um codon de início da tradução e terminar com um códon de terminação da tradução (<https://pt.wikipedia.org/wiki/C%C3%B3digo_gen%C3%A9tico>), é seu comprimento tem que ser múltiplo de três (3)
4. Daquele ORF mais comprido identificado, gerar o peptídeo codificado.
5. Imprimir o ORF mais comprido no arquivo ORF.fna. O formato da sequência tem que ser fasta, e seu identificador, tem que ser o identificador original, adicionando o string: “\_frameX\_START\_END”, onde X é a Fase de leitura (1,2,3,4,5 ou 6), e START e END são as coordenadas na sequencia
6. Imprimir o peptídeo correspondente ao ORF mais comprido no arquivo ORF.faa. O formato da sequência tem que ser fasta, e seu identificador, tem que ser o identificador original, adicionando o string: “\_frameX\_START\_END”, onde X é a Fase de leitura (1,2,3,4,5 ou 6), e START e END são as coordenadas na sequencia
7. Não esquece que tem que processar todos os registros do arquivo fasta original, i.e., tem que ter o mesmo número de entradas nos arquivos ORF.faa, ORF.fna que no arquivo original.
8. Carregue seu script, e os arquivos ORF.faa e ORF.fna no seu repositório.
9. Calcule o md5sum do seu script e coloque esse número como resposta desta pergunta.
10. Explique a diferença entre os dois sets:

set1=set('accggttaata')  
set2={'accggttaata'}

**R: A diferença entre esses dois exemplos está em que no primeiro caso (set1), a saída é única, ou seja, ela expõe todos os valores (strings) de forma única, {‘g’, ‘a’, ‘c’, ‘t’}, separada. No segundo caso, (set2), a saída são todas as strings presentes no dicionário de forma conjunta e não separada, {‘accggttaata’}. O “dicionário” e o “set” são formas de armazenamento de dados.**

1. Descreva com sus próprias palavras o que o seguinte pseudocódigo faz. Faz uma avaliação desse pseudocódigo e relate quaisquer problemas que possa ter.

Defina TOTAL com zero #total é que não tem nenhum número somado ainda

Defina CONTADOR NOTAS como zero #inicia-se sem notas

While CONTADOR NOTAS é menor ou igual a dez

      Entre (input) a seguinte nota

      Some a nota ao TOTAL

Defina a média da disciplina como o TOTAL dividido por dez (TOTAL/10)

Imprima na tela a média da disciplina

**R: O problema inicial está em função “while”, que faz com que seja feito um loop infinito, sem parada, com as notas adicionadas no input. Outro ponto, é se for colocado um valor float, flutuantes (reais) - exemplo de 2.5 - é dado um erro, o que impede, portanto, uma adição de notas, valores decimais. Na versão exercício 4.txt (github) foi adicionado o código, mas pode ser avaliado abaixo.**

**total= 0**

**contador\_de\_notas = 0**

**while contador\_de\_notas <= 10:**

**nota = int(input('coloque sua nota aqui:'))**

**somatória = nota + total**

**MÉDIA\_DA\_DISCIPLINA = (total/10)**

**print(MÉDIA\_DA\_DISCIPLINA)**

Implemente esse pseudocódigo num script de python e carregue ele no seu repositório da prova com o nome notas.py. Calcule o md5sum do seu script e coloque esse número também como resposta desta pergunta.

**R: calculo do md5sum = 7ed118f4abe2699ef5d694567d3241d0**