

Exercício para entender a complexidade de sequências de DNA e a ação de enzimas de restrição

As sequências de DNA são representadas por letras que correspondem às bases nitrogenadas (**A, G, T, C**). O primeiro formato padronizado de representar sequências de DNA e proteínas eram utilizados pelo programa denominado de FASTA. Neste formato **FASTA**, a primeira linha se inicia como um sinal ">" seguido da identificação e/ou descrição da sequência, que é então apresentada na linha subsequente, utilizando até 70 caracteres por linha, e a sequência se inicia na próxima linha. O mesmo formato é utilizado para sequências de proteínas e peptídeos, empregando os símbolos de uma letra para cada um dos 20 aminoácidos.

Neste exercício, gostaríamos que se familiarizasse com o formato FASTA da sequência abaixo. Trata-se da sequência de um mRNA real de cacaueteiro (*Theobroma cacao*) (expresso em bases de DNA) que codifica uma proteína muito abundante em sementes (proteína de semente com tamanho de 21 KiloDaltons). A sequência da região codificadora do gene possui 666 bases (nucleotídeos), mas a sequência abaixo possui 984 nucleotídeos.

Considerando que toda sequência codificadora de proteína se inicia com um **códon ATG** e termina com um códon de parada (*stop*), que podem ser **TAG, TAA** ou **TGA**, identifique o início e o final da região codificadora. **Informe que essa sequência não possui introns! Marque no texto estes pontos!**

Lembro que a sequência deve ser lida de três em três bases (equivalentes aos códons)!

Cada linha da sequência possui 70 bases!

Use o site <https://www.ncbi.nlm.nih.gov/orffinder/>

```
>NM_001404993.1 Theobroma cacao 21 kDa seed protein (LOC18604292), mRNA
GGCTTCCAGCCTCTATAAATTGAAGTATCAAAGCAAGTAACCGGCATCCAAATCAACACTTATCCAGCAA
CATTTCACTTAACCATGAAGACCGCAACAGCCGTAGTTTTACTCCTCTTCGCCCTTCACATCAAAATCATA
TTTTCTTTGGGGTAGCCAACGCTGCAAACCTCTCCTGTGCTTGACACTGATGGTGATGAGCTCCAAACTGGG
GTTCAATATTACGTCTTGTTCATCGATATCGGGTGTGGGGTGGAGGGCTAGCCCTAGGAAGGGCTACAG
GTCAAAGCTGCCAGAAATTGTTGTCCAAAGACGATCCGACCTTGACAATGGTACTCCTGTAATCTTTTC
AAATGCGGATAGCAAAGATGATGTTGTCCGCGTATCTACTGATGTAACATAGAGTTCGTTCCCATCAGA
GACAGACTCTGCTCAACGTCAACTGTGTGGAGGCTTGACAATTATGACAACCTCGGCAGGCAAATGGTGGG
TGACAACCTGATGGGGTTAAAGGTGAACCTGGTCCTAACACTTTGTGCAGTTGGTTTAAAGATTGAGAAGGC
CGGAGTACTCGGTTACAAATTCAGGTTCTGTCTTCTGTCTGTGATTCGTGCACAACCTTATGCAGCGAT
ATTGGAAGACATTCAGATGATGATGGACAAATACGTTTGGCTCTCAGTGACAATGAATGGGCATGGATGT
TTAAGAAAGCAAGTAAGACAATAAAACAAGTTGTTAACGCGAAGCATTAAATTTAAGTTTAAATGTACGAA
GTGTACGTCCAAAGCAGCAATACTAGCCGGTTCGTTACTTTCCACTAAAATAAAAGTTAAGTATGTGGTTCC
CAGCCCAGTGTTGTAATGCTATGCCTATGTAGTCAGTGTCTTGTGAGGGTGGAGATGCTTAAAGGGTG
TGTCTTCACAGCCCAGCTTCGTAGTCTTTCAGCTTTATGAATAAATGATTTGCCTCTTGCCCTTTTTAAA
AAAA
```

Enzimas de restrição reconhecem sequências palindrômicas e cortam no meio dela. Utilizando a mesma sequência de DNA acima, simule ela sendo digerida (cortada) com enzimas de restrição listadas abaixo. Use o site <https://tools.neb.com/REBsites/>. Considerando as enzimas abaixo e seus respectivos sítios de reconhecimento corte, **quantos e qual o tamanho dos fragmentos resultantes** do uso de cada uma das enzimas?

Digerindo com *AgsI* - CG|CG

Digerindo com *EcoRV* - GAT|ATC

Digerindo com *EcoRI* – -G|AATTC

O site apresenta uma simulação de um gel de agarose com diversas concentrações. Experimente trocar e analise os resultados.