

EXERCICIO 2

Sobre busca em Bancos de Dados Biológicos

Utilize a sequência (ID) abaixo de um gene/proteína de acordo com a inicial do seu primeiro nome (primeira coluna).

A a E - *Zea mays* mRNA for phosphoenolpyruvate carboxylase - GenBank AJ536629.1

F a G - *Triticum aestivum* NADP-dependent malic enzyme – GenBank XM_044499722.1

H a L - *Zea mays* phosphoenolpyruvate carboxykinase – GenBank NM_001348550

M - *Zea mays* 6-phosphofructokinase 2 – GenBank NM_001154608

N a P – *Zea mays* triosephosphate isomerase, cytosolic – GenBank NM_001153743.1

R a T – *Zea mays* Malic enzyme, mRNA – GenBank NM_001196190

V – *Zea mays* zein mRNA, complete cds - GenBank M60837

Busque as sequencias no NCBI na página [Analyze](#). Conduza as análises seguintes:

1. Apresentar a sequência no formato FASTA para nucleotídeos e proteína
2. Utilize a análise de BLAST (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) para encontrar sequencias similares a aquela assinalada a você e responda as seguintes questões:
 - a) Qual a ferramenta de comparação que você utilizou?
 - b) Qual o tipo de sequência foi utilizada como *query*?
 - c) Qual o tamanho da sequência *query*?
 - d) Qual o tamanho da sequência *subject*?
 - e) Qual a proteína codificada pela sequência *query* (produto do gene)?
 - f) A que organismo pertence a sequência *query*?
 - g) Qual a % de identidade encontrada entre a sequência *query* e *subject*?

3. A partir da sequência abaixo, informe qual seria a sua melhor sugestão de identificação para este gene? De qual espécie ou organismos seria? Qual a certeza estatística que você pode afirmar sobre essa identificação? Use [BLAST](#)

GGCGCTGGACGACTACCTGGACCGGCGCGTGGACGACATGGTGC GCGAGGGCATGGTGGAGGAGCTGCGG
 GAGTACTTCGCGTCCACGTCCCCGCGGAGCGCGCCTCCCACGCGGCCGGGCTGGGCAGGGCCATCGGCG
 TGCCGGAGCTGGGCGCCACTTCGCGGGCCGCAAGACCCTCGCCGCGGCGATCGACGAGATCAAGGTCAA
 CACGCGGGTGCTGGCCGCGGCGCAGGTGCGCAAGATCCGCGCATGGCCGACGCCCTGGGGCTGGCCCGTG
 CGGCGTCTAGACGCGACGCCACGATCCGCGCGCGGGCTGGCCGGCGCTGGCTGCGCCGCCGAGGCCGCCG
 CGTGGGAGCGCGACGTGGGCGGGCCTGGCCTCGCCGCCATGCGCCAGTTCCTCGACTGCCACGAGCAGCA
 GCAGCAGCAAAGGTGGCCGCTGGTGC GGAGGCAATGCCGTGGCATGGCTTGGTGACTCGTGATGACGCGT
 CGAAGGTCCGGGAGCCGAGTTTTTTTTT

4. Considere os dois alinhamentos possíveis. Qual o melhor deles?

Considere: **ACGGACT** e **ATCGGATCT**

A	-	C	-	G	G	-	A	C	T
A	T	C	G	G	A	T	-	C	T

Alinhamento 1

A	-	C	G	G	-	A	C	T	
A	T	C	G	G	A	T	C	T	

Alinhamento 2

Considere o seguinte esquema de score das identidades entre bases das duas sequencias de DNA (superior e inferior):

- +2 para igualdade (match)
- 1 para desigualdade (mismatch)
- 1 para penalidades (gaps)