

Atividades

1. Baixar e rodar programa que implementa o algoritmo básico de comparação de sequências por programação dinâmica
 - a. Alunos devem realizar pelo menos 3 testes, inventando 3 pares de sequências, que devem ter entre 10 e 30 bp

<http://www.codeproject.com/Articles/304772/DNA-Sequence-Alignment-using-Dynamic-Programming-A>

2. Rodar BLAST

- a. Blastn
 - i. comparar os mesmos 3 pares de sequências de (1) com blastn para 2 sequências (*Align two or more sequences*)
 - ii. Escolha uma das sequências de (1) com 30 bp e faça uma busca com blastn
 - iii. faça uma busca genérica com a sequência A, que se encontra abaixo
 - iv. faça uma busca genérica com a sequência B, abaixo
- b. Blastp
 - i. faça uma busca com a sequência C, que se encontra abaixo, especificando que quer que a busca seja contra o organismo *Methanococcus jannaschi*
 - ii. repita (i), mas especificando *Escherichia coli*
 - iii. repita (i), sem especificar organismo
- c. Blastx
 - i. Execute blastx com a sequência B e compare o resultado com aquele obtido em 2.a.iv

Prepare um relatório com os resultados obtidos, seguindo o seguinte formato:

- coloque os nomes e números USP dos membros da equipe
- Escreva um parágrafo descrevendo os resultados obtidos. Por exemplo, na atividade 2.a.iii, você deve indicar qual foi o organismo do primeiro hit
- Copie e cole o primeiro alinhamento obtido em cada uma das atividades

Envie seu relatório para daniela.bizinelli@usp.br , ANTES de sair da aula! não serão aceitos relatórios enviados fora da sala de aula

>A

g ttgaaatccttcactgtagaaaaaggaatctatcaaacttctaacttttagagtccaacc
aacagaatctattggttagatttcctaataattacaaacttgtgcccttttgggtgaagttt
taacgccaccagatttgcacatctgtttatgcttggaacaggaagagaatcagcaactgtgt
tgctgattattctgtcctatataattccgcatcattttccacttttaagtgttatggagt
gtctcctactaaattaaatgatctctgctttactaatgtctatgcagattcatttgtaat
tagaggtgatgaagtcagacaaatcgctccagggcaaactggaaagattgctgattataa
ttataaattaccagatgattttacaggctgctgttatagcttggaattctaacaatcttga
ttctaagggttgggtggaattataattacctgtatagattgttttaggaagtctaactcaa
accttttgagagagatatttcaactgaaatctatcaggccggtagcacaccttgtaatgg
tgttgaagggtttaattggttactttcctttacaatcatatgggtttccaaccactaatgg
tgttggttaccaaccatacagagtagtagtactttcttttgaacttctacatgcaccagc
aactgtttgtggacctaanaagtcactaatttgggttaaaaacaaatgtgtcaatttcaa
cttcaatgggttaacaggcacagggtgttcttactgagtcatacaaaaagtttctgcctt
ccaacaatttggcagagacattgctgacactactgatgctgtccgtgatccacagacact
tgagattcttgacattacaccatgttcttttgggtgggtgctcagtggtataacaccaggaac
aaatacttctaaccagggttgctgttctttatcaggatgttaactgcacagaagtcctgt
tgctattcatgcagatcaacttactcctacttggcgtgtttattctacagggttctaagt
ttttcaaacacgtgcaggctgtttaataggggctgaacatgtcaacaactcatatgagtg
tgacatacccattgggtgcaggatatagcgctagtatcagactcagactaattctcctcg
gcgggcacgtagtgtagctagtcaatccatcattgcctacactatgtcacttgggtgcaga
aaattcagttgcttactctaataactctattgccatacccacaaattttactattagtg
taccacagaaattctaccagtgctatgaccaagacatcagtagattgtacaatgtacat
ttgtgggtgattcaactgaatgcagcaatcttttgttgcaatatggcagtttttgtacaca
attaaaccgtgctttaactggaatagctgttgaacaagacaaaaacaccaagaagttt
tgcacaagtcaacaaatttacaaaacaccaccaattaaagattttgggtgggtttaattt
ttcaciaaatattaccagatccatcaaaaccaagcaagaggtcatttattgaagatctact
tttcaacaaagtgacacttgcagatgctggcttcatcaacaatatgggtgattgccttgg
tgatattgctgctagagac

>B

CGGCCTGCGCGAGCTTGACGACCTCGGCTTTGCTCTTGCCGGTGGCCTTGGCCAGCAGGT
CGTAGACGGGGATGCCGGCGTCGCGGAGCTGGTTGAGGTCTTCGCCGGTGATGCGGCCGG
CGGCACTCATCTGTTGCAGGGCGACGGTGGCCCGCTGGACGCCTTCCGAGCCGGTGCCCA
TGCCTGAGGTGACATCGCCCAAGGTGCGCATGATGGGGATGACCTTTGACGACTCAATGC
CTGCCGAGATCAGGGAGGATGCGGCCGGTCTGCAACTCGGGGAACTCAAACGGGGTTCGTGG
CGGCGAACTTCTGCATCTCCCGCAGGAACGAGTCGGCCCTTTCCGCTGAGCCCAGCATGG
TCGTGAACGAGATGCGGGCCTGTTTCGTTGCCGCTAGCAACCTCTAGTCCCATCGCGCCAC
CGGCCACACCGACCGCAGCGAGCCCTAGAGCGCCCGCTGCGCCGAGGCCAGGGCCGCCT
TGCCCAAGAACC CGCCTGCCCTGCCCGCGGCGCTCAGCGTCGATCCTGCGGCCCGTGACA
TGCCGGTGGCCATACTGCGCCGATGGACTGCGCCTGCCGCTCCAACCCAGCGCCGCGC
GCTTGATGGAGTCCTCGGCGGCATCAGCGAAACCTCGCCCAACCCCTTGCCCGCCCGCT
CACCTGCCGCTGCGTCAGATCAGCAAACCCGCGGACCCGCCGCTCAAACCCGGCCAGCG
TCCGCTCAGCGGGGCCAGCATCCATCGTCAGGGTGCCGCTGAGGGTGCCGAGGTCCAACA
AGGGTCAGTCCTCTCGGCTTGGCGGGGCCGAGGTTTCGATGATCGTTAACGGCTTCAACG
CCAATCCTAGCCGAGTGGACGGCTGGAAGATGAGCGCACCGCTAGGCGCAACCACCAGAG
TCGGCGGCCGGTTCAACAGTTGCTCAATATGCCAGCGCAACCAAAACCAGGACCGGGCCG
CCATCAGCTCACGATTGCCGAGGTCGAGACCGTACACCTCCAGTAGGTCCGAGGCTACGA
GGACTCGGTGCCGGATGAGTCCTCGCCAGTCGATGCGGGTCGGTTTCGTCGCCCGTGTTCGG
GGCTTCCTCCAGCGGCGGATGCCCGCTTGGCCGGTTGGTTGTCTTCCTCGCGGTCTT
CGGCTTGTGGTCTGAGGTGCCTTCGCCGGCCCCAACGCTTTTGGGTCAACGCTGGTCCA
CCATCGTTCCGCGTAGTCCTCGGAGATGGTCAGCCAGTAGAAGATGGTGAGGGAGGCATG
TTTGATCCACAGCCATTGGAGACCGTCCGCCACCATTTTCGGCGTAGGTGTCGCCCCAGCAG
GGATGCGTAGAAGTCGCGTTCCTCGTCGTCGGTCAGTTTCAGCCGGGCGAGCTGGGTTTC
GCTGAGCGTTTTGCCTTCGGATGCGGCAACCATCAGGTCTGCGATGTCTTGGAATGCTTG
CCCGGTGGCTGCGGAGGGCGACTGGATTTTCGTAAGTCTTGCCCGCGATGGGCAGGGGGAG
GTTTGGTTGTAGAAGTCGGCCAGGTCACGCCACGCCATTAGGACTCAACCCCCAGAGCC

>C

MTGQLVQYGRHRQRRSYARISEVLELPLNLIETSSYQWFLDEGLREMFQDISPIEDFTGNLSLEFIDYS
LGEPKYPVEESKERDVTYSAPLRVKVRLINKETGEVKDQDVFMGDFPIMTDTGTFIINGAERVIVSQLVR
SPSVYFSGKVDKNGKKGFTATVIPNRGAWLEYETDAKDVVYVRIDRTRKLPVTVLLRALGFGSDQEILDL
IGENEYLRNTLDKDNTESENKALLEIYERLRPGEPPTVENAKSLDSRFFDPKRYDLANVGRYKINKKLH
IKNRLFNQRLAETLVDPETGEILAEKQIILDRRTLDKVLPLENGIGFRKLYPNGGVVEDEVTLQSIKIF
APTDQEGEQVINVIGNAYIEEEIKNITPADISSISYFFNLLHGVGDTDDIDHLGNRRLRVSGELLQNQF
RIGLSRMERVVRRERMSIQDTNTITPQQLINIRPVIASIKEFFGSSQLSQFMDQTNPLAELTHKRRLSALG
PGGLTRERAGMEVRDVHYSHYGRMCP IETPEGPNIGLINSLSSYAKVNRFGFIETPYRRVDPETGKVTGR
IDYLTADEEDNYVVAQANARLDDEGAFIDDSIVARFRGENTVVSRRNRVDYMDVSPKQVVSAAATACIPFLE
NDDSNRALMGANMQRQAVPLMQPEAPFVGTGMEYVSGKDSGAAVICKHPGIVERVEAKNVVWRRYEEVDG
QKVKGNLDKYSLLKFVRSNQGTCYNQRPVSVGDEVVKGEILADGPSMELGELALGRNVMVGFMTWDGYN
YEDAIIMSERLVKDDVYTSIHIEEYESEARDTKLGPEEITRDIPNVGEDALRNLDLDRGIIRIGAEVKDGD
LLVGKVTGKGTTELTAERLLHAIFGEKAREVRDTSLRVPHGGGGIIHDVKVFNREDGDELPPGVNQLVR
VYIVQKRKISEGDKMAGRHNKGVISKILPEEDMPYLPDGTPIIDIMLNPLGVPSRMNIGQVLELHMGMAA
RYLGIHIASPVFDGAREEDVWETLEEAGMSRDAKTVLYDGRGTGEPFDNRVSVGIMYMIKLAHMVDDKLHA
RSTGPPYSLVTQQPLGGKAQFGGQRFGEVWALEAYGAAATLQEILTVKSDDVGRVKTYEAIKGDVNP
EPGVPESEFKVLIKELQSLGMDVKILSGDEEIEEMRLEDEEDAKQADGLALSGDEEPEETASADVERDVV
TKE