BIOINFORMATICS OPEN DAYS 2015 :: CROSSWORDS CHALLENGE

Solve the crosswords below by completing the tasks assigned to each bracket. In keeping with the spirit of the challenge, it is strongly recommended that you use scripts and/or bioinformatics tools when applicable.
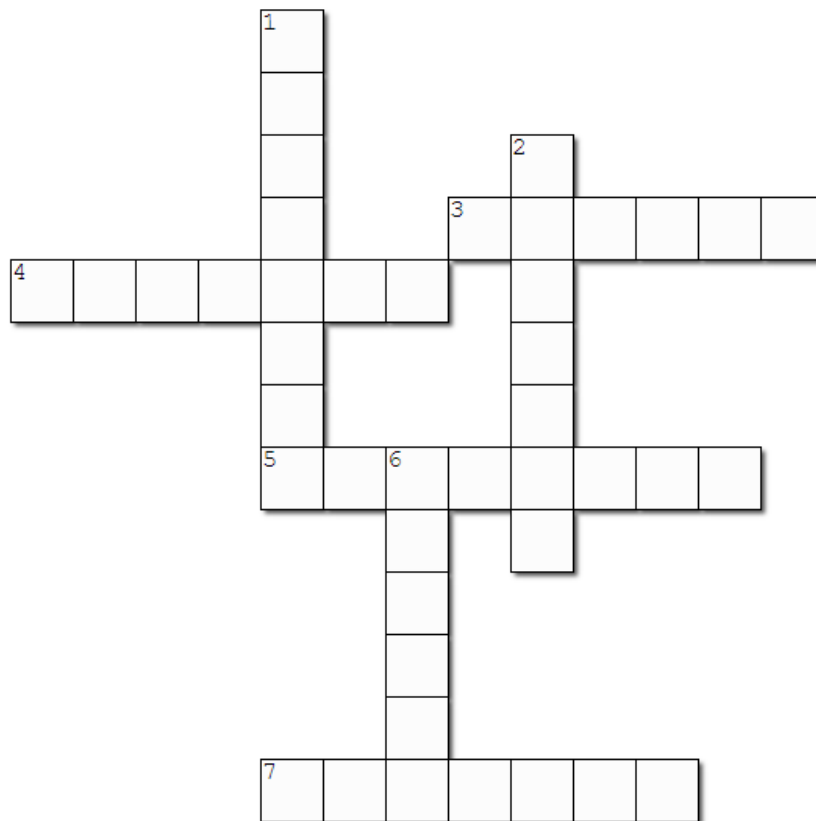
DOWN

1. In the following graph, the cost of moving from any given node to another is represented as a value ranging from 1 to 9. Which is the least costly path from the START to the END node?
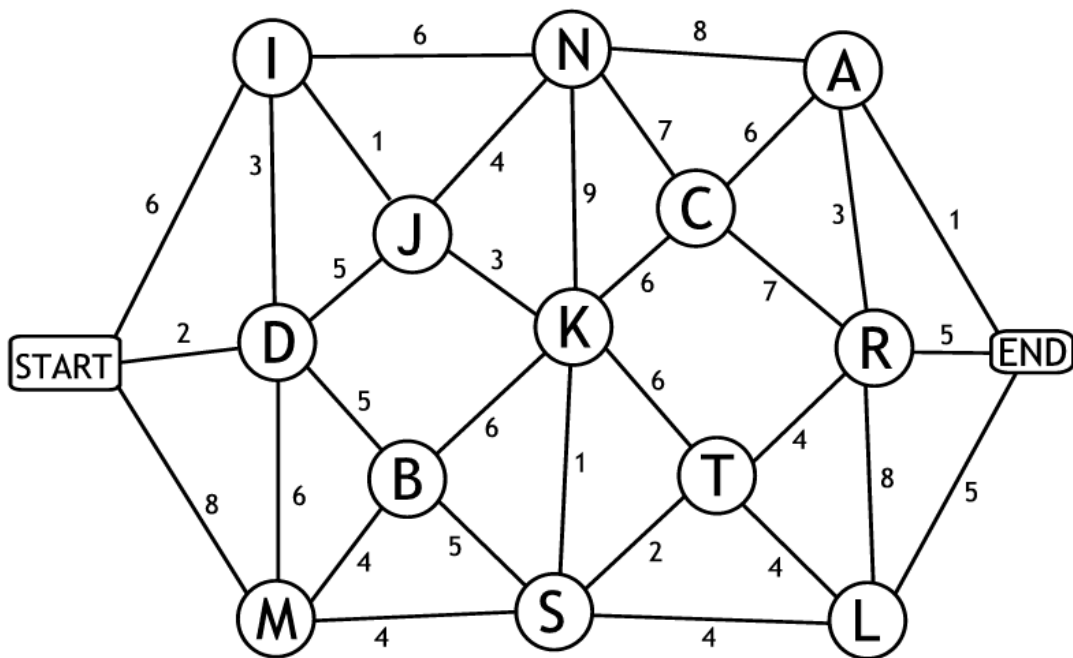
2. Search for the longest open reading frame (ORF) in the following sequence. A familiar field of study and a renowned name in the area should stand out. Complete this bracket with said name.

> sequence 1
CTACGTAGCTAGCTAGCTATATAGCTACGTACGTACGTACGTACGATCGATCGATCTACTAGCTAGCTAG
CTACGACTACTTAGACGCTCGGGGGGGGGGGGCTAGCCCCCCTATATATAACCCCCCCCCTAGCTAGCTAG
CTAGCTAGCTACGTAGCTAGCTAGCTATATAGGGGGGGGGGGGGGGTAGCGCTACGATCTAGCGATCGA
TCTTCGATCTAATCTACCTATCTAGATTCGTAGGGGGGGCTACCGATGGCCACTCATCTCGCACCCCTAGC
TTGCGAATAAATGACTCCTTAGCTTACTACGGGCGGCGGCGCGCGGCGAGGAAAAAAAAATATAAAAAA
AAATAAAATAGGCCCGGGGCCCCGTAAAATACTAGCTACTAAAACCGCCTTCCCGTACGATCGAGCTCCC
CCTTCGATAAAAAAAAAGGAAAAAAGAAAACGACTAGCTACGTAGTATAAGCTACGTCGCGCGCGGCG
GTATACGGGTCGATCGATCGTCGATCGAGCTAGTCTACTTTTTCCTACTCTAATATTCTCGTTTTCGTCTGT
CTATCTAAGTACTCGACGCCTATTCTTTATTCGCAAGCTAGGGGTGCGAGATGAGTGGCCATTAAGAATA
GGCGTCGAGTACTTAGATAGAAAGACGAAAACGAGAATATTAGAGTAGGAAAAAGTAGACTAGCTCGA
TCGACGATCGATCGACCCGTATACCGCCGCGCGCGACGTAGCTTATACTACGTAGCTAGTCGTTTTCTTTT
TTCCTTTTTTTTTTATCGAAGGGGGAGCTCGATCGTACGGGAAGGCGGTTTTAGTAGCTAGTATTTTACGG
GGCCCCGGGCCTATTTTATTTTTTTTTATATTTTTTTTTCCTCGCCGCGCGCCGCCGCCCGTAGTAAGCTAA
GGAGTAATCGGTAGCCCCCCCTACGAATCTAGATAGGTAGATTAGATCGAAGATCGATCGCTAGATCGT
AGCGCTACCCCCCCCCCCCCCTATATAGCTAGCTAGCTACGTAGCTAGCTAGCTAGCTAGCTAGGGGGGG
GGTTATATATAGGGGGGCTAGCCCCCCCCCCCGAGCGTCTAAGTAGTCGTAGCTAGCTAGCTAGTAGATC
GATCGATCGTACGTACGTACGTACGTAGCTATATAGCTAGCTAGCTACGTAG

6. Given the following pseudocode, what would the algorithm print out if implemented in Python?

```
alphabet = [A..Z]
array = [-7, 1, 3, 9, 5, -7]
SET p to 0
FOR n in array
        IF n // 2 = 3:
                p = p * n
        ELSE IF d mod 3 = 0:
                p = p - n
        ELSE:
                p = p + n
        PRINT alphabet[p]
```

ACROSS

3. Using the restriction enzymes EcoRI and HindIII, digest the following sequence and translate the shortest resulting fragment into the corresponding amino acid sequence. Join the letters in every odd-numbered position (starting from the first) to spell out the answer.

> sequence 2
AACGGTCCGCTACCTTACAGGAATTGAGACCGTCCTTTAATTTCCCTTGCATATATGTTGCGTTTCTTCGAC
CTTTTAACCGCTCCCTTAGGAGAAAGACAGATAGCCTCTTACCCGTACTCCACCGTTGGCAGCACGATCGC
ATGTCCCACGTGAACCATTGGTAAACCCTGTGGCCTGTGAGCGACAAAAGATTTAATGGGAAATTCGCGC
CCATAACTTGGTCCGAATACGGGTCCTAGCAACGTTCGTCTGAGTTTGATCTATATAATACGGGCGGTAT
GTCTGCTTTGATCAACCTCCAATACCTCGTATGATAGTGCACCCGCTGGTGATCACTCAATGATCTGGGCT
CCCCGTTGCAACTACGGGGATTTTTCGAGACCGACCTGCGTTCGGCATTGTGGGCACAGTGAAGTATTAG
CAAACGTTAAGTCCCGAACTAGATGTGACCTAACGGTAAGAGAATTTCATAATACGTCCTGCCGCACGCG
CAAGGTACATTTGGACAGTATTGAATGGACTCTGATCAACCTTCACACCGATCTAGAATCGAATGCGTAG
ATCAGCCAGGTGCAAACCAAAAATTCTAGGTTACTAGAAGTTTTGCGACGTTCTAAGTGTTGGACGAAAT
GAATCGCGACCCAGGATGAGGTCGCCCTAAAAAATAGATTTCTGCAACTCTCCTCGTGAGCAGTCTGGTG
TATCGAAAGTACAGGACTAGCCTTCCTAGCAACCGCGGGCTGGGAGTCTGAGACATCACTCAAGATATAT
GCTCGGTAACGTATGCTCTAGCCATCTAACTATTCCCTATGTCTTATAGGGGCCTACGTTATCTGCCTGTC
GAACCATAGGATCCGCGTCAGCGCGCAGGCTTGGATCGAGATGAAATCTCCGGAGCCTAAGACCACGAG
CGTCTGGCGTCTTGGCTAATCCCCCTACATGTTGTTATAAACAATCAGTGGAAACTCAGTGCTAGAGGGT
GGAGTGACCTTAAATCAAGCTTAGCCATCAACGGGGGTTATGAGAGTCGTAGAATTCCGCTGCAGCTGA
CTGAAGGACGATATTAATCGGAAGGAGTATTCAACGCAATGAAGTCGCAGGGTTGACGTGGGAATGGT
GCTTCTGTCCAAACAGGTAAGGGTATGAGGCCGCAACCGTCCCCCAAGCGTACAGGGTGCACTTTGCAA
CGATTTCGGAGTCCAAAGACTCGCTGTTTTCGAAATTTGCGCTCAAGGGCGAGTATTGAACCAGGCTTAC
GCCCAAGAACGTAGCAAGGTGACTCAAACAAAGTACATCTTGCCCGCGTTTCATATGAATCAAGTTAGAA
GTTATGGAGCATAATAACATGTGGATGGCCAGTGGTCGGTTGCTACACCCCTGCCGCAACGTTGAAGGT
CCCGGATTAGACTGGCTGGATCTATGCCGTGACACCCGTTATACTCCATTACCGTCTGTGGGTCACACCCC
CTTGTTGTGGACTGGATTGCCATTCTCTCAGTGTATTACGCAGGCCGGCGCACGGGTCCCATATAAACCT
GTCATAGTCTTACCTGACTCTACTTGGAAATGTGGCTAGGCCTTTGCCCACGCACCTGATCGGTCCTCGTT
TGCTTTTTAGGACCGGATGAACTACAGAGCATTGCAAGAATCTCTACCTGCTTTACAAAGTGCTGGATCCT
ATTCCAGCGGGATGTTTTATCTAAACACGATGAGAGGAGTATTCGTCAGGCCACATGGCTTTCTTGTTCT
GGTCGGATCCATCGTTGGCGCCCGACCCCCCCATTCCATAGTGAGTTCTTCGTCCGAGCCATTGTATGCCA
GATCGACAGACAGATAGCGGATCCAGTATATCCCTGGAAACTATAGACGCACAGGTTGGAATCTTAAGT
GAAGTCGCGCGTCCAAACCCAGCTCTATTTTAGTGGTCATGGGTTCTGGTCCCCCCGAGCCGCGGAACCG
ATTAGGACCATGTACAACAATACTTATTAGTCATCTTTTAGACACAATCTCCCTGCTCAGTGGTATATGGTT
TTTGCTATAATTAGCCACCCTCATAAG

4. Search for Bioinformatics articles with the following four IDs. Who or what do these articles have in common?

1577098   18613140   23455439   20430689

5. Align the sequences provided below. Upon inspecting the region where they overlap, identify the amino acids in the second sequence that differ from the first.

> sequence 3
PYAHYMRIDIGMYQDFRSGPMAMLNLTRDGVSAEPHDLTNNNPDNSKGGGEFRNLNELTSMRERKVFMA
FAMTSAIDGAQGAFKLVVEKKQGRSEDMCENDELVEVAAQNDIDIAVRNEQENDGKCQLTQFLYFLSDAIRR
QAVLVKASDIFIPEGVRWLKKDEADARVDDRGTVVVRRLNGEPKDEVVIFFDETETVPCEAEIYGTLWRIAVN
EFADGLLDLESITPDGKVDSGAHKCYWIAGDNKEADPKIKHWYSVFPSDVMVLSKFAEALRHFTFKGSGRGPS
VCISRESLNFSIHNGLYGLYKAKAPDVKENKTISPEGLAFNTAVYPNLGTQSFWGMTSRCPAGQKDYKTVVKM
TFIRAGNRSLVGATLEDLVNMKVGGTMKGEGEFLIFIHSLHAIEEIFDEHPIGRLETERAENDRAPPQIDEVNVL
LEKTGWDICQVKCTHLEEGYAESAGLPAANALYILYLVPGIDLLDVLRSEAEGEILATDLLVHYLLVLHSGAGLIR
QIVEPQKLTILKPIKVLDELTKAQIGGKLFWLGTVNFSHGVARDRMDANDPSTCYARATNKGVYAGKIKGEHD
PTILDSAGSHSQGTNPVQQPSSAILILKQLKFQIRGRETGETTTTRQCSDAGMILSVPVVVPADAYLLYKQPDLI
PNANEMAEAKGSSLLLSISIRTDNSGRPEERKSQAVKPMIQASASDDPTYYIVNALAPITYTAIDRYNTPVTIQR
KKLQQRMSDVGRGHTNETDSEKTKSPKTRVHEHYSFVSPQVLVGPSILYFIFQAVAIVALEGLATKVFNLLLRV
HKFHVQVQFPPQSRFKKEGLGRNFVMAFKKTECEGIGAGFSGKETRLPLLSAIQCLYENALSKSETHAAYLDRL
WGARALKSVLSHPLMVRNRSSTPKVPRNATESSIMIMILFYIHLLAQRDGLASGAWGECAILSFDRSIDRGWR
IVWLPLGLFQVFQEQQTLYLLQTARSSNAEAVMELSKMACPNLVP

> sequence 4
IKHWYSVFPSDVMVLSKFAEALRHFTFKGSGRGPSVCISRESLNFSIHNGLYGLYKAKAPDVKENKTISPEGLAF
NTAVYPNLGTQSFWGMTSRCPAGQKDYKTVVKMTFIRAGNRSLVGAALEDLVNMKVGGLMKGEGEFLIFIH
SLTAIEEIFDEHPISRLETERAENDRAPPQIDCVNVLLEKTGWDICQVKHTHLEEGYAESAGLUAANALYILYLVP
LIDLLDVLRSEAEGEILATDLLVHYLLVLHSGAGLIRQIVEPQKLTILKPIKVLDELTKAQIGGKLFWLGTVNFSHG
VARDRMDANDPSTCYARATNKGVYAGKIKGEHDPTILDSAGSHSQGTNPVQQPSSAILILKQLKFQIRGRETG
ETTTTRQCSDAGMILSVPVVVPADAYLLYKQPDLIPNANEMAEAKGSSLLLSISIRTDNSGRPEERKSQAVKPM
IQASASDDPTYYIVNALAPITYTAIDRYNTPVTIQRKKLQQRMSDVGRGHT

7. The following binary string encodes an interesting read into the early days of Bioinformatics. Who wrote it?

01010000 01001101 01000011 00110011 00110000
00110110 00111000 00111001 00110010 00110101