**IC251 – Basics of Bioinformatics (2022-23W)**

**Tutorial 4**  March 15, 2023

**Part A: Downloading spike protein sequences from UniProt database.**

3] Write the UniProt ID of the top 5 entries.

Solution :

1. K9NQ0Q4F25Q8
2. A3EXD0
3. A0A2R4KP93
4. A3EX94
5. Q0Q4F2

5] Repeat the steps 2 to 4 but here for “spike protein of SARS Coronavirus” (instead of MERS), write top 5 IDs and save first entry file (name the first entry as sequence2.txt).

Solution :

1. P59594
2. P0DTC2
3. Q3LZX1
4. Q3I5J5
5. P25190

6] Repeat again the steps 2 to 4 but here for “spike protein of SARS Coronavirus 2”, write top 5 IDs and save first entry (name the first entry as sequence3.txt).

1. [P5](https://www.uniprot.org/uniprotkb/P59594/entry)9594
2. P0DTC2
3. Q3LZX1
4. Q3I5J5
5. P25190

7] Repeat again the steps 2 to 4 but here for “spike protein of Murine Coronavirus”, write top 5 IDs and save first entry (name the first entry as sequence4.txt).

1. P11225
2. P11224
3. Q02385
4. P22432
5. Q9IKD1

8] Repeat again the steps 2 to 4 but here for “spike protein of Avian infectious bronchitis virus”, write top 5 IDs and save first entry (name the first entry as sequence5.txt).

1. P11223
2. P12651
3. P12650
4. P12722
5. P17662

10] Copy these sequences in one word file (for tutorial report purpose only).

**Sequence 1 :**

>sp|K9N5Q8|SPIKE\_MERS1 Spike glycoprotein OS=Middle East respiratory syndrome-related coronavirus (isolate United Kingdom/H123990006/2012) OX=1263720 GN=S PE=1 SV=1

MIHSVFLLMFLLTPTESYVDVGPDSVKSACIEVDIQQTFFDKTWPRPIDVSKADGIIYPQ

GRTYSNITITYQGLFPYQGDHGDMYVYSAGHATGTTPQKLFVANYSQDVKQFANGFVVRI

GAAANSTGTVIISPSTSATIRKIYPAFMLGSSVGNFSDGKMGRFFNHTLVLLPDGCGTLL

RAFYCILEPRSGNHCPAGNSYTSFATYHTPATDCSDGNYNRNASLNSFKEYFNLRNCTFM

YTYNITEDEILEWFGITQTAQGVHLFSSRYVDLYGGNMFQFATLPVYDTIKYYSIIPHSI

RSIQSDRKAWAAFYVYKLQPLTFLLDFSVDGYIRRAIDCGFNDLSQLHCSYESFDVESGV

YSVSSFEAKPSGSVVEQAEGVECDFSPLLSGTPPQVYNFKRLVFTNCNYNLTKLLSLFSV

NDFTCSQISPAAIASNCYSSLILDYFSYPLSMKSDLSVSSAGPISQFNYKQSFSNPTCLI

LATVPHNLTTITKPLKYSYINKCSRFLSDDRTEVPQLVNANQYSPCVSIVPSTVWEDGDY

YRKQLSPLEGGGWLVASGSTVAMTEQLQMGFGITVQYGTDTNSVCPKLEFANDTKIASQL

GNCVEYSLYGVSGRGVFQNCTAVGVRQQRFVYDAYQNLVGYYSDDGNYYCLRACVSVPVS

VIYDKETKTHATLFGSVACEHISSTMSQYSRSTRSMLKRRDSTYGPLQTPVGCVLGLVNS

SLFVEDCKLPLGQSLCALPDTPSTLTPRSVRSVPGEMRLASIAFNHPIQVDQLNSSYFKL

SIPTNFSFGVTQEYIQTTIQKVTVDCKQYVCNGFQKCEQLLREYGQFCSKINQALHGANL

RQDDSVRNLFASVKSSQSSPIIPGFGGDFNLTLLEPVSISTGSRSARSAIEDLLFDKVTI

ADPGYMQGYDDCMQQGPASARDLICAQYVAGYKVLPPLMDVNMEAAYTSSLLGSIAGVGW

TAGLSSFAAIPFAQSIFYRLNGVGITQQVLSENQKLIANKFNQALGAMQTGFTTTNEAFH

KVQDAVNNNAQALSKLASELSNTFGAISASIGDIIQRLDVLEQDAQIDRLINGRLTTLNA

FVAQQLVRSESAALSAQLAKDKVNECVKAQSKRSGFCGQGTHIVSFVVNAPNGLYFMHVG

YYPSNHIEVVSAYGLCDAANPTNCIAPVNGYFIKTNNTRIVDEWSYTGSSFYAPEPITSL

NTKYVAPQVTYQNISTNLPPPLLGNSTGIDFQDELDEFFKNVSTSIPNFGSLTQINTTLL

DLTYEMLSLQQVVKALNESYIDLKELGNYTYYNKWPWYIWLGFIAGLVALALCVFFILCC

TGCGTNCMGKLKCNRCCDRYEEYDLEPHKVHVH

**Sequence 2:**

>sp|P59594|SPIKE\_SARS Spike glycoprotein OS=Severe acute respiratory syndrome coronavirus OX=694009 GN=S PE=1 SV=1

MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFL

PFYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNS

TNVVIRACNFELCDNPFFAVSKPMGTQTHTMIFDNAFNCTFEYISDAFSLDVSEKSGNFK

HLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSP

AQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIY

QTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTF

FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCV

LAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLND

YGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTP

SSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQD

VNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASY

HTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDC

NMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFG

GFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGL

TVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYE

NQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLN

DILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSK

RVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVFN

GTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKN

HTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWL

GFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLHYT

**Sequence 3:**

>sp|P59594|SPIKE\_SARS Spike glycoprotein OS=Severe acute respiratory syndrome coronavirus OX=694009 GN=S PE=1 SV=1

MFIFLLFLTLTSGSDLDRCTTFDDVQAPNYTQHTSSMRGVYYPDEIFRSDTLYLTQDLFL

PFYSNVTGFHTINHTFGNPVIPFKDGIYFAATEKSNVVRGWVFGSTMNNKSQSVIIINNS

TNVVIRACNFELCDNPFFAVSKPMGTQTHTMIFDNAFNCTFEYISDAFSLDVSEKSGNFK

HLREFVFKNKDGFLYVYKGYQPIDVVRDLPSGFNTLKPIFKLPLGINITNFRAILTAFSP

AQDIWGTSAAAYFVGYLKPTTFMLKYDENGTITDAVDCSQNPLAELKCSVKSFEIDKGIY

QTSNFRVVPSGDVVRFPNITNLCPFGEVFNATKFPSVYAWERKKISNCVADYSVLYNSTF

FSTFKCYGVSATKLNDLCFSNVYADSFVVKGDDVRQIAPGQTGVIADYNYKLPDDFMGCV

LAWNTRNIDATSTGNYNYKYRYLRHGKLRPFERDISNVPFSPDGKPCTPPALNCYWPLND

YGFYTTTGIGYQPYRVVVLSFELLNAPATVCGPKLSTDLIKNQCVNFNFNGLTGTGVLTP

SSKRFQPFQQFGRDVSDFTDSVRDPKTSEILDISPCSFGGVSVITPGTNASSEVAVLYQD

VNCTDVSTAIHADQLTPAWRIYSTGNNVFQTQAGCLIGAEHVDTSYECDIPIGAGICASY

HTVSLLRSTSQKSIVAYTMSLGADSSIAYSNNTIAIPTNFSISITTEVMPVSMAKTSVDC

NMYICGDSTECANLLLQYGSFCTQLNRALSGIAAEQDRNTREVFAQVKQMYKTPTLKYFG

GFNFSQILPDPLKPTKRSFIEDLLFNKVTLADAGFMKQYGECLGDINARDLICAQKFNGL

TVLPPLLTDDMIAAYTAALVSGTATAGWTFGAGAALQIPFAMQMAYRFNGIGVTQNVLYE

NQKQIANQFNKAISQIQESLTTTSTALGKLQDVVNQNAQALNTLVKQLSSNFGAISSVLN

DILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKMSECVLGQSK

RVDFCGKGYHLMSFPQAAPHGVVFLHVTYVPSQERNFTTAPAICHEGKAYFPREGVFVFN

GTSWFITQRNFFSPQIITTDNTFVSGNCDVVIGIINNTVYDPLQPELDSFKEELDKYFKN

HTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYVWL

GFIAGLIAIVMVTILLCCMTSCCSCLKGACSCGSCCKFDEDDSEPVLKGVKLHYT

**Sequence 4:**

>sp|P11225|SPIKE\_CVMJH Spike glycoprotein OS=Murine coronavirus (strain JHM) OX=11144 GN=S PE=1 SV=1

MLFVFILLLPSCLGYIGDFRCIQTVNYNGNNASAPSISTEAVDVSKGRGTYYVLDRVYLN

ATLLLTGYYPVDGSNYRNLALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNTPTGAT

SYFPTIVIGSLFGNTSYTVVLEPYNNIIMASVCTYTICQLPYTPCKPNTNGNRVIGFWHT

DVKPPICLLKRNFTFNVNAPWLYFHFYQQGGTFYAYYADKPSATTFLFSVYIGDILTQYF

VLPFICTPTAGSTLAPLYWVTPLLKRQYLFNFNEKGVITSAVDCASSYISEIKCKTQSLL

PSTGVYDLSGYTVQPVGVVYRRVPNLPDCKIEEWLTAKSVPSPLNWERRTFQNCNFNLSS

LLRYVQAESLSCNNIDASKVYGMCFGSVSVDKFAIPRSRQIDLQIGNSGFLQTANYKIDT

AATSCQLYYSLPKNNVTINNYNPSSWNRRYGFKVNDRCQIFANILLNGINSGTTCSTDLQ

LPNTEVATGVCVRYDLYGITGQGVFKEVKADYYNSWQALLYDVNGNLNGFRDLTTNKTYT

IRSCYSGRVSAAYHKEAPEPALLYRNINCSYVFTNNISREENPLNYFDSYLGCVVNADNR

TDEALPNCNLRMGAGLCVDYSKSRRARRSVSTGYRLTTFEPYMPMLVNDSVQSVGGLYEM

QIPTNFTIGHHEEFIQIRAPKVTIDCAAFVCGDNAACRQQLVEYGSFCDNVNAILNEVNN

LLDNMQLQVASALMQGVTISSRLPDGISGPIDDINFSPLLGCIGSTCAEDGNGPSAIRGR

SAIEDLLFDKVKLSDVGFVEAYNNCTGGQEVRDLLCVQSFNGIKVLPPVLSESQISGYTA

GATAAAMFPPWTAAAGVPFSLNVQYRINGLGVTMNVLSENQKMIASAFNNALGAIQEGFD

ATNSALGKIQSVVNANAEALNNLLNQLSNRFGAISASLQEILTRLDAVEAKAQIDRLING

RLTALNAYISKQLSDSTLIKFSAAQAIEKVNECVKSQTTRINFCGNGNHILSLVQNAPYG

LCFIHFSYVPTSFKTANVSPGLCISGDRGLAPKAGYFVQDNGEWKFTGSNYYYPEPITDK

NSVAMISCAVNYTKAPEVFLNNSIPNLPDFKEELDKWFKNQTSIAPDLSLDFEKLNVTFL

DLTYEMNRIQDAIKKLNESYINLKEVGTYEMYVKWPWYVWLLIGLAGVAVCVLLFFICCC

TGCGSCCFRKCGSCCDEYGGHQDSIVIHNISAHED

**Sequence 5:**

>sp|P11223|SPIKE\_IBVB Spike glycoprotein OS=Avian infectious bronchitis virus (strain Beaudette) OX=11122 GN=S PE=1 SV=1

MLVTPLLLVTLLCALCSAVLYDSSSYVYYYQSAFRPPSGWHLQGGAYAVVNISSEFNNAG

SSSGCTVGIIHGGRVVNASSIAMTAPSSGMAWSSSQFCTAHCNFSDTTVFVTHCYKHGGC

PLTGMLQQNLIRVSAMKNGQLFYNLTVSVAKYPTFRSFQCVNNLTSVYLNGDLVYTSNET

IDVTSAGVYFKAGGPITYKVMREVKALAYFVNGTAQDVILCDGSPRGLLACQYNTGNFSD

GFYPFTNSSLVKQKFIVYRENSVNTTCTLHNFIFHNETGANPNPSGVQNIQTYQTKTAQS

GYYNFNFSFLSSFVYKESNFMYGSYHPSCKFRLETINNGLWFNSLSVSIAYGPLQGGCKQ

SVFKGRATCCYAYSYGGPSLCKGVYSGELDHNFECGLLVYVTKSGGSRIQTATEPPVITQ

NNYNNITLNTCVDYNIYGRTGQGFITNVTDSAVSYNYLADAGLAILDTSGSIDIFVVQGE

YGLNYYKVNPCEDVNQQFVVSGGKLVGILTSRNETGSQLLENQFYIKITNGTRRFRRSIT

ENVANCPYVSYGKFCIKPDGSIATIVPKQLEQFVAPLFNVTENVLIPNSFNLTVTDEYIQ

TRMDKVQINCLQYVCGSSLDCRKLFQQYGPVCDNILSVVNSVGQKEDMELLNFYSSTKPA

GFNTPVLSNVSTGEFNISLLLTNPSSRRKRSLIEDLLFTSVESVGLPTNDAYKNCTAGPL

GFFKDLACAREYNGLLVLPPIITAEMQALYTSSLVASMAFGGITAAGAIPFATQLQARIN

HLGITQSLLLKNQEKIAASFNKAIGHMQEGFRSTSLALQQIQDVVSKQSAILTETMASLN

KNFGAISSVIQEIYQQFDAIQANAQVDRLITGRLSSLSVLASAKQAEYIRVSQQRELATQ

KINECVKSQSIRYSFCGNGRHVLTIPQNAPNGIVFIHFSYTPDSFVNVTAIVGFCVKPAN

ASQYAIVPANGRGIFIQVNGSYYITARDMYMPRAITAGDVVTLTSCQANYVSVNKTVITT

FVDNDDFDFNDELSKWWNDTKHELPDFDKFNYTVPILDIDSEIDRIQGVIQGLNDSLIDL

EKLSILKTYIKWPWYVWLAIAFATIIFILILGWVFFMTGCCGCCCGCFGIMPLMSKCGKK

SSYYTTFDNDVVTEQYRPKKSV

**Part B: Performing local alignment using BLAST.**

5] For each alignment result, write the identity, positive and e- values in word file in Table format.

| Alignment | Identities | Positives | E-values |
| --- | --- | --- | --- |
| sequence1 with 2 | 366/1052(35%) | 545/1052(51%) | 2e-177 |
| sequence 1 with 3 | 366/1052(35%) | 545/1052(51%) | 2e-177 |
| sequence 1 with 4 | 424/1339(32%) | 632/1339(47%) | 0.0 |
| sequence 1 with 5 | 198/572(35%) | 310/572(54%) | 2e-96 |
| sequence 2 with 3 | 1255/1255(100%) | 1255/1255(100%) | 0.0 |
| sequence 2 with 4 | 409/1319(31%) | 619/1319(46%) | 2e-169 |
| sequence 2 with 5 | 193/534(36%) | 291/534(54%) | 1e-102 |
| sequence 3 with 4 | 409/1319(31%) | 619/1319(46%) | 2e-169 |
| sequence 3 with 5 | 193/534(36%) | 291/534(54%) | 1e-102 |
| sequence 4 with 5 | 206/553(37%) | 308/553(55%) | 4e-111 |