Multiple Sequence Alignment with 10 Sequences

A Minor project report

Submitted to -

Department Of Medical Laboratory Science

Ram Devi Jindal Group Of Institutions

Punjab

(Affiliated With)

I.K Gujral Punjab Technical University

Jalandhar, Kapurthala

For the award of Degree

Bachelor of Science [Medical Laboratory Sciences]

SESSION :2022-23

BY

Manisha Kumari (Roll.No.-2006812)

BSc.(MLS)-5th Semester

SUPERVISOR

Mrs. Anjali Sharma

RAM DEVI JINDAL GROUP OF PROFESSIONAL INSTITUTIONS

BASOLI (S.A.S NAGAR)140501(PUNJAB)

RAM DEVI JINDAL GROUP OF PROFESSIONAL INSTITUTIONS, BASOLI (140501) PUNJAB

Affiliated to IKG Punjab Technical University, Jalandhar

Ph. 9501688555, 9463823732

https://rdginstitutions.com, admission@rdginstitutions.com

CERTIFICATE

This is to certify that the minor project report entitled '*Multiple Sequence Alignment with 10 Sequences*' submitted for the fulfillment of the requirement for the degree of Bachelor of Science (Medical Laboratory Science) to I.K Gujral Punjab Technical University , Jalandhar embodies the work carried out by Mrs. Anjali Sharma in RDJ Group of professional Institutions, BASOLI.

Checked and Verified by

Supervisor

Mrs.Anjali Sharma

Assistant Professor

RDJ Group Of Institutions,

Basoli(S.A.S.Nagar)140501

Forwarded by

Head of Department

Mr.Vikash Kumar

Assistant Professor

RDJ Group Of Institutions,

Basoli(S.A.S.Nagar)140501

Office

Principal

A.K. Gupta

RDJ Group Of Institutions,

Basoli (S.A.S.Nagar)140501

External

ACKNOWLEDGEMENT

The completion of the project could not have been possible without the participation and assistance of so many people whose names may not all be enumerated. Their contribution is sincerely appreciated and gratefully acknowledged. However, I would like to express their deep appreciation and indebtedness, particularly to the following:

First, thanks to my parents for giving me encouragement, enthusiasm and invaluable assistance to me. Without this, I might not be able to complete this training properly. They helped me a lot in finalizing this project within the limited time frame.

Secondly, I would like to express my special thanks of gratitude to my teacher Mrs. Anjali Sharma as well as HOD sir Mr.Vikash Kumar and all the other teachers who gave me the golden opportunity to do this wonderful project on the topic '***Multiple Sequence Alignment with 10 Sequences*** 'which also helped in doing a lot of research and I came to know about so many new things .I am thankful to them.

I would also like to express my gratitude to my friends for their kind cooperation and encouragement which helps me in the completion of this project.

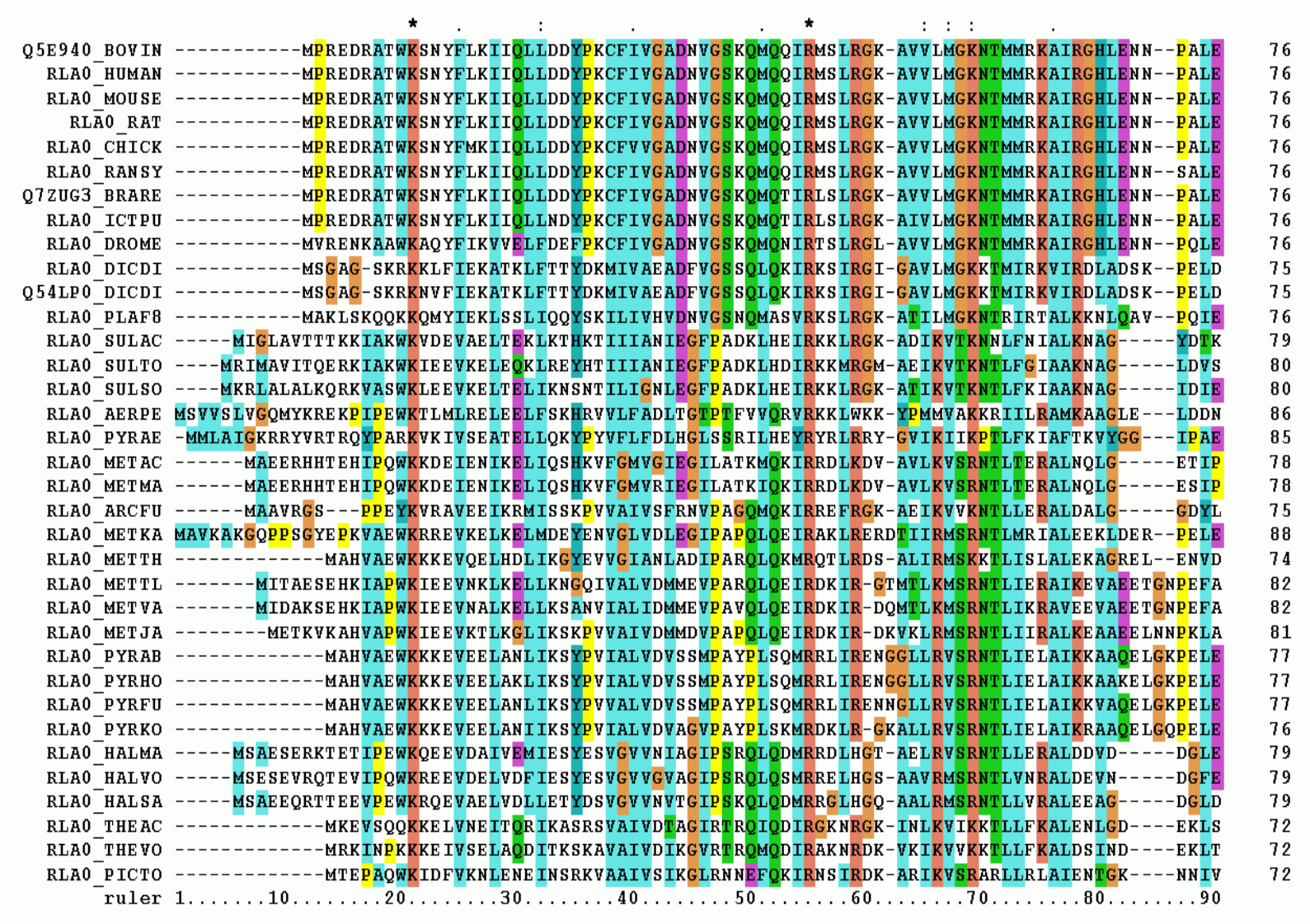
However, I accept the sole responsibility for any possible error and would be extremely grateful to the reader for this project report if they bring such mistake to my notice. I have tried my best to present this information as clearly as possible using basic terms.

Manisha Kumari
Roll No. 2006812
Semester 5th

INDEX

INTRODUCTIONTO MULTIPLE SEQUENCE ALIGNMENT (MSA)

Multiple sequence alignment (**MSA**) may refer to the process or the result of sequence alignment of three or more biological sequences, generally protein, DNA, or RNA. In many cases, the input set of query sequences are assumed to have an evolutionary relationship by which theyshare a linkage and are descended from a common ancestor. From the resulting MSA, sequence homology can be inferred and phylogenetic analysis can be conducted to assess the sequences' shared evolutionary origins. Visual depictions of the alignment as in the image at right illustrate mutation events such as point mutations (single amino acid or nucleotide changes) that appear as differing characters in a single alignment column, and insertion or deletion mutations (idles or gaps) that appear as hyphens in one or more of the sequences in the alignment. Multiple sequence alignment is often used to assess sequence conservation ofprotein domains, tertiary and secondary structures, and even individual amino acids or nucleotides.



First 90 positions of a protein multiple sequence alignment of instances of the acidic ribosomal protein P0 (L10E) from severalorganisms.

Computational algorithms are used to produce and analyze the MSAs due to the difficulty and intractability of manually processing the sequences given their biologically-relevant length. MSAs require more sophisticated methodologies than pairwise alignment because they are more computationally complex. Most multiple sequence alignment programsuse heuristic methods rather than global optimization because identifying the optimal alignment between more than a few sequences of moderate length is prohibitively computationally expensive. On the other hand, heuristic methods generally fail to give guarantees on the solution quality, with heuristic solutions shown to be often far below the optimalsolution on benchmark instances.

MULTIPLE SEQUENCE ALIGNMENT

1. Multiple sequence alignment (MSA) is generally the alignment of three or more biological sequence (protein or Nucleic acid) of similar length.

2. From the output, homology can be inferred and the evolutionary relationship between the sequences studied.

TYPES OF MSA

1. Dynamic Programming Approach
2. Progressive Method
3. Iterative Method

Dynamic Programming Approach

1. Dynamic Programming is applicable to align anynumber of sequences.
2. Computes an optimal alignment for a given score functions.
3. Because of its high running time, it is not typically used in practice.

Progressive Method

1. In this method, pair wise global alignment is performed for all the possible and these pairs are aligned together on the basis of their similarity.
2. The most similar sequences are aligned together and then less related sequences are added to it progressively one by one until a complete multiple query set is obtained.
3. This method is also called Hierarchical method or Tree method.

Iterative Method

1. A method of performing a series of steps to produce successively better approximation to align many sequences step by step is called iterative method.
2. Here the pair wise sequence alignment is totally avoided.
3. This method attempt to improve on the weak points of the progressive methods the heavy dependence on the accuracy of the initial pair wise alignment.

Tools Involved in MSA

- Crustal W
- Crustal W2
- Crustal Omega
- Kalign
- M View
- MAFFT
- MUSCLE
- T-Coffee
- Web PRANK
- MEME
- MACAW

1.Clustal Omega

New MSA tool that uses seeded guide trees and HMM profile techniques to generate alignments. Suitable for medium large alignments.

2.Cons (EMBOSS)

EMBOSS Cons creates a consensus sequence from a protein or nucleotide multiple alignment.

3.Kalign

Very fast MSA tool that concentrates on local regions. Suitable for large alignments.

4.MAFFT

MSA tool that uses Fast Fourier Transforms. Suitable formedium-large alignments.

5.MUSCLE

Accurate MSA tool, especially good with proteins. Suitablefor medium alignments.

6.MView

Transform a Sequence Similarity Search result into a Multiple Sequence Alignment or reformat a Multiple Sequence Alignment using the MView program.

7.T-Coffee

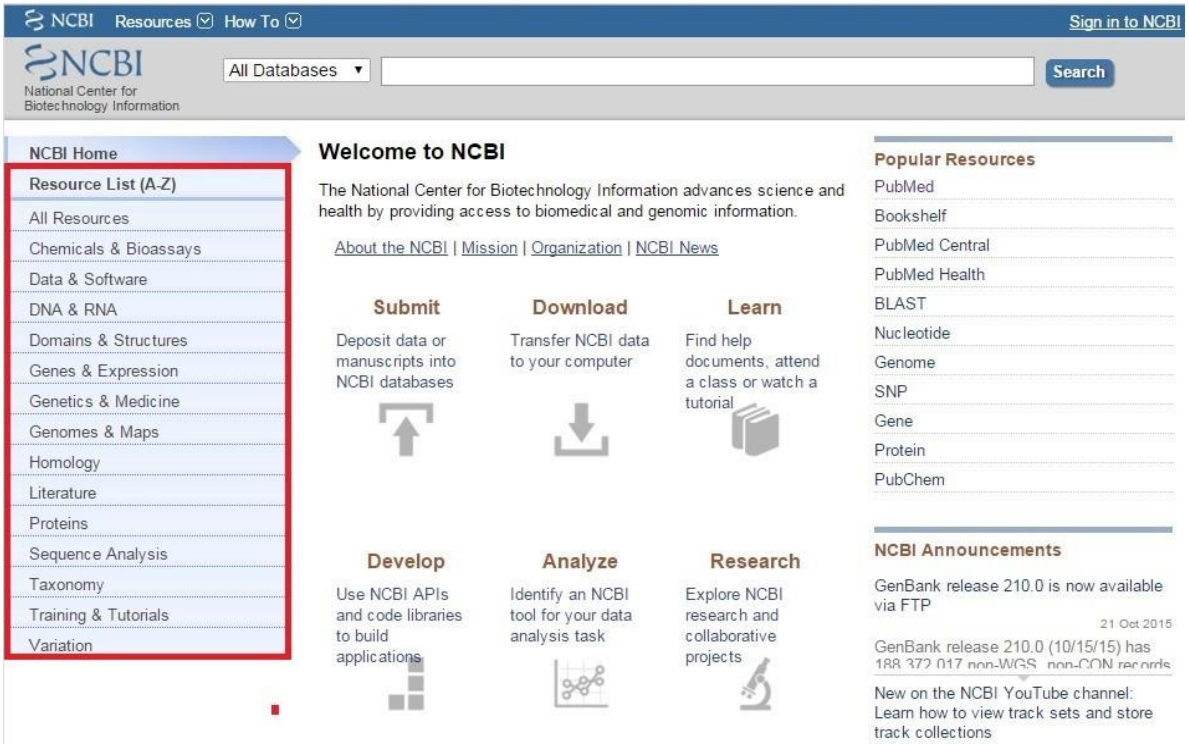
Consistency-based MSA tool that attempts to mitigate the pitfalls of progressive alignment methods. Suitable for small alignments.

8.WebPRANK

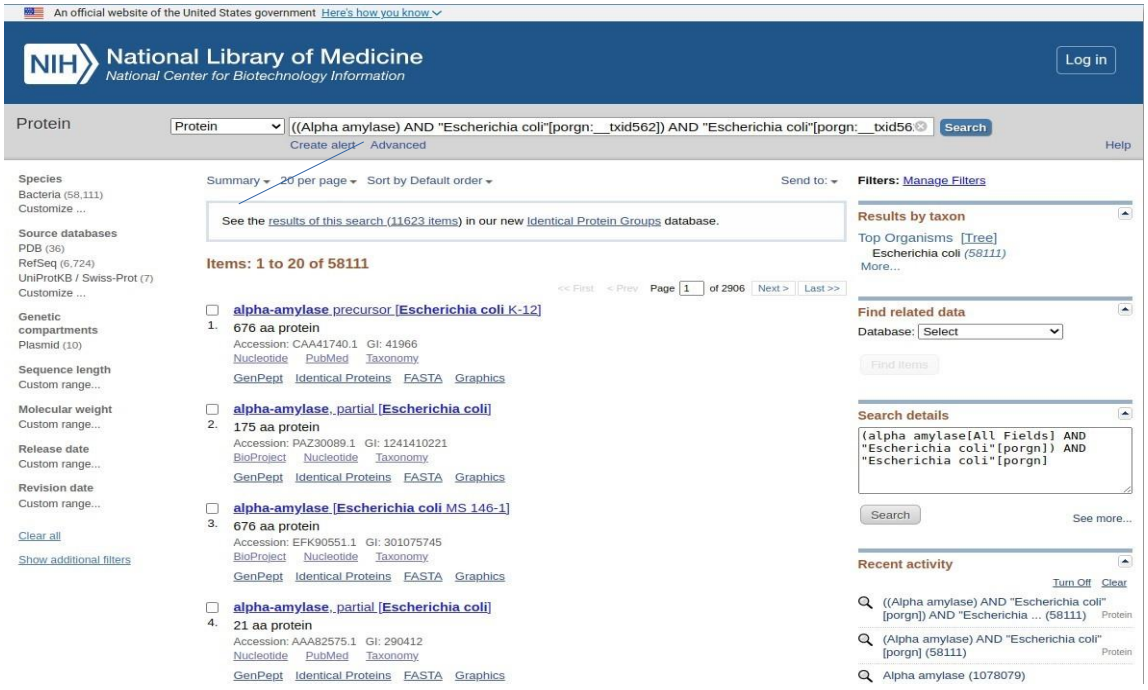
The EBI has a new phylogeny-aware multiple sequence alignment program which makes use of evolutionary information to help place insertions and deletions.

Steps

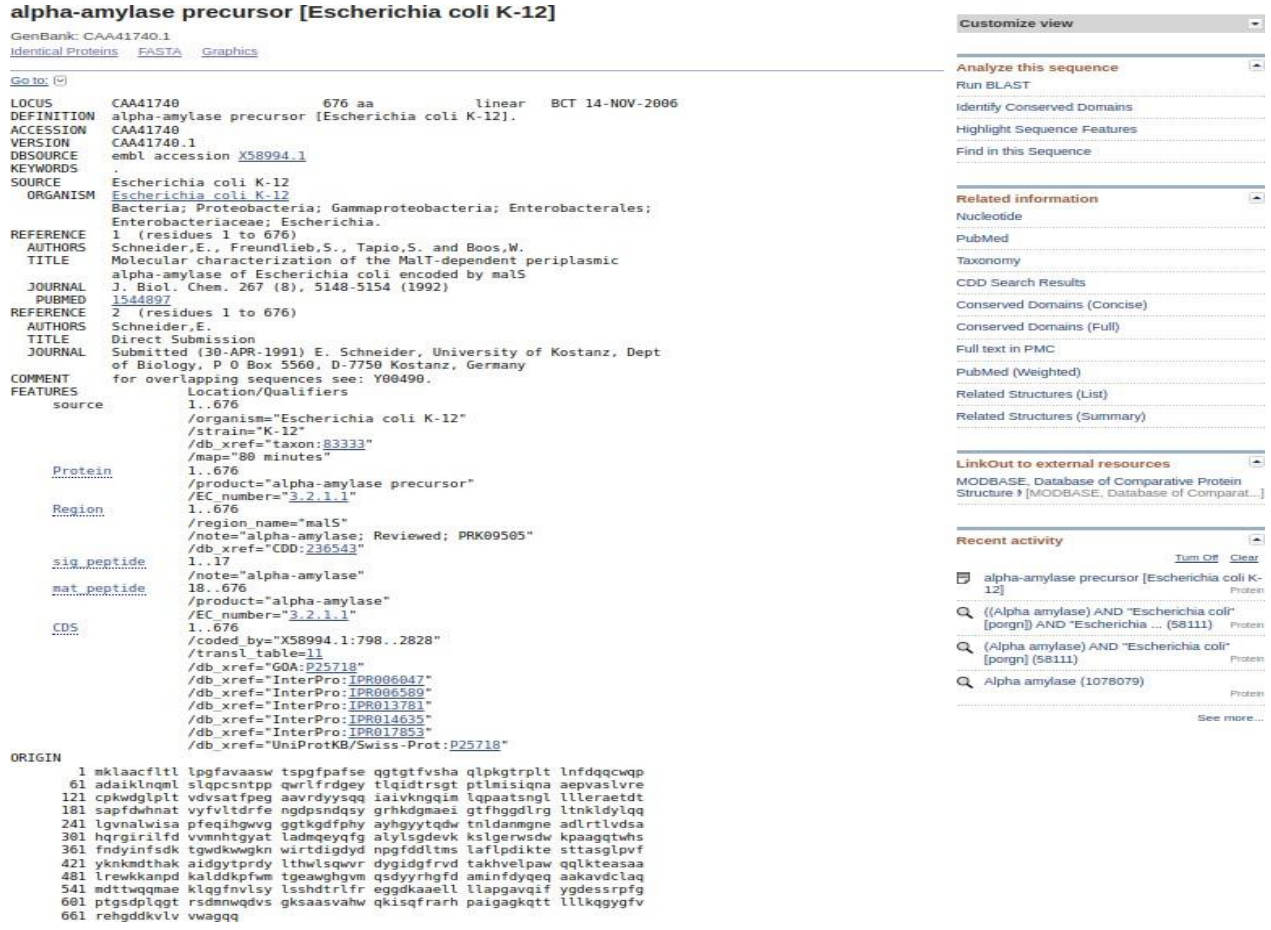
1. Search NCBI On Google .



2. Select Database to Search. Enter a Query sequence in the query box.



3.Click on FASTA to retrivre its FASTA seqence.



alpha-amylase precursor [Escherichia coli K-12]

GenBank: CAA41740.1

Identical Proteins FASTA Graphics

Go to:

```
LOCUS       CAA41740                676 aa            linear   BCT 14-NOV-2006
DEFINITION  alpha-amylase precursor [Escherichia coli K-12].
ACCESSION   CAA41740
VERSION     CAA41740.1
DBSOURCE    embl accession X58994.1
KEYWORDS    .
SOURCE      Escherichia coli K-12
  ORGANISM  Escherichia coli K-12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacterales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (residues 1 to 676)
  AUTHORS   Schneider,E., Freundlieb,S., Tapio,S. and Boos,W.
  TITLE     Molecular characterization of the MalT-dependent periplasmic
            alpha-amylase of Escherichia coli encoded by malS
  JOURNAL   J. Biol. Chem. 267 (8), 5148-5154 (1992)
   PUBMED   1544897
REFERENCE   2  (residues 1 to 676)
  AUTHORS   Schneider,E.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1991) E. Schneider, University of Kostanz, Dept
            of Biology, P O Box 5560, D-7750 Kostanz, Germany
COMMENT     for overlapping sequences see: Y00490.
FEATURES             Location/Qualifiers
     source          1..676
                     /organism="Escherichia coli K-12"
                     /strain="K-12"
                     /db_xref="taxon:83333"
                     /map="80 minutes"
     Protein         1..676
                     /product="alpha-amylase precursor"
                     /EC_number="3.2.1.1"
     Region          1..676
                     /region_name="malS"
                     /note="alpha-amylase; Reviewed; PRK09505"
                     /db_xref="CDD:236543"
     sig_peptide     1..17
                     /note="alpha-amylase"
     mat_peptide     18..676
                     /product="alpha-amylase"
                     /EC_number="3.2.1.1"
     CDS             1..676
                     /coded_by="X58994.1:798..2828"
                     /transl_table=11
                     /db_xref="GOA:P25718"
                     /db_xref="InterPro:IPR006047"
                     /db_xref="InterPro:IPR006589"
                     /db_xref="InterPro:IPR013781"
                     /db_xref="InterPro:IPR014635"
                     /db_xref="InterPro:IPR017853"
                     /db_xref="UniProtKB/Swiss-Prot:P25718"
ORIGIN
        1 mklaacfltl lpgfavaasw tspgfpafse qgtgtfvsha qlpkgtrplt lnfdqqcwqp
       61 adaiklnqml slqpcsntpp qwrlfrdgey tlqidtrsgt ptlmisiqna aepvaslvre
      121 cpkwdglplt vdvsatfpeg aavrdyysqq iaivkngqim lqpaatsngl llleraetdt
      181 sapfdwhnat vyfvltdrfe ngdpsndqsy grhkdgmaei gtfhggdlrg ltnkldylqq
      241 lgvnalwisa pfeqihgwvg ggtkgdfphy ayhgyytqdw tnldanmgne adlrtlvdsa
      301 hqrgirilfd vvmnhtgyat ladmqeyqfg alylsgdevk kslgerwsdw kpaagqtwhs
      361 fndyinfsdk tgwdkwwgkn wirtdigdyd npgfddltms laflpdikte sttasglpvf
      421 yknkmdthak aidgytprdy lthwlsqwvr dygidgfrvd takhvelpaw qqlkteasaa
      481 lrewkkanpd kalddkpfwm tgeawghgvm qsdyyrhgfd aminfdyqeq aakavdclaq
      541 mdttwqqmae klqgfnvlsy lsshdtrlfr eggdkaaell llapgavqif ygdessrpfg
      601 ptgsdplqgt rsdmnwqdvs gksaasvahw qkisqfrarh paigagkqtt lllkqgygfv
      661 rehgddkvlv vwagqq
```

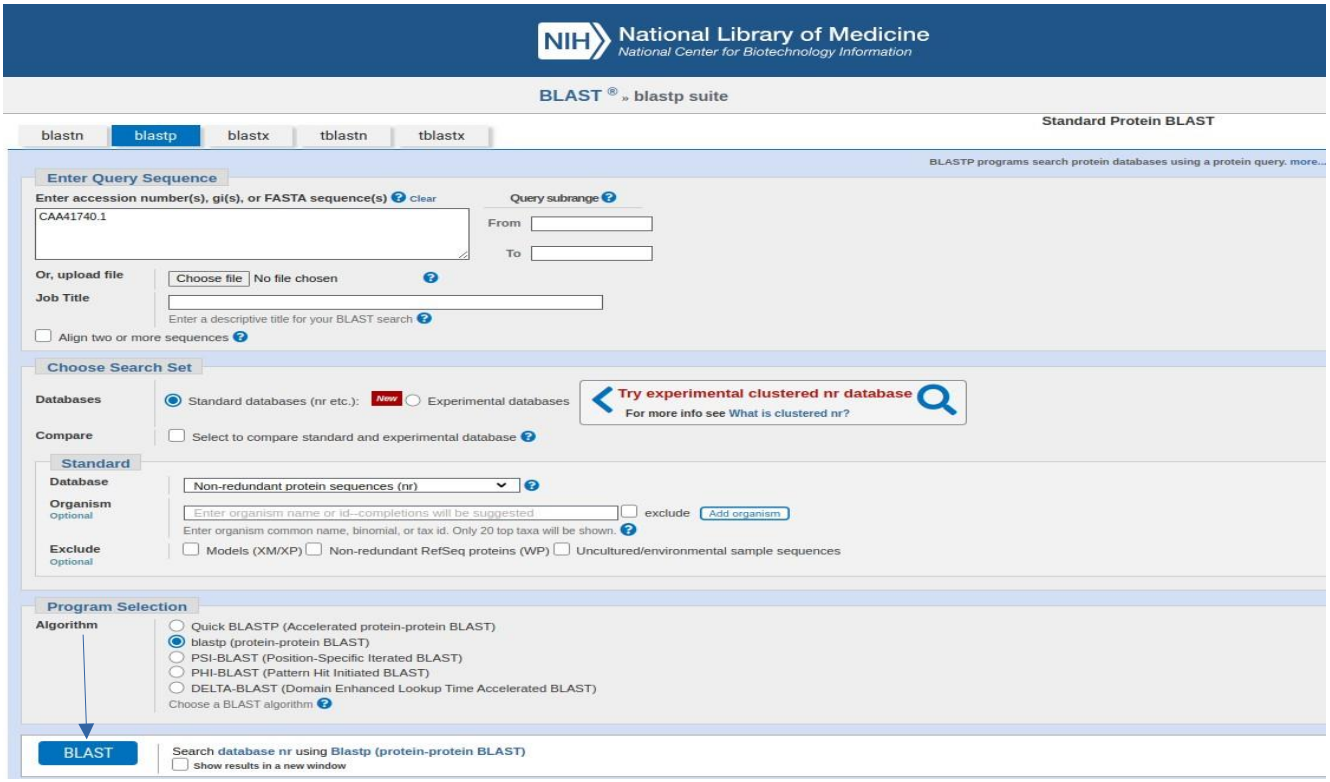
alpha-amylase precursor [Escherichia coli K-12]

GenBank: CAA41740.1

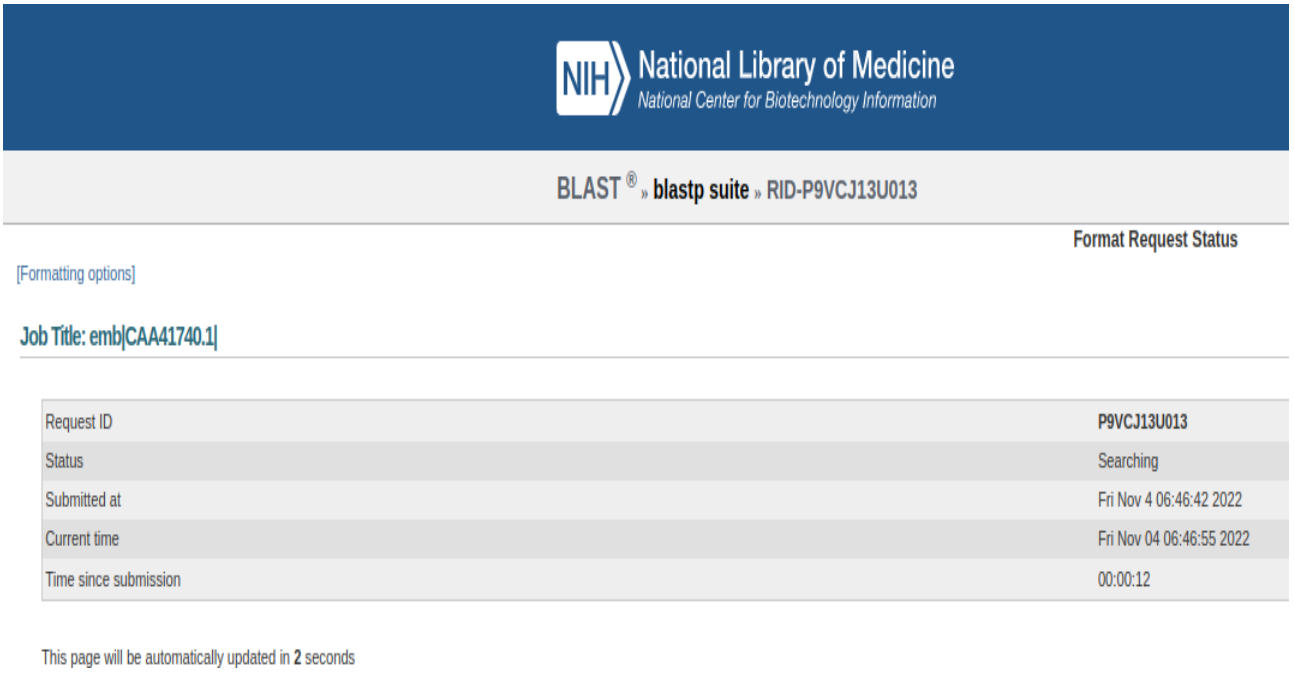
GenPept Identical Proteins Graphics

```
>CAA41740.1 alpha-amylase precursor [Escherichia coli K-12]
MKLAACFLTLLPGFAVAASWTSPGFPAFSEQGTGTFVSHAQLPKGTRPLTLNFDQQCWQPADAIKLNQML
SLQPCSNTPPQWRLFRDGEYTLQIDTRSGTPTLMISIQNAAEPVASLVRECPKWDGLPLTVDVSATFPEG
AAVRDYYSQQIAIVKNGQIMLQPAATSNGLLLLERAETDTSAPFDWHNATVYFVLTDRFENGDPSNDQSY
GRHKDGMAEIGTFHGGDLRGLTNKLDYLQQLGVNALWISAPFEQIHGWVGGGTKGDFPHYAYHGYYTQDW
TNLDANMGNEADLRTLVDSAHQRGIRILFDVVMNHTGYATLADMQEYQFGALYLSGDEVKKSLGERWSDW
KPAAGQTWHSFNDYINFSDKTGWDKWWGKNWIRTDIGDYDNPGFDDLTMSLAFLPDIKTESTTASGLPVF
YKNKMDTHAKAIDGYTPRDYLTHWLSQWVRDYGIDGFRVDTAKHVELPAWQQLKTEASAALREWKKANPD
KALDDKPFWMTGEAWGHGVMQSDYYRHGFDAMINFDYQEQAAKAVDCLAQMDTTWQQMAEKLQGFNVLSY
LSSHDTRLFREGGDKAAELLLLAPGAVQIFYGDESSRPFGPTGSDPLQGTRSDMNWQDVSGKSAASVAHW
QKISQFRARHPAIGAGKQTTLLLKQGYGFVREHGDDKVLVVWAGQQ
```

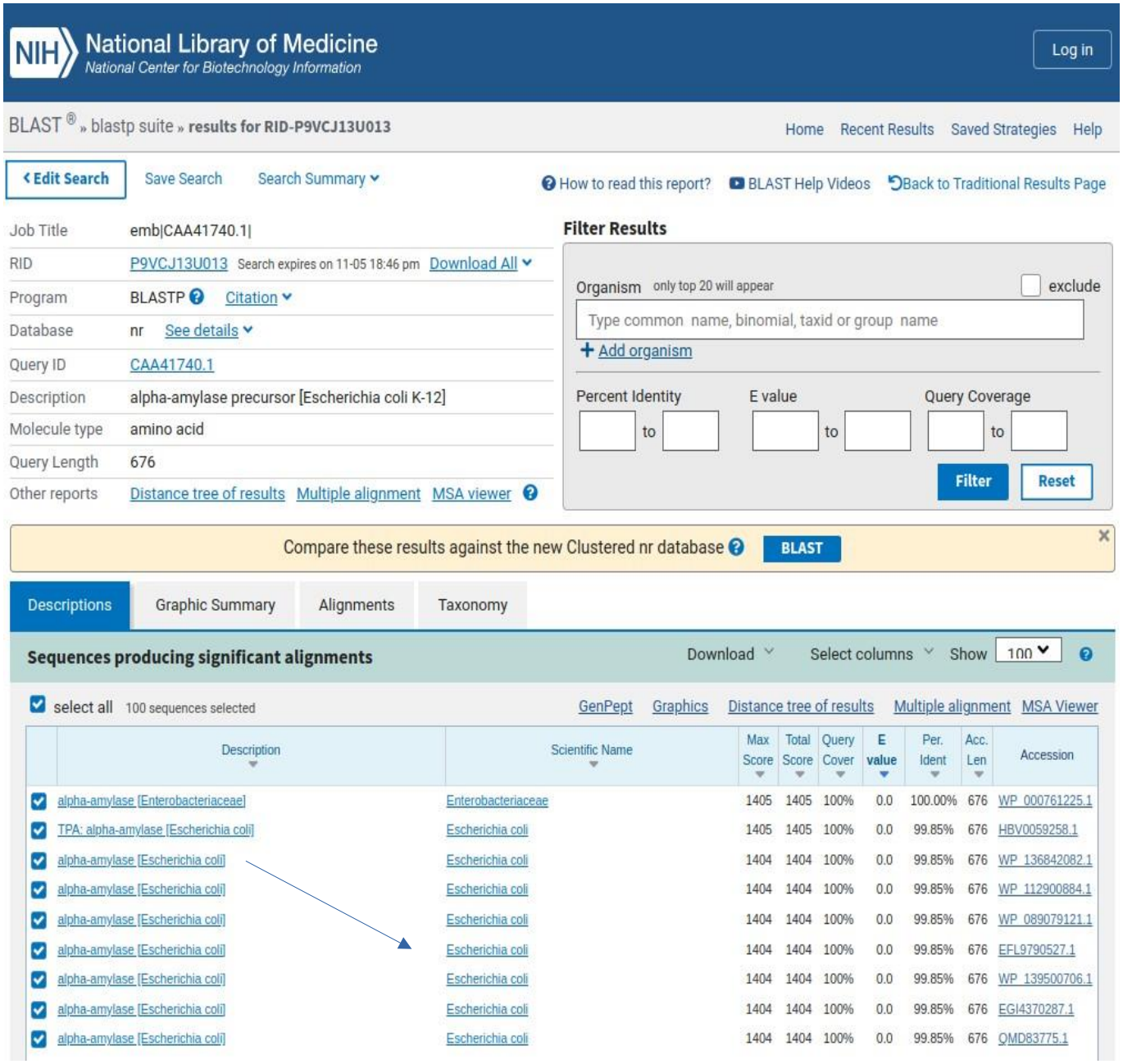
4.Run BLAST. Click on BLAST.



5.This screen will come up next . Finally (after a lenghty wait), a new window will appears showing any hits your seqence made . The result will be clour coded and annotated.



6. Select 10 seqences and download it. Download themultiple sequence alignment.



7. Search Clustal Omega and paste query sequence in query box.

Clustal Omega Result

```
CLUSTAL O(1.2.4) multiple sequence alignment


CAD6234996.1      ------------------------------------------------------------  0
CAD6235021.1      ------------------------------------------------------------  0
KAI3434954.1      ----------------------------------------MRPAITEALLAVLVCLVVG  19
GAX78569.1        ------------------------------------------------------------  0
KAF8065798.1      ----MAAGGRSAA-------LLALLWLGC-------SLMLLLPPARAARAPGELSSQLLL  42
KXZ42036.1        ----------------------------------------MFSRRRESSAEELLREVLS  19
KAG2493000.1      MVDRMQRGARQRQRACMRLALVTLLIAHQASALRLPSWGNLLRPKREKRDPEDILREVLS  60


CAD6234996.1      ------------------------------------------------------------  0
CAD6235021.1      ------------------------------------------------------------  0
KAI3434954.1      ANGATTAAQHRLAADRGLRLD-----------SSPNDTAAD---LAAWRAQF-EALPPSP  64
GAX78569.1        ------------------------------------------------------------  0
KAF8065798.1      SSGMHLLS---DQ----------------------SSLAPEGLLGLSAAQRASPLLANP  76
KXZ42036.1        PRSAYHAE---LAAGFGRSPAGGTAGTSAEGGATPLESAAAEAVLGSRHSAGRSIPLHDP  76
KAG2493000.1      PKSTVLAS---LHSSKGPRDAG----------ASPKASAQLQGSIALGDPAARDVCPSDP  107


CAD6234996.1      ------------------------------------------------------------  0
CAD6235021.1      ------------------------------------------------------------  0
KAI3434954.1      WWQRFW----PWRQPGLEPQHLVRVAPMRPDGGLAAPAELRLPAAALEAELGAGLAPAVN  120
GAX78569.1        ----------------------------------------MHGVEFLFQGRNE-GTPEVP  19
KAF8065798.1      ATRSVFAADSALARTATARDEAWR--------------------------RGGDAAAGRP  110
KXZ42036.1        AIATVFSRGASRRLPRLARGPSVSLSPL---PGQ---P-----------------------  108
KAG2493000.1      AIATIFSSQAARTLPRLARGPSVTLGGA---GDA---AHAAHGASSLAASKGTDFNPDRA  161


CAD6234996.1      ----------MGRRGLLTAALLLCFLAVCSGRELTIKQNPSTTIYNSTLAQILVEYAAA  49
CAD6235021.1      ----------MGRRGLLTAALLLCFLAVCSGRELTIKQNPSTTIYNSTLAQILVEYAAA  49
KAI3434954.1      AADPRDDSFYAYNSSAM-T-PLVPLWLSWMT---DAERMANRSHGLDFAVAWRLSSYVAI  175
GAX78569.1        AKLSED--LKASGLDTIATVFQRDIATRIIS---NSGSTDADWDNFDMDLAKTFALFTSI  74
KAF8065798.1      AA---------AGGG---R-AARRRGRGRS----------AALRGGFDLDVSATLAALQSI  148
KXZ42036.1        ------------------QPAPRPDGEHRSG---RIGAADPPPPELDVGEALVAGRLASI  147
KAG2493000.1      GV---------SGGQG--HGSGQPHTEPDSE---PLEAADPDPPDLDLDEAMLAGRLASI  207
                                                                 :   :     :


CAD6234996.1      IYTADLTELFTWTCARCGDLIKGFEMIEIIVDVENCLQAYVGFASDINAVVVVFRGTQEN  109
CAD6235021.1      IYTADLTELFTWTCARCGDLIKGFEMIEIIVDVENCLQAYVGFASDINAVVVVFRGTQEN  109
KAI3434954.1      SYCN-VSSIAGWNCTRCGGISADFTPEQVVFDSAWDLLGFVGYSEPLDAIVVAFRGTDSH  234
GAX78569.1        AYCSNTTTIKAWKCSRCQGVP-TFQVTDVVFDLKWDLIGYVGYWPERNAIVVAFRGSDSH  133
KAF8065798.1      SYCANLSDVAAWNCTRCARIP-NFQPSLVHFDEAWDLLGFAGFWPEVAAKVLVFRGTDSS  207
KXZ42036.1        AYCSRPDLISAWNCTRCADVP-GFTPYATVYDAIWDLAAYAGYYKPWNAIVLAFRGTDSS  206
KAG2493000.1      AYCAHPETLQSWNCTRCAEVP-GFVPHTVAFDSVWDLSAFAGYHKPSDTVWLSFRGTDSS  266
                   *      :  *.*:**  :   *       *    * .:.*:     :  : ***::.


CAD6234996.1      SIQNWIEDLLWKQLD--LDYPGMPEAM-VHRGFYAAYHNTTLRDGVVSGIQKTREAYGDI  166
CAD6235021.1      SIQNWIEDLLWKQLD--LDYPGMPEAM-VHRGFYAAYHNTTLRDGVVSGIQKTREAYGDI  166
KAI3434954.1      SIYNWVQNMRTWRTDLALGYPNAPMGALVHGGFFTSYNSSALASNITAGVTALRRRHPGS  294
GAX78569.1        SWTNWIENLRTWREDGMWPLPDAPNVK-VHAGFYNLWSKSDMASNITAAIQSVLVQHPDA  192
KAF8065798.1      SWYNWAENMRAWRTDASYPVPGAPRSLRIHSGFLLMWNASSMAATFHAAYAELRKAHPVG  267
KXZ42036.1        NWGQWAENMRYWRTDTMFPEPGFPHAL-IHSGFWTLWTGSRLQASFKDAVAELLAAHPGA  265
KAG2493000.1      NWSQWAENMRAWRTDHMYPVPNFPHAL-IHAGFYTLWTGSSLQKTFTDAVVELMAAHPRS  325
                  .  :* :::   : *     *. *    :* **   :  : :   .  .       :


CAD6234996.1      PIMITGHSMGGAMASFCALDLVVNYGLDGVKLMTFGQPRIGNAAFASYFKTYLPQAIRVT  226
CAD6235021.1      PIMITGHSMGGAMASFCALDLVVNYGLDGVKLMTFGQPRIGNAAFASYFKTYLPQAIRVT  226
KAI3434954.1      PVYVSGHSLGGAMATLCALDLRLNLKMPDVRVYTFGSPRVGNLIFAQWFEGVVQVHWRFT  354
GAX78569.1        SLYVTGHSMGAALAQICAVDSKIKLGVAQVHSFTFGSPRVGNAAFHVFYEGLMAESWRFT  252
KAF8065798.1      PTYVLGHSMGGALAHLAALDLAVTFGATDVRAVTFGSPRVGNAPFQEFFDGTVSQSWRFT  327
KXZ42036.1        RLVATGHSMGGALAQLAGLEAKLAHNDTHTSVWTFGAPRLGNLAYQQLFNQFVDVSWRFT  325
KAG2493000.1      RMVVTGHSMGGALAQLAGLEFKLSYNRTLTTVYTWGSPRVGNLAYAQLFSTFVDVSWRFT  385
                        ***:*.*:* :..::  :      .   *:* **:**  :   :.  :    *.*


CAD6234996.1      HAHDIVPHLPPYFAFFPQKTYHHFPREVWVHNVGLGTLV---YSAEKICDDSGEDPTCC-  282
CAD6235021.1      HAHDIVPHLPPYFAFFPQKTYHHFPREVWVHNVGLGTLV---YSAEKICDDSGEDPTCC-  282
KAI3434954.1      HNRDIVPSVPPGYM-----GFYHVAREVWVVDLLYG------HTLVGVCDASGEDMACHN  403
GAX78569.1        HNRDVVPSLPPWFL-----GFHHVAREVWTIDVEPEIPGSPPEFIALMCDTTGEDPSCHN  307
KAF8065798.1      HGRDIVPSVPPQLM-----GFRHVAREVWLVDVDGPRPGDPPAQRVVLCDDSGEDPSCHN  382
KXZ42036.1        HNRDIVPSVPLQLM-----GFQHVAREVWEVDVEDPSKAGGAERKLLLCDASGEDPSCHN  380
KAG2493000.1      HNRDIVPSVPLQLM-----GFQHVAREVWEVDVDDPAGGPPRDRKLLLCDGTGEDPSCHN  440
                  * :*:** :*          : *. ****  ::              :** :*** :*


CAD6234996.1      ---RSVIGNSVQDHIYYLGVSMHAEDWSSCRIVMDYSRLQYQMDLKGNLVLSKQPGLSND  339
CAD6235021.1      ---RSVIGNSVRDHIYYLGVSMHAEDWSSCRIVMDYSRLQYQMDLKVCKYKCRHLTNASE  339
KAI3434954.1      SMCHLGLCSSLADHLLYLSEMYSPHP-MGC-----------------------------  432
GAX78569.1        SACRLGLCTSIADHLTYLQHPMYRTE-GECSI---------------------------  338
KAF8065798.1      SACRLGLCTSVADHLTYLGAHMYRG--GEC-----------------------------  410
KXZ42036.1        SACYLGLCTSVADHLVYLGLHMYQDS-DEC-----------------------------  409
KAG2493000.1      SACYLGLCTSVADHLLYLGIPMYGDE-GEC-----------------------------  469
                        : .*: **: **           *


CAD6234996.1      RG----------FSAQ-----  345
CAD6235021.1      KSTVRPIWHEQQYNSEHSSFL  360
KAI3434954.1      ---------------------  432
GAX78569.1        ---------------------  338
KAF8065798.1      ---------------------  410
KXZ42036.1        ---------------------  409
KAG2493000.1      ---------------------  469
```

Result Of Multiple Sequence Alignment With 10Sequences-

KAF0245828.1:1-99 nucleotide [Planctomycetes bacterium]
MTIREFQSCIRAQYFTRDSARGVGGTFLWFVEEVGELAEALRHGTDEEKRA
EFADVLAWLSTLASMAGVDLEDA AAGKYG
KGCPRCGGTPCRCAEQGGP
>HED66435.1:1-79 nucleotide pyrophosphohydrolase [Planctomycetes bacterium]
MHIDEFQRLIEAIYFEKDNARGLAGTHMWFCEEVGELTRALRRGQTEELEG
EFADVLAWLSTLASIAGIDLEEAAR KKY
>MCE2725385.1:1-79 nucleotide pyrophosphohydrolase [Planctomycetaceae bacterium]
MTLREFQTLIRTMYHDKDAARGVEGTFMWLTEEIGELATALRSGTHEERAL
EFADVLAWLATIANVVDVDLEEA
VGRKY>HHI00078.1:1-79 nucleotide pyrophosphohydrolase [Thermococcus litoralis]
MEIREFQEMIRDIYFHKDSKRGVERTFLWFVEEIGELSEAIRKNDREAMEEEF
ADVLAWLASLANLVGVDLEEAAK KKY
>MBL6710039.1:1-79 nucleotide pyrophosphohydrolase [Planctomycetes bacterium]
MTLREFQQLIRNMYHDKDAARGVEGTFMWLTEEIGELATALRSGSPEEQAL
EFADVLAWLATIANVVDVDLEQ AIAKKY
>MBI1917951.1:3-81 nucleotide pyrophosphohydrolase [Planctomycetes bacterium]
ITLPQFQQLIRDLYGAKDSRRGVEGTFMWFMEEVGELSAALRGGTHEERVA
EFADVLAWLATLANLAGVDLESA VRAKY
>HDZ35607.1:1-79 nucleotide pyrophosphohydrolase [Thermococcus sp.]
MEIREFQGMIKDIYFHKDSKRGVNRTFLWFVEEIGELSEAIRKNDREAMEEE
FADVLAWLASLANLVGVDLEEAA KKKY
>MBL8829589.1:7-84 nucleotide pyrophosphohydrolase [Planctomycetaceae bacterium]
SIAGFQRLIRDMYFAKDQARGVDGTFMWFMEEVGELASCLRHGTHEERLG
EFADVLAWLTTMANVVGVDLTE AVRLKY
>NCO37222.1:1-74 nucleotide pyrophosphohydrolase [Armatimonadetes bacterium]
MAIREFQQLIEAIYFAKDSGRGVESTFMWFVEEVGELAQALRSGTVEQLEEE
FADVAAWLATLASLKGIDLETA
>MCP4193264.1:17-115 nucleotide pyrophosphohydrolase [Planctomycetaceae bacterium]
VTLADFQSLIRQMYFDKDKARGVDGTFMWLMEEVGELATALREGSAEDRS
EEFADVLAWLTTIANVSGVDLTEA VHRKYGAGCPGCQQFVCCCPDQGKP

APPLICATIONS OF MSA

1. Detecting similarities between sequences(closely or distinctly related).
2. Detecting conserved regions or motifs in sequences
3. Detecting of structural homologies.
4. Thus,assisting the improved prediction of secondary and tertiary structures of proteins.

ADVANTAGES OF MSA

Multiple sequence alignment has assumed a key role in comparative structure and function analysis of biological sequences.It often leads to fundamental biological insight into sequence-structure-function relationships of nucleotide or protein sequence families.Significant advances have been achieved in this field,and many useful tools have been developed for constructing alignments.It should be stressed however, that many complex biological and methodological issues are still open.It provides some background informations and considerations associated with MSA techniques,concentrating on the alignment of protein sequences.

CONCLUSION

This review is an attempt to put in context and cover the developments that have taken place in the field of multiple sequence alignments over the last decade or so far.The unprecedented pace of development makes it difficult to be truly exhaustive. We have nonetheless tried to provide the reader with an overview of the main aspects, and how they connect to one another. As shown in the figure the progressive alignment framework is the main algorithmic heuristic that has been adoped by almost all existing alignment methods.Further we can observe a clear clustering of the methods based on the type of sequences they are designed to align(RNA,DNA/Genomes or Proteins).It is also worth nothing that the current inflation in the number of available methods merely reflects the growing pace of data accumulation.MSA modeling in one of the most powerful ways to make sense of biological sequences.MSA methods, by their very approximate nature,are doomed to follow a red-queen evolutionary strategy and will need to keep evolving, faster and faster, in order to keep up with the processing of standard biological data.

REFERENCES

1. https://en.m.wikipedia.org
2. https://pubmed.ncbi.nlm.nih.gov
3. https://www.slideshare.net
4. https://www.ebi.ac.
5. https://pubmed.ncbi.nlm.nih.gov