- **Uploading your assignment as a PDF is mandatory. The system will only accept a single PDF file and won't accept anything else or more than one file.**

- [BLAST: Program Selection Guide](#)
    - Section 6.1 will be especially helpful.

- [BLAST: Frequently Asked Questions](#)

- You will have to add screenshots to your homework in this assignment. Remember the screenshot keyboard shortcuts. Feel free to use MS Powerpoint/GoogleSlides in your homework.

1) Use bl2seq to align **NM_000558.3** to **NP_000549.1**
    a) Which BLAST program did you choose? Why?
    b) What is the sequence type in the final alignment?
    c) What are the significance value and the total score of the alignment? Define them, and comment on the result you get.

2) Use bl2seq to align **NP_000509.1** to **NP_000549.1**

    a) Which BLAST program did you choose? Why?

    b) What is the sequence type in the final alignment?

    c) What are the significance value and the total score of the alignment? comment on the result you get.

3) Use Lalign to align **NM_000558.3** to **AF230076.1**

    a) What type of pairwise alignment approach Lalign algorithm is based on?

    b) Try **local, global** and **global with no end-gap penalty** alignments separately. Compare the results. What kind of differences do you observe?

        i)   Do you see any differences at the termini of the alignment?

        ii)   Do you get any errors or do all methods work?

4) Theoretical:

    a) What are the differences between **local, global,** and **global with no end-gap penalty** alignment methods?

    b) Which alignment gives biologically more relevant information (amongst **local** and **global** alignment)? Explain your reasoning.

Nucleotide

[Nucleotide ▾] [   ] [Search]

Advanced

FASTA ▾   Send to: ▾

Change region shown

Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA

NCBI Reference Sequence: NM_000558.3

Customize view

⚠ **This sequence has been updated. See current version.**

Analyze this sequence

GenBank   Graphics

Run BLAST

```
>NM_000558.3 Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA
ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTC
AAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTCC
TGTCCTTCCCCACCACCAAGACCTACTTCCCGCACTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGG
CCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTG
TCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCACT
GCCTGCTGGTGACCCTGGCCGCCCACCTCCCCGCCGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAA
GTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGGTGGCCATGCTT
CTTGCCCCTTGGGCCTCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGTGGTCTTTGAATAA
AGTCTGAGTGGGCGGC
```

Pick Primers

Highlight Sequence Features

Find in this Sequence

Recent activity

Turn Off

📄 HBA1 hemoglobin subunit alpha 1 [Ho
sapiens]

Protein   Protein ⌄ | Search

Advanced   Help

FASTA ⌄   Send to: ⌄   Change region shown   ⌄

hemoglobin subunit alpha [Homo sapiens]

NCBI Reference Sequence: NP_000549.1

GenPept   Identical Proteins   Graphics

```
>NP_000549.1 hemoglobin subunit alpha [Homo sapiens]
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSK
YR
```

Analyze this sequence   ▲

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Show in Genome Data Viewer

Protein 3D Structure   ▲

THE CRYSTAL
STRUCTURE OF HUMAN

1.bl2seq : BLASTX search protein subjects using a translated nucleotide query.

‹ **Edit Search**   Save Search   Search Summary ⌄   ❓ How to read this report?   ▶ BLAST Help Videos   ↺ Back to Traditional Results Page

Job Title	**NM_000558.3 Homo sapiens hemoglobin, alpha**
RID	STU06ZER114 *Search expires on 11-12 16:53 pm* Download All ⌄
Program	Blast 2 sequences Citation ⌄
Query ID	lcl|Query_13615 (dna)
Query Descr	NM_000558.3 Homo sapiens hemoglobin, alpha 1 (HBA ...
Query Length	576
Subject ID	lcl|Query_13617 (amino acid)
Subject Descr	NP_000549.1 hemoglobin subunit alpha [Homo sapiens]
Subject Length	142
Other reports	❓

Filter Results

Percent Identity		**E value**		**Query Coverage**	
[ ] to [ ]		[ ] to [ ]		[ ] to [ ]	

Filter   **Reset**

Descriptions	Graphic Summary	**Alignments**	Dot Plot

Other reports 🔵

Sequences producing significant alignments

Download ⌄   🆕 **Select columns** ⌄   Show [100 ⌄]

☑ select all   *1 sequences selected*

Graphics

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Access
☑	NP_000549.1 hemoglobin subunit alpha [Homo sapiens]		286	286	73%	2e-105	100.00%	142	Query_18(

how much of the query covered by targeted sequence

max score: highest alignment score
total score:sum of all alignment scores
no difference between total and max. score,
this means there is no gaps between the
alignment

Other reports ⓘ

| Descriptions | Graphic Summary | Alignments | Dot Plot |

Sequences producing significant alignments

Download ▾   🆕 Select columns ▾   Show 100 ▾

☑ select all   *1 sequences selected*   Graphics

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑ NP_000549.1 hemoglobin subunit alpha [Homo sapiens]		286	286	73%	2e-105	100.00%	142	Query_180-

E-values are just like p-values to evaluate
the statistical similarity but BLAST uses
E-value: significance value

how similar 2
sequences

final alignment: protein sequence

Alignment view [Pairwise ▾]   ❓ **Restore defaults**

1 sequences selected ❓

⬇ **Download** ▾   **Graphics**   ▼ Next ▲ Previous ◄

NP_000549.1 hemoglobin subunit alpha [Homo sapiens]

Sequence ID: **Query_180493**   Length: **142**   Number of Matches: **1**

Range 1: 1 to 142 Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps	Frame
286 bits(732)	2e-105	Compositional matrix adjust.	142/142(100%)	142/142(100%)	0/142(0%)	+2

```
Query  38    MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  217
             MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
Sbjct  1     MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  60

Query  218   KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP  397
             KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
Sbjct  61    KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP  120

Query  398   AVHASLDKFLASVSTVLTSKYR  463
             AVHASLDKFLASVSTVLTSKYR
Sbjct  121   AVHASLDKFLASVSTVLTSKYR  142
```

graphical summary, it means the resemblances between the sequences are very high

Descriptions	**Graphic Summary**	Alignments	Dot Plot

🖐 *hover to see the title* ➤ *click to show alignments*   Alignment Scores ■ < 40 ■ 40 - 50 ■ 50 - 80 ■ 80 - 200 ■ >= 200 ❓

1 sequences selected ❓

Distribution of the top 1 Blast Hits on 1 subject sequences

Query

| 1 | 100 | 200 | 300 | 400 | 500 |

matched sequences as dot plot:

Plot of lcl|Query_180491 vs lcl|Query_180493
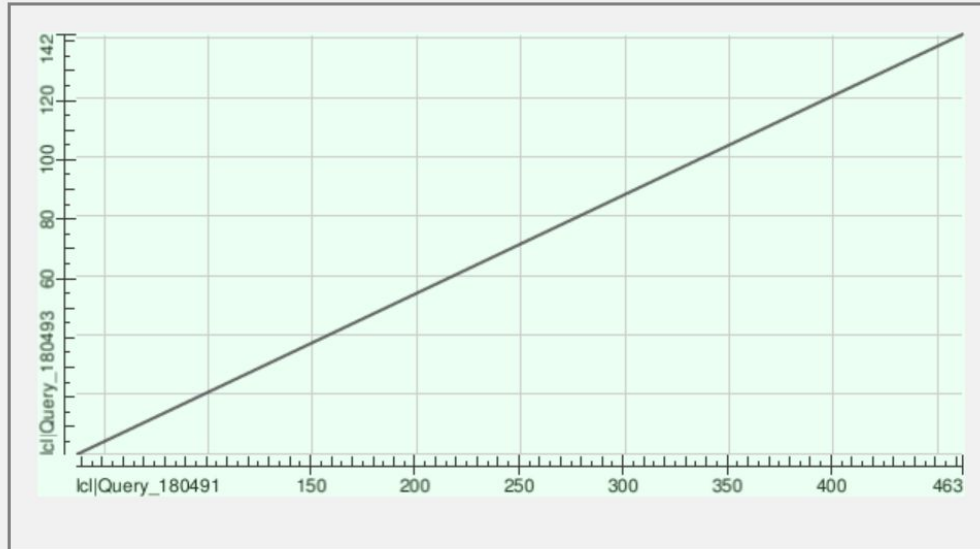
2.bl2seq: protein to protein

FASTA ▾

Send to

hemoglobin subunit alpha [Homo sapiens]

NCBI Reference Sequence: NP_000549.1

GenPept   Identical Proteins   Graphics

```
>NP_000549.1 hemoglobin subunit alpha [Homo sapiens]
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNA
VAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSK
YR
```

Blastp was used : protein to protein

COVID-19 Information

Public health information (CDC) | Research information (NIH) | SARS-CoV-2 data (NCBI) | Prevention and treatment information (HHS) | Español

BLAST ® » blastp suite-2sequences » results for RID-STVD02VC11N

‹ Edit Search   Save Search   Search Summary ❯   ❓ How to read this report?   ▶ BLAST Help Videos   ↻ Back to Traditional Results Page

Job Title	NP_000509.1 hemoglobin subunit beta [Homo...
RID	STVD02VC11N Search expires on 11-12 17:17 pm
	Download All ❯
Program	Blast 2 sequences Citation ❯
Query ID	lcl|Query_382385 (amino acid)
Query Descr	NP_000509.1 hemoglobin subunit beta [Homo sapiens]
Query Length	147
Subject ID	lcl|Query_382387 (amino acid)
Subject Descr	NP_000549.1 hemoglobin subunit alpha [Homo sapiens]
Subject Length	142

Filter Results

Percent Identity		E value		Query Coverage	
___	to ___	___	to ___	___	to ___

Filter   Reset

Multiple alignment   MSA viewer ❓

https://combatcovid.hhs.gov

📧 Feedback

13:17

Multiple alignment   MSA Viewer ❶

Descriptions	Graphic Summary	Alignments	Dot Plot

Sequences producing significant alignments

Download ⌄   **New** Select columns ⌄   Show [100 ▾]   ❓

☑ select all   *1 sequences selected*   Graphics   Multiple alignment   **New** MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	NP_000549.1 hemoglobin subunit alpha [Homo sapiens]		114	114	97%	2e-38	43.45%	142	Query_382387

high resemblances (but not higher as NM_000558.3&NP_000549.1 query)

final alignment as protein seq.

⬇ **Download** ⌄   **Graphics**   ▼ Next ▲ Previous ◄ Des

NP_000549.1 hemoglobin subunit alpha [Homo sapiens]

Sequence ID: **Query_382387**   Length: **142**   Number of Matches: **1**

Range 1: 3 to 141 Graphics   ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
114 bits(286)	2e-38	Compositional matrix adjust.	63/145(43%)	88/145(60%)	8/145(5%)

```
Query  4    LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV   61
            L+P +K+ V A WGKV  +  E G EAL R+ + +P T+ +F  F DLS       G+ +V
Sbjct  3    LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSAQV   56

Query  62   KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK   121
            K HGKKV  A ++ +AH+D++    + LS+LH  KL VDP NF+LL + L+  LA H
Sbjct  57   KGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPA   116

Query  122  EFTPPVQAAYQKVVAGVANALAHKY   146
            EFTP V A+  K +A V+  L  KY
Sbjct  117  EFTPAVHASLDKFLASVSTVLTSKY   141
```

dot plot: not straightly linear as NM_000558.3&NP_000549.1 query, high resemblance but 3 deletions / mismatches between two sub-sequences among query & target sequences

3. Lalign: William Pearson's lalign program

Nucleotide

Nucleotide ▾ [   ] [Search]

Advanced   Help

FASTA ▾

Send to: ▾

Change region shown ▾

Customize view ▾

Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA

NCBI Reference Sequence: NM_000558.3

⚠ This sequence has been updated. See current version.

GenBank   Graphics

```
>NM_000558.3 Homo sapiens hemoglobin, alpha 1 (HBA1), mRNA
ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTC
AAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTCC
TGTCCTTCCCCACCACCAAGACCTACTTCCCGCACTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGG
CCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTG
TCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAACTTCAAGCTCCTAAGCCACT
GCCTGCTGGTGACCCTGGCCGCCCACCTCCCCGCCGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAA
GTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGGTGGCCATGCTT
CTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGTGGTCTTTGAATAA
AGTCTGAGTGGGCGGC
```

Analyze this sequence ▲

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Recent activity ▲

▤ Homo sapiens alpha-2-globin (HBA2) gene, complete cds   Nucleotide

FASTA ▾

Send to: ▾

Homo sapiens alpha-2-globin (HBA2) gene, complete cds

GenBank: AF230076.1

[GenBank](#)   [Graphics](#)

```
>AF230076.1 Homo sapiens alpha-2-globin (HBA2) gene, complete cds
CGCCCGGCCGGGCGTGCCCCCGCGCCCCAAGCATAAACCCTGGCGCGCTCGCGGCCCGGCACTCTTCTGG
TCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCT
GGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAAGTGAGGCTCCCTCCCCTG
CTCCGACCCGGGCTCCTCGCCCGCCCGGACCCACAGGCCACCCTCAACCGTCCTGGCCCCGGACCCAAAC
CCCACCCCTCACTCTGCTTCTCCCCGCAGGATGTTCCTGTCCTTCCCCACCACCAAGACCTACTTCCCGC
ACTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCAA
CGCCGTGGCGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTT
CGGGTGGACCCGGTCAACTTCAAGGTGAGCGGCGGGCCGGGAGCGATCTGGGTCGAGGGGCGAGATGGCG
CCTTCCTCTCAGGGCAGAGGATCACGCGGGTTGCGGGAGGTGTAGCGCAGGCGGCGGCTGCGGGCCTGGG
CCGCACTGACCCTCTTCTCTGCACAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCCACCTCCC
CGCCGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTGGCTTCTGTGAGCACCGTGCTGACC
TCCAAATACCGTTAAGCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTCCCAACGGGCCCTCC
TCCCCTCCTTGCACCGGCCCTTCCTGGTCTTTGAATAAAGTCTGAGTGGGCGGCAGCCTGTGTGTGCCTG
GGTTCTCTCTGTCCCGGAATGTGCCAACAA
```

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the HBA2 gene

Dual proteome-scale networks reveal cell-specific remodeling of the human inte [Cell. 2021]

Molecular and Hematological Analysis of Alpha- and Beta-Thal [Genet Test Mol Biomarkers. 2021]

Molecular Characterization and Hematological Aspects of Hb E-Myanmar [β; [Hemoglobin. 2020]

See all

Change region shown ▾

Customize view ▾

local:

results:

```
# bin/lalign36 -E 10.0 -f -12 -g -2 10196.1.seq 10196.2.seq -J -K 3
LALIGN finds non-overlapping local alignments
 version 36.3.5e Nov, 2012(preload8)
Please cite:
 X. Huang and W. Miller (1991) Adv. Appl. Math. 12:373-381

*** Warning - unrecognized residue at 7:I - 73
*** Warning - unrecognized residue at 8:E - 69
*** Warning - unrecognized residue at 12:E - 69
*** Warning - unrecognized residue at 15:L - 76
*** Warning - unrecognized residue at 17:I - 73
*** Warning - unrecognized residue at 20:L - 76
Parameters not available for: +5/-4: -12/-2
Query: 10196.1.seq
  1>>>unknown 611 bp - 607 nt
Library: 10196.2.seq
       967 residues in     1 sequences

Statistics: (shuffled [500]) MLE statistics: Lambda= 0.0539;  K=0.02506
 statistics sampled from 1 (1) to 500 sequences
Threshold: E() < 10 score: 135
Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010)
Parameters: +5/-4 matrix (5:-4), open/ext: -12/-2
 Scan time:  0.110

>>unknown 982 bp                                    (967 nt)
 Waterman-Eggert score: 2157;  172.9 bits; E(1) <  5.1e-47
66.7% identity (66.7% similar) in 835 nt overlap (32-607:88-921)

             40        50        60        70        80        90
unknow ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
unknow ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAA
         90        100       110       120       130       140

             100       110       120       130       140       150
unknow GACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
unknow GACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGA
         150       160       170       180       190       200
```

The E indicates the matching between the sequences, the score expands the treshold

Algorithm: Smith-Waterman

results:

scores for locally optimal alignments

local similarities

matching sequences

```
>>unknown 982 bp                                    (967 nt)
Waterman-Eggert score: 185;  19.7 bits; E(1) <  0.5
53.3% identity (53.3% similar) in 569 nt overlap (606-83:138-662)

              600       590       580       570       560
unknow CCGCCCACTCAGACTTTATTCAAAGACCACGGGGGTACGG-----GTGCAGGAAGGGGAG
       : ::: ::  ::::     :::: : :  : :::::: ::    : ::: :  :: :::
unknow CTGCCGACA-AGACCAACGTCAAGGCCGCCTGGGGTAAGGTCGGCGCGCACGCTGGCGAG
           140       150       160       170       180       190

              550       540       530       520       510       500
unknow GAGGGGCTGGGGGGAGGCCCAAGGGGCAAGAAGCATGGC---CACCGAGGCTCCAGCTTA
       : ::      : :::::::: :    ::::: :::  : ::   ::::: :
unknow TATGGT----GCGGAGGCCCTGG-----AGAAGTGAGGCTCCCTCCCCTGCTCCGACC--
          200       210       220       230       240

              490       480       470       460       450       440
unknow ACGGTATTTGGAGGTCAGCACGGTGCTCACAGAAGCCAGG---AACTTGTCCAGGGAGGC
       :::  :    : : :: :::  : :::::  ::::    ::: :::: ::
unknow -CGGGCTCCTC--GCCCGCCCGGACC-CACAG--GCCACCCTCAACC-GTCCTGGCCCCG
           250       260       270       280       290

                 430       420            410       400
unknow GTGC---ACCGCAGGGGTGAACTCGGCGG---------GGAGGTGGGCGGCCAGGGTCAC
        : :   ::: ::  : : :: ::      :: ::  :: :::   :::
unknow GACCCAAACCCCACCCCTCA-CTCTGCTTCTCCCCGCAGGATGTTCCTGTCCTTCCCCAC
      300       310       320       330       340       350

              390       380       370       360       350       340
unknow CAGCAGGCAGTGGCTTAGGAGCTTGAAGTTGACC---GGGTCCACCCGAAGCTT--GTGC
       :: :: :  :    :   :  ::: :  ::: :   :: :: ::: :: ::   :::
unknow CACCAAGACCTACTTCCCGCACTTCGACCTGAGCCACGGCTCTGCCC--AGGTTAAGGGC
          360       370       380       390       400       410

                 330       320       310       300       290
unknow ----GCGTGCAGGTCGCTCAGG-GCGGAC-AGCGCGTTGG-GCATGTCGTCCACGTGCGC
         ::  : :::: ::  : : :: ::: : ::: ::: ::: :: :  :: ::: :
unknow CACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCCGTGGCGCACGTGGACGACATGCCC
          420       430       440       450       460       470

          280       270       260       250       240       230
unknow CACGGCGTTGGTCAGCGCGTCGGCCACCTTCTTGCCGTGGCCCTTAACCT--GGGCAGAG
       :::: :: :: :: :: :: : :  :: :::: :  : ::    :: :: :::  ::
unknow AACG-CGCTG-TCCGC-CCTGAGCGACCTGCACGC----GCAC--AAGCTTCGGGTGGAC
          480       490       500       510       520
```

gobal:

result:

Readseq version 2.1.30 (12-May-2010) Readseq version 2.1.30 (12-May-2010)

lalign output for UNKNOWN vs. UNKNOWN

[EMBnet-Server] Date: Thu Nov 11 8:51:18 2021

```
# bin/ggsearch36 -E 10.0 -f -12 -g -2 5804.1.seq 5804.2.seq
GGSEARCH performs a global/global database searches
 version 36.3.5e Nov, 2012(preload8)
*** Warning - unrecognized residue at 7:I - 73
*** Warning - unrecognized residue at 8:E - 69
*** Warning - unrecognized residue at 12:E - 69
*** Warning - unrecognized residue at 15:L - 76
*** Warning - unrecognized residue at 17:I - 73
*** Warning - unrecognized residue at 20:L - 76
Parameters not available for: +5/-4: -12/-2
Query: 5804.1.seq
  1>>>unknown 611 bp - 607 nt
Library: 5804.2.seq
      967 residues in     1 sequences

Statistics:  Unscaled normal statistics: mu= 50.0000  var=10.0000 Ztrim: 0
 statistics sampled from 0 (0) to 0 sequences
Algorithm: Global/Global affine Needleman-Wunsch (SSE2, Michael Farrar 2010) (6.0 April 2007)
Parameters: +5/-4 matrix (5:-4), open/ext: -12/-2
 Scan time:  0.000
!! No sequences with E() < 10

607 residues in 1 query    sequences
967 residues in 1 library sequences
 Tcomplib [36.3.5e Nov, 2012(preload8)] (4 proc in memory [0G])
 start: Thu Nov 11 08:51:19 2021 done: Thu Nov 11 08:51:19 2021
 Total Scan time:  0.000 Total Display time:  0.000

Function used was GGSEARCH [36.3.5e Nov, 2012(preload8)]
```

falling the recognize residues

Algorithm: Global/Global affine
Needleman-Wunsch

statistics of the normal distribution

The small value of E indicates that in practice these two sequences do not match globally.

global
with
no
end-gap
penalty

Choose the alignment method: ○ local (default) ○ global ● global without end-gap penalty

Number of reported sub-alignments: 3 ▾   E-value threshold: 10.0 (default 10.0)

Scoring matrix: BLOSUM50 (default) ▾

Opening gap penalty: -12 (default -12)   Extending gap penalty: -2 (default -2)

First sequence title (optional): [   ]

Input sequence format: Plain Text ▾

1st Query sequence or ID or AC or GI (see above for valid formats):
```
CCCTGGACAA
GTTCCTGGCTTCTGTGAGCACCGTGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGGT
GGCCATGCTT
CTTGCCCCTTGGGCCTCCCCCCAGCCCCTCCTCCCCTTCCTGCACCCGTACCCCCGTGGT
CTTTGAATAA
AGTCTGAGTGGGCGGC
```

Second sequence title (optional): [   ]

Input sequence format: Plain Text ▾

2nd Query sequence or ID or AC or GI (see above for valid formats):
```
CGTGCTGACC
TCCAAATACCGTTAAGCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTCCCAA
CGGGCCCTCC
TCCCCTCCTTGCACCGGCCCTTCCTGGTCTTTGAATAAAGTCTGAGTGGGCGGCAGCCTG
TGTGTGCCTG
GGTTCTCTCTGTCCCGGAATGTGCCAACAA
```

[Run lalign] [Clear Input]

results:

```
# bin/glsearch36 -E 10.0 -f -12 -g -2 9779.1.seq 9779.2.seq
GLSEARCH performs a global-query/local-library search
 version 36.3.5e Nov, 2012(preload8)
*** Warning - unrecognized residue at 7:I - 73
*** Warning - unrecognized residue at 8:E - 69
*** Warning - unrecognized residue at 12:E - 69
*** Warning - unrecognized residue at 15:L - 76
*** Warning - unrecognized residue at 17:I - 73
*** Warning - unrecognized residue at 20:L - 76
Parameters not available for: +5/-4: -12/-2
Query: 9779.1.seq
  1>>>unknown 611 bp - 607 nt
Library: 9779.2.seq
      967 residues in      1 sequences

Statistics: (shuffled [500]) Unscaled normal statistics: mu= 92.3571  var=33.1703 Ztrim: 0
 statistics sampled from 1 (1) to 160 sequences
Algorithm: Global/Local affine Needleman-Wunsch (SSE2, Michael Farrar 2010) (6.0 April 2007)
Parameters: +5/-4 matrix (5:-4), open/ext: -12/-2
 Scan time:  0.080

The best scores are:                                n-w bits E(1)
unknown 982 bp                              ( 967) [f] 2144 679.1       0
unknown 982 bp                              ( 967) [r]  148 37.9 2.2e-22

>>unknown 982 bp                                      (967 nt)
 n-w opt: 2144  Z-score: 3612.3  bits: 679.1 E(1):     0
global/local score: 2144; 64.4% identity (65.7% similar) in 866 nt overlap (1-607:57-921)

                                        10        20        30
unknow                          NMHMSARXXNSHXMGXBXNAXRHAHBAMRN
                                 .. :.   ..    .    .. . ..
unknow SCGCCCGGCCGGGCGTGCCCCCGCGCCCCAAGCATAAACCCTGGCGCGCTCGCGGCCCGG
         30        40        50        60        70        80

              40        50        60        70        80        90
unknow AACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACA
       ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
unknow CACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACA
         90       100       110       120       130       140
```

Algorithm: Needleman-Wunsch

in global, there is error due to fail the global alignment, but local alignment and the global with no end-gap penalty are worked. Since global method is too restricted and comparing the sequences as a whole, the result of it gives no significant matching, this can be considered as error. But local alignment works better than global since it searches the local similarities, and finding the local similarities is more appropriate to align two sequences if the sequences are not very close. In the no end-gap penalty, the algorithm is not penalising the end gaps, ant it can be used to identify the terminal domains better than the global one.

in no end-gap, a global query-local library search was performed. in local, non-overlapping local alignments were searched. in global, a global to global db search was done. global ones uses the same algorithm whereas the local one uses a different algorithm.

```
              40                  30        20        10
unkno-  CA-------GAAGAGTTNYKTVDTDYNTNNVNCKNDSNNNYTSKDKN
         ::         :   :::::      .  :..      .  :.  .    .:...
unknow  CACCTCCCCGCCGAGTTCACCCCTGCGGTGCACGC-CTCCCTGGACAAGTTCCTGGCTTC
     720        730        740        750        760        770

unknow  TGTGAGCACCGTGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGGTAGCCGTTCCTCC
     780        790        800        810        820        830
```

```
     80        70        60        50
unknow  GACAGCACCATGGTGGGTTCTCTCTGAGT-CTGTGGGGACC
         :  ::    ::       :::  ::::  :    :::    :  :::    ::::
unknow  CA-AGTTCC----TGGCTTCTGT--GAGCACCGTGCTGACC
              770                780              790
```

4. theoretical

Local: local alignment methods search only the best matching && high scoring regions among the sequence. it can start and end anywhere. it searches largest similar sub domains in the sequences.

Global: alignements match as many character as possible in the whole sequences. it requires 4 termini to process.

Global with no end-gap penalty uses the same algorithm as global alignment and try to match as many character as possible among the whole sequence but it requires 2 / 4 termini to process. it is also called semi-global alignment. it ignores end-gaps and therefore can find a more appropriate termini, but do the same thing with global alignment in general.

4. theoretical

the choose of the which alignment technique is used for the alignment is depend on the sequences. if the sequences are more similar, global alignment methods can be used but if there is less assumption about the sequences, the local alignment should be used firsty, and after that the global alignments can be proceeded. Since each gap and mismatch, etc. represent a evolutionary relatedness of divergence, it is irrelevant to do global alignment for the the unknown sequences. These differences in the sequences represents evolutionary distances between the organisms // sequences, therefore the selection process of local && global context-dependent and based on the assumptions about the sequences.

references

https://www.metagenomics.wiki/tools/blast/evalue

https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Investigations_in_Molecular_Cell_Biology_(O'Connor)/09%3A_Protein_Conservation/9.07%3A_The_BLASTP_algorithm

https://resources.qiagenbioinformatics.com/manuals/clcgenomicsworkbench/650/_E_value.html

https://www.cs.purdue.edu/homes/ayg/TALKS/BLORE10/lecture4.pdf

https://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html#head3