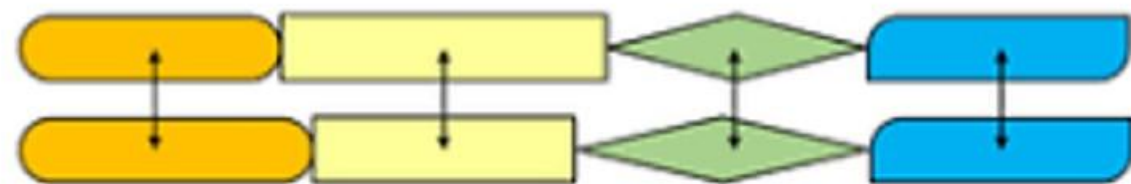


Bioinformatics Lec5 /2025

By

Dr Delveen R. Ibrahim

Global and Local alignments



Global Alignment



Local Alignment

What is sequence alignment??

Sequence alignment is the procedure of **comparing two (pair-wise alignment) or more (multiple) sequences** by searching for a **series of individual characters or patterns that are in the same order in the sequences.**

[Download](#) [GenBank](#) [Graphics](#) [Next](#) [Previous](#) [Descriptions](#)

Nannizzia sp. strain GZUIFR 21.329 internal transcribed spacer 1, partial sequence; 5.8S ribosomal RNA gene and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence

Sequence ID: [MZ019683.1](#) Length: 630 Number of Matches: 1

Range 1: 13 to 624 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1112 bits(602)	0.0	609/612(99%)	1/612(0%)	Plus/Plus
Query 6	GCATACGGCCCCCGCCCGGATGCGTCCGAGGGCGGTGCCGCGGCCACACGCCCATTCCT	65		
Sbjct 13	GCAGACGGCCCCCGCCCGGATGCGTCCGAGGGCGGTGCCGCGGCCACACGCCCATTCCT	72		
Query 66	GTCTATCTACCTGGTTGCCTCGGCGGGCCGCGCACTCGTGCCGCGCGGGGATGTTTATCT	125		
Sbjct 73	GTCTATCTACCTGGTTGCCTCGGCGGGCCGCGCACTCGTGCCGCGCGGGGATGTTTATCT	132		
Query 126	CCCTGGACCGCGCCCGCCGAGGAGCGATTAAAACCATGAATACTGTTTCACTGAGCG	185		
Sbjct 133	CCCTGGACCGCGCCCGCCGAGGAGCGATTAAAACCATGAATACTGTTTCACTGAGCG	192		

Differences between Global and local alignments

Global Alignment

Difference No 1: Why this Name?

```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
   |||||         |||||   ||||| ||||| ||||| |||||
5' ACTACTAGATT----ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'
```

'end to end' alignment

✓ In Global alignment; an attempt is made to align the **entire sequence**.

Local alignment

```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
      |||||  ||||| ||||| ||||| ||||| |||||
5' TACTCACGGATGAGGTACTTTAGAGGC 3'
```

✓ In local alignment; an attempt is made to **align local regions** with the highest level of similarity

Global Alignment

HBA_HUMAN	1	MVLSPADKTNVKAANGKVGAAHAGEYGAEALERMFLSFPTTKTYFPHFDLS	50
HBA_MOUSE	1	MVLSGEDKSNIKAAANGKIGGHGAIEYGAEALERMFASFPTTKTYFPHFDVS	50
HBA_HUMAN	51	HGSAQVKGHGKKVADALTNAAVAHVDDMPNALSALSDDLHAHKLRVDPVNFK	100
HBA_MOUSE	51	HGSAQVKGHGKKVADALASAAGHLDDLPGALSALSDDLHAHKLRVDPVNFK	100
HBA_HUMAN	101	LLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR	142
HBA_MOUSE	101	LLSHCLLVTLASHHPADFTPAVHASLDKFLASVSTVLTSKYR	142

Difference No 2: Requirement?

Suitable for aligning two closely related sequences, homologous genes

Local alignment

Multiple sequence alignment of H-RAS, N-RAS, and K-RAS sequences. The alignment shows amino acid positions (1 to 60) and conservation scores. The sequences are aligned as follows:

Input sequences	Amino acid position	Sequence	Conservation score
H-RAS	1	MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQV	60
N-RAS	10	MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQV	60
K-RAS	20	MTEYKLVVVGAGGVGKSALTIQLIQNHVFDEYDPTIEDSYRKQV	60

Conservation score

* Conserved sequence (identical)
: Conservative mutation
· Semi-conservative mutation
() Non-conservative mutation
- Gap

- Suitable for aligning more **divergent sequences**
- Used for finding out conserved patterns in DNA or protein sequences

Global Alignment

Difference No 3: Tools & Algorithms?

EMBOSS Needle

Input form Web services Help & Documentation Also in this section

Tools > Pairwise Sequence Alignment > EMBOSS Needle

API
in your application

Use BLAST in the cloud
Start an instance at a cloud provider

Pairwise Sequence Alignment

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

Global Align

CD-search

Needleman–Wunsch algorithm **Tools:** EMBOSS Needle, BLAST Global Align

Local alignment

EMBOSS Water

Input form Web services Help & Documentation Also in this section

Tools > Pairwise Sequence Alignment > EMBOSS Water

Pairwise Sequence Alignment

EMBOSS Water uses the Smith-Waterman algorithm (modified for speed on two sequences).

Basic Local Alignment Search Tool

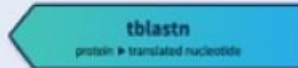
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. [Learn more](#)

ElasticBLAST 1.0.0 is Now available!
ElasticBLAST version 1.0.0 has support for faster cheaper disks at AWS and better supports Kubernetes on GCP.

Mon, 09 Jan 2023

[More BLAST news...](#)

Web BLAST



Smith–Waterman algorithm. **Tools:** BLAST, EMBOSS Water

Global alignment penalizes all gaps, since it forces full-length alignment.

Local alignment applies gap penalties only inside the matching region, making it more flexible.

Feature	Global Alignment (Needleman-Wunsch)	Local Alignment (Smith-Waterman)
Applies Gap Penalty?	✓ Yes, throughout the sequence	✓ Yes, but only inside matching region
Forces full alignment?	✓ Yes, aligns from start to end	✗ No, aligns only the best matching part
Effect of Gaps	Penalized everywhere in the sequence	Penalized only within the matching region

- The **Needleman-Wunsch algorithm** is a method used in biology to compare two sequences of DNA or proteins. It helps find the best way to align them by showing similarities and differences. This method was created by Saul B. Needleman and Christian D. Wunsch in 1970.
- The **Smith-Waterman algorithm** is another method used to compare sequences, but it focuses on finding the best matching parts instead of aligning everything. It is useful for studying DNA, RNA, or proteins. This method was developed by Temple F. Smith and Michael S. Waterman in 1981.

In Summary:

Feature	Global Alignment 	Local Alignment 
What it does?	Compares the whole sequence.	Finds only the best matching part.
Algorithm Used	Needleman-Wunsch	Smith-Waterman
Example Usage	Comparing entire genes or proteins .	Finding a small part of a gene inside a big sequence.
Gaps?	Gaps are penalized and avoided.	Gaps are allowed to find the best match.
Best For?	When two sequences are almost the same .	When looking for a small part inside a long sequence .
Simple Example	"Comparing two full books  "	"Finding one paragraph in a big book  "

Dr Delveen R. Ibrahim

Comparing BLAST with Global BLAST

Practical Lab 5

Example1

- You have been given **two FASTA sequences** from *Homo sapiens* (1) and *Rattus rattus* (2).
- **Tasks:**
 1. Perform **BLASTn** using **local alignment (multiple)**.
 2. Identify the **gene** to which the query sequences belong.
 3. Determine the **description, similarity percentage, and coverage percentage** between the two sequences using **both local and global alignment**.
 4. Display the alignment in **pairwise format**, using **dots (.)** to indicate identical bases.
 5. Generate and present a **dot plot** for both sequences.

Answers

Job Title	Homo sapiens
RID	XC2S7HDX013 Search expires on 02-22 21:39 pm Download All ▼
Program	BLASTN ? Citation ▼
Database	nt See details ▼
Query ID	lcl Query_6674293
Description	Homo sapiens
Molecule type	dna
Query Length	870
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity: to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions	Graphic Summary	Alignments	Taxonomy
---------------------	-----------------	------------	----------

Sequences producing significant alignments

[Download](#) ▼ [Select columns](#) ▼ Show [?](#)

☒ select all 100 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#) [MSA Viewer](#)

	Q1	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒		Homo sapiens lysophosphatidic acid receptor 1 (LPAR1), transcript variant 13, mRNA	Homo sapiens	1214	1214	100%	0.0	91.10%	3935	NM_001351407.2

Answers

Descriptions Graphic Summary Alignments Dot Plot

Sequences producing significant alignments Download Select columns Show 100 ?

☒ select all 1 sequences selected [Graphics](#) [MSA Viewer](#)

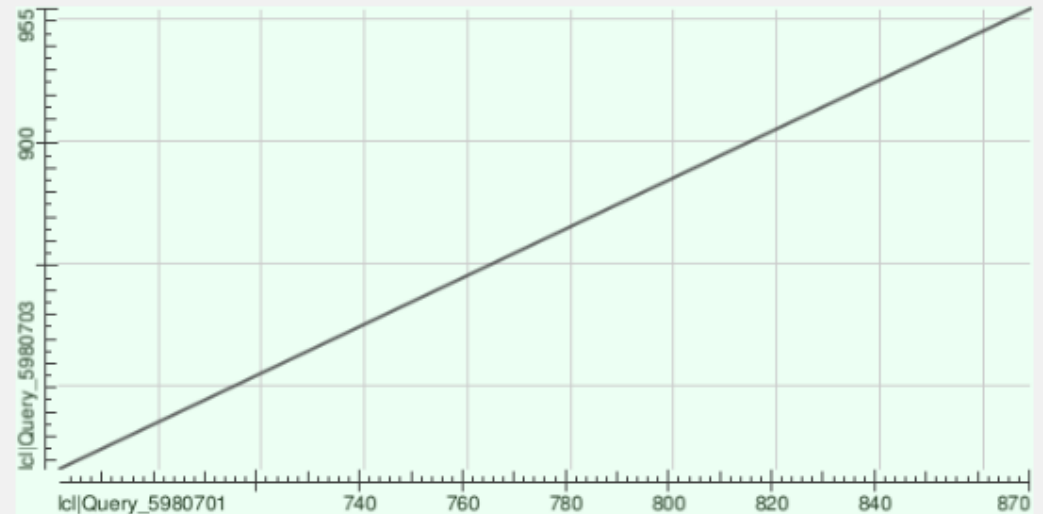
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒	Rattus norvegicus		255	255	21%	7e-72	91.01%	955	Query_5980703

Rattus norvegicus

Sequence ID: **Query_5980703** Length: **955** Number of Matches: **1**

Range 1: 767 to 955 [Graphics](#)

Score	Expect	Identities	Gaps	Strand
255 bits(138)	7e-72	172/189(91%)	0/189(0%)	Plus/Plus
Query 682	AATTCTGCCATGAACCCCATCATTTACTCCTACCGCGACAAAGAAATGAGCGCCACCTTT	741		
Sbjct 767	..C..G..T.....C.....G.....C	826		
Query 742	AGGCAGATCCTCTGCTGCCAGCGCAGTGAGAACCCACCGGCCCCACAGAAGGCTCAGAC	801		
Sbjct 827	G..T.....AC.....A.....G.....T...	886		
Query 802	CGCTCGGCTTCCTCCCTCAACCACACCATCTTGGCTGGAGTTCACAGCAATGACCACTCT	861		
Sbjct 887	C.....T..TC.....	946		
Query 862	GTGGTTTAG	870		
Sbjct 947		955		



Global alignment answers:

Descriptions

Graphic Summary

Alignments

Dot Plot

Sequences producing significant alignments

Download Manage columns Show 100 ?

☒ select all 1 sequences selected

GraphicsMSA Viewer

Description		Score	Percent Ident	Accession
Rattus norvegicus		783	76.00%	Query_5999029

Sequence ID: Query_5999029 Length: 955 Number of Matches: 1

Range 1: 1 to 955 Graphics

Next Match

NW Score	Identities	Gaps	Strand
783	736/972(76%)	119/972(12%)	Plus/Plus
Query 1	AGTCAGCAAGCTGGTGATGGGACTTGAATCACTGTTGTATCTTCATCATGTTGGCCAA	60	
Sbjct 1	.TG---.CCTC---.CC.GC---.G---.C..C.	44	
Query 61	CCTATTGGTCATGGTGGCA-ATCTATG--TCAACCGCCGCTTCCATTTTCTATT-TATT	116	
Sbjct 45	GT.CACA.C...AAC.A.C.A.AG..CT..T..AA--.AG..T..CGC.T.C..C..A	103	
Query 117	ACCTAA-TGGCTA--ATCTGGCTGCTG-----CAGACTTCTTTGCTGGGTTGGCC	163	
Sbjct 104	..GG.G...AA.GT...A..CA.A.AATGGAACAC.....C.....AC.....	163	
Query 164	TACTTCTATCTCATGTTCAACACAGGACCAATACTCGGAGACTGACTGTTAGCACATGG	223	
Sbjct 164	C..G.....G.....T.....C.....G.....	223	
Query 224	CTCCTTCGTCAGGGCCTCATTGACACAGCCTGACGGCATCTGTGGCCAACCTACTGG--	281	
Sbjct 224	..T..C..G.....C.....T.....C.....G.....CC	283	
Query 282	-----TTTCCGCATGCAGCTCCACACAGGATGAGCAAC	316	
Sbjct 284	ATTGCCATCGAGAGGCACATCACAG.....A.....T.....A.....	343	
Query 317	CGGCG-GTAGTGGTGGTCATTGTGGTCATCTGGACTATGGCCATCGTTATGGGTGCTATA	375	
Sbjct 344	..A..T..G.....G.....A.....T.....G.....C.....	403	
Query 376	CCCAGTGTGGGCTGGAAGTGTATCTGTGATATTGAAATTTGTCCAACATGGCACCCCTC	435	
Sbjct 404	C.....C.....TC.....G.....	463	
Query 436	TACAGTGACTCTTACTTAGTCTTCTGGGCCATTTTCAACTTGGTGACCTTTGTGGTAATG	495	
Sbjct 464	C.....C.....	523	
Query 496	GTGGTTCTCTATGCTCACATCTTTGGC-----AGAGGACTATGAGAATGTCTCGG	545	
Sbjct 524	C.....TATGTTTCGCC.....C.....	583	
Query 546	CATAGTTCTGGACCCCGCGGAATCGGGATACCATGATGAGTCTTCTGAAG-----	596	
Sbjct 584	A..A.....C.....C.....ACTGTGGTC	643	
Query 597	-----GCCTTTATCATCTGCTGGACTCCTGGATTGGTTTGTACTTCTAGAC	644	
Sbjct 644	ATTGTGCTGGGT..TG.....G.....C.....C.....G..C..T	703	
Query 645	GTGTGCTGCCACAGTGCAGCTGCTGGCCTATGAGAAATCTGCC-----	690	
Sbjct 704	T..C..G.....T..C.....G...T..TCCTCCTGGCCGAG	763	
Query 691	-----ATGAACCCCATCTTACTCTACCGCAGACAAAGAAATGAGCGCCACC	738	
Sbjct 764	TTCAACTCGGCT.....C.....G.....	823	
Query 739	TTTAGGCAGATCCTCTGCTGCCAGCGCAGTGAGAACCCACCGGCCACAGAAAGGCTCA	798	
Sbjct 824	C.....G..T.....AC.....A.....G.....T	883	

Example 2

- You have been provided by 2 sequences (FASTA format) present in some *E. coli* strains , the sequences are CTX-1 and CTX-15.
- **Tasks:**
 1. Perform **BLASTn** using **local alignment (multiple)**.
 2. Identify the **gene** to which the query sequences belong.
 3. Determine the **description, similarity percentage, and coverage percentage** between the two sequences using **both local and global alignment**.
 4. Display the alignment in **pairwise format**, using **dots (.)** to indicate identical bases.
 5. Generate and present a **dot plot** for both sequences.

Blastn

Descriptions

Graphic Summary

Alignments

Dot Plot

Sequences producing significant alignments

Download

Select columns

Show

100

?

☒ select all

1 sequences selected

[Graphics](#)

[MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☒ LC583827.1 Escherichia coli CTX-M-15 gene, complete cds (sequence 2)		446	446	43%	2e-129	87.12%	525	Query_32503

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

150

300

450

600

750

Descriptions

Graphic Summary

Alignments

Dot Plot

Plot of lcl|Query_32501 vs lcl|Query_32503

Blastn

Alignment view Pairwise CDS feature Restore

1 sequences selected

[Download](#) [Graphics](#)

LC583827.1 Escherichia coli CTX-M-15 gene, complete cds (sequence 2)
Sequence ID: Query_32503 Length: 525 Number of Matches: 1

Range 1: 25 to 420 [Graphics](#) [Next Match](#) [Prev](#)

Score	Expect	Identities	Gaps	Strand
446 bits(241)	2e-129	345/396(87%)	4/396(1%)	Plus/Plus
Query 283	TGAAAGCGAACCG-AATCT-GTTAAATCA-G-CGAGTTGAGATCAAAAAATCTGACTTGG	338		
Sbjct 25	TGCAAGCTTACCGCAATCTCGTTACATCATGTCGAGTTCAGATCAAAAAATCTCACCTTG	84		
Query 339	TTAACTATAATCCGATTGCGGAAAAGCACGTCGATGGGACGATGTCACTGGCTGAGCTTA	398		
Sbjct 85	TTCACTATACTCCGATTGCGGAAAAGCACGTCGATGGGACGATGTCACTGGCTCAGCTTC	144		
Query 399	GCGCGGCCGCGCTACAGTACAGCGATAACGTGGCGATGAATAAGCTGATTTCTCACGTTG	458		
Sbjct 145	GCGCGGCCGCGCTACAGTACAGCGATAACGTGGCGATGAATAAGCTGATTTCTCACGTTG	204		
Query 459	GCGGCCCGGCTAGCGTCACCGCGTTGCGCCGACAGCTGGGAGACGAAACGTTCCGTCTCG	518		
Sbjct 205	GCGGCCCGGCTAGCGTCACCGCGTTGCGCCGACAGCTGGGAGACTACACGTTCCGTATCG	264		
Query 519	ACCGTACCGAGCCGACGTTAAACACCGCCATTCCGGGCGATCCGCGTGATACCACTTCAC	578		
Sbjct 265	ACCGTACCGAGCCGACGTTAAACACCGCCATTCCGGGCAATCCGCGTCATACCACTTCAC	324		
Query 579	CTCGGGCAATGGCGCAAATCTGCGTAATCTGACGCTGGGTAAAGCATTGGGTGACAGCC	638		
Sbjct 325	CTCGGGTACTGGGCAAACTCTGCGGAATCTCACGCTACGTACAGCTTTGGGAGAGAGCC	384		
Query 639	AACGGGCGCAGCTGGTGACATGGATGAAAGGCAATA	674		
Sbjct 385	GCCGGTCCCATCTGGGGTTATGGTTCAAAGGAAATA	420		

Dr Delveen R. Ibrahim

Global alignments

Descriptions

Graphic Summary

Alignments

Dot Plot

Sequences producing significant alignments

Download

Manage columns

Show

100



☒ select all 1 sequences selected

[Graphics](#)

[MSA Viewer](#)

Description	Score	Percent Ident	Accession
LC583827.1 Escherichia coli AAK16 blaCTX-M-15 gene for extended-spectrum beta-lactamase, complete cds	-192.0	50.00%	Query_55015

Descriptions

Graphic Summary

Alignments

Dot Plot

hover to see the title click to show alignments

Alignment Scores

< 40

40 - 50

50 - 80

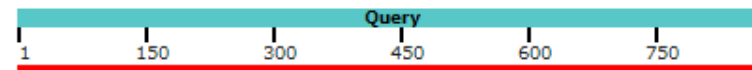
80 - 200

>= 200



1 sequences selected ?

Distribution of the top 1 Blast Hits on 1 subject sequences



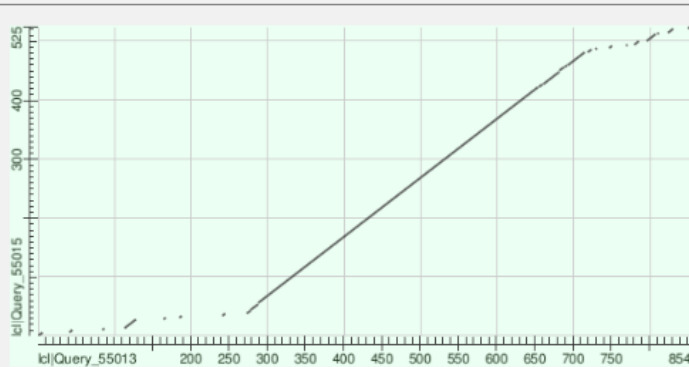
Descriptions

Graphic Summary

Alignments

Dot Plot

Plot of lcl|Query_55013 vs lcl|Query_55015 ?



Dr Delveen R. Ibrahim

Global Alignments

1 sequence selected

Download	Graphics	Next	Previous	Descriptions
--------------------------	--------------------------	----------------------	--------------------------	------------------------------

LC583827.1 Escherichia coli AAK16 blaCTX-M-15 gene for extended-spectrum beta-lactamase, complete cds

Sequence ID: Query_55015 Length: 525 Number of Matches: 1

Range 1: 1 to 525 [Graphics](#) [Next Match](#) [Previous Match](#)

NW Score	Identities	Gaps	Strand
-192	442/881 (50%)	361/881 (40%)	Plus/Plus
Query 1	ATGGTTAAAAATCACTGCGTCAGTTACGCTGATGGCGACGGCAACCGTCACGCTGTTG	60	
Sbjct 1	ATGGTT-----GGCA-----	10	
Query 61	TTAGGAAGTGTGCGCTGTATGCGCAACGGCGSACGTACAGCAAAACTTGCCGAATTA	120	
Sbjct 11	-----CAA-----CGATATC	20	
Query 121	GAGCGGCAGTCGGGAGGAAGACTGGGTGTGGCATTGATTAACACAGCAGATAATTCGCAA	180	
Sbjct 21	GAGCTGCA-----AGC-----	31	
Query 181	ATACTTATCGTGCTGATGAGCGCTTTGCGATGTGCAGCACCAGTAAAGTGATGGCCGTG	240	
Sbjct 32	-----TTA-----	34	
Query 241	GCGCGGTGCTGAAGAAAAGTGAAGCGAACCGAATCT-GTTAAATCAG--CGAGTTGAG	297	
Sbjct 35	-CCGC-----AATCTCGTTACATCATGTGCGAGTTGAG	65	
Query 298	ATCAAAAAATCTGACTTGGTTAACTATAATCCGATTGCGGAAAACACGTCGATGGGACG	357	
Sbjct 66	ATCAAAAAATCTCAGCTTGTTCACATATACCCGATTGCGGAAAACACGTCGATGGGACG	125	
Query 358	ATGTCACTGGCTGAGCTTAGCGCGGCCGCGCTACAGTACAGCGATAACGTGGCGATGAAT	417	
Sbjct 126	ATGTCACTGGCTCAGCTTCGCGCGGCCGCGCTACAGTACAGCGATAACGTGGCGATCAAT	185	
Query 418	AAGCTGATTTCTCACGTTGGCGGCCCGGCTAGCGTCACCGCGTTCCGCCGACAGCTGGGA	477	
Sbjct 186	ACGCTCATTTGCTCACGTTGGCGGCCCGGCTAGCGTCACCGCGTTCCGCCGACAGCTGGGA	245	
Query 478	GACGAAACGTTCCGTCTGACCGTACCGAGCCGACGTAAACACCGCATTCGGGGCAT	537	
Sbjct 246	GACTACACGTTCCGTATCGACCGTACCGAGCCGACGTCAACACCGCATTCGGGGCAAT	305	
Query 538	CCGCGTGATACCACTTCACTCGGGCAATGGCGCAAACTCTGCGTAATCTGACGCTGGGT	597	
Sbjct 306	CCGCGTGATACCACTTCACTCGGGTACTGGGGCAAACTCTGCGTAATCTGACGCTACGT	365	
Query 598	AAAGCATTGGGTGACAGCAACCGGGCGAGCTGGTGACATGGAAGGCAATACACC	657	
Sbjct 366	ACAGCTTTGGGAGAGAGCGCGGTCCCATCTGGGTATGGTTCAAGGAAATAAT-CC	424	
Query 658	GGTGACGAGCAATTCAAGCTGGAC--TGCGTGCCTTCCCTGGGTTGTGGGGGATAAAACCG	715	
Sbjct 425	GST-CACCGCGGTTTTTTCATTACAAATGCCCG-TTC-TGTGTTTTTGGTCATCAAAACCG	481	
Query 716	GCAGCGGTGACTATGGCAACCAACGATATCGCGGTGATCTGGCCAAAAGATCGTGGC	775	
Sbjct 482	-----GG-----GC GG-----GT-----	494	
Query 776	CGCTGATTCGTGTCACCTTACTTCAACCGACCTCAACCTAAGGCAGAAAGCCGTCGCGATG	835	
Sbjct 495	-----ATTGA-----TCAACCGGGTT--CC-----CCGTGGC-----	521	
Query 836	TATTAGCGTCGGCGGCTAAAACTGTCACCAACGGTTGTAA	876	
Sbjct 522	-----CTGA-----	525	

Dr Delveen R. Ibrahim

Lab 5 example 1

> Homo sapiens

```
AGTCAGCAAGCTGGTGATGGGACTTGGAATCACTGTTTGTATCTTCATCATGTTGGCCAACCTATTGGTCATGGT
GGCAATCTATGTCAACCGCCGCTTCCATTTTCCTATTTATTACCTAATGGCTAATCTGGCTGCTGCAGACTTCTTTG
CTGGGTTGGCCTACTTCTATCTCATGTTCAACACAGGACCCAATACTCGGAGACTGACTGTTAGCACATGGCTCC
TTCGTCAGGGCCTCATTGACACCAGCCTGACGGCATCTGTGGCCAACCTTACTGGTTTTTCCGCATGCAGCTCCAC
ACACGGATGAGCAACCGGCGGTAGTGGTGGTCATTGTGGTCATCTGGACTATGGCCATCGTTATGGGTGCTATA
CCCAGTGTGGGCTGGAAGTGTATCTGTGATATTGAAAATTGTTCCAACATGGCACCCCTCTACAGTGAAGTCTTAC
TTAGTCTTCTGGGGCCATTTTCAACTTGGTGACCTTTGTGGTAATGGTGGTTCTCTATGCTCACATCTTTGGCAGA
GGACTATGAGAATGTCTCGGCATAGTTCTGGACCCCGGCGGAATCGGGATAACCATGATGAGTCTTCTGAAGGCC
TTTATCATCTGCTGGACTCCTGGATTGGTTTTTGTACTTCTAGACGTGTGCTGTCCACAGTGCGACGTGCTGGCC
TATGAGAAATTCTGCCATGAACCCCATCATTTACTCCTACCGCGACAAAGAAATGAGCGCCACCTTTAGGCAGAT
CCTCTGCTGCCAGCGCAGTGAGAACCCCAACCGGCCCCACAGAAGGCTCAGACCGCTCGGCTTCCTCCCTCAAC
CACACCATCTTGGCTGGAGTTCACAGCAATGACCACTCTGTGGTTTAG
```

> Rattus norvegicus

ATGGCAGCTGCCTCTACTTCCAGCCCTGTGATTTACAGCCCCAGTTCACAGCCATGAACGAACAACAGT
GCTTCTACAACGAGTCTATCGCCTTCTTCTATAACCGGAGTGGAAAGTATCTAGCCACAGAATGGAACAC
CAGACTTCTTCGCTGGACTGGCCTACTTCTACCTGATGTTCAACACGGGACCTAATACCCGGAGACTGAC
CGTGAGCACATGGCTTCTCCGGCAGGGCCTCATCGACACCAGCCTGACGGCTTCTGTGGCCAACCTGCTG
GCCATTGCCATCGAGAGGCACATCACAGTTTTCCGAATGCAGCTCCATACACGAATGAGCAACCGACGTG
TGGTGGTGGTGATTGTAGTCATCTGGACTATGGCCATTGTGATGGGTGCCATACCCAGTGTGGGCTGGAA
CTGCATCTGTGATATCGATCATTGTTCCAACATGGCGCCCCTCTACAGTGACTCCTACTTAGTCTTCTGG
GCCATTTTCAACCTGGTGACCTTTGTGGTCATGGTGGTTCTCTACGCTCACATCTTTGGCTATGTTCGCC
AGAGGACTATGAGAATGTCCCGGCATAGTTCTGGACCCAGGAGGAATCGGGACACCATGATGAGCCTTCT
GAAGACTGTGGTCATTGTGCTGGGTGCCTTTATTGTCTGCTGGACTCCGGGATTGGTCTTGCTACTGCTC
GATGTGTGTTGCCCCGAGTGCGATGTCCTGGCCTATGAGAAGTTCTTCCTCCTCCTGGCCGAGTTCAACT
CGGCTATGAACCCCATCATCTACTCCTACCGCGACAAAGAGATGAGCGCCACCTTCAGGCAGATCCTGTG
TTGCCAGCGCAACGAGAACCCCAACGGCCCCACGGAAGGCTCTGACCGCTCGGCCTCCTCCCTCAACCAC
ACTATTCTGGCTGGAGTTCACAGCAATGACCACTCTGTGGTTTAG

Example 2

```
>HQ456393.1 Escherichia coli CTX-M-1, complete cds (sequence 1)
ATGGTTAAAAAATCACTGCGTCAGTTCACGCTGATGGCGACGGCAACCGTCACGCTGTTGTTAGGAAGTG
TGCCGCTGTATGCGCAAACGGCGGACGTACAGCAAAAACCTTGCCGAATTAGAGCGGCAGTCGGGAGGAAG
ACTGGGTGTGGCATTGATTAACACAGCAGATAATTCGCAAATACTTTATCGTGCTGATGAGCGCTTTGCG
ATGTGCAGCACCGAGTAAAGTGATGGCCGTGGCCGCGGTGCTGAAGAAAAGTGAAAGCGAACCGAATCTGT
TAAATCAGCGAGTTGAGATCAAAAAATCTGACTTGGTTAACTATAATCCGATTGCGGAAAAGCACGTCGA
TGGGACGATGTCACTGGCTGAGCTTAGCGCGGCCGCGCTACAGTACAGCGATAACGTGGCGATGAATAAG
CTGATTTCTCACGTTGGCGGCCCGGCTAGCGTCACCGCGTTTCGCCCCGACAGCTGGGAGACGAAACGTTCC
GTCTCGACCGTACCGAGCCGACGTTAAACACCGCCATTCCGGGCGATCCGCGTGATACCACTTCACCTCG
GGCAATGGCGCAAACCTCTGCGTAATCTGACGCTGGGTAAAGCATTGGGTGACAGCCAACGGGCGCAGCTG
GTGACATGGATGAAAGGCAATACCACCGGTGCAGCGAGCATTCAGGCTGGACTGCCTGCTTCCTGGGTG
TGGGGGATAAAACCGGCAGCGGTGACTATGGCACCAACGATATCGCGGTGATCTGGCCAAAAGATCG
TGCGCCGCTGATTCTGGTCACTTACTTCACCCAGCCTCAACCTAAGGCAGAAAGCCGTCGCGATGTATTA
GCGTCGGCGGCTAAAATCGTCACCAACGGTTTGTA
```

>LC583827.1 Escherichia coli CTX-M-15 gene, complete cds (sequence 2)

ATGGTTGGCACAACGATATCGAGCTGCAAGCTTACCGCAATCTCGTTACATCATGTCGAGTTCAGATCAA
AAAATCTCACCTTGTTCACTATACTCCGATTGCGGAAAAGCACGTCAATGGGACGATGTCACTGGCTCAG
CTTCGCGCGGGCCGCGCTACAGTACAGCGATACCGTGGCGATCAATACGCTCATTGCTCACGTTGGCGGGCC
CGGCTCGCGTCACCGCGTTTCGCCCCGACAGCTGGGAGACTACACGTTCCGTATCGACCGTACCGAGCCGAC
GTTCAACACCGCCATTCCGGGGCAATCCGCGTCATACCACTTCACCTCGGGTACTGGGGCAAACCTCTGCGG
AATCTCACGCTACGTACAGCTTTGGGAGAGAGCCGCCGGTCCCATCTGGGGTTATGGTTCAAAGGAAATA
ATCCGGTCACCGCGGTTTTTTTCATTTACAATGCCCGTTCTGTGTTTTTGGTCATCAAACCGCGGGGGGGGC
GGGTATTGTATCACCCGGGGTTCCCCGTGGCCTGA

Lab 5 task

- Lab 6 task:
- You have been given **two FASTA sequences** from *Macaca mulatta* (1) and *Homo sapiens* (2).
- **Tasks:**
 1. Identify the **gene** to which the query and subject sequences belong.
 2. Determine the **description, similarity percentage, and coverage percentage** between the two sequences using **both local and global alignment**.
 3. Display the alignment in **pairwise format**, using **dots (.)** to indicate identical bases for both **global and local alignments**.
 4. Generate and present a **dot plot** for both **local and global alignments**.

Task lab 6

>Macaca mulatta (sequence 1)

```
GGTCGGCGGGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTTCCTGTCCTTCCCCACCACCAAGACCTACTT  
CGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCCTCGCCGTGGGGGCAC  
GTGGACGACATGCCCCAAGCGCTGTCCGCGCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTGCTCCTGAG  
CCTGCTGCTGGTGACTCTGGCCGCTCACCTCCCCGCCGAGTTCACCCCTGCGGTGCACGCCTCCCTGGACAAGTTCCTG  
GCTTCTGTGAGGCTGACCTCCAAATACCGTTAAGCTGGAGCCTCGGTGGCCATGCTTCT
```

> Homo sapiens (sequence 2)

```
ACTCTTCTGGTCCCCACAGACTCAGAGAGAACCCACCATGGTGCTGTCTCCTGCCGACAAGACCAACGTCAAGGCCGCCTG  
GGGTAAGGTCGGCGCGCACGCTGGCGAGTATGGTGCGGAGGCCCTGGAGAGGATGTTTCCTGTCCTTCCCCACCACCAAGA  
CCTCTTCGACCTGAGCCACGGCTCTGCCCAGGTTAAGGGCCACGGCAAGAAGGTGGCCGACGCGCTGACCAACGCCGTGG  
CGCACGTGGACGACATGCCCAACGCGCTGTCCGCCCTGAGCGACCTGCACGCGCACAAGCTTCGGGTGGACCCGGTCAAC  
TTCAAGCTCCTAAGCCACTGCCTGCTGGTGACCCTGGCCGCCACCGTTACCCCTGCGGTGCACGCCTCCCTGGACAAGTT  
CCTGGCTTCTGTGAGCACCGTGCTGACCT
```