

Sequence Alignment Project

Dot Matrix Analysis & Pairwise Alignment

Task1: Compare two sequences of your interest using Dot Matrix

- Target sequence-Human Insulin (AJ009655.1), Subject sequence- Owl Monkey Insulin (J02989)
- They were pasted in the YASS genomic similarity search tool
- Check window size and threshold value parameters and select “Run YASS”
- Once result was obtained, zoomed Dot Plot was selected
- The Red lines represented reverse alignment and green lines represented forward alignments

The screenshot shows the YASS web interface with the title "YASS :: genomic similarity search tool". It has a navigation bar with links: YASS, web server, help, download, iedera, and retrieve result with an ID. Below the navigation bar, there is a "Data [?]" section with instructions: "(1/2) Select the genomic (DNA/RNA) sequence(s) to be compared with YASS (1 sequence self-compared is no main diagonal = put it twice)". There are three input methods: 1. "Upload your sequence(s)" with a "Choose File" button and a "select" button. 2. "Paste your sequence(s)" with a text area containing two sequences: "CACACCGCCGCTCTGCA" and "TGGGACAGCCCGCCGCTCT", and "select" buttons. 3. "Use proposed sequence(s)" with a dropdown menu for "choose species/database" and a "select" button.

Figure 1: Sequences inserted in YASS

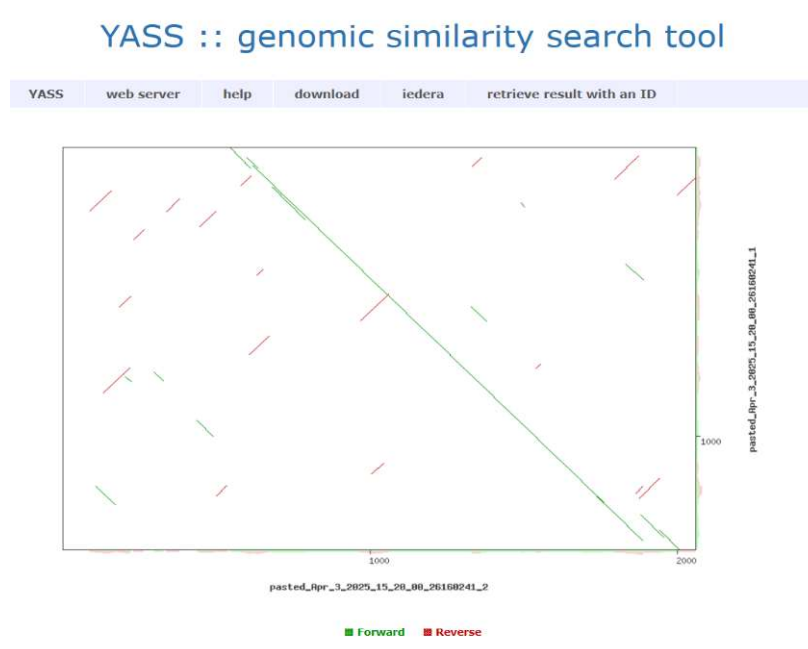


Figure 2: Zoomed Dot Plot

Dot matrix plots with explanations.

1) Mapped Dot Plot

- Straight Diagonal line indicated regions of high similarity or exact matches.
- Slight deviation in diagonal led to breaking of one alignment into two due to insertion, deletion or some other mutations.
- Parallel lines indicated sequence duplications (annotated by arrows in mapped dot plot), Short dots represented random matches and noise.

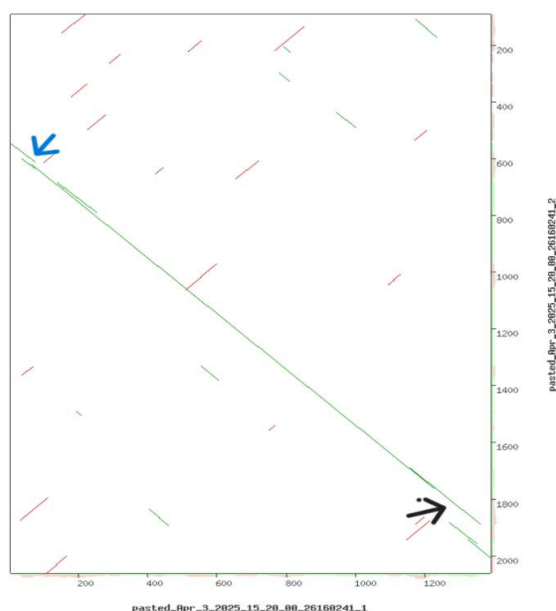


Figure 3: Mapped Dot Plot

2) Tabular Result:

Gives us information of alignment, size, statistics, and information regarding mutations

Size Length of aligned sequences	The first three entries in the table show a greater length of aligned sequences. This means more residues are aligned, there is better conservation of sequences, and alignment is reliable
Sense	
1. Number of forward alignments	14 forward alignments – indicate homology and shared conserved regions in insulin sequence of human and monkey
2. Number of Reverse alignments-	18 reverse alignments – indicate presence of inverted repeats and complementary sequences between target and subject sequences
Statistical Info	
1.E value: Range $2.31742e^{-270}$ to 8.41428	Low E value: $2.31742e^{-270}$ – alignment is statistically significant High E value: 8.41428- Less significant alignment

2. Score: Range 3281 to 57	High score- 3281 indicates high sequence identity and similarity, homology Lowest score, 57 indicates low similarity and poor alignment quality
3. Bit Score: Range 116.55 to 18.42	116.50 - High quality alignment, statistically significant 18.42 - Less significant alignment
Mutations Info	
ts- Transition substitutions of purines by purines(A,G) or of pyrimidines by pyrimidines (T,C)	ts/tv ratio gives information regarding substitution pattern (transition/transversion) Transition (ts/tv>2) Entries:2, Neutral/slight transition (ts/tv=1-2) Entries:1,4,5,7,11,15,22,28, Transversion(ts/tv<1) Entries:3,8,9,10,12,13,14,16,17,18,19,20,23,24,25,26,27,29,30,31
tv- Transversion substitutions between purines and pyrimidines	
Bias- represents (match,mis match,deletion) respectively	First entry has maximum match,mismatch and deletions – (15,15,24) 21 st and 32 nd entry has minimum match,mismatch and deletions –(1,0,0)

Positions					Statistics				Mutations			
#	pasted.Apr_3_2025_17_18_43_96271454.1	pasted.Apr_3_2025_17_18_43_96271454.2	size	sense	Evalue	score	bit-score	entropy	#ts	#tv	bias	bias.pr
	...sapiens_ins_gene_partial	...sulin_gene_complete_cds.										
1	(65-1228)	(620-1761)	1164/1142	forward	2.31742e-270	3281	917.2	5.35706	33	21	15, 15, 24	3.74e-03
2	(1158-1362)	(1691-1888)	205/198	forward	2.42642e-29	409	116.55	4.95539	24	10	14, 7, 13	6.47e-03
3	(139-252)	(683-790)	114/108	forward	1.03595e-10	187	54.66	4.19824	8	10	6, 5, 7	3.80e-02
4	(1272-1350)	(1882-1957)	79/76	forward	4.00667e-10	180	52.71	4.31497	9	9	6, 6, 6	4.43e-02
5	(1-72)	(546-611)	72/66	forward	2.00603e-05	124	37.09	4.09023	10	6	7, 5, 4	3.35e-02
6	(37-73)	(599-635)	37/37	forward	0.000114189	115	34.59	3.78714	7	0	4, 2, 1	4.80e-02
7	(719-654)	(609-672)	66/64	reverse	0.021061	88	27.06	3.82687	12	12	11, 5, 8	1.14e-02
8	(599-510)	(972-1062)	90/91	reverse	0.14544	78	24.27	3.74644	12	28	12, 16, 12	1.40e-02
9	(554-604)	(1331-1380)	51/50	forward	0.14544	78	24.27	3.37356	1	15	8, 4, 4	2.09e-02
10	(110-32)	(1797-1875)	79/79	reverse	0.315041	74	23.16	3.23593	10	24	13, 10, 11	1.96e-02
11	(1130-1095)	(1007-1045)	36/39	reverse	0.382197	73	22.88	3.32636	2	1	1, 0, 2	1.11e-01
12	(1326-1393)	(1944-2000)	68/65	forward	0.382197	73	22.88	4.02193	9	10	9, 3, 7	9.54e-03
13	(850-766)	(134-218)	85/85	reverse	0.682416	70	22.04	3.79465	12	21	12, 8, 13	1.30e-02
14	(133-99)	(500-614)	35/35	reverse	0.827885	69	21.76	2.85539	1	9	2, 4, 4	5.33e-02
15	(224-179)	(338-381)	46/44	reverse	1.00436	68	21.48	3.54659	10	5	4, 7, 4	3.14e-02
16	(166-107)	(2000-2060)	60/61	reverse	1.00436	68	21.48	2.98523	4	18	7, 10, 5	1.63e-02
17	(1215-1148)	(1877-1944)	68/68	reverse	1.21846	67	21.2	3.375	9	19	7, 11, 10	1.83e-02
18	(319-288)	(232-264)	32/33	reverse	1.4782	66	20.93	3.27761	1	7	3, 1, 4	4.27e-02
19	(1175-1236)	(109-172)	62/64	forward	1.4782	66	20.93	3.39275	8	15	9, 5, 9	1.74e-02
20	(553-517)	(185-221)	37/37	reverse	1.7933	65	20.65	3	4	10	4, 5, 5	5.27e-02
21	(766-751)	(1542-1557)	16/16	reverse	1.7933	65	20.65	3.27761	1	0	1, 0, 0	3.33e-01
22	(946-1000)	(437-490)	55/54	forward	1.7933	65	20.65	2.85539	10	10	8, 5, 7	2.86e-02
23	(67-37)	(1333-1363)	31/31	reverse	2.17557	64	20.37	2.92193	1	8	4, 3, 2	6.40e-02
24	(780-810)	(297-327)	31/31	forward	2.17557	64	20.37	2.84644	3	6	2, 2, 5	3.84e-02
25	(404-459)	(1834-1892)	56/59	forward	3.8845	61	19.53	3.03915	7	13	7, 6, 7	3.82e-02
26	(1199-1175)	(1865-1809)	25/25	reverse	4.71254	60	19.25	2.84535	2	4	2, 1, 3	8.23e-02
27	(443-423)	(632-652)	21/21	reverse	5.7171	59	18.97	2.92193	1	3	1, 1, 2	1.40e-01
28	(220-152)	(89-157)	69/69	reverse	6.9358	58	18.69	3.27761	13	18	12, 8, 11	1.73e-02
29	(277-225)	(446-498)	53/53	reverse	6.9358	58	18.69	3.08496	9	13	10, 6, 6	1.90e-02
30	(1207-1172)	(500-535)	36/36	reverse	6.9358	58	18.69	3	3	9	4, 3, 5	5.22e-02
31	(794-813)	(205-224)	20/20	forward	8.41428	57	18.42	2.80735	1	3	1, 0, 3	4.94e-02
32	(193-206)	(1492-1505)	14/14	forward	8.41428	57	18.42	3.16992	1	0	1, 0, 0	3.33e-01

Figure 4: Tabular Result of alignment of Human and Owl Monkey Insulin sequences

3) Raw Data

Raw data of some alignments-

[illegible]

Figure 4: Raw Result of alignment of Human and Owl Monkey Insulin sequence

Task 2: Perform pairwise sequence alignment using BLAST for the same sequences.

Align Sequences Nucleotide BLAST

BLASTN programs search nucleotide subjects using a nucleotide query. more...

Reset page Bookmark

Enter Query Sequence

Enter accession number(s), g(i)s, or FASTA sequence(s) Clear

Query subrange

From To

Or, upload file Choose File No file chosen

Job Title ENA|AJ009655|AJ009655.1 Homo sapiens ins gene,...

Enter a descriptive title for your BLAST search

☒ Align two or more sequences

Enter Subject Sequence

Enter accession number(s), g(i)s, or FASTA sequence(s) Clear

Subject subrange

From To

Or, upload file Choose File No file chosen

Program Selection

Optimize for

☒ Highly similar sequences (megablast)

☐ More dissimilar sequences (discontiguous megablast)

☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

BLAST Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

Figure 5: Sequence pasted in blast n

BLAST® » blastn suite-2sequences » results for RID-YXD95SHJ114

Home Recent Results Saved Strategies Help

< Edit Search Save Search Search Summary

How to read this report? BLAST Help Videos Back to Traditional Results Page

Job Title ENA|AJ009655|AJ009655.1 Homo sapiens ins gene,...

RID YXD95SHJ114 Search expires on 04-05 00:20 am Download All

Program Blast 2 sequences Citation

Query ID Icl|Query_5538337 (dna)

Query Descr ENA|AJ009655|AJ009655.1 Homo sapiens ins gene, partial

Query Length 1393

Subject ID Icl|Query_5538339 (dna)

Subject Descr ENA|J02989|J02989.1 Owl monkey (A.trivirgatus) insulin c ...

Subject Length 2113

Other reports MSA viewer

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
<input checked="" type="checkbox"/> ENA J02989 J02989.1 Owl monkey (A.trivirgatus) insulin gene, complete cds		850	850	62%	0.0	84.59%	2113	Query_5538339

Figure 6: Result obtained from blast n

Task 3: Interpret the E-value and alignment score for BLAST results.

1) Alignment Score – 850

It is based on matches, mismatches, gaps, and scoring matrices (BLOSUM) **Higher score means better alignment.** A score of **850** is **high**, alignment is of good quality with few mismatches or gaps.

2)E value- 0

Alignment is statistically significant,not by chance.The match is **not random**,sequences are **very likely related**.

3) Percentage Identity- 84.59%

84.59% of the aligned nucleotides (or amino acids) are **exact matches**.For DNA criteria for homology is Over ~70–75% identity. The obtained identity is above that hence the two sequences are homologous in nature

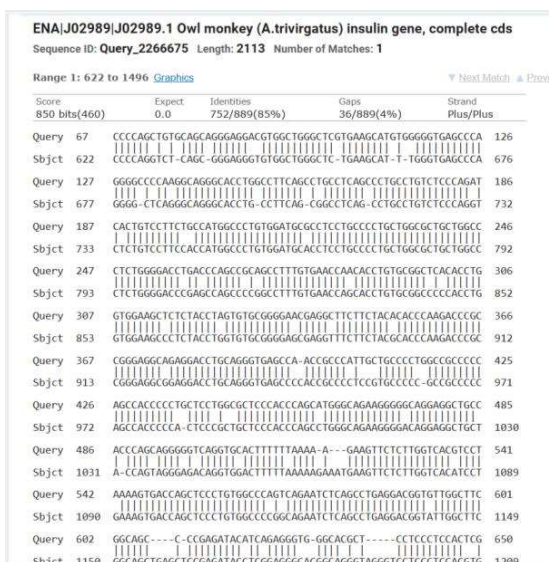


Figure 7: Sequence Alignment from blast n

Score	Expect	Identities	Gaps
850 bits(460)	0.0	752/889(85%)	36/889(4%)

Discussion on similarity/homology between the sequences.

- The alignment between the two sequences revealed an alignment score of **850**, an **E-value of 0**, and a percentage identity of **84.59%**.
- These values indicate a **highly significant similarity** between the sequences.
- The high alignment score and E-value confirm that the match is not random and is significant
- Gap of 4% might be due to deletion, insertions or mismatches
- A high percentage of identity indicates conservation at the nucleotide level.
- Sequences are likely **homologous**, they are from a common ancestor.

- This may represent **orthology** (from different species human and owl monkey, implying functional conservation because of speciation)

Conclusion

The observed results show evolutionary relationship between Human Insulin and Owl Monkey Insulin sequences.