

MOLECULAR METHODS AND BIOINFORMATICS

LM Evolutionary Biology, University of Padova
AA 2025/2026

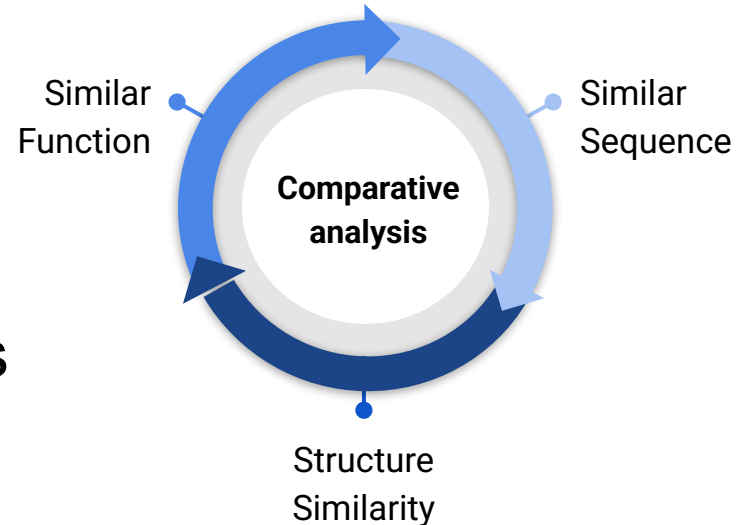
Enrico Gaffo, Silvia Orsi,
Prof. Stefania Bortoluzzi

“Comparison of biosequences and multiple alignments”

Padua, October 29, 2025

Why to align?

- To provide a measure of how much two nucleotide or amino acid sequences have in common
- It allows us to make biological inferences in terms of:
 - structural relations
 - functional relations
 - evolutionary relationships



Nucleic Acid Alignment: Why?

My DNA (or RNA, cDNA) sequence:

- is already present in nucleic acid databases (EMBL/Genbank), even if only in part, or is completely new?
- It is similar to other genes/sequences present in the databases?
- Can be aligned to a genome?
- Can encode a known protein or a protein similar to known proteins in other species?

Protein Alignment: Why?

- Is my protein "new" or it is already known?
- Is my protein similar to any known protein, also in other species?
- Can we recognize in my protein any sequence matching known functional/structural conserved domain?
- Can we use it in evolutionary studies of protein families?
- Can we do comparative modeling of protein fold?

Biosequence Alignment: Programs

Pairwise sequence alignment (PSA) and similarity search:

- **BLAST** (NCBI)
- **BLAT** (UCSC)

Multiple sequence alignment (MSA):

- **Muscle**

BLAST (Basic Local Alignment Search Tool, NCBI)

- Allows you to compare biological sequences (DNA, RNA or protein sequences) against constantly updated sequence databases
- Alignment not as precise as Needleman-Wunsch and Smith-Waterman exact algorithms but much faster, as explained in the course
- It comes in a series of sub-programs that vary according to the type of sequences to be compared and the database chosen (blastn, blastp, tblastn, tblastx) and to specific purposes

What can BLAST be useful for?

- **Identify sequence similarity indicative of homology**
- Detect orthologues or paralogs of a test sequence
- Locating the **functional domains** of a protein
- Mapping a DNA/RNA/protein sequence to find the **chromosomal location of the corresponding gene or similar genes** in the same species or in others
- **Locate similar genes** between two related species (can be used to “map” annotations from one organism to another)

BLAT (Blast-like Alignment Tool, UCSC)

*Pairwise sequence alignment algorithm developed by **Jim Kent** at the University of California Santa Cruz (UCSC) in the early 2000s to assist in the assembly and annotation of the human genome.*

It was primarily designed to reduce the time required to align millions of mouse genomic reads and expressed sequence tags against the human genome sequence.

- BLAST-like algorithm
- Aligns DNA, RNA or protein sequences with genomic sequences
- Search for similarities between a sequence under examination and the sequence of a genome
- Effective for short sequences (at least 40 characters and maximum 25,000 nucleotides or 10,000 amino acids)

What can BLAT be useful for?

- **Alignment of mRNA, RNA, or protein sequences** to the genome to:
 - find the coordinates of the **gene** that encodes it
 - Determine **the distribution** of exons and introns in a gene by
- **Cross-species analysis:** alignment of a protein or mRNA from one species against a sequence database from another species

Practical lesson 2: Nucleotide sequence alignment

1. Find out **if** the query nucleotide sequence “*seqA*” is part of a known gene;
2. Visualize the “ *seq1*” and “ *seq2*” *sequences* as they align on the genome and what their functional role is (coding/non-coding/regulatory)

1 - Download the file from **Guide** in “**II Bioinformatics practical session**”

http://compgen.bio.unipd.it/~stefania/Didattica/AA2025-2026/MMOL_BIOINFO_EB/Practical_session_2.zip

2 - unzip the file Practical_session_2.zip and open the file 3seqs.txt with the text editor

3 - Write Blast on Google and click on Nucleotide BLAST:

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](#)

Web BLAST

Nucleotide BLAST
nucleotide ► nucleotide

blastx
translated nucleotide ► protein

tblastn
protein ► translated nucleotide

Protein BLAST
protein ► protein

BLAST Genomes

Enter organism common name, scientific name, or tax id

Human Mouse Rat Microbes

BLAST now supports the new RefSeq Select databases.

NEWS

The RefSeq Select data-set consists of a representative or "Select" transcript for every protein-coding gene.

Tue, 13 Oct 2020 12:00:00 EST [More BLAST news...](#)

2 - Enter seqA as query sequence and call the analysis “SeqA”:

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

No file selected. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

FASTA Sequence:

```
>SeqA
CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCG
GGGACACTTTGCGTTCGGGC
TGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA
```

3 – Make sure that the selected database is “Nucleotide collection (nr/nt)” and that the type is “Standard databases” (click on the question mark to see the database features chosen); choose “mammalia” as the organism:

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

>SeqA
CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCG
GGGACACTTTGCGTTCGGGC
TGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCTGGATTGGCAGCCA

Or, upload file [?](#)

No file selected.

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus ☐ Experimental databases

Core nucleotide database (core_nt) [?](#)

Organism [Add Organism](#)

Optional Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Optional

Limit to ☐ Sequences from type material

Optional

Entrez Query

Optional Enter an Entrez query to limit search [?](#) [YouTube](#) [Create custom database](#)

3 – Make sure that the selected database is “Nucleotide collection (nr/nt)” and that the type is “Standard databases” (click on the question mark to see the database features chosen); choose “mammalia” as the organism:

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

BLASTn programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#) **Query subrange** [?](#)

>SeqA
CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCG
GGGACACTTTGCGTTTCGGGC
TGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCTGGATTGGCAGCCA

From

To

Or, upload file No file selected. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus ☐ Experimental databases

Core nucleotide database (core_nt) [?](#)

Organism [Add Organism](#)

Optional

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Title:Core nucleotide BLAST database
Description:The core nucleotide BLAST database consists of GenBank+EMBL+DDBJ+PDB+RefSeq sequences, but excludes EST, STS, GSS, WGS, TSA, patent sequences as well as phase 0, 1, and 2 HTGS sequences and most eukaryotic chromosome sequences. The database is non-redundant. Identical sequences have been merged into one entry, while preserving the accession, GI, title and taxonomy information for each entry.
Molecule Type:mixed DNA
Update date:2025/10/05
Number of sequences:119087540

Megablast is optimized to align nearly identical sequences, whose differences may arise from sequencing errors or polymorphisms.

discontiguous megablast is mainly used for **comparative genomics projects** : sequences from related species are compared, whose sequences do not diverge as much as those of evolutionarily distant species, but more than identical sequences (*apart from polymorphisms*) from the same species.

Program Selection

Optimize for

- ☒ Highly similar sequences (megablast)
- ☐ More dissimilar sequences (discontiguous megablast)
- ☐ Somewhat similar sequences (blastn)

Choose a BLAST algorithm ?

BLAST

Search **database core_nt** using **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window

The **difference** between megablast and blastn is in the choice of the "**word size**", that is, the **minimum length of the string of contiguous residues considered for the purpose of identity evaluation**, which in Megablast is optimal with values greater than or equal to 16 (multiples of 4). This makes the acceptance of identities more selective and Megablast up to 10 times faster , which is **therefore better suited to use with very long and extremely similar sequences** ;

4 - Optimize for megablast, leave the default algorithm parameters and start the search:

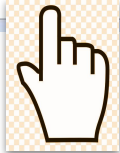
Program Selection

Optimize for

☒ Highly similar sequences (megablast)
☐ More dissimilar sequences (discontiguous megablast)
☐ Somewhat similar sequences (blastn)
Choose a BLAST algorithm ?

BLAST

Search **database core_nt** using **Megablast (Optimize for highly similar sequences)**
☐ Show results in a new window



BLAST result:

return to the search page to be able to redo it by changing parameters

save the parameters used for a possible future search

shows a table with the parameters used for the search

[← Edit Search](#)

[Save Search](#)

[Search Summary ▼](#)

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

i Your search is limited to records **include: Mammalia (taxid:40674)**

Job Title	SeqA
RID	G1W4FWVK015 Search expires on 10-30 00:34 am Download All ▼
Program	BLASTN ? Citation ▼
Database	core_nt See details ▼
Query ID	lcl Query_7910667
Description	SeqA
Molecule type	dna
Query Length	2512
Other reports	Distance tree of results MSA viewer ?

Research information

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](#)

filters

Download selected hits in various formats

allows you to choose which columns to display

allows you to choose how many rows to display

Descriptions	Graphic Summary	Alignments	Taxonomy					
Sequences producing significant alignments								
Download Select columns Show 100 ?								
<input checked="" type="checkbox"/> select all 100 sequences selected								
GenBank Graphics Distance tree of results MSA Viewer								
Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA	Homo sapiens	4639	4639	100%	0.0	100.00%	2512	NM_000546.6
<input checked="" type="checkbox"/> TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)	Homo sapiens	4639	4639	100%	0.0	100.00%	2586	HG975427.1
<input checked="" type="checkbox"/> Homo sapiens tumor protein p53 (TP53), transcript variant 2, mRNA	Homo sapiens	4619	4619	100%	0.0	99.88%	2509	NM_001276761.3
<input checked="" type="checkbox"/> Homo sapiens p53 protein mRNA, complete cds	Homo sapiens	4599	4599	100%	0.0	99.88%	2521	AF307851.1
<input checked="" type="checkbox"/> Homo sapiens tumor protein p53, mRNA (cDNA clone MGC:646 IMAGE:3544714), complete cds	Homo sapiens	4580	4580	99%	0.0	99.88%	2508	BC003596.1
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes tumor protein p53 (TP53), transcript variant X2, mRNA	Pan troglodytes	4553	4553	100%	0.0	99.36%	2537	XM_001172077.6
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla gorilla tumor protein p53 (TP53), transcript variant X2, mRNA	Gorilla gorilla g...	4545	4545	100%	0.0	99.36%	2529	XM_004058511.5

Max score: The highest calculated score of the alignment, counting prizes and penalties.

Total score: the sum of the alignment scores of all segments of the same sequence.

E - Value (Expected value): the number of alignment expected by chance with the calculated score. For significant alignments the E-value should be very close to 0.

5 - Click on “Taxonomy” to see the the taxonomic distribution of matched subject sequences in the database.

<div> <div>Descriptions</div> <div>Graphic Summary</div> <div>Alignments</div> <div>Taxonomy</div> </div>				
<div> <div>Reports</div> <div>Lineage</div> <div>Organism</div> <div>Taxonomy</div> </div>				
100 sequences selected ?				
Organism	Blast Name	Score	Number of Hits	Description
Euarchontoglires	placentals		115	
• Simiiformes	primates		113	
• Catarrhini	primates		103	
• Hominoidea	primates		73	
• Hominidae	primates		68	
• Homininae	primates		66	
• Homo sapiens	primates	4639	60	Homo sapiens hits
• Pan troglodytes	primates	4553	2	Pan troglodytes hits
• Gorilla gorilla gorilla	primates	4545	2	Gorilla gorilla gorilla hits
• Pan paniscus	primates	4521	2	Pan paniscus hits
• Pongo abelii	primates	4327	1	Pongo abelii hits
• Pongo pygmaeus	primates	4311	1	Pongo pygmaeus hits
• Nomascus leucogenys	primates	4135	1	Nomascus leucogenys hits
• Symphalangus syndactylus	primates	4124	3	Symphalangus syndactylus hits
• Hylobates moloch	primates	4117	1	Hylobates moloch hits
• Ptilocolobus tephrosceles	primates	3982	1	Ptilocolobus tephrosceles hits
• Macaca fascicularis	primates	3978	2	Macaca fascicularis hits
• Cercopithecus atys	primates	3971	1	Cercopithecus atys hits

i Your search is limited to records **include: Mammalia (taxid:40674)**

Job Title	SeqA
RID	G1W4FWVK015 <small>Search expires on 10-30 00:34 am</small> Download All ▾
Program	BLASTN ? Citation ▾
Database	core_nt See details ▾
Query ID	Ic Query_7910667
Description	SeqA
Molecule type	dna
Query Length	2512
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

Blast Tree View

This tree was produced using BLAST pairwise alignments. [more...](#)

[Reset Tree](#)

BLAST RID [G1TJG5ZC014](#)

Query ID |cl|Query_5842903

Database core_nt

Tree method

Fast Minimum Evolution

Max Seq Difference

0.75

Sequence Label

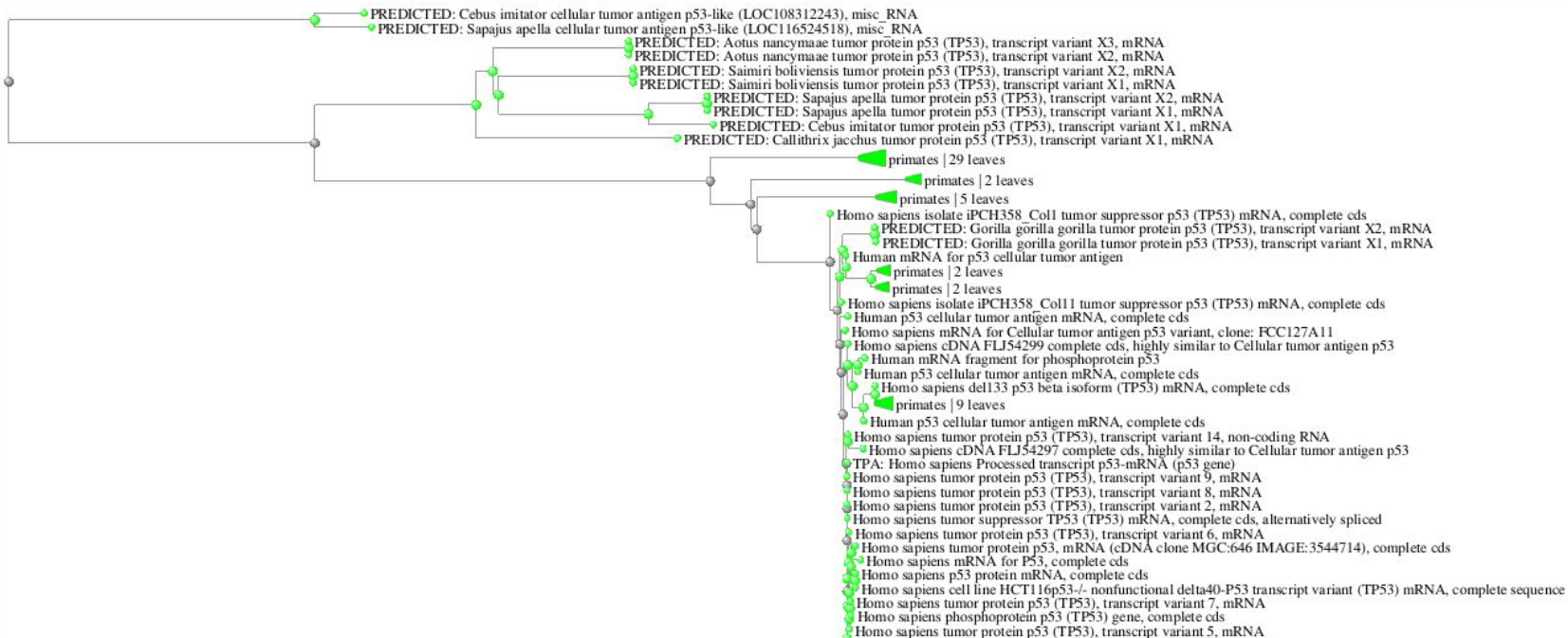
Sequence Title (if avail)

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

[Hide legend](#)

Find: all

Tools Upload



Label color map	
 	query
 	from type material
Blast names color map	
 	primates
 	rodents
 	unknown

[← Edit Search](#)[Save Search](#)[Search Summary ▾](#)[? How to read this report?](#)[▶ BLAST Help Videos](#)[↶ Back to Traditional Results Page](#)

i Your search is limited to records **include: Mammalia (taxid:40674)**

Job Title	SeqA
RID	G1W4FWVK015 <small>Search expires on 10-30 00:34 am</small> Download All ▾
Program	BLASTN ? Citation ▾
Database	core_nt See details ▾
Query ID	Ic Query_7910667
Description	SeqA
Molecule type	dna
Query Length	2512
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear

☐ exclude[+ Add organism](#)

Percent Identity

 to

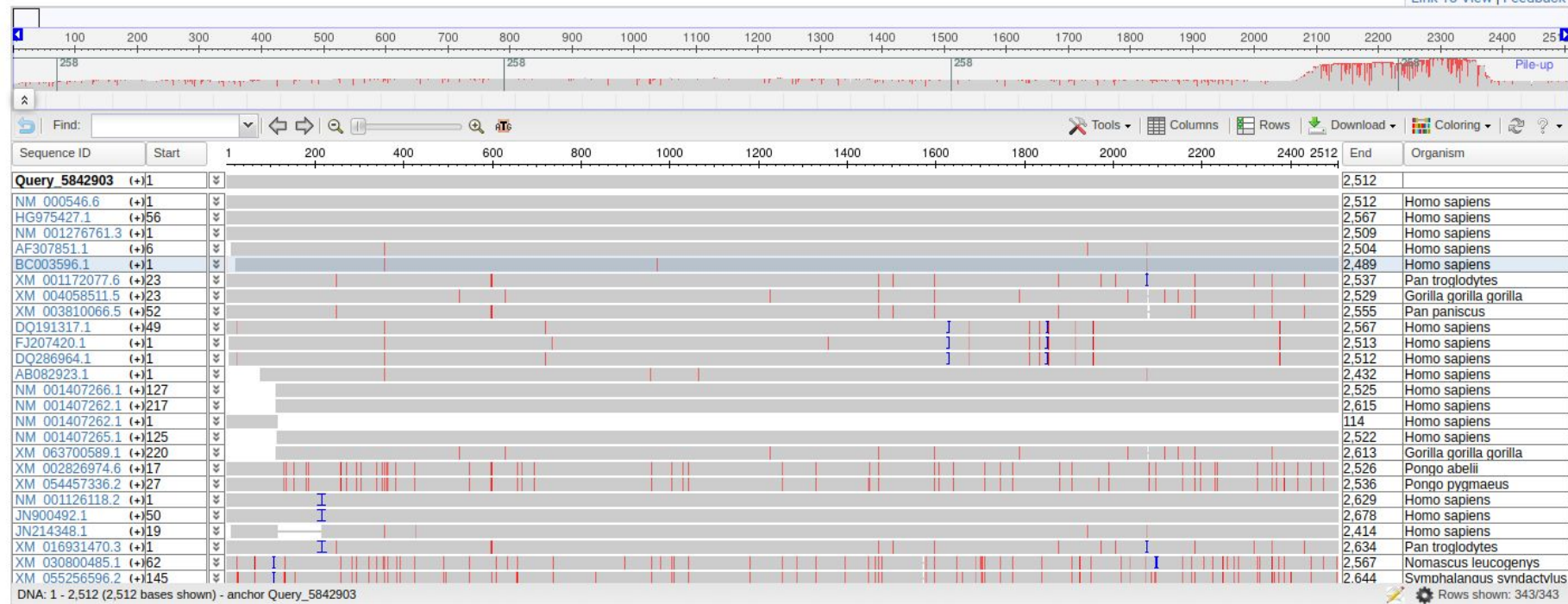
E value

 to

Query Coverage

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

Alignment

[Link To View](#) | [Feedback](#)

What is the best alignment?

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

Sequences producing significant alignments
Download
Select columns
Show
100

☒ select all 100 sequences selected
GenBank
Graphics
Distance tree of results
MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA	Homo sapie...	4639	4639	100%	0.0	100.00%	2512	NM_000546.6
<input checked="" type="checkbox"/>	TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)	Homo sapie...	4639	4639	100%	0.0	100.00%	2586	HG975427.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 2, mRNA	Homo sapie...	4619	4619	100%	0.0	99.88%	2509	NM_001276761.3
<input checked="" type="checkbox"/>	Homo sapiens p53 protein mRNA, complete cds	Homo sapie...	4599	4599	100%	0.0	99.88%	2521	AF307851.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53, mRNA (cDNA clone MGC:646 IMAG...	Homo sapie...	4580	4580	99%	0.0	99.88%	2508	BC003596.1
<input checked="" type="checkbox"/>	PREDICTED: Pan troglodytes tumor protein p53 (TP53), transcript vari...	Pan troglod...	4553	4553	100%	0.0	99.36%	2537	XM_001172077.6
<input checked="" type="checkbox"/>	PREDICTED: Gorilla gorilla gorilla tumor protein p53 (TP53), transcript...	Gorilla gorill...	4545	4545	100%	0.0	99.36%	2529	XM_004058511.5
<input checked="" type="checkbox"/>	PREDICTED: Pan paniscus tumor protein p53 (TP53), transcript varian...	Pan paniscus	4521	4521	100%	0.0	99.20%	2555	XM_003810066.5
<input checked="" type="checkbox"/>	Homo sapiens p53 protein (TP53) mRNA, complete cds, alternatively s...	Homo sapie...	4518	4518	100%	0.0	99.09%	2584	DQ191317.1

What is the best alignment?

Descriptions									
Graphic Summary									
Alignments									
Taxonomy									
Sequences producing significant alignments									
Download									
Select columns									
Show 100									
?									
select all 100 sequences selected									
GenBank									
Graphics									
Distance tree of results									
MSA Viewer									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA	Homo sapie...	4639	4639	100%	0.0	100.00%	2512	NM_000546.6
<input checked="" type="checkbox"/>	TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)	Homo sapie...	4639	4639	100%	0.0	100.00%	2586	HG975427.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 2, mRNA	Homo sapie...	4619	4619	100%	0.0	99.88%	2509	NM_001276761.3
<input checked="" type="checkbox"/>	Homo sapiens p53 protein mRNA, complete cds	Homo sapie...	4599	4599	100%	0.0	99.88%	2521	AF307851.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53, mRNA (cDNA clone MGC:646 IMAG...	Homo sapie...	4580	4580	99%	0.0	99.88%	2508	BC003596.1
<input checked="" type="checkbox"/>	PREDICTED: Pan troglodytes tumor protein p53 (TP53), transcript vari...	Pan troglod...	4553	4553	100%	0.0	99.36%	2537	XM_001172077.6
<input checked="" type="checkbox"/>	PREDICTED: Gorilla gorilla gorilla tumor protein p53 (TP53), transcript ...	Gorilla gorill...	4545	4545	100%	0.0	99.36%	2529	XM_004058511.5
<input checked="" type="checkbox"/>	PREDICTED: Pan paniscus tumor protein p53 (TP53), transcript varian...	Pan paniscus	4521	4521	100%	0.0	99.20%	2555	XM_003810066.5
<input checked="" type="checkbox"/>	Homo sapiens p53 protein (TP53) mRNA, complete cds, alternatively s...	Homo sapie...	4518	4518	100%	0.0	99.09%	2584	DQ191317.1

6 - Click on the first “Description” link to see the alignment of the query sequence with this target sequence:

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise ☐ CDS feature ☒ Restore defaults

100 sequences selected

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

Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA
Sequence ID: [NM_000546.6](#) Length: 2512 Number of Matches: 1
[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 2512 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
4639 bits(2512)	0.0	2512/2512(100%)	0/2512(0%)	Plus/Plus
Query 1	CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT	60		
Sbjct 1	CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT	60		
Query 61	GCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCTGGATTGGCAGCCA	120		
Sbjct 61	GCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCTGGATTGGCAGCCA	120		
Query 121	GACTGCCCTTCGGGTCACTGCCATGGAGGAGCCGCACTCAGATCCTAGCGTCGAGCCCC	180		
Sbjct 121	GACTGCCCTTCGGGTCACTGCCATGGAGGAGCCGCACTCAGATCCTAGCGTCGAGCCCC	180		
Query 181	TCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCTGAAACAACTTCTGTCT	240		
Sbjct 181	TCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCCTGAAACAACTTCTGTCT	240		
Query 241	CCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCGGACGATATTGAACAATG	300		
Sbjct 241	CCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCGGACGATATTGAACAATG	300		
Query 301	GTTCACTGAAGACCCAGGTCCAGATGAAGTCCCAGAATGCCAGAGGCTGTCCCCCGT	360		
Sbjct 301	GTTCACTGAAGACCCAGGTCCAGATGAAGTCCCAGAATGCCAGAGGCTGTCCCCCGT	360		
Query 361	GGCCCCTGACACGAGCTCCTACACGGCGGCCCTTGACACAGCCCCCTCCTGGCCCC	420		
Sbjct 361	GGCCCCTGACACGAGCTCCTACACGGCGGCCCTTGACACAGCCCCCTCCTGGCCCC	420		
Query 421	GTCATCTTCTGTCCCTTCCAGAAAACCTACAGGGCAGTACGGTTTCGGTCTGGGCTT	480		
Sbjct 421	GTCATCTTCTGTCCCTTCCAGAAAACCTACAGGGCAGTACGGTTTCGGTCTGGGCTT	480		

alignment display
mode

target sequence
information

alignment information

7 - Click on “GenBank” to see the GenBank information for this gene:

Descriptions

Graphic Summary

Alignments

Taxonomy

Alignment view Pairwise ☐ CDS feature [Restore defaults](#)

100 sequences selected [?](#)

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

Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA
Sequence ID: [NM_000546.6](#) Length: 2512 Number of Matches: 1
[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 2512 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
4639 bits(2512)	0.0	2512/2512(100%)	0/2512(0%)	Plus/Plus
Query 1	CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT	60		
Sbjct 1	CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT	60		
Query 61	GCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA	120		
Sbjct 61	GCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA	120		
Query 121	GACTGCCCTTCGGGTCACTGCCATGGAGGAGCCGCGAGTCAGATCCTAGCGTCGAGCCCC	180		
Sbjct 121	GACTGCCCTTCGGGTCACTGCCATGGAGGAGCCGCGAGTCAGATCCTAGCGTCGAGCCCC	180		
Query 181	TCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTCTGAAAACACGTTCTGTC	240		
Sbjct 181	TCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTCTGAAAACACGTTCTGTC	240		
Query 241	CCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCGGACGATATTGAACAATG	300		
Sbjct 241	CCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCGGACGATATTGAACAATG	300		
Query 301	GTTCACTGAAGACCCAGGTCCAGATGAAGTCCCAGAATGCCAGAGGCTGTCCCCCGT	360		
Sbjct 301	GTTCACTGAAGACCCAGGTCCAGATGAAGTCCCAGAATGCCAGAGGCTGTCCCCCGT	360		
Query 361	GGCCCCTGACACGAGCTCCTACACGGCGGCCCTTGACACAGCCCCCTCCTGGCCCC	420		
Sbjct 361	GGCCCCTGACACGAGCTCCTACACGGCGGCCCTTGACACAGCCCCCTCCTGGCCCC	420		
Query 421	GTCATCTTCTGTCCCTTCCAGAAAACCTACAGGGAGTACGGTTTCGCTCTGGGCTT	480		
Sbjct 421	GTCATCTTCTGTCCCTTCCAGAAAACCTACAGGGAGTACGGTTTCGCTCTGGGCTT	480		

7 - Click on “GenBank” to see the GenBank information for this gene:

GenBank ▼

Send to: ▼

Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

NCBI Reference Sequence: NM_000546.6

[FASTA](#) [Graphics](#)

Go to: ☐

LOCUS	NM_000546	2512 bp	mRNA	linear	PRI 12-JUN-2025
DEFINITION	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA.				
ACCESSION	NM_000546				
VERSION	NM_000546.6				
KEYWORDS	RefSeq; MANE Select.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.				

8 - click on “Gene” to see NCBI Gene information about this gene:

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

Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA

Sequence ID: [NM_000546.6](#) Length: 2512 Number of Matches: 1

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 1 to 2512 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
4639 bits(2512)	0.0	2512/2512(100%)	0/2512(0%)	Plus/Plus
Query 1	CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT	60		
Sbjct 1	CTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACTTT	60		
Query 61	GCGTTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA	120		
Sbjct 61	GCGTTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGCCA	120		

Related Information
[Gene](#) - associated gene details
[Genome Data Viewer](#) - aligned genomic context

Summary exercises

- Can you find possible orthologs of P53 gene in Dolphins?
 - Use seq1 as query
 - Subset the database to *Tursiops truncatus*
 - Examine results. Can you describe the best hit alignment scores?

Practical lesson 2: Nucleotide sequence alignment

1. Find out **if** the query nucleotide sequence “*seqA*” is part of a known gene;
2. Visualize the “ *seq1*” and “ *seq2*” *sequences* as they align on the genome and what their functional role is (coding/non-coding/regulatory)

Connect to UCSC Genome Browser, BLAT and enter Seq1 in the space specific. Check that the Genome Assembly is that of Feb. 2009 and that the species both Human. Start search:

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Projects](#) [Help](#) [About Us](#)

Human BLAT Search

BLAT Search Genome

Genome: ☐ Search all Assembly: Feb. 2009 (GRCh37/hg19) ☐ Query type: BLAT's guess ☐ Sort output: query,score ☐ Output type: hyperlink ☐

>Seq1
actggcgctaaaagtgttgagctttcctttacaattctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctccggggacactttgcttcgggctgggagcgtgctttccacgacggtgacacgcttcctg
gattggcagccagactgccttcgggtcactgcatggaggagcgcagtcagatcctagctgcagcccccctgagtcaggaaacattttcagacctatggaaactacttctgaaaacaacgttctgtccccccttgccgtc
ccaagcaatggatgatttgatgctgtccccggacgatattgaacaatggttctactgaagaccagggtccagatgaagctccagaatgccagaggctgctcccccggtggccctgacacgacgctcctacaccggcgcc
tgaccagccccctcctggccccctgtcatcttctgtccctccagaaaaactaccaggggacctacggtttccgctggtgcttcttgcatctgggacagccaagtctgtgacttgacgctactccctgcctcaacaagat
gtttgccaactggccaagacctgcccgtgtgagctgtgggttattccacacccccgcccggctttacgcgcatggccatctacaagcagtcacagcacatgacggagggttgtagggcgctgccccaccatgagcgctgt
cagatagcgtggtgtgccccctcctcagcatcttatccgagtggaaaccttgatttgcgtgtggagtatttggatgacagaaaacactttcgacatagtggtggtggtgcccctttaccatgctgaggttggtctgactgtac
caccatccactcatataactacatgtgtaacagttcctgcatggcgccatgaaccggaggccatcctcaccatcatcacactggaagactccagtggttaactactggtggagcgaacagctttgaggtgctgtgttgcctgt
cctgggagagaccggcgccacagaggaagagaatctccgaagaaaggggagcctcaccacagagctgccccaggagcactaagcgagcactgccaacaacaccag

☐ All Results (no minimum matches)



Human (hg19) BLAT Results

BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

Which is the best alignment?

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Projects](#)[Help](#)[About Us](#)

Human (hg19) BLAT Results

BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

Which is the best alignment?

[Genomes](#)[Genome Browser](#)[Tools](#)[Mirrors](#)[Downloads](#)[My Data](#)[Projects](#)[Help](#)[About Us](#)

Human (hg19) BLAT Results

BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

3 - Click on “details” to see the alignment:

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Projects](#) [Help](#) [About Us](#)

Human (hg19) BLAT Results

BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

4 - Click on “together” in the left menu:

Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<<< ||||| <<<<<<<
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 0000084
<<<<<<< ||||| <<<<<<<
7590810 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 7590761
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<<< ||||| <<<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<<<<<<< ||||| <<<<<<<
7590710 cgcttccctggattgg 7590695
```

```
0000151 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 0000200
<<<<<<< ||||| <<<<<<<
7579940 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 0000250
<<<<<<< ||||| <<<<<<<
7579890 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 7579841
```

```
0000251 ct 0000252
<<<<<<< || <<<<<<<
7579840 ct 7579839
```

Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<<< ||||| <<<<<<<
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084
<<<<<<< ||||| <<<<<<<
7590810 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 7590761
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<<< ||||| <<<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<<<<<<< ||||| <<<<<<<
7590710 cgcttccctggattgg 7590695
```

```
0000151 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 0000200
<<<<<<< ||||| <<<<<<<
7579940 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaa 0000250
<<<<<<< ||||| <<<<<<<
7579890 ctagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaa 7579841
```

```
0000251 ct 0000252
<<<<<<< || <<<<<<<
7579840 ct 7579839
```

Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<<< ||||| <<<<<<<
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 0000084
<<<<<<< ||||| <<<<<<<
7590810 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 7590761
```

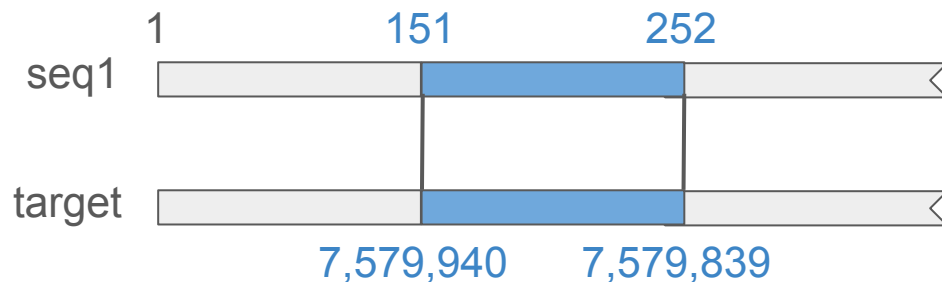
```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<<< ||||| <<<<<<<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<<<<<<< ||||| <<<<<<<
7590710 cgcttccctggattgg 7590695
```

```
0000151 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 0000200
<<<<<<< ||||| <<<<<<<
7579940 cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc 7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaa 0000250
<<<<<<< ||||| <<<<<<<
7579890 ctagcgtcgagccccctctgagtcaggaaacattttcagacctatggaaa 7579841
```

```
0000251 ct 0000252
<<<<<<< || <<<<<<<
7579840 ct 7579839
```



Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<<<<<<< |||<
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 0000084
<<<<<<< |||<
7590810 ttctcaaaagtctagagccaccgtccagggagcaggtagctgctgggctc 7590761
```

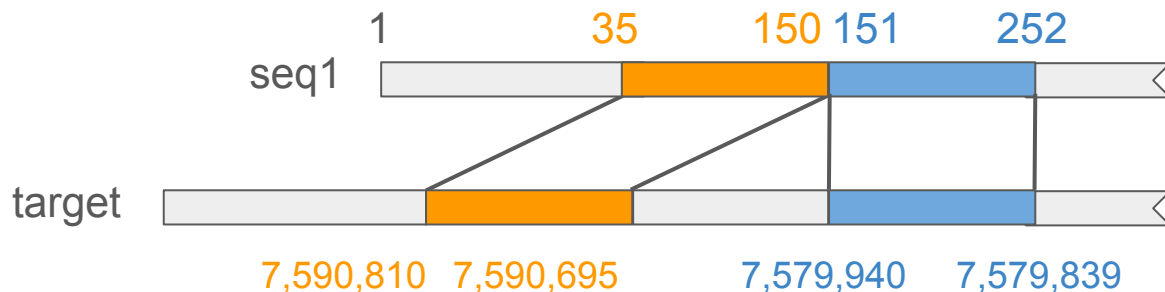
```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<<<<<<< |||<
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<<<<<<< |||<
7590710 cgcttccctggattgg 7590695
```

```
0000151 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 0000200
<<<<<<< |||<
7579940 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 7579891
```

```
0000201 ctacgctcgagccccctctgagtcaggaaacattttcagacctatggaaa 0000250
<<<<<<< |||<
7579890 ctacgctcgagccccctctgagtcaggaaacattttcagacctatggaaa 7579841
```

```
0000251 ct 0000252
<<<<<<< ||<
7579840 ct 7579839
```



Side by Side Alignment

```
0000001 actggcgctaaaagttttgagc 0000022
<----- |||-----<-----
7590832 actggcgctaaaagttttgagc 7590811
```

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084
<----- |||-----<-----
7590810 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 7590761
```

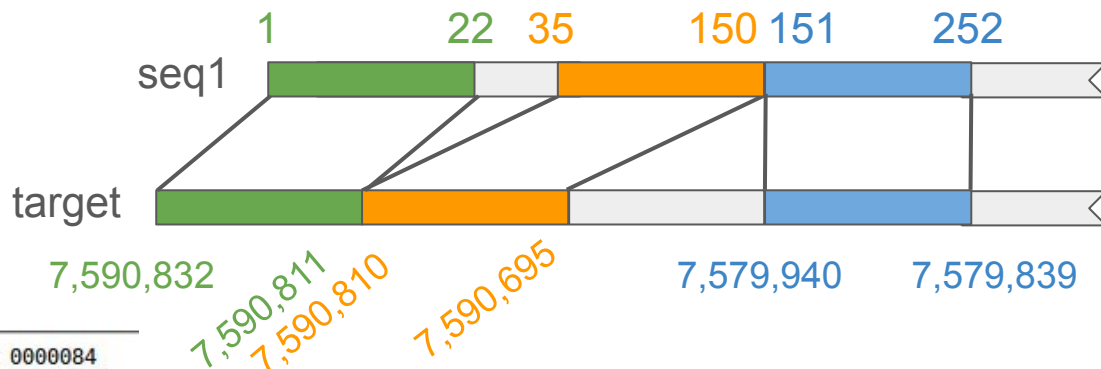
```
0000085 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 0000134
<----- |||-----<-----
7590760 cggggacactttgcgttcgggctgggagcgtgctttccacgacggtgaca 7590711
```

```
0000135 cgcttccctggattgg 0000150
<----- |||-----<-----
7590710 cgcttccctggattgg 7590695
```

```
0000151 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 0000200
<----- |||-----<-----
7579940 cagccagactgccttcgggtcactgcatggaggagccgcagtcagatc 7579891
```

```
0000201 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 0000250
<----- |||-----<-----
7579890 ctagcgtcgagccccctctgagtcaggaacattttcagacctatggaaa 7579841
```

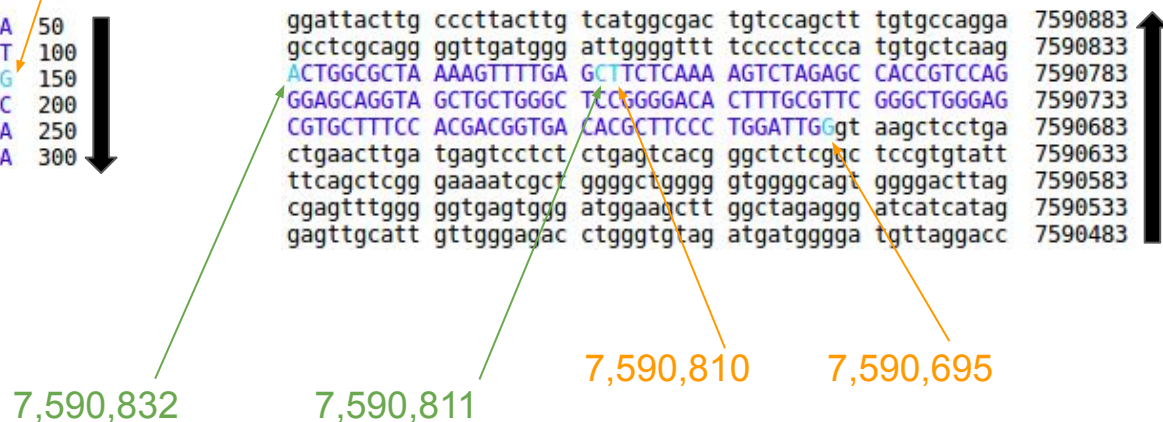
```
0000251 ct 0000252
<----- || <-----
7579840 ct 7579839
```



5 - Click on "Seq1" in the left menu:



Genomic chr17 (reverse strand):



cDNA Seq1

ACTGGCGCTA AAAGTTTTGA Gctttccttt acaaTTCTCA AAAGTCTAGA 50
GCCACCGTCC AGGGAGCAGG TAGCTGCTGG GCTCCGGGGA CACTTTGCGT 100
TCGGGCTGGG AGCGTGCTTT CCACGACGGT GACACGCTTC CCTGGATTGG 150
CAGCCAGACT GCCTTCCGGG TCACTGCCAT GGAGGAGCCG CAGTCAGATC 200
CTAGCGTCGA GCCCCCTCTG AGTCAGGAAA CATTTTCAGA CCTATGGAAA 250
CTACTTCCTG AAAACAACGT TCTGTCCCCC TTGCCGTCCC AAGCAATGGA 300

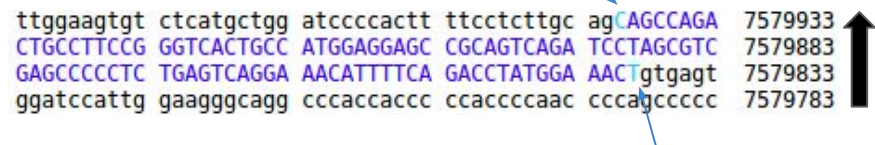


151

252

7,579,940

ttggaagtgt ctcattgctgg atccccactt ttcctcttgc agCAGCCAGA 7579933
CTGCCTTCCG GGTCACTGCC ATGGAGGAGC CGCAGTCAGA TCCTAGCGTC 7579883
GAGCCCCCTC TGAGTCAGGA AACATTTTCA GACCTATGGA AACTgtgagt 7579833
ggatccattg gaagggcagg cccaccaccc ccaccccaac cccagcccc 7579783



7,579,839

6 - Go back and click on “browser” to see the sequence in the Genome Browser:

[Home](#) [Genomes](#) [Genome Browser](#) [Tools](#) [Mirrors](#) [Downloads](#) [My Data](#) [Projects](#) [Help](#) [About Us](#)

Human (hg19) BLAT Results

BLAT Search Results

Go back to [chrX:15,578,261-15,621,068](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22

UCSC Genome Browser on Human Feb. 2009 (GRCh37/hg19) Assembly

move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x 100x

chr17:7,576,908-7,590,832 13,925 bp. enter position, gene symbol, HGVS or search terms

go

chr17 (p13.1) 13.3 13.2 13.1 17p12 17p11.2 17q11.2 17q12 21.31 17q22 24.2q24.3q25.1 17q25.3

Scale 5 kb hg19
chr17: 7,576,000|7,579,000|7,580,000|7,581,000|7,582,000|7,583,000|7,584,000|7,585,000|7,586,000|7,587,000|7,588,000|7,589,000|7,590,000|
Seq1 blat on Seq1

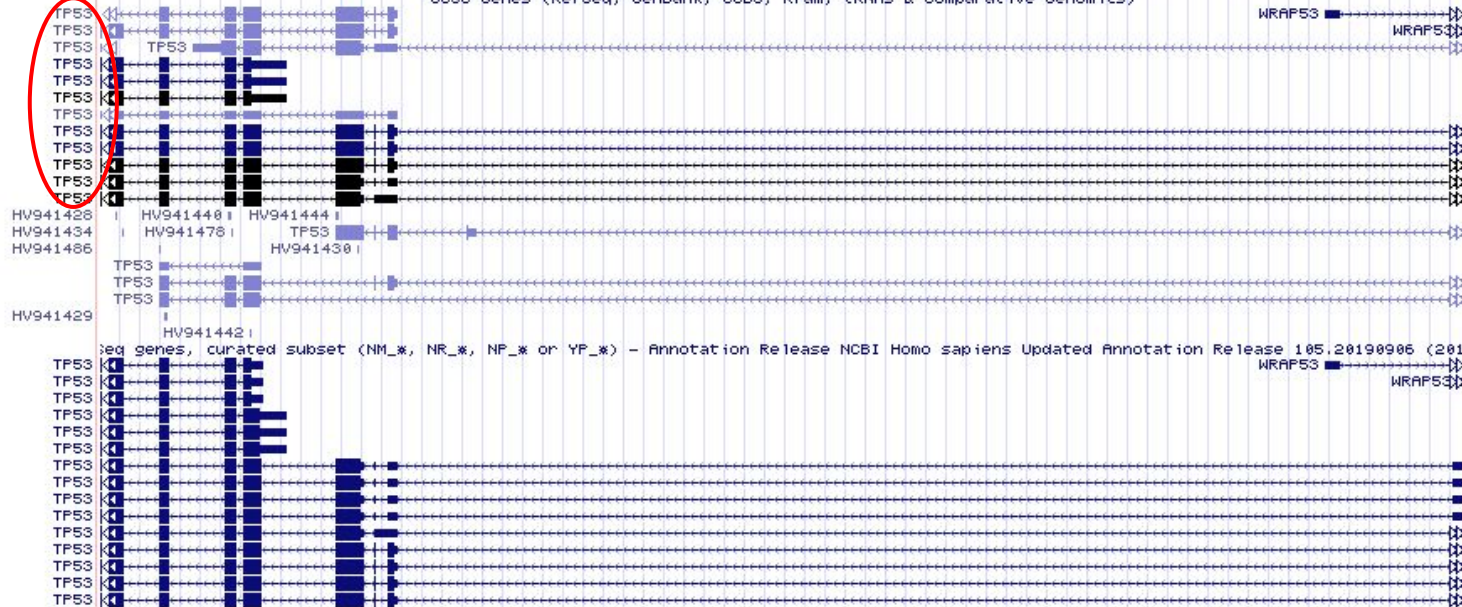
Alt Haplotypes

Seq1

Reference Assembly Fix Patch Sequence Alignments
Reference Assembly Alternate Haplotype Sequence Alignments

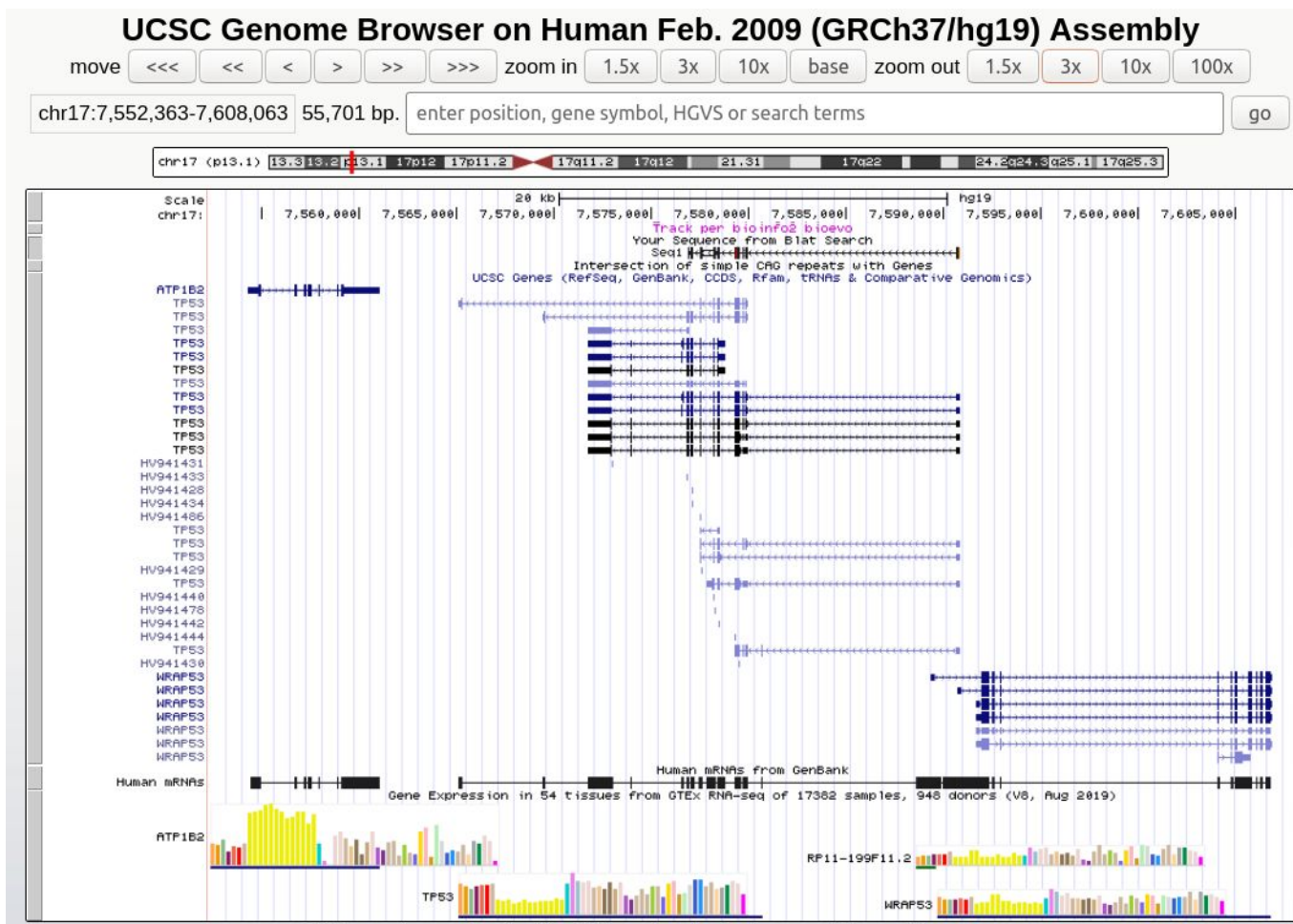
Your Sequence from Blat Search

UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)



Publications: Sequences in Scientific Articles

7 - Zoom out 2 times to 3x to see to which portion of the gene the query sequence corresponds to:



8 - Go back and insert the second sequence into the form:

Human BLAT Search

BLAT Search Genome

Genome: Search all

Assembly:

Query type:

Sort output:

Output type:

Human

Feb. 2009 (GRCh37/hg19) ▾

BLAT's guess

query,score

hyperlink

```
>Seq1
```

[illegible]

```
>Seq2
```

[illegible]☐ All Results (no minimum matches)

Submit

I'm feeling lucky

Clear

Paste in a query sequence to find its location in the the genome. Multiple sequences may be searched if separated by lines starting with '>' followed by the sequence name.

File Upload: Rather than pasting a sequence, you can choose to upload a text file containing the sequence.

Upload sequence: Sfoglia... Nessun file selezionato. submit file



Human (hg19) BLAT Results

BLAT Search Results

Go back to [chr17:7576908-7590832](#) on the Genome Browser.

Custom track name:

Custom track description:

[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22
browser details	Seq2	910	1	996	996	97.1%	chr17	-	7571736	7572719	984
browser details	Seq2	171	551	856	996	85.0%	chr5	+	49950222	49950507	286
browser details	Seq2	167	547	854	996	85.6%	chr10	-	27408468	27408791	324
browser details	Seq2	158	566	844	996	86.0%	chr22	+	28176923	28177196	274
browser details	Seq2	156	552	834	996	83.6%	chr2	+	37248781	37249047	267
browser details	Seq2	155	553	865	996	85.0%	chr6	-	38263541	38263850	310
browser details	Seq2	149	547	865	996	86.4%	chr10	-	104340658	104365268	24611
browser details	Seq2	148	550	856	996	83.6%	chr15	-	43833643	43833952	310
browser details	Seq2	148	561	865	996	82.6%	chr2	+	26597435	26597736	302
browser details	Seq2	147	550	856	996	86.3%	chr4	+	101191092	101191390	299
browser details	Seq2	146	565	857	996	82.5%	chr9	+	80078922	80079207	286
browser details	Seq2	144	548	851	996	89.2%	chr3	-	73757750	73758057	308
browser details	Seq2	143	552	855	996	85.3%	chr6	+	155350920	155351215	296
browser details	Seq2	143	550	844	996	84.8%	chr15	+	75466123	75466409	287
browser details	Seq2	141	551	855	996	85.8%	chr3	-	167045434	167045723	290

seq1

seq2



Human (hg19) BLAT Results

BLAT Search Results

Go back to [chr17:7576908-7590832](#) on the Genome Browser.

Custom track name:

Custom track description:

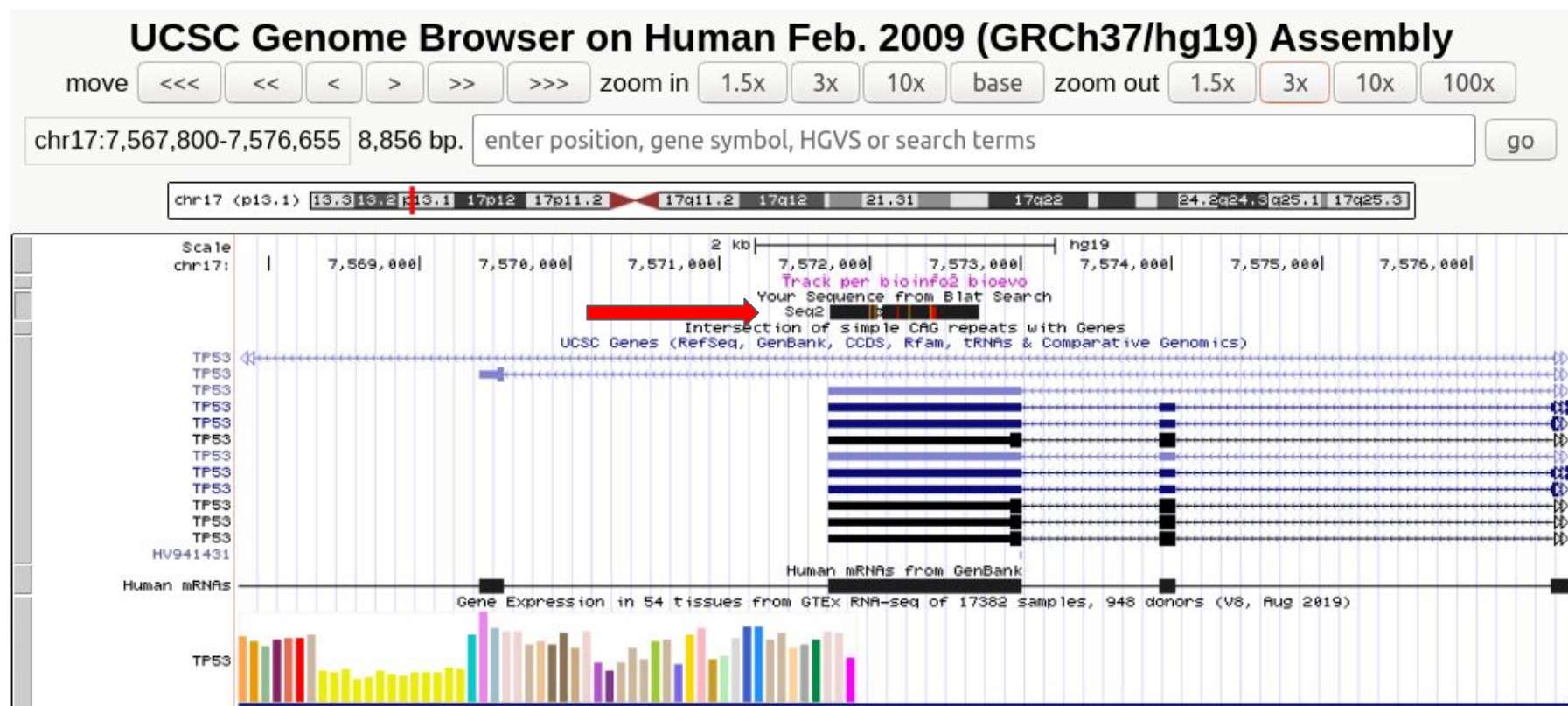
[Build a custom track with these results](#)

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHROM	STRAND	START	END	SPAN
browser details	Seq1	1057	1	1115	1115	99.4%	chr17	-	7576908	7590832	13925
browser details	Seq1	38	471	810	1115	95.4%	chr2	-	37029480	37047471	17992
browser details	Seq1	30	401	450	1115	66.7%	chr1	-	223362777	223362812	36
browser details	Seq1	29	999	1043	1115	96.8%	chr4	-	145830377	145830766	390
browser details	Seq1	25	214	253	1115	70.4%	chr3	+	136385181	136385209	29
browser details	Seq1	21	530	550	1115	100.0%	chr4	-	5763145	5763165	21
browser details	Seq1	20	367	388	1115	95.5%	chr2	+	63944551	63944572	22
browser details	Seq2	910	1	996	996	97.1%	chr17	-	7571736	7572719	984
browser details	Seq2	171	551	856	996	85.0%	chr5	+	49950222	49950507	286
browser details	Seq2	167	547	854	996	85.6%	chr10	-	27408468	27408791	324
browser details	Seq2	158	566	844	996	86.0%	chr22	+	28176923	28177196	274
browser details	Seq2	156	552	834	996	83.6%	chr2	-	37248781	37249047	267
browser details	Seq2	155	553	865	996	85.0%	chr6	-	38263541	38263850	310
browser details	Seq2	149	547	865	996	86.4%	chr10	-	104340658	104365268	24611
browser details	Seq2	148	550	856	996	83.6%	chr15	-	43833643	43833952	310
browser details	Seq2	148	561	865	996	82.6%	chr2	+	26597435	26597736	302
browser details	Seq2	147	550	856	996	86.3%	chr4	+	101191092	101191390	299
browser details	Seq2	146	565	857	996	82.5%	chr9	+	80078922	80079207	286
browser details	Seq2	144	548	851	996	89.2%	chr3	-	73757750	73758057	308
browser details	Seq2	143	552	855	996	85.3%	chr6	+	155350920	155351215	296
browser details	Seq2	143	550	844	996	84.8%	chr15	+	75466123	75466409	287
browser details	Seq2	141	551	855	996	85.8%	chr3	-	167045434	167045723	290

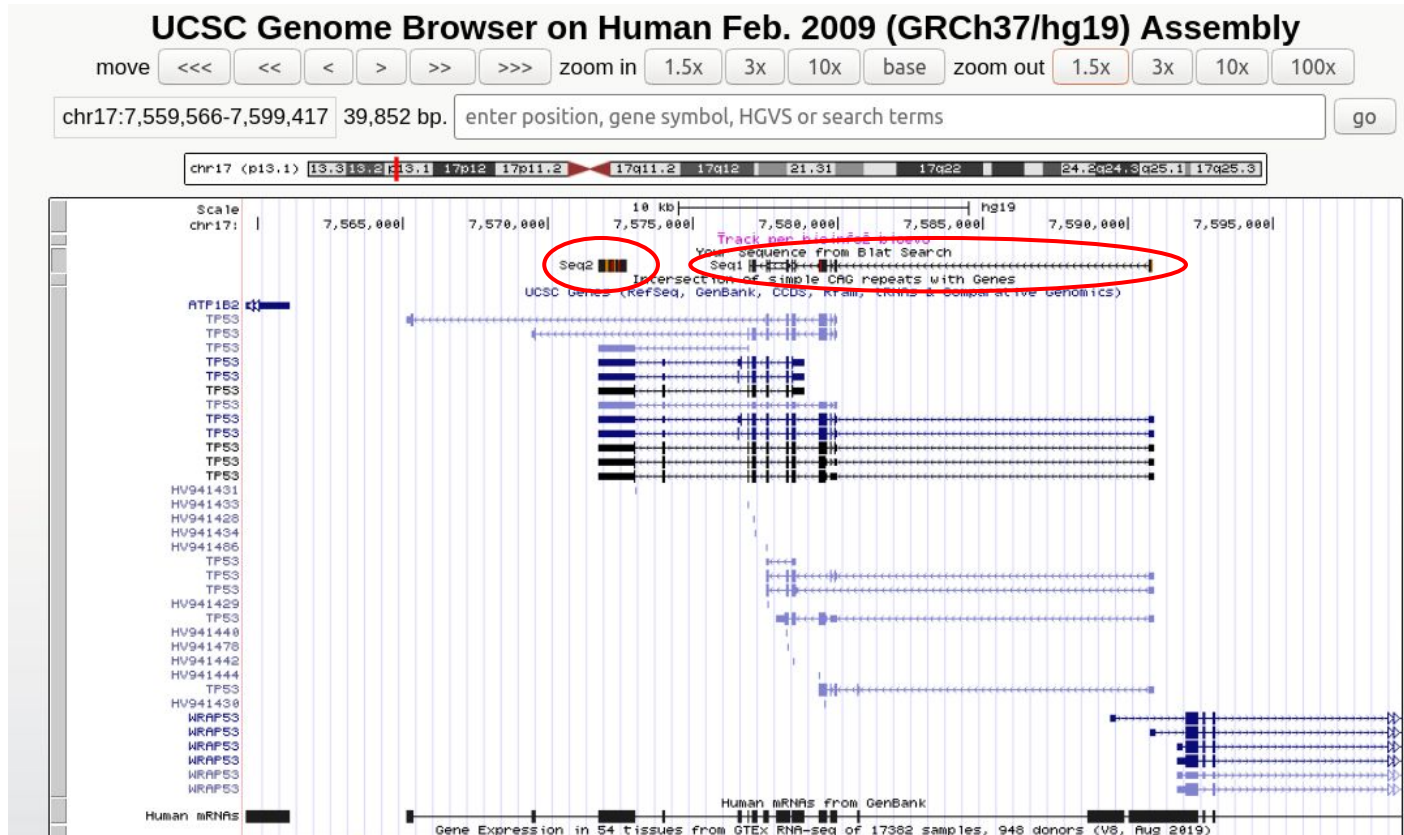
seq1

seq2

9 - Zoom 2 times to 3x to see at which portion of the gene the query sequence corresponds:



10 - Move the browser a little to the left and zoom first to 3x and then to 1.5x to see how the two sequences are positioned on the TP53 gene:



Practical lesson 2: Multiple protein sequence alignment

1. **Given a protein sequence query from *Homo sapiens*, identify the protein to which it belongs and highlight if it belongs to a protein family**
2. Align the protein with some other members of the family also belonging to other species (*Tursiops truncatus*, *Rattus norvegicus*, *Pan troglodytes*) present in the Reference proteins database (RefSeq protein)
3. Obtain and analyse the cladogram and the phylogenetic tree resulting from the multiple alignment

1 - Enter Protein BLAST:

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](#)

NEW

BLAST now supports the new RefSeq Select databases.

The RefSeq Select data-set consists of a representative or “Select” transcript for every protein-coding gene.

Tue, 13 Oct 2020 12:00:00 EST

[More BLAST news...](#)

Web BLAST



BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Human

Mouse

Rat

Microbes

2 - Paste sequence 3 in FASTA format, choose the UniProtKB/Swiss-Prot database and start the search with default parameters:

Standard Protein BLAST

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query sequence

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

`>seqP
VDRPGRIPITSSRCNTIPDDNOVVSLSPGSRVYVLRPVCFEKGNTYVRLLELPYITSSSDVESPYT
LIDSLVLMYPYCKSLDTFTYGGSGGVYTNASAWETEDRYRCLENSRSVYKTPMTDVCNIIIESIALHQ
TGLACECDPQGLSSVCDPNGGQCCRPNVVGRITCNRCAPGTFEGGPGCKPCECHLG`

[Clear](#) Query subrange

From

To

Or, upload file Nessun file selezionato.

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database UniProtKB/Swiss-Prot(**swissprot**)

Organism Optional ☐ exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Exclude Optional ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

BLAST Search database **swissprot** using **Blastp** (protein-protein BLAST)

☐ Show results in a new window

[+ Algorithm parameters](#)

Note: Parameter values that differ from the default are highlighted in yellow and marked with sign

3 - Analyze the result of the similarity search, to understand which human protein corresponds to the sequence fragment under analysis (see “Descriptions”, “Alignments”, “GenPept” and “Gene”):

<div> <div>Descriptions</div> <div>Graphic Summary</div> <div>Alignments</div> <div>Taxonomy</div> </div>									
Sequences producing significant alignments						Download ▾	Manage columns ▾	Show	100 ▾ ?
<input checked="" type="checkbox"/> select all 72 sequences selected						GenPept	Graphics	Distance tree of results	Multiple alignment
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-1; AltName: Full=Laminin B1 chain; AltName: Full=Laminin-1 subunit beta; AltName: Full=Laminin-10 subunit	410	635	100%	4e-132	100.00%	P07942.2		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-1; AltName: Full=Laminin B1 chain; AltName: Full=Laminin-1 subunit beta; AltName: Full=Laminin-10 subunit	393	615	100%	3e-126	94.42%	P02469.3		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-2; AltName: Full=Laminin-11 subunit beta; AltName: Full=Laminin-14 subunit beta; AltName: Full=Laminin-15	176	361	100%	9e-50	45.64%	Q61292.2		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-2; AltName: Full=Laminin chain B3; AltName: Full=Laminin-11 subunit beta; AltName: Full=Laminin-14 subunit	175	362	100%	1e-49	45.69%	P15800.1		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-2; AltName: Full=Laminin B1s chain; AltName: Full=Laminin-11 subunit beta; AltName: Full=Laminin-14 subunit	164	354	100%	9e-46	44.28%	P55268.2		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-4; Flags: Precursor [Danio rerio]	124	410	94%	1e-31	37.43%	Q8JHV6.1		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-1; AltName: Full=Laminin B1 chain; Flags: Precursor [Drosophila melanogaster]	120	347	98%	1e-30	37.50%	P11046.4		
<input checked="" type="checkbox"/>	RecName: Full=Laminin subunit beta-4; AltName: Full=Laminin beta-1-related protein; Flags: Precursor [Homo sapiens]	102	341	88%	4e-24	36.57%	A4D0S4.1		

3 - Analyze the result of the similarity search, to understand which human protein corresponds to the sequence fragment under analysis (see “Descriptions”, “Alignments”, “GenPept” and “Gene”):

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignmentsDownload Manage columns Show 100 ?

LAMB1 laminin subunit beta 1 [Homo sapiens (human)]

Gene ID: 3912, updated on 20-Sep-2020

Summary

Official Symbol

LAMB1 provided by HGNC

Official Full Name

laminin subunit beta 1 provided by HGNC

Primary source

HGNC:HGNC:6486

See related

Ensembl:ENSG00000091136 MIM:150240

Gene type

protein coding

RefSeq status

REVIEWED

Organism

Homo sapiens

Lineage

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo

Also known as

CLM; LIS5

Summary

Laminins, a family of extracellular matrix glycoproteins, are the major noncollagenous constituent of basement membranes. They have been implicated in a wide variety of biological processes including cell adhesion, differentiation, migration, signaling, neurite outgrowth and metastasis. Laminins are composed of 3 non identical chains: laminin alpha, beta and gamma (formerly A, B1, and B2, respectively) and they form a cruciform structure consisting of 3 short arms, each formed by a different chain, and a long arm composed of all 3 chains. Each laminin chain is a multidomain protein encoded by a distinct gene. Several isoforms of each chain have been described. Different alpha, beta and gamma chain isomers combine to give rise to different heterotrimeric laminin isoforms which are designated by Arabic numerals in the order of their discovery, i.e. alpha1beta1gamma1 heterotrimer is laminin 1. The biological functions of the different chains and trimer molecules are largely unknown, but some of the chains have been shown to differ with respect to their tissue distribution, presumably reflecting diverse functions in vivo. This gene encodes the beta chain isoform laminin, beta 1. The beta 1 chain has 7 structurally distinct domains which it shares with other beta chain isomers. The C-terminal helical region containing domains I and II are separated by domain alpha, domains III and V contain several EGF-like repeats, and domains IV and VI have a globular conformation. Laminin, beta 1 is expressed in most tissues that produce basement membranes, and is one of the 3 chains constituting laminin 1, the first laminin isolated from Engelbreth-Holm-Swarm (EHS) tumor. A sequence in the beta 1 chain that is involved in cell attachment, chemotaxis, and binding to the laminin receptor was identified and shown to have the capacity to inhibit metastasis. [provided by RefSeq, Aug 2011]

Expression

Broad expression in placenta (RPKM 75.7), fat (RPKM 65.2) and 23 other tissues See more

Orthologs

mouse all

Exercise 2: Multiple protein sequence alignment

1. Given a protein sequence query from *Homo sapiens*, identify the protein to which it belongs and highlight if it belongs to a protein family
2. **Align the protein with some other members of the family also belonging to other species (*Tursiops truncatus*, *Rattus norvegicus*, *Pan troglodytes*) present in the Reference proteins database (RefSeq protein)**
3. Obtain and analyse the cladogram and the phylogenetic tree resulting from the multiple alignment

1 - Retrieve the reference sequence (Accession number, FASTA), save it in a text file and change the header to make it more readable, e.g. ">LAMB1_HUMAN_precursor":

```
*Documento senza titolo 1
>LAMB1_HUMAN_precursor
MGLLQLLAFSFLALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHL
QEDKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLEAEFHFTHLIMTF
KTRPAAMLIERSDFGKTWGVYRYFAYDCEASFGISTGPMKKVDDIICDSRYSDIEPSTEGEVIFRAL
DPAFKIEDPYPRIQNLLKITNLRIKFVKLHLLGDNLLDSRMEIREKYYYAVYDMVVRGNCFCYGHASEC
APVDGFEVEGVMVGHGCMCRHNTKGLNCEL CMDFYHDL PWRPAEGRNSNACKKCNCEHSISCHFDMAV
YLATGNVSGGVCDQCQNTMGRNCEQCKPFYQHPERDIRDPNFCERCTCDPAGSQNEGICDSYTDFTG
LIAGQCRCKLNVEGEHCDVCKEGFYDLSSDPFGCKSCACNPLGTIPGGNPCDSETHCYCKRLVTGQHC
DQCLPEHWGLSNDLDGCRPCDCDLGGALNNSCFAESGQCSCRPHMIGRCNEVEPGYYFATLDHYLYEAE
EANLPGVSIIVERQYIQDRIPSWTGAGFVRVPEGAYLEFFIDNIPYSMEYDILIRYEPQLPDHWEKAVIT
VQRPGRIPITSSRCGNTIPDDDNQVVSLSPGSRVVLPRPVCFEKGTNYTVRLELPQYTSDDSDVESPYTL
IDSLVLMFYCKSLDIFTVGGSGDGVVTSNAWETFQRYRCLENSRSVVKTPMTDVCRNIIIFSISALLHQTG
LACECDPQGSLSVCDPNGGQCQCRPNVVGRTCNRCAPGTGFGPSGCKPCECHLQGSVNAFCNPVTGQC
HCFQGVYARQCDRCLPGHWGFPSCQPCQCNHADDQDPTGECNLCQDYTMGHNCERCLAGYYGDP IIGS
GDHCRPCPCPDGPDGSRQFARSCYQDPVTLQLACVCDPGYIGSRCDDCASGYFGNPSEVGGSCQPCQCHN
NIDTTDPEACDKETGRCLKLYHTEGEHCQFCRFGYYGDALQQDCRKCVCNYLGTVEHCNGSDCQCDKA
TGQCLCLPNVIGQNCDCAPNTWQLASGTGDCPCNCAAHSGPSCNEFTGQCQCMFGFGGRTCTSECQEL
FWGDPDVECRACDCDPRGIETPQCDQSTGQCVCVEGVEGPRCDKCTRGYSGVFPDCTPCHQCFALWDVII
AELTNRTHRFLKAKALKISGVIGPYRETVDSEVERKVSEIKDILAQSPAAEPLKNIGNLFEAEKLIKDV
TEMAQVEVKLSDTTSQSNSTAKELDSLQTEAESLDNTVKELAEQLEFIKNSDIRGALDSITKYFQMSLE
AEERVNASTTEPNSTVEQSALMRDRVEDVMERESQFKEKQEEQARLLDELAGKLQSLDLASAAEMTCGT
PPGASCSETECGGPNCRTDEGERKCGGPGCGGLVTVAHNAWQKAMDLDQDVL SALAEVEQLSKMVSEAKL
RADEAKQSAEDILLKTNATKEKMDKSNEELRNLIKQIRNFLTQDSADLDSIEAVANEVLKMEMPSTPQQL
QNLTEDIRERVESLSQVEVILQHSAAIARAEMLL EEAKRASKSATDVKVTADMVKEALEEAEKAQVAAE
KAIKQADEDIQGTQNLTSIESETAASEETLFNASQRISIELERNVEELKRKAAQNSGEAEYIEKVYTVK
QSAEDVKKTLDGELDEKYKKVENLIAKKTEESADARRKAEMLQNEAKTLLAQANSKLQLLKDLERKYEDN
QRYLEDKAQELARLEGEVRSLLKDISQKVAVYSTCL
```

2 - Return to Protein BLAST, use the sequence just retrieved as the query, limiting the search to “Reference proteins (refseq protein)” of proteins in the taxonomic groups *Homo Sapiens* , *Tursiops truncatus* , *Rattus norvegicus* and *Pan troglodytes* and start the search:

Standard Protein BLAST

blastn **blastp** blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

MGLLQLLAFLALCRARVRAQEPEFSYGAEGSCYPATGDLIGRAQKLSV
TSTCGLHKPEPYCIVSHLQEDKKCFICNSQDPYHETLNPDSHLIENVVTFAP
NRLKIWWQSENGVENVTIQDLAEAFHFTHLIMTFKTRPAAMLIERSSDFC

Or, upload file Nessun file selezionato [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database [?](#)

Organism Optional

☐ exclude [Add organism](#)

☐ exclude

☐ exclude

☐ exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Optional

☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

☐ PSI-BLAST (Position-Specific Iterated BLAST)

☐ PHI-BLAST (Pattern Hit Initiated BLAST)

☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

QuickBLASTP is an accelerated version of BLASTP that is very fast and works best if the target percent identity is 50% or more.

BlastP simply compares a protein query to a protein database.

PSI-BLAST allows the user to build a PSSM (position-specific scoring matrix) using the results of the first BlastP run.

PHI-BLAST performs the search but limits alignments to those that match a pattern in the query.

DELTA-BLAST constructs a PSSM using the results of a Conserved Domain Database search and searches a sequence database.

3 - Select the sequences of *Homo sapiens*, *Pan troglodytes*, *Tursiops truncatus*, *Rattus norvegicus* of the laminin **beta 1 and 3 family** (if there are more isoforms, choose “precursor” for *Hs* and “X1” for *Rn*)

FASTA format (“**Download**”) and copy them into a text file.

5- Modify the sequence headers to make them short, informative and unique (e.g. “>Hs_lamb1_pre”, “>Pt_lamb1”)

Sequences producing significant alignments									
Download ▼ New Select columns ▼ Show 100 ▼ ?									
<input type="checkbox"/> select all	8 sequences selected								
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	laminin subunit beta-1 precursor [Homo sapiens]	Homo sapiens	3705	3705	100%	0.0	100.00%	1786	NP_002282.2
<input checked="" type="checkbox"/>	laminin subunit beta-1 [Pan troglodytes]	Pan troglodytes	3691	3691	100%	0.0	99.61%	1786	XP_001165667.4
<input type="checkbox"/>	laminin subunit beta-1 isoform X1 [Homo sapiens]	Homo sapiens	3678	3678	99%	0.0	99.72%	1810	XP_016867690.1
<input checked="" type="checkbox"/>	laminin subunit beta-1 [Tursiops truncatus]	Tursiops truncatus	3492	3492	100%	0.0	93.17%	1786	XP_033718774.1
<input type="checkbox"/>	laminin subunit beta-1 isoform X2 [Rattus norvegicus]	Rattus norvegicus	3438	3438	100%	0.0	92.61%	1786	XP_006240058.1
<input checked="" type="checkbox"/>	laminin subunit beta-1 isoform X1 [Rattus norvegicus]	Rattus norvegicus	3437	3437	100%	0.0	92.61%	1834	XP_003750185.1
<input type="checkbox"/>	laminin subunit beta-1 isoform X2 [Homo sapiens]	Homo sapiens	2368	2688	65%	0.0	99.56%	1212	XP_016867691.1
<input type="checkbox"/>	laminin subunit beta-2 isoform X2 [Tursiops truncatus]	Tursiops truncatus	1907	1907	98%	0.0	51.28%	1803	XP_033720801.1
<input type="checkbox"/>	laminin subunit beta-2 isoform X1 [Rattus norvegicus]	Rattus norvegicus	1905	1905	99%	0.0	51.18%	1801	XP_006243771.1
<input type="checkbox"/>	laminin subunit beta-2 precursor [Rattus norvegicus]	Rattus norvegicus	1905	1905	99%	0.0	51.18%	1801	NP_037106.1
<input type="checkbox"/>	laminin subunit beta-2 isoform X2 [Rattus norvegicus]	Rattus norvegicus	1901	1901	99%	0.0	51.18%	1800	XP_038936809.1
<input type="checkbox"/>	laminin subunit beta-2 isoform X1 [Tursiops truncatus]	Tursiops truncatus	1899	1899	98%	0.0	51.10%	1809	XP_033720799.1
<input type="checkbox"/>	laminin subunit beta-2 [Pan troglodytes]	Pan troglodytes	1882	1882	98%	0.0	50.65%	1798	XP_016796574.2
<input type="checkbox"/>	laminin subunit beta-2 precursor [Homo sapiens]	Homo sapiens	1878	1878	98%	0.0	50.65%	1798	NP_002283.3
<input type="checkbox"/>	laminin subunit beta-2 isoform X3 [Rattus norvegicus]	Rattus norvegicus	1763	1763	93%	0.0	50.51%	1667	XP_038936810.1
<input type="checkbox"/>	laminin subunit beta-4 isoform 1 precursor [Homo sapiens]	Homo sapiens	1358	1358	98%	0.0	41.31%	1761	NP_001304975.1
<input type="checkbox"/>	laminin subunit beta-4 isoform X1 [Homo sapiens]	Homo sapiens	1336	1336	95%	0.0	41.32%	1772	XP_011514277.1
<input type="checkbox"/>	laminin subunit beta-4 isoform X2 [Homo sapiens]	Homo sapiens	1336	1336	95%	0.0	41.32%	1753	XP_016867368.1
<input type="checkbox"/>	laminin subunit beta-4 isoform X3 [Homo sapiens]	Homo sapiens	1334	1334	95%	0.0	41.31%	1723	XP_011514280.1
<input type="checkbox"/>	laminin subunit beta-4 isoform X4 [Homo sapiens]	Homo sapiens	1334	1334	96%	0.0	41.22%	1703	XP_011514281.1
<input type="checkbox"/>	laminin subunit beta-4 isoform X6 [Homo sapiens]	Homo sapiens	1312	1312	86%	0.0	43.34%	1564	XP_011514282.1
<input type="checkbox"/>	laminin subunit beta-4 isoform X5 [Homo sapiens]	Homo sapiens	1311	1311	86%	0.0	43.34%	1567	XP_016867369.1
<input type="checkbox"/>	laminin subunit beta-4 isoform 2 precursor [Homo sapiens]	Homo sapiens	1081	1375	64%	0.0	49.77%	1101	NP_001304976.1
<input type="checkbox"/>	laminin subunit beta-4 [Pan troglodytes]	Pan troglodytes	739	1274	62%	0.0	49.93%	772	XP_016813513.2
<input type="checkbox"/>	laminin subunit beta-4 isoform 3 precursor [Homo sapiens]	Homo sapiens	736	1271	62%	0.0	49.79%	772	NP_001304977.1
<input checked="" type="checkbox"/>	laminin subunit beta-3 precursor [Rattus norvegicus]	Rattus norvegicus	416	1133	87%	3e-121	43.54%	1172	NP_001094311.1
<input checked="" type="checkbox"/>	laminin subunit beta-3 [Pan troglodytes]	Pan troglodytes	410	1046	87%	2e-119	41.56%	1171	XP_003949737.2
<input checked="" type="checkbox"/>	laminin subunit beta-3 precursor [Homo sapiens]	Homo sapiens	408	1091	87%	1e-118	41.19%	1172	NP_000219.2
<input checked="" type="checkbox"/>	laminin subunit beta-3 [Tursiops truncatus]	Tursiops truncatus	400	927	85%	5e-116	42.25%	1172	XP_033708792.1

Exercise 2: Multiple protein sequence alignment

1. Given a protein sequence query from *Homo sapiens*, identify the protein to which it belongs and highlight if it belongs to a protein family
2. Align the protein with some other members of the family also belonging to other species (*Tursiops truncatus*, *Rattus norvegicus*, *Pan troglodytes*) present in the Reference proteins database (RefSeq protein)
3. **Obtain and analyse the cladogram and the phylogenetic tree resulting from the multiple alignment**

Muscle

- Performs multiple protein sequence alignments
- Highly optimized both in terms of time and accuracy of results (can be used to align up to 1000+ sequences)
- Based on progressive alignment followed by iterative refinements

1 - Connect to the EBI **Muscle** home page dedicated to multiple alignments
(<https://www.ebi.ac.uk/jdispatcher/msa/muscle?style=protein>) .

2 - Select the fasta file with the 8 sequences and start the analysis with the default parameters:

The screenshot shows the EBI Muscle web interface. At the top, there's a header with the 'Muscle' logo and 'Multiple Sequence Alignment (MSA)'. Below the header, there are navigation links: 'Job Dispatcher', 'Help & Privacy', 'Your Jobs', and 'Input form'. A 'Feedback' button is also present. A yellow banner message reads: 'Welcome to the new Job Dispatcher website. We'd love to hear your feedback about the new webpages!'. The main content area explains that MUSCLE stands for Multiple Sequence Comparison by Log-Expectation and provides details about its accuracy and speed compared to ClustalW2 and T-Coffee. It also states the tool can align up to 500 sequences or a maximum file size of 1 MB. The 'Input sequence' section includes a text area for pasting sequences, a file selection button labeled 'Sfoglia...' with 'seqdump(1).txt' as the chosen file, and buttons for 'Use the example' and 'Clear sequence'. A link for 'More example inputs' is also available. The 'Parameters' section shows the 'OUTPUT FORMAT' dropdown menu set to 'ClustalW'.

Muscle
Multiple Sequence Alignment (MSA)

Job Dispatcher Help & Privacy Your Jobs **Input form** **Feedback**

Welcome to the new **Job Dispatcher** website. We'd love to hear your [feedback](#) about the new webpages! ×

MUSCLE stands for Multiple Sequence Comparison by Log-Expectation. MUSCLE is claimed to achieve both better average accuracy and better speed than ClustalW2 or T-Coffee, depending on the chosen options. This tool can align up to 500 sequences or a maximum file size of 1 MB.

Input sequence ⓘ **Paste your sequence here - or use the example sequence**

Sfoglia... seqdump(1).txt Use the example Clear sequence More example inputs

Parameters **OUTPUT FORMAT** ⓘ
ClustalW ▼
[More options](#) ↗

3 - Click on “Show colors” to display a color for each group of amino acids (polar, non-polar, acids):

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

```
Rattus_norvegicus_LAMB1      MERPLSSLPTFAFSPLYLSRKEDTKRRAACLP RPSPRAASRLQGLDMGVLQVFAFGVL
Tursiops_truncatus_LAMB1    -----MGRLLQMFASFSL
Homo_sapiens_LAMB1          -----MGLLQLLAFSFL
Pan_troglodytes_LAMB1       -----MGLLQLLAFSFL
Rattus_norvegicus_LAMB3     -----MMTALFLLWLALP
Tursiops_truncatus_LAMB3    -----MRPLLLLYFVLP
Pan_troglodytes_LAMB3       -----MRPFLLCFALP
Homo_sapiens_LAMB3          -----MRPFLLCFALP
```

“*” = same amino acid
for all sequences

```
Rattus_norvegicus_LAMB1      ALWGRVCAQEPEFSYGC AEGSCYPATGDLLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
Tursiops_truncatus_LAMB1    ALCGAQVSAQEPEFSYGC AEGSCYPATGDLLIGRAQKLSVTSTCGLYKPEPYCIVSHLQE
Homo_sapiens_LAMB1          ALCRARVRAQEPEFSYGC AEGSCYPATGDLLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
Pan_troglodytes_LAMB1       ALCRARVRAQEPEFSYGC AEGSCYPATGDLLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
Rattus_norvegicus_LAMB3     GL----LCAQQP-----CSRGACYPVGDLLIGRTQLLRASSTCGLTKPETYC-TQYQGW
Tursiops_truncatus_LAMB3    SV----LCAQQA-----CSRGACYPVGDLLIGRTRFLRASSTCGLAKPETYC-TQYGEW
Pan_troglodytes_LAMB3       GL----LHAQQA-----CSRGACYPVGDLLVGRTRFLRASSTCGLTKPETYC-TQYGEW
Homo_sapiens_LAMB3          GL----LHAQQA-----CSRGACYPVGDLLVGRTRFLRASSTCGLTKPETYC-TQYGEW
.:      : **: .      * : ** : . : ** : ** : . : ** : ** : . : :
```

“:” = conservative
substitution

```
Rattus_norvegicus_LAMB1      DKKCFICDSRDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTQLDLAEAFH
Tursiops_truncatus_LAMB1    DKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
Homo_sapiens_LAMB1          DKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
Pan_troglodytes_LAMB1       DKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLAEAFH
Rattus_norvegicus_LAMB3     QMKCKCKDSRLPR---NYYSHRVENVVSSSGPMR---WWQSQNDVSPVSLQLDLDRRFQ
Tursiops_truncatus_LAMB3    QMKCKCKDSRLPH---NYYSHRVENVVSSSGPMR---WWQSQNDVSPVSLQLDLDRRFQ
Pan_troglodytes_LAMB3       QMKCKCKDSRQPH---NYYSHRVENVVSSSGPMR---WWQSQNDVNPVSLQLDLDRRFQ
Homo_sapiens_LAMB3          QMKCKCKDSRQPH---NYYSHRVENVVSSSGPMR---WWQSQNDVNPVSLQLDLDRRFQ
: ** *: . *      * ** : ** : . : * *      ** : ** : . : ** : ** : :
```

“.” = semi-conservative
substitution

“ ” = non-conservative
substitution

4 - Click on “Phylogenetic Tree” to view the phylogenetic tree:

Tool Output	Alignments	Phylogenetic Tree	Results Viewers	Result Files	Submission Details
-------------	------------	-------------------	-----------------	--------------	--------------------

Phylogenetic Tree

```
(
  Rattus_norvegicus_LAMB1:0.03710,
  (
    (
      Tursiops_truncatus_LAMB1:0.03657,
      (
        Homo_sapiens_LAMB1:0.00149,
        Pan_troglodytes_LAMB1:0.00299)
        :0.03118)
    :0.00937,
    (
      Rattus_norvegicus_LAMB3:0.08783,
      (
        Tursiops_truncatus_LAMB3:0.07624,
        (
          Pan_troglodytes_LAMB3:0.00809,
          Homo_sapiens_LAMB3:0.01049)
          :0.06524)
        :0.01733)
    :0.51843);
```

Phylogram

