

MOLECULAR METHODS AND BIOINFORMATICS

**LM Evolutionary Biology, University of Padova
AA 2024/2025**

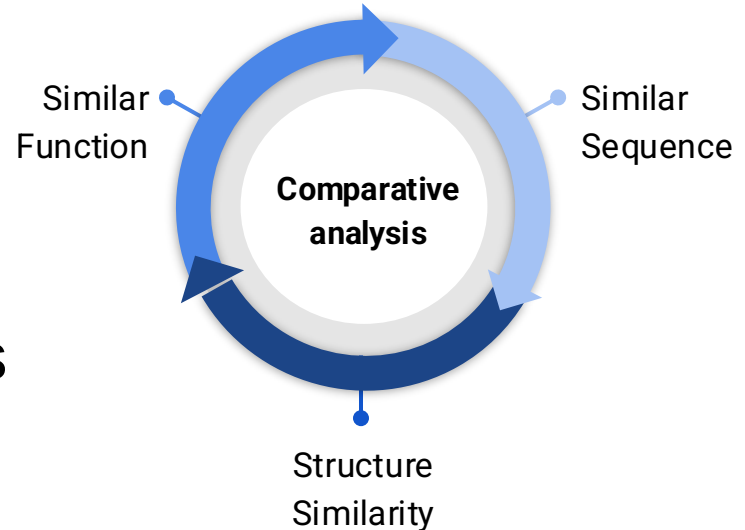
**Enrico Gaffo, Silvia Orsi,
Prof. Stefania Bortoluzzi**

“Comparison of biosequences and multiple alignments”

Padua, November 4, 2024

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?

I discovered a new gene and I want to see if:

- the gene is already present in nucleic acid databases (EMBL/Genbank), even if only in part
- the gene is similar to other genes present in the databases
- the gene is similar to a particular sequence

Protein Alignment: Why?

- I'm not sure if the protein I discovered is a "new" protein, is it already known?
- the protein I discovered is a “new” one, but does it resemble any known protein?
- In the protein I am studying (which is not similar to any other protein), are there portions similar to known proteins, or functional domains known in other species?

Biosequence Alignment: Programs

Pairwise sequence alignment (PSA):

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

Multiple sequence alignment (MSA):

- **Muscle**

BLAST (Basic Local Alignment Search Tool, NCBI)

- Align DNA, mRNA or protein sequences
- Allows you to compare biological sequences against constantly updated sequence databases
- Alignment not as precise as Needleman-Wunsch and Smith-Waterman algorithms but much faster
- It includes a series of programs that vary according to the type of sequences to be compared (blastn, blastp, tblastn, tblastx).

What can BLAST be useful for?

- **Identify a species or a homologous species** (for example, a DNA sequence of an unknown species is known)
- Locating the **functional domains** of a protein
- Mapping a DNA sequence to find the **chromosomal location**
- **Locate common 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.*

- Align DNA, mRNA or protein sequences with genomic sequences
- BLAST-like algorithm
- 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 sequences** on the genome to find the coordinates of the gene that encodes it
- Show the **coding sequences** of a gene
- Determine **the distribution** of exons and introns in a gene
- **Cross-species analysis**: alignment of a protein or mRNA from one species against a sequence database from another species
- **Align two protein sequences** (less precise than BLAST)

Exercise 2: Nucleotide sequence alignment

1. Find out **if** the query nucleotide sequence “ *seq1*” of *Homo Sapiens* 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)

Exercise 2: Nucleotide sequence alignment

1. Find out **if** the query nucleotide sequence “ *seq1*” of *Homo Sapiens* 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

Nucleotide Blast (NCBI)

2

BLAT (UCSC)

3

Protein Blast and Muscle

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

http://compgen.bio.unipd.it/~stefania/Didattica/AA2024-2025/MMOL_BIOINFO_EB/Practical_session_3.zip

2 - unzip the file Practical_session_3.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](#)

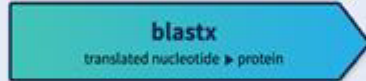
BLAST now supports the new RefSeq Select databases.

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 taxid

Search

Human

Mouse

Rat

Microbes

2 - Enter seq1 as query sequence and call the analysis “Seq1”:

BLAST[®] » blastn suite [Home](#)

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 Nessun file selezionato. [?](#)


Job Title


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

☐ Align two or more sequences [?](#)

FASTA sequence:

```
>Seq1
actggcgcctaaaagtgttgacgtttcctttacaattctcaaaagctcagagccaccgtccagggagcaggtagctg
ctgggtccggggacactttcgttcgggctgggagcgtgcttccacgacggtagacagcttccctgattggc
agccagacttccttcgggtcactgccatggaggagccgcagtcagatcctagcgtcagacccctctgagtc
```





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 “Homo sapiens” 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

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

[New](#) ☐ Experimental databases [Try experimental taxonomic nt databases](#) [Download](#)

For more info see [What are taxonomic nt databases?](#)

Nucleotide collection (nr/nt) [?](#)

Homo sapien (taxid:9606) ☐ [Exclude organism](#)

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

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

☐ Sequences from type material

Exclude ☐ Limit to [YouTube](#) Create custom database

Entrez Query Enter an Entrez query to limit search [?](#)

3 - Check 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 “Homo sapiens” 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 ?

>Seq1
 atggcctttaaagatgacttcctttcaatttcaaagtcttagagccacotccagggaacaggaagctc
 ctgggcctcgaaacactttagcttgcgaactggaagcctgcttcacgaacggaacaccttccttaaatgac
 agccagaactacccttcgaatcactgcacatgaagagcccpavtcagatcctaagcttcgaagccccctctgatgc

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

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

New ☐ Experimental databases

Try experimental taxonomic nt databases
 For more info see What are taxonomic nt databases? [Download](#)

Nucleotide collection (nr/nt) ▼ ?

Homo sapien (taxid:9606) ☐ experimental organism

Organism
Optional

Exclude
Optional

Limit to
Optional

Entrez Query
Optional

Title: Nucleotide collection (nr)

Description: The nucleotide collection consists of GenBank+EMBL+DDBJ+PDB+RefSeq sequences, but excludes EST, STS, GSS, W, 0, 1, and 2 HTGS sequences and sequences longer than 100Mb. The database is non-redundant. Identical sequences have been merged by accession, GI, title and taxonomy information for each entry.

Molecule Type: mixed DNA

Update date: 2023/10/30

Number of sequences: 99985738

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 **Nucleotide collection (nr/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 Nucleotide collection (nr/nt)** using **Megablast (Optimize for highly similar sequences)**

☐ Show results in a new window

+ Algorithm parameters



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

BLAST[®] » blastn suite » results for RID-TZSWRRGG013

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](#)

i Your search is limited to records that include: Homo sapiens (taxid:9606)

Job Title	Seq1
RID	TZSWRRGG013 Search expires on 11-03 17:46 pm Download All ▼
Program	BLASTN ? Citation ▼
Database	nt See details ▼
Query ID	lcl Query_32607
Description	Seq1
Molecule type	dna
Query Length	1115
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](#)

Research information

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 Manage columns Show 100

☒ select all 100 sequences selected

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

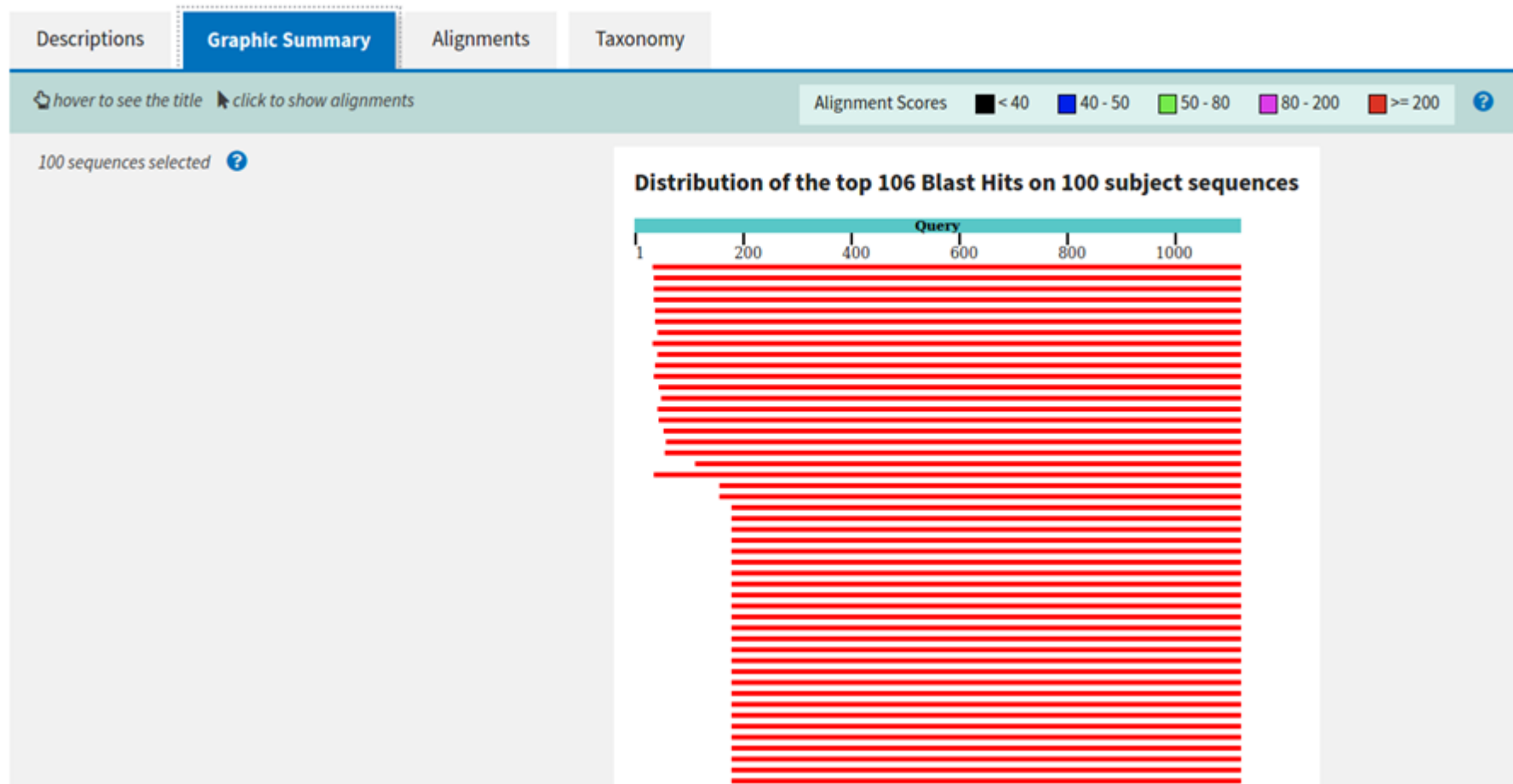
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)	1858	1858	96%	0.0	97.70%	HG975427.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA	1855	1855	96%	0.0	97.69%	NM_000546.6
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 3, mRNA	1855	1855	96%	0.0	97.69%	NM_001276696.3
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 4, mRNA	1855	1855	96%	0.0	97.69%	NM_001276695.3
<input checked="" type="checkbox"/>	Homo sapiens mRNA for Cellular tumor antigen p53 variant, clone: FCC127A11	1844	1844	96%	0.0	97.59%	AK225838.1
<input checked="" type="checkbox"/>	Human p53 cellular tumor antigen mRNA, complete cds	1844	1844	96%	0.0	97.59%	K03199.1

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 “Graphic Summary” to see the quality of the alignments:



What is the best alignment?

Descriptions							Graphic Summary		Alignments		Taxonomy	
Sequences producing significant alignments							Download		Manage columns		Show 100	
<input checked="" type="checkbox"/> select all 100 sequences selected							GenBank		Graphics		Distance tree of results	
	Description						Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)						1858	1858	96%	0.0	97.70%	HG975427.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA						1855	1855	96%	0.0	97.69%	NM_000546.6
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 3, mRNA						1855	1855	96%	0.0	97.69%	NM_001276696.3
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 4, mRNA						1855	1855	96%	0.0	97.69%	NM_001276695.3
<input checked="" type="checkbox"/>	Homo sapiens mRNA for Cellular tumor antigen p53 variant, clone: FCC127A11						1844	1844	96%	0.0	97.59%	AK225838.1
<input checked="" type="checkbox"/>	Human p53 cellular tumor antigen mRNA, complete cds						1844	1844	96%	0.0	97.59%	K03199.1

What is the best alignment?

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage columns

Show100

☒

select all

100 sequences selected

GenBank

Graphics

Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)	1858	1858	96%	0.0	97.70%	HG975427.1
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 1, mRNA	1855	1855	96%	0.0	97.69%	NM_000546.6
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 3, mRNA	1855	1855	96%	0.0	97.69%	NM_001276696.3
<input checked="" type="checkbox"/>	Homo sapiens tumor protein p53 (TP53), transcript variant 4, mRNA	1855	1855	96%	0.0	97.69%	NM_001276695.3
<input checked="" type="checkbox"/>	Homo sapiens mRNA for Cellular tumor antigen p53 variant, clone: FCC127A11	1844	1844	96%	0.0	97.59%	AK225838.1
<input checked="" type="checkbox"/>	Human p53 cellular tumor antigen mRNA, complete cds	1844	1844	96%	0.0	97.59%	K03199.1

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

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

TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

Sequence ID: [HG975427.1](#) Length: 2586 Number of Matches: 1

Range 1: 54 to 1135 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1858 bits(1006)	0.0	1061/1086(98%)	9/1086(0%)	Plus/Plus

Query	35	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGCACT	94
Sbjct	54	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGCACT	113
Query	95	TTGCGTTCGGGCTGGGAGCGTGCTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	154
Sbjct	114	TTGCGTTCGGGCTGGGAGCGTGCTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	173
Query	155	CAGACTGCCCTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	214
Sbjct	174	CAGACTGCCCTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	233
Query	215	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAAACACGTTCTG	274
Sbjct	234	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAACTACTTCTGAAAAACACGTTCTG	293
Query	275	TCCCCCTTCCCGTCCCAAGCAATGGATGATTGATGCTGTCCCGGACGATATTGAACAA	334
Sbjct	294	TCCCCCTTCCCGTCCCAAGCAATGGATGATTGATGCTGTCCCGGACGATATTGAACAA	353
Query	335	TGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAAGATGCCAGAGGCTGCTCCCCC	394
Sbjct	354	TGGTTCACTGAAGACCCAGGTCCAGATGAAGCTCCCAAGATGCCAGAGGCTGCTCCCCC	413
Query	395	GTGGCCCTGCACACAGAGCTCCTACACGGCGGCCCTGCACACGCCCTCTCTGGCCC	454
Sbjct	414	GTGGCCCTGCACACAGAGCTCCTACACGGCGGCCCTGCACACGCCCTCTCTGGCCC	473

alignment display
mode

target sequence
information

alignment information

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

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

[Next](#) [Previous](#) [Descriptions](#)

TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

Sequence ID: [HG975427.1](#) Length: 2586 Number of Matches: 1

Range 1: 54 to 1135 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1858 bits(1006)	0.0	1061/1086(98%)	9/1086(0%)	Plus/Plus
Query 35	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	94		
Sbjct 54	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	113		
Query 95	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	154		
Sbjct 114	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	173		
Query 155	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	214		
Sbjct 174	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGAGTCAGATCCTAGCGTCGAGCCC	233		
Query 215	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTGAAAACAACGTTCTG	274		
Sbjct 234	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTGAAAACAACGTTCTG	293		
Query 275	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA	334		
Sbjct 294	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA	353		

Related Information

[Gene](#) - associated gene details
[PubChem BioAssay](#) - bioactivity screening

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

TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

GenBank: HG975427.1

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

[Go to:](#) 

LOCUS HG975427 2586 bp RNA linear PRI 03-MAR-2015
DEFINITION TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene).
ACCESSION HG975427
VERSION HG975427.1
DBLINK BioProject: [PRJEB6238](#)
KEYWORDS Third Party Data; TPA; RNACentral; TPA:specialist_db.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Amaral,P.P., Clark,M.B., Gascoigne,D.K., Dinger,M.E. and
Mattick,J.S.
TITLE lncRNAdb: a reference database for long noncoding RNAs
JOURNAL Nucleic Acids Res 39 (DATABASE ISSUE), D146-D151 (2011)
PUBMED [21112873](#)
REFERENCE 2 (bases 1 to 2586)
CONSTRM lncRNAdb and RNACentral
TITLE Direct Submission
JOURNAL Submitted (30-APR-2014)
COMMENT lncRNAdb; 163; p53 mRNA.

Specialist DB : lncRNAdb (The long non-coding RNA database) URL
: <http://www.lncrnadb.org/>

PRIMARY	TPA_SPAN	PRIMARY_IDENTIFIER	PRIMARY_SPAN	COMP
	1-77	DA932053.1	1-77	
	78-411	BC003596.1	1-334	
	412-1379	X02469.1	350-1317	
	1380-2567	AK223026.1	478-1665	
	2568-2586	DQ186650.1	2257-2275	

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

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

[Next](#) [Previous](#) [Descriptions](#)

TPA: Homo sapiens Processed transcript p53-mRNA (p53 gene)

Sequence ID: [HG975427.1](#) Length: 2586 Number of Matches: 1

Range 1: 54 to 1135 [GenBank](#) [Graphics](#)

[Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1858 bits(1006)	0.0	1061/1086(98%)	9/1086(0%)	Plus/Plus
Query 35	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	94		
Sbjct 54	TTCTCAAAAGTCTAGAGCCACCGTCCAGGGAGCAGGTAGCTGCTGGGCTCCGGGGACACT	113		
Query 95	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	154		
Sbjct 114	TTGCGTTCGGGCTGGGAGCGTGCTTTCCACGACGGTGACACGCTTCCCTGGATTGGCAGC	173		
Query 155	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCC	214		
Sbjct 174	CAGACTGCCTTCCGGGTCACTGCCATGGAGGAGCCGCAGTCAGATCCTAGCGTCGAGCCC	233		
Query 215	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTGAAAACAACGTTCTG	274		
Sbjct 234	CCTCTGAGTCAGGAAACATTTTCAGACCTATGGAAACTACTTCTGAAAACAACGTTCTG	293		
Query 275	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA	334		
Sbjct 294	TCCCCCTTGCCGTCCCAAGCAATGGATGATTTGATGCTGTCCCCGGACGATATTGAACAA	353		

Related Information

[Gene](#) - associated gene details
[PubChem BioAssay](#) - bioactivity screening

Summary exercises

- To derive this information: In which tissues is TP53 most highly expressed?
- Go back and select the sequences with at least 97% identity and 90% coverage, and with E-value = 0 (***filters***). How many are there?

Exercise 2: Nucleotide sequence alignment

1. Find out if the query nucleotide sequence “ *seq1*” of *Homo Sapiens* 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

Nucleotide Blast (NCBI)

2

BLAT (UCSC)

3

Protein Blast and Muscle

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:

Human BLAT Search

BLAT Search Genome

Genome: ☐ Search all Human

Assembly: Feb. 2009 (GRCh37/hg19)

Query type: BLAT's guess

Sort output: query,score

Output type: hyperlink

>Seq1

```
actggcgctaaaagttttgagctttcctttacaattctcaaaagtctagagccaccgtccaggaggacaggttagctgctgggctccggggacactttgcgttcgggctgggagcgtgctttccacgacggtgacacgcttccctg
gattggcagccagactgccttcgggtcactgcatggaggagcgcagtcagatcctagcgtcgagccctctgagtcaggaaacattttcagacctatggaaactacttctgaaaacaacgcttctgtcccttgcgctc
ccaagcaatggatgatttgatgctgtccccggacgatattgaacaatggttcaactgaagaccaggtccagatgaagctcccagaatgccagaggctgctcccccggtggccctgcaccagcagctcctacaccggcgccccc
tgaccagccccctcctggccccctgtcatcttctgtcccttccagaaaaactaccaggggacctacggtttccgctcgggcttcttgcattctgggacagccaagctgtgacttgccagctactccccctgccctcaacaagat
gttttgccaactggccaagacctgccctgtgcagctgtgggttgattccacacccccgccggctttacgcgcatggccatctacaagcagtcacagcacatgacggagggttgtagggcgctgccccaccatgagcgctgct
cagatagcgatggtctggccccctcctcagcatcttatccgagtggaaaccttgatttgcgtgtggagtatttgatgacagaaacacttttcgacatagtggtggtggtgccctttaccatgctgaggttggtctgactgtac
caccatccactcattaactacatgtgtaacagttcctgcattggggcgcatgaaccggaggcccatcctcaccatcatcacactggaagactccagtggtaatctactgggacggaacagctttgaggtgcgtgtttgtgcctgt
cctgggagagaccggcgccacagaggaagagaatctccgaagaaaggggagcctcaccacagactgccccaggagcactaagcgagcactgccaacaacaccag
```

☐ All Results (no minimum matches)

Submit I'm feeling lucky Clear



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:

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 ttctcaaaagcttagagccaccgtccagggagcaggtagctgctgggctc 0000084
<<<<<<< ||||| <<<<<<<
7590810 ttctcaaaagcttagagccaccgtccagggagcaggtagctgctgggctc 7590761
```

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

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

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

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

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

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

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084  

<<<<<< ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| <<<<<<  

7590810 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 7590761
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgcttccacgacggtgaca 0000134  
<<<<<< ||||| ||||| ||||| ||||| ||||| ||||| ||||| <<<<<<  
7590760 cggggacactttgcgttcgggctgggagcgtgcttccacgacggtgaca 7590711
```

```
0000135 cgcttcctggattgg 0000150
<<<<<<< |||||  <<<<<<<
7590710 cgcttcctggattgg 7590695
```

```
0000151   cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc   0000200  

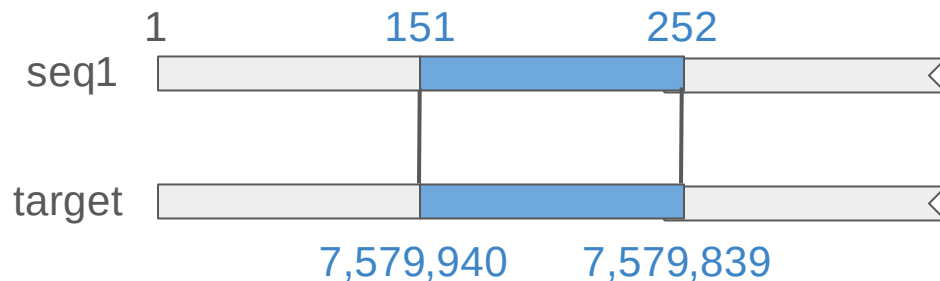
<<<<<< ||| | | | | | | | | | | | | | | | | | | | | <<<<<<  

7579940   cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc   7579891
```

```
0000201 ctagcgtcgagccccctctgagttaggaacattttcagacctatggaaa 0000250  

<<<<<< |||||<br>
7579890 ctagcgtcgagccccctctgagttaggaacattttcagacctatggaaa 7579841
```

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



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

```
0000035 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 0000084  

<<<<<< ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| <<<<<<  

7590810 ttctcaaaagtctagagccaccgtccaggagcaggtagctgctgggctc 7590761
```

```
0000085 cggggacactttgcgttcgggctgggagcgtgcttccacgacggtgaca 0000134  
<<<<<< ||||| ||||| ||||| ||||| ||||| ||||| <<<<<<  
7590760 cqqqgacactttgcgttcgggctgggagcgtgcttccacgacggtgaca 7590711
```

```
0000135 cgcttcctggattgg 0000150
<<<<<<< |||||  <<<<<<<
7590710 cgcttcctggattgg 7590695
```

```
0000151   cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc   0000200  

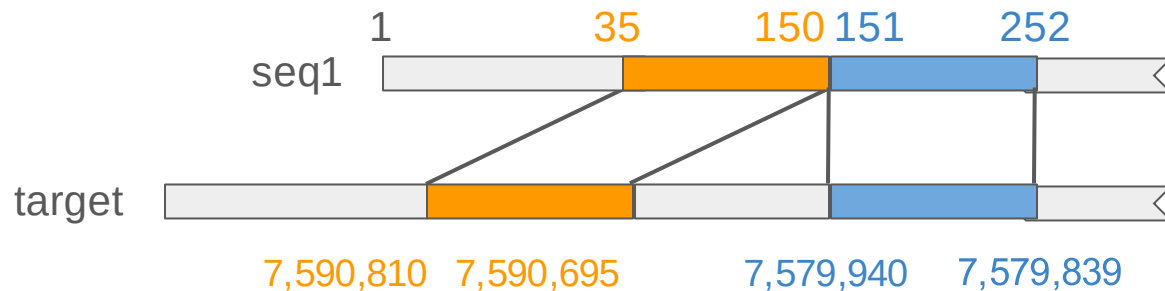
<<<<<< ||||| | | | | | | | | | | | | | | | | | | | <<<<<<  

7579940   cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc   7579891
```

```
0000201 ctagcgtcgagccccctctgagtcatgaacattttcagacctatggaaa 0000250  

<<<<<< |||||<br>
7579890 ctagcgtcgagccccctctgagtcatgaacattttcagacctatggaaa 7579841
```

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



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

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

```
0000135 cgcttcctggattgg 0000150
<<<<<<< |||||  <<<<<<<
7590710 cgcttcctggattgg 7590695
```

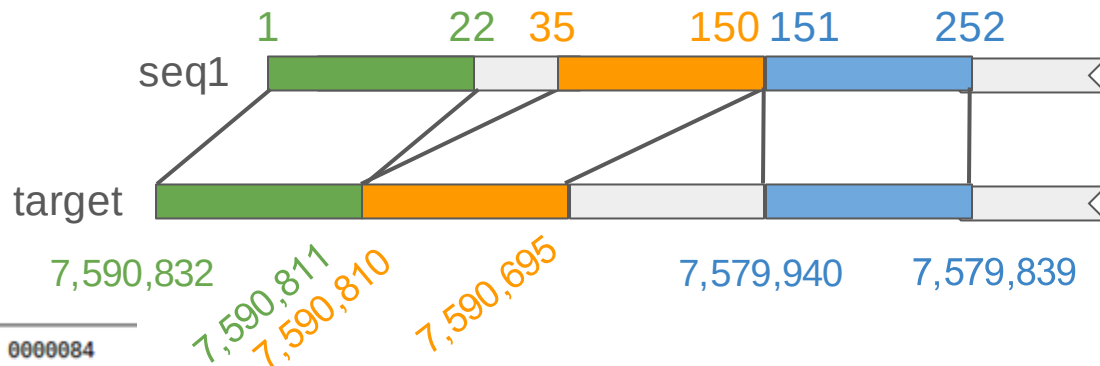
```
0000151   cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc   0000200  

<<<<<< ||||| |<br>
7579940   cagccagactgccttcgggtcactgccatggaggagccgcagtcagatc   7579891
```

```
0000201 ctagcgtcgagccccctctgagtcatgaataacattttcagacctatggaaa 0000250  

<<<<<< |||||<br>
7579890 ctagcgtcgagccccctctgagtcatgaataacattttcagacctatggaaa 7579841
```

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



5 - Click on “Seq1” in the left menu:



Genomic chr17 (reverse strand):



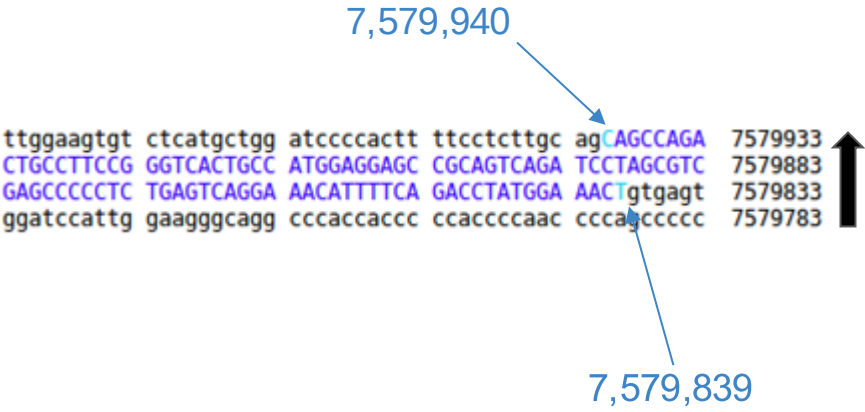
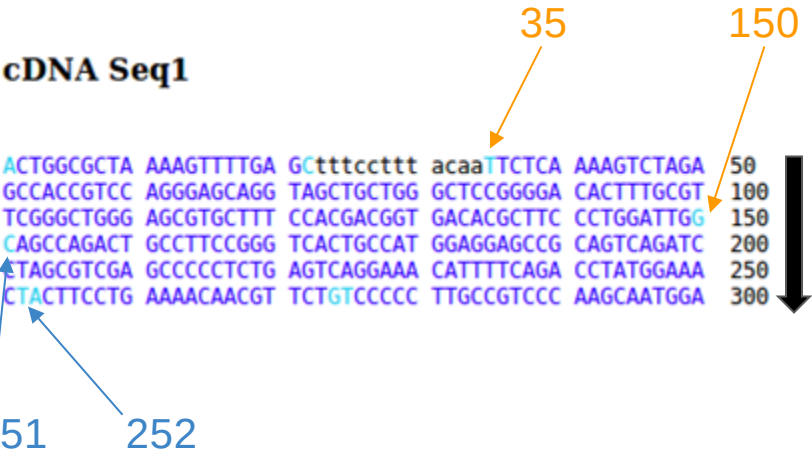
7,590,832

7,590,811

7,590,810

7,590,695

cDNA Seq1



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:

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 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
5 kb hg19
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)

WRAP53

WRAP53

HV941428
HV941434
HV941486

HV941448
HV941478
HV941430

TP53

HV941429

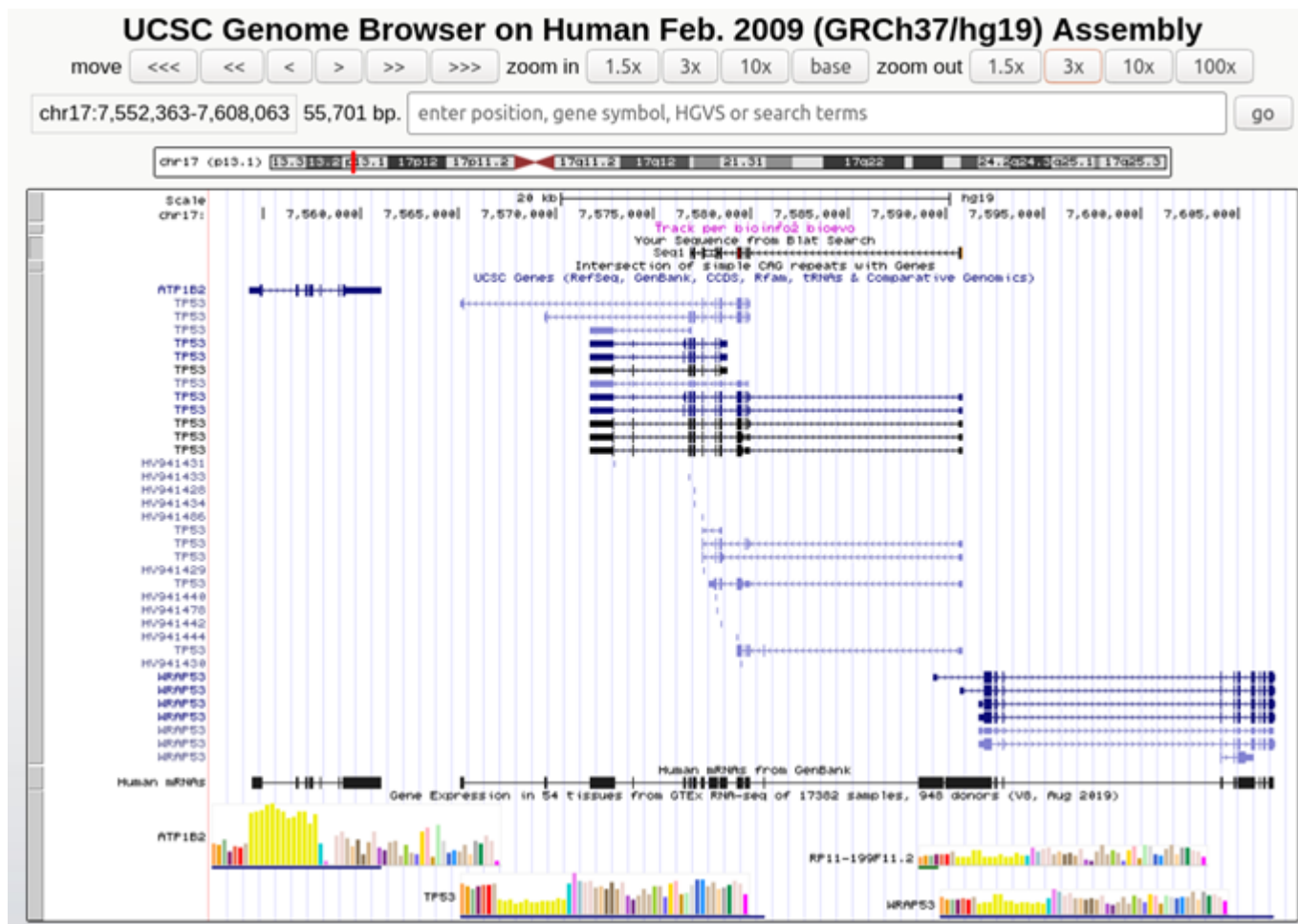
HV941442

seq genes, curated subset (NM_*, NR_*, NP_* or YP_*) - Annotation Release NCBI Homo sapiens Updated Annotation Release 105.20100906 (20100906)

WRAP53

WRAP53

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) v

BLAT's guess

query.score

hyperlink

>Seq1

actggcgctaaaaagtitttgagcttttctttacaaatcccaaaagcttagagccaccgtccaggagcaggtagctgtctgggctccggggacactttgcgttcgggctgggagcgtctttccacgacggtgacacgcttc
cctggattgagcagcagactgctcttcgggtcactgcatggagagccgagctcagattcctagcgtcagacccccctcgaagtcaggaaacattttcagacctatggaacactacttcttgaaaacaacgcttctctctcc
ttgccgtcccaagcaatggatatttgatgctgtcccgacgaatgaacaaatggctcagagaagccagggtccagatgaagctctccagattgccagaggtgctctcccccggtggccctcgacacagcagctctcact
acggcgcccttcagcagacccccctcggccctgctatcttctctctcccgaaacactaccaggggagctcagttttcgtctgggctcttgcgattctgggacagcaagctgtgagattgcacgtactccc
ctgccctcaacaagatgttttcccaactggccaagacctgcccctgtgcagctgtgggttgattccacacccccggcctttacgcgccattggccatctacaagcagtcacagcacatgacggaggttgtgaagcgtg
ccccaccactacgctctcagatagcgaatgcttgcccctctcagcattctatccgaattggaacacttgatttgcctctggagatttgnatgacagaaacacttttcgacattgattggtgtgcttctcttacc
tctgagaggtggcctcgaactcaccactcactcttaactacatctgtgaacagttcctgcattggcgggcctgaacacggagggccattctcaccatcagacactctggaagactcagttggtgaattctctgggacggaa
cagctttgagatctgattttctctctctctctggagagacccgcacacagaaagaaagaaatctcccaagaaagaaagaaacctcaccacgaactcccccgaggaacactaagcgaacactgcccaacacaccag

>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:

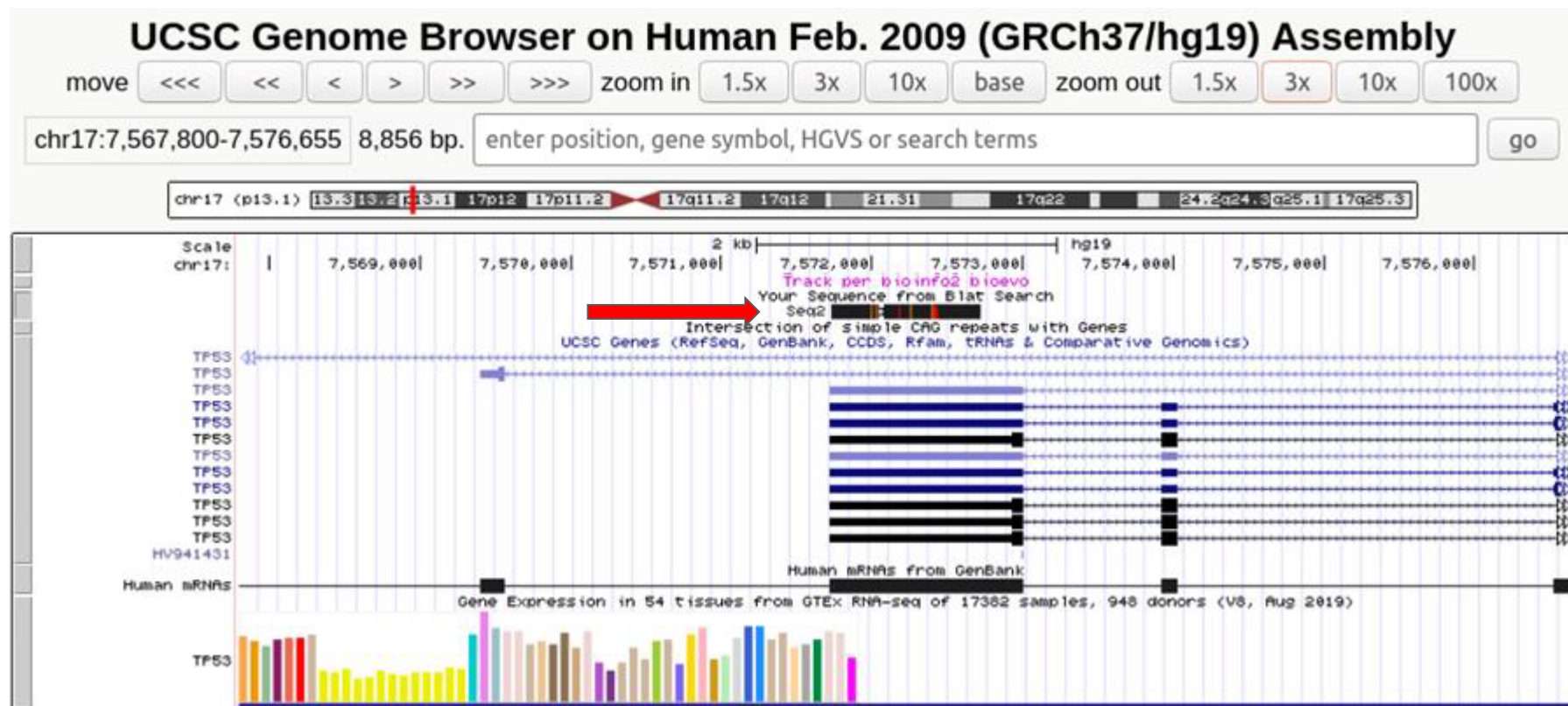
[Build a custom track with these results](#)

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

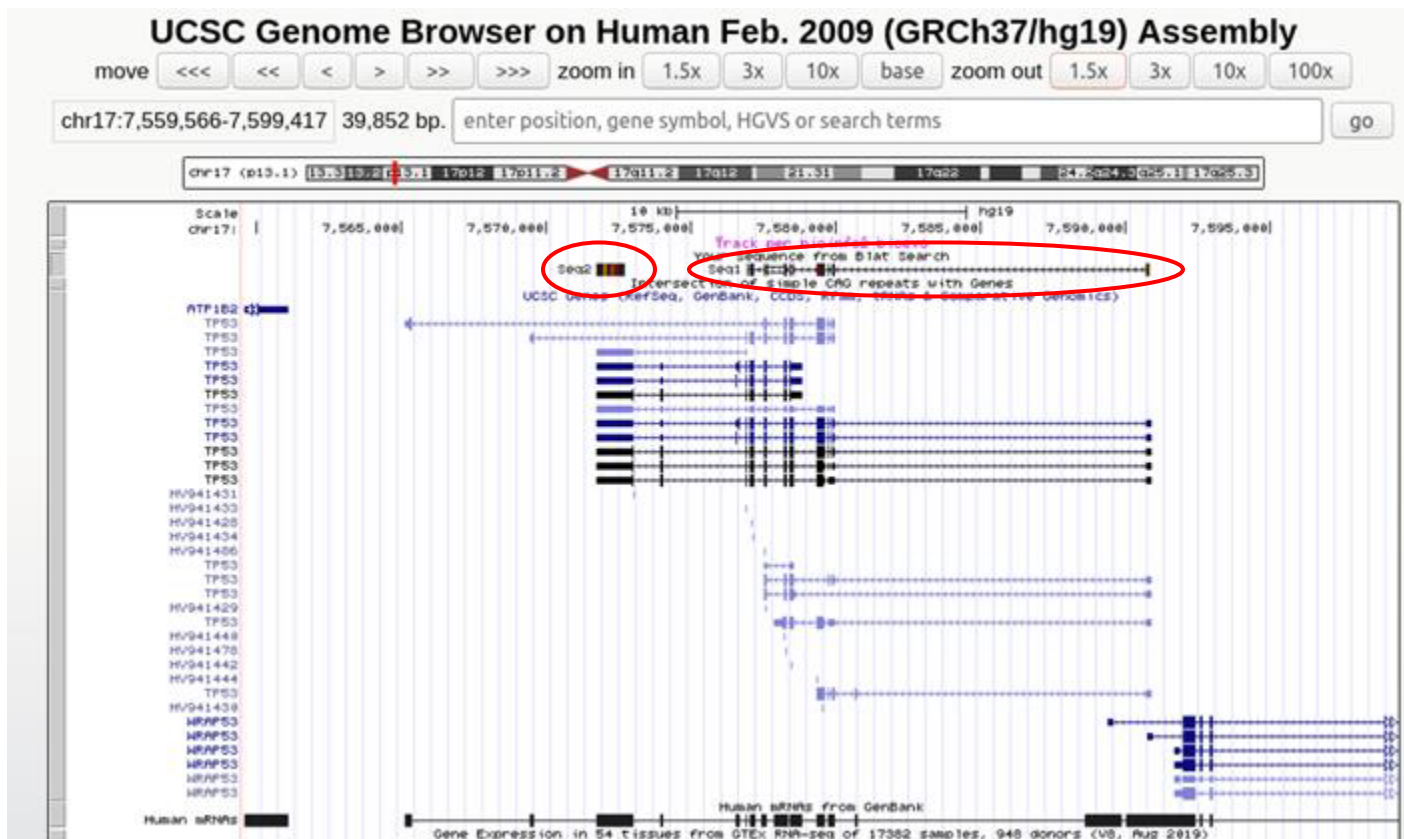
seq1

seq2

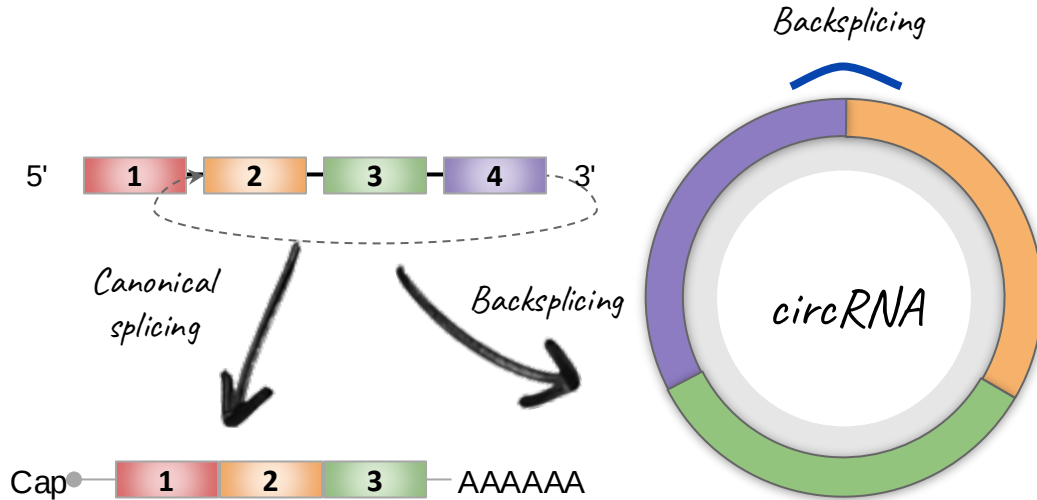
9 - Zoom 2 times to 3x to see at which portion of the gene the query sequence corresponds:



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:

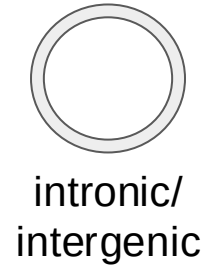
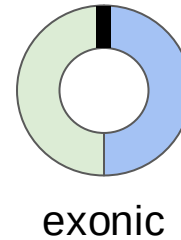
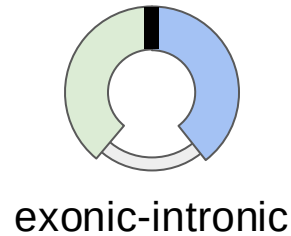


circular RNA



- Covalently closed RNA molecules
- Ubiquitous
- CircRNA structure evolutionarily conserved
- Expression varied with tissue and developmental stage
- More stable than linear RNAs
- Detected in plasma and saliva
- Conferred to extracellular bodies eg exosomes

- Where do they map on the human genome?
- Do they arise from intronic/exonic/intergenic regions?



Summary exercise

STEP 1

For each nucleotide
sequence of the
backsplicing region

- **circ1** and **circ2** in
the Practical_session_3
folder with nucleotide
sequences obtained from the
reference genome **hg38**



STEP 1

For each nucleotide
sequence of the
backsplicing region

STEP 2

Obtaining
sequence
alignment in **BLAT**

STEP 1

For each nucleotide
sequence of the
backsplicing region

STEP 2

Obtaining
sequence
alignment in **BLAT**

STEP 3

Analyze the result
in browser to
characterize
circRNAs

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

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

Nucleotide Blast (NCBI)

2

BLAT (UCSC)

3

Protein Blast and Muscle

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](#)

NEWS

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

BLASTP programs search protein databases using a protein query sequence.

Enter Query Sequence

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

>seqP
VDRPGRIPTSSRCNTIPDQNDVYSLSPGSRVYVLPKPKCFKAGTNYTVLELPQYTSOSQVESPYT
LISGLVLPKPKKSLDIPYVGGSGGQVYTNAMTTFQRYCLNGASVYKTPMDVCEBIIPTSLALLND
TGLACEIDPQVSLSSVCDPAGWGCCKRPVYVRTTNRICAPSTFLFSGSKPKCEHLDG

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”):

Descriptions	Graphic Summary	Alignments	Taxonomy						
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 alignmentsDownloadManage columnsShow100?

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; Haplorhini; 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
MGLLQLLAFSFLALCRARVRAQEPEFSYGCAEGSCYPATGDLLIGRAQKLSVTSTCGLHKPEPYCIVSHL
QEDKKCFICNSQDPYHETLNPDSHLIENVVTFAPNRLKIWWQSENGVENVTIQLDLEAEFHFTHLIMTF
KTRFPAAMLIERSDFGKTWGVYRYFAYDCEASFPGISTGPMKKVDDIICDSRYSDIEPSTEGEVIFRAL
DPAFKIEDPYPRIQNLKITNLRIKFVKLHTLGDNLDSRMEIREKYYYAVYDMVVRGNCFCYGHASEC
APVDGFGNEEVEGMVHGCMCRHNTKGLNCEL CMDFYHDL PWRPAEGRNSNACKKCNENEHSISCHFDMAV
YLATGNVSGGVCDQCQNTMGRNCEQCKPFYQHPERDIRDPNFCERCTCDPAGSQNEGICDSYTDFTSG
LIAGQCRCKLNVEGEHCDVCKEGFYDLSSDPFGCKSCACNPLGTIPGGNPCDSETHCYCKRLVTGQHC
DQCLPEHWGLSNDLDGCRPCDCDLGGALNNSCFAESGQCSCRPHMIGRQCNEVEPGYYFATLDHYLYEAE
EANLPGVSIIVERQYIQDRIPSWTGAGFVRVPEGAYLEFFIDNIPYSMEYDILIRYEPQLPDHWEKAVIT
VQRPGRIPTSSRCGNTIPDDDNQVVSLSPGSRVVLPRPVCFEKGTNYTVRLELPQYTSDDSDVESPYTL
IDSLVLMPYCKSLDIFTVGGSGDGVVTNSAWETFRQYRCLENSRSVVKTPMTDVCRNIIIFSISALLHQTG
LACECDPQGSLSVCDPNGGQCQCRPNVVGRTCNRCAPGTGFGPSGCKPCECHLQGSVNAFCNPVTGQC
HCFQGVYARQCDRCLPGHWGFPSCQPCQCNHADDQDVTGECNQCQDYMGNHCERCLAGYYGDPPIGS
GDHCRPCPCPDGPDGRQFARSCYQDPVTLQLACVCDPGYIGSRCDDCASGYFGNPSEVGGSCQPCQCHN
NIDTTDPEACDKETGRCLKLYHTEGEHCQCFRFGYYGDALQQDCRKCVCNYLGTVQEHNGSDCQCDKA
TGQCLCLPNVIGQNCDCAPNTWQLASGTGDCPCNCAAHSGPSCNEFTGQCQCMFGGGRTCSECQEL
FWGDPDVECRACDCDPRGIETPQCDQSTGQCVCVEGVEGPRCDKCTRGYSGVFPDCTPCHQCFALWDVII
AELTNRTHRFLKAKALKISGVIGPYRETVDSEVERKVSEIKDILAQSPAAEPLKNIGNLFEEAEKLIKDV
TEMAQVEVKLSDTTSQSNSTAKELDSLQTEAESLDNTVKELAEQLEFIKNSDIRGALDSITKYFQMSLE
AEERVNASTTEPNSTVEQSALMRDRVEDVMERESQFKEKQEEQARLLDELAGKLQSLDLAAAEMTCGT
PPGASCSETECGGPNCRTDEGERKCGGPGCGGLVTVAHNAWQKAMDLDQDVL SALAEVEQLSKMVSEAKL
RADEAKQSAEDILLKTNATKEKMDKSNEELRNLIKQIRNFLTQDSADLDSIEAVANEVLKMEMPSTPQQL
QNLTEDIRERVESLSQVEVILQHSAAIARAEMLL EAKRASKSATDVKVTADMVKEALEEAEKAQVAE
KAIKQADEDIQGTQNL LTSI ESETAASEETLFNASQRISELERNVEELKRKAAQNSGEAEYIEKVYTVK
QSAEDVKKTLDGELDEKYKKVENLIAKKT EESADARRKAEMLQNEAKTL LAQANSKLQLLKD LERKYEDN
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

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

☐ 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

Select columns

Show

100

☐ select all

sequences selected

GenPept

Graphics

Distance tree of results

Multiple alignment

MSA Viewer

	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_001355667.4
<input type="checkbox"/>	laminin subunit beta-1 isoform X1 (Homo sapiens)	Homo sapiens	3678	3678	99%	0.0	99.72%	1810	XP_016867090.1
<input checked="" type="checkbox"/>	laminin subunit beta-1 (Tursiops truncatus)	Tursiops truncatus	3492	3492	100%	0.0	93.17%	1786	XP_033728774.1
<input type="checkbox"/>	laminin subunit beta-1 isoform X2 (Rattus norvegicus)	Rattus norvegicus	3438	3438	100%	0.0	92.61%	1786	XP_006340058.1
<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	2668	65%	0.0	99.56%	1212	XP_016867091.1
<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_006343771.1
<input type="checkbox"/>	laminin subunit beta-2 precursor (Rattus norvegicus)	Rattus norvegicus	1905	1905	99%	0.0	51.18%	1801	NP_037109.1
<input type="checkbox"/>	laminin subunit beta-2 isoform X2 (Rattus norvegicus)	Rattus norvegicus	1901	1901	99%	0.0	51.18%	1800	XP_039366079.1
<input type="checkbox"/>	laminin subunit beta-2 isoform X1 (Tursiops truncatus)	Tursiops truncatus	1899	1899	96%	0.0	51.10%	1809	XP_033720799.1
<input type="checkbox"/>	laminin subunit beta-2 (Pan troglodytes)	Pan troglodytes	1882	1882	96%	0.0	50.65%	1798	XP_016796574.2
<input type="checkbox"/>	laminin subunit beta-2 precursor (Homo sapiens)	Homo sapiens	1878	1878	96%	0.0	50.65%	1798	NP_002283.3
<input type="checkbox"/>	laminin subunit beta-2 isoform X2 (Rattus norvegicus)	Rattus norvegicus	1763	1763	93%	0.0	50.51%	1667	XP_039366810.1
<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_011514290.1
<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_016832513.2
<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_001064311.1
<input checked="" type="checkbox"/>	laminin subunit beta-3 (Pan troglodytes)	Pan troglodytes	410	1046	87%	2e-119	41.56%	1171	XP_002946131.2
<input checked="" type="checkbox"/>	laminin subunit beta-3 precursor (Homo sapiens)	Homo sapiens	408	1091	87%	1e-118	41.19%	1172	NP_000218.2
<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 biological 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's a navigation bar with 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 is claimed to achieve better accuracy and speed than ClustalW2 or T-Coffee. It also states that the tool can align up to 500 sequences or a maximum file size of 1 MB. Under the heading 'Input sequence', there's a text input field with the placeholder 'Paste your sequence here - or use the example sequence'. Below the input field, there's a file selection button labeled 'Sfoglia...' and a text input field containing 'seqdump(1).txt'. To the right of the input field, there are two buttons: 'Use the example' and 'Clear sequence', followed by a link 'More example inputs'. Under the heading 'Parameters', there's a section for 'OUTPUT FORMAT' with a dropdown menu currently set to 'ClustalW'. A link 'More options' is visible below the dropdown.

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):

Rn_b1X1	MERPLSSLPTFAFSPLYLSRKEDTKKRAACLPRPPSRAASPRLLQGLDMGVLQVFAFGVL
Tt_b1	-----MGRLQMFAFSFL
Hs_b1p	-----MGLLQLLAFSFL
Pt_b1	-----MWLLQLLAFSFL
Rn_b3p	-----MMTALFLLWLALP
Tt_b3	-----MRPLLLLYFVL
Pt_b3	-----MRPFLLCFALP
Hs_b3p	-----MRPFLLCFALP

“*” = same amino acid
for all sequences

Rn_b1X1	ALWGTRVCAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
Tt_b1	ALCGAQVSAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLYKPEPYCIVSHLQE
Hs_b1p	ALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
Pt_b1	ALCRARVRAQEPEFSYGCAEGSCYPATGDLIGRAQKLSVTSTCGLHKPEPYCIVSHLQE
Rn_b3p	GL----LCAQQP----CSRGAQYPPVGDLLIGRTQLLRASSTCGLTKPETYC-TQYGQW
Tt_b3	SV----LCAQQA----CSRGAQYPPVGDLLIGRTFLRASSTCGLAKPETYC-TQYGEW
Pt_b3	GL----LHAQQA----CSRGAQYPPVGDLLVGRTRFLRASSTCGLTKPETYC-TQYGEW
Hs_b3p	GL----LHAQQA----CSRGAQYPPVGDLLVGRTRFLRASSTCGLTKPETYC-TQYGEW

“:” = conservative
substitution

Rn_b1X1	DKKCFICDSRDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLEAEFH
Tt_b1	DKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLEAEFH
Hs_b1p	DKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLEAEFH
Pt_b1	DKKCFICNSQDPYHETLNPDSHLIENVVTTFAPNRLKIWWQSENGVENVTIQLDLEAEFH
Rn_b3p	QMKCKCKDSRLPR---NYNSHRVENVVSSSGPMR---WWQSQNDVSPVSLQLDLDRKFQ
Tt_b3	QMKCKCKDSRLPH---NYNSHRVENVVSSSGPMR---WWQSQNDVSPVSLQLDLDRKFQ
Pt_b3	QMKCKCKDSRQPH---NYYSHRVENVASSSGPMR---WWQSQNDVNPVSLQLDLDRRFQ
Hs_b3p	QMKCKCKDSRQPH---NYYSHRVENVASSSGPMR---WWQSQNDVNPVSLQLDLDRRFQ

“.” = semi-conservative
substitution

“ ” = non-conservative
substitution

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

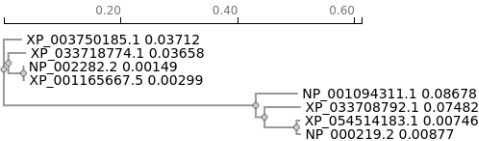
Tool Output	Alignments	Phylogenetic Tree	Results Viewers	Result Files	Submission Details
-------------	------------	-------------------	-----------------	--------------	--------------------

Phylogenetic Tree

```
(
  XP_003750185.1:0.03712,
  (
    XP_033718774.1:0.03658,
    (
      NP_002282.2:0.00149,
      XP_001165667.5:0.00299)
    :0.03117)
  :0.00935,
  (
    NP_001094311.1:0.00678,
    (
      XP_033708792.1:0.07482,
      (
        XP_054514183.1:0.00746,
        NP_000219.2:0.00877)
      :0.06602)
    :0.01811)
  :0.51949);
```

Phylogram

Radial



Selected 0 branches with current label