

Phyloinformatics

Assignment-2

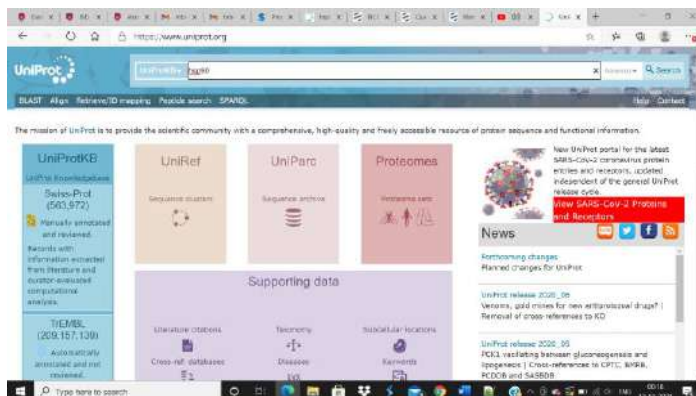
Q1-Using BLAST to identify similar sequences

Answer: **FOR EXAMPLE:**

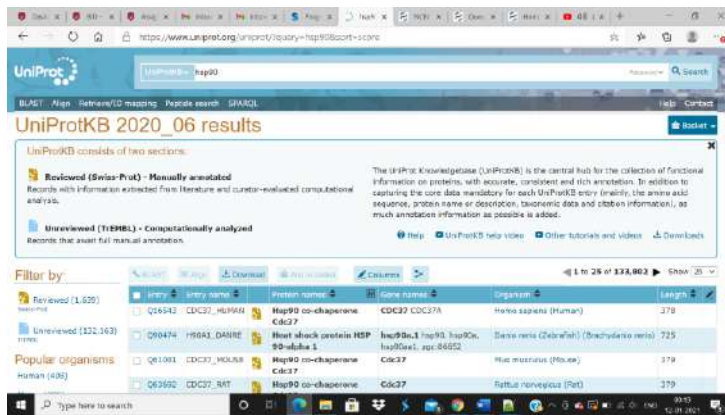
UniProt:

The UniProt Knowledgebase (UniProtKB) is the central hub for the collection of functional information on proteins, with accurate, consistent and rich annotation. In addition to capturing the core data mandatory for each UniProtKB entry (mainly, the amino acid sequence, protein name or description, taxonomic data and citation information), as much annotation information as possible is added. This includes widely accepted biological ontologies, classifications and cross-references, and clear indications of the quality of annotation in the form of evidence attribution of experimental and computational data. The UniProt Knowledgebase consists of two sections: a section containing manually-annotated records with information extracted from literature and curator-evaluated computational analysis, and a section with computationally analyzed records that await full manual annotation. For the sake of continuity and name recognition, the two sections are referred to as "UniProtKB/Swiss-Prot" (reviewed, manually annotated) and "UniProtKB/TrEMBL" (unreviewed, automatically annotated), respectively.

Step 1: Go to the homepage of UniProtKB and in the search bar type 'hsp90' and click on search button.



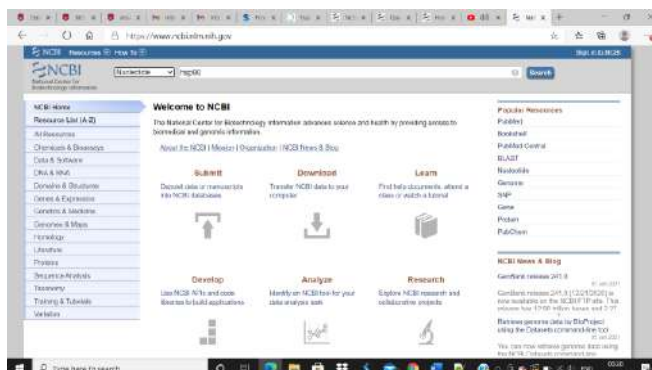
Step 2: You will see 1 to 25 of 133,802 number of hits.



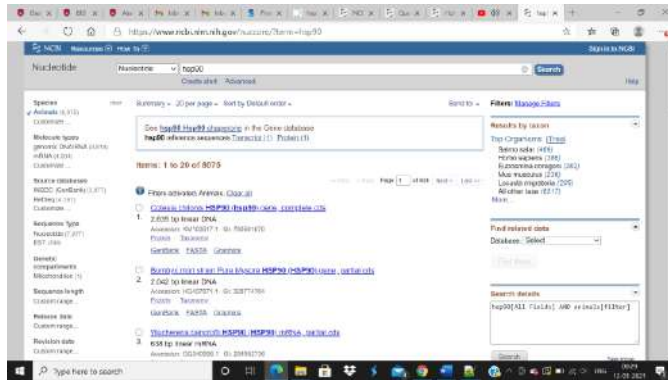
NCBI:

The National Center for Biotechnology Information (NCBI) is part of the United States National Library of Medicine (NLM), a branch of the National Institutes of Health (NIH). The NCBI is located in Bethesda, Maryland and was founded in 1988 through legislation sponsored by Senator Claude Pepper. The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. Major databases include GenBank for DNA sequences and PubMed, a bibliographic database for biomedical literature. Other databases include the NCBI Epigenomics database. All these databases are available online through the Entrez search engine. NCBI was directed by David Lipman, one of the original authors of the BLAST sequence alignment program and a widely respected figure in bioinformatics. He also led an intramural research program, including groups led by Stephen Altschul (another BLAST co-author), David Landsman, Eugene Koonin, John Wilbur, Teresa Przytycka, and Zhiyong Lu. David Lipman stood down from his post in May 2017.

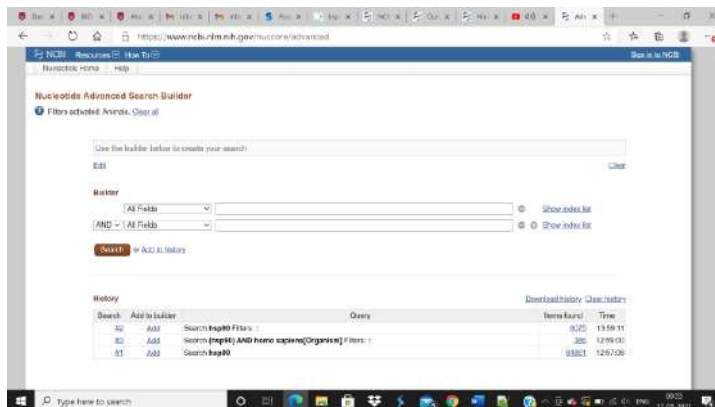
Step 3: Go to homepage of NCBI in the search bar type 'hsp90' and search. Select the database as 'nucleotide' in the drop box.



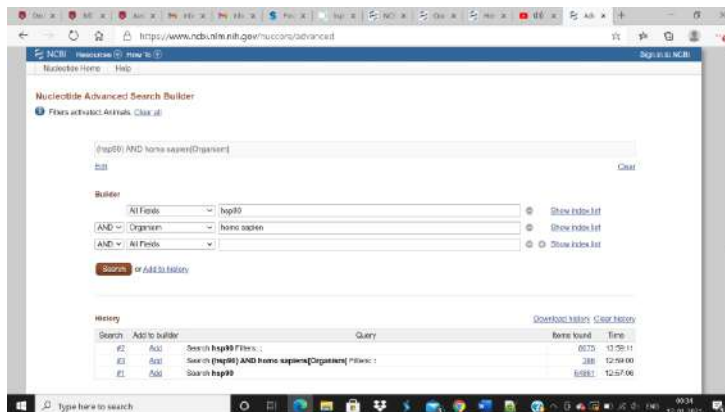
Step 4: In the extreme left corner select the species according to your choice. I have selected 'animal' and after clicking you will see 1 to 20 of 8075 number of hits.



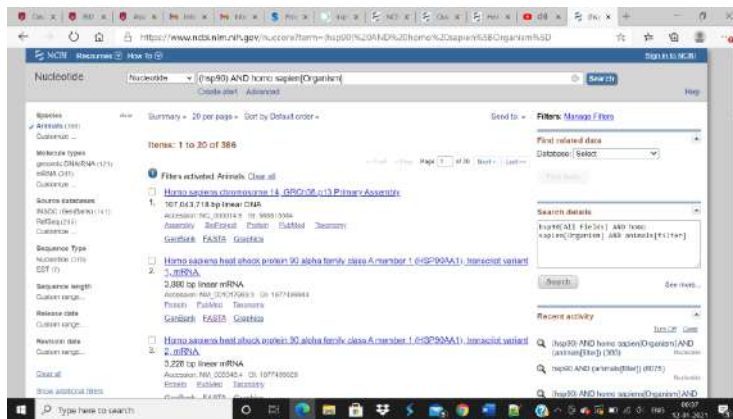
Step 5: Click on 'Advanced' below the search bar on the ncbi page.



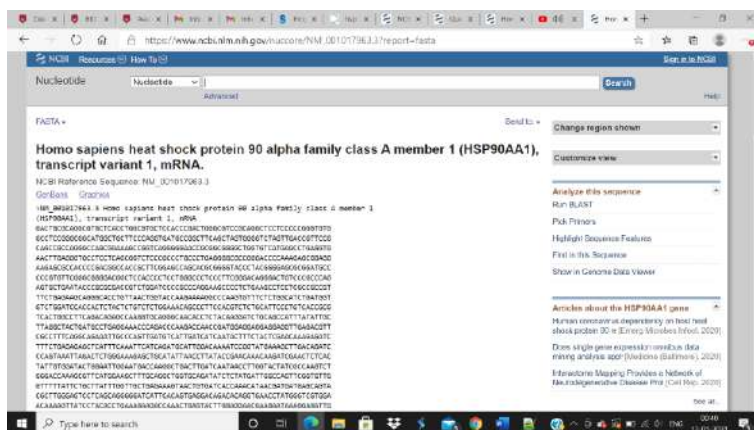
Step 6: Under the builder heading type 'hsp90' and in the second drop box select 'organism' as the category and 'homo sapien' as the name and click on search.



Step 7: You will get 1 to 20 of 386 as maximum number of hits.



Step 8: Select any one type , for example I am selecting the second one : Homo sapiens heat shock protein 90 alpha family class A member 1 (HSP90AA1) and click on its FASTA formate.



>Sample_1

```
CGAGGGACCTTTACGGCGTAATCCTGGAAACCATGACAAATCCAGAACCCCAAGGCTCCCCTCTT
CAGCTGATGTAGAATTTTGCCTGAGTTTGACCCATGGAAGGATTTGCTAGTCCACTTACTGGGAT
AGCGGATGCCTCTCAAAGAGCATGCACAATGCCTTGACATCTATATGAATGGAACAATGTCCCA
GGCAGGGATCTGCCAACGATCCTATCTTCTTCTTACCATGCATTGTTGACAGTATTTTTGAG
CAGTGGCTCCGAAGGCACCGTCTCTTCAAGAAGTTTATCCAGAAGCCAATGCACCCATTGGACA
TAACCGGGAATCCTACATGGTTCTTATACCACTGTACAGAAATGGTGATTTCTTTATTTTCATCCA
AAGATCTGGGCTATGACTATAGCTATCTACAAGATTGACACCCAGACTCTTTTCAAGACTACATT
AAGTCTATTTGGAACAAGCGAGTCGGATCTGGTCATGGCTCCTTGGGGCGGCGATGGTAGGGGC
CGTCTCATCTGCCCTGCTGGCGGGCTTGTGAGCTTGCTGTGTCGTACAAGAGAAAGCAGCTTCC
TGAAGAAAAGCAGCCACTCCTCATGGAGAAAAGAGGATTACCACAGCTTGTATCAGAGCCATTAT
AAAAGGCTTAGGCAATAGAGTAGGGCCAAAAGCCTGACCTCACTCTAACTCAAAGTAATGTCCA
GGTCCCAGAGAATATCTGCTGGTATTTTTCTGTAAAGACCATTGCAAAATTGTAACCTAATAC
AAAGTGATAGCCTTCTTCAACTCAGGTAGAACACACCTGTCTTGTCTGCTGTTTTCACTCAGC
CTTTTAAACATTTTCCCTAAGCCCATATGTCTAAGGAAAGGATGCTATTTGGTAATGAGGAAT
GTTATTTGTATGTGAATTAAGTGCTCTTATTTTAAAAAATTGAAATAATTTGATTTTGCCTT
CTGATTATTTAAGATCTATATATGTTTTATTGGCCCTTCTTATTTTAAATAAACAGTGAGAA
ATCT
```

- **Select the BLAST program that is required protein / DNA**

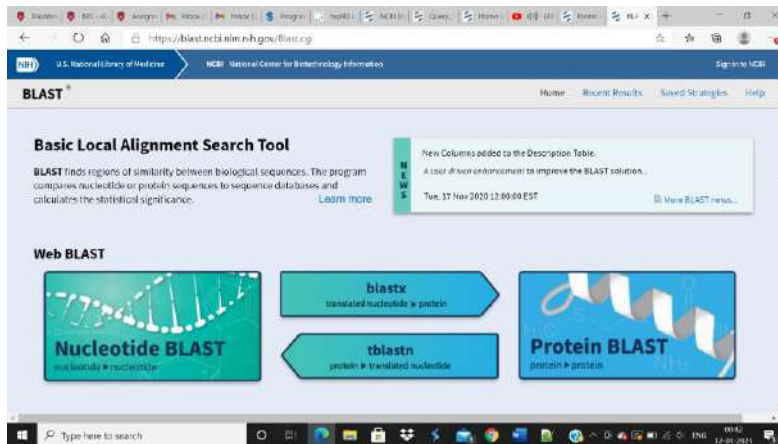
- **Copy and paste your DNA** into the search box.
- Select the required **Database**
- Select the Megablast for DNA
- For the first time, **leave the Options section settings** on the **Default** values and keep **homo sapiens** as organism
- **Click** on the **Blast button** at the bottom or top of the screen.
- **A new window will appear** gives an estimate of how long the search will take and which lists conserved domains in your query sequence.
- **Check your result** - check **graphic** - red, green, blue, black lines and their meaning
- **Identify the top 10 similar sequence** and analyse **Accession No. E value, coverage, identity** with other sequences (use a table)
- What is the source of the given query sequence?
- Show the Alignment, Graphic summary and taxonomy of top ten similar sequences
- List out the set of sequences showed good similarity with query sequence

BLAST:

In bioinformatics, BLAST is an algorithm and program for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA and/or RNA sequences. A BLAST search enables a researcher to compare a subject protein or nucleotide sequence with a library or database of sequences, and identify library sequences that resemble the query sequence above a certain threshold. **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. PHI-BLAST stands for Pattern-hit initiated BLAST. The program uses an input sequence and a defined pattern to query a protein database. **The BLAST Homepage** : The BLAST homepage consists of several sections, each provides a specific set of functions:

1. The common header , present in most BLAST-related pages, provides easy access to other content or functions not directly accessible from the homepage.
2. The “Recent Results” lists a set of links to recently completed search results.
3. Pages with web forms for submitting searches are listed as links in the body of the BLAST homepage. These links are organized into three categories, “Web BLAST” , used to be called “Basic BLAST”, “BLAST Genomes” , and “Specialized searches” .
4. The search box in the “BLAST Genomes” section takes the name of an organism as input and suggests a list of candidates. Selecting from the suggested list and clicking the “Search” button locates the best genomic sequence dataset for BLAST alignment purposes.

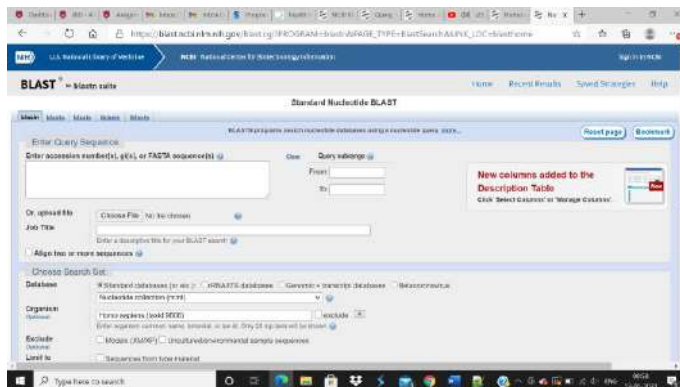
Step 1: Copy the give 'Sample1' nucleotide sequence. Open the blast homepage and click on Nucleotide blast..



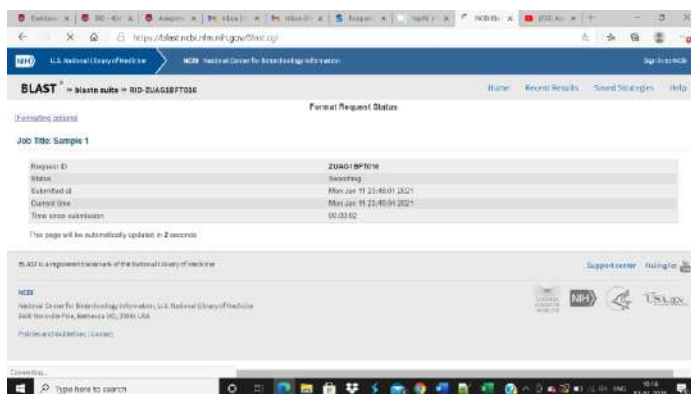
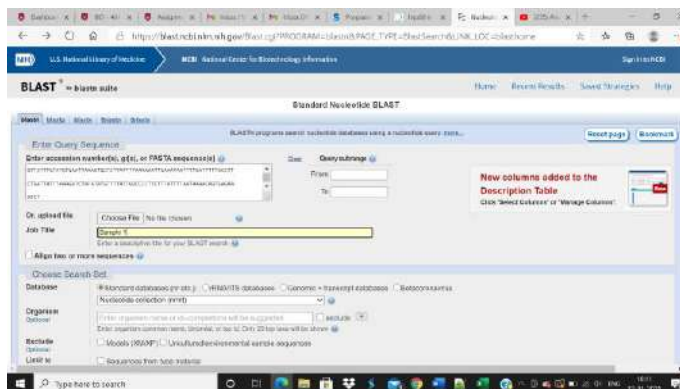
Nucleotide blast:

- Enter query sequence(s) in the text area. It automatically determines the format of the input. To allow this feature, certain conventions are required with regard to the input of identifiers.
- Enter coordinates for a subrange of the query sequence. The BLAST search will apply only to the residues in the range. Sequence coordinates are from 1 to the sequence length. The range includes the residue at the To coordinate.
- Use the browse button to upload a file from your local disk. The file may contain a single sequence or a list of sequences. The data may be either a list of database accession numbers, NCBI gi numbers, or sequences in FASTA format.
- Title: Nucleotide collection (nt)
- Description: The nucleotide collection 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 sequences longer than 100Mb. 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: 2021/01/10
 - Number of sequences: 65815883
- Start typing in the text box, then select your taxid. Use the 'plus' button to add another organism or group, and the 'exclude' checkbox to narrow the subset. The search will be restricted to the sequences in the database that correspond to your subset.
- You can use Entrez query syntax to search a subset of the selected BLAST database. This can be helpful to limit searches to molecule types, sequence lengths or to exclude organisms.
- Megablast is intended for comparing a query to closely related sequences and works best if the target percent identity is 95% or more but is very fast. Discontiguous megablast uses an initial seed that ignores some bases (allowing mismatches) and is

intended for cross-species comparisons. BlastN is slow, but allows a word-size down to seven bases.



Step 2: Paste the nucleotide sequence and click on blast. You will see 100 sequences selected.



U.S. National Library of Medicine
National Center for Biotechnology Information

BLAST⁺ » blastn suite » results for RID:ZUAG18FT016

Job Title: Sample 1
RID: ZUAG18FT016
Program: BLASTN
Database: nt
Query ID: JctQuery_24353
Description: None
Molecule type: dna
Query Length: 1044

Filter Results

Organism: only top 20 will appear
Type common name, accession, build or group name:
+ Add organism

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

Filter Reset

Sequences producing significant alignments

Download Select columns Show Feedback

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments

Download Select columns Show 100 Feedback

☒ select all 100 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens tyrosinase (TYR). mRNA	Homo sapiens	1712	1696	100%	0.0	99.28%	2062	NM_000712.5
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus tyrosinase (TYR). mRNA	Pan paniscus	1690	1674	100%	0.0	98.84%	2485	XM_00852389.2
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla tyrosinase (LOC101188488). mRNA	Gorilla gorilla gorilla	1685	1663	100%	0.0	98.74%	2485	XM_031336263.1
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes tyrosinase (TYR). mRNA	Pan troglodytes	1676	1659	100%	0.0	98.53%	4893	XM_001336041.2
<input checked="" type="checkbox"/> PREDICTED: Hylobates moloch tyrosinase (TYR). mRNA	Hylobates moloch	1578	1749	99%	0.0	96.68%	2486	XM_032168104.1
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii tyrosinase (TYR). mRNA	Pongo abelii	1535	1700	95%	0.0	97.35%	2486	XM_00252337.3
<input checked="" type="checkbox"/> Human tyrosinase (TYR). mRNA, complete cds	Homo sapiens	1528	1712	90%	0.0	99.18%	2384	M27168.1
<input checked="" type="checkbox"/> Human mRNA for tyrosinase (EC 1.14.18.1)	Homo sapiens	1528	1712	90%	0.0	99.18%	1886	Y00018.1
<input checked="" type="checkbox"/> PREDICTED: Ptilinopus leucostictus tyrosinase (TYR). mRNA	Ptilinopus leucostictus	1489	1651	100%	0.0	94.99%	2486	XM_023209125.1
<input checked="" type="checkbox"/> Human tyrosinase and melanin tyrosinase, complete cds	Homo sapiens	1474	1638	90%	0.0	97.91%	1929	M25215.1
<input checked="" type="checkbox"/> PREDICTED: Trachypithecus francoisi tyrosinase (TYR). mRNA	Trachypithecus francoisi	1472	1645	100%	0.0	94.68%	2484	XM_033003068.1
<input checked="" type="checkbox"/> PREDICTED: Cebus apella tyrosinase (TYR). mRNA	Cebus apella	1471	1642	99%	0.0	94.50%	2485	XM_011928172.1
<input checked="" type="checkbox"/> PREDICTED: Rhinopithecus roosei tyrosinase (TYR). transcript variant X1. mRNA	Rhinopithecus roosei	1467	1640	100%	0.0	94.57%	2488	
<input checked="" type="checkbox"/> PREDICTED: Ceropterus aha tyrosinase (TYR). mRNA	Ceropterus aha	1463	1630	100%	0.0	94.57%	2485	

Step 3: By default the 100 sequences are selected already. We have to select on 10 sequences as shown below:

Descriptions Graphic Summary Alignments Taxonomy

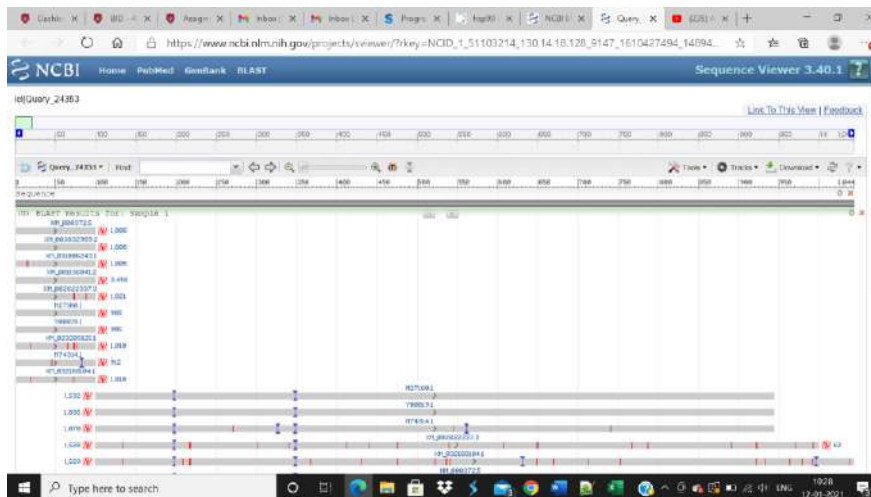
Sequences producing significant alignments

Download Select columns Show 100 Feedback

☐ select all 10 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens tyrosinase (TYR). mRNA	Homo sapiens	1712	1696	100%	0.0	99.28%	2062	NM_000712.5
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus tyrosinase (TYR). mRNA	Pan paniscus	1690	1674	100%	0.0	98.84%	2485	XM_00852389.2
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla tyrosinase (LOC101188488). mRNA	Gorilla gorilla gorilla	1685	1663	100%	0.0	98.74%	2485	XM_031336263.1
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes tyrosinase (TYR). mRNA	Pan troglodytes	1676	1659	100%	0.0	98.53%	4893	XM_001336041.2
<input checked="" type="checkbox"/> PREDICTED: Hylobates moloch tyrosinase (TYR). mRNA	Hylobates moloch	1578	1749	99%	0.0	96.68%	2486	XM_032168104.1
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii tyrosinase (TYR). mRNA	Pongo abelii	1535	1700	95%	0.0	97.35%	2486	XM_00252337.3
<input checked="" type="checkbox"/> Human tyrosinase (TYR). mRNA, complete cds	Homo sapiens	1528	1712	90%	0.0	99.18%	2384	M27168.1
<input checked="" type="checkbox"/> Human mRNA for tyrosinase (EC 1.14.18.1)	Homo sapiens	1528	1712	90%	0.0	99.18%	1886	Y00018.1
<input checked="" type="checkbox"/> PREDICTED: Ptilinopus leucostictus tyrosinase (TYR). mRNA	Ptilinopus leucostictus	1489	1651	100%	0.0	94.99%	2486	XM_023209125.1
<input checked="" type="checkbox"/> Human tyrosinase and melanin tyrosinase, complete cds	Homo sapiens	1474	1638	90%	0.0	97.91%	1929	M25215.1

Step 4: To get the result of graphics, we have to click on the 'Graphics' besides GenBank you will get - red, green, blue, black lines and their meaning as shown below:



(U) BLAST Results for: Sample 1 (Non-NCBI data)

Alignment: Query_24353 x NM_000372.5
Anchor: Query_24353 (1..99)
Query: NM_000372.5
 (958..1,056)
 Homo sapiens tyrosinase (TYR), mRNA
Relative orientation: forward
Span on Query_24353: 99
Segments: 1
Coverage: 4.8%
Identity: 100.0%
Mismatches: 0
Gaps: 0

Download FASTA: [Query_24353](#)

Links & Tools

BLAST nr: [NM_000372.5](#)
BLAST to Genome: [NM_000372.5](#)
FASTA record: [NM_000372.5](#)
GenBank record: [NM_000372.5](#)

Alignment: Query_24353 x M27160.1
Anchor: Query_24353 (100..944)
Query: M27160.1 (1,533..2,384)
 Human tyrosinase (TYR) mRNA, complete cds
Relative orientation: forward
Span on Query_24353: 852
Segments: 1
Coverage: 35.4%
Identity: 100.0%
Mismatches: 0
Gaps: 4

Unaligned regions: 5
Query_24353 position: 196
M27160.1 position: 1,630

Download FASTA: Query_24353

Links & Tools

BLAST nr: M27160.1

BLAST to Genome: M27160.1

FASTA record: M27160.1

GenBank record: M27160.1

Step 5: What is the source of the given query sequence?

Select seq NM_000372.5 Homo sapiens tyrosinase (TYR), mRNA

Sequences producing significant alignments

Download Select columns Show 100

select all 10 sequences selected

Description	Source Name	Max Score	Raw Score	Query Cover	Ident	Pos	Exp	Acc	Accession
<input checked="" type="checkbox"/> Homo sapiens: nucleoside (TYR). mRNA	Homo sapiens	1712	1000	100%	0.0	89,259	2.0E-362	100	NM_000732.3
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus: nucleoside (TYR). mRNA	Pan paniscus	1000	1000	100%	0.0	56,649	3.0E-268	100	XM_001330869.2
<input checked="" type="checkbox"/> PREDICTED: Gallus gallus: nucleoside (TYR). mRNA	Gallus gallus	1000	1000	100%	0.0	18,181	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes: nucleoside (TYR). mRNA	Pan troglodytes	1078	1000	100%	0.0	56,520	4.8E-361	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Erythrina speciosa: nucleoside (TYR). mRNA	Erythrina speciosa	1038	1000	100%	0.0	49,900	2.0E-358	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1028	1000	100%	0.0	57,285	3.4E-361	100	XM_013306241.1
<input checked="" type="checkbox"/> Homo sapiens: nucleoside (TYR). mRNA	Homo sapiens	1028	1000	100%	0.0	88,195	2.0E-361	100	M22706.1
<input checked="" type="checkbox"/> Homo sapiens: nucleoside (TYR). mRNA	Homo sapiens	1028	1000	100%	0.0	88,195	2.0E-361	100	M22706.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1000	1000	100%	0.0	54,000	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> Homo sapiens: nucleoside (TYR). mRNA	Homo sapiens	1474	1000	100%	0.0	57,215	3.4E-361	100	M22706.1
<input checked="" type="checkbox"/> PREDICTED: Homo sapiens: nucleoside (TYR). mRNA	Homo sapiens	1474	1000	100%	0.0	54,000	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Gallus gallus: nucleoside (TYR). mRNA	Gallus gallus	1471	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1
<input checked="" type="checkbox"/> PREDICTED: Pteropus alectes: nucleoside (TYR). mRNA	Pteropus alectes	1467	1000	100%	0.0	54,526	2.4E-51	100	XM_013306241.1

Database X | RefSeq X | Aligner X | View X | Index X | Program X | ArgLine X | NCBI Help X | (22) X +

[Home](#) [Help](#) [Feedback](#)

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise ▾ CDS feature ⓘ Restores defaults Download ▾

10 sequences selected ⓘ

Download ▾ GenBank Graphics Sort by: E value ▾ Next < Previous > Descriptions

Homo sapiens tyrosinase (TYR), mRNA

Sequence ID: [NM_009722.5](#) Length: 2052 Number of Matches: 2

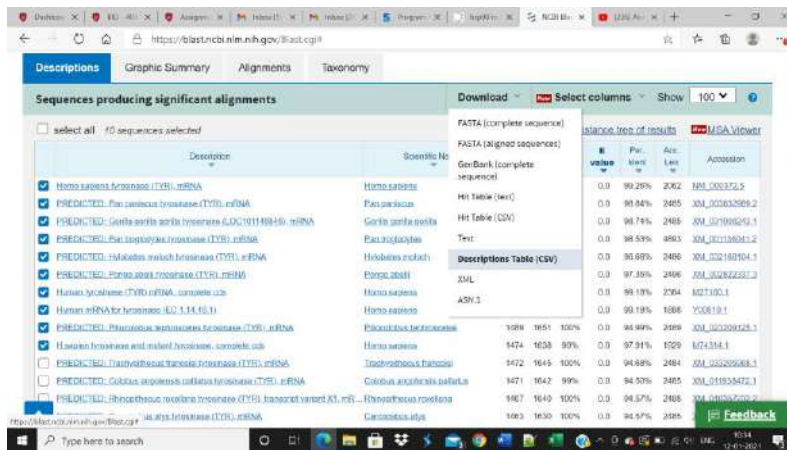
Range 1: 1110 -> 2061 [GenBank](#) [Graphics](#) ▾ [Next Match](#) A Threshold Match

Score	Expect	Identical	Gaps	Strand
1712 bits(927)	0.0	945 / 952 (99%)	7 / 652 (1%)	Plus/Minus
Query 180	TGAGAAAGTTTTCATGGCTCTTC	TGC TGGGACATCGCATGTCCCTCAGAA-ACAATCAACA	154	
Subject 1510	TGAGAAAGTTTTCATGGCTCTTC	TGC TGGGACATCGCATGTCCCTCAGAA-ACAATCAACA	1550	
Query 159	ATGCTCTTGACATCTATATGTA	GGACATATGTCCTCAGG-CAGGGGTCTGCAACGATC	216	
Subject 1178	ATGCTCTTGACATCTATATGTA	GGACATATGTCCTCAGG-CAGGGGTCTGCAACGATC	1178	
Query 227	CTATTCTTCTCTTCACATCATT	TTCACACTATTTTGACCATCTGCTCCGAAGGC	276	
Subject 1238	CTATTCTTCTCTTCACATCATT	TTCACACTATTTTGACCATCTGCTCCGAAGGC	1238	
Query 277	ACCCTCTCTCTCAGAAATTTT	CAGAACGCCATGACCATCTTGACATATACCGAAT	336	
Subject 1290	ACCCTCTCTCTCAGAAATTTT	CAGAACGCCATGACCATCTTGACATATACCGAAT	1240	
Query 357	CTCACATGGTCTC----	TTATACCATATACGGAATTTGATTTTATTCTTCACGAG	393	

There have no search results.

Feedback

Step 6: To identify the top 10 similar sequence and analyse Accession No. E value, coverage, identity with other sequences (use a table) we need to click on ‘Download’ and then click on ‘Descriptions Table (CSV)’.



Step 7: To identify the top 10 similar sequence and analyse Accession No. E value, coverage, identity with other sequences (use a table)

	Description	Scientific Name	E-value	Percent Identity	Accession	Score
1	Human sapiens sapiens (TYR) mRNA	Human sapiens	0.0	99.29%	XM_000372.5	1712
2	PREDICTED: Pan paniscus tyrosinase (TYR) mRNA	Pan paniscus	0.0	98.84%	XM_003832989.2	1650
3	PREDICTED: Gorilla gorilla gorilla tyrosinase (TYR) mRNA	Gorilla gorilla gorilla	0.0	98.74%	XM_031006243.1	1585
4	PREDICTED: Pan troglodytes tyrosinase (TYR) mRNA	Pan troglodytes	0.0	98.53%	XM_001136041.2	1676
5	PREDICTED: Hylobates moloch tyrosinase (TYR) mRNA	Hylobates moloch	0.0	96.89%	XM_032168104.1	1578
6	PREDICTED: Pongo abelii tyrosinase (TYR) mRNA	Pongo abelii	0.0	97.35%	XM_002822337.3	1535
7	Human tyrosinase (TYR) mRNA, complete cds	Human sapiens	0.0	99.18%	M27160.1	1528
8	Human mRNA for tyrosinase (LOC 114161)	Human sapiens	0.0	99.18%	Y00819.1	1528
9	PREDICTED: Pan paniscus tyrosinase (TYR) mRNA	Pan paniscus	0.0	94.99%	XM_023209125.1	1489
10	Human tyrosinase and related tyrosinase, complete cds	Human sapiens	0.0	97.91%	M74314.1	1474

Accession No:

Accession
NM_000372.5
XM_003832989.2
XM_031006243.1
XM_001136041.2
XM_032168104.1
XM_002822337.3
M27160.1
Y00819.1
XM_023209125.1
M74314.1

E value:

The Expect value (E) is a parameter that describes the number of hits one can 'expect' to see by chance when searching a database of a particular size. It decreases exponentially as the Score (S) of the match increases. Essentially, the E value describes the random background

noise. For example, an E value of 1 assigned to a hit can be interpreted as meaning that in a database of the current size one might expect to see 1 match with a similar score simply by chance. The lower the E-value, or the closer it is to zero, the more 'significant' the match is." E value and the matrix score/identity score go inversely., exponential graph.

E value
0
0
0
0
0
0
0
0
0
0
0

Coverage:

Query Coverage
100%
100%
100%
100%
99%
95%
90%
90%
100%
90%

Identity:

Per. Identity
99.26
98.84
98.74
98.53
96.66
97.35
99.18
99.18
94.99
97.91

Step 8: To show the Alignment, Graphic summary and taxonomy of top ten similar sequences we need to click on 'Alignment'

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment View: Pairwise CDS feature Restore defaults Download

10 sequences selected

Download GenBank Graphics Sort by: E value Next Previous Descriptions

Homo sapiens tyrosinase (TYR). mRNA
Sequence ID: NM_009372.5 Length: 2052 Number of Matches: 2

Range: 1:1110 to 2061 GenBank GenBank

Score	Expect	Identities	Gaps	Strand
1712 bits(927)	0.0	945/952(99%)	7/952(0%)	Plus/Plus

Query 188 TGGAGGATTTTCCTAGTCCTCTTACTTGGATGACGAGTCCCTCCGAG-AGCTGACCA 158
 Sbjct 1110 TGGAGGATTTTCCTAGTCCTCTTACTTGGATGACGAGTCCCTCCGAG-AGCTGACCA 158

Query 190 ATGCTTTCACATCTATATGAATGACAAATCTCCAG-CAAGATCTCCACGATC 226
 Sbjct 1178 ATGCTTTCACATCTATATGAATGACAAATCTCCAG-CAAGATCTCCACGATC 226

Query 217 CTATCTTCTTCTTCCACATGCTTTTTCACATATTTTAAAGCAGTGGCTCCAGAGC 276
 Sbjct 1236 CTATCTTCTTCTTCCACATGCTTTTTCACATATTTTAAAGCAGTGGCTCCAGAGC 276

Query 277 ACCGTCTCTTCAAGAGTTTATCCAGAGCAGTCCACCATTTGACATACCGGGAAT 336
 Sbjct 1298 ACCGTCTCTTCAAGAGTTTATCCAGAGCAGTCCACCATTTGACATACCGGGAAT 336

Query 337 CCTACATGATTC---TTATACCACTGTGAGAGATGATTTCTTTATTTATCCAGAA 393

Feedback

‘Graphic summary’

Descriptions **Graphic Summary** Alignments Taxonomy

Hover to see the title click to show alignments Alignment Scores: < 40 40 - 60 60 - 80 80 - 200 >= 200

10 sequences selected

Distribution of the top 20 Blast Hits on 10 subject sequences

Query

1 200 400 600 800 1000

Twitter Facebook YouTube LinkedIn Messenger Blog Support Center

National Center for Biotechnology Information Popular PubMed Resources Literature Actions Submit Feedback

Taxonomy:

Query Length: 1044 Filter Reset

Other reports: Distance tree of results MSV viewer

Descriptions Graphic Summary Alignments **Taxonomy**

Reports Lineage Organism Taxonomy

10 sequences selected

Taxonomy	Number of hits	Number of Organisms	Description
Eukaryota	10	7	
Chromista	8	6	
Opisthokonta	8	6	
Amniota	7	4	
Mammalia	5	1	Homo sapiens hits
Primates	2	2	
Hominoidea	1	1	Pan paniscus hits
Homini	1	1	Pan troglodytes hits
Gorilla gorilla gorilla	1	1	Gorilla gorilla gorilla hits
Pongo	1	1	Pongo abelii hits
Hylobates moloch	1	1	Hylobates moloch hits
Ptilocobus leproscelus	1	1	Ptilocobus leproscelus hits

Feedback

Step 9: List out the set of sequences showed good similarity with query sequence

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Select columns Show 100

☐ select all 10 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 tyrosinase (TYR) mRNA	<i>Homo sapiens</i>	1712	1898	100%	0.0	99.20%	2062	NM_000372.5
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus tyrosinase (TYR) mRNA	<i>Pan paniscus</i>	1680	1874	100%	0.0	98.84%	2485	XM_003832989.2
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla gorilla tyrosinase (LOC101148886) mRNA	<i>Gorilla gorilla gorilla</i>	1685	1863	100%	0.0	98.74%	2485	XM_031026263.1
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes tyrosinase (TYR) mRNA	<i>Pan troglodytes</i>	1676	1859	100%	0.0	98.53%	4893	XM_001136091.2
<input checked="" type="checkbox"/> PREDICTED: Hylobates moloch tyrosinase (TYR) mRNA	<i>Hylobates moloch</i>	1578	1749	99%	0.0	96.66%	2486	XM_032168104.1
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii tyrosinase (TYR) mRNA	<i>Pongo abelii</i>	1535	1708	85%	0.0	97.35%	2490	XM_002822337.3
<input checked="" type="checkbox"/> Human tyrosinase (TYR) mRNA, complete cds	<i>Homo sapiens</i>	1528	1712	99%	0.0	99.18%	2384	M27189.1
<input checked="" type="checkbox"/> Human mRNA for tyrosinase (EC 1.14.18.1)	<i>Homo sapiens</i>	1528	1712	99%	0.0	99.18%	1888	Y00019.1
<input checked="" type="checkbox"/> PREDICTED: Ptilinopus tephrocetes tyrosinase (TYR) mRNA	<i>Ptilinopus tephrocetes</i>	1489	1651	100%	0.0	94.90%	2489	XM_023209125.1
<input checked="" type="checkbox"/> Haasien tyrosinase and mutant tyrosinase, complete cds	<i>Homo sapiens</i>	1474	1638	99%	0.0	97.91%	1929	M74314.1
<input type="checkbox"/> PREDICTED: Trachypithecus francoisi tyrosinase (TYR) mRNA	<i>Trachypithecus francoisi</i>	1472	1645	100%	0.0	94.60%	2484	XM_033205968.1
<input type="checkbox"/> PREDICTED: Colobus angolensis palliatus tyrosinase (TYR) mRNA	<i>Colobus angolensis palliatus</i>	1471	1642	99%	0.0	94.50%	2485	XM_011838472.1
<input type="checkbox"/> PREDICTED: Rhinopithecus roosei tyrosinase (TYR), transcript variant X1, mRNA	<i>Rhinopithecus roosei</i>	1467	1640	100%	0.0	94.57%	2488	
<input type="checkbox"/> PREDICTED: Cercopithecus aethiops tyrosinase (TYR) mRNA	<i>Cercopithecus aethiops</i>	1463	1630	100%	0.0	94.57%	2485	

Feedback