

Βιοπληροφορική

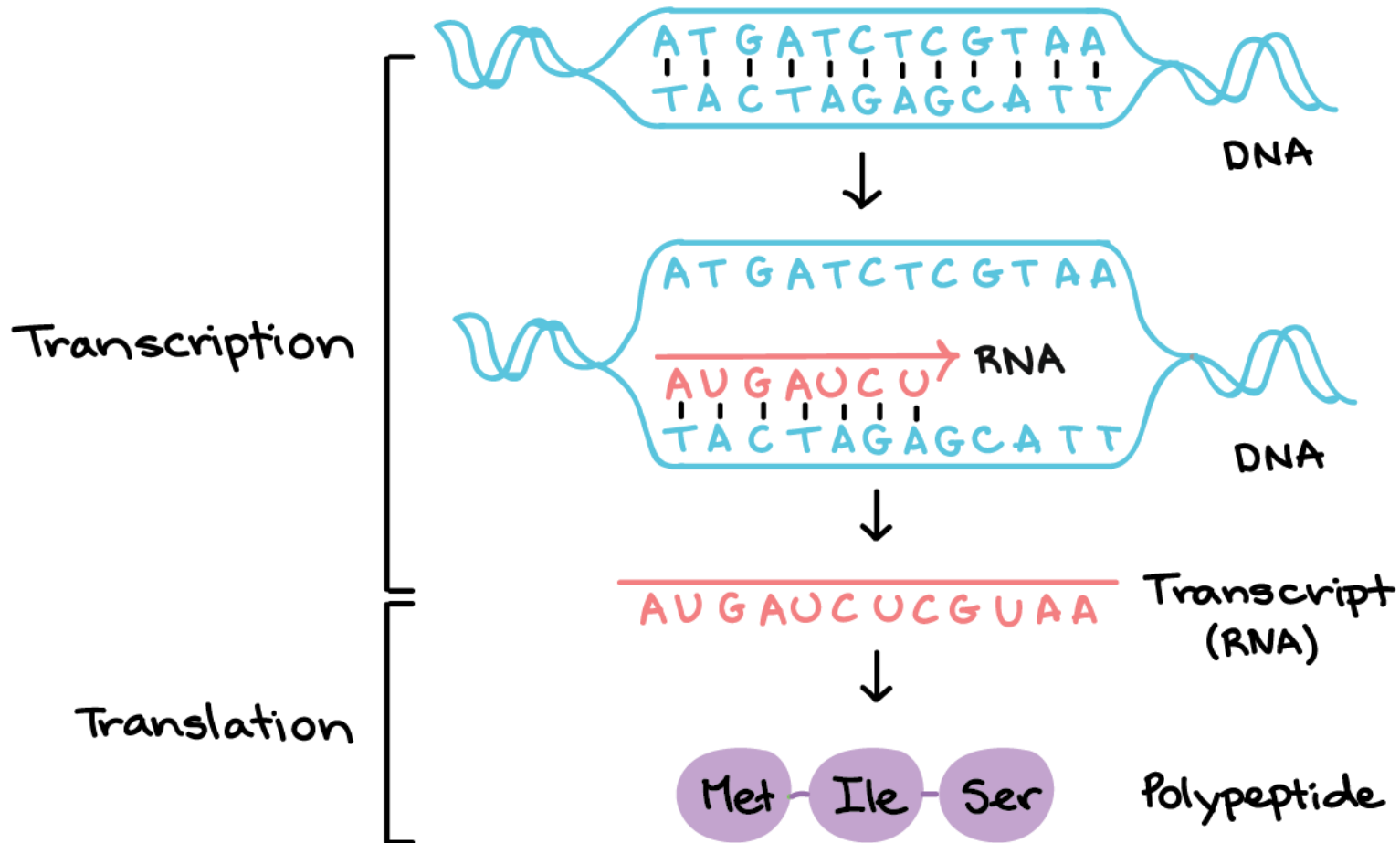
# Αλληλουχία

- Είναι μια σειρά από νουκλεοτίδια ή αμινοξέα όπως αυτά εμφανίζονται σε ένα μόριο DNA ή RNA και μια πρωτεΐνη αντίστοιχα

Πχ CGATCGTCTAGCCATGCTCATG

# Μεταγραφή

Οι Jacob και Monod ανακάλυψαν επίσης ότι η αλληλουχία ενός mRNA μορίου είναι συμπληρωματική με ένα τμήμα μίας από τις δύο αλυσίδες DNA.



# Μετάφραση

Η αλληλουχία των βάσεων του DNA δίνει τις οδηγίες, ώστε να συνδυαστούν τα κατάλληλα αμινοξέα στη σωστή σειρά. Οι οδηγίες δίνονται με τη μορφή κώδικα, ο οποίος καλείται **γενετικός κώδικας**. Κάθε αμινοξύ ορίζεται από τρεις βάσεις του DNA. Ο συνδυασμός των τριών βάσεων καλείται **τριπλέτα** ή **κωδικόνιο**

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } <b>AUG Met</b>	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

- <https://web.expasy.org/translate/>

# Expasy translate



ExpASY  
Bioinformatics Resource Portal

## Translate tool

---

**Translate** is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

---

Please enter a DNA or RNA sequence in the box below (numbers and blanks are ignored).

Output format:

Genetic code:

or

---

# One letter code

## Abbreviations for amino acids

<i>Amino acid</i>	<i>Three-letter abbreviation</i>	<i>One-letter symbol</i>
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Asparagine or aspartic acid	Asx	B
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glutamine or glutamic acid	Glx	Z
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

# Ανοιχτά πλαίσια ανάγνωσης

## Translate Tool - Results of translation

Open reading frames are highlighted in red. Please select one of the following frames - in the next page, you will be able to select your initiator and retrieve your amino acid sequence:

### 5'3' Frame 1

FQNPKEKPKETPTAAQWTLSSRVAARSSVESTYCRAGALDWVCARRGRPPNGPSPEVGASPYQRRASLRA  
GTGLSLVTLGPF **Met** ATRAVDGGDGGSCPPAALALIGAPTTGEPATWVRSSIP **Stop** PVASPTSWGTFPLWE  
HRLAVSRL **Stop** H **Met** VLGQSKTGST **Met** QPETFLVAPFLSSFWHFSRASPRPPRQSTIKTALVYTILP **Met** TAL  
RVLS **Met** RRTVSSCTALDA

### 5'3' Frame 2

SKTPKKNQKKHQPPPNGR **Stop** VPGWRPDRRWSLLIAAQQP **Stop** IGCARDEEDLRTVPAQR **Stop** APAHTKGA  
PVSGPALGSAWLPLAPLWQRGLW **Met** GR **Met** APVPPRLSP **Stop** LGPQRPPAKIPQH **Stop** GHRFPNLWLRP  
HGVHSRCGSTAWRCRGCFCSTWC **Stop** GNRRRDQLCNRKPSWLLLFYLPSTSLVPHHARLGSQ **Stop** KQLW  
YIPSYQ **Stop** LP **Stop** FEYCL **Stop** GGQCHHAQPW **Met** R

### 5'3' Frame 3

PKPQRKTKRNTNRRP **Met** DVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSEERSQPRGRRQPIPKARQSQG  
RHWAGPYWPPLYGNEGCGWAGWLLSPRGSRPNWGPNDRRRSRN **Met** GKVIDSLTCGFADL **Met** GYIPVV  
GAPLGGVAAALAHGVRAIEDGINYATGNLPGCSFSIFLLALLSCLTTPASAVNYKNSSGIYHLTNDCPNSSIV  
YEADSVI **Met** HSPGC

### 3'5' Frame 1

THPGLC **Met** **Met** TLSAS **Stop** TILELGQSLVRWYIPELFL **Stop** LTAEAGVVRHERSARRKIEKEQPGRFPVA **Stop** LI  
PSSIALTPCAKAAATPPSGAPTTG **Met** YP **Met** RSAKPQVRES **Met** TLP **Met** LRDLRRGSLGPQLGREPRGDRSH  
**PAHPQPSLP** **Stop** RGQG **Stop** PG **Stop** AQCRP **Stop** DWRAFG **Met** GWRLPLGWDRSEVFLVARTPNLGLPLRGNK  
**Stop** TPPTIWPPP GNLTSIGRRLVFLVFLWGF

### 3'5' Frame 2

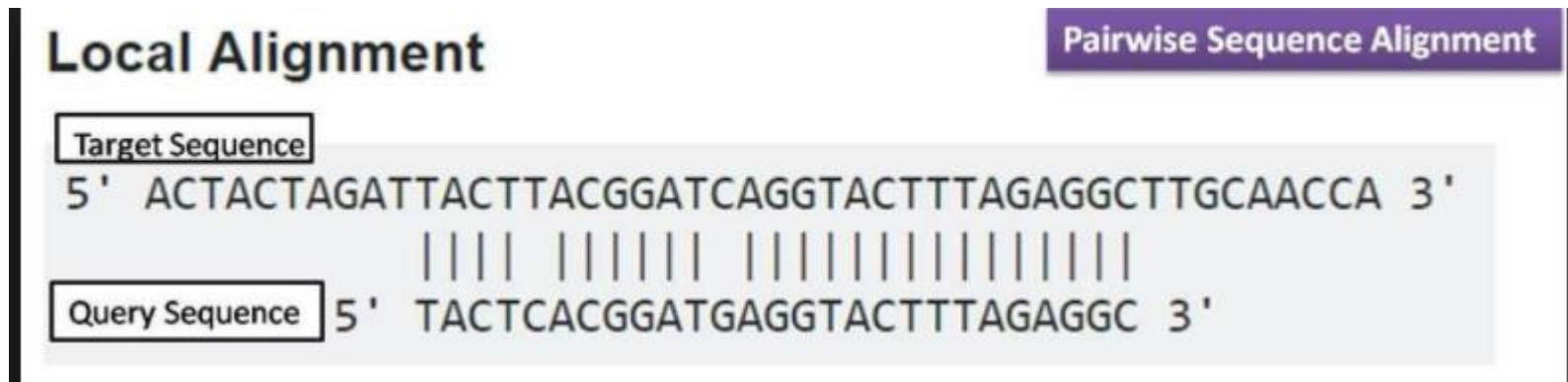
RIQGC **Stop** **Stop** HCPPHRQYSN **Stop** GSHW **Stop** DGIYQSCFYS **Stop** LPRRAW **Stop** GTREVPEGR **Stop** KRSNQE  
GFRLHS **Stop** SRLRLP **Stop** HHVLKQPRHRQAVLPQRECTP **Stop** GRRSHRLGNR **Stop** PYPCCGIFAGGRWGP  
**Stop** GESRGGTGAILPIHSPRCHKGAKGNQAEPGAPETGAPLVWAGAYLWAGTVRRSSSSRAHPI **Stop** GPC  
AAISRLHRRSGRHPGT **Stop** RPLGGGWCFWFFFGVLE

### 3'5' Frame 3

ASRAVHDDTVRLIDNTRIRAVIGK **Met** VYTRAVFIVDCRGGRGEAREKCKQKEDRKGATRKYVSGCIVDPVFDG  
**PNT** **Met** C **Stop** SSRDTAKRCSHNGNVPHEVGEATG **Stop** GIDDLTHVAGSSPGVVGAPIRARAAGGQEPSCPST  
ALVAIKGPRVTRLSPVPALRLARLWYGLAPTSGLGPFGLPRAHTQSRAPARQ **Stop** VDSTDDLAATRELN  
VHWAAVGVSFGFSLGF

# Στοίχιση και πολλαπλή στοίχιση

- Εύρεση της μέγιστης δυνατής ομοιότητας μεταξύ δύο αλληλουχιών





# Στοίχιση και πολλαπλή στοίχιση

- DNA

<1> 28~68	AACAAGCA-A-ACTTTTATCCATGGTCGTGGTACAGAGGGGGTC
<4> 333~373	AACAAGCA-A-ACTTTTATCCATGGTCGTGGTACAGAGGGGGTC
<8> 154~193	AACAAGCA-A-ACTTTTATCCATGGTCGTGGTACAGAGGGGGT-
<6> 615~647	AACAAGCAGA-ACTTTTATCCATGGTCGTGGTAC-----
<4> 502~533	AACAAGCA-ACCCTTTTATCCATGGTCGTGGTA-----
<1> 844~872	AACAAGCA-A-ACTTTTATCCATGGTCGTGG-----
<8> 194~220	-----A-ACTTTTATCCATGGTCGTGGTACAGA-----
<5> 451~480	-----CTTTCA-ACGTGGTCGTGGTACAAAGGGGGTC

# Στοίχιση και πολλαπλή στοίχιση

- Πρωτεΐνη

**Histone H1** (residues 120-180)

HUMAN	KKASKPKKAASKAPT	KKPKATPVKKA	KKKLAATPKKAKKPK	TVKAKPNKASKPKKAKPVK
MOUSE	KKAAKPKKAASKAPS	KKPKATPVKKA	KKKRAATPKKAKKPK	VVKVPVKASKPKKAKTVK
RAT	KKAAKPKKAASKAPS	KKPKATPVKKA	KKKRAATPKKAKKPK	IVKVPVKASKPKKAKPVK
COW	KKAAKPKKAASKAPS	KKPKATPVKKA	KKKRAATPKKTKKPK	TVKAKPNKASKPKKTKPVK
CHIMP	KKASKPKKAASKAPT	KKPKATPVKKA	KKKLAATPKKAKKPK	TVKAKPNKASKPKKAKPVK
	***	:*****	:*****	*****:***** **

NON-CONSERVED  
AMINO ACIDS

Conservative

Conservative

Non-conservative

Conservative

Non-conservative

Semi-conservative

Conservative

Non-conservative

# Clustal Omega

[Input form](#)[Web services](#)[Help & Documentation](#)[Feedback](#)[Share](#)

[Tools](#) > [Multiple Sequence Alignment](#) > [Clustal Omega](#)

## Service Retirement

[Wise2DBA](#) and [Promoterwise](#) are scheduled for retirement on 15th April 2018. Alternatives can be found at [Exonerate](#), [BWA](#) or [BLAT](#). If you have any concerns, please contact us via [support](#).

## Multiple Sequence Alignment

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between **three or more** sequences. For the alignment of two sequences please instead use our [pairwise sequence alignment tools](#).

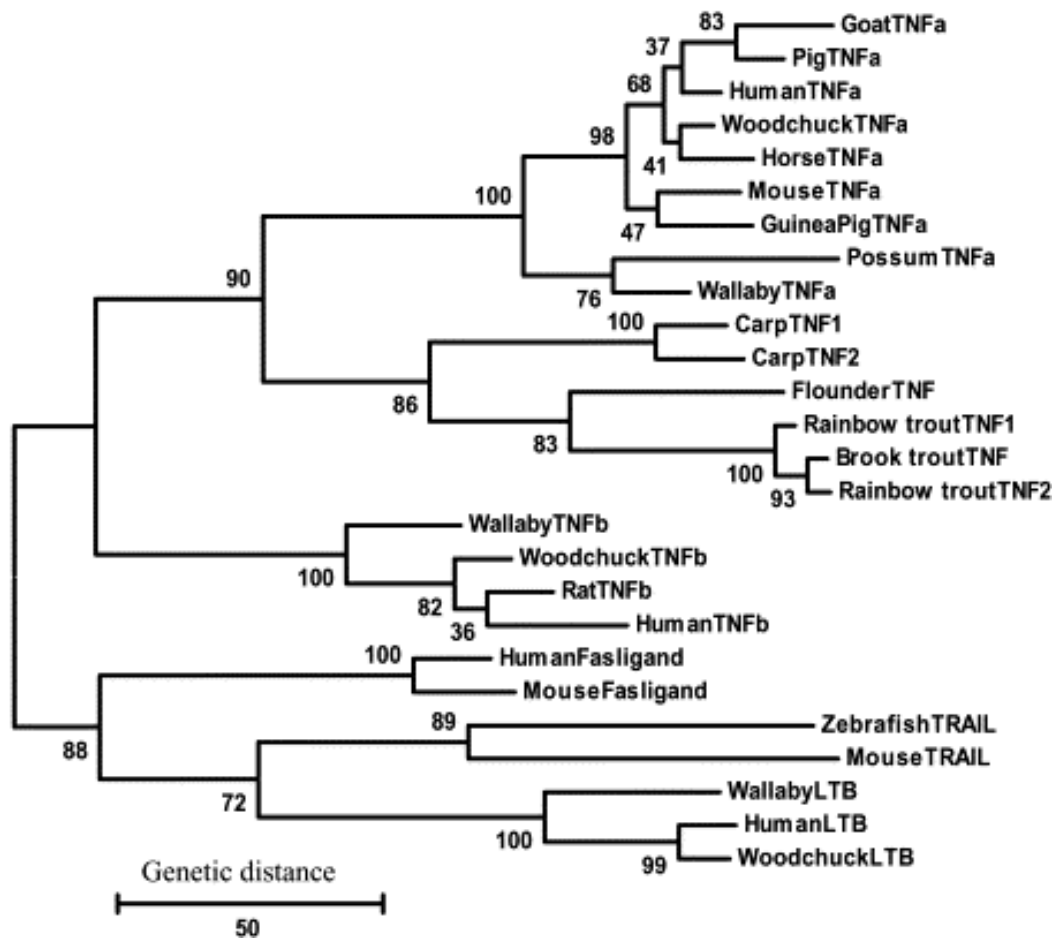
**Important note:** This tool can align up to 4000 sequences or a maximum file size of 4 MB.

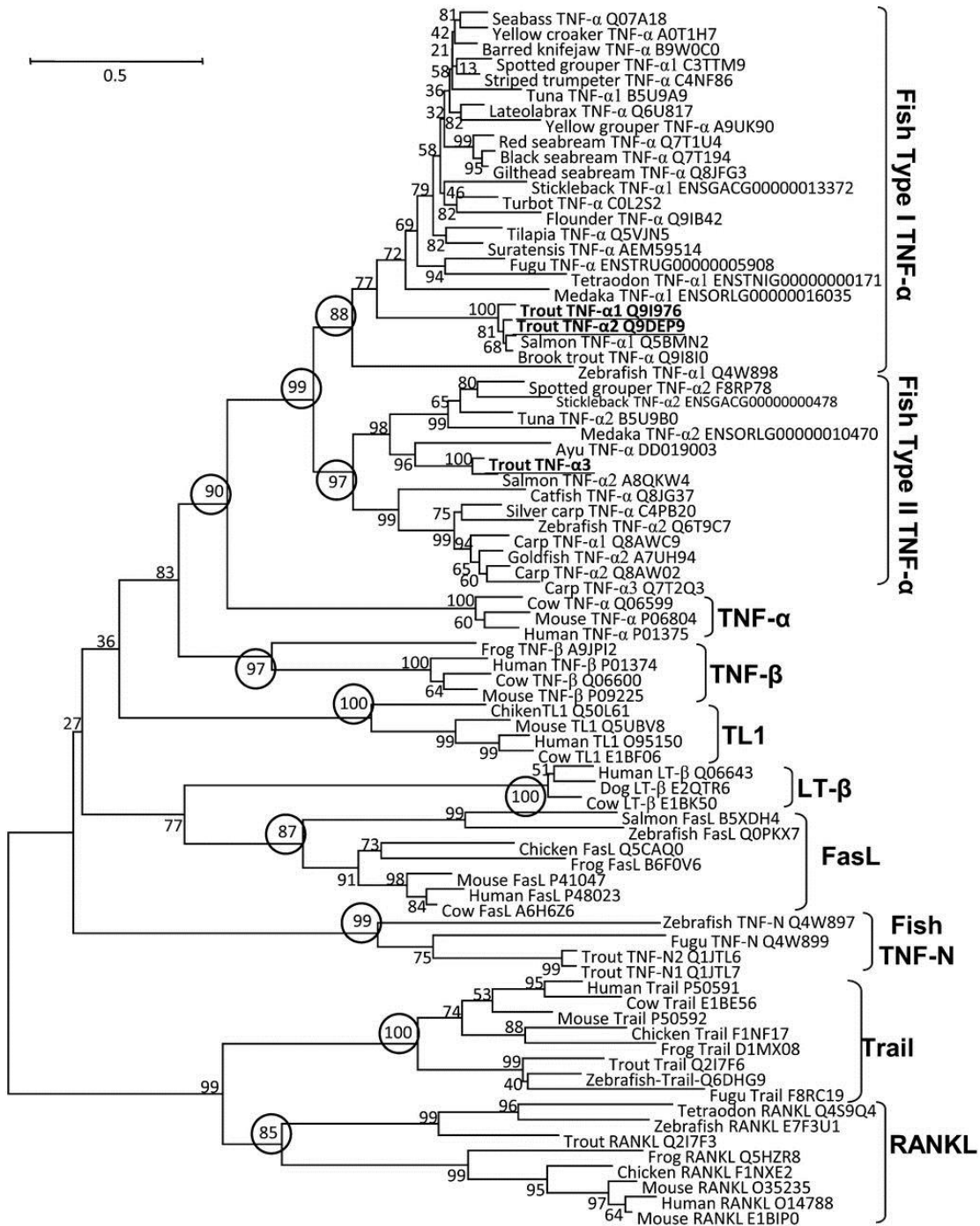
### STEP 1 - Enter your input sequences

[Enter or paste a set of](#)

PROTEIN

[sequences](#) in any supported format:





# Pubmed - GenBank

Nucleotide

Nucleotide

Search

[Advanced](#)

[Help](#)

## Nucleotide

The Nucleotide database is a collection of sequences from several sources, including GenBank, RefSeq, TPA and PDB. Genome, gene and transcript sequence data provide the foundation for biomedical research and discovery.

### Using Nucleotide

[Quick Start Guide](#)

[FAQ](#)

[Help](#)

[GenBank FTP](#)

[RefSeq FTP](#)

### Nucleotide Tools

[Submit to GenBank](#)

[LinkOut](#)

[E-Utilities](#)

[BLAST](#)

[Batch Entrez](#)

### Other Resources

[GenBank Home](#)

[RefSeq Home](#)

[Gene Home](#)

[SRA Home](#)

[INSDC](#)

LOCUS KY500678 1796 bp DNA linear PRI 28-APR-2017  
DEFINITION Homo sapiens isolate RSH TNF gene, complete cds.  
ACCESSION KY500678  
VERSION KY500678.2  
KEYWORDS .  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;  
Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1796)  
AUTHORS Norman,P.J., Norberg,S.J., Nemat-Gorgani,N., Ronaghi,M. and  
Parham,P.  
TITLE CDS alleles of MHC region genes derived from homozygous individuals  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 1796)  
AUTHORS Norman,P.J., Norberg,S.J., Nemat-Gorgani,N., Ronaghi,M. and  
Parham,P.  
TITLE Direct Submission  
JOURNAL Submitted (17-JAN-2017) Structural Biology, Stanford University  
School of Medicine, 299 Campus Drive West, Stanford, CA 94305, USA  
REFERENCE 3 (bases 1 to 1796)  
AUTHORS Norman,P.J., Norberg,S.J., Nemat-Gorgani,N., Ronaghi,M. and  
Parham,P.  
TITLE Direct Submission  
JOURNAL Submitted (28-APR-2017) Structural Biology, Stanford University  
School of Medicine, 299 Campus Drive West, Stanford, CA 94305, USA  
REMARK Sequence update by submitter  
COMMENT On Apr 28, 2017 this sequence version replaced [gi:1159611448](https://www.ncbi.nlm.nih.gov/nuccore/gi/1159611448).

FEATURES	Location/Qualifiers
source	1..1796 /organism="Homo sapiens" /mol_type="genomic DNA" /isolate="RSH" /db_xref="taxon: <a href="#">9606</a> "
<a href="#">mRNA</a>	join(<1..186,793..838,1026..1073,1375..>1796) /product="TNF"
<a href="#">CDS</a>	join(1..186,793..838,1026..1073,1375..1796) /codon_start=1 /product="TNF" /protein_id=" <a href="#">A0Y77150.1</a> " /translation="MSTESMIRDVELAEEALPKKTGGPQGSRRCLFLSLFSFLIVAGA TTLFCLLHFGVIGPQREEFPRDLSLISPLAQAVRSSSRTPSDKPVAHVVANPQAEQQL QWLNRRANALLANGVELRDNQLVVPSEGLYLIYSQVLFKGQGCPSTHVLLTHTISRIA VSYQTKVNLLSAIKSPCQRETPEGAEAKPWYEPYIYLGGVFQLEKGDRLSAEINRPDYL DFAESGQVYFGIIAL"



ORIGIN

```

1 atgagcactg aaagcatgat ccgggacgtg gagctggccg aggaggcgct cccaagaag
61 acaggggggc cccagggctc caggcgttgc ttgttctca gcctcttctc cttcctgatc
121 gtggcaggcg ccaccacgct cttctgcctg ctgcactttg gagtgatcgg cccccagagg
181 gaagaggtga gtgcctggcc agccttcac cactctcca cccaagggga aatggagacg
241 caagagaggg agagagatgg gatgggtgaa agatgtgctg tgataggagg ggatggagag
301 aaaaaaacgt ggagaaagac ggggatgcag aaagagatgt ggcaagagat ggggaagaga
361 gagagagaaa gatggagaga caggatgtct ggcacatgga aggtgctcac taagtgtgta
421 tggagtgaat gaatgaatga atgaatgaac aagcagatat ataaataaga tatggagaca
481 gatgtggggt gtgagaagag agatggggga agaaacaagt gatatgaata aagatgggtga
541 gacagaaaaga gcgggaaata tgacagctaa ggagagagat gggggagata aggagagaag
601 aagatagggt gtctggcaca cagaagacac tcagggaaag agctgttgaa tgcctggaag
661 gtgaatacac agatgaatgg agagagaaaa ccagacacct cagggctaag agcgcaggcc
721 agacaggcag ccagctgttc ctctttaag ggtgactccc tcgatgttaa ccattctcct
781 tctcccaac agttccccag ggacctctct ctaatcagcc ctctggcca ggagtcagt
841 aagtgtctcc aaacctcttt cctaattctg ggtttgggtt tggggtagg gttagtaccg
901 gtatggaagc agtgggggaa atttaaagtt ttggtcttgg gggaggatgg atggaggtga
961 aagtaggggg gtattttcta ggaagtttaa gggctcagc ttttctttt ctctctctc
1021 ttcaggatca tcttctcgaa ccccgagtga caagcctgta gcccattgtg taggtaagag
1081 ctctgaggat gtgtcttggg acttggaggg ctaggatttg gggattgaag cccggctgat
1141 ggtaggcaga acttggagac aatgtgagaa ggactcgctg agctcaaggg aagggtggag
1201 gaacagcaca ggccttagtg ggatactcag aacgtcatgg ccagggtggga tgtgggatga
1261 cagacagaga ggacaggaac cggatgtggg gtgggcagag ctcgagggcc aggatgtgga
1321 gagtgaaccg acatggccac actgactctc ctctccctct ctccctccct ccagcaaac
1381 ctcaagctga ggggcagctc cagtggctga accgccgggc caatgccctc ctggccaatg
1441 gcgtggagct gagagataac cagctggtgg tgccatcaga gggcctgtac ctcatctact
1501 cccaggtcct cttcaagggc caaggctgcc cctccacca tgtgctctc acccacacca
1561 tcagccgcat cgccgtctcc taccagacca aggtcaacct cctctctgcc atcaagagcc
1621 cctgccagag ggagacccca gagggggctg aggccaagcc ctggtatgag cccatctatc
1681 tgggaggggt cttccagctg gagaaggggtg accgactcag cgctgagatc aatcggcccg
1741 actatctcga ctttgccgag tctgggcagg tctactttgg gatcattgcc ctgtga

```

//

# Ψάχνοντας στις βάσεις

- Basic Local Alignment Search Tool (BLAST)

Το βασικό εργαλείο αναζήτησης τοπικής στοίχισης (BLAST) είναι ένα πρόγραμμα που αναφέρει περιοχές τοπικής μεταξύ αλληλουχίας αναζήτησης και αλληλουχιών σε μια βάση δεδομένων. Η ικανότητα ανίχνευσης της ομολογίας αλληλουχίας μας επιτρέπει να προσδιορίσουμε εάν ένα γονίδιο ή μια πρωτεΐνη σχετίζεται με άλλα γνωστά γονίδια ή πρωτεΐνες. Η ανίχνευση της ομολογίας αλληλουχίας διευκολύνει επίσης την αναγνώριση των διατηρημένων τομέων που μοιράζονται τα πολλαπλά γονίδια και την αναγνώριση των μελών μιας οικογένειας γονιδίων



- Homo sapiens
- Pan troglodytes
- Mus musculus
- Ovis aries
- Felis catus
- Canis lupus
- Aedes albopictus
- Drosophila melanogaster
- Solanum tuberosum
- Trypanosoma cruzi
- Nicotiana tabacum



# NM\_001101 human

- >pan
- 1 ataaaaaacc ggcggcgcga cgcgccacca cgcggagac cgcgtccgcc ccgcgagcac
- 61 agagcctcgc ctttgccgat cgcgcgccg tccacaccg ccgccagctc accatgggatg
- 121 atgatatcgc cgcgctcgtt gtcgacaacg gctccggcat gtgcaaggcc ggcttcgctgg
- 181 gcgacgatgc cccccgggcc gtcttcccct ccatcgtggg gcgccccagg caccagggcg
- 241 tgatgggtggg catgggtcag aaggattcct atgtgggcga cgaggcccag agcaagagag
- 301 gcatcctcac cctgaagtac cctatcagc acggcatcgt caccaactgg gacgacatgg
- 361 agaagatctg gcaccacacc ttctacaatg agctgcgtgt ggctcccag agcacccccg
- 421 tgcctctcac cgaggcccc ctgaaccca aggccaaccg cgagaagatg acccagatca
- 481 tgtttgagac ctcaacacc ccagccatgt acgttgctat ccaggctgtg ctatccctgt
- 541 acgcctctgg ccgtaccat ggcatcgtga tggactccgg tgacggggtc acccacactg
- 601 tgcccacta cgaggggat gcctcccc atgccatcct gcgtctggac ctggctggcc
- 661 gggacctgac tgactacct atgaagatcc tcaccgagcg cggctacagc tcaccacca
- 721 cggccgagcg ggaatcgtg cgtgacatta aggagaagct gtgctacgtc gcctggact
- 781 tcgagcagga gatggccacg gctgcttcca gctcctcct ggagaagagc tacgagctgc
- 841 ctgacggcca ggtcatcacc attgcaatg agcgggtccg ctgccctgag gcactcttc
- 901 agccttctt cctgggcatg gagtctgtg gcacacga aactacctc aactccatca
- 961 tgaagtgtga cgtggacatc cgcaaagacc tgtacgcaa cacagtgtg tctggcggca
- 1021 ccaccatgta cctggcatc gccgacagga tgcagaagga gatcactgcc ctggcaccca
- 1081 gcacaatgaa gatcaagatc attgctcctc ctgagcgcaa gtactccgtg tggatcggcg
- 1141 gctccatcct ggctcgtcgt tccacctcc acgagatgtg gatcagcaag caggagatg
- 1201 acgagtccgg ccctccatc gtccaccgca aatgcttcta ggcggactgt gacttagttg
- 1261 cgttacacc tttcttgaca aaacctaaat tgcgcagaaa acaagatgag attggcatgg
- 1321 ctttattgt ttttttgt ttgtttgt tttttttt ttttggctg actcaggatt
- 1381 taaaaactgg aacggtgaag gtgacagcag tcggttggag cgagcatccc caaagtca
- 1441 caatgtggcc gaggacttg attgcacatt gttctttt taatagtcac tccaaatag
- 1501 agatgcattg ttacaggaag tcccttgcca tcctaaaagc caccctactt ctctcaagg
- 1561 agaattggccc agtcctcctc cgagtccaca caggggaggt gatagcattg cttttgtga
- 1621 aattatgtaa tgcaaaatt ttttaactt cgccttaata ctttttatt ttgtttatt
- 1681 tgaatgatg agccttcgtg cccccctc cccttttt gtccccaac ttgagatga
- 1741 tgaaggctt ttgtctcctt gggagtgggt ggaggcagcc agggcttacc tgtactactg
- 1801 ctgagacca gttgaataaa agtgcacacc ttaaaaaaa aaaaaaaaa