

BLASTing Through the Kingdom of Life

5/28/2010

Chris Overall
Department of Bioinformatics & Genomics
University of North Carolina - Charlotte

BLASTing Through the Kingdom of Life

Learning Objectives

- Learn about sequence alignment.
 - What is sequence alignment?
 - What is biological sequence alignment?
 - Why is biological sequence alignment important?
- Learn how to use BLAST for biological sequence alignment.

BLASTing Through the Kingdom of Life

What is a Sequence Alignment?

- Arrange two or more words (sequences of letters) so that you can identify regions of similarity between them.
- For the current discussion, the alphabet (set of letters) is the standard English alphabet (size = 26).
- Consider these four words: bat, hat, batter, hatter.
- How similar are **bat** and **hat**?
- Which word is more similar to **bat**: **hat** or **batter**? What about **hatter**? Why?
- Which pair of words is more similar to each other: (**bat, hat**) or (**batter, hatter**)? Why?

BLASTing Through the Kingdom of Life

Sequence Alignments → Longest Common Subsequence

- Let's be a bit more formal about determining similarity.
- For a particular alignment, we will score them as follows:
 - Perfect match = 1 (A – A, B – B, etc.)
 - Mismatch = 0 (A – B, B – A, etc.)
- Using this scoring scheme, we'll find the longest common subsequence in each word pair and score them to determine the most similar sequences.

BLASTing Through the Kingdom of Life

Sequence Alignments → Longest Common Subsequence Examples

Alignment 1

BAT

||

HAT

Score = 2

Alignment 2

HAT

||

BATTER

Score = 2

Alignment 3

BAT

||

HATTER

Score = 2

Alignment 4

BAT

|||

BATTER

Score = 3

Alignment 5

HAT

|||

HATTER

Score = 3

Alignment 6

BATTER

|||||

HATTER

Score = 5

BLASTing Through the Kingdom of Life

Sequence Alignments → Gapped Alignments

- Consider these three sequences:
 1. THE CATS IN THE HAT
 2. THE CAT IN THE HAT
 3. THE CAT IS A HAT
- Which phrase (2 or 3) is the most similar to phrase 1?

BLASTing Through the Kingdom of Life

Sequence Alignments → Gapped Alignments

- Again, let's be a bit more formal about determining similarity.
- For a particular alignment, we will score it as follows:
 - Perfect match = 1 (A – A, B – B, etc.)
 - Mismatch = 0 (A – B, B – A, etc.)
 - Gap open = -1
 - Gap extension = 0
- Using this scoring scheme, we'll find the longest gapped subsequence in each word pair and score them to determine the most similar sequences.

BLASTing Through the Kingdom of Life

Sequence Alignments → Gapped Alignment Examples

Alignment 1

```
THECATSINTHEHAT
|||||
THECATINTHEHAT
```

Score = 6

Alignment 2

```
THECATSINTHEHAT
||||| |||||
THECAT-INTHEHAT
```

Score = 14 - 1 = 13

Alignment 3

```
THECATSINTHEHAT
|||||
THECATISAHAT
```

Score = 6

Alignment 4

```
THECATSINTHEHAT
||||| |||
THECAT---ISAHAT
```

Score = 6 - 1 + 3 = 8

BLASTing Through the Kingdom of Life

Sequence Alignments → Global vs. Local Alignment

- **Global alignment** → find the single best alignment across the entire length of both sequences.
- **Local alignment** → find one or more highly similar local regions between both sequences.



BLASTing Through the Kingdom of Life

What is a Biological Sequence Alignment?

- Arrange two or more molecular sequences (DNA, RNA, protein) to identify regions of similarity between them.
 - Matches, mismatches and gaps.
 - Pairwise vs. multiple sequence alignment (MSA)
- We'll focus on DNA.
- DNA is a sequence of four possible nucleotides: adenine, guanine, cytosine and thymine.
 - DNA alphabet (letters) = {A, C, G, T}
- Much smaller than English alphabet.

BLASTing Through the Kingdom of Life

Biological Sequence Alignment Example

- How would you globally align these two sequences?
 - Sequence 1 → ACTG
 - Sequence 2 → ACGTG

BLASTing Through the Kingdom of Life

Biological Sequence Alignment Example

Alignment 1

```
ACTG-  
  ||  
ACGTG
```

Alignment 2

```
-ACTG  
  ||  
ACGTG
```

Alignment 3

```
AC-TG  
||  ||  
ACGTG
```

BLASTing Through the Kingdom of Life

Why are Biological Sequence Alignments Important?

- The more similar two molecular sequences are, the more likely that the molecules are also similar in structure, function or evolutionary history.
- DNA sequences
 - Mismatch = point mutation
 - Gap = insertion or deletion (indel)
- When developing primers for a PCR reaction (you'll learn about this later), need to make sure that the primer is unique to the organism's genome.

BLASTing Through the Kingdom of Life

Why Do We Need Computers and Algorithms to Find Alignments?

- Why not find all biological sequence alignments manually?
- How many times can you find the **query** sequence ATCGGCCATTAC in the following **target** sequence? Is it there at all? If so, is it unique?

```
ATCACTGTAGTAGTAGCTGGAAAGAGAAATCTGTGACTCCAATTAGCCAGTTCCTGCAGACCTTGTGAGGACTAG
AGGAAGAATGCTCCTGGCTGTTTTGTACTGCCTGCTGTGGAGTTTCCAGACCTCCGCTGGCCATTTCCCTAGAGC
CTGTGTCTCCTCTAAGAACCTGATGGAGAAGGAATGCTGTCCACCGTGGAGCGGGGACAGGAGTCCCTGTGGCCA
GCTTTCAGGCAGAGGTTTCTGTGAGAATATCCTTCTGTCCAATGCACCACTTGGGCCTCAATTTCCCTTCACAGG
GGTGGATGACCGGGAGTCGTGGCCTTCCGTCTTTTATAATAGGACCTGCCAGTGCTCTGGCAACTTCATGGGATT
CAACTGTGGAAACTGCAAGTTTGGCTTTTGGGGACCAAATGCACAGAGAGACGACTCTTGGTGAGAAGAAACAT
CTTCGATTTGAGTGCCCCAGAGAAGGACAAATTTTTTGCCTACCTCACTTTAGCAAAGCATAACCATCAGCTCAGA
CTATGTCATCCCCATAGGGACCATTTGGCCAAATGAAAAATGGATCAACACCCATGTTTAAACGACATCAATATTTA
TGACCTCTTTGTCTGGATGCATTATTATGTGTCAATGGATGCACTGCTTGGGGGATCTGAAATCTGGAGAGACAT
TGATTTTGCCCATGAAGCACCAGCTTTTCTGCCTTGGCATAGACTCTTCTTGTTGCGGTGGGAACAAGAAATCCA
GAAGCTGACAGGAGATGAAAACCTCACTATTCCATATTGGGACTGGCGGGATGCAGAAAAGTGTGACATTTGCAC
AGATGAGTACATGGG
```

BLASTing Through the Kingdom of Life

What is BLAST?

- Basic Local Alignment Search Tool
- Very fast tool for finding local regions of similarity between sequences.
 - Given a **query sequence**, search one or more **target databases** for matches.
- Target databases are extremely large; millions of sequences.
 - **nt** = non-redundant nucleotide sequence database
 - **nr** = non-redundant protein sequence database

BLASTing Through the Kingdom of Life

Interpreting BLAST Results

- For a given query sequence, you will receive a list of matching target sequences with corresponding alignments.
 - The query sequence can match more than one region for a target sequence (local alignments).
- Each alignment will have an E-value; statistical significance of alignment.
 - Number of sequence matches you would expect to find in a target database composed entirely of random sequences.

BLASTing Through the Kingdom of Life

Interpreting BLAST Results

- Low E-value means that the match is not random.
- What does an E-value = 4 mean?
- A couple observations about E-values to remember:
 - Greater chance of finding a random match in a larger target database (larger E-value).
 - Greater chance of finding a random match for a shorter query sequence (larger E-value).

BLASTing Through the Kingdom of Life

Hands-On Tutorial

- Go to <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

BLASTing Through the Kingdom of Life

Hands-On Tutorial

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

[Learn more](#) about how to use the new BLAST design

BLAST Assembled Genomes

Choose a species genome to search, or [list all genomic BLAST databases](#).

- [Human](#)
- [Mouse](#)
- [Rat](#)
- [Arabidopsis thaliana](#)
- [Oryza sativa](#)
- [Bos taurus](#)
- [Danio rerio](#)
- [Drosophila melanogaster](#)
- [Gallus gallus](#)
- [Pan troglodytes](#)
- [Microbes](#)
- [Apis mellifera](#)

Basic BLAST

Choose a BLAST program to run.

- [nucleotide blast](#) Search a **nucleotide** database using a **nucleotide** query
Algorithms: blastn, megablast, discontinuous megablast
- [protein blast](#) Search **protein** database using a **protein** query
Algorithms: blastp, psi-blast, phi-blast
- [blastx](#) Search **protein** database using a **translated nucleotide** query
- [tblastn](#) Search **translated nucleotide** database using a **protein** query
- [tblastx](#) Search **translated nucleotide** database using a **translated nucleotide** query

Specialized BLAST

Copyright © [Digital World Biology](#) All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

- Copy the first query sequence (> Example) from <http://www.digitalworldbiology.com/BLAST/62000sequences.html>
- Paste target sequence into textbox on BLAST start screen.
- Additional step:
 - Algorithm parameters → Set max target sequences to 500.

The screenshot shows the NCBI BLAST web interface. The 'Enter Query Sequence' section contains a text box with a nucleotide sequence: GAATCGGAGAGTGTGTCACCTAGCGCGGGGAACATCGAGCAATTC AAGATGACCATTTG CACGACAAGCAGTTTCAGGC ACTGAAGTTGTTGAGAAGCTCAGCGTAGCCGCCACTGGTGAGCCA GTTCCTGCAGACCAGATCGACGAAAGGCTTAGAAACATCACAA. The 'Choose Search Set' section has 'Database' set to 'Nucleotide collection (nr/nt)' and 'Others (nr etc.)' selected. The 'Program Selection' section has 'More dissimilar sequences (discontiguous megablast)' selected. A green arrow points to the 'BLAST' button. A yellow box with a list of instructions is overlaid on the bottom right of the interface.

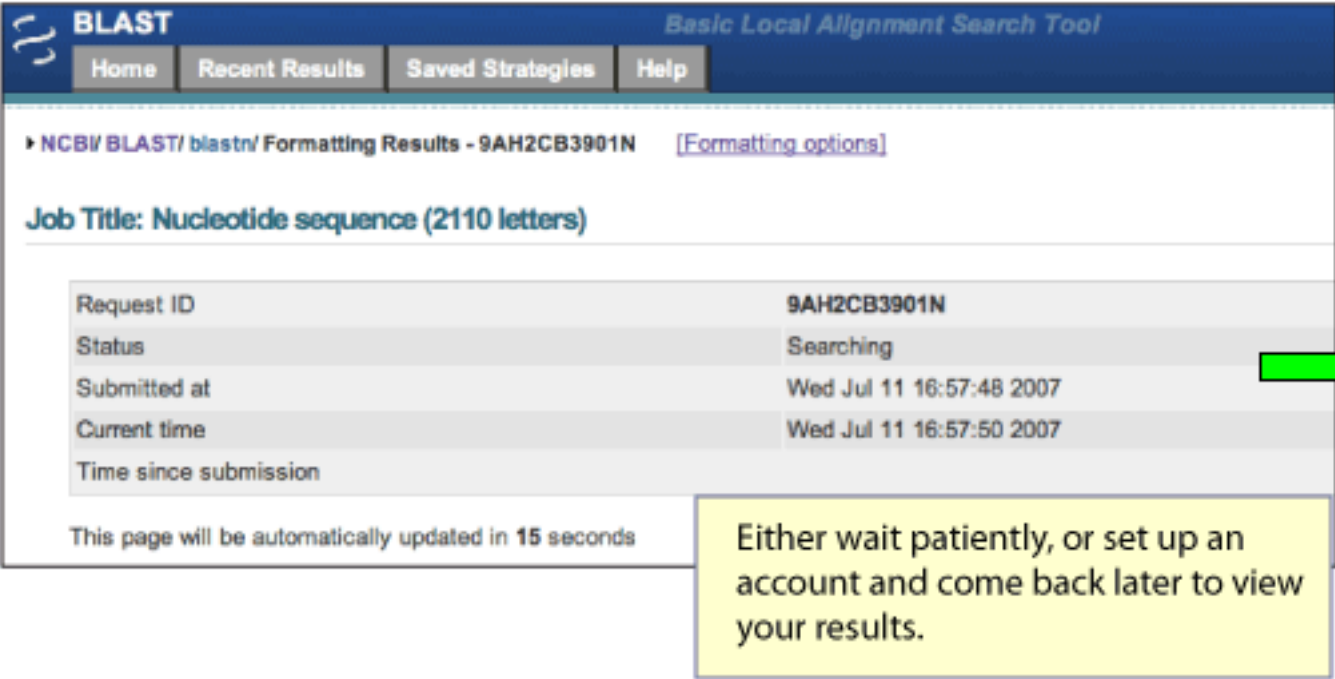
For the BLAST activities:

1. Change the database to "Others"
2. Select "nucleotide collection" from the pull-down menu.
3. Change the program selection to optimize for more dissimilar sequences.

Copyright © [Digital World Biology](http://www.digitalworldbiology.com) All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial



The screenshot shows the NCBI BLAST web interface. At the top, there is a navigation bar with links for Home, Recent Results, Saved Strategies, and Help. Below this, the page title is "NCBI/BLAST/blastn/ Formatting Results - 9AH2CB3901N" with a link for [Formatting options]. The job title is "Nucleotide sequence (2110 letters)". A table displays the job details:

Request ID	9AH2CB3901N
Status	Searching
Submitted at	Wed Jul 11 16:57:48 2007
Current time	Wed Jul 11 16:57:50 2007
Time since submission	

Below the table, it states "This page will be automatically updated in 15 seconds". A yellow box contains the text: "Either wait patiently, or set up an account and come back later to view your results." A green arrow points to the right from the "Searching" status in the table.

Copyright © [Digital World Biology](#) All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

- How many sequences are in the database?
- How many letters are in the database?
- How long is the query sequence?

Query ID |c|55449
Description Example
Molecule type nucleic acid
Query Length 840

Database Name nr
Description All GenBank+EMBL+DBJ+PDB sequences (but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences)
Program BLASTN 2.2.23+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#)

Search Parameters	
Program	blastn
Word size	7
Expect value	10
Hitlist size	500
Match/Mismatch scores	2,-3
Gapcosts	5,2
Low Complexity Filter	Yes
Filter string	L;m;
Genetic Code	1

Database	
Posted date	Apr 25, 2010 5:43 PM
Number of letters	31,036,146,654
Number of sequences	11,514,615
Entrez query	none

Karlin-Altschul statistics		
Lambda	0.633731	0.625
K	0.408146	0.41
H	0.912438	0.78

Results Statistics	
Length adjustment	36
Effective length of query	804
Effective length of database	30621620514
Effective search space	24619782893256
Effective search space used	24619782893256

BLASTing Through the Kingdom of Life

Hands-On Tutorial

The thick bar represents the sequence that was used to search the database. This sequence is called the "query" sequence.

[Blast Hits on the Query Sequence](#)
click to show alignments

Color key for alignment scores

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Magenta
>=200	Red

Query
0 400 800 1200 1600 2000

This scale shows the nucleotide position in the query sequence. The first base is position 1, the last base, position 2110

Each bar indicates a region where a sequence from the database matches the query sequence. The color of the bar corresponds to the length of the matching region and how well the sequences match each other.

Hashed areas represent gaps or regions that differ between the two sequences.

Scroll down

Copyright © Digital World Biology All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

- What is the most likely identity of your sequence? What is its accession number?
- What data supports this conclusion?
- What organism is the source of the target sequence?
- Approximately how many alignments have an E-value less than 0.01?
- For which other organisms are there matches?

A description of the sequence

The total score includes scores from non-contiguous portions of the subject sequence that match the query.

The Max % ident corresponds to the match to a subject sequence with the highest percentage of identical bases.

Legend for links to other resources: [U](#) UniGene [E](#) EMBL [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
X15893.1	Tarantula mRNA for hemocyanin subunit a	4532	4057	100%	0.0	100%	
AJ247807.1	Nephila inaurata madagascariensis mRNA for hemocyanin	662	662	79%	0.0	73%	
AJ247811.1	Nephila inaurata madagascariensis mRNA for hemocyanin	202	319	43%	5e-48	88%	
AJ247809.1	Nephila inaurata madagascariensis mRNA for hemocyanin	185	241	14%	7e-43	83%	
AJ307908.1	Cupiennius salei mRNA for hemocyanin subunit 5' (hc-5')	125	298	19%	6e-40	84%	
AJ277492.1	Euryelma californicum mRNA for hemocyanin subunit g	171	278	22%	8e-39	79%	

The Accession number is linked to the GenBank record.

A score that indicates how well the sequences match. For nucleotide sequences, this is approximately equal to twice the length of the matching region.

The Max score is linked to data that show where the sequences match.

The query coverage corresponds to the fraction of the query sequence that matches a subject sequence.

See the next page to learn more about the E value.

Copyright © Digital World Biology All rights reserved.

BLASTing Through the Kingdom of Life


Hands-On Tutorial

Legend for links to other resources: [U](#) UniGene [G](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
X16893.1	Tarantula mRNA for hemocyanin subunit a	4057	4057	100%	0.0	100%	
AJ547807.1	Nephila inaurata madagascariensis mRNA for hemocyanin	662	662	79%	0.0	73%	
AJ547811.1	Nephila inaurata madagascariensis mRNA for hemocyanin	312	319	43%	5e-48	88%	
AJ547809.1	Nephila inaurata madagascariensis mRNA for hemocyanin	285	241	14%	7e-43	83%	
AJ302908.1	Cupiennius salei mRNA for hemocyanin subunit 5' (hc-5')	275	298	19%	6e-40	84%	
AJ277492.1	Eurypelma californicum mRNA for hemocyanin subunit g	271	278	22%	8e-39	79%	

The Max score is linked to data that show where the sequences match.

See where the sequences align. 

Copyright © [Digital World Biology](#) All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

Sequences producing significant alignments:

	Score (bits)	E Value
gi 92266 emb X16893.1 ECHEMSUA Tarantula mRNA for hemocyanin...	4183	0.0
gi 28569681 emb AJ547807.1 NIN547807 Nephila inaurata madag...	202	2e-48
gi 15027023 emb AJ307908.1 CSA307908 Cupiennius salei mRNA ...	121	6e-24

Look at more alignments

>gi|28569681|emb|AJ547807.1|NIN547807 Nephila inaurata madagascariensis mRNA (gene)
Length = 2089

Score = 202 bits (102), Expect = 2e-48
Identities = 153/170 (90%)
Strand = Plus / Plus

Query: 1027 cacaactggggacatgttatgatggcttacattcatgatcctgatggcagattcagggaa 1086
Sbjct: 1013 cacaactggggacacgtaaatgatggcctacatccatgatcctgatggcagattcagggaa 1072

Query: 1087 acaccaggtgtcatgactgacacagccacaagtcttagggatccaatcttctacagatac 1146
Sbjct: 1073 actccaggtgtcatgaccgatactgctacaagtcttcgggatcccattttctacgattc 1132

Query: 1147 cacagattcatcgacaacgttttccaagaatacaagaaaactctgccagt 1196
Sbjct: 1133 cacagattcatcgataacgttttccaggaatacaagaaaactctccagt 1182

Score = 111 bits (56), Expect = 2e-48
Identities = 200/248 (80%)
Strand = Plus / Plus

Query: 467 ctggagatgagagcgaca 527
Sbjct: 453 ctggagatgagagcgatg 513

Query: 527 aatacaaaactcgctact 587
Sbjct: 513 aatacaagctggcttact 573

Query: 587 atgttgtttaccctcta 647
Sbjct: 573 acgttgtctaccctctgtatacagattccaattctttggaagaagaaggacaggactg 632

Query: 647 gagagctcttctattacatgcacagatgtgtgccagatagcactgtgagcgttgt 706
Sbjct: 633 gtgaactttctattacatgcacagatgtgtgccagatagcattgcaagattgt 692

Query: 707 ctaatggc 714
Sbjct: 693 ccaatggc 700

blastn did not detect matching sequences in the dashed regions

80% of the nucleotides in this section match the query sequence

90% of these nucleotides match the query

Copyright © Digital World Biology All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

- Find the sequence "NM_000372.4"
- Is your sequence expressed?
- How do you know?
- How is this sequence related to the first alignment in the results?

Legend for links to other resources: [U](#) UniGene [E](#) EBI [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
X16893.1	Tarantula mRNA for hemocyanin subunit a	4052	4057	100%	0.0	100%	
AF04807.1	Nephila inaurata madagascariensis mRNA for hemocyanin	662	662	79%	0.0	73%	
AF04808.1	Nephila inaurata madagascariensis mRNA for hemocyanin	202	319	43%	5e-48	88%	
AF04809.1	Nephila inaurata madagascariensis mRNA for hemocyanin	185	241	14%	7e-43	93%	
AF04808.1	Cupiennius salei mRNA for hemocyanin subunit 5' (hc-5')	175	298	19%	6e-40	84%	
AF04892.1	Eurytelma californicum mRNA for hemocyanin subunit g	171	278	22%	8e-39	79%	

Look at the GenBank record.

Copyright © [Digital World Biology](#) All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

- On which chromosome is the matching target sequence located?

NCBI Nucleotide

Search Nucleotide for [] Go Clear

Accession number [] Limits Preview/Index History Clipboard

Display GenBank Save Text Add to Clipboard

1: X16893.1 Tarantula mRNA for...[gi:9266]

LOCUS ECHEMSUA 2110 bp mRNA INV 12-SEP-1993

DEFINITION Tarantula mRNA for hemocyanin subunit a.

ACCESSION X16893

VERSION X16893.1 GI:9266

KEYWORDS hemocyanin; hemocyanin subunit a

SOURCE Aphonopelma sp.

ORGANISM [Aphonopelma sp.](#) Link to taxonomy database
Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
Mygalomorphae; Theraphosidae; Aphonopelma.

REFERENCE 1 (bases 1 to 2110)

AUTHORS Voit, R.

TITLE Direct Submission

JOURNAL Submitted (12-OCT-1989) Voit R., Zoologisches Institut,
Universitaet Muenchen, Luisenstrasse 14, D-8000 Muenchen 2, FRG

REFERENCE 2 (bases 1 to 2110)

AUTHORS Voit, R. and Feldmaier-Fuchs, G.

TITLE Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs
encoding the tarantula hemocyanin subunits a and e

JOURNAL J. Biol. Chem. 265 (32), 19447-19452 (1990)

MEDLINE [91060544](#)

COMMENT Data kindly reviewed (26-MAR-1990) by Voit R.

FEATURES Location/Qualifiers

source 1..2110
/organism="Aphonopelma sp."
/db_xref="taxon:29932"
/clone="lambda-K1"
/clone_lib="lambda-gt10"
52..1947
/note="hemocyanin subunit a (AA 1-631)"
/codon_start=1

[CDS](#)

in_id="CAA34771.1"
ef="GI:9267"
ef="SWISS-PROT:P14750"
lation="MTILHDKVQALKLFEKLSVAATGPEVFPADQIDERLRNITTLGP
FYPDHLROAKRYVEFECHAAENFDEYSLAKDARSEFNSTLPAFSAEYALHRR

The amino acid sequence for the predicted protein

CDS is short for "coding sequence". The region of DNA between nucleotides 52 and 1947 is predicted to code for the hemocyanin subunit a protein.

Copyright © Digital World Biology All rights reserved.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

The screenshot shows the NCBI Entrez Gene website interface. The browser address bar displays the URL: http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein&cmd=Link&LinkName=protein_gene&from_uid=4507753. The page title is "Entrez Gene". The search bar contains the text "Gene" and "for". The search results show the entry for "1: TYR tyrosinase (oculocutaneous albinism IA) [Homo sapiens]" with GeneID: 7299, updated 10-Apr-2010.

Summary

Official Symbol	TYR
Official Full Name	tyrosinase (oculocutaneous albinism IA)
Primary source	HGNC:12442
See related	Ensembl:ENSG00000077498 ; HPRD:06086 ; MIM:606933
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota ; Metazoa ; Chordata ; Craniata ; Vertebrata ; Euteleostomi ; Mammalia ; Eutheria ; Euarchontoglires ; Primates ; Haplorrhini ; Catarrhini ; Hominidae ; Homo
Also known as	OCA1A; OCAIA; SHEP3; TYR

Summary The enzyme encoded by this gene catalyzes the first 2 steps, and at least 1 subsequent step, in the conversion of tyrosine to melanin. The enzyme has both tyrosine hydroxylase and dopa oxidase catalytic activities, and requires copper for function. Mutations in this gene result in oculocutaneous albinism, and nonpathologic polymorphisms result in skin pigmentation variation. The human genome contains a pseudogene similar to the 3' half of this gene. [provided by RefSeq]

Genomic regions, transcripts, and products

(plus) Go to [reference sequence details](#) [Try our new Sequence Viewer](#)

Genomic context

chromosome: 11; Location: 11q14-q21

The diagram shows the genomic context of the TYR gene on chromosome 11. The gene is located at 11q14-q21. The diagram shows the gene structure with exons represented by red boxes and introns by lines. The gene is labeled TYR. The diagram shows the gene structure with exons represented by red boxes and introns by lines. The gene is labeled TYR.

BLASTing Through the Kingdom of Life

Hands-On Tutorial

The screenshot shows the PubMed website interface. At the top, there is the PubMed logo and the National Library of Medicine (NLM) logo. Below the logos, there are navigation tabs for Nucleotide, Protein, Genome, Structure, and PopSet. A search bar is present with a "Go" button and a "Clear" button. Below the search bar, there are options for "Limits", "Preview/Index", "History", and "Clipboard". A "Display" dropdown menu is set to "Citation", with buttons for "Save", "Text", "Order", and "Add to Clipboard".

The search result is for a paper: **1 : J Biol Chem** 1990 Nov 15;265(32):19447-52. The title is **Arthropod hemocyanins. Molecular cloning and sequencing of cDNAs encoding the tarantula hemocyanin subunits a and e.** The author is **Voit R, Feldmaier-Fuchs G**. The affiliation is **Zoologisches Institut, Universitat Munchen, Federal Republic of Germany.**

The abstract text reads: "cDNA clones comprising the entire coding region of two out of the seven heterogeneous subunits of hemocyanin from the tarantula, *Eurypelma californicum*, were isolated from four cDNA libraries constructed from total RNA from the heart tissue of single spiders. Hybridization was first carried out using a tarantula hemocyanin subunit e partial cDNA, and several positive clones were isolated, including one containing a 2.2-kilobase full-length cDNA (lambda M1). The cDNA comprises an open reading frame for 623 amino acids, 34 nucleotides of the 5'-noncoding region, and 286 nucleotides of the 3'-noncoding region. To select for other hemocyanin subunits, two 17-mer oligonucleotide mixtures, corresponding to the conserved regions in the copper A and copper B oxygen-binding site of chelicerate hemocyanins, were used as probes. Among the positive clones obtained, full-length cDNAs coding for subunit a were identified. The cDNA sequence determined from clone lambda K1 provides an open reading frame coding for 630 amino acids and includes the 5'- and 3'-noncoding regions. Northern blot analysis revealed single transcripts for subunits a and e, each 2.3 kilobases long. The cDNAs for subunits a and e were both found to lack any leader peptide sequence. This supports the idea that the mature protein accumulates in the cytoplasm and is released by cell rupture."

MeSH Terms:

- o Amino Acid Sequence
- o Animal
- o Base Sequence
- o Binding Sites
- o Cloning, Molecular*
- o Codon
- o Copper/metabolism

At the bottom right of the result, there is a button labeled "Back to the beginning" with a green arrow pointing right, and a blue link labeled "Try a BLAST search".

Copyright © [Digital World Biology](#) All rights reserved.