Mini Courses NCBI Resources one course at a time

BLAST Quick Start

Course Developed by Medha Bhagwat
NCBI

Basic Local Alignment Search Tool

Compares protein and nucleic acid sequence(s)

- to those in NCBI protein and nucleotide databases
- to local databases (standalone BLAST)
- to a single protein or nucleotide sequence (BLAST 2)

Basic Local Alignment Search Tool

- local alignments; suitable for detecting isolated regions of similarity
- fast and sensitive
- breaks the query sequence into “words “
- word hits to database sequences extended in both directions

Nucleotide Words

ATGCTGCTAGTCGATGACGTAGCTA

ATGCTGCTAGT
  TGCTGCTAGTC
    GCTGCTAGTCG

Make a lookup table based on the word size

Protein Words

AIEKCYTGCTLAQEADDTA

AIE
IEK LEK, IDK, IQK, IER, IDR, etc
EKC
KCY
CYT

Make a lookup table of neighboring words based on the word size and matrix

Finding the Word Hits

Nucleotide: One exact match

ATGCTGCTAGTCGATGACGTAGCTA
GCTGCTAGTCG

Protein: Two neighboring matches within 40 residues

AIEKCYTGCTLAQEADDTA
IDK EAD

BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

BLAST Programs

		Word Size
blastn	nucleotide X nucleotide	11
blastp	protein X protein	3

6 frame translated nucleotide searches

blastx	nucleotide X protein	3
tblastx	nucleotide X nucleotide	3
tblastn	protein X nucleotide	3

BLAST Programs

	Word Size
BLAST 2 compares 2 sequences	
Mega BLAST	28
- batch nucleotide queries	
- for comparison of sequences that differ slightly	
discontiguous Mega BLAST	11 matches out of 18
- batch nucleotide queries	
- for comparison of diverged sequences	

BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options

Some BLAST Databases

Nucleotide Databases

nr	GenBank+EMBL+DDBJ+PDB sequences (no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS)
refseq_rna	RefSeq RNA sequences
month	new or revised GenBank+EMBL+DDBJ+PDB sequences released in the last 30 days
est	Expressed Sequence Tags (EST division)
gss	Genome Survey Sequence
htgs	Unfinished High Throughput Genomic Sequences: phases 0, 1 and 2 (finished phase 3 sequences in nr)

HTGS

Phase 0	single-few pass reads of a single clone
Phase 1	unfinished, unordered and contain gaps
Phase 2	unfinished, ordered and contain gaps
Phase 3	high quality finished sequences no gaps

Some BLAST Databases

Nucleotide Databases

dbsts	Sequence Tag Site sequences (STS division)
pdb	Protein Data Bank sequences
chromosome	Complete Genomes, Complete Chromosome, or contigs

Major BLAST Databases

Protein Databases

nr	non-redundant GenBank CDS translations+PDB +SwissProt+PIR+PRF
refseq	RefSeq proteins
month	new or revised GenBank CDS translation+PDB +SwissProt+PIR+PRF released in the last 30 days
Swissprot	Last major release of the SWISS-PROT database
pdb	Protein Data Bank protein sequences

BLAST Databases

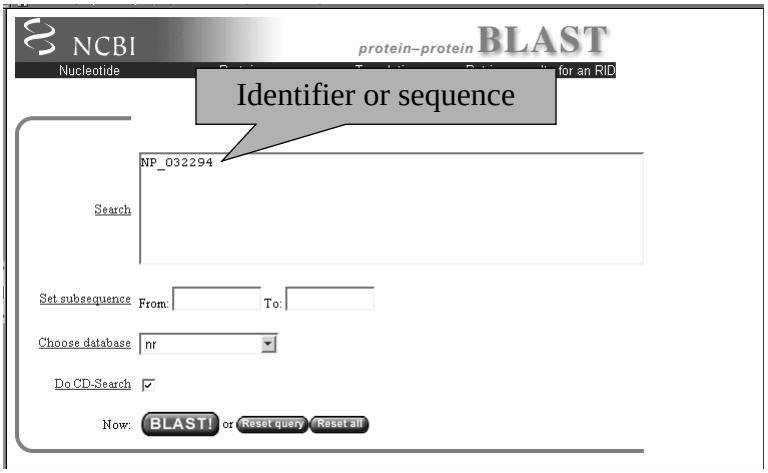
Genome-specific Databases

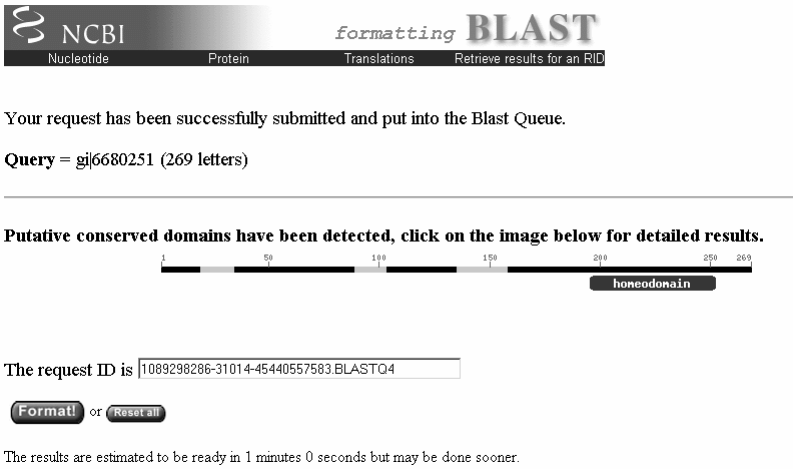
- Human Genome
- Mouse Genome
- Rat Genome
- Other eukaryotes
- Microbial Genomes
- Arabidopsis thaliana
- Insects
- Plants

BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

Setting up a BLAST Search





BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

BLAST Output

NCBI *results of* **BLAST**

BLASTP 2.2.2 [Dec-14-2001]

Reference:
Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

RID: 1013108703-18948-4263

Query= gi|6680251|ref|NP_032294.1| homeo box B5 [Mus musculus]
(269 letters)

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF
865,418 sequences; 272,484,131 total letters

If you have any problems or questions with the results of this search please refer to the **BLAST FAQs**

Taxonomy reports

Taxonomy Reports

- Lineage Report
- Organism Report
- Taxonomy Report
- Help

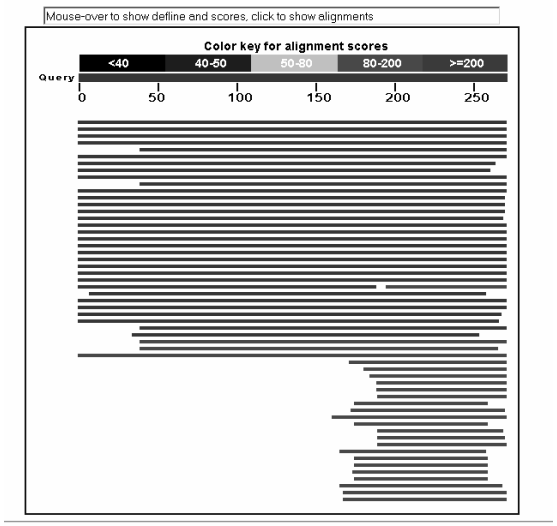
Lineage Report

root
. Bilateria [animals]
. . Coelomata [animals]
. . . Deuterostomia [animals]
. . . . Chordata [chordates]
. . . . . Vertebrata [vertebrates]
. . . . . Gnathostomata [vertebrates]
. . . . . Euteleostomi [vertebrates]
. . . . . Sarcopterygii [vertebrates]
. . . . . Tetrapoda [vertebrates]
. . . . . Amniota [vertebrates]
. . . . . Eutheria [mammals]
. . . . . Euarchontoglires [mammals]
. . . . . Muroidea [rodents]
. . . . . Muridae [rodents]
. . . . . Mus [rodents]

. . . . . Mus musculus (mouse)	361	40 hits	[rodents]	homeo box B5 [Mus musculus] gi|1522035|gb|AAA37842.1| homeo
. . . . . Mus sp. (mice)	128	2 hits	[rodents]	Hox-4 [Mus sp.] >gi|19485660|gb|AAD30705.2| Hox-4 [Mus sp.]
. . . . . Rattus norvegicus (brown rat)	360	10 hits	[rodents]	PREDICTED: similar to Homeobox protein Hox-B5 (Hox-2.1) (HU
. . . . . Rattus sp. (rats)	115	2 hits	[rodents]	Hox-B|Hox-2 [Rattus sp.]
. . . . . Homo sapiens (man)	355	41 hits	[primates]	homeobox B5 [Homo sapiens] >gi|11466635|ref|XP_001173004.1
. . . . . Macaca mulatta (rhesus macaque)	211	10 hits	[primates]	PREDICTED: similar to Homeobox protein Hox-A5 (Hox-1.3) (HU
. . . . . Pan troglodytes	206	11 hits	[primates]	PREDICTED: hypothetical protein, isoform 1 [Pan troglodytes]
. . . . . Canis familiaris (dogs)	258	19 hits	[mammals]	PREDICTED: similar to Homeobox protein Hox-B5 (Hox-2A) (HUG
. . . . . Bos taurus (cow)	313	12 hits	[mammals]	PREDICTED: similar to Homeobox protein Hox-B5 (Hox-2A) (HUG
. . . . . Ovis aries (domestic sheep)	112	3 hits	[mammals]	Homeobox protein Hox-A7 >gi|114859005|gb|AAB04755.1| Hox-7
. . . . . Gallus gallus (bantam)	306	22 hits	[birds]	homeo box B5 [Gallus gallus] >gi|57205402|gb|AAW40409.1| ho
. . . . . Paleudrus sinensis	268	1 hit	[turtles]	Hox5-5 [Paleudrus sinensis]
. . . . . Coturnix japonica	117	1 hit	[birds]	Homeobox protein Hox-A7 (Quox-1) >gi|2135618|gb|AAA49501.1|
. . . . . Xenopus tropicalis	208	1 hits	[amphibians]	homeo box A5 [Xenopus tropicalis] >gi|154759615|gb|AAX86772.
. . . . . Xenopus laevis (clawed frog)	194	20 hits	[amphibians]	Homeobox protein Hox-B5 (XlHbox-4) (Xhox-2B) >gi|12296521|em
. . . . . Ambystoma mexicanum	172	1 hit	[amphibians]	Homeobox protein Hox-A5 >gi|17454218|gb|AAA91634.1| homeobox
. . . . . Notophthalmus viridescens (red-spotted newt)	124	4 hits	[amphibians]	homeobox protein
. . . . . Pleurodeles waltl	117	1 hit	[amphibians]	transcription factor HoxA4 [Pleurodeles waltl]
. . . . . Latimeria menadoensis (king of the sea)	270	20 hits	[coelacanths]	HoxB5 [Latimeria menadoensis]
. . . . . Danio rerio (zebra fish)	268	22 hits	[bony fishes]	homeo box B5a [Danio rerio] >gi|16039240|sp|P09214|HUBB5A BR
. . . . . Oryzias latipes (Japanese rice fish)	255	16 hits	[bony fishes]	hoxB5b [Oryzias latipes] >gi|10016957|dbj|BAC07402.1| hoxB5
. . . . . Sphoeroides nephelus	251	3 hits	[bony fishes]	HoxB5b [Sphoeroides nephelus]
. . . . . Takifugu rubripes (fugufugu)	250	10 hits	[bony fishes]	homeobox protein HoxB5b [Takifugu rubripes]

BLAST Output

Distribution of 514 Blast Hits on the Query Sequence



BLAST Output

Sequences producing significant alignments:	Score (Bits)	E Value	
gi|6680251|ref|NP_032294.1| homeo box B5 [Mus musculus] >gi|5...	361	2e-98	U G
gi|62656894|ref|XP_573183.1| PREDICTED: similar to Homeobox p...	360	5e-98	G
gi|73966238|ref|XP_548176.2| PREDICTED: similar to Homeobox p...	358	2e-97	U G
gi|4504469|ref|NP_002138.1| homeobox B5 [Homo sapiens] >gi|11...	355	1e-96	U G
gi|12848710|dbj|BAB28059.1| unnamed protein product [Mus musculu	321	2e-86	U G
gi|76644496|ref|XP_584774.2| PREDICTED: similar to Homeobox p...	311	3e-83	G
gi|70778859|ref|NP_001020526.1| homeo box B5 [Gallus gallus] ...	305	1e-81	U G
gi|62526079|dbj|BAD95556.1| Hoxb-5 [Gallus gallus]	294	2e-78	U G
gi|28629649|gb|AAO43030.1| HoxB5 [Latimeria menadoensis]	270	6e-71	
gi|54261740|ref|NP_571176.2| homeo box B5a [Danio rerio] >gi|...	268	2e-70	U G
gi|62526089|dbj|BAD95561.1| Hoxb-5 [Pelodiscus sinensis]	268	2e-70	
gi|62537|emb|CAA48320.1| homeodomain protein [Danio rerio] >g...	267	3e-70	G
gi|60392405|sp|P09013|HXB5B_BRARE Homeobox protein Hox-B5b (H...	265	2e-69	
gi|53749656|ref|NP_571612.2| homeo box B5b [Danio rerio] >gi|...	265	2e-69	U G
gi|4322074|gb|AAD15946.1| homeobox protein [Danio rerio]	262	1e-68	G
gi|74267553|dbj|BAE44275.1| hoxB5b [Oryzias latipes] >gi|8301...	255	1e-66	
gi|34484445|gb|AAQ72845.1| Hoxb5b [Sphoeroides nephelus]	251	3e-65	
gi|94482807|gb|ABF22424.1| homeobox protein HoxB5b [Takifugu rub	250	4e-65	
gi|62125375|gb|AAX63757.1| HoxB5bi [Oncorhynchus mykiss]	232	2e-59	
gi|94482797|gb|ABF22415.1| homeobox protein HoxB5a [Takifugu rub	224	3e-57	
gi|47216999|emb|CAG01627.1| unnamed protein product [Tetraodon n	223	6e-57	
gi|60392398|sp|Q6B3N0|HXA5_CHICK Homeobox protein Hox-A5 >gi|...	221	3e-56	
gi|15029697|gb|AAH11063.1| Homeo box A5 [Mus musculus]	214	5e-54	U G
gi|32879829|gb|AAP88745.1| homeo box A5 [synthetic construct]...	212	1e-53	
gi|109472056|ref|XP_001064984.1| PREDICTED: similar to Homeob...	212	2e-53	G
gi|387668|gb|AAA58663.1| HOX 1.3	211	2e-53	G
gi|24497517|ref|NP_061975.2| homeobox A5 [Homo sapiens] >gi|5...	211	2e-53	U G
gi|94482760|gb|ABF22380.1| homeobox protein HoxA5a [Takifugu rub	211	3e-53	

BLAST Statistics

Bit Score
- derived from raw alignment score
- statistical properties of the scoring system taken into account
- can be used to compare different searches

Expect value E
- number of alignments expected to occur by chance

BLAST Output

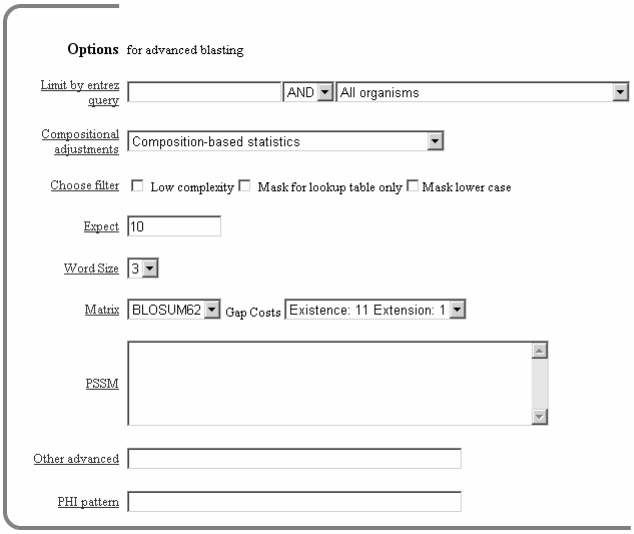
```
> gi|6680251|ref|NP_032294.1|  homeo box B5 [Mus musculus]
 gi|522335|gb|AAA37842.1|  homeobox protein
Length=269

 Score =  361 bits (926),  Expect = 2e-98, Method: Composition-based stats.
 Identities = 269/269 (100%), Positives = 269/26

Query  1    MSSYFVNSFSGRYPNGPDyqllnygsgsslsgsyRD
            MSSYFVNSFSGRYPNGPDYQLLNYGSGSSLSGSYRDPAAMH                  HDLSVNRSS
Sbjct  1    MSSYFVNSFSGRYPNGPDYQLLNYGSGSSLSGSYRDP                  SYGYNYNGMDLSVNRSS  60

Query  61   ASSSHFGAVGESSRAFPASAKEPRFRQAtsscslsspeslpctNGDSHGAKPSASSPSDQ  120
            ASSSHFGAVGESSRAFPASAKEPRFRQATSSCSLSSPESLPCTNGDSHGAKPSASSPSDQ
Sbjct  61   ASSSHFGAVGESSRAFPASAKEPRFRQATSSCSLSSPESLPCTNGDSHGAKPSASSPSDQ  120

Query  121  ATPASSSANFTEIDeasassepeeaasqlsspslaraqPEPMATSTAAPEGQTPQIFPWM  180
            ATPASSSANFTEIDEASASSEPEEAASQLSSPSLARAQPEPMATSTAAPEGQTPQIFPWM
Sbjct  121  ATPASSSANFTEIDEASASSEPEEAASQLSSPSLARAQPEPMATSTAAPEGQTPQIFPWM  180

Query  181  RKLHISHDMTGPDGKRARTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKI  240
            RKLHISHDMTGPDGKRARTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKI
Sbjct  181  RKLHISHDMTGPDGKRARTAYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLSERQIKI  240

Query  241  WFQNRRMKWKKDNKLKSMSLATAGSAFQP  269
            WFQNRRMKWKKDNKLKSMSLATAGSAFQP
Sbjct  241  WFQNRRMKWKKDNKLKSMSLATAGSAFQP  269
```

Low complexity region

BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

Options



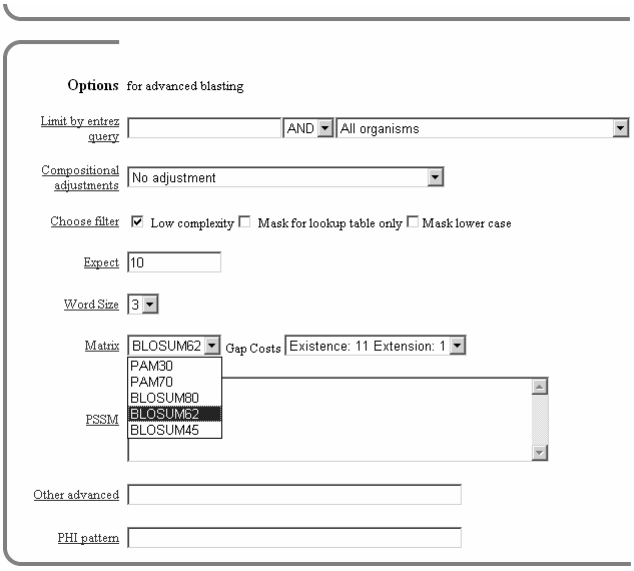
Other Advanced Options

-e expect value
-W word size
-v descriptions
-b alignments

For short query: increase e and lower W

Protein word size 2 or 3
Nucleotide word size minimum 7 and no maximum limit

BLAST Options



Matrices

Nucleotide search

- identity matrix

	A	T	C	G
A	1	-3	-3	-3
T	-3	1	-3	-3
C	-3	-3	1	-3
G	-3	-3	-3	1

Matrices

Protein search
- substitution matrix

Percent Accepted Mutation (PAM)
Blocks Substitution Matrix (BLOSUM)

PAM Matrices

based on global alignments of closely related proteins

PAM1 is calculated from comparisons of sequences with no more than 1% divergence

Other PAM matrices are extrapolated from PAM1

Examples: PAM30 and PAM70

BLOSUM Matrices

based on local alignments

based on observed alignments; not extrapolated from comparisons of closely related proteins

BLOSUM 62 calculated from comparisons of sequences with no more than 62% identity

Examples: BOLSUM45, BLOSUM62 and BLOSUM80

BLOSUM62 is among the best of the available matrices for detecting weak protein similarities

BLOSUM62 is the default matrix in BLAST

Blosum 62

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	X
A	4																				
R	-1	5																			
N	-2	0	6																		
D	-2	-2	1	6																	
C	0	-3	-3	-3	9																
Q	-1	1	0	0	-3	5															
E	-1	0	0	2	-4	2	5														
G	0	-2	0	-1	-3	-2	-2	6													
H	-2	0	1	-1	-3	0	0	-2	8												
I	-1	-3	-3	-3	-1	-3	-3	-4	-3	4											
L	-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4										
K	-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5									
M	-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	-1	5								
F	-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6							
P	-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7						
S	1	-1	1	0	-1	0	0	0	-1	-2	-2	0	-1	-2	-1	4					
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	1	5				
W	-3	-3	-4	-4	-2	-2	-3	-2	-2	-3	-2	-3	-1	1	-4	-3	-2	11			
Y	-2	-2	-2	-3	-2	-1	-2	-3	2	-1	-1	-2	-1	3	-3	-2	-2	2	7		
V	0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	0	-3	-1	4	
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1

frequently observed substitution

rare residue

Substitution Matrices

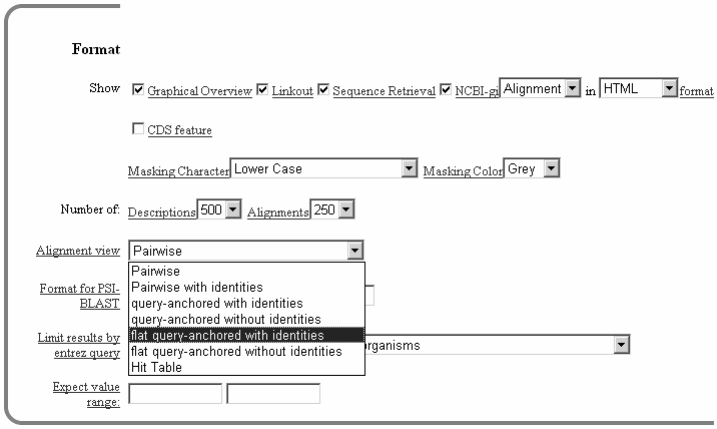
A Provisional Table

Query Length	Substitution Matrix
<35	PAM30
35-50	PAM70
50-85	BLOSUM80
>85	BLOSUM62

BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

Formatting Options



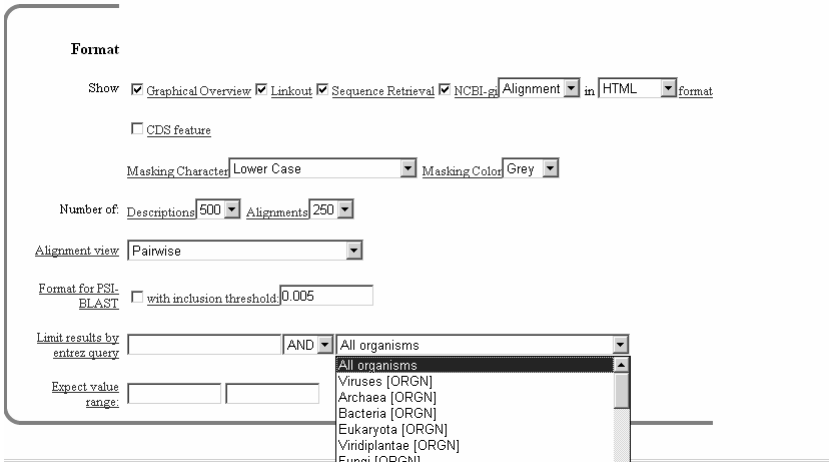
Flat Query-anchored with Identities

```
Query      1    MSSYFVNSFSGRYPNGPDXXXXXXXXXXXXXXXX--RDP--AAMH--TGSYG-YNYNGMD  53
4504469    1    ....................................--...--....--.....-.......  53
12848710   1                                                ..--.....-.......  14
28629649   1                                                ..--SS...-.......  14
26984637   1    ........................TS..AMNASY-..S--GT..--S....-.......  54
22316136   1    .....L..........S........TNGAMNA..--..S--TS..--S..F.-.....I.  53
18858837   1    ........................TS..AMNASY-..S--GT..--S....-.......  54
24119211   1    .....L..........S........TNGAMNA..--..S--TS..--S..F.-N....I.  53
34484445   1    ................S........TNGAV...F--...--GT..--S..F.-.......  53
47216999   1    ........................AS.GALNGGTY..SST.T..HA.....-.T.....  59
15029697   1    .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  53
32879829   1    .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  53
34855875   112  .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  164
14916595   1    .........C...........H...DH..V.EQ.--..S--.T..--SSR..-.G.....  53
24497517   1    .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  53
123225     1    .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  53
14916601   1    .........C......A.F..H...DH.TANDQ.--..S--T...--SSR..-.G.....  53
6754232    1    .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  53
49457548   1    .........C...........H...DH..V.EQF--..S--.S..--S.R..-.G.....  53
123297     18                                            .Q--P.A..S...T...  32
47229432   1                                             ..--SSR..-.G.....  14
14916584   1    .....L.......Q..H....QS..NAA.VGEHF--.ES--TMLS--.-R..-.D.T.I.  52
24119205   1                                             ..--S.R.A-CG.....  14
1708345    14                                       ..S--.S..--S.R..-.G.....  32
```

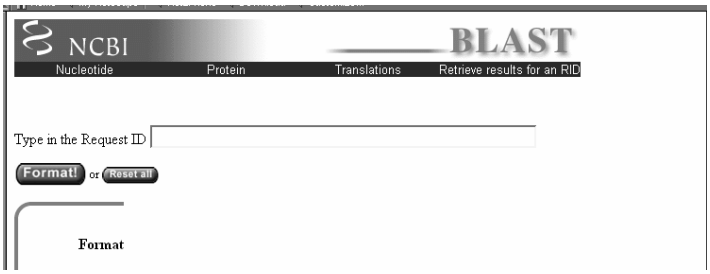
HitTable

```
# BLASTP 2.2.2 [Dec-14-2001]
# Database: nr
# Query: gi|6680251|ref|NP_032294.1| homeo box B5 [Mus musculus]
# Fields: Query id, Subject id, % identity, alignment length, mismatches, gap openings, q. start, q. end, s. start, s. end,
gi|6680251|ref|NP_032294.1|     gi|6680251|ref|NP_032294.1|     100.00  269     0       0       1       269     1       269
gi|6680251|ref|NP_032294.1|     gi|4504469|ref|NP_002138.1|     98.51   269     4       0       1       269     1       269
gi|6680251|ref|NP_032294.1|     gi|12848710|dbj|BAB28059.1|     99.57   230     1       0       40      269     1       230
gi|6680251|ref|NP_032294.1|     gi|123245|sp|P09014|HXB5_BRARE  70.50   278     70      6       1       269     1       275
gi|6680251|ref|NP_032294.1|     gi|4322074|gb|AAD15946.1|       65.00   280     81      5       1       268     1       275
gi|6680251|ref|NP_032294.1|     gi|15029697|gb|AAH11063.1|AAH11063      56.25   288     89      9       1       269     1
gi|6680251|ref|NP_032294.1|     gi|9506791|ref|NP_061975.1|     54.70   287     95      7       1       269     1       270
gi|6680251|ref|NP_032294.1|     gi|14751311|ref|XP_040801.1|    54.90   286     96      6       1       269     1       270
gi|6680251|ref|NP_032294.1|     gi|14916595|sp|Q9IA23|HXA5_HETFR        51.38   290     105     4       1       269     1
gi|6680251|ref|NP_032294.1|     gi|14916601|sp|Q9PWD3|HXA5_MORSA        53.07   277     115     4       1       264     1
gi|6680251|ref|NP_032294.1|     gi|6754232|ref|NP_034583.1|     55.40   287     93      8       1       269     1       270
gi|6680251|ref|NP_032294.1|     gi|123297|sp|P09019|HXB5_XENLA  58.87   231     76      6       40      269     18      230
gi|6680251|ref|NP_032294.1|     gi|14916584|sp|Q9IA11|HXD5_HETFR        45.93   270     127     6       1       269     1
gi|6680251|ref|NP_032294.1|     gi|1708344|sp|P50208|HXA5_AMBME 89.00   100     10      1       171     269     49      148
gi|6680251|ref|NP_032294.1|     gi|4322062|gb|AAD15940.1|       51.08   231     108     4       40      269     1       227
gi|6680251|ref|NP_032294.1|     gi|5441871|dbj|BAAB2371.1|      95.35   86      4       0       184     269     1       86
gi|6680251|ref|NP_032294.1|     gi|87743|pir||A24777    100.00  82      0       0       188     269     37      118     6.5
gi|6680251|ref|NP_032294.1|     gi|123246|sp|P14838|HXB5_CHICK  98.78   82      1       0       188     269     1       82
gi|6680251|ref|NP_032294.1|     gi|1708345|sp|P52949|HXA5_RAT   50.00   236     83      7       35      252     14      232
gi|6680251|ref|NP_032294.1|     gi|62528|emb|CAA30293.1|        100.00  81      0       0       189     269     1       81
gi|6680251|ref|NP_032294.1|     gi|89868|pir||JT0273    97.56   82      2       0       188     269     37      118     7.5
gi|6680251|ref|NP_032294.1|     gi|90629|pir||B24777    100.00  82      0       0       188     269     37      118     7.5
gi|6680251|ref|NP_032294.1|     gi|6252867|dbj|BAA86245.1|      97.47   79      2       0       189     267     1       79
gi|6680251|ref|NP_032294.1|     gi|4322038|gb|AAD15929.1|       77.78   99      20      2       172     268     204     302
gi|6680251|ref|NP_032294.1|     gi|1334621|emb|CAA30294.1|      100.00  76      0       0       194     269     1       76
```

Limiting Options



Retrieve Results for an RID



BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

Examples

Protein searches more sensitive than nucleotide searches

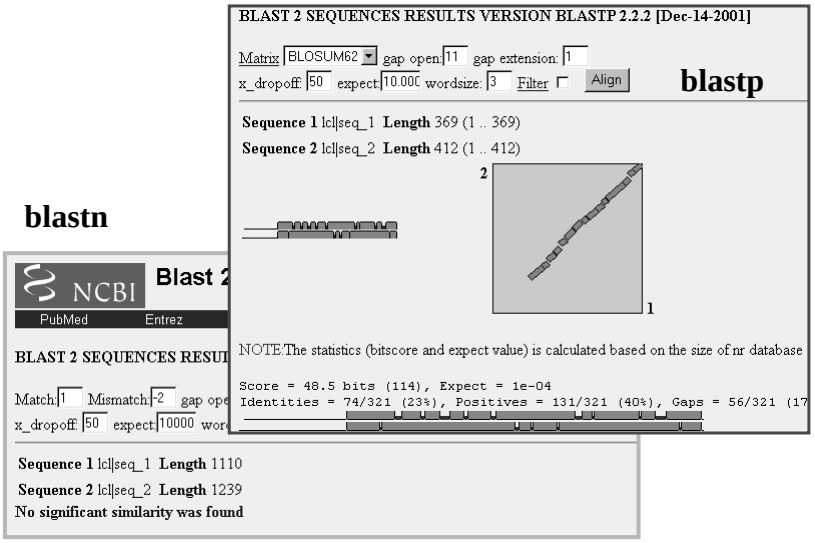
- redundancy of the genetic code allows the same information to be encoded differently

blastn best to search for the same sequence from the same organism

blastn and blastp

YadC, a quickly evolving gene from K12 and O157

Compare nucleotide and protein sequences using BLAST 2



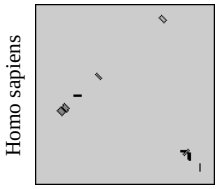
blastn and tblastx

Compare using BLAST 2

Anopheles gambiae mitochondrion, complete genome.
ACCESSION NC_002084 GI:5834911

Homo sapiens mitochondrion, complete genome.
ACCESSION NC_001807 GI:17981852

blastn



tblastx

BLAST Basics

- Programs
- Databases
- Setting up a search
- Interpret the results
- BLAST options
- Formatting options
- Examples

Hands-On

http://www.ncbi.nlm.nih.gov/Class/minicourses/quickblast.html

BLAST Statistics

Bit Score S'

$$\frac{\lambda S - \ln K}{\ln 2}$$

Expect value E

$$mn\ 2^{-S'}$$

BLOSUM 80	BLOSUM 62	BLOSUM 45
PAM 1	PAM 120	PAM 250

Less divergent ←——→ *More divergent*