Mini Courses NCBI Resources one course at a time

NCBI Minicourses

BLAST Quick Start

blast-help@ncbi.nlm.nih.gov

NCBI Minicourses

Algorithm Basics

- Introduction
- Words & extensions

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

NCBI Minicourses

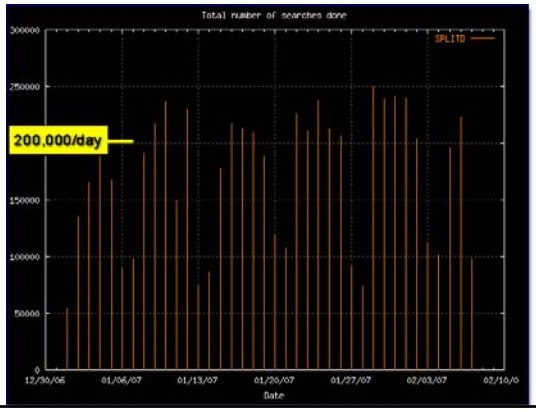
Basic Local Alignment Search Tool

Compare protein or nucleic acid sequences to protein or nucleic acid databases

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

(BLAST 2 Sequences, or pairwise BLAST)

BLAST Web Searches, 2007



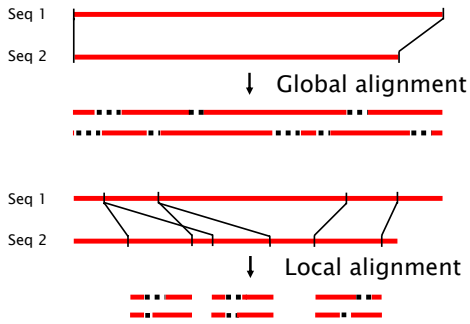
NCBI Minicourses

Basic Local Alignment Search Tool

- local alignments; isolated regions of similarity
- fast and sensitive
- breaks the query sequence into “words”
- word matches to database sequences are extended in both directions

NCBI Minicourses

Global vs Local Alignment



NCBI Minicourses

Global vs Local Alignment

```
Seq1:  WHEREISWALTERNOW      (16aa)
Seq2:  HEWASHEREBUTNOWISHERE (21aa)
```

Global

```
Seq1:  1                W--HEREISWALTERNOW 16
                        W   HERE
Seq2:  1  HEWASHEREBUTNOWISHERE            21
```

Local

```
Seq1: 1  W--HERE 5        Seq1:  1 W--HERE  5
         W   HERE                  W   HERE
Seq2: 3  WASHERE 9        Seq2: 15 WISHERE 21
```

NCBI Minicourses

How BLAST Works

1. Make lookup table of "words" for query
2. Scan database for hits
3. Extend alignment both directions
   - Ungapped extensions of hits (initial HSPs)
   - Gapped extensions (no traceback)
   - Gapped extensions (traceback – alignment details)

NCBI Minicourses

Nucleotide Words

Make a lookup table based on the word size.

11-mer

ATGCTGCTAGTCGATGACGTAGCTA

ATGCTGCTAGT
 TGCTGCTAGTC
  GCTGCTAGTCG
   . . .

NCBI Minicourses

Protein Words

AIEKCYTGCTLAQEADDTA

AIE
IEK
EKC
KCY
CYT
…

LEK, IDK, IQK, IER, IDR, etc

Neighborhood words

Lookup table, including neighborhood words, is based on word size, score matrix, and threshold.

NCBI Minicourses

Scoring Systems - Proteins (BLOSUM62)

	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

E / E

I / I

L / I

K / K

IEK = 14

LEK = 12

IEK: keep LEK? (threshold = 11)

NCBI Minicourses

Word Hits & Extensions

Nucleotide: one exact match

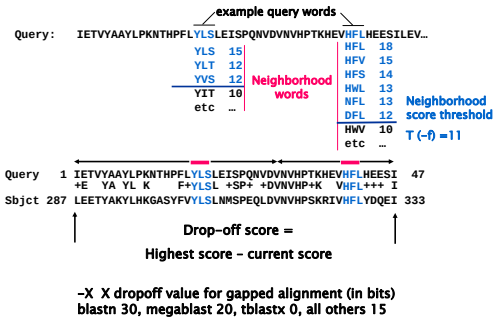
⇦ GCTGCTAGTCG ⇨

Protein: two matches within 40 residues

⇦IDK ⇨ ⇦ EAD ⇨

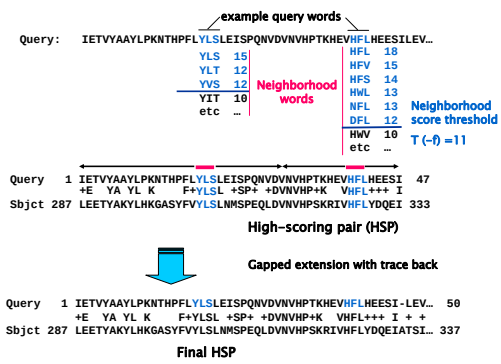
NCBI Minicourses

BLASTP Summary



NCBI Minicourses

BLASTP Summary



NCBI Minicourses

BLAST

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

NCBI Minicourses

BLAST Programs

What is your goal?

blastn nucleotide X nucleotide
blastp protein X protein

6 frame, translated nucleotide searches

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

NCBI Minicourses

Word Size

Trade-off: sensitivity vs speed

NCBI Minicourses

More BLAST Programs

	Word Size
MegaBLAST - batch nucleotide queries - very similar sequences	28
Discontiguous MegaBLAST - batch nucleotide queries - divergent sequences	11 matches out of 18

NCBI Minicourses

Templates for Discontiguous Words

NCBI Minicourses

```
W = 11, t = 16, coding:        110110110110110 1
W = 11, t = 16, non-coding:    111001011011011 1
W = 12, t = 16, coding:        111110110110110 1
W = 12, t = 16, non-coding:    111011011011011 1
W = 11, t = 18, coding:        1011011001011011 01
W = 11, t = 18, non-coding:    1110100101100101 11
W = 12, t = 18, coding:        1011011011011011 01
W = 12, t = 18, non-coding:    1110101100101101 11
W = 11, t = 21, coding:        100101100101100101101
W = 11, t = 21, non-coding:    111010010100010010111
W = 12, t = 21, coding:        100101101101100101101
W = 12, t = 21, non-coding:    111010010110010010111
```

W = word size; # matches in template
t = template length

Reference: Ma, B, Tromp, J, Li, M. PatternHunter: faster and more sensitive homology search. Bioinformatics March, 2002; 18(3):440-5

BLAST

Search Basics

NCBI Minicourses

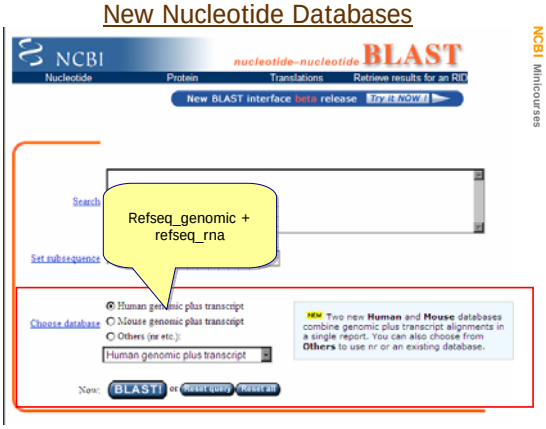
- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

Nucleotide BLAST Databases

NCBI Minicourses

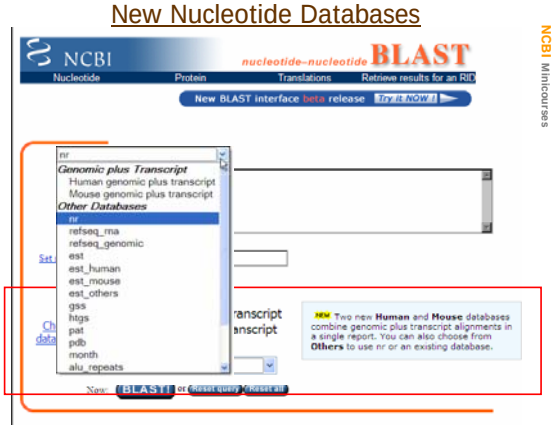
- nr (nt)
  - Traditional GenBank Divisions
  - NM_ and XM_ RefSeqs
- refseq_rna
  - NM_ , XM_ , NR_
- refseq_genomic
  - NC_ , NT_ , NG_
- est
  - EST Division
- htgs
  - HTG division
- dbsts
  - STS Division
- chromosome
  - NC genomic records
- gss
  - GSS division
- pat
  - PAT Division
- wgs
  - wgs entries from traditional divisions
- pdb
  - Nucleotide sequences from structures
- env_nt
  - environmental samples

New Nucleotide Databases



NCBI Minicourses

New Nucleotide Databases



NCBI Minicourses

Protein BLAST Databases

Protein

- nr
  traditional GenBank records
- refseq = NP_, XP_
- swissprot
- pdb
- pat
- env_nr

nr ≠ nr

NCBI Minicourses

BLAST Databases: Genome-specific

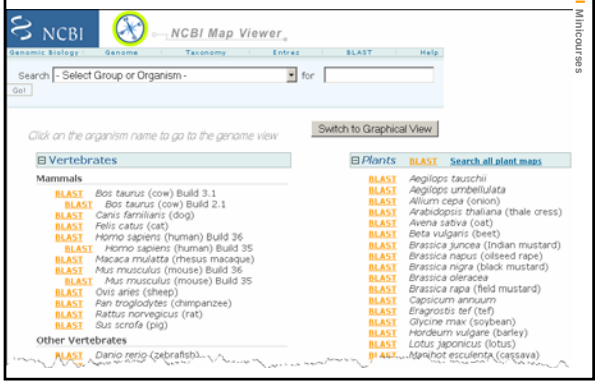
NCBI Minicourses

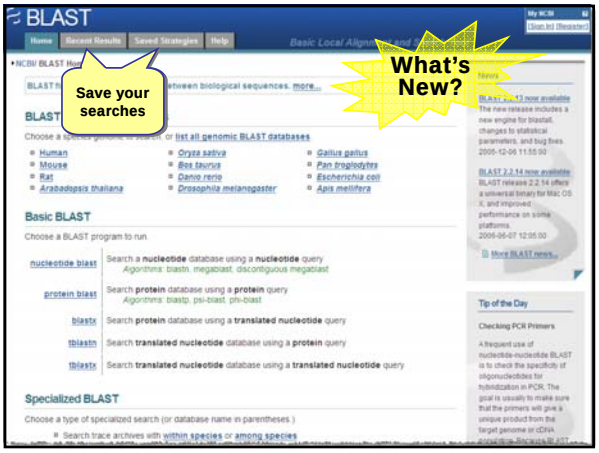
Genomes

- Human, mouse, rat, chimp, cow, pig, dog, sheep, cat
- Chicken, puffer fish, zebrafish
- Fly, honey bee, other insects
- Microbes, environmental samples
- Plants, nematodes
- Fungi, protozoa, other eukaryotes

Genome-specific: Map Viewer

NCBI Minicourses





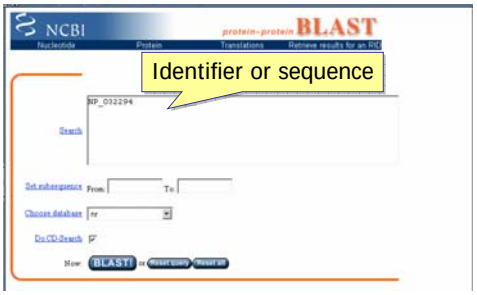
BLAST

Search Basics

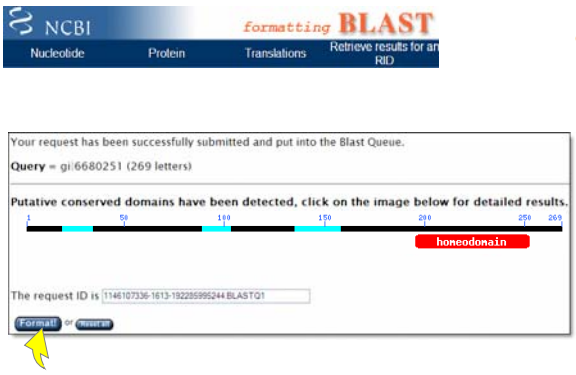
- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

NCBI Minicourses

Setting up a BLAST Search



NCBI Minicourses



BLAST Output

NCBI results of BLAST

BLASTP 2.2.13 [Nov-27-2005]

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: 1146107336-1613-192285995244.BLASTQ1

Database: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples
3,605,113 sequences; 1,238,276,235 total letters

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

Tax BLAST Report

Index

- Lineage Report
- Organism Report
- Taxonomy Report
- Help

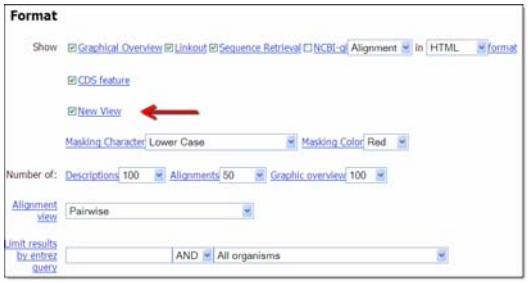
Lineage Report

```
root
- Bilateria          [animals]
. - Deuterostomia     [animals]
. . - Chordata        [chordates]
. . . - Vertebrata    [vertebrates]
. . . . - Gnathostomata   [vertebrates]
. . . . . - Euteleostomi  [vertebrates]
. . . . . . - Sarcopterygii   [vertebrates]
. . . . . . . - Tetrapoda     [vertebrates]
. . . . . . . . - Amniota       [vertebrates]
. . . . . . . . . - Eutheria      [mammals]
. . . . . . . . . . - Eutheria-Rodentia [mammals]
. . . . . . . . . . . - Muroidea        [mammals]
. . . . . . . . . . . . - Mus           [mammals]
. . . . . . . . . . . . . - Mus musculus (mouse) ---------- 419 46 hits [mammals]
. . . . . . . . . . . . . - Mus sp. (mice) ..... 148 4 hits [mammals]
. . . . . . . . . . . . . - Rattus norvegicus (brown rat) -------- 195 11 hits [mammals]
. . . . . . . . . . . . . - Rattus sp. (rats) ..... 117 2 hits [mammals]
. . . . . . . . . . . . - Macaca fascicularis (cynomolgus monkey) - 416 1 hit [mammals]
. . . . . . . . . . . . - Homo sapiens (man) ..... 412 43 hits [mammals]
. . . . . . . . . . . . - Pan troglodytes ..... 240 9 hits [mammals]
. . . . . . . . . . . . - Bos taurus (cow) ..... 365 17 hits [mammals]
. . . . . . . . . . . . - Canis familiaris (dogs) ..... 257 15 hits [mammals]
. . . . . . . . . . . . - Ovis aries (domestic sheep) ..... 124 2 hits [mammals]
. . . . . . . . . . - Gallus gallus (chickens) ---------- 353 23 hits [birds]
. . . . . . . . . . - Pelodiscus sinensis ..... 316 2 hits [turtles]
. . . . . . . . . . - Coturnix japonica ..... 132 1 hit [birds]
. . . . . - Xenopus tropicalis ---------- 249 3 hits [amphibians]
. . . . . - Xenopus laevis (clawed frog) ..... 221 18 hits [amphibians]
. . . . . - Ambystoma mexicanum ..... 185 2 hits [amphibians]
. . . . . - Pleurodeles waltl ..... 125 1 hit [amphibians]
. . . . . - Notophthalmus viridescens (red-spotted newt) .. 125 4 hits [amphibians]
. . . . . - Latimeria menadoensis (king of the sea) -------- 312 10 hits [coelacanths]
```

Tax BLAST Report

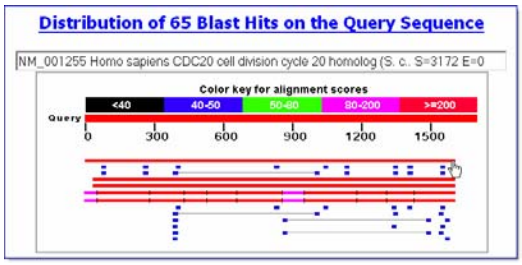
```
. . . . . - Mus musculus (mouse) ---------- 419 46 hits [mammals]
. . . . . - Mus sp. (mice) ..... 148 4 hits [mammals]
. . . . . - Rattus norvegicus (brown rat) -------- 195 11 hits [mammals]
. . . . . - Rattus sp. (rats) ..... 117 2 hits [mammals]
. . . . - Macaca fascicularis (cynomolgus monkey) - 416 1 hit [mammals]
. . . . - Homo sapiens (man) ..... 412 43 hits [mammals]
. . . . - Pan troglodytes ..... 240 9 hits [mammals]
. . . - Bos taurus (cow) ---------- 365 17 hits [mammals]
. . . - Canis familiaris (dogs) ..... 257 15 hits [mammals]
. . . - Ovis aries (domestic sheep) ..... 124 2 hits [mammals]
. . - Gallus gallus (chickens) ---------- 353 23 hits [birds]
. . - Pelodiscus sinensis ..... 316 2 hits [turtles]
. . - Coturnix japonica ..... 132 1 hit [birds]
. - Xenopus tropicalis ---------- 249 3 hits [amphibians]
. - Xenopus laevis (clawed frog) ..... 221 18 hits [amphibians]
. - Ambystoma mexicanum ..... 185 2 hits [amphibians]
. - Pleurodeles waltl ..... 125 1 hit [amphibians]
. - Notophthalmus viridescens (red-spotted newt) .. 125 4 hits [amphibians]
Latimeria menadoensis (king of the sea) -------- 312 10 hits [coelacanths]
```

New Output View



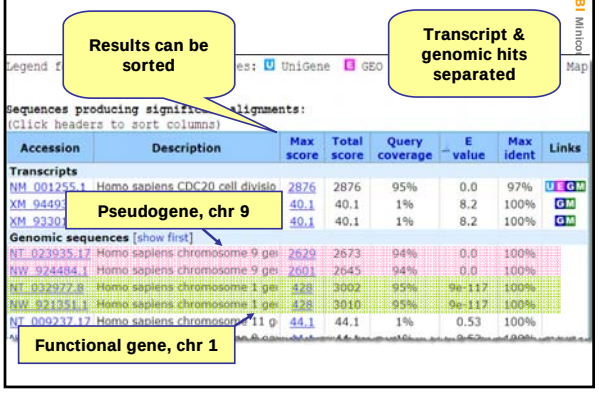
NCBI Minicourses

Graphic Overview



NCBI Minicourses

New Output View



NCBI Minicourses

Sorting Results

NCBI Minicourses

Resorted by Total score

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

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell divisi	2876	2876	95%	0.0	97%	U E G M
XM_933013.2	PREDICTED: Homo sapiens chrc	40.1	40.1	1%	8.2	100%	G M
XM_944935.2	PREDICTED: Homo sapiens chrc	40.1	40.1	1%	8.2	100%	G M
Genomic sequences [show first]							
NW_921351.1	Homo sapiens chromosome 1 ge	428	3010	95%	9e-117	100%	
NT_032977.8	Homo sapiens chromosome 1 ge	428	3002	95%	9e-117	100%	
NT_023935.17	Homo sapiens chromosome 9 ge	2629	2673	94%	0.0	100%	
NW_924484.1	Homo sapiens chromosome 9 ge	2601	2645	94%	0.0	100%	
NW_921807.1	Homo sapiens chromosome 3 ge	44.1	44.1	1%	0.53	100%	

Functional gene now first

Sorting Hits: by Score

NCBI Minic

ref|NT_032977.8|Hs1_33153 Homo sapiens chromosome 1 genomic contig, reference assembly
Length=73835825

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features in this part of subject sequence:
cell division cycle 20

Score = 428 bits (216), Expect = 9e-117
Identities = 231/236 (97%), Gaps = 0/236 (0%)
Strand=Plus/Plus

Query 965 CTCCAGTGGTTCACGTTCTGGCCACATCCACCACCATGATGTTCGGGTAGCAGAACACCA 1024
Sbjct 13798316 CTCCAGTGGTTCACGTTCTGGCCACATCCACCACCATGATGTTCGGGTAGCAGAACACCA 13798375

Query 1028 TGTGGCTACACTGAGTGGCCACAGCCAGGAAGTGTGTGGGCTGCGCTGGGCCCAGATGG 1086

Sort by Score: longest exon usually first

Sorting Hits: by Query Start

NCBI Minicou

ref|NT_032977.8|Hs1_33153 Homo sapiens chromosome 1 genomic contig, reference assembly
Length=73835825

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Features flanking this part of subject sequence:
6149 bp at 5' side: myeloproliferative leukemia virus oncogene
223 bp at 3' side: cell division cycle 20

Score = 89.7 bits (45), Expect = 1e-14
Identities = 51/53 (96%), Gaps = 0/53 (0%)
Strand=Plus/Plus

Query 1 AGCGGAGAGTTTAAGAGGCGTAAGCGAGGCGTGTTAAACCCGGTCGGAACTGC 53
Sbjct 13796530 AGCGGAGAGTTTAAGAGGCGTAAGCCAGGCGTGTTAAAGCCGGTCGGAACTGC 13796582

Sort by Query start: Proper exon order

BLAST

Search Basics

NCBI Minicourses

- Programs
- Databases
- Submit a search
- Interpret the results ...con't
- BLAST options
- Format options
- Examples

Where do Scores Come From?

NCBI Minicourses

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

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell divisi	2876	2876	95%	0.0	97%	U E G M
XM_933013.2	PREDICTED: Homo sapiens chrc	40.1	40.1	1%	8.2	100%	G M
XM_944935.2	PREDICTED: Homo sapiens chrc	40.1	40.1	1%	8.2	100%	G M
Genomic sequences [show first]							
NW_921351.1	Homo sapiens chromosome 1 ge	428	3010	95%	9e-117	100%	
NT_032977.8	Homo sapiens chromosome 1 ge	428	3002	95%	9e-117	100%	
NT_023935.17	Homo sapiens chromosome 9 ge	2629	2673	94%	0.0	100%	
NW_924484.1	Homo sapiens chromosome 9 ge	2601	2645	94%	0.0	100%	
NW_921807.1	Homo sapiens chromosome 3 ge	44.1	44.1	1%	0.53	100%	

Scoring Matrices

NCBI Minicourses

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

```
CAGGTAGCAAGCTTGCATGTCA
|| |||||||||||  |||||     raw score = 19-9* = 10*
CACGTAGCAAGCTTG-GTGTCA
```

* ignores gap costs

Scoring Matrices

Protein search: substitution matrix

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

NCBI Minicourses

PAM Matrices

- global alignments of closely related proteins
- PAM1 calculated from sequences with no more than 1% divergence
- Other PAM matrices are extrapolated from PAM1
- Examples: PAM30 and PAM70

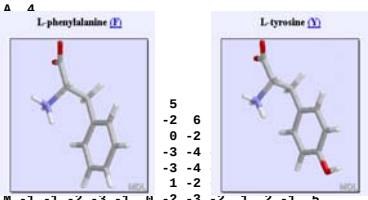
NCBI Minicourses

BLOSUM Matrices

- local alignments
- all matrices based on observed alignments; not extrapolated
- BLOSUM 62 calculated from sequences with no more than 62% identity
- Examples: BLOSUM45, BLOSUM62 and BLOSUM80
- BLOSUM62: very good for detecting weak protein similarities
- BLOSUM62 is the default matrix in BLAST

NCBI Minicourses

BLOSUM62




```
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
   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  X
```

Negative for less likely substitutions

Positive for more likely substitutions

NCBI Minicourses

BLOSUM62

```
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
   A  R  N  D  C  Q  E  G  H  I  L  K  M  F  P  S  T  W  Y  V  X
```

J (leucine or isoleucine) and O (pyrrolysine)

NCBI Minicourses

Substitution Matrices

A Provisional Table

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

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

Less divergent ← → *More divergent*

NCBI Minicourses

Where do Expect Values Come From?

NCBI Minicourses

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

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell divisi	2876	2876	95%	0.0	97%	U E G M
XM_933013.2	PREDICTED: Homo sapiens chrc	40.1	40.1	1%	8.2	100%	G M
XM_944935.2	PREDICTED: Homo sapiens chrc	40.1	40.1	1%	8.2	100%	G M
Genomic sequences [show first]							
NW_921351.1	Homo sapiens chromosome 1 ge	428	3010	95%	9e-117	100%	
NT_032977.8	Homo sapiens chromosome 1 ge	428	3002	95%	9e-117	100%	
NT_025355.12	Homo sapiens chromosome 9 ge	2629	2673	94%	0.0	100%	
NW_924484.1	Homo sapiens chromosome 9 ge	2601	2645	94%	0.0	100%	
NW_921807.1	Homo sapiens chromosome 3 ge	44.1	44.1	1%	0.53	100%	

NCBI Minicourses

Expect Value

E = number of database hits you expect to find by chance, ≥ S

$E = Kmne^{-\lambda S}$ or $E = mn2^{-S'}$

K = scale for search space
λ = scale for scoring system
λS' = bitscore = (λS – lnK)/ln2
m = query length
n = database length

E is dependent on m x n (search space)

More info: The Statistics of Sequence Similarity Scores

E is dependent on m x n (search space)

NCBI Minicourses

Short query =

low score

high Expect

```
Score = 30.2 bits (15),  Expect =    19
Identities = 15/15 (100%), Gaps = 0/15 (0%)
Strand=Plus/Plus

Query  4         GTAGCAAGCTTGCAT  18
                 ||||||||||||||
Sbjct  40347983  GTAGCAAGCTTGCAT  40347997
```

```
CAGGTAGCAAGCTTGCATGTCA
|| |||||||||||  |||||     raw score = 19-9 = 10
CACGTAGCAAGCTTG-GTGTCA
```

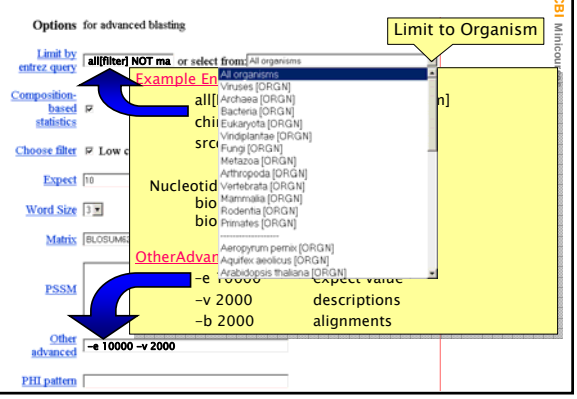
More info: The Statistics of Sequence Similarity Scores

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

NCBI Minicourses

Advanced Options



NCBI Minicourses

Other Advanced Options

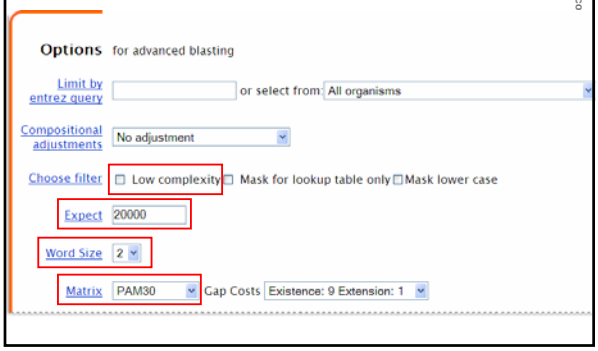
For short queries: increase e and lower W

Other advanced | **– e 10000 –W 7 –v 2000 –b 1500**

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

NCBI Minicourses

BLASTP – "…short, nearly exact matches"



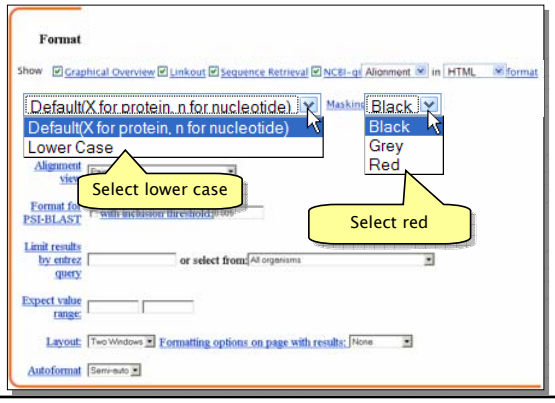
NCBI Minico

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

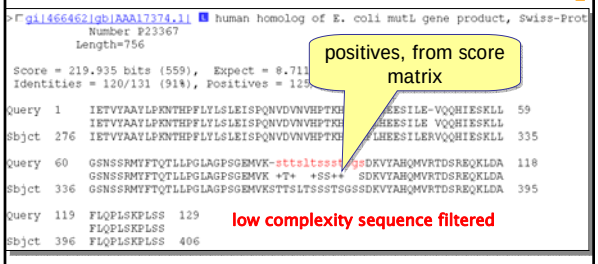
NCBI Minicourses

'New' Formatter

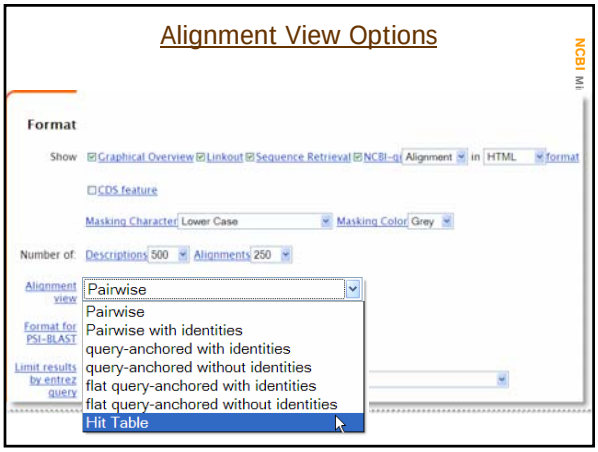


NCBI Minicourses

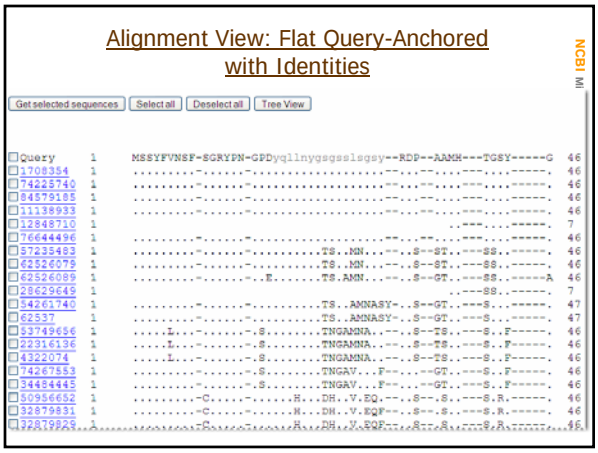
BLAST Output: Alignments & Filter



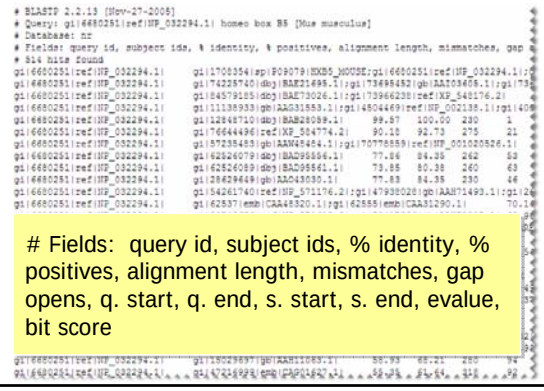
Alignment View Options



Alignment View: Flat Query-Anchored with Identities



Alignment View: Hit Table



Fields: query id, subject ids, % identity, % positives, alignment length, mismatches, gap opens, q. start, q. end, s. start, s. end, evalue, bit score

NCBI Minicourses

BLAST

Search Basics

- Programs
- Databases
- Submit a search
- Interpret the results
- BLAST options
- Format options
- Examples

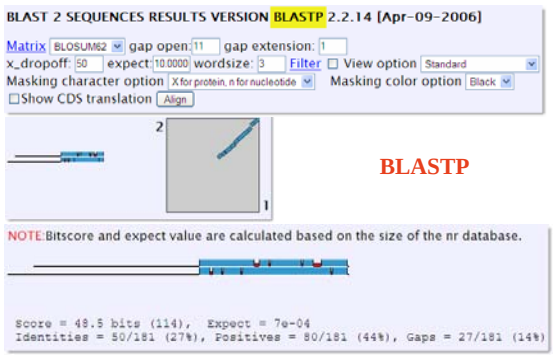
NCBI Minicourses

Examples

- Protein searches more sensitive than nucleotide searches
  - redundancy of the genetic code
- blastn, megablast best when searching within same organism

NCBI Minicourses

BLASTN vs BLASTP



BLASTP

BLAST is a shortcut . . .

An alignment BLAST cannot make:

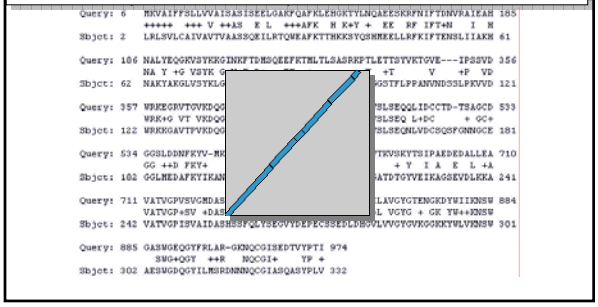
```
  1 GAATATATGAAGACCAAGATTGCAGTCCTGCTGGCCTGAACCACGCTATTCTTGCTGTTG
    || | || || || |  || || ||   ||  |  ||| |||||| | | || | ||| |
  1 GAGTGTACGATGAGCCCGAGTGTAGCAGTGAAGATCTGGACCACGGTGTACTCGTTGTCG

 61 GTTACGGAACCGAGAATGGTAAAGACTACTGGATCATTAAGAACTCCTGGGGAGCCAGTT
    | || ||     ||  ||| ||  | |||||| || | ||||||  |||||  |     |
 61 GCTATGGTGTTAAGGGTGGGAAGAAGTACTGGCTCGTCAAGAACAGCTGGGCTGAATCCT

121 GGGGTGAACAAGGTTATTTCAGGCTTGCTCGTGGTAAAAAC
    |||| || ||||| ||  ||    |  | ||||  || |||
121 GGGGAGACCAAGGCTACATCCTTATGTCCCGTGACAACAAC
```

Reason:

no contiguous exact match of 7 bp.

An Alignment BLAST Can Make

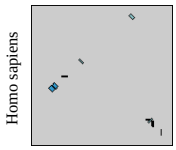
```
Score = 290 bits (741), Expect = 7e-77
Identities = 147/331 (44%), Positives = 206/331 (61%), Gaps = 8/331 (2%)
Frame = +3
```

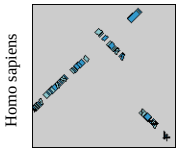


BLASTN vs TBLASTX

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

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





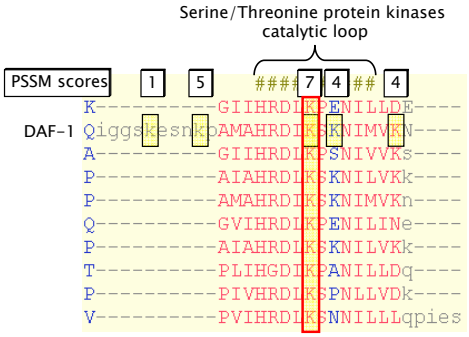
NCBI Minicourses

Hands-On

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

NCBI Minicourses

Position-Specific Score Matrix



NCBI Minicourses

Position-Specific Score Matrix

NCBI | Minicourses

catalytic loop

		A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
435	K	-1	0	0	-1	-2	3	0	3	0	-2	-2	1	-1	-1	-1	-1	-1	-1	-1	-2
436	E	0	1	0	2	-1	0	2	-1	0	-1	-1	0	0	0	-1	0	0	-1	-1	-1
437	S	0	0	-1	0	1	1	0	1	1	0	-1	0	0	0	2	0	-1	-1	0	-1
438	N	-1	0	-1	-1	1	0	-1	3	3	-1	-1	1	-1	0	0	-1	-1	1	1	-1
439	K	-2	1	1	-1	-2	0	-1	-2	-2	-1	-2	5	1	-2	-2	-1	-1	-2	-2	-1
440	P	-2	-2	-2	-2	-3	-2	-2	-2	-2	-1	-2	-1	0	-3	7	-1	-2	-3	-1	-1
441	A	3	-2	1	-2	0	-1	0	1	-2	-2	-2	0	-1	-2	3	1	0	-3	-3	0
442	M	-3	-4	-4	-4	-3	-4	-4	-5	-4	7	0	-4	1	0	-4	-4	-2	-4	-1	2
443	A	4	-4	-4	-4	0	-4	-4	-3	-4	4	-1	-4	-2	-3	-4	-1	-2	-4	-3	4
444	H	-4	-2	-1	-3	-5	-2	-2	-4	10	-6	-5	-3	-4	-3	-2	-3	-4	-5	0	-5
445	R	-4	8	-3	-4	0	-1	-2	-3	-2	-5	-4	0	-3	-2	-4	-3	-3	0	-4	-5
446	D	-4	-4	-1	8	-6	-2	0	-3	-3	-5	-6	-3	-5	-6	-4	-2	-3	-7	-5	-5
447	I	-4	-5	-6	-6	-3	-4	-5	-6	-5	3	5	-5	1	1	-5	-5	-3	-4	-3	1
448	K	0	0	1	-3	-5	-1	-1	-3	-3	-5	-5	7	-4	-5	-3	-1	-2	-5	-4	-4
449	S	0	-3	-2	-3	0	-2	-2	-3	-3	-4	-4	-2	-4	-5	2	6	2	-5	-4	-4
450	K	0	3	0	1	-5	0	0	-4	-1	-4	-3	4	-3	-2	2	1	-1	-5	-4	-4
451	N	-4	-3	8	-1	-5	-2	-2	-3	-1	-6	-6	-2	-4	-5	-4	-1	-2	-6	-4	-5
452	I	-3	-5	-5	-6	0	-5	-5	-6	-5	6	2	-5	2	-2	-5	-4	-3	-5	-3	3
453	M	-4	-4	-6	-6	-3	-4	-5	-6	-5	0	6	-5	1	0	-5	-4	-3	-4	-3	0
454	V	-3	-3	-5	-6	-3	-4	-5	-6	-5	3	3	-4	2	-2	-5	-4	-3	-5	-3	5
455	K	-2	1	1	4	-5	0	-1	-2	1	-4	-2	4	-3	-2	-3	0	-1	-5	-2	-3
456	N	1	1	3	0	-4	-1	1	0	-3	-4	-4	3	-2	-5	-2	2	-2	-5	-4	-4
457	D	-3	-2	5	5	-1	-1	1	-1	0	-5	-4	0	-2	-5	-1	0	-2	-6	-4	-5
458	L	-3	-1	0	-3	0	-3	-2	3	-4	-2	3	0	1	1	-2	-2	-3	5	-1	-3