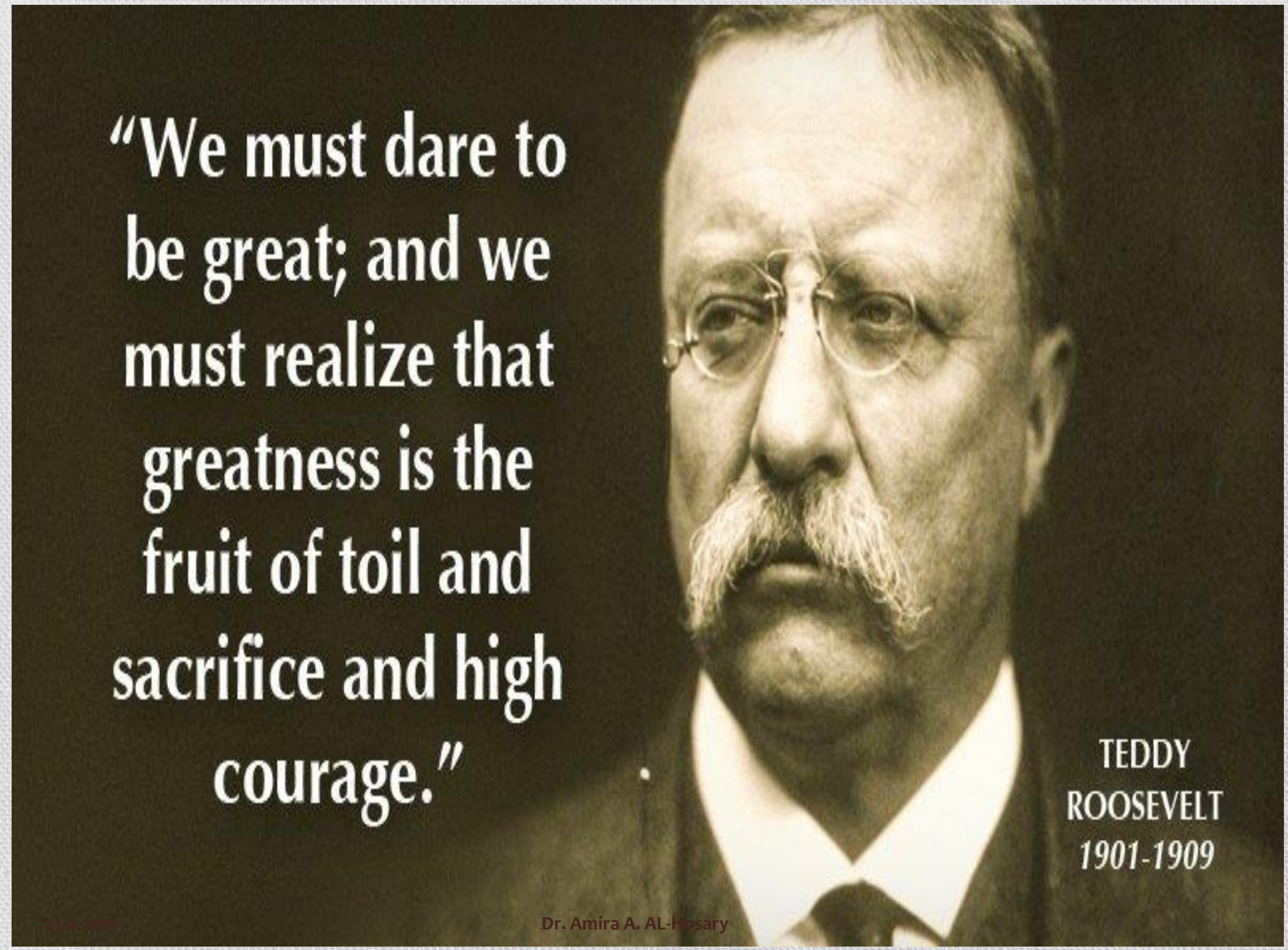

"We must dare to be great; and we must realize that greatness is the fruit of toil and sacrifice and high courage."
TEDDY ROOSEVELT
1901-1909

BLAST & FASTA

Amira A. AL-Hosary
PhD of infectious diseases
Department of Animal Medicine
(Infectious Diseases)
Faculty of Veterinary Medicine
Assiut University-Egypt

WHAT IS A DATABASE?

A *database* is a computerized archive used to store and organize data in such a way that information can be retrieved easily via a variety of search criteria.

Each record, also called an *entry*, should contain a number of fields that hold the actual data items.

Types of biological data:

Primary data: that includes simple raw data

-Nucleotides (DNA) -Amino Acids (Protein).

Secondary data: (protein): includes secondary data like motifs (regular expressions, blocks, profiles, finger print.

-2nd structures of protein like alpha-helix, Beta-strand.

Tertiary data: (Protein) tertiary data:

-Atomic co-ordination (tertiary protein structure).

The data bases may be classified by different ways one of them may be according to its data as above

Types of biological data bases

Nucleotides (INSDC)

- EMBL= Europe Molecular biology Lab, Europe =
- GENBANK= USA, NCBI
- DDBJ = DNA Data bank of Japan

Protein

- Sequence (uniport, PIR, SWISS Port) = 1 ry DB
- Interaction Bio grid string (Protein/Protein interaction.
- Structure
- PDB = protein database bank
- CATH = class architecture topology homology
- SCOPE = structure classification of protein

Whole genome DB

- ENSEMBL

Specialized DB

- Inherited disease database
- OMIM (Online Mendelian Inherited of man)

Biological databases:

Current biological databases use all three types of database structures: flat files, relational, and object oriented.

Flat File

Name, States, Course number, Course name|John Smith, Texas, Biol 689, Bioinformatics|Jane Doe, Kansas, Bich 441, Biochemistry|William Brown, Illinois, Chem 289, Organic Chemistry|Jennifer Taylor, New York, Hort 201, Horticulture|Howard Douglas, Texas, Math 172, Calculus

Table A

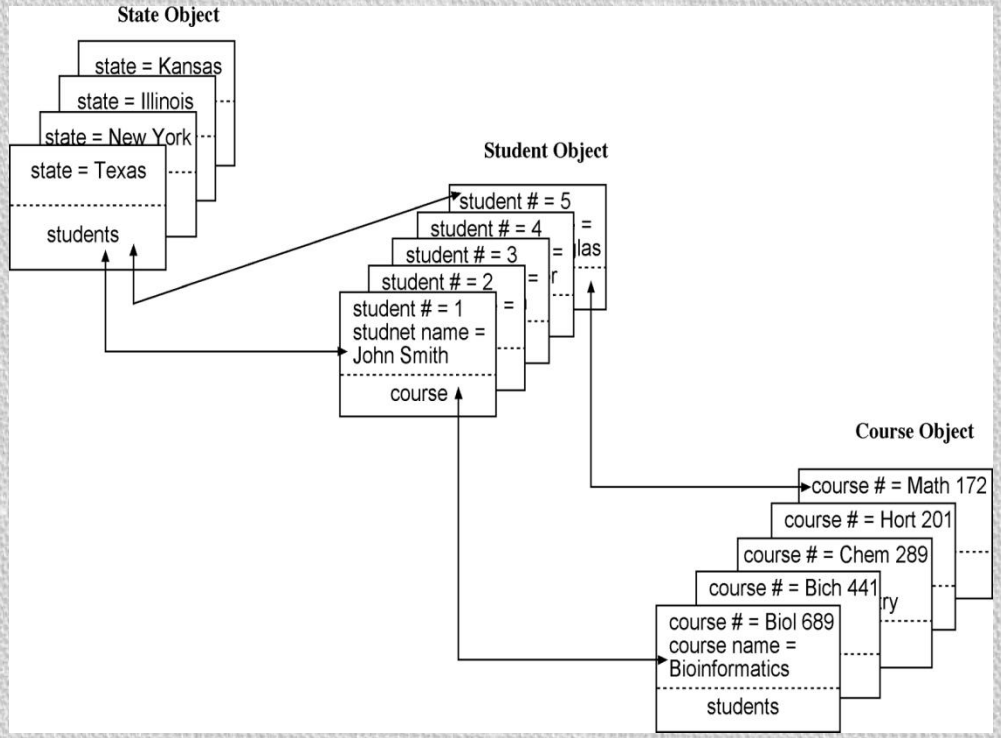
Student #	Name	State
1	John Smith	Texas
2	Jane Doe	Kansas
3	William Brown	Illinois
4	Jennifer Taylor	New York
5	Howard Douglas	Texas

Table B

Student #	Course #
1	Biol 689
2	Bich 441
3	Chem 289
4	Hort 201
5	Math 172

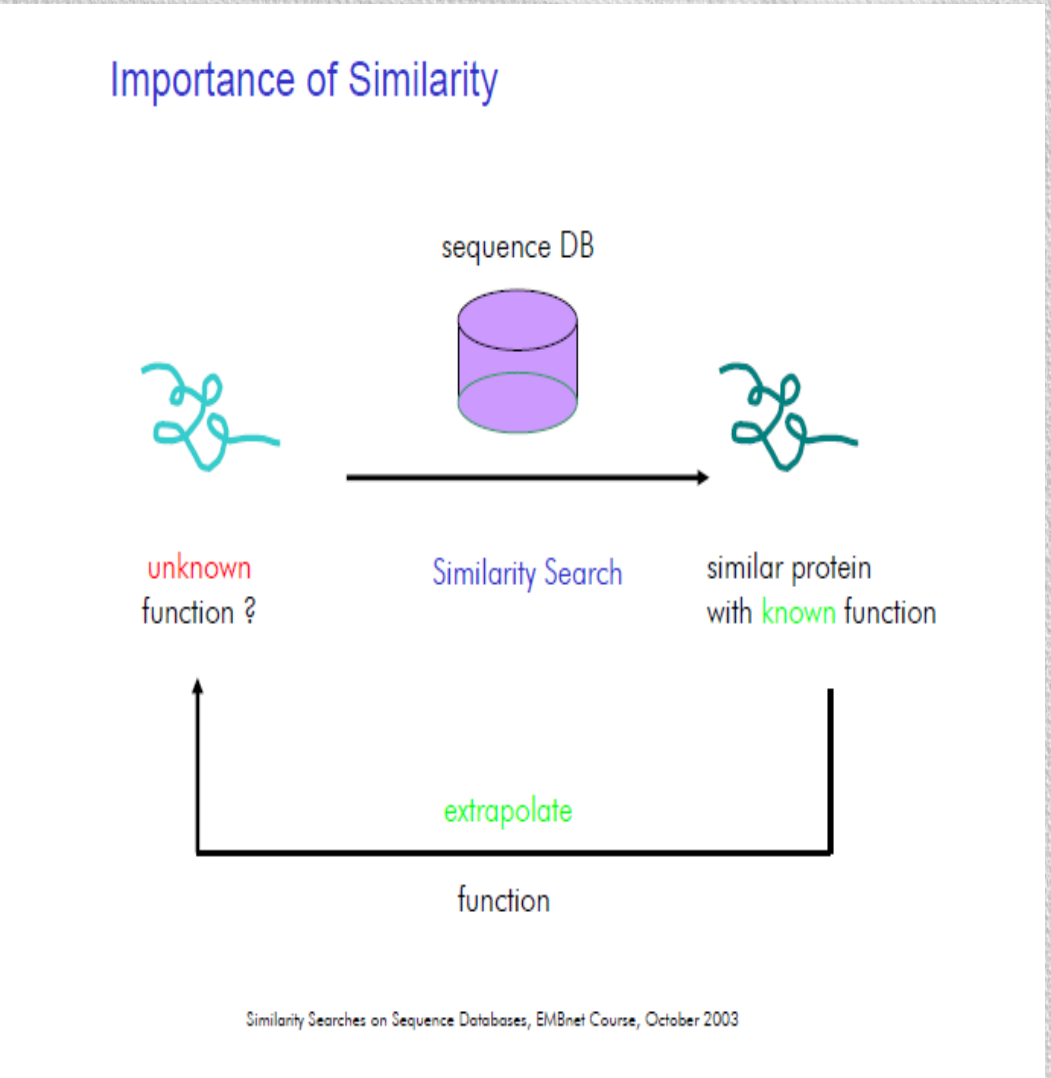
Table C

Course #	Course name
Biol 689	Bioinformatics
Bich 441	Biochemistry
Chem 289	Organic chemistry
Hort 201	Horticulture
Math 172	Calculus



Why we need data bases

Similar sequences/proteins: probably have the same ancestor, share the same structure, and have a similar biological function.



Search on biological data bases:

Searching on biological data bases depends mainly on sequences alignment.

Types of sequences alignment:

Global alignment

- Depends on full sequence alignment.

Local alignment

- Depends on partial sequence alignment.

Local alignment

1. Compare short sequence to long one.
2. Compare single sequence to entire database.
3. Compare partial sequence to whole.
4. Identified new determined sequence.
5. Compare new gens to new ones.
6. Guess function for entire genomes full of ORFs of unknown function.

BLAST & FASTA

FASTA	BLAST
First fast sequence algorithm for comparing query sequence to database sequences.	Improvement of FASTA, it usually search speed, easy to use, statistical rigor.

What is an algorithm= (خوارزميه) نظام حسابي؟

“Procedure for accomplishing some task”

– Set of well defined instructions

Cookbook recipe

– Produce result from initial data

Input data set > output data set

All software implements algorithms

The beginning of the story:

1985 : FASTP (D. Lipman and W. Pearson)

Global gapped alignments

1988 : FASTA (W. Pearson and D. Lipman)

Local gapped alignments

1990 : BLAST1

(S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman)

Local ungapped alignments

Heuristic Sequence Alignment BLAST, FASTA

Heuristic Methods

(BLAST, FASTA) they prune the search space by using fast approximate methods to select the sequences of the database that are likely to be similar to the query and to locate the similarity region inside them

Principle

Dynamic Programming, computational method that provide in mathematical sense the best alignment between two sequences, given a scoring system.

Both of them search for local sequence alignment but using different algorithms and statistical approaches.

FASTA

- FASTA is a DNA and protein sequence alignment software package first described (as FASTP) by David J. Lipman and William R. Pearson in 1985.
- Its legacy is the FASTA format which is now ubiquitous in bioinformatics.

FASTA

Its idea:

A good alignment contains subsequences of absolute identity (short lengths of exact matches).

Steps:

First: identify very short exact matches.

Next: the best short hits from the first step are extended to longer regions of similarity.

Finally: the best hits are optimized.

FASTA

FASTA algorithm:

It depends on logic dot plot method.

Compute best diagonals from all frames of alignment.

This method looks for exact matches between words in query and test sequence.

FASTA word:

DNA = 6 Nucleotides

Protein = 2 AA (amino acids).

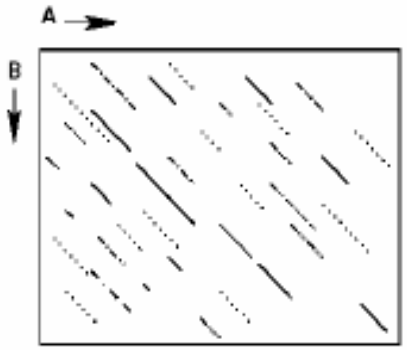
How FASTA algorithm works?

- **FASTA algorithm has five steps:**

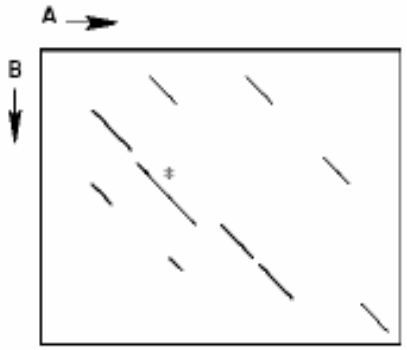
1. Identify common K-words between query and test sequences.
2. Score diagonals with K-words matches, identify 10 best diagonals.
3. Rescore the initial regions with a substitution score matrix.
4. Join the initial regions using gaps, penalize for gaps.
5. Perform dynamic programming to find final alignment.

FASTA: algorithm (4 steps)

Localize the 10 best regions of similarity between the two seq. Each identity between two "word" is represented by a dot



Identify all k-tuple matches



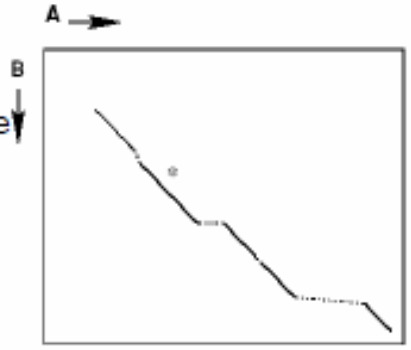
score the 10 best scoring regions using a scoring matrix

→ Init1 score

Each diagonal: ungapped alignment

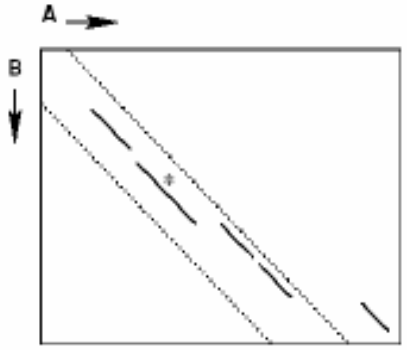
The smaller the k, The sensitive the method but slower

Find the best combination of the diagonals-> compute a score.
Only those sequences with a score higher than a threshold will go to the fourth step



Apply joining procedure

→ Initn score



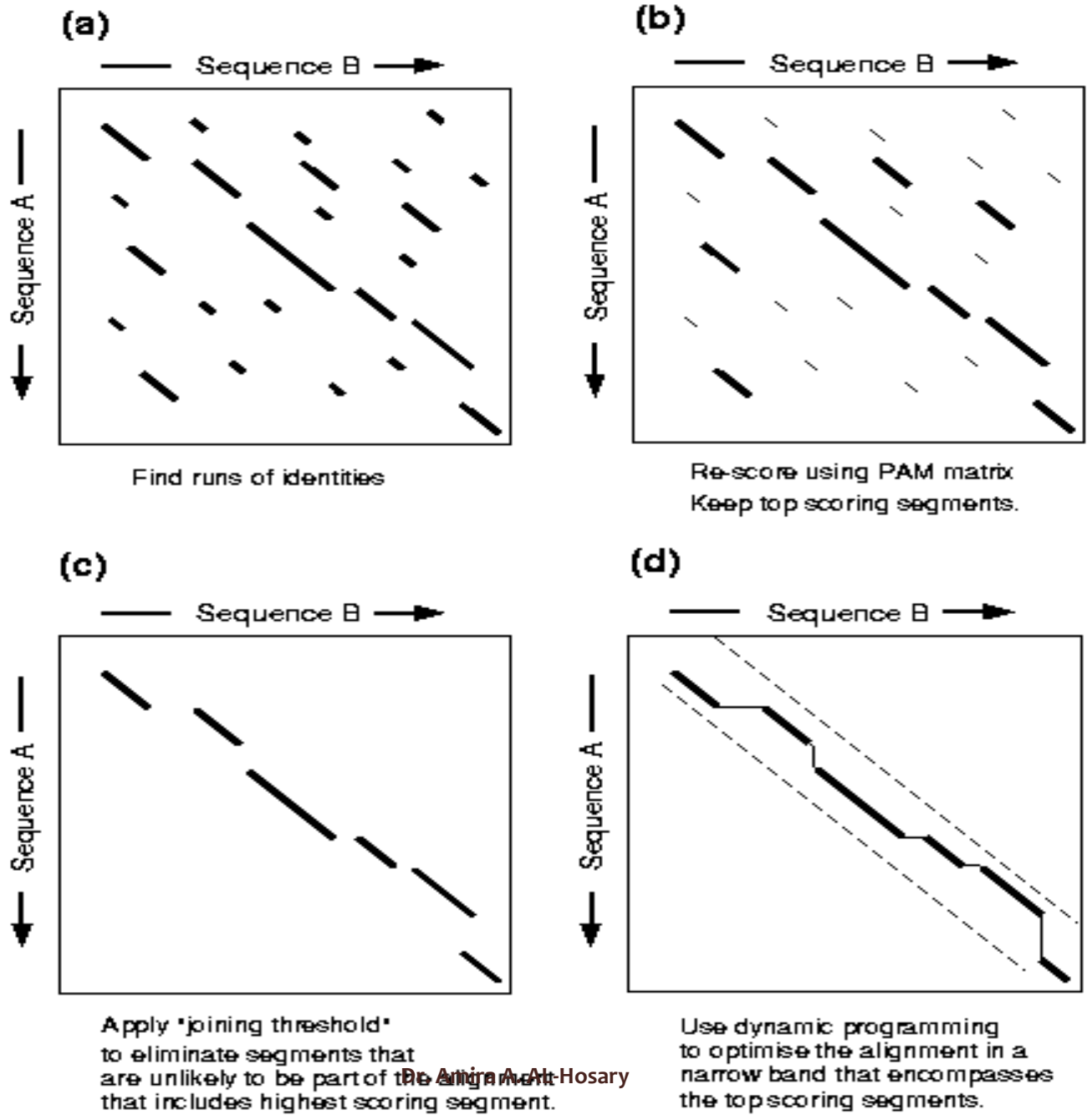
Apply limited DP

→ Opt score

DP applied around The best scoring diagonal.

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

FASTA Algorithm



How FASTA algorithm works?

↑Deviation→ ↑Z value → ↑E score

E- value

↓10^{-6} →Probably statistically

↑10^{-6} but ↓10^{-3} deserve a second look

↑10^{-6} should be thrown out with great force

FASTA Format:

FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which nucleotides or amino acids are represented using single-letter codes. The format also allows for sequence names and comments to precede the sequences.

A sequence in FASTA format start with (>) then a single-line description, followed by lines of sequence data.

```
>gi|18203677|sp|Q9ZGE9|BCHN
MERVERENGCFHTFCPIASVAWLHRKIKDSFFLIVGTHTCAHFIQTALDVMVYAHSRFGFAVLEESDLVS
ASPTEELGKVVQQVVDEWHPKVIFVLSTCSVDILKMDLEVSCKDLSTRFGFPVLPASTSGIDRSFTQGED
AVLHALLPFVPKEAPAVEPVEEKKPRWFSFGKESEKEKAEPARNLVLIGAVTDSTIQQLQWELKQLGLPK
VDVFPDGDIRKMPVINEQTVVVPLQPYLNDTLATIRRERRAKVLSTVFPIGPDGTARFLEAICLEFGLDT
SRIKEKEAQAWRDLEPQLQILRGKKIMFLGDNLLELPLARFLTSCDVQVVEAGTPYIHSKDLQQELELLK
ERDVRIVESPDFTKQLQRMQEYKPDLVVAGLGICNPLEAMGFTTAWSIEFTFAQIHGFVNAIDLIKLFTK
PLLKRQALMEHGWAEAGWLE
```


I hope you
always find
a reason to Smile...

Basic Local Alignment Search Tool

BLAST

Altschul, et al. 1990, 1994, 1997

BLAST

- Heuristic method for local alignment.
- Designed specifically for database searches.
- Based on the same assumption of FASTA that good alignments contain short length of exact matches.

Advantages of BLAST:

1. Speed.
2. User friendly.
3. More Sensitive.
4. Statistical rigor.

BLAST

- Like FASTA it depends on K-words= Q word.
- DNA word = 11 Nucleotides.
- Protein = 3 AA (amino acids).

How it Calculate number of word in given sequence?

number of word in query sequence= L – W + 1

L= sequence length.

W= word equal to 11 nucleotides

Example:

query sequence length is 120

(120 – 11) + 1 = 110 Q word

How BLAST algorithm works?

For example, suppose that the sequence contains the following stretch of letters, **GLKFA**.

If a BLASTp was being conducted under default conditions, the word size would be 3 letters.

In this case, using the given stretch of letters, the searched words would be **GLK, LKF, KFA**.

Once both words and neighborhood words are assembled and compiled, they are compared to the sequences in the database in order to find matches. The threshold score *T* determines whether or not a particular word will be included in the alignment.

How BLAST algorithm works?

1. Calculate the words in both query and database sequences.
2. Identification of the exact word (breaks query and database sequences in to words then match both of them).
3. Maximum Segment pair alignment (MSP): for each word match extend the alignment in both directions to find alignment that score greater than the threshold of value S.

How BLAST algorithm works?

HSPS

It is aligned regions, the results of the word matching and attempts to extend the alignments are segments.

It called HSPS (High scoring segment Paris).

BLAST often produces several short HSPS rather than single aligned regions.

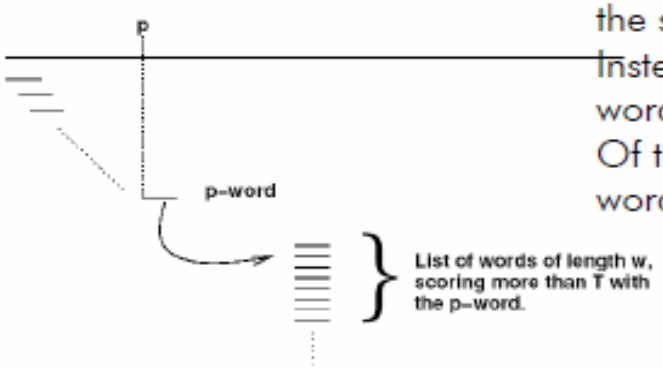
How BLAST algorithm works?

BLAST1: Algorithm

With w=2 :
(20x20=400
Possible words,
w=3, 8000
Possible words,…)

First step:

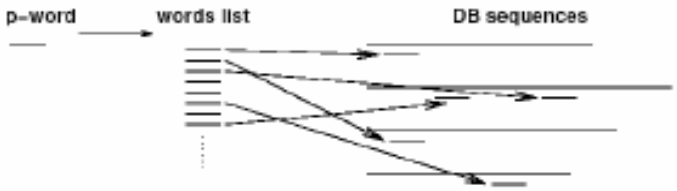
For each position p of the query, find the list or words of length w scoring more than T when paired with the word starting at p:



Quickly locate ungapped similarity regions between the sequences.
Instead of comparing each word of the query with each word Of the DB: create a list of "similar" words.

Second step:

For each words list, identify all exact matches with DB sequences:



Similarity Searches on Sequence Databases, EMBnet Course, October 2003

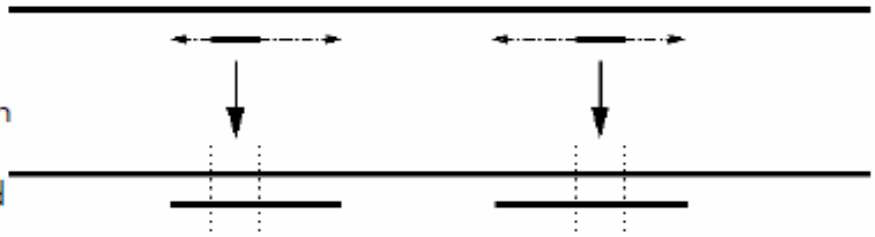
How BLAST algorithm works?

BLAST1: Algorithm

Third step:

For each word match («hit»), extend ungapped alignment in both directions. Stop when S decreases by more than X from the highest value reached by S.

Each match is then extended. The extension is stopped as soon as the score decreases more then X when compared with the highest value obtained During the extension process



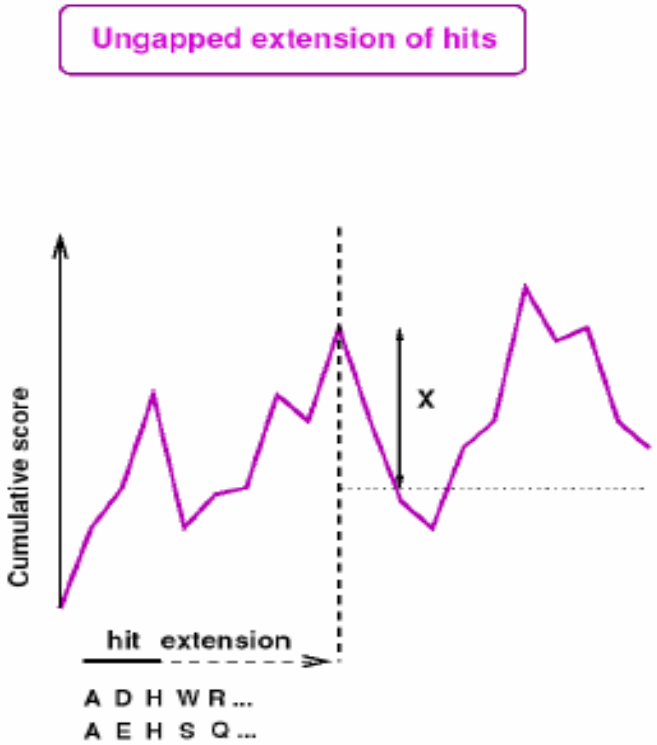
HSP = High Scoring Segment Pair

Reports all HSPs having score S above a threshold, or equivalently, having E–value below a threshold.

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

How BLAST algorithm works?

BLAST1: Algorithm



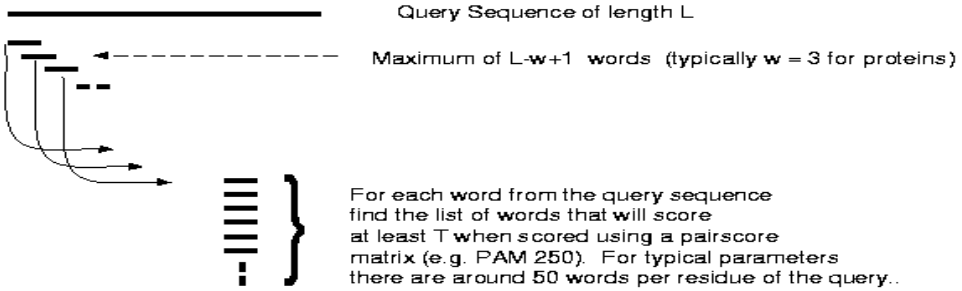
Each match is then extended. The extension is stopped as soon as the score decreases more then X when compared with the highest value obtained During the extension process

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

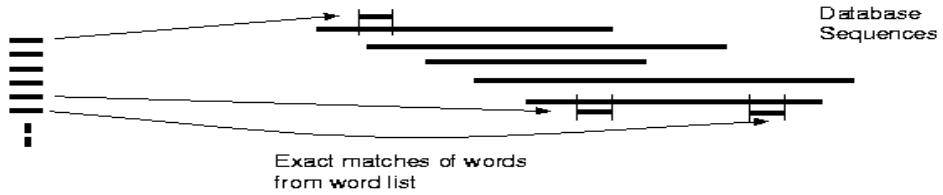
How BLAST algorithm works?

BLAST Algorithm

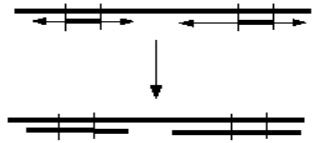
(1) For the query find the list of high scoring words of length w.



(2) Compare the word list to the database and identify exact matches.



(3) For each word match, extend alignment in both directions to find alignments that score greater than score threshold S.



How BLAST algorithm works?

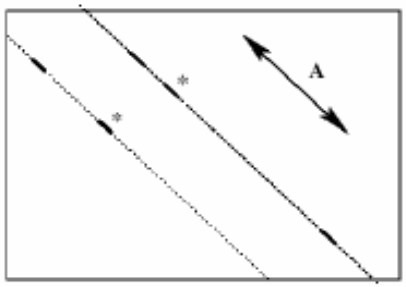
BLAST2: (NCBI)

The «two–hits» requirement

First step: as with BLAST1, generate lists of words scoring more than T with words of the query.

Second step: generation of hits: identify all word matches in DB sequences

Third step: extension of hits: requires a second hit on the same diagonal at a distance of less than A.



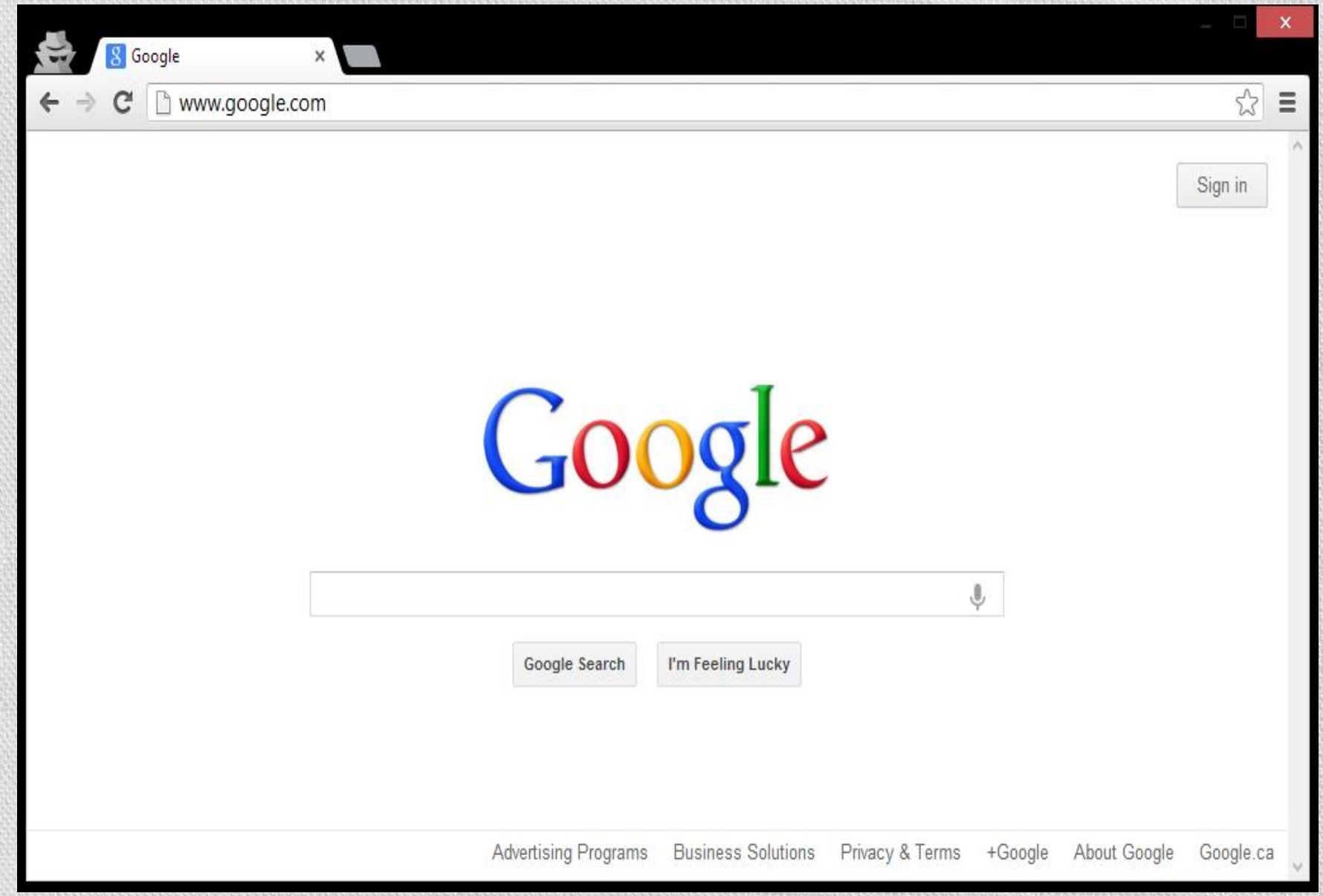
Additional step:
Gapped extension of the hits slower-> therefore: requirement of a second hits on the diagonal. (hits not joined by ungapped extensions could be part of the same gapped alignmnet)

This step generates ungapped HSPs

Fourth step: gapped extension of HSPs having score above a threshold S_g

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

Practical work:


Google
www.google.com
Sign in
Google
Google Search
I'm Feeling Lucky
Advertising Programs
Business Solutions
Privacy & Terms
+Google
About Google
Google.ca

Input:

1. Input sequences are in FASTA or Genbank format and weight matrix.
2. FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which nucleotides or amino acids are represented using single-letter codes.
3. The format also allows for sequence names and comments to precede the sequences.
4. The format originates from the FASTA software package, but has now become a standard in the field of bioinformatics.

1. A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column.
2. The word following the ">" symbol is the identifier of the sequence, and the rest of the line is the description (both are optional).
3. There should be no space between the ">" and the first letter of the identifier. It is recommended that all lines of text be shorter than 80 characters. The sequence ends if another line starting with a ">" appears; this indicates the start of another sequence.
4. A simple example of one sequence in FASTA format:

```
>gi|18203677|sp|Q9ZGE9|BCHN
MERVERENGCFHTFCPIASVAWLHRKIKDSFFLIVGTHTCAHFIQTALDVMVYAHSRFGFAVLEESDLVS
ASPTEELGKVVQQVVDEWHPKVIFVLSTCSVDILKMDLEVSCKDLSTRFGFPVLPASTSGIDRSFTQGED
AVLHALLPFVPKEAPAVEPVEEKKPRWFSFGKESEKEKAEPARNLVLIGAVTDSTIQQLQWELKQLGLPK
VDVFPDGDIRKMPVINEQTVVVPLQPYLNDTLATIRRERRAKVLSTVFPIGPDGTARFLEAICLEFGLDT
SRIKEKEAQAWRDLEPQLQILRGKKIMFLGDNLLELPLARFLTSCDVQVVEAGTPYIHSKDLQQELELLK
ERDVRIVESPDFTKQLQRMQEYKPDLVVAGLGICNPLEAMGFTTAWSIEFTFAQIHGFVNAIDLIKLFTK
PLLKRQALMEHGWAEAGWLE
```

BLAST: Basic Local Alignment Search Tool

blast.ncbi.nlm.nih.gov/

The **Basic Local Alignment Search Tool** (**BLAST**) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to **...**

Align two or more - Protein BLAST: ***search ... - Nucleotide BLAST Rat - sequences

Nucleotide **BLAST**: Search nucleotide databases using a nucleotide **...**

blast.ncbi.nlm.nih.gov/Blast.cgi?...blastn...BlastSearch...

No BLAST database contains all the sequences at NCBI. BLAST databases **...**

BLAST - Wikipedia, the free encyclopedia

en.wikipedia.org/wiki/BLAST

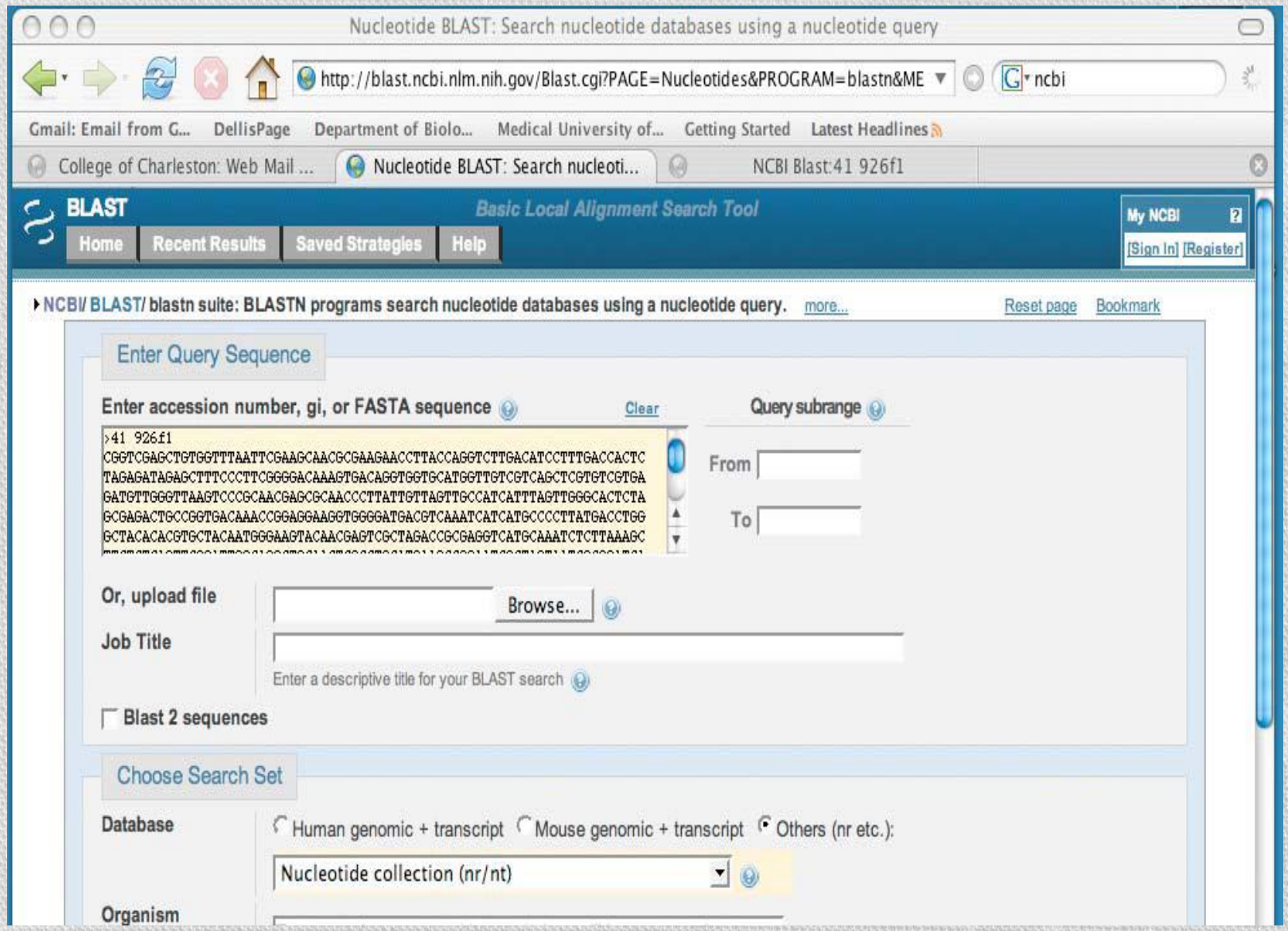
In bioinformatics, **Basic Local Alignment Search Tool**, or **BLAST**, is an algorithm for comparing primary biological sequence information, such as the amino-acid **...**

Process - Output - Input - Background

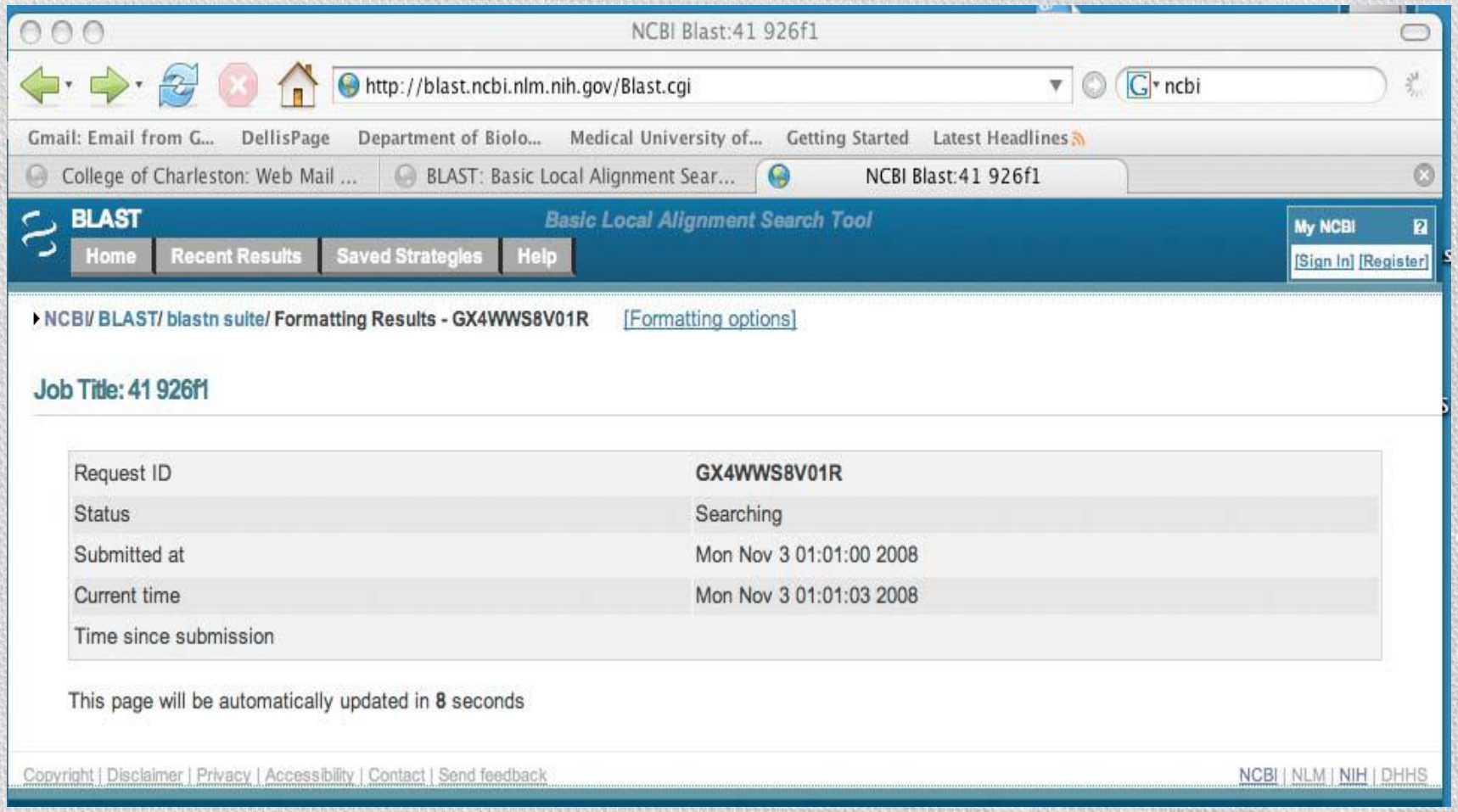
Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous 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



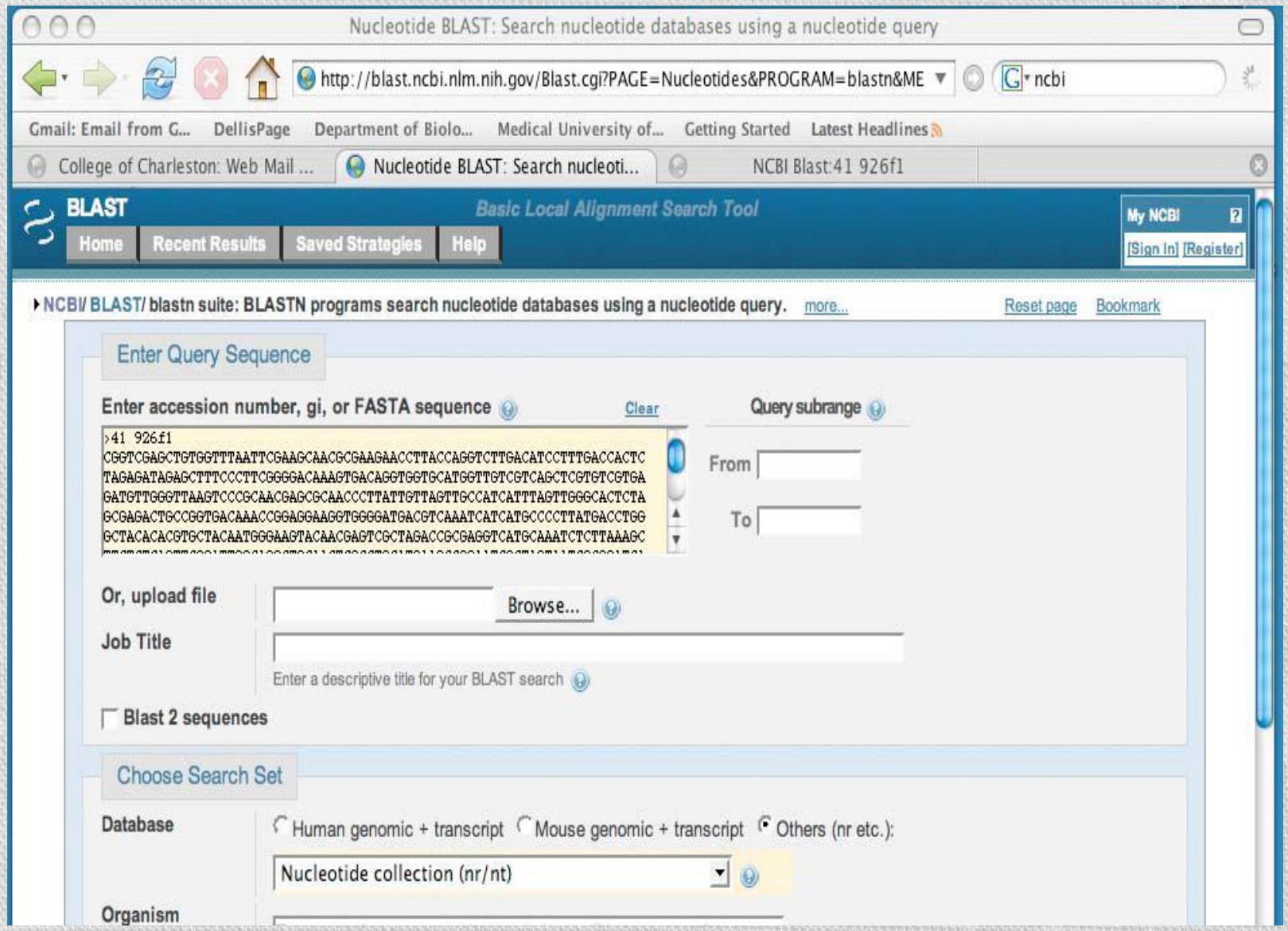
Click the "Blast!" button at the bottom to submit your sequence data.



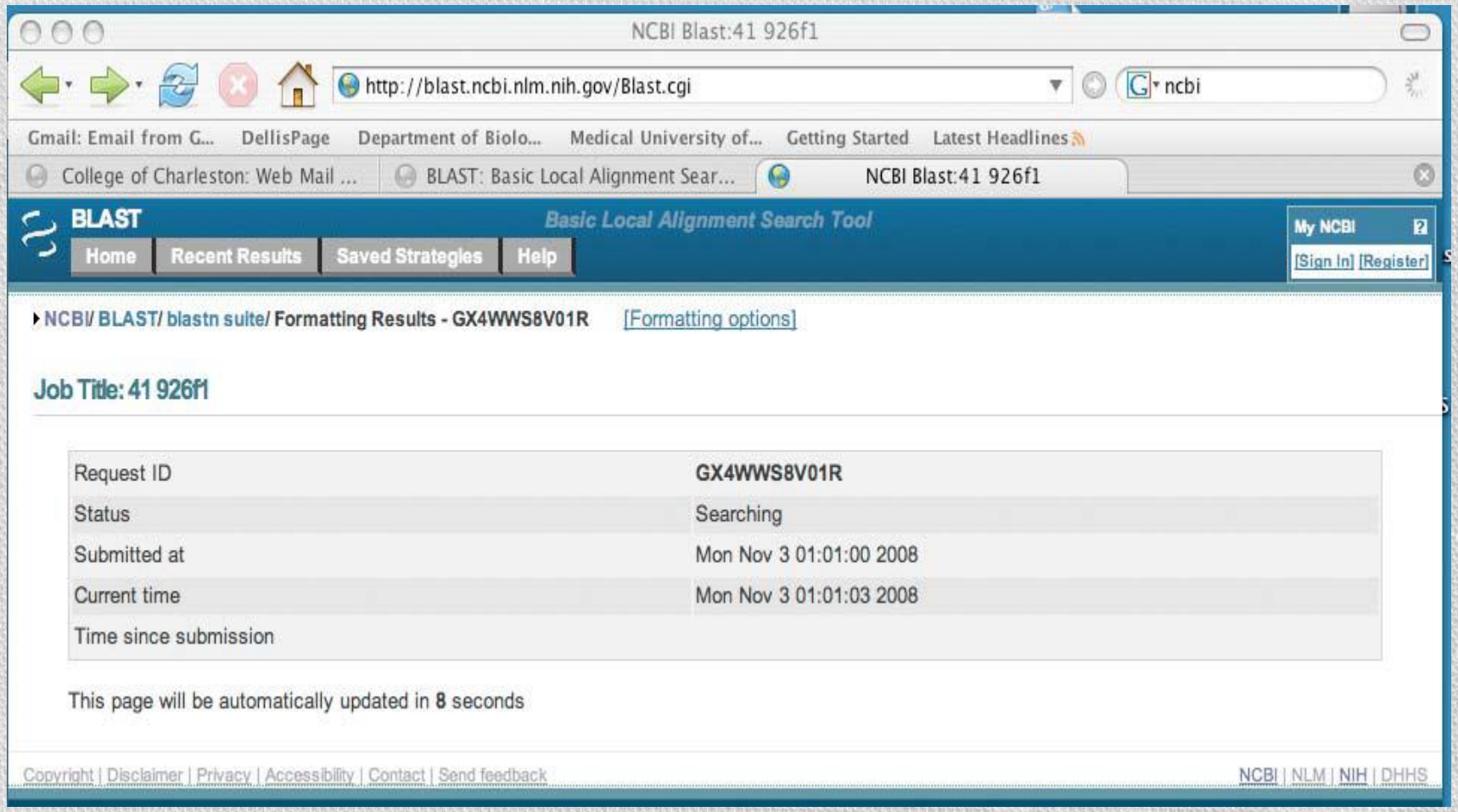
This screen will come up next. Finally (sometimes after a lengthy wait), a new window will appear showing any "hits" your sequence made. The results will be color coded and annotated

Output:

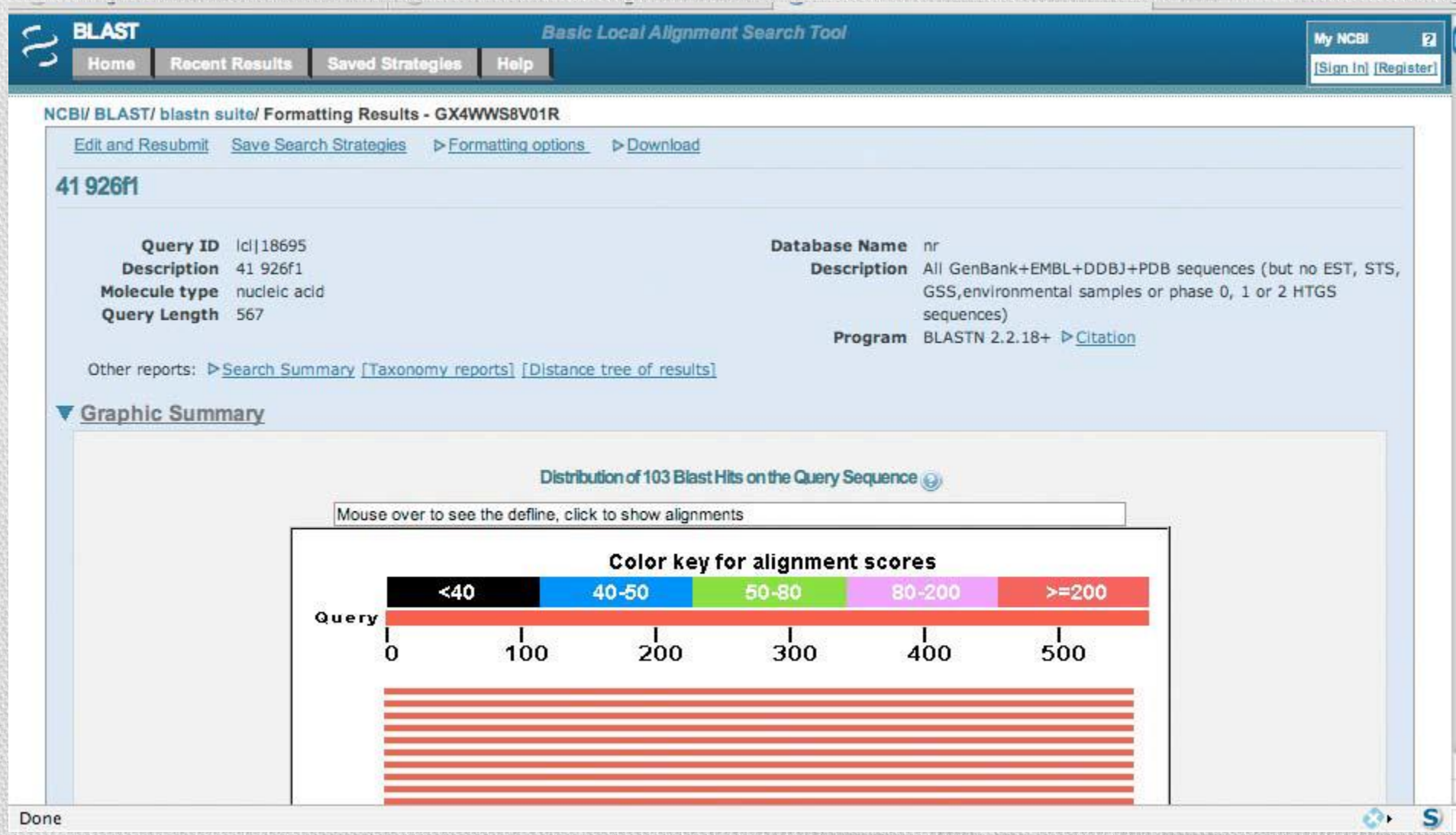
BLAST output can be delivered in a variety of formats. These formats include HTML, plain text, and XML formatting.



Click the "Blast!" button at the bottom to submit your sequence data.



This screen will come up next. Finally (sometimes after a lengthy wait), a new window will appear showing any "hits" your sequence made. The results will be color coded and annotated



The bars show what places along your sequence are similar to other published sequences; the colors indicate how many bases were involved in homology determination.

Descriptions

Legend for links to other resources: U UniGene E 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
EU557008.1	Uncultured bacterium clone C56 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557006.1	Uncultured bacterium clone C59 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557004.1	Uncultured bacterium clone C62 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557001.1	Uncultured bacterium clone C66 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU557000.1	Uncultured bacterium clone C72 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556999.1	Uncultured bacterium clone C75 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556998.1	Uncultured bacterium clone C80 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU556996.1	Uncultured bacterium clone C99 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EU285587.1	Enterococcus faecalis strain C19315led5A 16S ribosomal RNA gene, partial s	946	946	98%	0.0	97%	
EU547775.1	Enterococcus faecalis strain IJ-07 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
AB362599.1	Enterococcus faecalis gene for 16S rRNA, partial sequence, strain: NRIC 011	946	946	98%	0.0	97%	
EF653454.1	Enterococcus faecalis strain 47/3 16S ribosomal RNA gene, partial sequence	946	946	98%	0.0	97%	
EF608536.1	Uncultured bacterium clone PCD-8 16S ribosomal RNA gene, partial sequenc	946	946	98%	0.0	97%	
AM697463.1	Uncultured bacterium partial 16S rRNA gene, isolate BF0001D078	946	946	98%	0.0	97%	

Clicking on a "gi" link at the beginning of any line will take you to the GenBank accession page for a sequence showing similarity to yours. There you can find a wealth of information about the published sequence to which yours showed some homology.

>gb|EU285587.1| Enterococcus faecalis strain C19315led5A 16S ribosomal RNA gene, partial sequence
Length=1456

Score = 946 bits (512), Expect = 0.0
Identities = 550/566 (97%), Gaps = 12/566 (2%)
Strand=Plus/Plus

```
Query  1     CGGTCGAGC-TGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCC  59
             |||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  893   CGGTGGAGCATGTGGTTTAATTCGAAGCAACGCGAAGAACCTTACCAGGTCTTGACATCC  952

Query  60    TTTGACCACTCTAGAGATAGAGCTTTCCCTTCGGGGACAAAGTGACAGGTGGTGCATGGT  119
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  953   TTTGACCACTCTAGAGATAGAGCTTTCCCTTCGGGGACAAAGTGACAGGTGGTGCATGGT  1012

Query  120   TGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATT  179
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1013  TGTCGTCAGCTCGTGTCGTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCTTATT  1072

Query  180   GTTAGTTGCCATCATTTAGTTGGGCACTCTAGCGAGACTGCCGGTGACAAACCGGAGGAA  239
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1073  GTTAGTTGCCATCATTTAGTTGGGCACTCTAGCGAGACTGCCGGTGACAAACCGGAGGAA  1132

Query  240   GGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAAT  299
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1133  GGTGGGGATGACGTCAAATCATCATGCCCCTTATGACCTGGGCTACACACGTGCTACAAT  1192

Query  300   GGGAAGTACAACGAGTCGCTAGACCGCGAGGTCATGCAAATCTCTTAAAGCTTCTCTCAG  359
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1193  GGGAAGTACAACGAGTCGCTAGACCGCGAGGTCATGCAAATCTCTTAAAGCTTCTCTCAG  1252

Query  360   TTCGGATTGGCAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCGGATC  419
             |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1253  TTCGGATTG-CAGGCTGCAACTCGCCTGCATGAAGCCGGAATCGCTAGTAATCGCGGATC  1311

Query  420   AGCACGCCGCGGTGAATACGTTGCCGGGGCCTTGTACACACCGCCCGTCACACCACGAGA  479
             |||||||||||||||||||||| ||||| ||||||||||||||||||||||||||||||||
Sbjct  1312  AGCACGCCGCGGTGAATACGTTCCCGGG-CCTTGTACACACCGCCCGTCACACCACGAGA  1370

Query  480   GTTTGTAACACCCGAAGTCGG-GAGGTACCCTTTT-GGAGC-A-CCGCCTTAGGTGG-AT  534
             |||||||||||||||||||||| |||||| |||||| ||||| | |||||| |||||| ||
Sbjct  1371  GTTTGTAACACCCGAAGTCGGTGAGGTAACCTTTTTGGAGCCAGCCGCCTAAGGTGGGAT  1430

Query  535   AGATGAT-GGGGTGA-GTTC-TAACA  557
             ||||||| ||||||| || | |||||
Sbjct  1431  AGATGATTGGGGTGAAGT-CGTAACA  1455
```

Output:

Understanding your BLAST output

1. Graphic display:
   shows you where your query is similar to other sequences

2. Hit list:
   the name of sequences similar to your query, ranked by similarity

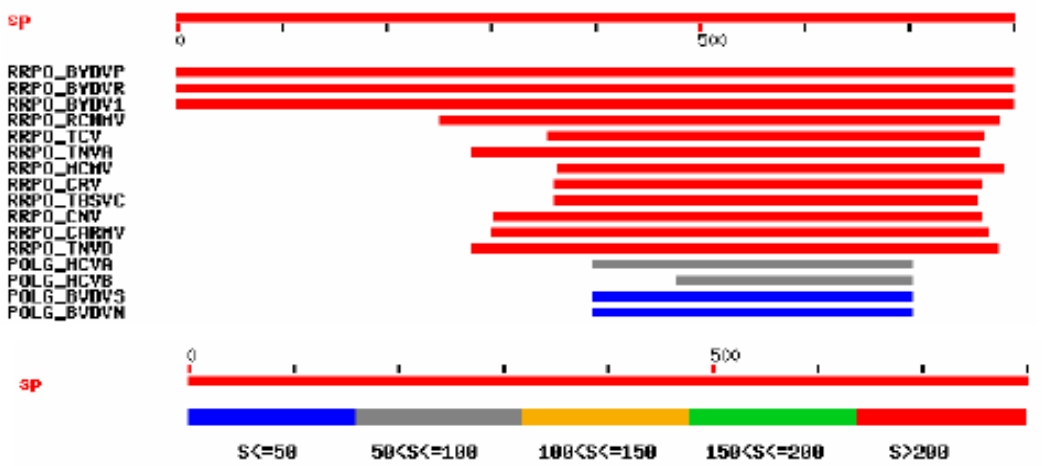
3. The alignment:
   every alignment between your query and the reported hits

4. The parameters:
   a list of the various parameters used for the search

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

Output:

Understanding your BLAST output: 1. Graphic display



query sequence

Portion of another sequence similar to your query sequence:

red, green, ochre, matches: good
grey matches: intermediate
blue: bad, (twilight zone)

The display can help you see that some matches do not extend over the entire length of your sequence => useful tool to discover domains.

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

Output:

Understanding your BLAST output: 2. Hit list

Sequences producing significant alignments:	Score (bits)	E Value
sp|P09505|RRPO_BYDVP Putative RNA-directed RNA polymerase (EC 2....	1652	0.0
sp|P29045|RRPO_BYDVR Putative RNA-directed RNA polymerase (EC 2....	1635	0.0
sp|P29044|RRPO_BYDV1 Putative RNA-directed RNA polymerase (EC 2....	1625	0.0
sp|P22956|RRPO_RCNMV Putative RNA-directed RNA polymerase (EC 2....	367	e-101
sp|P17460|RRPO_TCV Probable RNA-directed RNA polymerase (EC 2.7....	286	1e-76
sp|P22958|RRPO_TNVA RNA-directed RNA polymerase (EC 2.7.7.48) [C...	280	1e-74

Sequence ac number and name | Description | Bit score | E-value

- Sequence ac number and name: Hyperlink to the database entry: useful annotations
- Description: better to check the full annotation

- Bit score (normalized score) : A measure of the similarity between the two sequences: the higher the better (matches below 50 bits are very unreliable)

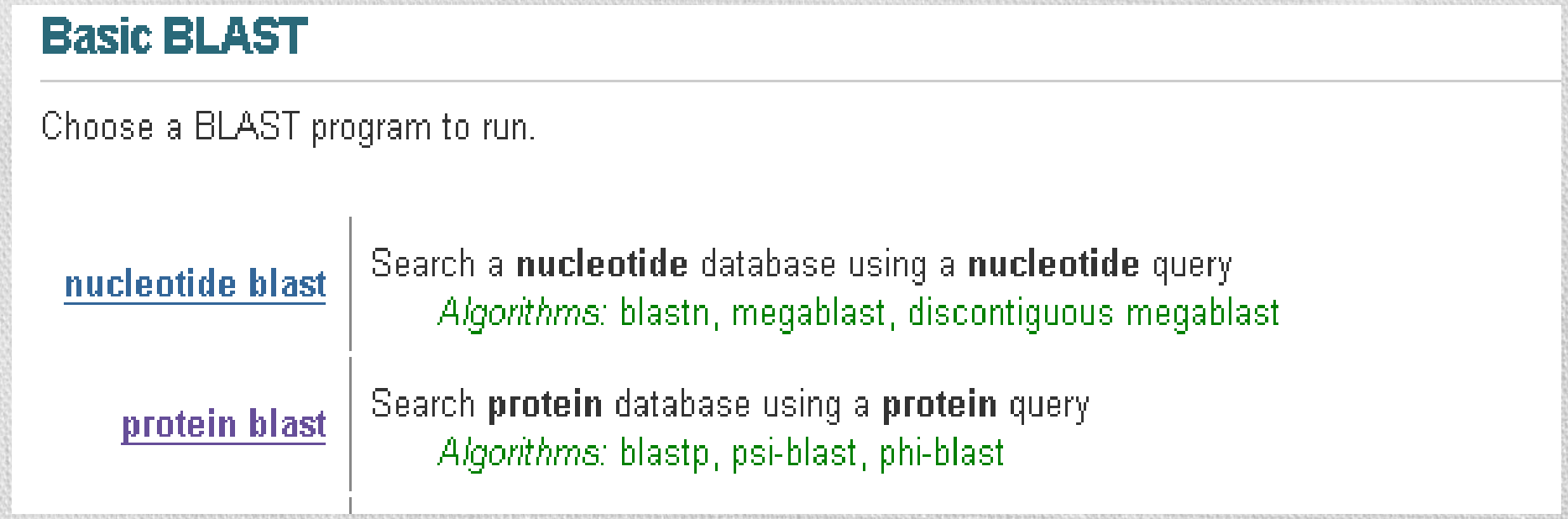
- E-value: The lower the E-value, the better. Sequences identical to the query have an E-value of 0. Matches above 0.001 are often close to the twilight zone. As a rule-of-thumb an E-value above 10-4 (0.0001) is not necessarily interesting. If you want to be certain of the homology, your E-value must be lower than 10^{-4}

Similarity Searches on Sequence Databases, EMBnet Course, October 2003

Basic BLAST

Choose a BLAST program to run.

nucleotide blast	Search **nucleotide** database using a **nucleotide** query *Algorithms:* blastn, megablast, discontiguous 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



Nucleotide blast:

compares a nucleotide query sequence to nucleotide database.

protien blast:

compares a protien query sequence to protein sequence database.

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

blastx:

compares a nucleotide query sequence translated in all reading frames against a protein sequence database.

tblastn:

Compares a protein query sequence against a nucleotide sequence database dynamically translated in all reading frames.

tblastx:

Compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

Conversion of sequence format:

Readseq < Sequence Format *Conversion* < EMBL-EBI

www.ebi.ac.uk/Tools/sfc/readseq/

Pubmed:

It is biomedical literature database and one of the accessible data bases, which contains abstracts and in some cases the full text articles fromnearly 4,000 journals.

An important feature of PubMed is the retrieval of information based on medical subject headings (MeSH) terms.

MeSH system consists of a collection of more than 20,000 controlled and standardized vocabulary terms used for indexing articles.

In other words, it is a the saurus that helps convert search *keywords into standardized terms to describe a concept.*

By doing so, it allows "smart" searches in which a group of accepted synonyms are employed so that the user not only gets exact matches, but also related matches on the same topic that otherwise might have been missed.

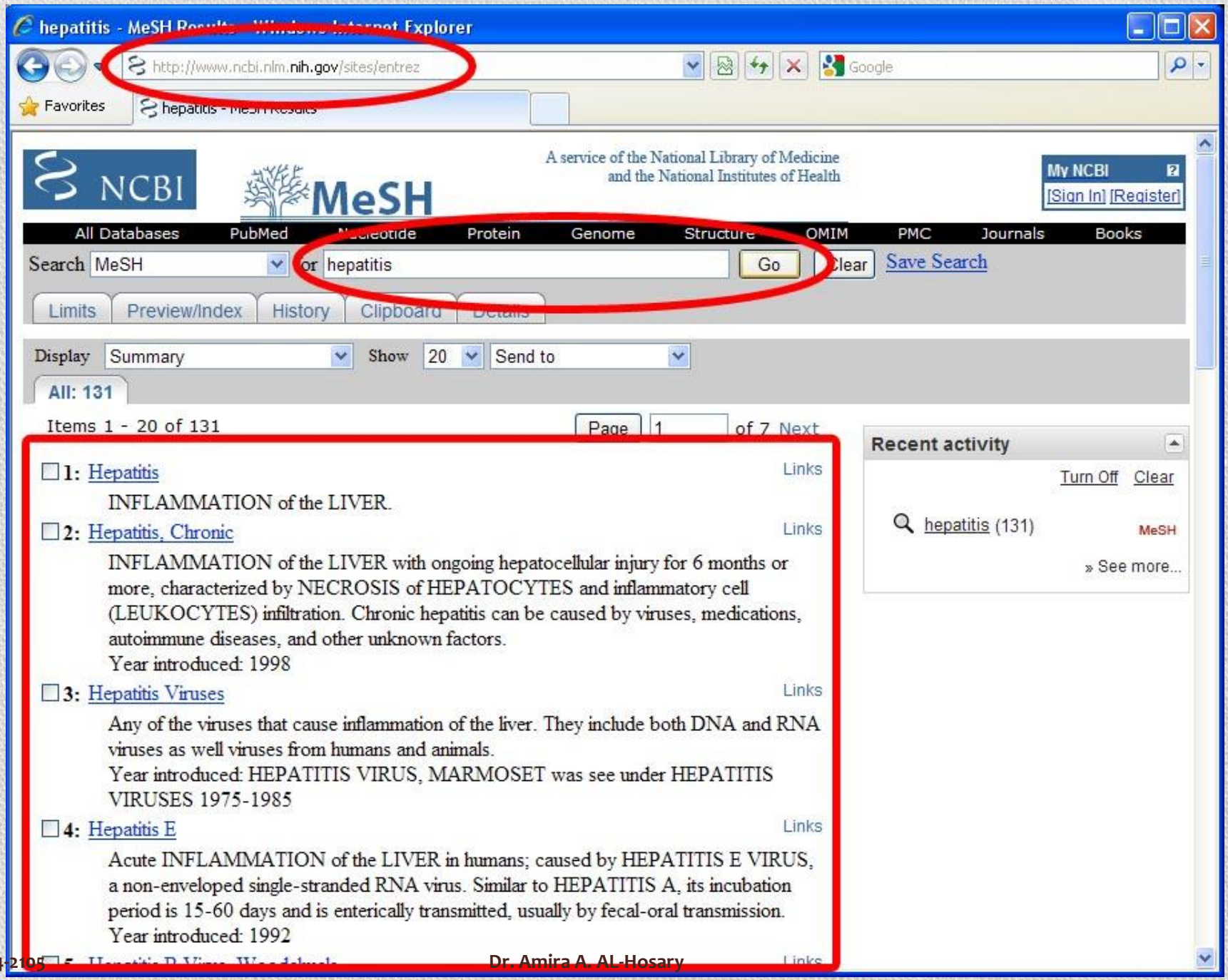
Pubmed:

Another way to broaden the retrieval is by using the **"Related Articles" option**.

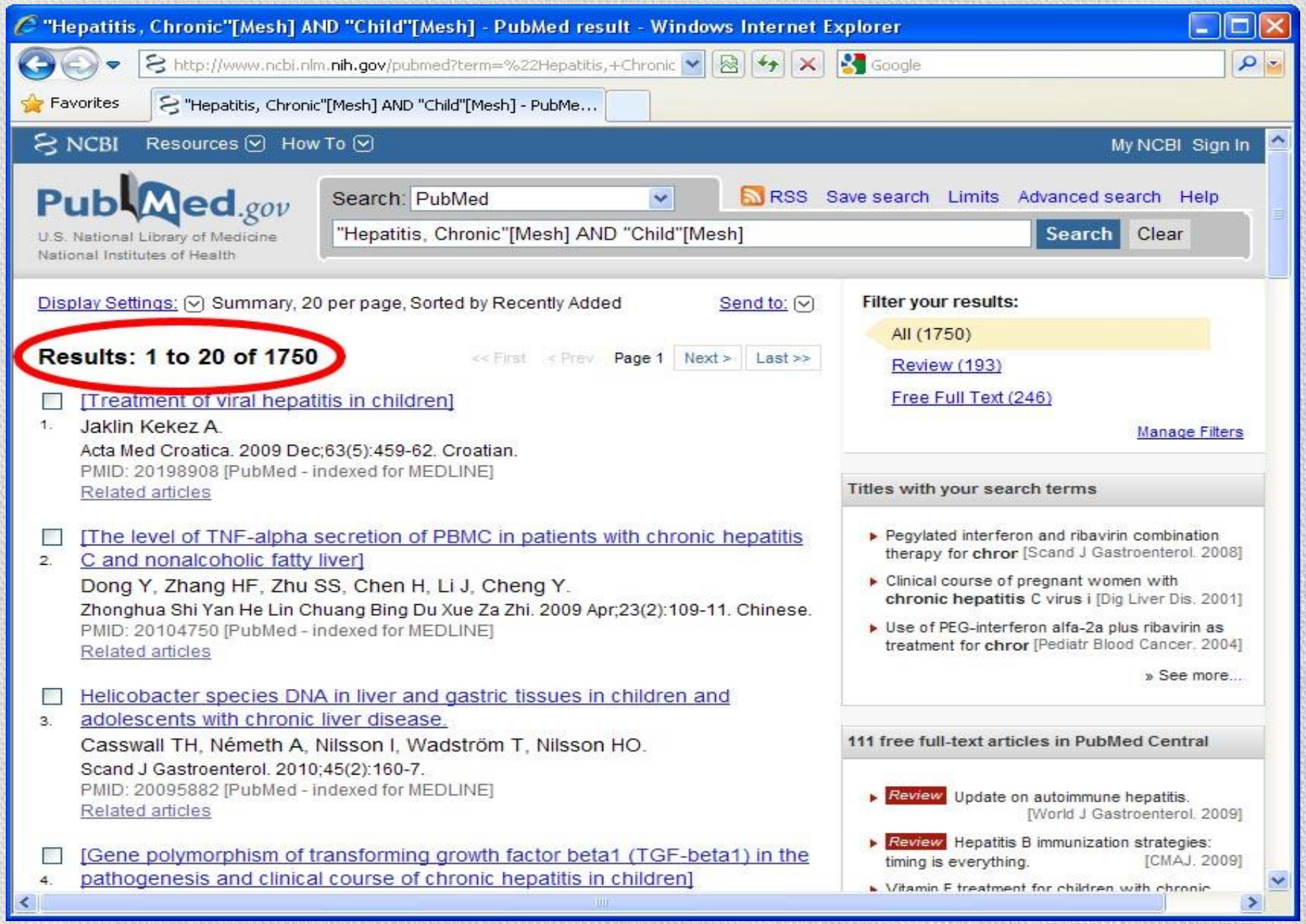
PubMed uses a word weight algorithm to identify related articles with similar words in the titles, abstracts, and MeSH. By using this feature, articles on the same topic that were missed in the original search can be retrieved.

Several Selected PubMed Tags and Their Brief Descriptions		
Tag	**Name**	**Description**
AB	Abstract	Abstract
AD	Affiliation	Institutional affiliation and address of the first author and grant numbers
AID	Article identifier	Article ID values may include the PII (controlled publisher identifier) or doi (digital object identifier)
AU	Author	Authors
DP	Publication date	The date the article was published
JID	Journal ID	Unique journal ID in the National Library ofMedicine's catalog of books, journals, and audiovisuals
LA	Language	The language in which the article was published
PL	Place of publication	Journal's country of publication
PT	Publication type	The type of material the article represents.
SO	Source	Composite field containing bibliographic information
TA	Journal title abbreviation	Standard journal title abbreviation
TI	Title	The title of the article
VI	Volume	Journal volume

Source: www.ncbi.nlm.nih.gov/entrez/query/static/help/pmhelp.html.


hepatitis - MeSH Results - Windows Internet Explorer
http://www.ncbi.nlm.nih.gov/sites/entrez
Google
Favorites
hepatitis - MeSH Results
NCBI
MeSH
A service of the National Library of Medicine and the National Institutes of Health
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search MeSH for hepatitis Go Clear Save Search
Limits Preview/Index History Clipboard Details
Display Summary Show 20 Send to
All: 131
Items 1 - 20 of 131
Page 1 of 7 Next
1: Hepatitis Links
INFLAMMATION of the LIVER.
2: Hepatitis, Chronic Links
INFLAMMATION of the LIVER with ongoing hepatocellular injury for 6 months or more, characterized by NECROSIS of HEPATOCYTES and inflammatory cell (LEUKOCYTES) infiltration. Chronic hepatitis can be caused by viruses, medications, autoimmune diseases, and other unknown factors.
Year introduced: 1998
3: Hepatitis Viruses Links
Any of the viruses that cause inflammation of the liver. They include both DNA and RNA viruses as well viruses from humans and animals.
Year introduced: HEPATITIS VIRUS, MARMOSET was see under HEPATITIS VIRUSES 1975-1985
4: Hepatitis E Links
Acute INFLAMMATION of the LIVER in humans; caused by HEPATITIS E VIRUS, a non-enveloped single-stranded RNA virus. Similar to HEPATITIS A, its incubation period is 15-60 days and is enterically transmitted, usually by fecal-oral transmission.
Year introduced: 1992
Recent activity
Turn Off Clear
hepatitis (131) MeSH
» See more...


hepatitis - MeSH Results - Windows Internet Explorer
http://www.ncbi.nlm.nih.gov/sites/entrez
Google
Favorites
hepatitis - MeSH Results
NCBI
MeSH
A service of the National Library of Medicine and the National Institutes of Health
My NCBI
[Sign In] [Register]
All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books
Search MeSH for hepatitis
Go
Clear
Save Search
Limits Preview/Index History Clipboard Details
"Hepatitis, Chronic"[Mesh]
Search PubMed
Clear
Display Summary
Show 20
Send to
Send to
Text
File
Printer
Clipboard
Search Box with AND
Search Box with OR
Search Box with NOT
All: 131
Items 1 - 20 of 131
of 7 Next
1: Hepatitis
Links
INFLAMMATION of the LIVER.
2: Hepatitis, Chronic
Links
INFLAMMATION of the LIVER with ongoing hepatocellular injury for 6 months or more, characterized by NECROSIS of HEPATOCYTES and inflammatory cell (LEUKOCYTES) infiltration. Chronic hepatitis can be caused by viruses, medications, autoimmune diseases, and other unknown factors.
Year introduced: 1998
3: Hepatitis Viruses
Links
Any of the viruses that cause inflammation of the liver. They include both DNA and RNA viruses as well viruses from humans and animals.
Year introduced: HEPATITIS VIRUS, MARMOSET was see under HEPATITIS VIRUSES 1975-1985
4: Hepatitis E
Links
Acute INFLAMMATION of the LIVER in humans caused by HEPATITIS E VIRUS
Recent activity
Turn Off Clear
hepatitis (131) MeSH
» See more...


"Hepatitis, Chronic"[Mesh] AND "Child"[Mesh] - PubMed result - Windows Internet Explorer
http://www.ncbi.nlm.nih.gov/pubmed?term=%22Hepatitis,+Chronic
Google
Favorites
"Hepatitis, Chronic"[Mesh] AND "Child"[Mesh] - PubMe...
NCBI Resources How To
My NCBI Sign In
PubMed.gov
U.S. National Library of Medicine
National Institutes of Health
Search: PubMed
RSS Save search Limits Advanced search Help
"Hepatitis, Chronic"[Mesh] AND "Child"[Mesh]
Search Clear
Display Settings: Summary, 20 per page, Sorted by Recently Added
Send to:
Results: 1 to 20 of 1750
<< First < Prev Page 1 Next > Last >>
[Treatment of viral hepatitis in children]
1. Jaklin Kekez A.
Acta Med Croatica. 2009 Dec;63(5):459-62. Croatian.
PMID: 20198908 [PubMed - indexed for MEDLINE]
Related articles
[The level of TNF-alpha secretion of PBMC in patients with chronic hepatitis C and nonalcoholic fatty liver]
2. Dong Y, Zhang HF, Zhu SS, Chen H, Li J, Cheng Y.
Zhonghua Shi Yan He Lin Chuang Bing Du Xue Za Zhi. 2009 Apr;23(2):109-11. Chinese.
PMID: 20104750 [PubMed - indexed for MEDLINE]
Related articles
Helicobacter species DNA in liver and gastric tissues in children and adolescents with chronic liver disease.
3. Casswall TH, Németh A, Nilsson I, Wadström T, Nilsson HO.
Scand J Gastroenterol. 2010;45(2):160-7.
PMID: 20095882 [PubMed - indexed for MEDLINE]
Related articles
[Gene polymorphism of transforming growth factor beta1 (TGF-beta1) in the pathogenesis and clinical course of chronic hepatitis in children]
4.
Filter your results:
All (1750)
Review (193)
Free Full Text (246)
Manage Filters
Titles with your search terms
Pegylated interferon and ribavirin combination therapy for chror [Scand J Gastroenterol. 2008]
Clinical course of pregnant women with chronic hepatitis C virus i [Dig Liver Dis. 2001]
Use of PEG-interferon alfa-2a plus ribavirin as treatment for chror [Pediatr Blood Cancer. 2004]
» See more...
111 free full-text articles in PubMed Central
Review Update on autoimmune hepatitis. [World J Gastroenterol. 2009]
Review Hepatitis B immunization strategies: timing is everything. [CMAJ. 2009]
Vitamin E treatment for children with chronic

The accession number

It is a unique number assigned to a piece of DNA when it was first submitted to GenBank and is permanently associated with that sequence.

This is the number that should be cited in publications.

It has two different formats: two letters with five digits or one letter with six digits.

Header

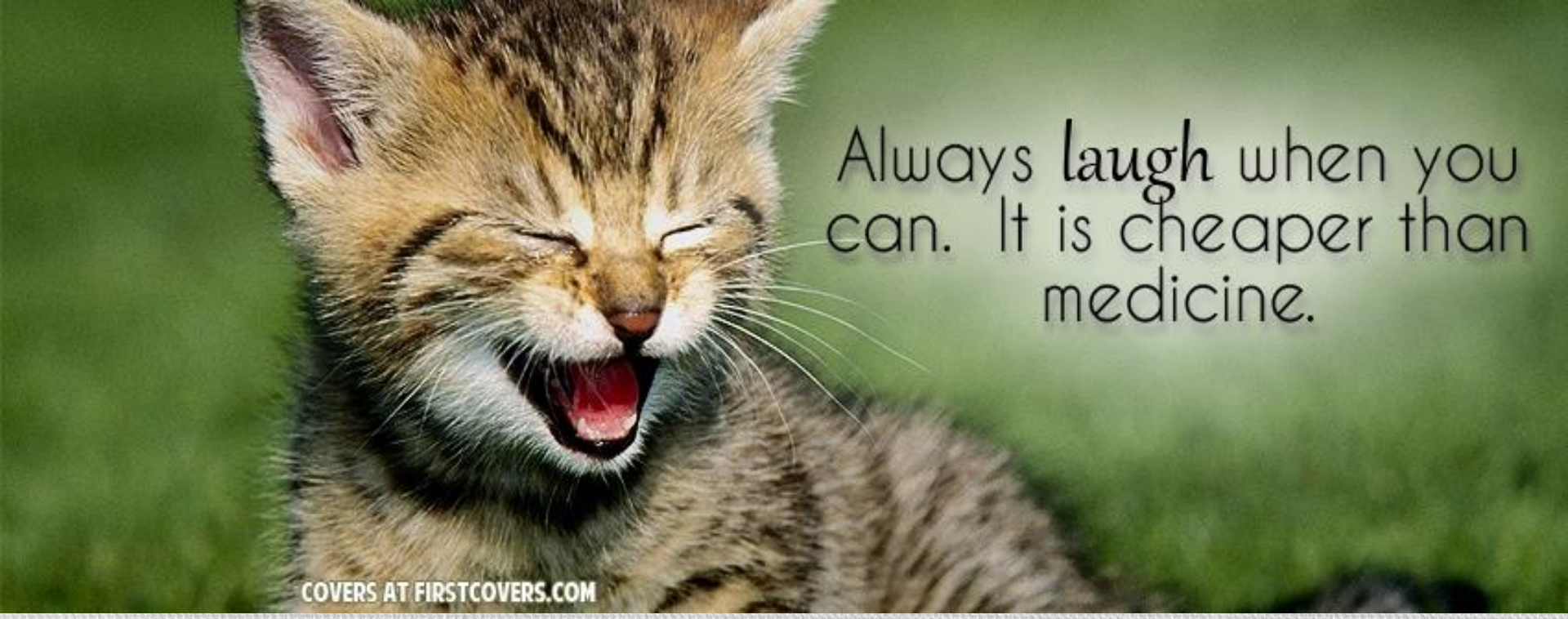
```
LOCUS       Q9ZGE9                   440 aa            linear   BCT 15-JUN-2002
DEFINITION  Light-independent protochlorophyllide reductase subunit N (LI-POR
            subunit N) (DPOR subunit N).
ACCESSION   Q9ZGE9
VERSION     Q9ZGE9  GI:18203677
DBSOURCE    swissprot: locus BCHN_HELMO, accession Q9ZGE9;
            class: standard.
            created: Oct 16, 2001.
            sequence updated: Oct 16, 2001.
            annotation updated: Jun 15, 2002.
            xrefs: gi: 3820536, gi: 3820556
            xrefs (non-sequence databases): InterProIPR000510, PfamPF00148
KEYWORDS    Photosynthesis; Bacteriochlorophyll biosynthesis; Oxidoreductase.
SOURCE      Heliobacillus mobilis
  ORGANISM  Heliobacillus mobilis
            Bacteria; Firmicutes; Clostridia; Clostridiales; Heliobacteriaceae;
            Heliobacillus.
REFERENCE   1  (residues 1 to 440)
  AUTHORS   Xiong,J., Inoue,K. and Bauer,C.E.
  TITLE     Tracking molecular evolution of photosynthesis by characterization
            of a major photosynthesis gene cluster from Heliobacillus mobilis
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 95 (25), 14851-14856 (1998)
  MEDLINE   99061957
   PUBMED   9843979
  REMARK    SEQUENCE FROM N.A.
COMMENT     -------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            -------------------------------------------------------------------.
            [FUNCTION] Uses Mg-ATP and reduced ferredoxin to reduce ring D of
            protochlorophyllide (Pchlide) to form chlorophyllide a (Chlide) (By
            similarity). This reaction is light-independent.
            [PATHWAY] Light-independent bacteriochlorophyll biosynthesis.
            [SUBUNIT] Protochlorophyllide reductase is thought to be composed
            of three subunits; bchL, bchN and bchB. Could form a heterotetramer
            of two bchB and two bchN subunits.
            [SIMILARITY] BELONGS TO THE BCHN / CHLN FAMILY.
```

Features

```
FEATURES             Location/Qualifiers
     source          1..440
                     /organism="Heliobacillus mobilis"
                     /db_xref="taxon:28064"
     gene            1..440
                     /gene="BCHN"
     Protein         1..440
                     /gene="BCHN"
                     /product="Light-independent protochlorophyllide reductase
                     subunit N"
                     /EC_number="1.18.-.-"
```

Sequence

```
ORIGIN
        1 merverengc fhtfcpiasv awlhrkikds fflivgthtc ahfiqtaldv mvyahsrfgf
       61 avleesdlvs aspteelgkv vqqvvdewhp kvifvlstcs vdilkmdlev sckdlstrfg
      121 fpvlpastsg idrsftqged avlhallpfv pkeapavepv eekkprwfsf gkesekekae
      181 parnlvliga vtdstiqqlq welkqlglpk vdvfpdgdir kmpvineqtv vvplqpylnd
      241 tlatirrerr akvlstvfpi gpdgtarfle aiclefgldt srikekeaqa wrdlepqlqi
      301 lrgkkimflg dnllelplar fltscdvqvv eagtpyihsk dlqqelellk erdvrivesp
      361 dftk[illegible] [illegible] [illegible]pleam gfttawsief tfaqihgfvn aidliklftk
      421 pllkrqalme hgwaeagwle
//
```



Thanks a lot

with my Best Regards and My Best wishes

Amira A. AL-Hosary
E-mail: Amiraelhosary @yahoo.com
Mob. (002) 01004477501