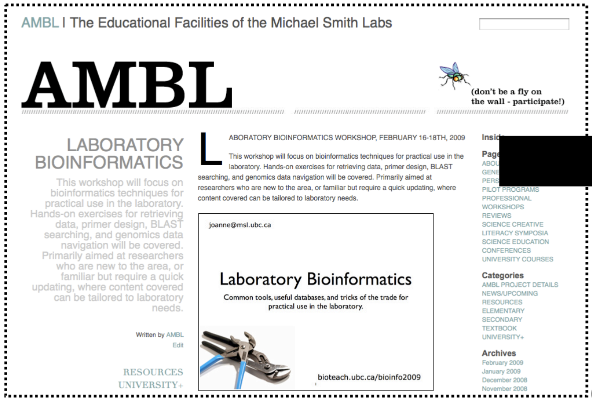
BLAST

PRACTICAL EXERCISE: The Jurassic Park Detective Story

navigate to:
bioteach.ubc.ca/
bioinfo2009#BLASTexercises



AMBL | The Educational Facilities of the Michael Smith Labs

AMBL

(don't be a fly on
the wall - participate!)

LABORATORY BIOINFORMATICS

This workshop will focus on bioinformatics techniques for practical use in the laboratory. Hands-on exercises for retrieving data, primer design, BLAST searching, and genomics data navigation will be covered. Primarily aimed at researchers who are new to the area, or familiar but require a quick updating, where content covered can be tailored to laboratory needs.

Written by AMBL
Edit

RESOURCES
UNIVERSITY+

LABORATORY BIOINFORMATICS WORKSHOP, FEBRUARY 16-18TH, 2009

This workshop will focus on bioinformatics techniques for practical use in the laboratory. Hands-on exercises for retrieving data, primer design, BLAST searching, and genomics data navigation will be covered. Primarily aimed at researchers who are new to the area, or familiar but require a quick updating, where content covered can be tailored to laboratory needs.

joanne@msl.ubc.ca

Laboratory Bioinformatics

Common tools, useful databases, and tricks of the trade for practical use in the laboratory.

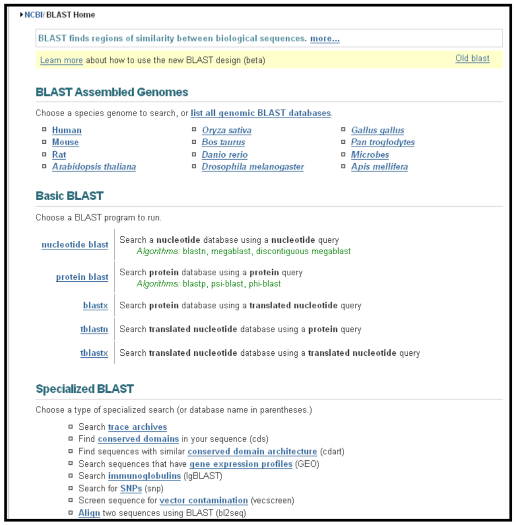
bioteach.ubc.ca/bioinfo2009

Insid

Page
ABOU
GENE
PERS
PILOT PROGRAMS
PROFESSIONAL
WORKSHOPS
REVIEWS
SCIENCE CREATIVE
LITERACY SYMPOSIA
SCIENCE EDUCATION
CONFERENCES
UNIVERSITY COURSES

Categories
AMBL PROJECT DETAILS
NEWS/UPCOMING
RESOURCES
ELEMENTARY
SECONDARY
TEXTBOOK
UNIVERSITY+

Archives
February 2009
January 2009
December 2008
November 2008



NCBI/ BLAST Home

BLAST finds regions of similarity between biological sequences. more...

Learn more about how to use the new BLAST design (beta) Old blast

BLAST Assembled Genomes

Choose a species genome to search, or list all genomic BLAST databases.

- Human
- Mouse
- Rat
- *Arabidopsis thaliana*
- *Oryza sativa*
- *Bos taurus*
- *Danio rerio*
- *Drosophila melanogaster*
- *Gallus gallus*
- *Pan troglodytes*
- *Microbes*
- *Apis mellifera*

Basic BLAST

Choose a BLAST program to run.

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

Specialized BLAST

Choose a type of specialized search (or database name in parentheses.)

- Search trace archives
- Find conserved domains in your sequence (cds)
- Find sequences with similar conserved domain architecture (cdart)
- Search sequences that have gene expression profiles (GEO)
- Search immunoglobulins (IgBLAST)
- Search for SNPs (snp)
- Screen sequence for vector contamination (vecscreen)
- Align two sequences using BLAST (bl2seq)

Let's compare
our results

Get the sequences from the webpage and carry out BLAST searches

Can you identify the Dinosaur sequences?

Search #1:
Jurassic Park
sequence

use blastn

Search #2:
The Lost World
sequence

use blastx

Try some BLAST searches with your own sequence of interest…

Explore what happens when you change advanced parameters…

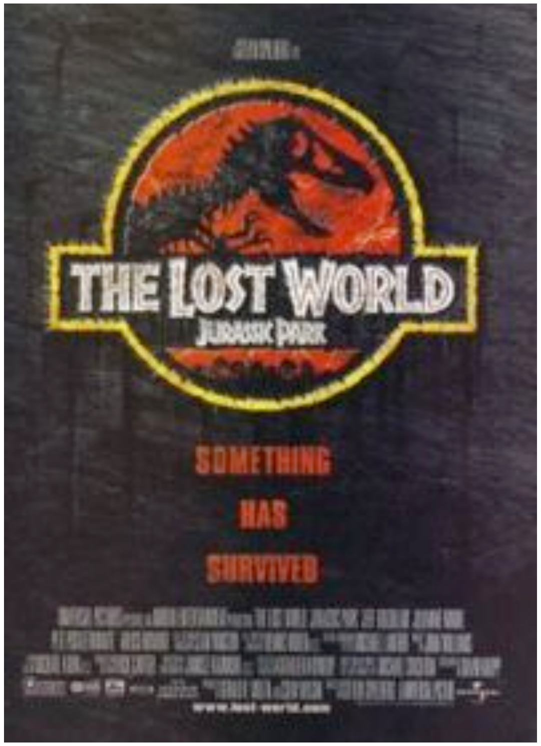
Search #1 - blastn against nr

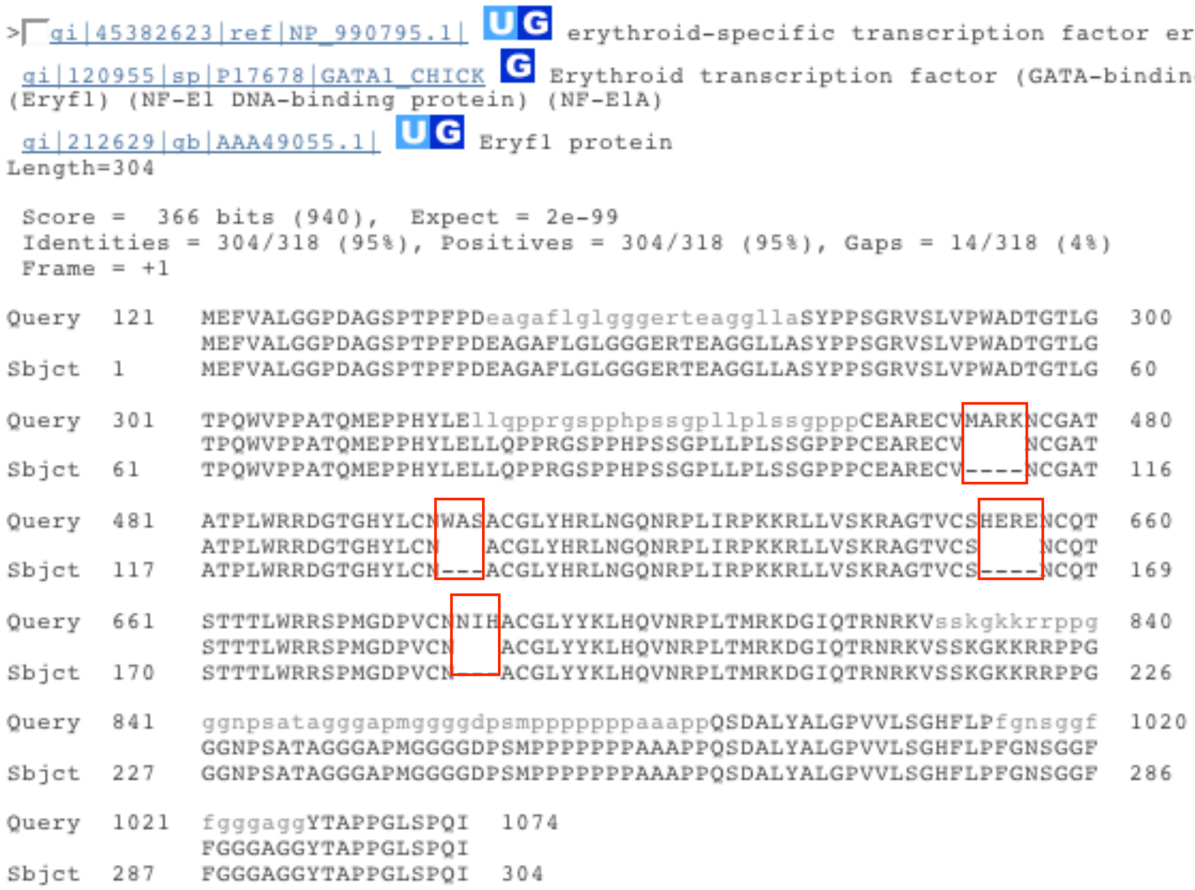
- Most common use of blastn
  - ✓ Sequence identification
  - ✓ Establish whether an exact match for a sequence is already present in the database

```
>gi|157064989|gb|EU118176.1|  Cloning vector pCM433, complete sequence
Length=8081

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

 Score =  437 bits (484),  Expect = 4e-119
 Identities = 297/340 (87%), Gaps = 40/340 (11%)
 Strand=Plus/Plus

Query  1     GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7309  GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGC  7368

Query  61    ----------GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA  110
                       ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  7369  TCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGA  7428

Query  111   AGCTCCCTCG----------TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT  160
             ||||||||||          ||||||||||||||||||||||||||||||||||||||||
Sbjct  7429  AGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTT  7488

Query  161   CTCCCTTCGGGAAGCGTGGC----------TGCTCACGCTGTACCTATCTCAGTTCGGTG  210
             ||||||||||||||||||||           ||||||||||||  |||||||||||||||
Sbjct  7489  CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTG  7548

Query  211   TAGGTCGTTCGCTCCAAGCTGGGCTGTGTG----------CCGTTCAGCCCGACCGCTGC  260
             ||||||||||||||||||||||||||||||          ||||||||||||||||||||
Sbjct  7549  TAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGC  7608

Query  261   GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAA  300
             ||||||||||||||||||||||||||||||||||||||||
Sbjct  7609  GCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAA  7648


 Score =  536 bits (594),  Expect = 6e-149
 Identities = 360/410 (87%), Gaps = 50/410 (12%)
 Strand=Plus/Plus

Query  302   GTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAG-  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Sbjct  3591  GTAGGACAGGTGCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGC  3650

Query  361   ---------ATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCC  411
                      |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3651  GCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTTGCACGCCCTCGCTCAAGCC  3710

Query  412   TTCGTCACT----------CCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATG  461
             |||||||||          |||||||||||||||||||||||||||||||||||||||||
Sbjct  3711  TTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATG  3770

Query  462   GCGGCCGACGCGCTGGGCT----------GGCGTTCGCGACGCGAGGCTGGATGGCCTTC  511
             |||||||||||||||||||          |||||||||||||||||||||||||||||||
Sbjct  3771  GCGGCCGACGCGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTC  3830

Query  512   CCCATTATGATTCTTCTCGCTTCCGGCG----------GCCCGCGTTGCAGGCCATGCTG  561
             ||||||||||||||||||||||||||||          ||||||||||||||||||||||
Sbjct  3831  CCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCCCGCGTTGCAGGCCATGCTG  3890

Query  562   TCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAA----------CGGCTCTTACC  611
             |||||||||||||||||||||||||||||||||||||||          |||||||||||
Sbjct  3891  TCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACC  3950

Query  612   AGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGC  661
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3951  AGCCTAACTTCGATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGC  4000
```

Search #2 - blastx against nr

- Translating BLAST programs (blastx, tblastn, tblastx)
  - ✓ Look for similar proteins
  - ✓ Identify potential homologs in other species





```
>gi|45382623|ref|NP_990795.1| UG erythroid-specific transcription factor eryf1 [Gallus gallus]
 gi|120955|sp|P17678|GATA1_CHICK G Erythroid transcription factor (GATA-binding factor 1) (GATA-1)
(Eryf1) (NF-E1 DNA-binding protein) (NF-E1A)
 gi|212629|gb|AAA49055.1| UG Eryf1 protein
Length=304

 Score =  366 bits (940),  Expect = 2e-99
 Identities = 304/318 (95%), Positives = 304/318 (95%), Gaps = 14/318 (4%)
 Frame = +1

Query  121   MEFVALGGPDAGSPTPFPDeagaflglgggerteaggllaSYPPSGRVSLVPWADTGTLG  300
             MEFVALGGPDAGSPTPFPDEAGAFLGLGGGERTEAGGLLASYPPSGRVSLVPWADTGTLG
Sbjct  1     MEFVALGGPDAGSPTPFPDEAGAFLGLGGGERTEAGGLLASYPPSGRVSLVPWADTGTLG  60

Query  301   TPQWVPPATQMEPPHYLEllqpprgspphpssgpllplssgpppCEARECVMARKNCGAT  480
             TPQWVPPATQMEPPHYLELLQPPRGSPPHPSSGPLLPLSSGPPPCEARECV    NCGAT
Sbjct  61    TPQWVPPATQMEPPHYLELLQPPRGSPPHPSSGPLLPLSSGPPPCEARECV----NCGAT  116

Query  481   ATPLWRRDGTGHYLCNWASACGLYHRLNGQNRPLIRPKKRLLVSKRAGTVCSHERENCQT  660
             ATPLWRRDGTGHYLCN   ACGLYHRLNGQNRPLIRPKKRLLVSKRAGTVCS    NCQT
Sbjct  117   ATPLWRRDGTGHYLCN---ACGLYHRLNGQNRPLIRPKKRLLVSKRAGTVCS----NCQT  169

Query  661   STTTLWRRSPMGDPVCNNIHACGLYYKLHQVNRPLTMRKDGIQTRNRKVsskgkkrrppg  840
             STTTLWRRSPMGDPVCN   ACGLYYKLHQVNRPLTMRKDGIQTRNRKVSSKGKKRRPPG
Sbjct  170   STTTLWRRSPMGDPVCN---ACGLYYKLHQVNRPLTMRKDGIQTRNRKVSSKGKKRRPPG  226

Query  841   ggnpsatagggapmggggdpsmppppppppaaappQSDALYALGPVVLSGHFLPfgnsggf  1020
             GGNPSATAGGGAPMGGGGDPSMPPPPPPPPAAAPPQSDALYALGPVVLSGHFLPFGNSGGF
Sbjct  227   GGNPSATAGGGAPMGGGGDPSMPPPPPPPPAAAPPQSDALYALGPVVLSGHFLPFGNSGGF  286

Query  1021  fgggaggYTAPPGLSPQI  1074
             FGGGAGGYTAPPGLSPQI
Sbjct  287   FGGGAGGYTAPPGLSPQI  304
```

Mark was here, NIH