APPENDIX 2 - BIOINFORMATICS (PARTS I AND II)

HC70AL Spring 2004

An Introduction to Bioinformatics -- Part I

By

Brandon Le

April 6, 2004

What is a Gene?

An <u>ordered</u> sequence of nucleotides

What are the 4 Nucleotides in DNA?

A - Adenine
T - Thymine
C - Cytosine
G - Guanine

What are the Characteristics of a Gene?

· **An <u>ordered</u> sequence of nucleotides**
· **A unique position/location in the genome**
· **Polarity (5' to 3')**
· **Exons and Introns**

What are the Anatomical Features of Genes?

· **Discrete beginning and discrete end**
· **Two strands of DNA**
· **Double helical**
· **Strand one (5' to 3')**
· **Strand two (3' to 5')**
· **Sense strand (5' to 3')**
     - specifies the trait
· **Nonsense strand (3' to 5')**
     - template for transcription

Sense Strand

```
5' – ACGTCAGTCGATGCATGCTAGCTAGC – 3'
3' – TGCAGTCAGCTACGTACGATCGATCG – 5'
```

Nonsense Strand

Genes Have a Unique Position in the Genome!

Task: Where is your gene located in the genome?

Tools: The Arabidopsis Information Resources (TAIR)
(http://www.arabidopsis.org)

01	AT2G22800
02	AT2G23290
03	AT2G37120
04	AT3G09735
05	AT3G12840
06	AT3G50060
07	AT3G53370
08	AT4G37260
09	AT4G37790
10	AT5G03220
11	AT5G03500
12	AT5G19490
13	AT5G67300

Procedure:

1. **Select Seqviewer**
2. **Enter gene number (ex. AT1G18260)**
3. **Submit**

Results/Question:

1. **What chromosome is your gene in?**
2. **What other genes/markers are next to your gene?**
3. **What is the exact position of your gene in the genome?**

Genes Have a Unique Order of Nucleotides!

Task: What is the order of nucleotides for your gene?

Tools: The Arabidopsis Information Resources (TAIR)
(http://www.arabidopsis.org)

Procedure: **(Continue from previous slide)**

1. **Click on Location**

Results/Question:

1. **What are your neighbor genes?**
2. **What is the orientation of your gene?**
3. **How big is your gene?**

Genes Have Exons and Introns!

Task: How many exons and introns does your gene have?

Tools: The Arabidopsis Information Resources (TAIR)
(http://www.arabidopsis.org)

Procedure: (Continue from previous slide)

1. Click on gene information on the right

Results/Question:

1. How many exons/introns in your gene?
2. What are exons?
3. What are introns?

Gene Encodes a Protein

Task: Determine the protein encoded by gene?

Tools: The Arabidopsis Information Resources (TAIR)
(http://www.arabidopsis.org)

Results/Question:

1. How large is your protein?
2. What are the anatomy of a protein?

N-terminal   [red bar]   C-terminal

What is the identity of your gene?

Task: What does your gene code for?

Tools: NCBI BLAST Tools
(http://www.ncbi.nlm.nih.gov/BLAST)

What is BLAST?

Basic Local Alignment Search Tool (BLAST)

What does BLAST do?

A family of programs that allows you to input a query sequence and compare it to DNA or protein sequences in db.

What are the steps to performing BLAST search?

Paste sequence of interest into BLAST input box
Select BLAST program
Select db
Select Optional Parameters

What are the different BLAST Programs?

Fastest

blastp - protein query vs protein db

blastn - DNA query vs DNA db

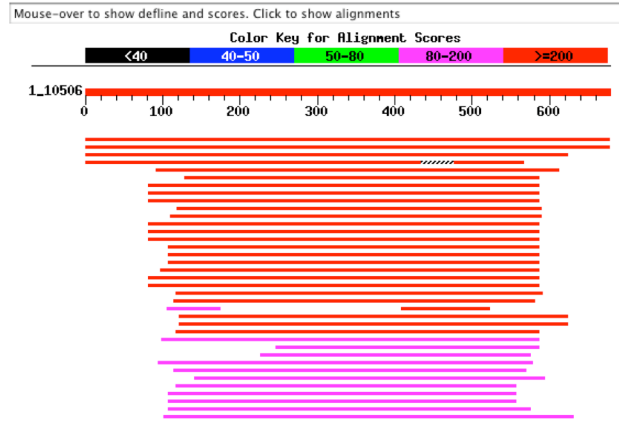**blastx - translated DNA query vs protein db**

tblastx - protein query vs translated DNA db

Slowest

tblastn - translated DNA query vs translated DNA db

Anatomy of a BLAST Result -- Part I

Distribution of 339 Blast Hits on the Query Sequence

Anatomy of a BLAST Result -- Part II

```
                                                                  Score      E
Sequences producing significant alignments:                      (bits)   Value

gi|14532716|gb|AAK64159.1|    unknown protein [Arabidopsis tha...   1206    0.0
gi|18394588|ref|NP_564049.1|    suppressor of lin-12-like prot...   1209    0.0
gi|15219499|ref|NP_177498.1|    suppressor of lin-12-like prot...    877    0.0
gi|11120786|gb|AAG30966.1|    hypothetical protein, 3' partial...    426    e-118
gi|41151276|ref|XP_046437.5|    chromosome 20 open reading fra...    291    3e-77  L
gi|13559241|emb|CAB65792.2|    dJ842G6.2 (novel protein imilar...    282    2e-74  L
gi|19923669|ref|NP_005056.3|    sel-1 suppressor of lin-12-lik...    268    4e-70  L
gi|6851089|gb|AAF29413.1|    SEL1L [Homo sapiens] >gi|17646138...    268    4e-70  L
gi|9967440|dbj|BAB12403.1|    SEL1L [Mesocricetus auratus]           264    4e-69
gi|31203035|ref|XP_310466.1|    ENSANGP00000019196 [Anopheles ...    263    1e-68
gi|21355295|ref|NP_651179.1|    CG10221-PA [Drosophila melanog...    263    1e-68  L
gi|20857527|ref|XP_127076.1|    Sel1 (suppressor of lin-12) 1 ...    261    4e-68  L
gi|4159995|gb|AAD05210.1|    SEL1L [Mus musculus] >gi|20073079...    259    1e-67  L
gi|29336095|ref|NP_808794.1|    Sel1 (suppressor of lin-12) 1 ...    259    2e-67  L
gi|29612522|gb|AAH49959.1|    Sel1h protein [Mus musculus]          258    4e-67  L
gi|17563256|ref|NP_506144.1|    Suppressor/Enhancer of Lin-12 ...    247    9e-64  L
gi|1255199|gb|AAC47112.1|    sel-1 gene product                      247    9e-64
```

Anatomy of a BLAST Result -- Part III

```
>gi|14532716|gb|AAK64159.1|    unknown protein [Arabidopsis thaliana]
          Length = 678

 Score = 1206 bits (3120), Expect = 0.0
 Identities = 614/678 (90%), Positives = 614/678 (90%)

Query: 1     MRILSYGIVILSLLVFSFIEFGVHARPVVLVXXXXXXXXXXXXXXXVXXXXXXXXXXXXXX 60
             MRILSYGIVILSLLVFSFIEFGVHARPVVLV              V
Sbjct: 1     MRILSYGIVILSLLVFSFIEFGVHARPVVLVLSNDDLNSGGDDNGVGESSDFDEFGESEP 60

Query: 61    XXXXXLDPGSWRSIFEPDDSTVQAASPQYYSGLKKILSAASEGNFRLMEEAVDEIEAASS 120
                  LDPGSWRSIFEPDDSTVQAASPQYYSGLKKILSAASEGNFRLMEEAVDEIEAASS
Sbjct: 61    KSEEELDPGSWRSIFEPDDSTVQAASPQYYSGLKKILSAASEGNFRLMEEAVDEIEAASS 120

Query: 121   AGDPHAQSIMGFVYGIGMMREKSKSKSFLHHNFAAAGGNMQSKMALAFTYLRQDMHDKAV 180
             AGDPHAQSIMGFVYGIGMMREKSKSKSFLHHNFAAAGGNMQSKMALAFTYLRQDMHDKAV
Sbjct: 121   AGDPHAQSIMGFVYGIGMMREKSKSKSFLHHNFAAAGGNMQSKMALAFTYLRQDMHDKAV 180

Query: 181   QLYAELAETAVNSFLISKDSPVVEPTRIHSGTEENKGALRKSRGEEDEDFQILEYQAQKG 240
             QLYAELAETAVNSFLISKDSPVVEPTRIHSGTEENKGALRKSRGEEDEDFQILEYQAQKG
Sbjct: 181   QLYAELAETAVNSFLISKDSPVVEPTRIHSGTEENKGALRKSRGEEDEDFQILEYQAQKG 240

Query: 241   NANAMYKIGLFYYFGLRGLRRDHTKALHWFLKAVDKGEPRSMELLGEIYARGAGVERNYT 300
             NANAMYK GLFYYFGLRGLRRDHTKALHWFLKAVDKGEPRSMELLGEIYARGAGVERNYT
Sbjct: 241   NANAMYKNGLFYYFGLRGLRRDHTKALHWFLKAVDKGEPRSMELLGEIYARGAGVERNYT 300
```
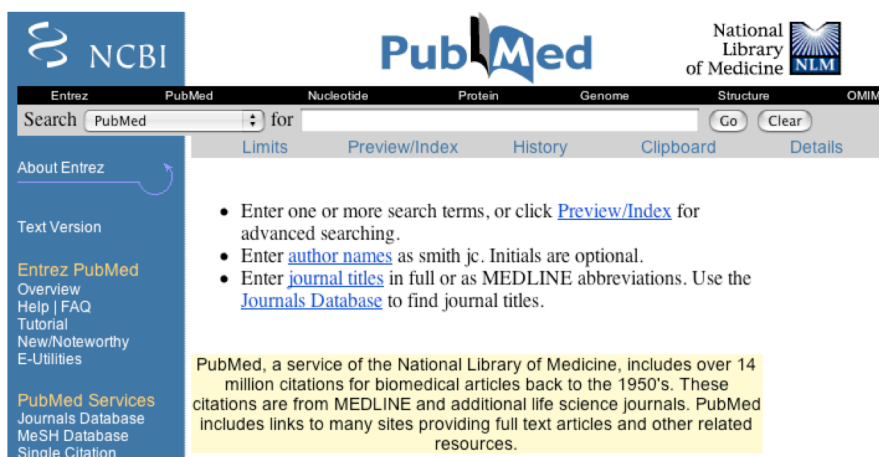
PubMed - Endless Resources

NCBI   PubMed   National Library of Medicine NLM

Entrez   PubMed   Nucleotide   Protein   Genome   Structure   OMIM

Search [PubMed] for [   ] (Go) (Clear)

Limits   Preview/Index   History   Clipboard   Details

About Entrez

Text Version

Entrez PubMed
Overview
Help | FAQ
Tutorial
New/Noteworthy
E-Utilities

PubMed Services
Journals Database
MeSH Database
Single Citation

- Enter one or more search terms, or click Preview/Index for advanced searching.
- Enter author names as smith jc. Initials are optional.
- Enter journal titles in full or as MEDLINE abbreviations. Use the Journals Database to find journal titles.

PubMed, a service of the National Library of Medicine, includes over 14 million citations for biomedical articles back to the 1950's. These citations are from MEDLINE and additional life science journals. PubMed includes links to many sites providing full text articles and other related resources.

HC70AL Spring 2004

An Introduction to Bioinformatics -- Part II

By
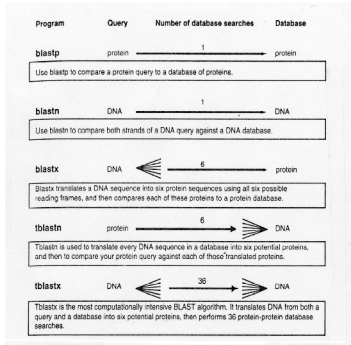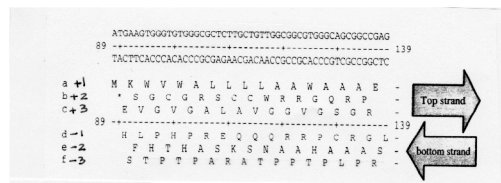
Brandon Le

April 8, 2004

Review of BLAST Search

1. What is the purpose of running BLAST Search?

2. What are the steps to performing BLAST search?

3. What does the e-value from a blast result tell you?

4. How may BLAST program can you perform?

5. What BLAST program(s) takes the least computational time?

6. What BLAST program(s) takes the most computational time? Why?

What are the Five BLAST Search Programs?

Program	Query	Number of database searches	Database
blastp	protein	1	protein

Use blastp to compare a protein query to a database of proteins.

blastn	DNA	1	DNA

Use blastn to compare both strands of a DNA query against a DNA database.

blastx	DNA	6	protein

Blastx translates a DNA sequence into six protein sequences using all six possible reading frames, and then compares each of these proteins to a protein database.

tblastn	protein	6	DNA

Tblastn is used to translate every DNA sequence in a database into six potential proteins, and then to compare your protein query against each of those translated proteins.

tblastx	DNA	36	DNA

Tblastx is the most computationally intensive BLAST algorithm. It translates DNA from both a query and a database into six potential proteins, then performs 36 protein-protein database searches.

·How many proteins can a short DNA sequence **potentially** encode?

```
              ATGAAGTGGGTGTGGGCGCTCTTGCTGTTGGCGGCGTGGGCAGCGGCCGAG
   89 -----------+----------+----------+----------+--------- 139
              TACTTCACCCACACCCGCGAGAACGACAACCGCCGCACCCGTCGCCGGCTC

   a +1   M  K  W  V  W  A  L  L  L  L  A  A  W  A  A  A  E  -
   b +2   *  S  G  C  G  R  S  C  C  W  R  R  G  Q  R  P   -     Top strand
   c +3   E  V  G  V  G  A  L  A  V  G  G  V  G  S  G  R   -
   89 -+----------+----------+----------+----------+--------- 139
   d -1   H  L  P  H  P  R  E  Q  Q  R  R  P  C  R  G  L  -
   e -2     F  H  T  H  A  S  K  S  N  A  A  H  A  A  A  S  -    bottom strand
   f -3     S  T  P  T  P  A  R  A  T  P  P  T  P  L  P  R   -
```

Question:

You have <u>DNA</u> Sequence. You want to know which protein in the main <u>protein</u> database is most similar to some <u>protein</u> encoded by your DNA.

Which BLAST program should you use?

Suppose you have a <u>protein</u> sequence.
Which BLAST program should you use?

HOW to interpret BLAST results?

Review of gene transcription

1. What product is made after transcription?

2. How is the product similar/different from the gene?

3. What is cDNA?

4. What important information does a cDNA tell you about a gene?

5. What are ESTs?

6. What important information does ESTs tell you about a gene?

Annotation of your gene

1. What chromosome is your gene in?
2. How "big" is your gene?
3. How many exons and introns in your gene?
4. What orientation is your gene in the genome?
5. What is the specific position of your gene in the genome?
6. What gene is "upstream" of your gene?
7. What gene is "downstream" of your gene?
8. How far are the other genes (6 & 7) from your gene?
9. What is the "structure" of your gene?
10. What is the size of the protein in your gene encodes?
11. What protein does your gene encode
12. Is your gene structure predicted by a program?

Webbook -
A Virtual Lab Notebook

Webbook is a web lab notebook

Purpose/goal: To have access to experiments carried out b
Lab members, etc… from anywhere
Also serves as a repository for protocols, stocks/reagents

Created by: Harry Hahn
Brandon Le
Bob Goldberg

http//estdb.biology.ucla.edu/webbook

Using the Webboook

1. Username: email username
    Password: 9 digit student id

2. Check message board for important news/updates

3. An overview of the different sections

    Projects   - list of experiments

    Stocks   - catalog of stocks/reagent in the lab

    Protocols - procedures carried out in the lab (pdf format)

    Calendar   - calendar to plant your experiments

    Browse   - search and look at other members experiments

    Contact   - email for help

    Logout   - will logout if idle for 30 min

Webbook Login Page

webBOOK Login

Username: ble   Password: [   ]   (Login)

Help Login

Last modified August 03 2003 21:16:09.

Copyrighted by the University of California (2003)

Created by Harry Hahn and Brandon Le, Laboratory of Bob Goldberg, UCLA

Creating Projects / Experiments

1. **Title of project**

2. **Questions/Purpose of project**

3. **Summary of project (ideas)**

Entering Gene Information

Entering Experiments Information
Part 1

Experiments

Fields marked with a red asterisk (*) are **REQUIRED**

Title:*	
Goal:*	
Background Info:*	
Approach:*	
Controls:*	
Discussion:	
Next:	

Entering Experiment Information
Part II

Materials	**Primer ***
	AT2G22800-FW AT2G22800-RV AT2G23290-FW AT2G23290-RV AT2G37120-FW AT2G37120-RV AT3G09735-FW AT3G09735-RV
Protocols:	**Protocols ***
	*Sequencing Using SPPCR Alkali Lysis Plasmid Isolation Arabidopsis Tissue Harvest For GeneChip Experiment Bacteria Chromosome Mini-Prep Bacteriophage Chromatin Immunoprecipitation with Leaves from Arabidopsis
Attach a file:	Title: File: Choose File no file selected Description:

All files must have a file name extension. Images must end in .jpg, .png, or .gif. Additional files can be attached by later editing this record.

Entering References Relating to your Gene