United States Department of Agriculture | National Institute of Food and Agriculture

Solanaceae Coordinated Agricultural Project

THE OHIO STATE UNIVERSITY OARDC

Bioinformatics 101

Part I: the tools

David Francis

The Ohio State University, OARDC

SolCAP workshop

This module Introduces some basic tools used for bioinformatics. After following this module, you should be able to:

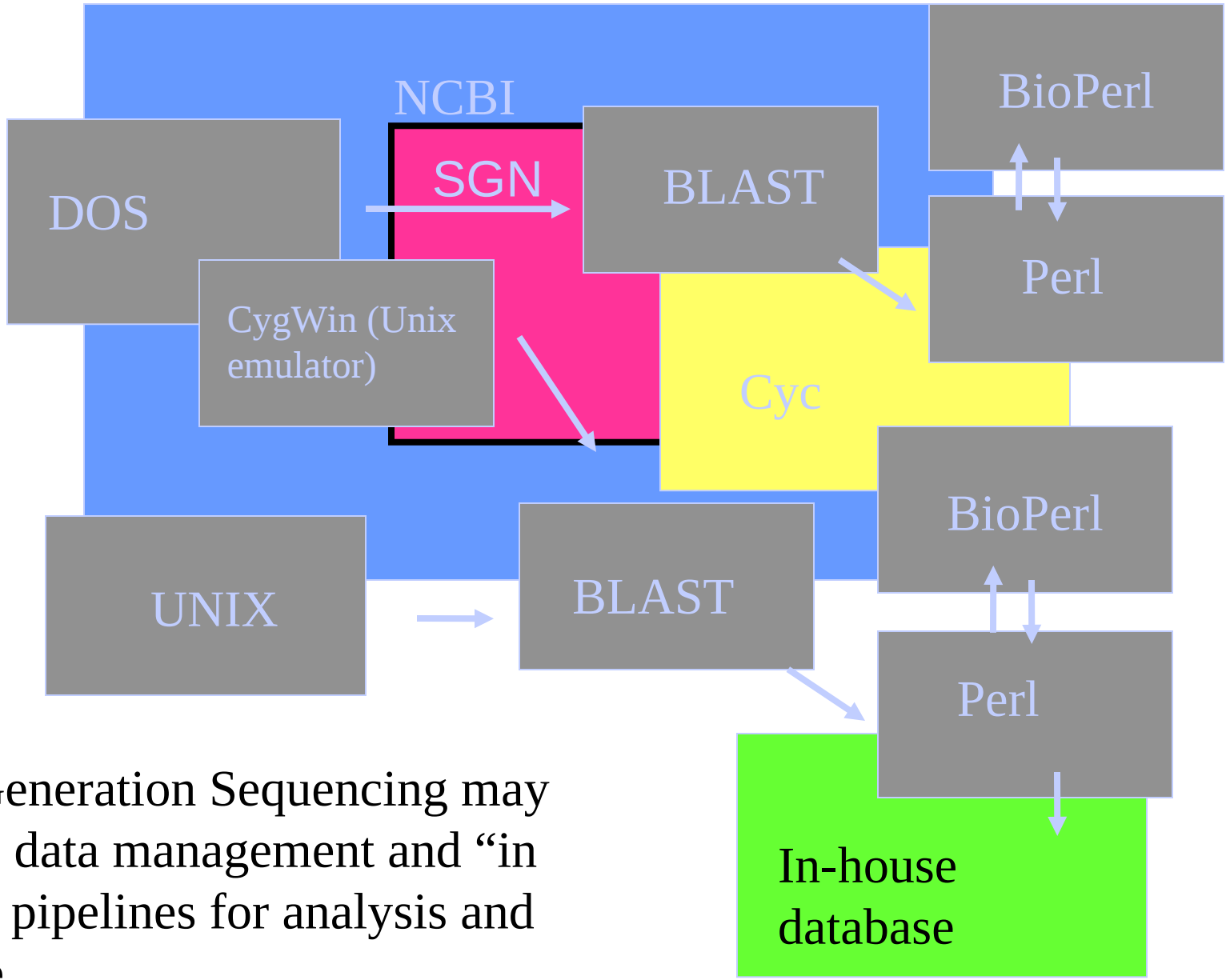
Describe the purpose of BLAST, perl and BioPerl in building pipelines for marker discovery

Find and install BLAST

Format a FASTA file as a database for BLAST searches

Perform BLAST searches

Use PERL and BioPerl to parse BLAST searches



Next Generation Sequencing may require data management and “in house” pipelines for analysis and storage

National Center for Biotechnology Information (NCBI) – 1988
BLAST – Basic Local Alignment Search Tool

Blast finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases

NCBI maintains databases that are freely available to the public for download.
YOU MAY WANT TO USE YOUR OWN OR CREATE ONE TO ADDRESS SPECIFIC QUESTIONS

MCIC BLAST ACTGATTACGAA STANDALONE

Basic Local Alignment Search Tool SETUP TUTORIAL

The BLAST Family

blastp
protein/protein

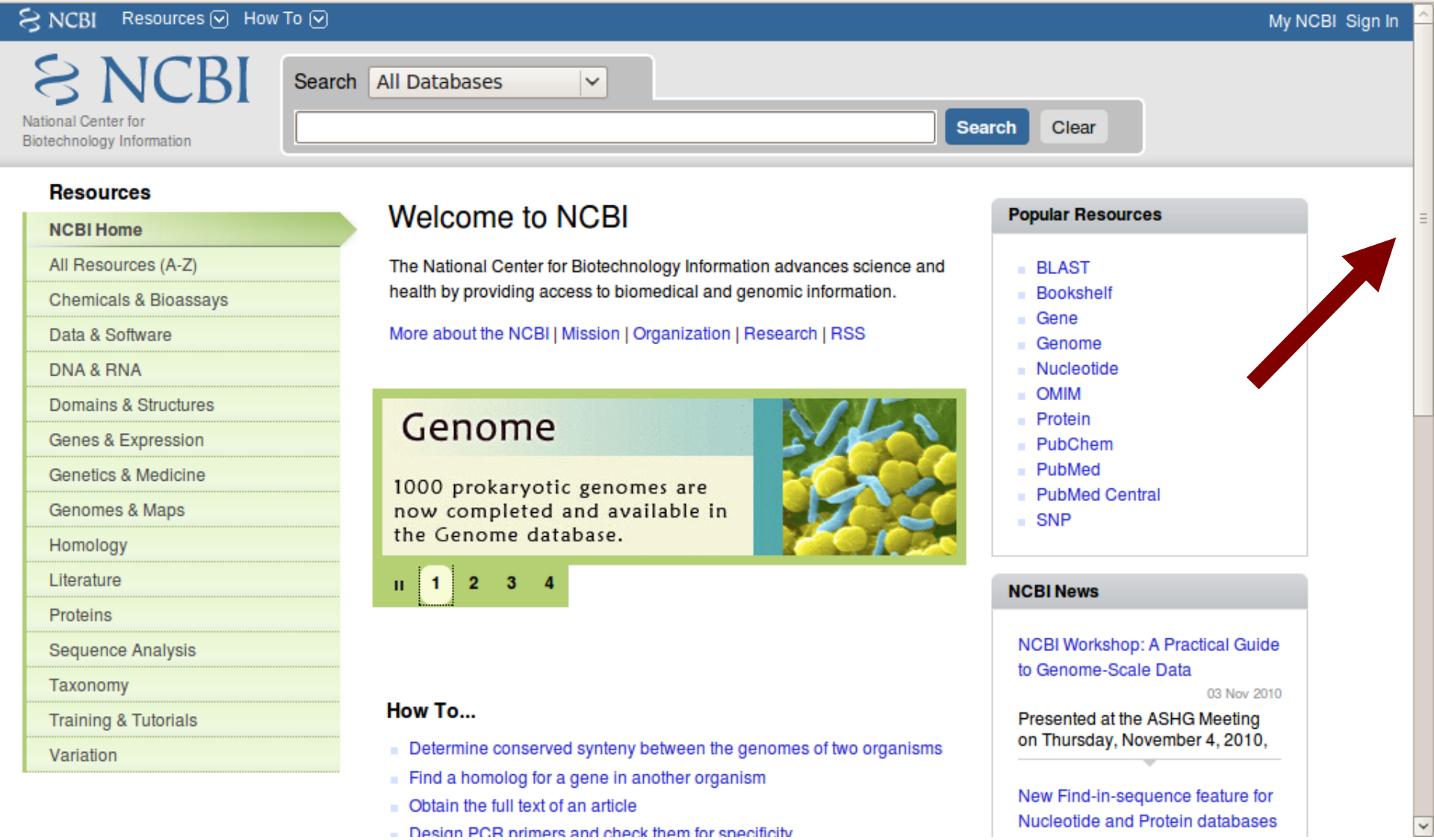
tblastx
Translated nucleotide vs Translated nucleotide

blastn
nucleotide/nucleotide

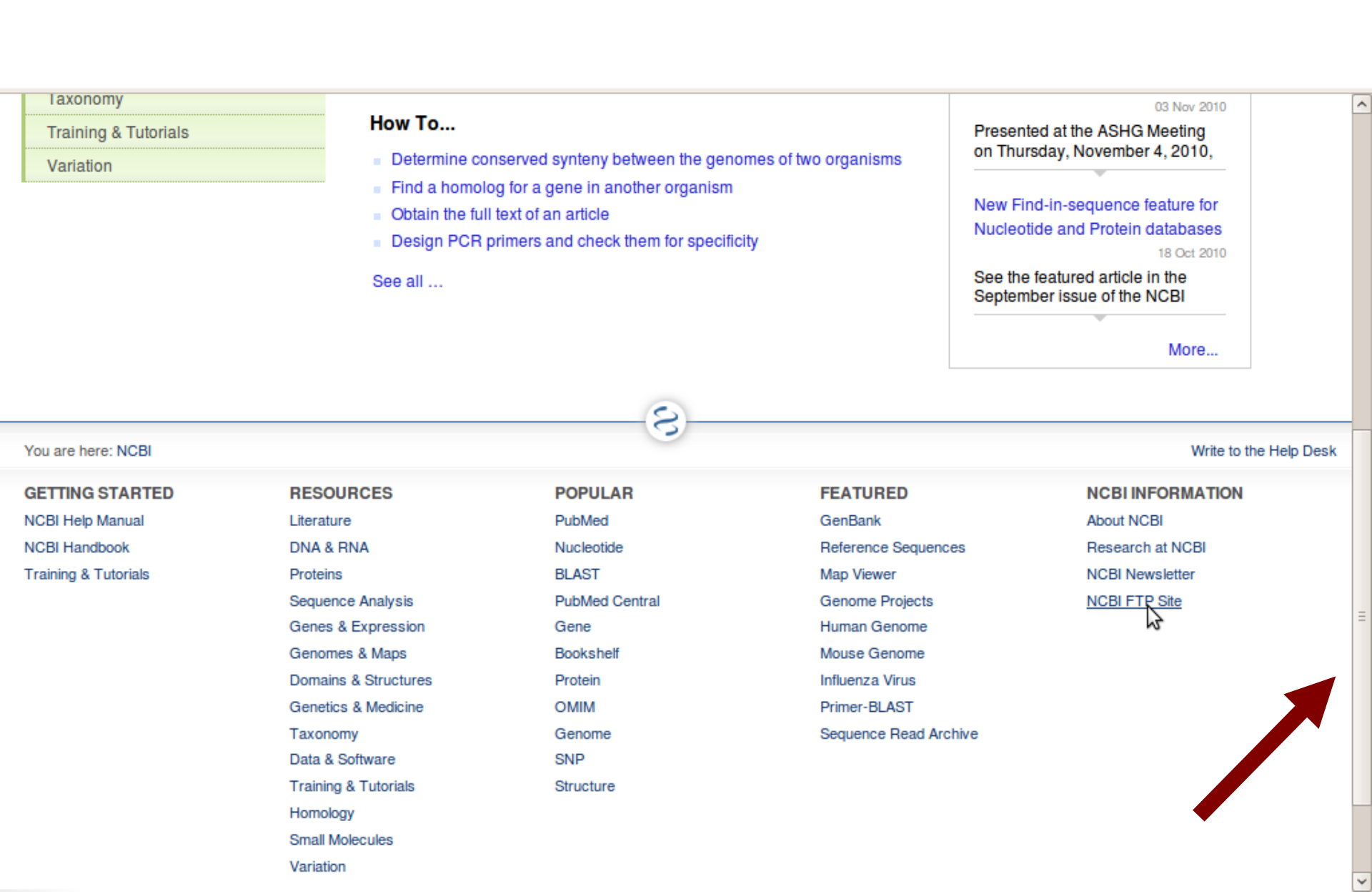
tblastn
protein/nucleotide

blastx
nucleotide/protein

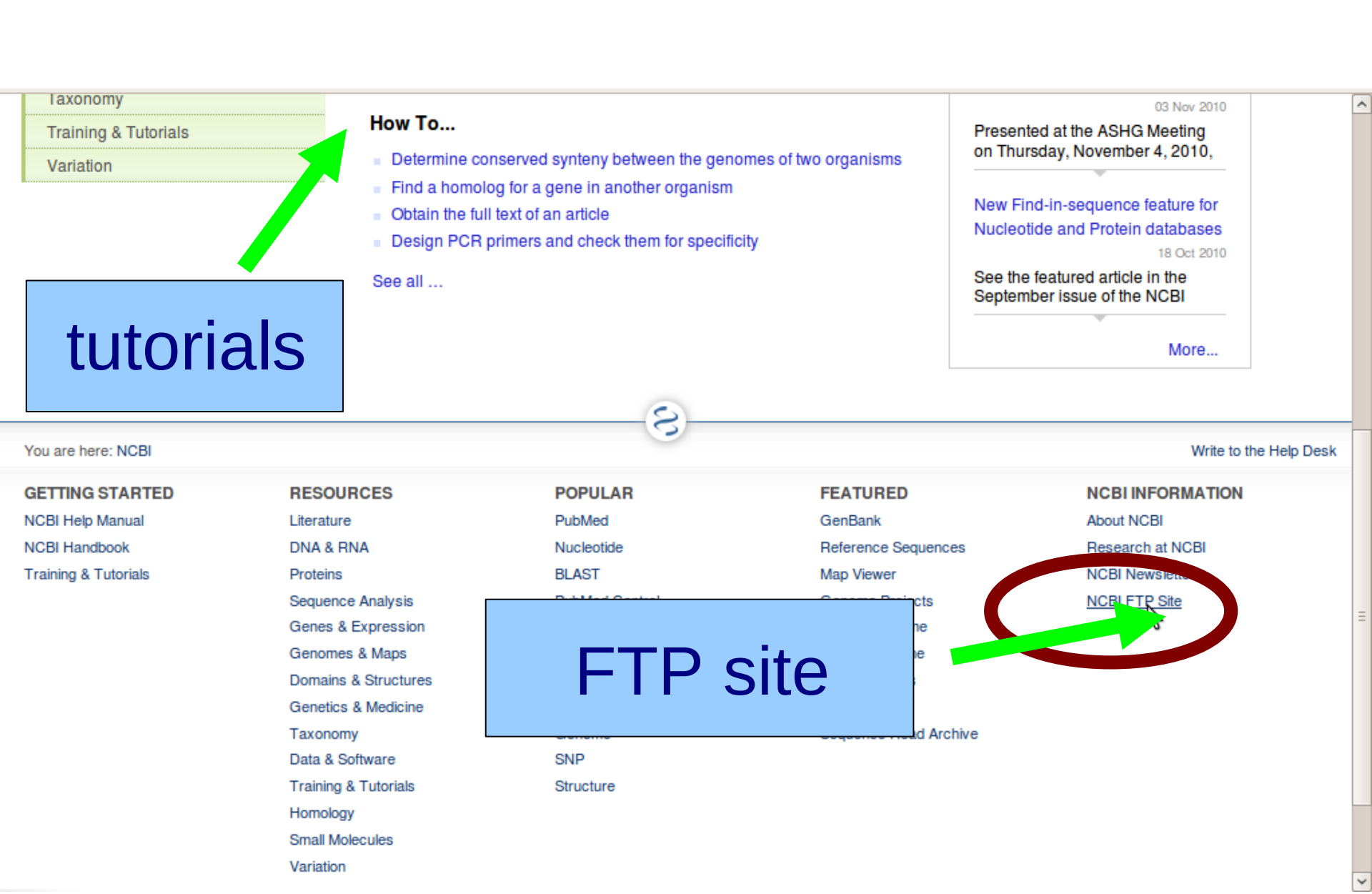
Similarity search programs



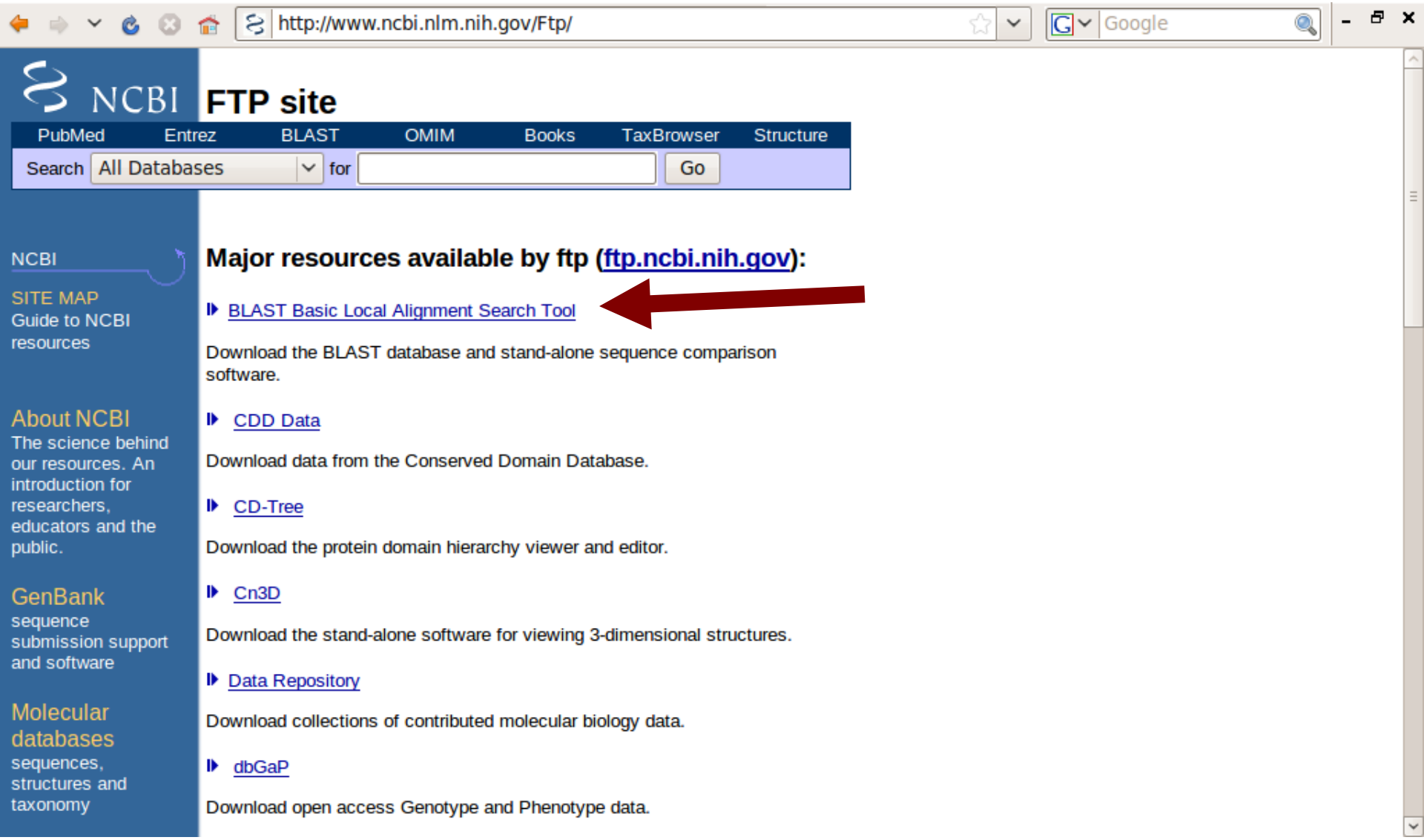
‘http://www.ncbi.nlm.nih.gov/’



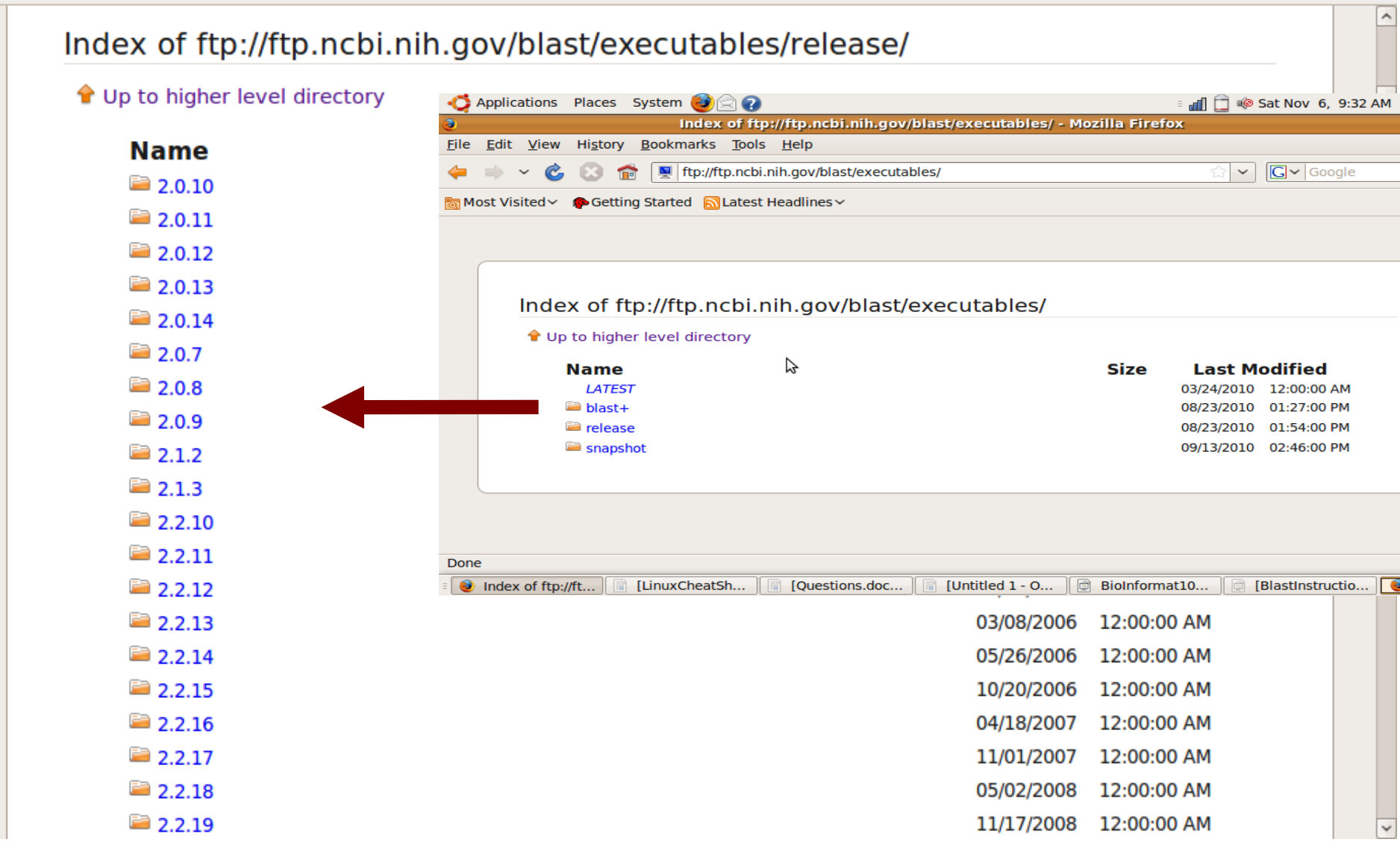
‘http://www.ncbi.nlm.nih.gov/’



‘http://www.ncbi.nlm.nih.gov/’



http://www.ncbi.nlm.nih.gov/Ftp/



Example 2.2.20 (April 2009)

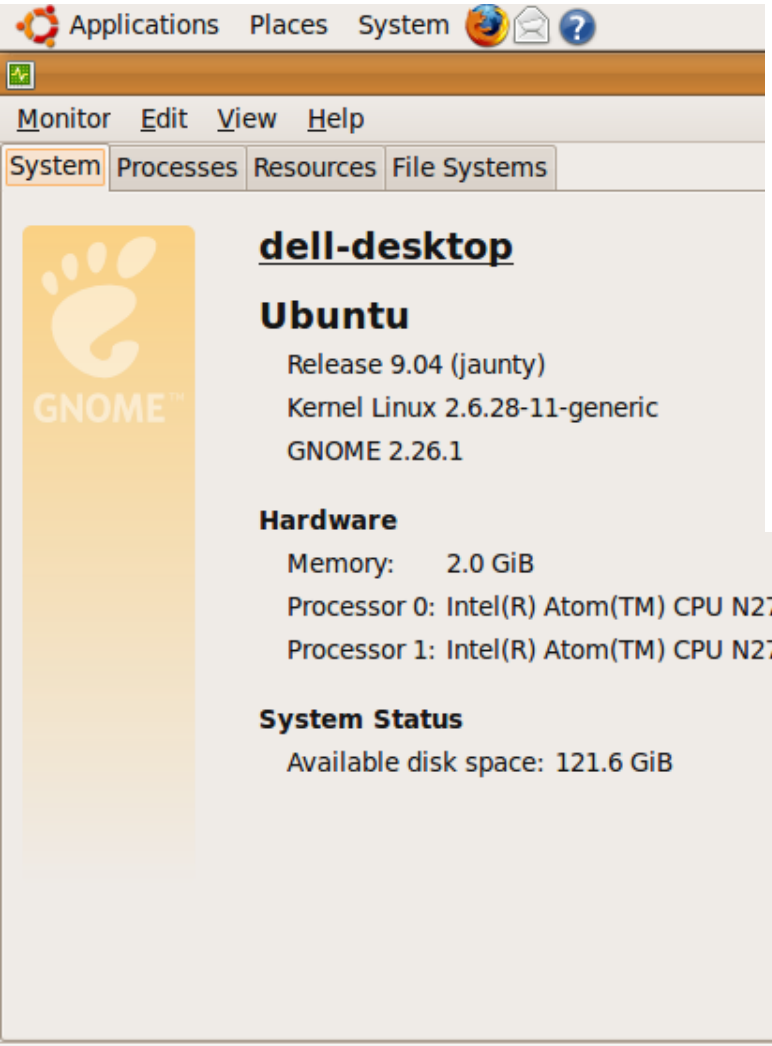
Index of ftp://ftp.ncbi.nih.gov/blast/executables/release/2.2.20/

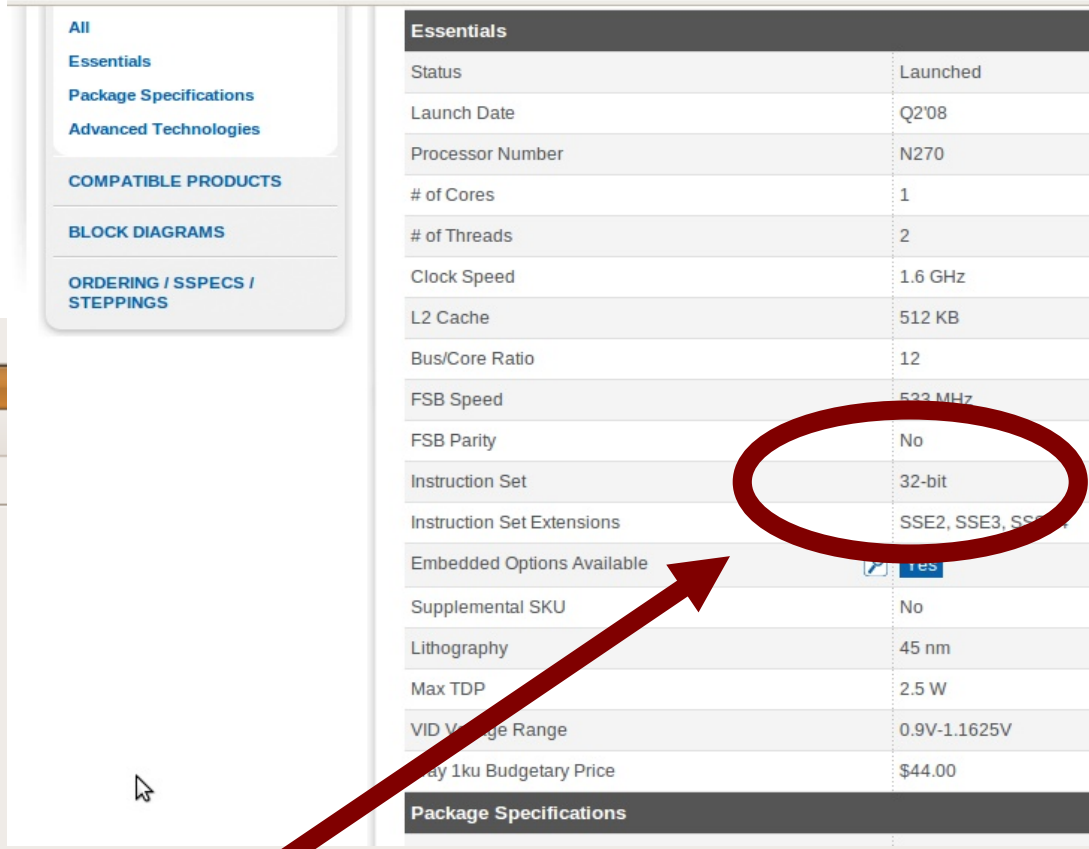
Up to higher level directory

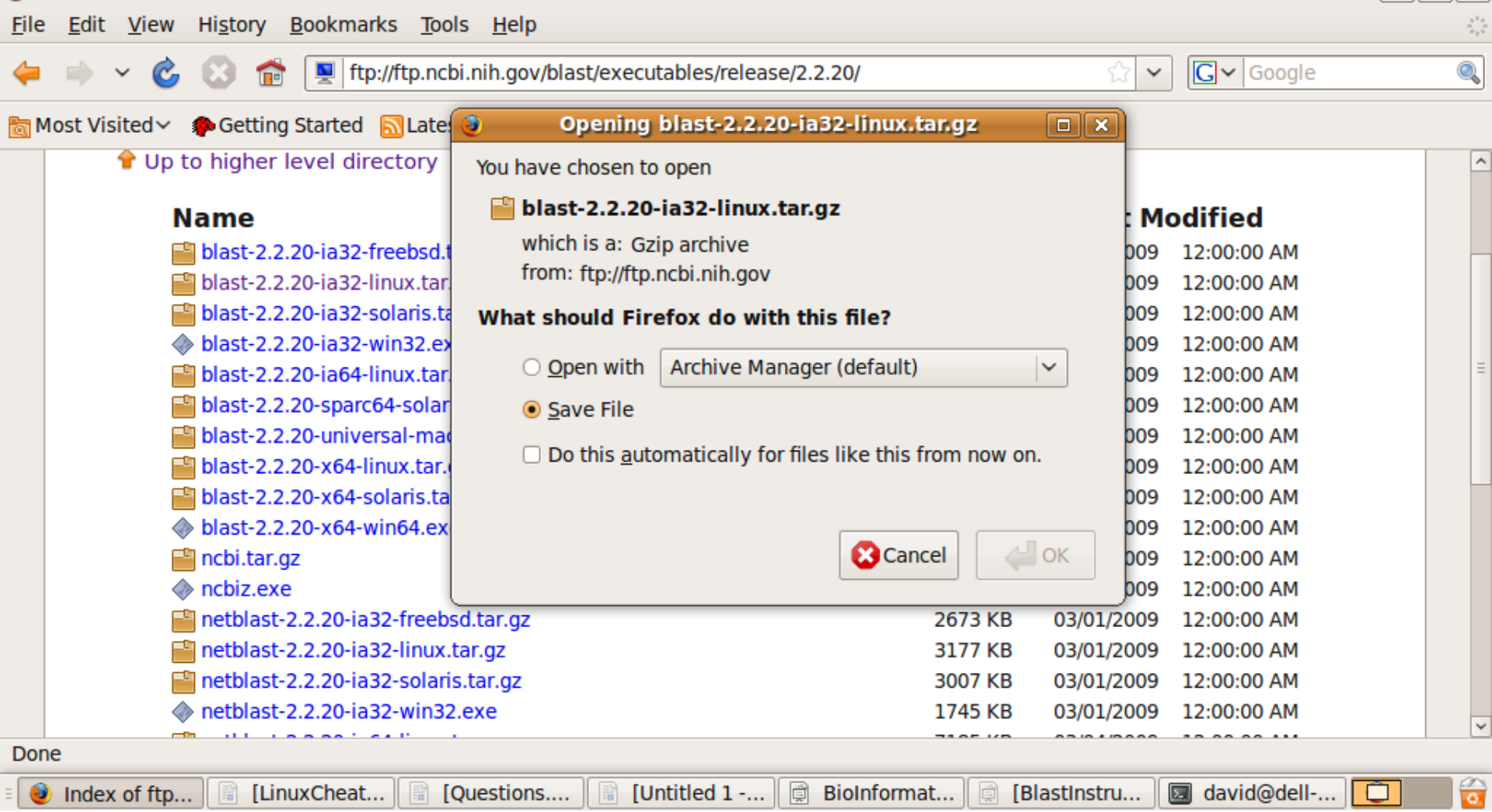
Name	Size	Last Modified
blast-2.2.20-ia32-freebsd.tar.gz	25514 KB	03/01/2009 12:00:00 AM
blast-2.2.20-ia32-linux.tar.gz	31357 KB	03/01/2009 12:00:00 AM
blast-2.2.20-ia32-solaris.tar.gz	29056 KB	03/01/2009 12:00:00 AM
blast-2.2.20-ia32-win32.exe	14364 KB	03/01/2009 12:00:00 AM
blast-2.2.20-ia64-linux.tar.gz	77572 KB	03/04/2009 12:00:00 AM
blast-2.2.20-sparc64-solaris.tar.gz	43674 KB	03/01/2009 12:00:00 AM
blast-2.2.20-universal-macosx.tar.gz		
blast-2.2.20-x64-linux.tar.gz		
blast-2.2.20-x64-solaris.tar.gz		
blast-2.2.20-x64-win64.exe		
ncbi.tar.gz		
ncbiz.exe		
netblast-2.2.20-ia32-freebsd.tar.gz		
netblast-2.2.20-ia32-linux.tar.gz		
netblast-2.2.20-ia32-solaris.tar.gz		
netblast-2.2.20-ia32-win32.exe		
netblast-2.2.20-ia64-linux.tar.gz		
netblast-2.2.20-sparc64-solaris.tar.gz		
netblast-2.2.20-universal-macosx.tar.gz		
netblast-2.2.20-x64-linux.tar.gz		
netblast-2.2.20-x64-solaris.tar.gz		
netblast-2.2.20-x64-win64.exe		

Select version for your OS; Generally Win32 for Windows; ia32-linux for Linux or MAC (32 or 64 bit depends on system's processor)

My notebook processor is 32 bit





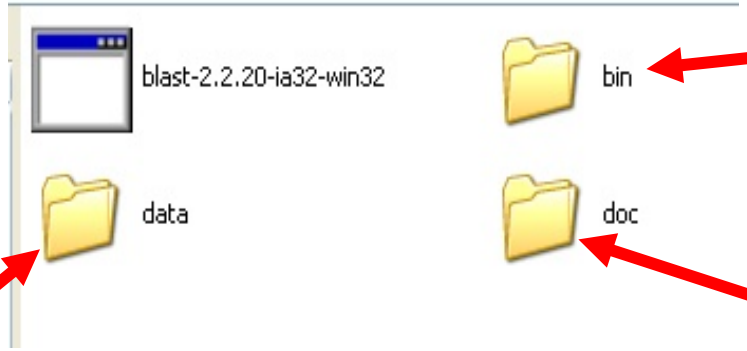


Create a folder called ‘BLAST’ (preferably on your main drive to simplify PATH statements e.g. C:\) and save the file to that folder.

- Go to the Blast directory: cd c: /Blast
- Windows Install: ./blast-2.2.20-ia32.win32.exe
- Unix (incl. MAC) Install: ./tar zxf blast-2.2.20-ia32-linux.tar.gz

(MAC: you may find that an additional folder was created "blast-2.2.20". In there you will find the folders below)

- In the Blast directory there will now be 3 new folders



All Blast files are located here

Algorithms for statistical analysis and sample search databases

Blast documentation in html format. Best viewed in a web browser

This ends our discussion of BLAST

Next up: BioPerl

BioPerl

Windows users will first need to install perl

Perl comes installed with MAC and Linux operating systems

Perl is: Practical Extraction and Report Language; a programing language for easily manipulating text, files, and processes

BioPerl

BioPerl is an open source project that develops modules for biological data in Perl.

A Perl module is a reusable package defined in a library file.

BioPerl modules are stable and “easy” to use.

Modules include objects for sequence files, alignment files and database searching. These objects can interact: the objects provide a coordinated and extensible framework for computational biology.

BioPerl

BioPerl module names minimize ‘namespace’ collisions by separating parts of a name by a double colon (::). For example:.

The module ‘Bio::DB::GenBank’; instructs Perl to go to the Database GenBank

This module can automate retrieval of a set of sequences

The ‘Bio::SearchIO’ module is used for parsing an input file and creating an output file with the specified information.

This module can be used to create tables that summarize results from BLAST searches

Google

bioPerl Buscar

Aproximadamente 266.000 resultados (0,87 segundos) Búsqueda avanzada

Todo
Libros
Más

Granadero Baigorria, Santa Fe
Cambiar ubicación

La Web
Páginas en español
Páginas de Argentina

Más herramientas

BioPerl - [Traducir esta página]
15 Oct 2010 ... The **Bioperl** Project is an international association of developers of open source Perl tools for bioinformatics, genomics and life science ...
www.**bioperl**.org/ - En caché

Installing BioPerl — Browse Modules
BioPerl Tutorial — Installing Bioperl for Unix
HOWTOs — Getting Started
Downloads — SeqIO

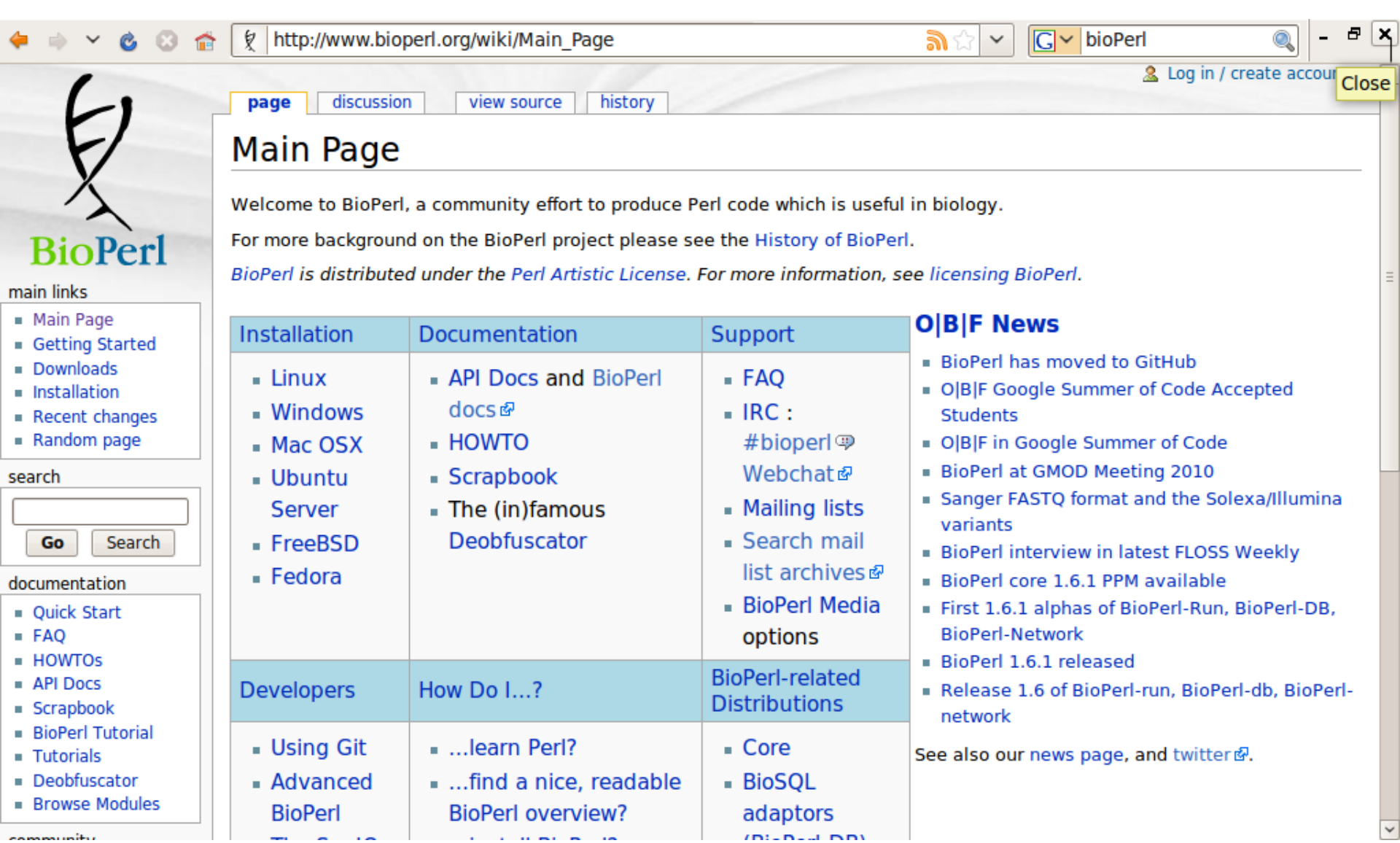
Más resultados de bioperl.org »

BioPerl Tutorial - **BioPerl** - [Traducir esta página]
25 May 2010 ... This tutorial is an overview of **Bioperl**, it includes snippets of code and text from various **Bioperl** documents including module documentation ...
www.**bioperl**.org/wiki/**BioPerl**_Tutorial - En caché - Similares

BioPerl - Wikipedia, la enciclopedia libre
BioPerl es una colección de módulos de Perl que facilitan el desarrollo de scripts en Perl para aplicaciones de bioinformática. ...
es.wikipedia.org/wiki/**BioPerl** - En caché - Similares

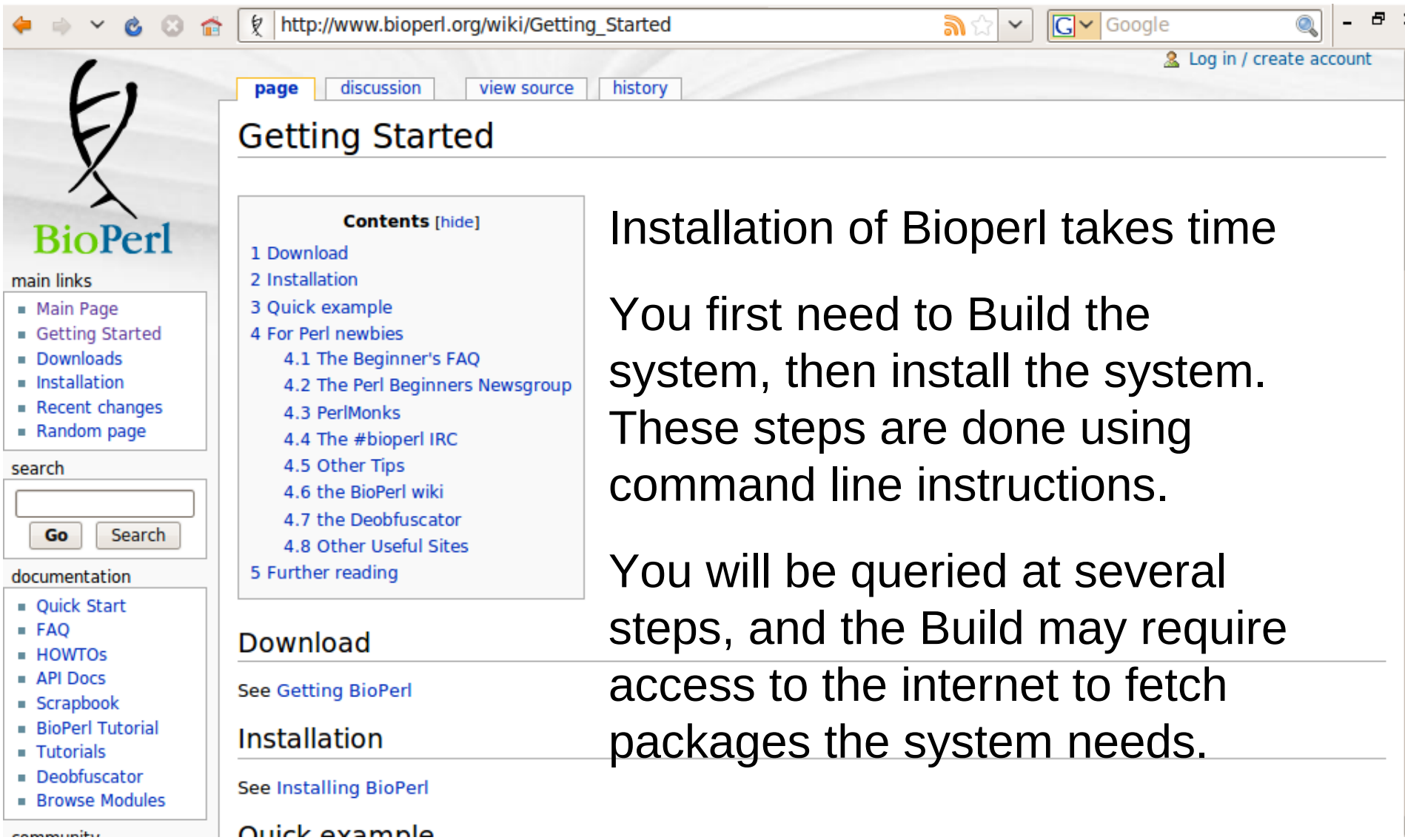
Bioperl course - [Traducir esta página]
Introduction to **Bioperl** (www.**bioperl**.org). This course introduces to the **bioperl** modules with examples and exercises. The course content has been upgraded ...
www.pasteur.fr/recherche/unites/sis/.../**bioperl**/ - En caché - Similares

[PDF] Taller de (**bio**)**Perl**
Formato de archivo: PDF/Adobe Acrobat - Vista rápida
de B Contreras-Moreira - 2010
2 Ene 2010 ... que necesitamos, con dos subrutinas de **BioPerl** podemos obtenerlas: .....



http://www.bioperl.org/wiki/Main_Page

http://www.bioperl.org/wiki/Getting_Started



Installation of Bioperl takes time

You first need to Build the system, then install the system. These steps are done using command line instructions.

You will be queried at several steps, and the Build may require access to the internet to fetch packages the system needs.

USDA United States Department of Agriculture National Institute of Food and Agriculture

Solanaceae Coordinated Agricultural Project

THE OHIO STATE UNIVERSITY OARDC

Bioinformatics 101

Part II: using the tools to for marker discovery

David Francis

The Ohio State University, OARDC

SolCAP workshop

GAII (Next Generation Sequence) PI 128216

H1706 Whole Genome Sequence Draft

BioPerl

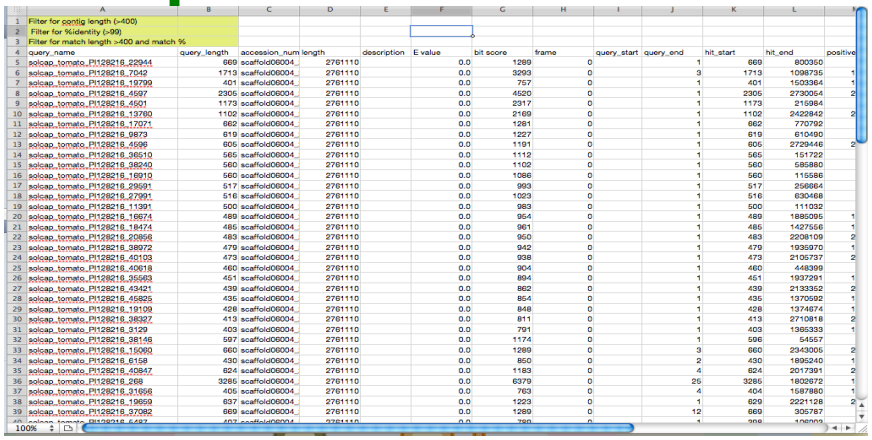
BLAST

BLAST

(2) WGS Scaffold

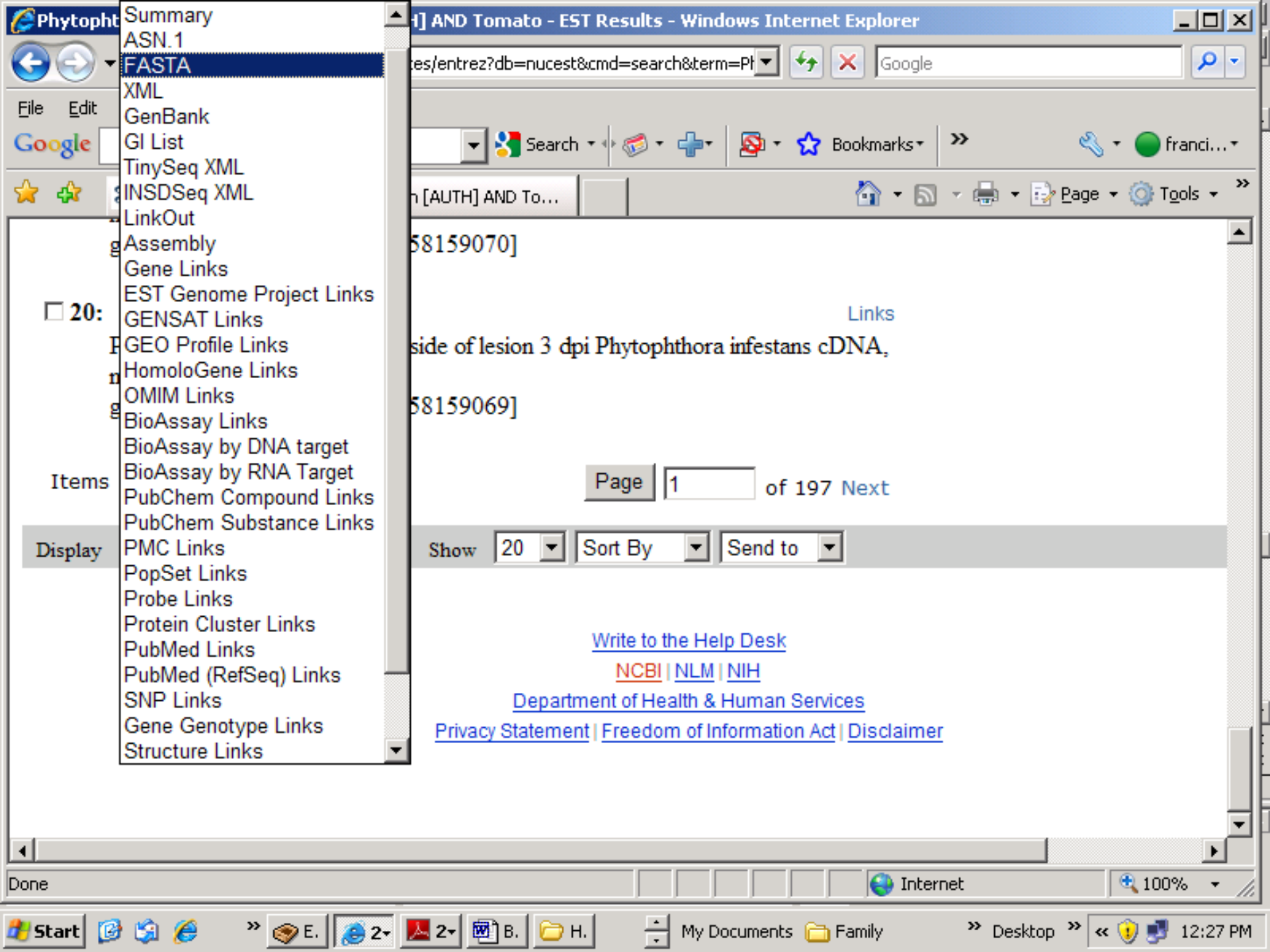
BioPerl

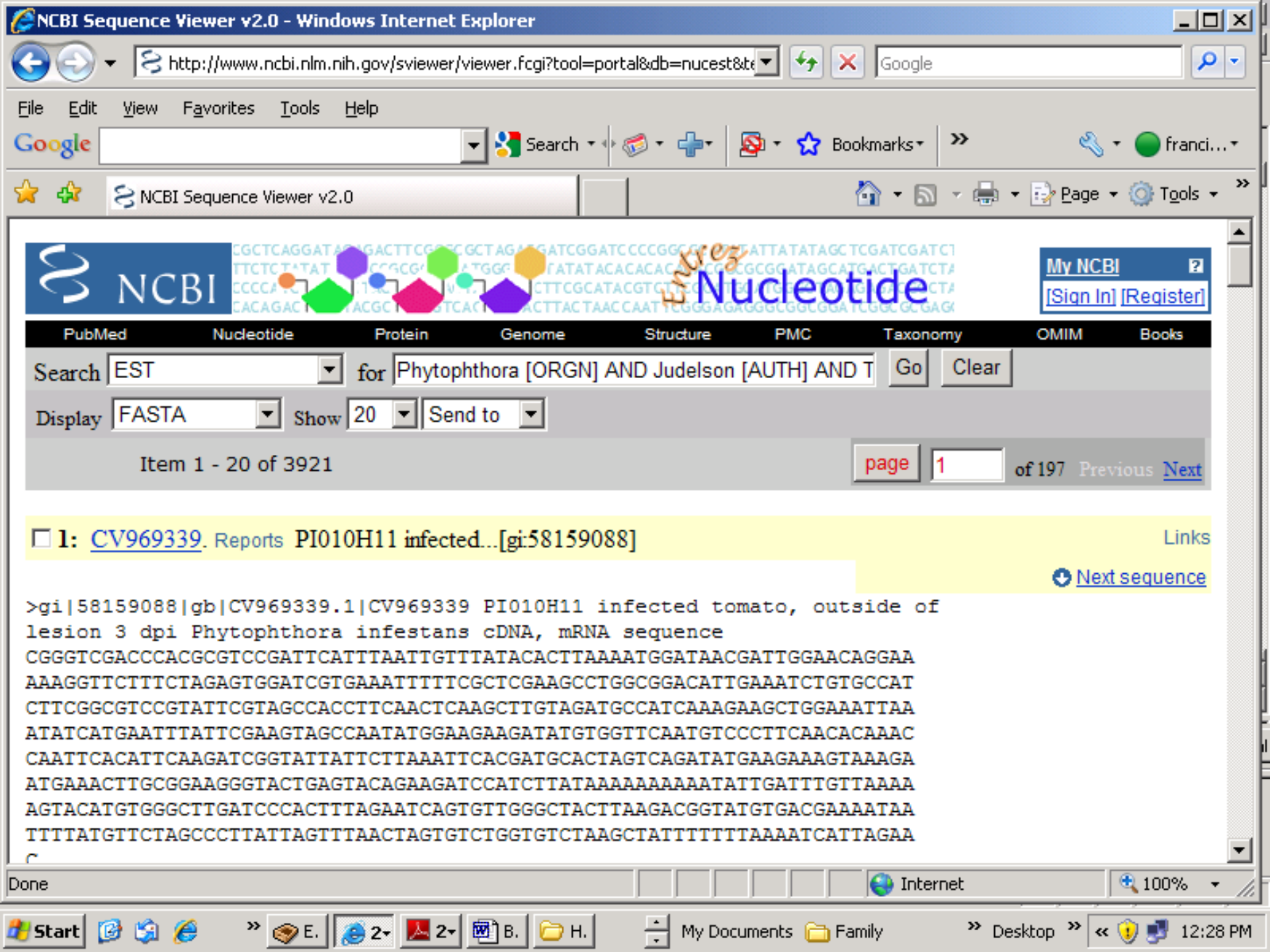
(1) Markers loosely Linked to the Resistant gene



	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Filter for contig length (>400)												
2	Filter for %identity (>99)												
3	Filter for match length >400 and match %												
4	query_name	query_length	accession_num	length	description	E value	bit score	frame	query_start	query_end	hit_start	hit_end	positive
5	solcap_tomato_PI128216_22944	669	scaffold06004_	2761110		0.0	1289	0		1	669	800360	
6	solcap_tomato_PI128216_7042	1713	scaffold06004_	2761110		0.0	3293	0		3	1713	1098735	1
7	solcap_tomato_PI128216_19799	401	scaffold06004_	2761110		0.0	757	0		1	401	1503364	1
8	solcap_tomato_PI128216_4597	2305	scaffold06004_	2761110		0.0	4520	0		1	2305	2730054	2
9	solcap_tomato_PI128216_4501	1173	scaffold06004_	2761110		0.0	2317	0		1	1173	215984	
10	solcap_tomato_PI128216_13780	1102	scaffold06004_	2761110		0.0	2169	0		1	1102	2422842	2
11	solcap_tomato_PI128216_17071	662	scaffold06004_	2761110		0.0	1261	0		1	662	770792	
12	solcap_tomato_PI128216_9873	619	scaffold06004_	2761110		0.0	1227	0		1	619	610490	
13	solcap_tomato_PI128216_4596	605	scaffold06004_	2761110		0.0	1191	0		1	605	2729446	2
14	solcap_tomato_PI128216_38510	565	scaffold06004_	2761110		0.0	1112	0		1	565	151722	
15	solcap_tomato_PI128216_38240	560	scaffold06004_	2761110		0.0	1102	0		1	560	585880	
16	solcap_tomato_PI128216_16910	560	scaffold06004_	2761110		0.0	1086	0		1	560	115586	
17	solcap_tomato_PI128216_29691	517	scaffold06004_	2761110		0.0	993	0		1	517	256864	
18	solcap_tomato_PI128216_27991	516	scaffold06004_	2761110		0.0	1023	0		1	516	630468	
19	solcap_tomato_PI128216_11391	500	scaffold06004_	2761110		0.0	983	0		1	500	111032	
20	solcap_tomato_PI128216_16674	489	scaffold06004_	2761110		0.0	954	0		1	489	1885095	1
21	solcap_tomato_PI128216_18474	485	scaffold06004_	2761110		0.0	961	0		1	485	1427556	1
22	solcap_tomato_PI128216_20898	483	scaffold06004_	2761110		0.0	950	0		1	483	2208109	2
23	solcap_tomato_PI128216_38972	479	scaffold06004_	2761110		0.0	942	0		1	479	1935970	1
24	solcap_tomato_PI128216_40103	473	scaffold06004_	2761110		0.0	938	0		1	473	2105737	2
25	solcap_tomato_PI128216_40618	460	scaffold06004_	2761110		0.0	904	0		1	460	448399	
26	solcap_tomato_PI128216_35593	451	scaffold06004_	2761110		0.0	894	0		1	451	1937291	1
27	solcap_tomato_PI128216_43421	439	scaffold06004_	2761110		0.0	862	0		1	439	2133352	2
28	solcap_tomato_PI128216_45825	436	scaffold06004_	2761110		0.0	854	0		1	436	1370592	1
29	solcap_tomato_PI128216_19109	428	scaffold06004_	2761110		0.0	848	0		1	428	1374674	1
30	solcap_tomato_PI128216_36927	413	scaffold06004_	2761110		0.0	811	0		1	413	2710818	2
31	solcap_tomato_PI128216_3129	403	scaffold06004_	2761110		0.0	791	0		1	403	1365333	1
32	solcap_tomato_PI128216_38146	597	scaffold06004_	2761110		0.0	1174	0		1	596	54557	
33	solcap_tomato_PI128216_15060	660	scaffold06004_	2761110		0.0	1289	0		3	660	2343005	2
34	solcap_tomato_PI128216_6158	430	scaffold06004_	2761110		0.0	850	0		2	430	1895240	1
35	solcap_tomato_PI128216_40847	624	scaffold06004_	2761110		0.0	1183	0		4	624	2017391	2
36	solcap_tomato_PI128216_268	3286	scaffold06004_	2761110		0.0	6379	0		25	3286	1802672	1
37	solcap_tomato_PI128216_31698	406	scaffold06004_	2761110		0.0	763	0		4	404	1587880	1
38	solcap_tomato_PI128216_19659	637	scaffold06004_	2761110		0.0	1223	0		1	629	2221128	2
39	solcap_tomato_PI128216_37082	669	scaffold06004_	2761110		0.0	1289	0		12	669	305787	
40	solcap_tomato_PI128216_6487	407	scaffold06004_	2761110		0.0	780	0		1	398	126009	

1.8 Mb, 172 SNPs, 132 pass Illumina Design Criteria; 60 unique genes


Phytopht
] AND Tomato - EST Results - Windows Internet Explorer
es/entrez?db=nucest&cmd=search&term=Ph
Google
File Edit
Google
Search
Bookmarks
franci...
[AUTH] AND To...
Page
Tools
Summary
ASN.1
FASTA
XML
GenBank
GI List
TinySeq XML
INSDSeq XML
LinkOut
Assembly
Gene Links
EST Genome Project Links
GENSAT Links
GEO Profile Links
HomoloGene Links
OMIM Links
BioAssay Links
BioAssay by DNA target
BioAssay by RNA Target
PubChem Compound Links
PubChem Substance Links
PMC Links
PopSet Links
Probe Links
Protein Cluster Links
PubMed Links
PubMed (RefSeq) Links
SNP Links
Gene Genotype Links
Structure Links
58159070]
20:
Links
side of lesion 3 dpi Phytophthora infestans cDNA,
58159069]
Items
Page 1 of 197 Next
Display
Show 20
Sort By
Send to
Write to the Help Desk
NCBI | NLM | NIH
Department of Health & Human Services
Privacy Statement | Freedom of Information Act | Disclaimer
Done
Internet
100%
Start
My Documents
Family
Desktop
12:27 PM


NCBI Sequence Viewer v2.0 - Windows Internet Explorer
http://www.ncbi.nlm.nih.gov/sviewer/viewer.fcgi?tool=portal&db=nucest&te
Google
File Edit View Favorites Tools Help
Search Bookmarks franci...
NCBI Sequence Viewer v2.0
Page Tools
NCBI
Entrez Nucleotide
My NCBI
[Sign In] [Register]
PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Books
Search EST for Phytophthora [ORGN] AND Judelson [AUTH] AND T Go Clear
Display FASTA Show 20 Send to
Item 1 - 20 of 3921
page 1 of 197 Previous Next
1: CV969339. Reports PI010H11 infected...[gi:58159088]
Links
Next sequence
>gi|58159088|gb|CV969339.1|CV969339 PI010H11 infected tomato, outside of
lesion 3 dpi Phytophthora infestans cDNA, mRNA sequence
CGGGTCGACCCACGCGTCCGATTCATTTAATTGTTTATACACTTAAAATGGATAACGATTGGAACAGGAA
AAAGGTTCTTTCTAGAGTGGATCGTGAAATTTTTCGCTCGAAGCCTGGCGGACATTGAAATCTGTGCCAT
CTTCGGCGTCCGTATTCGTAGCCACCTTCAACTCAAGCTTGTAGATGCCATCAAAGAAGCTGGAAATTAA
ATATCATGAATTTATTCGAAGTAGCCAATATGGAAGAAGATATGTGGTTCAATGTCCCTTCAACACAAAC
CAATTCACATTCAAGATCGGTATTATTCTTAAATTCACGATGCACTAGTCAGATATGAAGAAAGTAAAGA
ATGAAACTTGCGGAAGGGTACTGAGTACAGAAGATCCATCTTATAAAAAAAAAATATTGATTTGTTAAAA
AGTACATGTGGGCTTGATCCCACTTTAGAATCAGTGTTGGGCTACTTAAGACGGTATGTGACGAAAATAA
TTTTATGTTCTAGCCCTTATTAGTTTAACTAGTGTCTGGTGTCTAAGCTATTTTTTTAAAATCATTAGAA
C
Done
Internet
100%
Start
E.
2
2
B.
H.
My Documents
Family
Desktop
12:28 PM

USE the Unix grep command to verify that you downloaded the entire file

NOTE: a ‘cheet sheet of UNIX commands’ is available on the PBGWorks wiki at:

http://pbgworks.org/node/901

Example: $ **grep -c ‘>’** will count the number of times the ‘>’ occurs, and therefore the number of sequences in a FASTA file.

For the file downloaded using following the NCBI EST database search using ENTREZ: **Lycopersicum [ORGN] AND TA496** grep returns 116711, which matches our expectations

Lycopersicum [ORGN] AND Rio Grande

21973

Lycopersicum [ORGN] AND Rio Fuego

171

Lycopersicum [ORGN] AND MicroTom

120462

Lycopersicum [ORGN] AND TA496

116711

Lycopersicum [ORGN] AND Moneymaker

833

Phytophthora [ORGN] AND Judelson [AU] AND Tomato

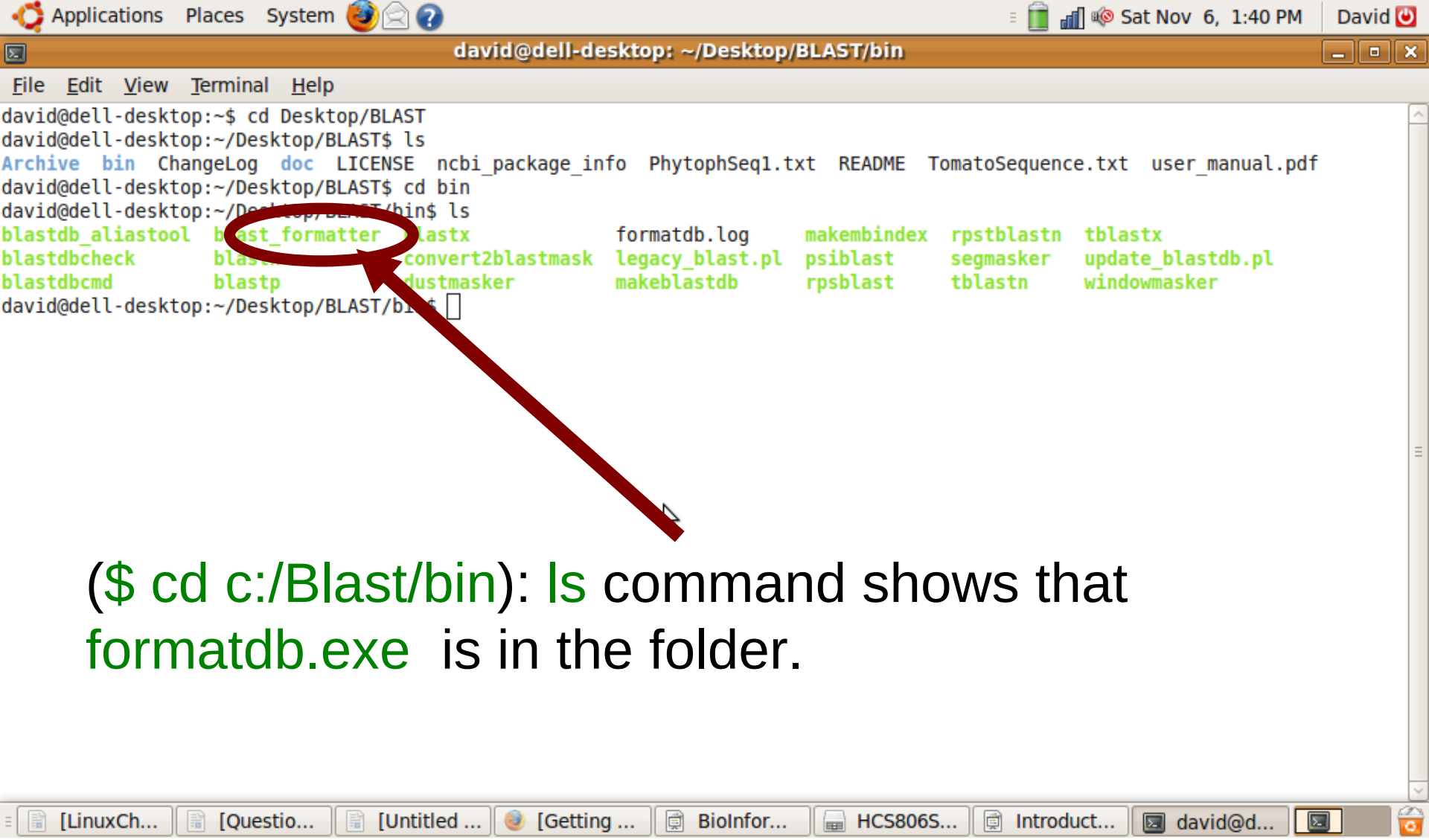
3921

Format your database for BLAST

1) Use the formatdb command for this task.

2) formatdb.exe is located in the bin folder that was created when you unpacked BLAST. So, if you saved BLAST to a folder named Blast, bin will be located within Blast.

3) You need to tell the computer where to look for formatdb.exe , and where to look for the file that you want to format. This means specifying a PATH. Use cd to navigate to the bin folder ($ cd c:/Blast/bin). The pwd and ls commands can be used to verify that you are in the proper path and that formatdb.exe is in the folder.



($ cd c:/Blast/bin): ls command shows that formatdb.exe is in the folder.

SYNTAX of the command:

$/**formatdb -i ./DatabaseName –p F**

Note:
the ./ implies the database input file is in the bin folder with the formatdb.exe

If the database is in another folder, you must specify the path (C:/BLAST/DF/)

-p asks if the file contains protein data. Our answer is False, because the file contains DNA sequence.

Use ls to list the files in the folder containing the database.

You should now see three new files:

- DatabaseName.nhr
- DatabaseName.nin
- DatabaseName.nsq

Now we're ready to run a stand alone BLAST.
SYNTAX of the command:

$/blastall -p blastn -d ./DatabaseName -i ./QueryFile.txt –o Output

Note:
the ./ implies all files are in the bin folder

-p asks which program. We are using blastn
-d asks for the database (must be formatted)
-i asks for the input or query file (FASTA format)
–o Tell BLAST what you want to name the output file

Viewing results of a BLAST search will depend on the search:

A simple search may be viewed by opening the output file in a text editor

Some BLAST searches will return very large files. These are best examined with some basic UNIX commands (grep, head, tail, and less), and then parsed to organize the data.

Viewing the output file

File Edit View Terminal Help

```
BLASTN 2.2.18 [Mar-02-2008]


Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
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.

Query=

Database: ./TA496Seq1.txt
           116,711 sequences; 60,705,274 total letters

Searching.................................................done



                                                                 Score    E
Sequences producing significant alignments:                      (bits) Value

gi|16229829|gb|BI925062.1|BI925062 EST544951 tomato flower, buds...  1530   0.0
gi|18263929|gb|BM412299.1|BM412299 EST586626 tomato breaker frui...  1516   0.0
gi|13778955|gb|BG643730.1|BG643730 EST511924 tomato shoot/merist...  1507   0.0
gi|18262786|gb|BM411156.1|BM411156 EST585483 tomato breaker frui...  1471   0.0
gi|18262599|gb|BM410969.1|BM410969 EST585296 tomato breaker frui...  1453   0.0
gi|16232161|gb|BI925777.1|BI925777 EST545666 tomato flower, buds...  1449   0.0
gi|18262889|gb|BM411259.1|BM411259 EST585586 tomato breaker frui...  1447   0.0
gi|16248431|gb|BI933959.1|BI933959 EST553848 tomato flower, anth...  1433   0.0
gi|16236252|gb|BI927083.1|BI927083 EST546972 tomato flower, 3 - ...  1423   0.0
gi|14683291|gb|BI205567.1|BI205567 EST523607 cTOS Solanum lycope...  1403   0.0
gi|12627674|gb|BG127486.1|BG127486 EST473132 tomato shoot/merist...  1390   0.0
:
```

Use the UNIX “less” command (followed by q to quit)

Parsing the output file

```
#  usage:
# perl program  <BLAST-report-file>  # to extract <output_name >
#

use strict;
use warnings;
use lib "/home/users/David/lib/perl5";
use Bio::SearchIO;
```

Parsing the output file

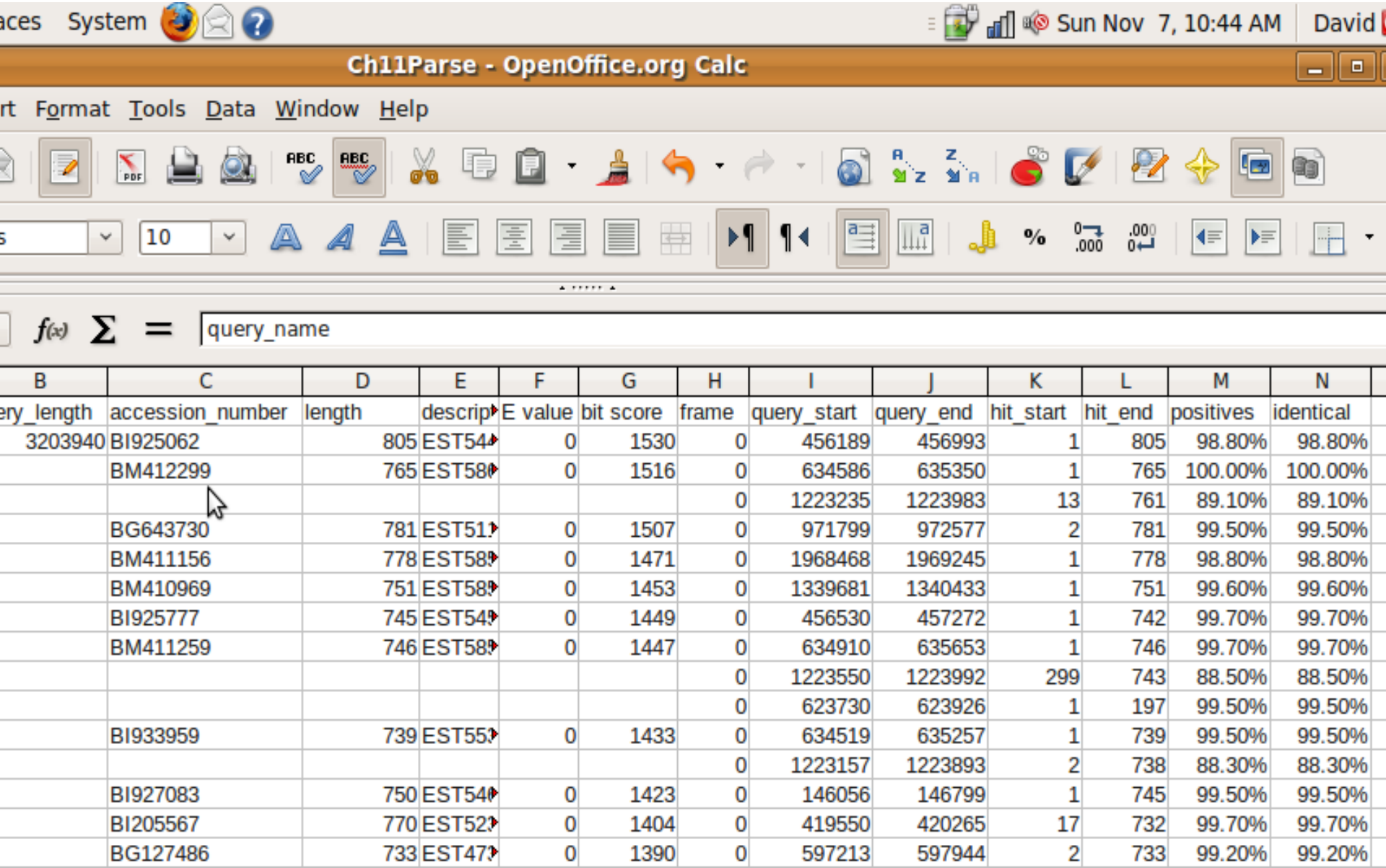
```perl
# Usage information
die "Usage: $0 <BLAST-report-file> <number-of-top-hits> <output-file>\n", if (@ARGV != 3);

my ($infile,$numHits,$outfile) = @ARGV;
print "Parsing the BLAST result ...";
my $in = Bio::SearchIO->new(-format => 'blast', -file => $infile);
open (OUT,">$outfile") or die "Cannot open $outfile: $!";

# print the header info for tab-deliminated columns
print OUT "query_name\tquery_length\taccession_number\tlength\tdescription\tE value\tbit score\tframe\tquery_start\t";
print OUT "query_end\thit_start\thit_end\tpositives\tidentical\n";

# extraction of information for each result recursively
while ( my $result = $in->next_result ) {
        # the name of the query sequence
        print OUT $result->query_name . "\t";

        # the length of the query sequence
        print OUT $result->query_length;

        # output "no hits found" if there is no hits
        if ( $result->num_hits == 0 ) {
                print OUT "\tNo hits found\n";
        } else {
                my $count = 0;

                # process each hit recursively
                while (my $hit = $result->next_hit) {
                        print OUT "\t" if ($count > 0);
                        # get the accession numbers of the hits
                        print OUT "\t" . $hit->accession . "\t";
                        # get the lengths of the hit sequences
                        print OUT $hit->length . "\t";
                        # get the description of the hit sequences
                        print OUT $hit->description . "\t";
```

The Perl script checks for the expected three arguments (input file, number of hits to extract, output file); then it uses Bio::SearchIO to pull information from the blast report and put that information into a tab delimited file

Key Commands

```
perl   blast_parsing_pl1.pl
  out_Ch11_blast   100  Ch11Parse
```

Results (open in EXCEL)



B	C	D	E	F	G	H	I	J	K	L	M	N
ery_length	accession_number	length	descrip	E value	bit score	frame	query_start	query_end	hit_start	hit_end	positives	identical
3203940	BI925062	805	EST54	0	1530	0	456189	456993	1	805	98.80%	98.80%
	BM412299	765	EST58	0	1516	0	634586	635350	1	765	100.00%	100.00%
						0	1223235	1223983	13	761	89.10%	89.10%
	BG643730	781	EST51	0	1507	0	971799	972577	2	781	99.50%	99.50%
	BM411156	778	EST58	0	1471	0	1968468	1969245	1	778	98.80%	98.80%
	BM410969	751	EST58	0	1453	0	1339681	1340433	1	751	99.60%	99.60%
	BI925777	745	EST54	0	1449	0	456530	457272	1	742	99.70%	99.70%
	BM411259	746	EST58	0	1447	0	634910	635653	1	746	99.70%	99.70%
						0	1223550	1223992	299	743	88.50%	88.50%
						0	623730	623926	1	197	99.50%	99.50%
	BI933959	739	EST55	0	1433	0	634519	635257	1	739	99.50%	99.50%
						0	1223157	1223893	2	738	88.30%	88.30%
	BI927083	750	EST54	0	1423	0	146056	146799	1	745	99.50%	99.50%
	BI205567	770	EST52	0	1404	0	419550	420265	17	732	99.70%	99.70%
	BG127486	733	EST47	0	1390	0	597213	597944	2	733	99.20%	99.20%

Next Step:

Retrieve the desired sequence:

a) Directly from the FASTA file

b) from GenBank (using BioPerl)

For (b) create a text file with the gb id of the sequences you want to retrieve:

BG643730

AF536200

AF536199...

(Chr11EST.txt = file containing a single column of gb id numbers)

Retrieving Sequence Data

```
#This script will extract sequences from Genbank
# Only 2 arguments are required
# an input file (with the accession numbers) and output file

use strict;
use warnings;
use lib "/home/users/David/lib/perl5";
use Bio::DB::GenBank;
use Bio::SearchIO;
```

Key Commands

perl GenbankSearch2.pl Chr11EST.txt Chr11_FASTA

Workshop Resources:

http://pbgworks.org/node/901

Perl script to test if BioPerl has been properly installed (returns “it works!)

Perl script that will parse a BLAST search

Perl scripts that allow user defined criteria to be considered during the BLAST parsing

Perl script that will retrieve specified sequences from GenBank (NCBI)

Questions?