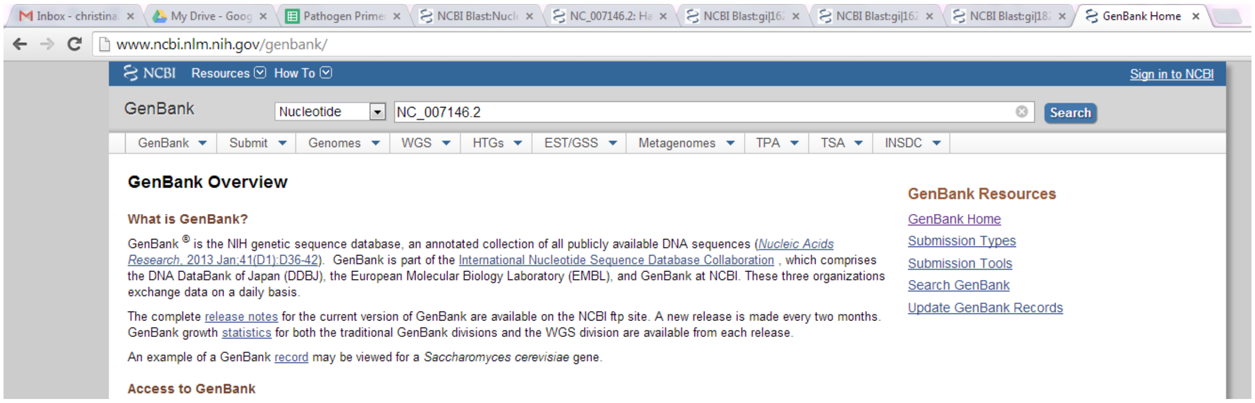
Reflection Questions for the DNA Project

1. What was the DNA Project about and what skills were you able to apply to the project?
2. What new knowledge and skills did you learn from this interdisciplinary collaborative project?
3. What were the areas that you faced challenges and feel you need improvement?
4. How well were you able to adapt to the research project's ambiguity?
5. What near term goals can you set for the project?

BLAST search Protocol:

Enter NC Number into Nucleotide (http://www.ncbi.nlm.nih.gov/genbank/) and click Search.



GenBank Overview

What is GenBank?

GenBank ® is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences (*Nucleic Acids Research*, 2013 Jan;41(D1):D36-42). GenBank is part of the International Nucleotide Sequence Database Collaboration , which comprises the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI. These three organizations exchange data on a daily basis.

The complete release notes for the current version of GenBank are available on the NCBI ftp site. A new release is made every two months. GenBank growth statistics for both the traditional GenBank divisions and the WGS division are available from each release.

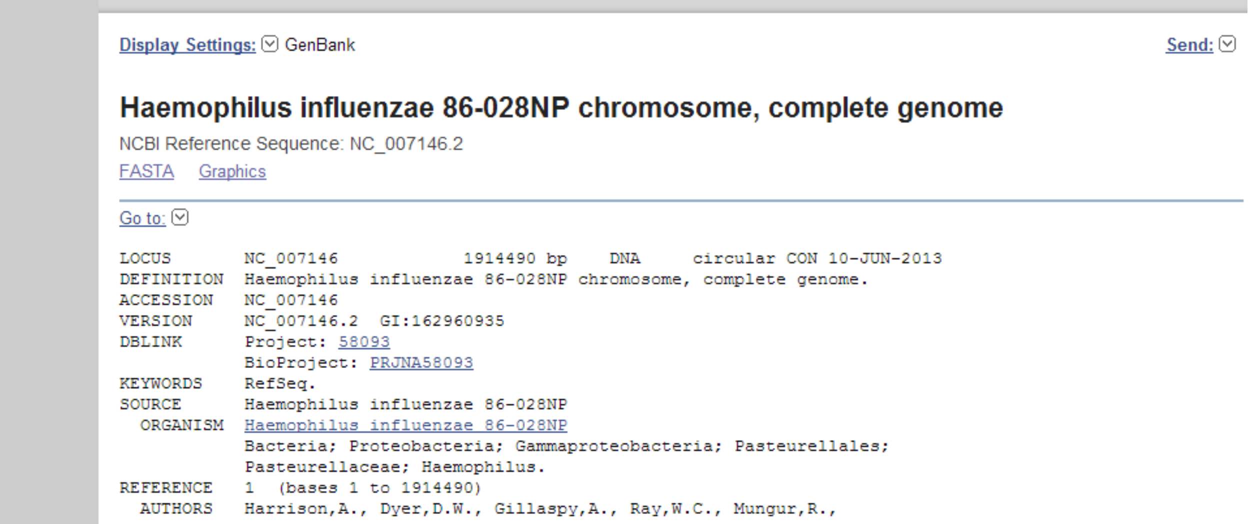
An example of a GenBank record may be viewed for a *Saccharomyces cerevisiae* gene.

Access to GenBank

GenBank Resources

- GenBank Home
- Submission Types
- Submission Tools
- Search GenBank
- Update GenBank Records

Click on the Graphics view

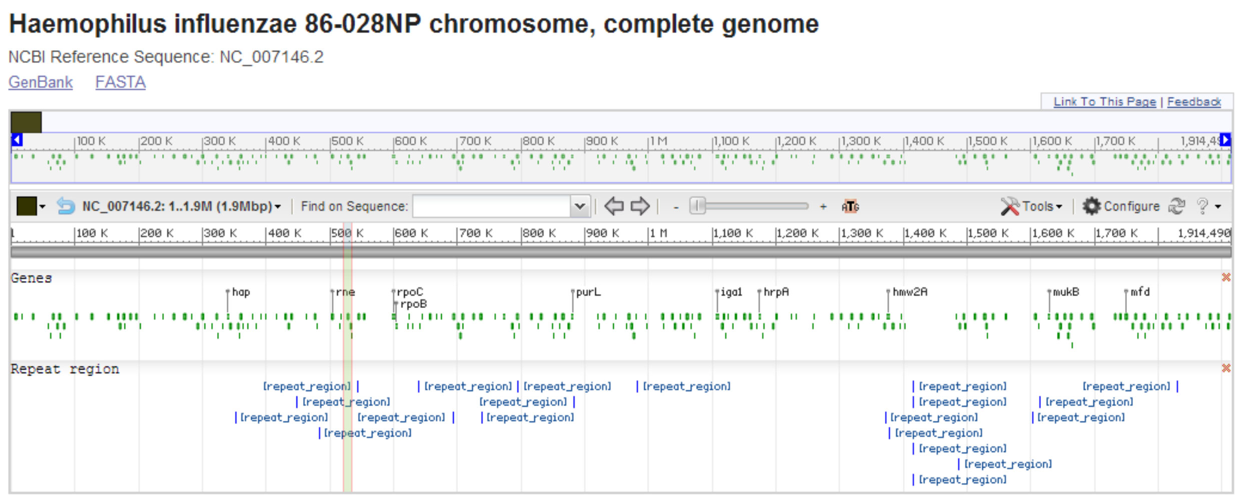


Haemophilus influenzae 86-028NP chromosome, complete genome

NCBI Reference Sequence: NC_007146.2

```
LOCUS       NC_007146            1914490 bp    DNA     circular CON 10-JUN-2013
DEFINITION  Haemophilus influenzae 86-028NP chromosome, complete genome.
ACCESSION   NC_007146
VERSION     NC_007146.2  GI:162960935
DBLINK      Project: 58093
            BioProject: PRJNA58093
KEYWORDS    RefSeq.
SOURCE      Haemophilus influenzae 86-028NP
  ORGANISM  Haemophilus influenzae 86-028NP
            Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
            Pasteurellaceae; Haemophilus.
REFERENCE   1  (bases 1 to 1914490)
  AUTHORS   Harrison,A., Dyer,D.W., Gillaspy,A., Ray,W.C., Mungur,R.,
```

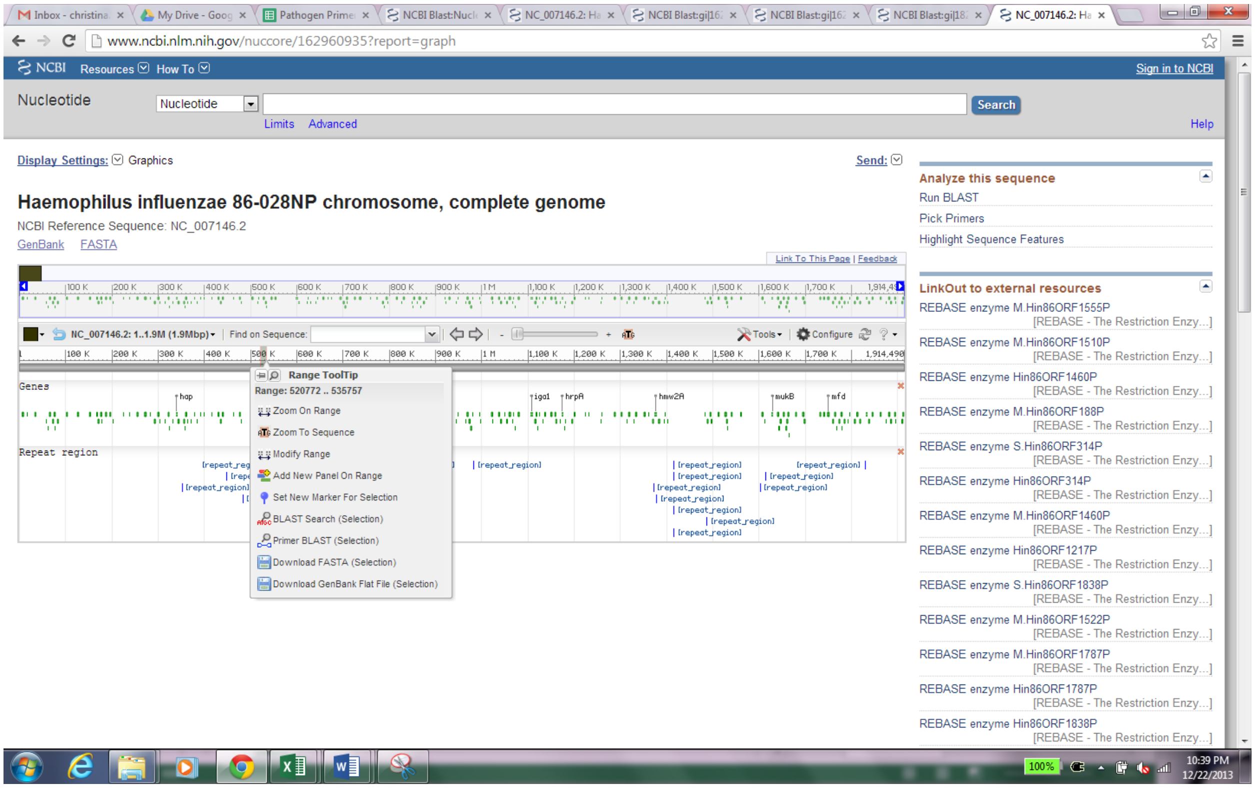
Click and Drag on the Nucleotide bar.



Haemophilus influenzae 86-028NP chromosome, complete genome

NCBI Reference Sequence: NC_007146.2

The Range ToolTip Menu will come up and Select Modify Range



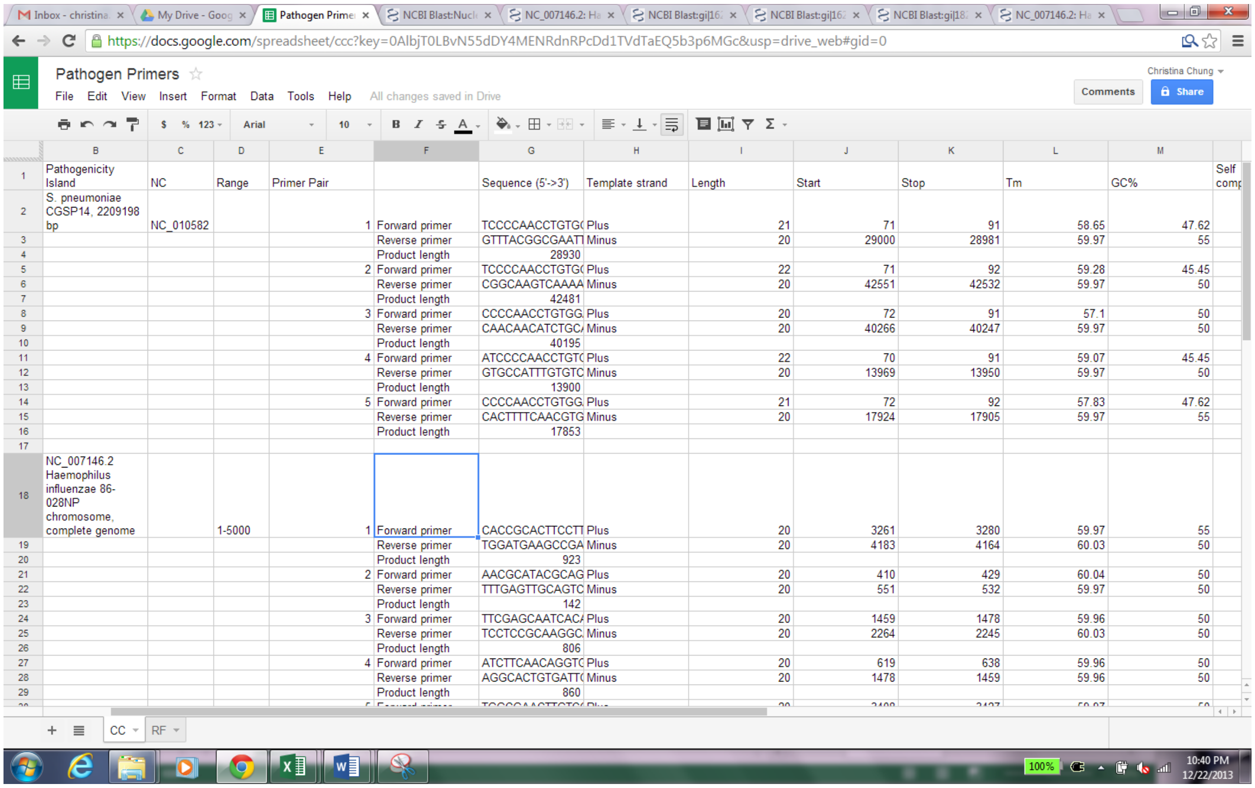
www.ncbi.nlm.nih.gov/nuccore/162960935?report=graph

Haemophilus influenzae 86-028NP chromosome, complete genome

NCBI Reference Sequence: NC_007146.2

Range: 520772 .. 535757

Go back to the results from primer searches:

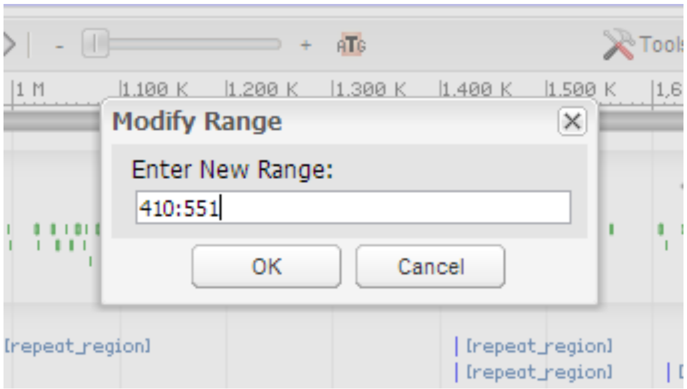


Pathogen Primers

https://docs.google.com/spreadsheet/ccc?key=0AlbjT0LBvN55dDY4MENRdnRPcDd1TVdTaEQ5b3p6MGc&usp=drive_web#gid=0

	B	C	D	E	F	G	H	I	J	K	L	M	
1	Pathogenicity Island	NC	Range	Primer Pair		Sequence (5'->3')	Template strand	Length	Start	Stop	Tm	GC%	Self comp
2	S. pneumoniae CGSP14, 2209198 bp	NC_010582		1	Forward primer	TCCCCAACCTGTG	Plus	21	71	91	58.65	47.62	
3					Reverse primer	GTTTACGGCGAAT	Minus	20	29000	28981	59.97	55	
4					Product length	28930							
5				2	Forward primer	TCCCCAACCTGTG	Plus	22	71	92	59.28	45.45	
6					Reverse primer	CGGCAAGTCAAAA	Minus	20	42551	42532	59.97	50	
7					Product length	42481							
8				3	Forward primer	CCCCAACCTGTGG	Plus	20	72	91	57.1	50	
9					Reverse primer	CAACAACATCTGC	Minus	20	40266	40247	59.97	50	
10					Product length	40195							
11				4	Forward primer	ATCCCCAACCTGT	Plus	22	70	91	59.07	45.45	
12					Reverse primer	GTGCCATTTGTGTC	Minus	20	13969	13950	59.97	50	
13					Product length	13900							
14				5	Forward primer	CCCCAACCTGTGG	Plus	21	72	92	57.83	47.62	
15					Reverse primer	CACTTTTCAACGTG	Minus	20	17924	17905	59.97	55	
16					Product length	17853							
17													
18	NC_007146.2 Haemophilus influenzae 86-028NP chromosome, complete genome		1-5000	1	Forward primer	CACCGCACTTCCT	Plus	20	3261	3280	59.97	55	
19					Reverse primer	TGGATGAAGCCGA	Minus	20	4183	4164	60.03	50	
20					Product length	923							
21				2	Forward primer	AACGCATACGCAG	Plus	20	410	429	60.04	50	
22					Reverse primer	TTTGAGTTGCAGTC	Minus	20	551	532	59.97	50	
23					Product length	142							
24				3	Forward primer	TTCGAGCAATCAC	Plus	20	1459	1478	59.96	50	
25					Reverse primer	TCCTCCGCAAGGC	Minus	20	2264	2245	60.03	50	
26					Product length	806							
27				4	Forward primer	ATCTTCAACAGGT	Plus	20	619	638	59.96	50	
28					Reverse primer	AGGCACTGTGATT	Minus	20	1478	1459	59.96	50	
29					Product length	860							

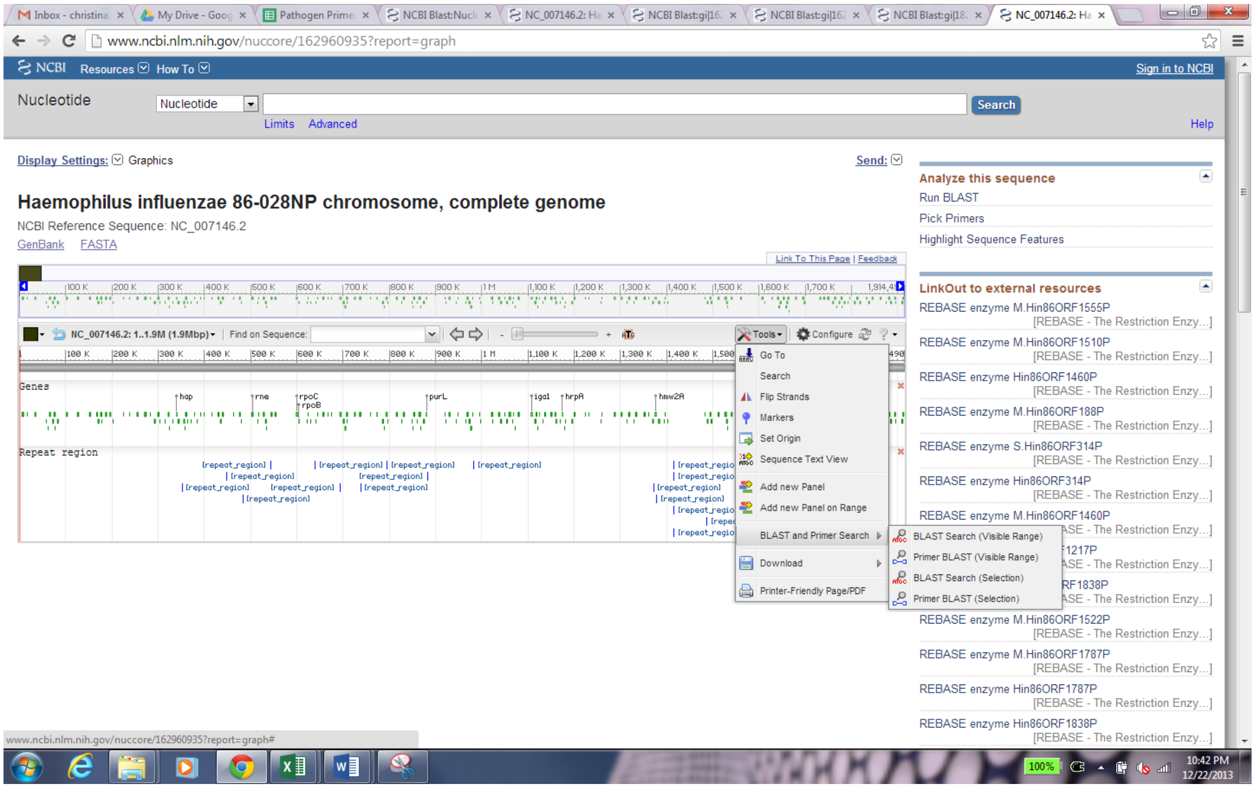
Select the range from the start points for the primers.



Modify Range

Enter New Range: 410:551

Tools>BLAST and Primer Search>BLAST Search (selection)



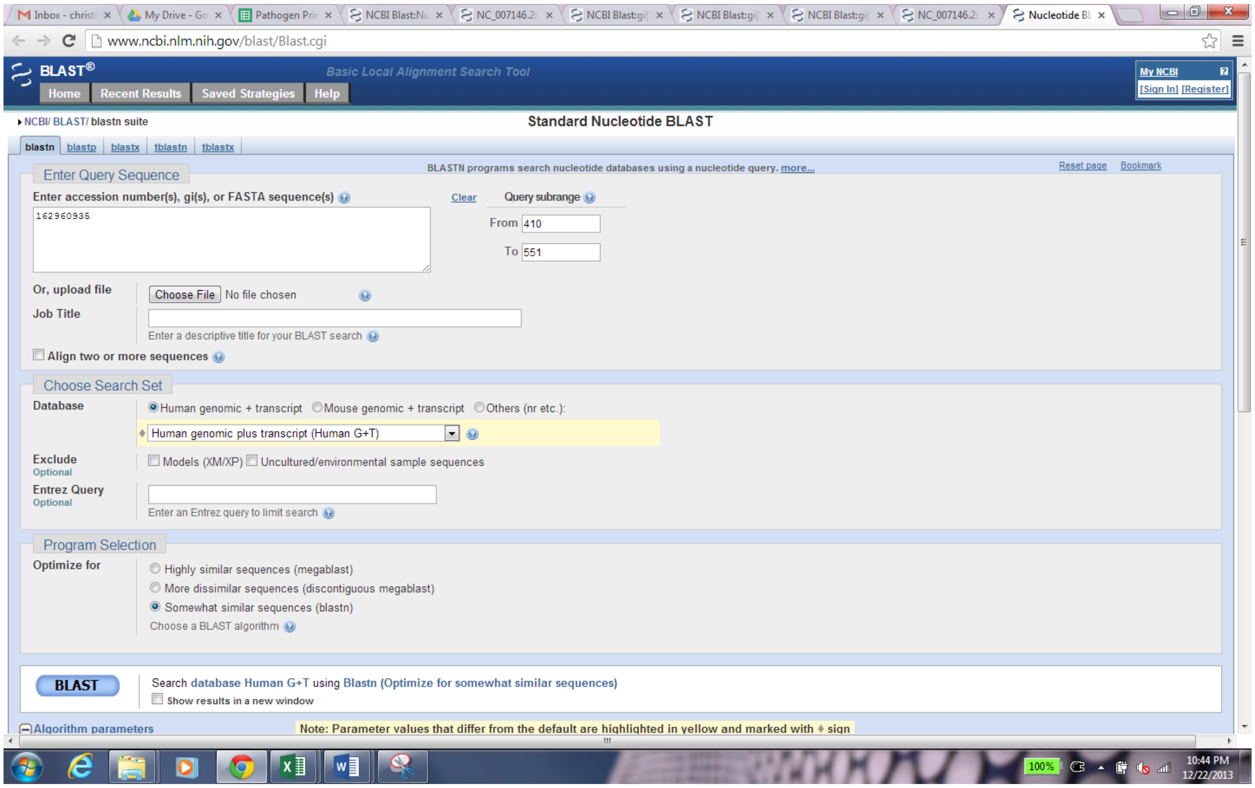
www.ncbi.nlm.nih.gov/nuccore/162960935?report=graph

Display Settings: Graphics

Haemophilus influenzae 86-028NP chromosome, complete genome

NCBI Reference Sequence: NC_007146.2

Make sure the database is Human and click BLAST.



www.ncbi.nlm.nih.gov/blast/Blast.cgi

Standard Nucleotide BLAST

Enter accession number(s), gi(s), or FASTA sequence(s): 162960935

Query subrange: From 410 To 551

Database: ● Human genomic + transcript ○ Mouse genomic + transcript ○ Others (nr etc.):

Human genomic plus transcript (Human G+T)

Exclude: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Optimize for:
○ Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
● Somewhat similar sequences (blastn)

Search database Human G+T using Blastn (Optimize for somewhat similar sequences)

☐ Show results in a new window