joanne@msl.ubc.ca

Bioinformatics

Common tools, useful databases, and tricks of the trade.

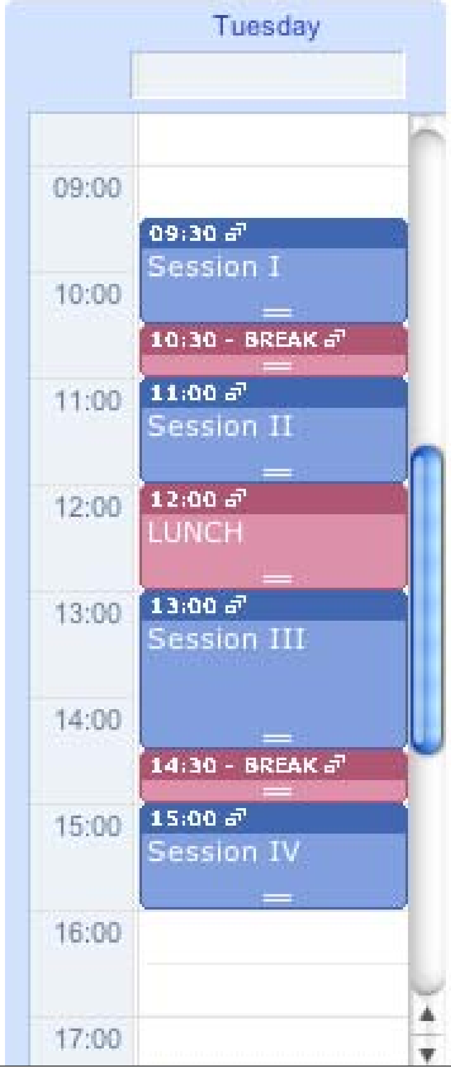
bioteach.ubc.ca/bioinfo2008

Workshop Schedule

- Laptops, available here for your use 9am - 4:30pm
- wireless login

  mslguest

  4myguest
- Vancouver guide books available



Tuesday

09:30 Session I

10:30 - BREAK

11:00 Session II

12:00 LUNCH

13:00 Session III

14:30 - BREAK

15:00 Session IV

Today’s Topics

- **BLAST -** Finding Function by Sequence Similarity
- **GUIDED TOUR** - Advanced Tips & Tricks for Using BLAST
- **PRACTICAL EXERCISES** - The Jurassic Park Detective Story
- **Genome Browsers -** Accessing Genome Annotations
- **PRACTICAL EXERCISES** - Three different views of the BRCA1 gene

BLAST

Finding Function By Sequence Similarity

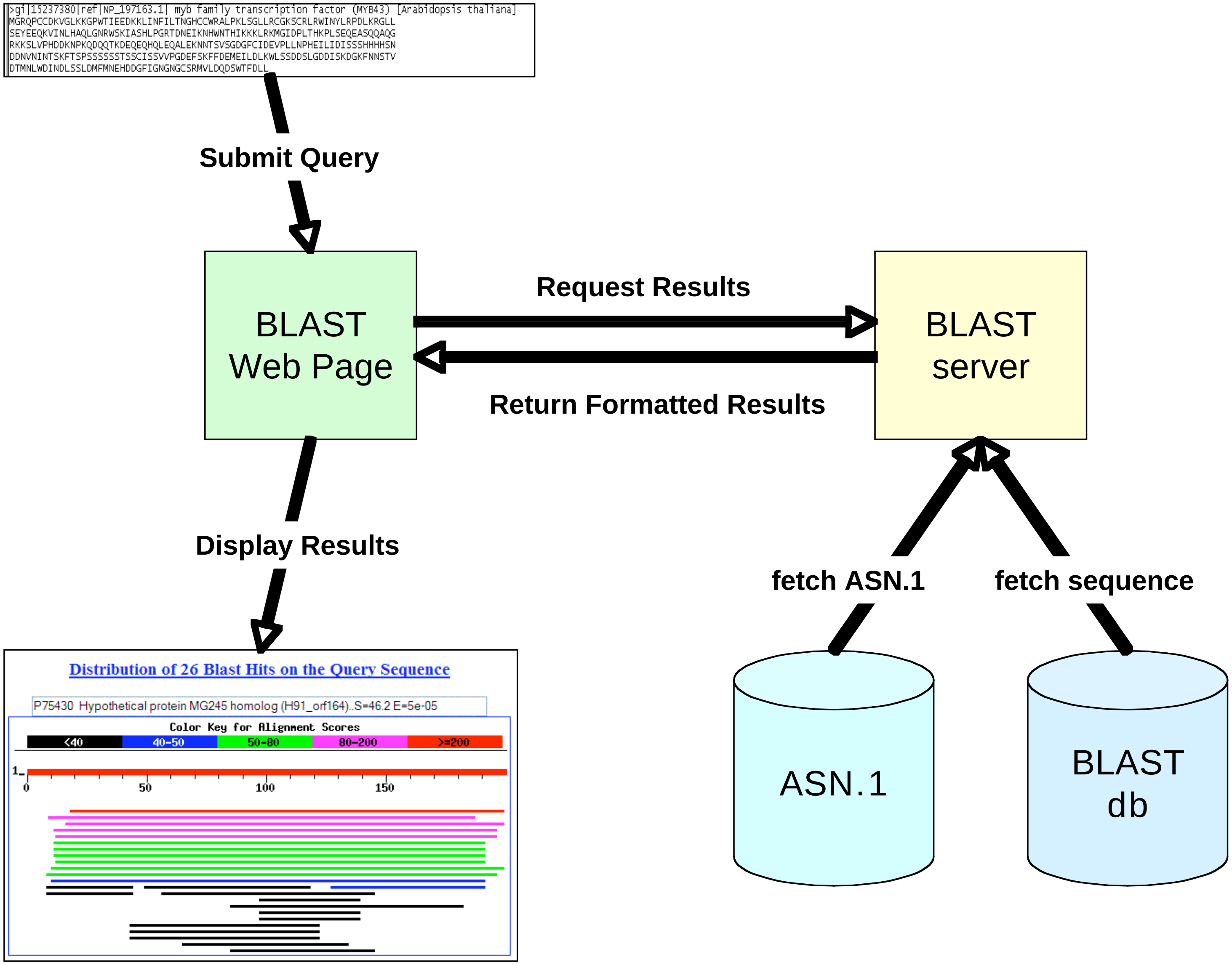
Concepts of Sequence Similarity Searching

- The premise:

  One sequence by itself is not informative; it must be analyzed by comparative methods against existing sequence databases to develop hypothesis concerning relatives and function.

The BLAST algorithm

- The BLAST programs (Basic Local Alignment Search Tools) are a set of sequence comparison algorithms introduced in 1990 that are used to search sequence databases for optimal local alignments to a query.
  - Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) “Basic local alignment search tool.” J. Mol. Biol. 215:403-410.
  - Altschul SF, Madden TL, Schaeffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) “Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.” NAR 25:3389-3402.



```
>gi|15237380|ref|NP_197163.1| myb family transcription factor (MYB43) [Arabidopsis thaliana]
MGRQPCCDKVGLKKGPWTIEEDKKLINFILTNGHCCWRALPKLSGLLRCGKSCRLRWINYLRPDLKRGLL
SEYEEQKVINLHAQLGNRWSKIASHLPGRTDNEIKNHWNTHIKKKLRKMGIDPLTHKPLSEQEASQQAQG
RKKSLVPHDDKNPKQDQQTKDEQEQHQLEQALEKNNTSVSGDGFCIDEVPLLNPHEILIDISSSHHHHSN
DDNVNINTSKFTSPSSSSSSTSSCISSVVPGDEFSKFFDEMEILDLKWLSSDDSLGDDISKDGKFNNSTV
DTMNLWDINDLSSLDMFMNEHDDGFIGNGNGCSRMVLDQDSWTFDLL
```

Submit Query

BLAST
Web Page

Request Results

Return Formatted Results

BLAST
server

Display Results

fetch ASN.1

fetch sequence

ASN.1

BLAST
db

Distribution of 26 Blast Hits on the Query Sequence

P75430 Hypothetical protein MG245 homolog (H91_orf164)..S=46.2 E=5e-05

What BLAST tells you ...

- BLAST reports surprising alignments
  - Different than chance
- Assumptions
  - Random sequences
  - Constant composition
- Conclusions
  - Surprising similarities imply evolutionary homology

Evolutionary Homology: descent from a common ancestor
Does not always imply similar function

Basic **L**ocal **A**lignment **S**earch **T**ool

- Widely used similarity search tool
- Heuristic approach based on Smith Waterman algorithm
- Finds best local alignments
- Provides statistical significance
- www, standalone, and network clients

BLAST programs

Program	Description
blastp	Compares an amino acid query sequence against a protein sequence database.
blastn	Compares a nucleotide query sequence against a nucleotide sequence database.
blastx	Compares a nucleotide query sequence translated in all reading frames against a protein sequence database. You could use this option to find potential translation products of an unknown nucleotide sequence.
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.

more BLAST programs

Program		Notes
Megablast	Contiguous	Nearly identical sequences
	Discontiguous	Cross-species comparison
Position Specific	PSI-BLAST	Automatically generates a position specific score matrix (PSSM)
	RPS-BLAST	Searches a database of PSI-BLAST PSSMs

🟨 nucleotide only

🟦 protein only

BLAST Algorithm

- Scoring of matches done using scoring matrices
- Sequences are split into words (default n=3)
  - Speed, computational efficiency
- BLAST algorithm extends the initial “seed” hit into an HSP
  - HSP = high scoring segment pair = Local optimal alignment

Sequence Similarity Searching – The statistics are important

Discriminating between real and artifactual matches is done using an estimate of probability that the match might occur by chance.

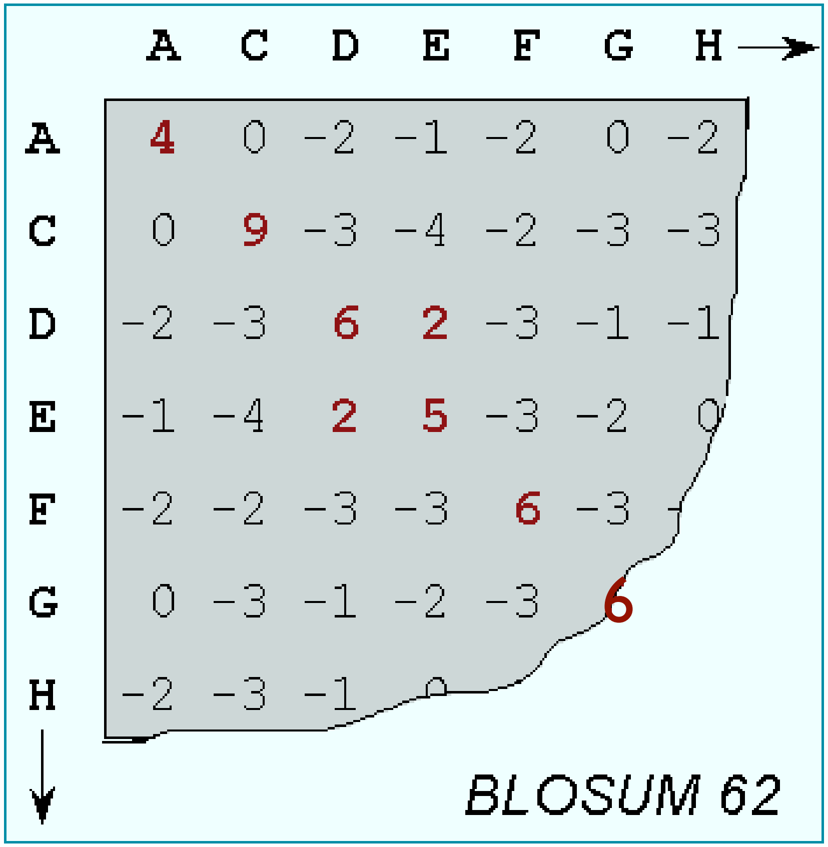
We'll talk more about the meaning of the scores (S) and e-values (E) that are associated with BLAST hits

Where does the score (S) come from?

- The quality of each pair-wise alignment is represented as a score and the scores are ranked.
- Scoring matrices are used to calculate the score of the alignment base by base (DNA) or amino acid by amino acid (protein).
- The alignment score will be the sum of the scores for each position.

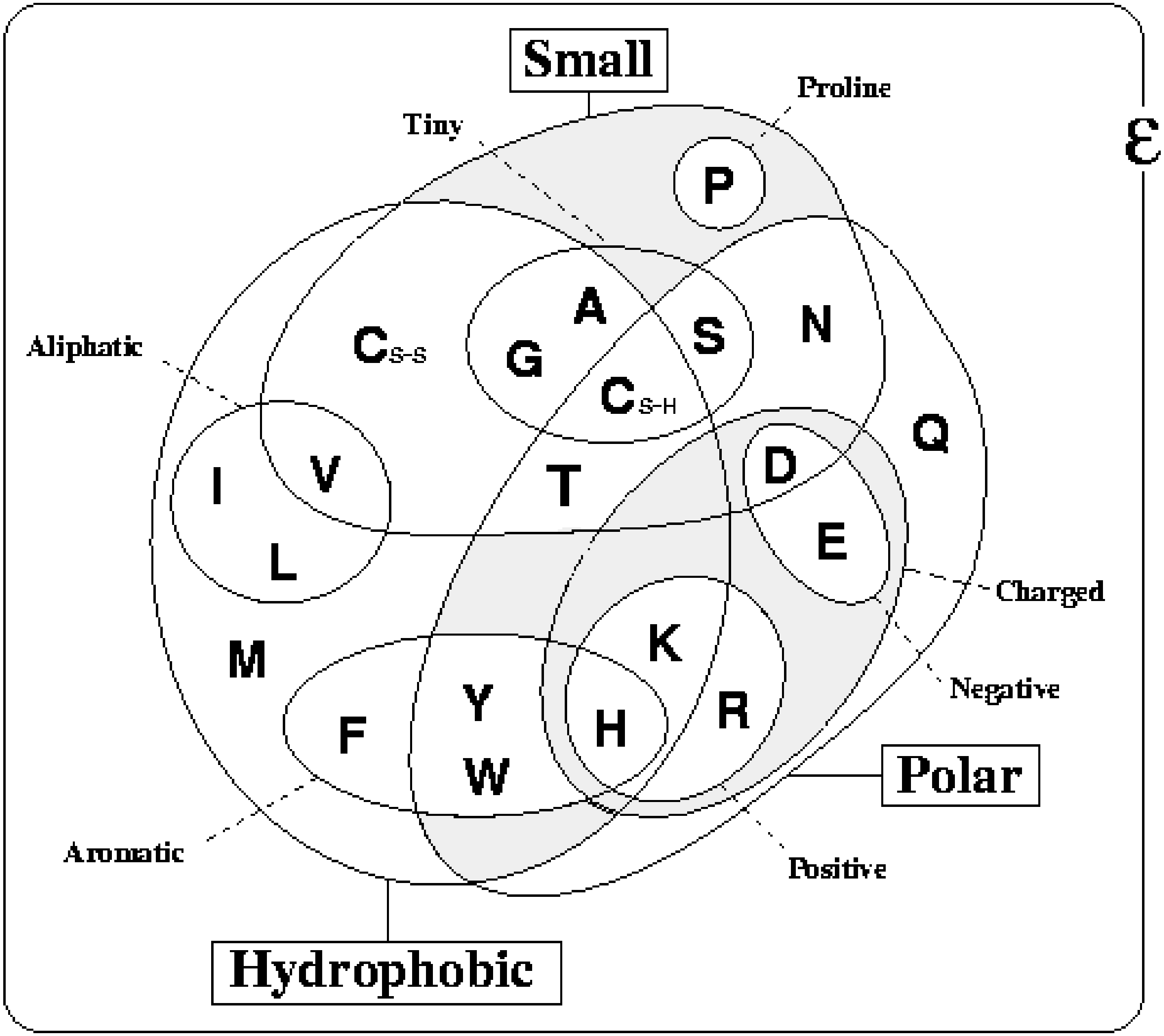
What's a scoring matrix?

- Substitution matrices are used for amino acid alignments.
  - each possible residue substitution is given a score
- A simpler unitary matrix is used for DNA pairs (+1 for match, -2 mismatch)



	A	C	D	E	F	G	H
A	4	0	-2	-1	-2	0	-2
C	0	9	-3	-4	-2	-3	-3
D	-2	-3	6	2	-3	-1	-1
E	-1	-4	2	5	-3	-2	0
F	-2	-2	-3	-3	6	-3	-
G	0	-3	-1	-2	-3	6	
H	-2	-3	-1	0			

BLOSUM 62


Small
Proline
Tiny
ε
P
A
S
N
G
Cs-s
Cs-H
Aliphatic
Q
I
V
T
D
E
L
Charged
M
K
Negative
Y
R
F
H
W
Polar
Aromatic
Positive
Hydrophobic

BLOSUM vs PAM

BLOSUM 45	**BLOSUM 62**	**BLOSUM 90**
PAM 250	**PAM 160**	**PAM 100**
More Divergent		***Less Divergent***

- BLOSUM 62 is the default matrix in BLAST 2.0. Though it is tailored for comparisons of moderately distant proteins, it performs well in detecting closer relationships. A search for distant relatives may be more sensitive with a different matrix.

What do the Score and the e-value really mean?

- The quality of the alignment is represented by the Score (S).

  The score of an alignment is calculated as the sum of substitution and gap scores. Substitution scores are given by a look-up table (PAM, BLOSUM) whereas gap scores are assigned empirically .

- The significance of each alignment is computed as an E value (E).

  Expectation value. The number of different alignments with scores equivalent to or better than S that are expected to occur in a database search by chance. The lower the E value, the more significant the score.

Notes on E-values

- Low E-values suggest that sequences are homologous
  - Can’t show non-homology
- Statistical significance depends on both the size of the alignments and the size of the sequence database
  - Important consideration for comparing results across different searches
  - E-value increases as database gets bigger
  - E-value decreases as alignments get longer

Homology: Some Guidelines

- Similarity can be indicative of homology
- Generally, if two sequences are significantly similar over entire length they are likely homologous
- Low complexity regions can be highly similar without being homologous
- Homologous sequences not always highly similar

Suggeste T
utoffs

Take Home Message:
Always look at your alignments

- Source: Chapter 11 – Bioinformatics: A Practical Guide to the Analysis of Genes and Proteins
- For nucleotide based searches, one should look for hits with E-values of 10-6 or less and sequence identity of 70% or more
- For protein based searches, one should look for hits with E-values of 10-3 or less and sequence identity of 25% or more

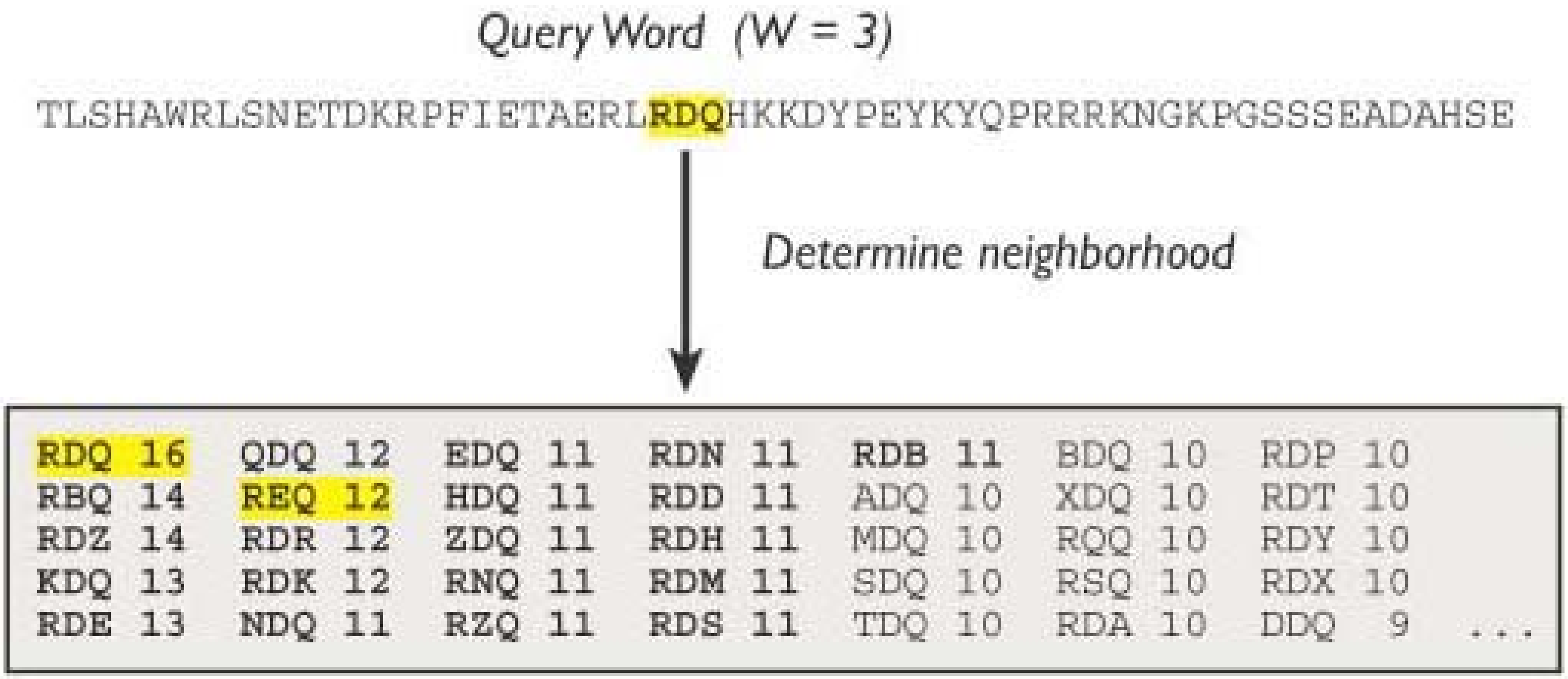
BLAST Algorithm

- Scoring of matches done using scoring matrices
- Sequences are split into words (default n=3)
  - Speed, computational efficiency
- BLAST algorithm extends the initial “seed” hit into an HSP
  - HSP = high scoring segment pair = Local optimal alignment

How Does BLAST Really Work?

- The BLAST programs improved the overall speed of searches while retaining good sensitivity (important as databases continue to grow) by breaking the query and database sequences into fragments ("words"), and initially seeking matches between fragments.
- Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S".

BLAST Algorithm



Query Word (W = 3)

TLSHAWRLSNETDKRPFIETAERL**RDQ**HKKDYPEYKYQPRRRKNGKPGSSSEADAHSE

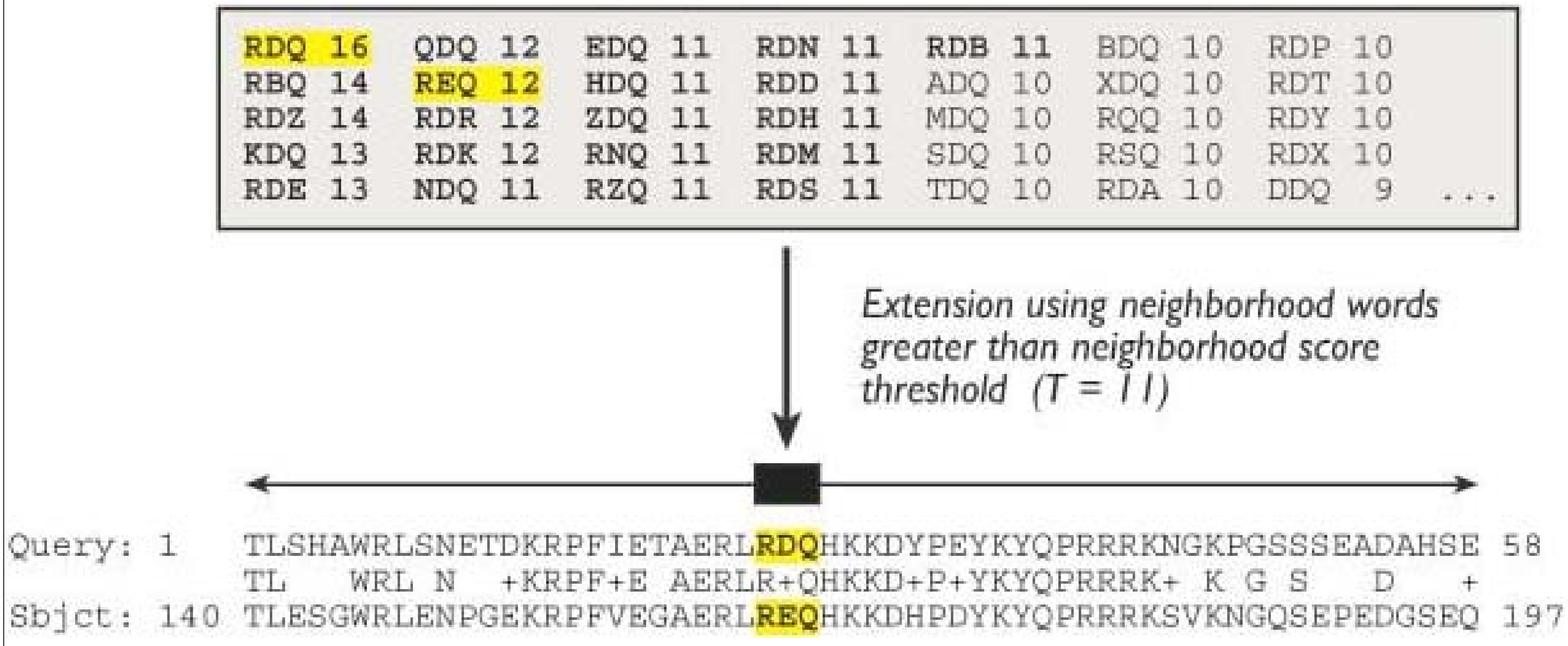
Determine neighborhood

RDQ 16	QDQ 12	EDQ 11	RDN 11	RDB 11	BDQ 10	RDP 10	
RBQ 14	REQ 12	HDQ 11	RDD 11	ADQ 10	XDQ 10	RDT 10	
RDZ 14	RDR 12	ZDQ 11	RDH 11	MDQ 10	RQQ 10	RDY 10	
KDQ 13	RDK 12	RNQ 11	RDM 11	SDQ 10	RSQ 10	RDX 10	
RDE 13	NDQ 11	RZQ 11	RDS 11	TDQ 10	RDA 10	DDQ 9	...

How Does BLAST Really Work?

- The BLAST programs improved the overall speed of searches while retaining good sensitivity (important as databases continue to grow) by breaking the query and database sequences into fragments ("words"), and initially seeking matches between fragments.
- Word hits are then extended in either direction in an attempt to generate an alignment with a score exceeding the threshold of "S".

BLAST Algorithm

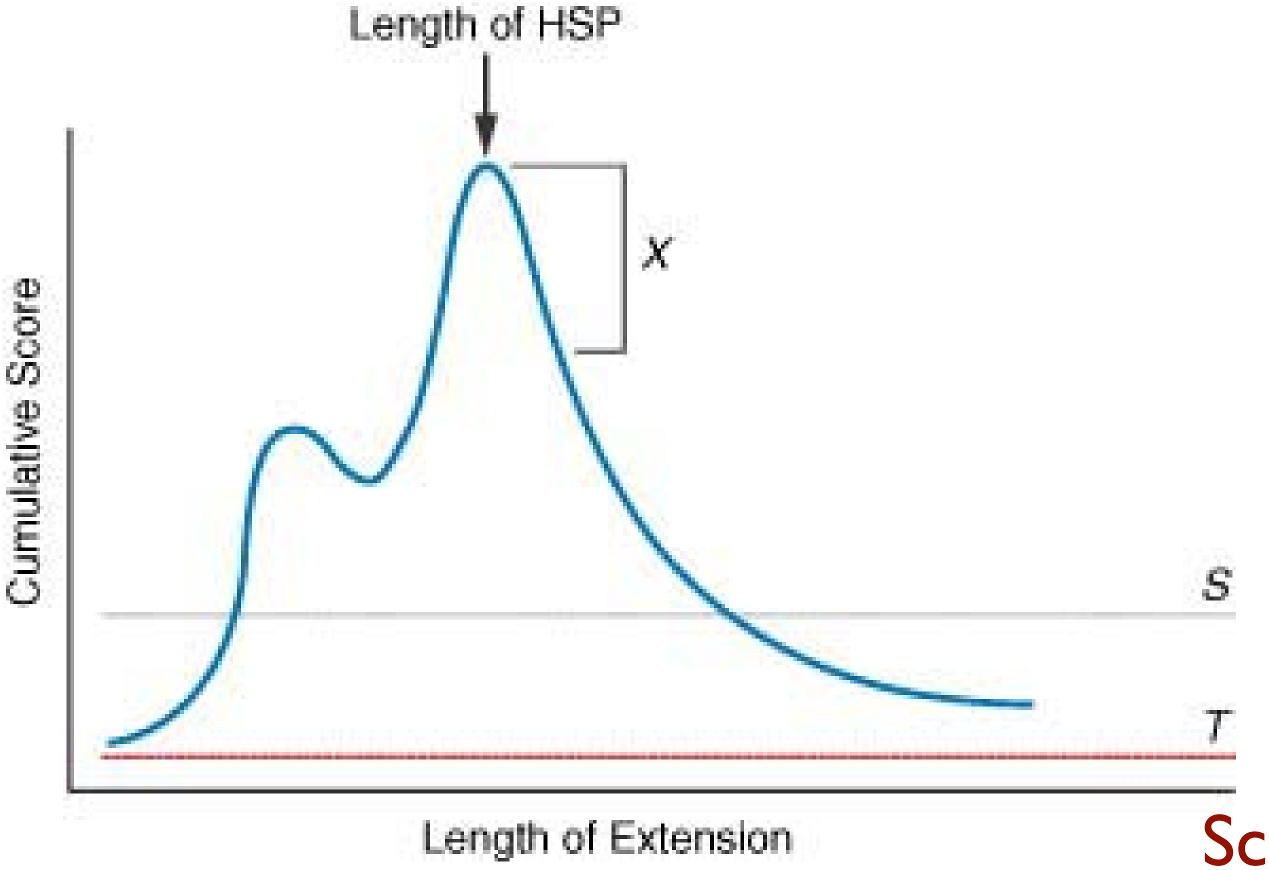


```
RDQ 16   QDQ 12   EDQ 11   RDN 11   RDB 11   BDQ 10   RDP 10
RBQ 14   REQ 12   HDQ 11   RDD 11   ADQ 10   XDQ 10   RDT 10
RDZ 14   RDR 12   ZDQ 11   RDH 11   MDQ 10   RQQ 10   RDY 10
KDQ 13   RDK 12   RNQ 11   RDM 11   SDQ 10   RSQ 10   RDX 10
RDE 13   NDQ 11   RZQ 11   RDS 11   TDQ 10   RDA 10   DDQ  9   ...
```

Extension using neighborhood words greater than neighborhood score threshold (T = 11)

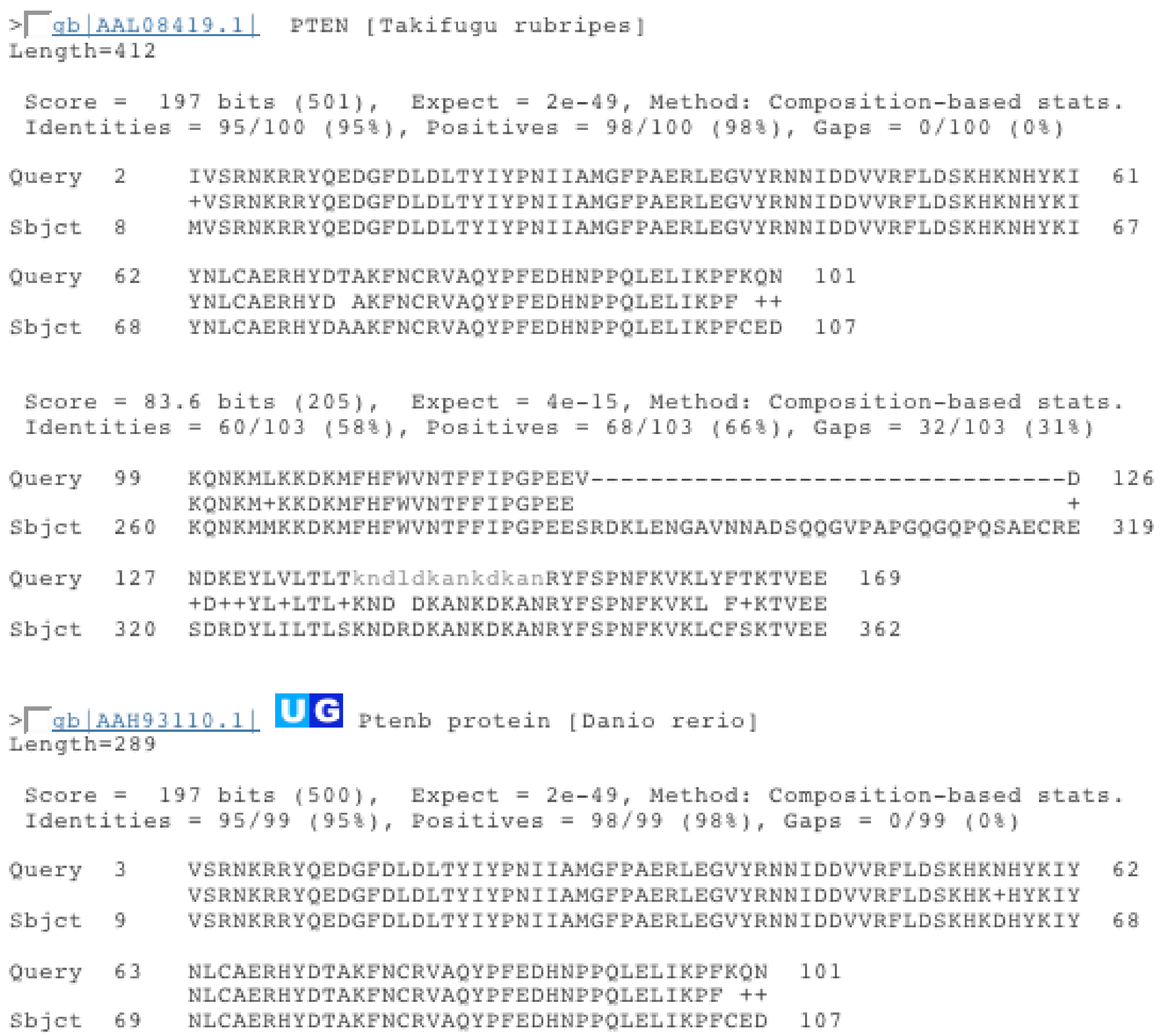
```
Query: 1   TLSHAWRLSNETDKRPFIETAERLRDQHKKDYPEYKYQPRRRKNGKPGSSSEADAHSE 58
           TL   WRL N  +KRPF+E AERLR+QHKKD+P+YKYQPRRRK+ K G S   D   +
Sbjct: 140 TLESGWRLENPGEKRPFVEGAERLREQHKKDHPDYKYQPRRRKSVKNGQSEPEDGSEQ 197
```

Extending the High Scoring Segment Pair (HSP)



Minimum Score (S)

Neighborhood Score Threshold (T)



```
>gb|AAL08419.1|  PTEN [Takifugu rubripes]
Length=412

 Score =  197 bits (501),  Expect = 2e-49, Method: Composition-based stats.
 Identities = 95/100 (95%), Positives = 98/100 (98%), Gaps = 0/100 (0%)

Query  2    IVSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKI  61
            +VSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKI
Sbjct  8    MVSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKI  67

Query  62   YNLCAERHYDTAKFNCRVAQYPFEDHNPPQLELIKPFKQN  101
            YNLCAERHYD AKFNCRVAQYPFEDHNPPQLELIKPF ++
Sbjct  68   YNLCAERHYDAAKFNCRVAQYPFEDHNPPQLELIKPFCED  107


 Score = 83.6 bits (205),  Expect = 4e-15, Method: Composition-based stats.
 Identities = 60/103 (58%), Positives = 68/103 (66%), Gaps = 32/103 (31%)

Query  99   KQNKMLKKDKMFHFWVNTFFIPGPEEV--------------------------------D  126
            KQNKM+KKDKMFHFWVNTFFIPGPEE                                 +
Sbjct  260  KQNKMMKKDKMFHFWVNTFFIPGPEESRDKLENGAVNNADSQQGVPAPGQGQPQSAECRE  319

Query  127  NDKEYLVLTLTkndldkankdkanRYFSPNFKVKLYFTKTVEE  169
            +D++YL+LTL+KND DKANKDKANRYFSPNFKVKL F+KTVEE
Sbjct  320  SDRDYLILTLSKNDRDKANKDKANRYFSPNFKVKLCFSKTVEE  362


>gb|AAH93110.1| UG Ptenb protein [Danio rerio]
Length=289

 Score =  197 bits (500),  Expect = 2e-49, Method: Composition-based stats.
 Identities = 95/99 (95%), Positives = 98/99 (98%), Gaps = 0/99 (0%)

Query  3    VSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKNHYKIY  62
            VSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHK+HYKIY
Sbjct  9    VSRNKRRYQEDGFDLDLTYIYPNIIAMGFPAERLEGVYRNNIDDVVRFLDSKHKDHYKIY  68

Query  63   NLCAERHYDTAKFNCRVAQYPFEDHNPPQLELIKPFKQN  101
            NLCAERHYDTAKFNCRVAQYPFEDHNPPQLELIKPF ++
Sbjct  69   NLCAERHYDTAKFNCRVAQYPFEDHNPPQLELIKPFCED  107
```

BLAST Algorithm

- Scoring of matches done using scoring matrices
- Sequences are split into words (default n=3)
  - Speed, computational efficiency
- BLAST algorithm extends the initial "seed" hit into an HSP
  - HSP = high scoring segment pair = Local optimal alignment

Credits

- Materials for this presentation have been adapted from the following sources:

  NCBI HelpDesk - Field Guide Course Materials

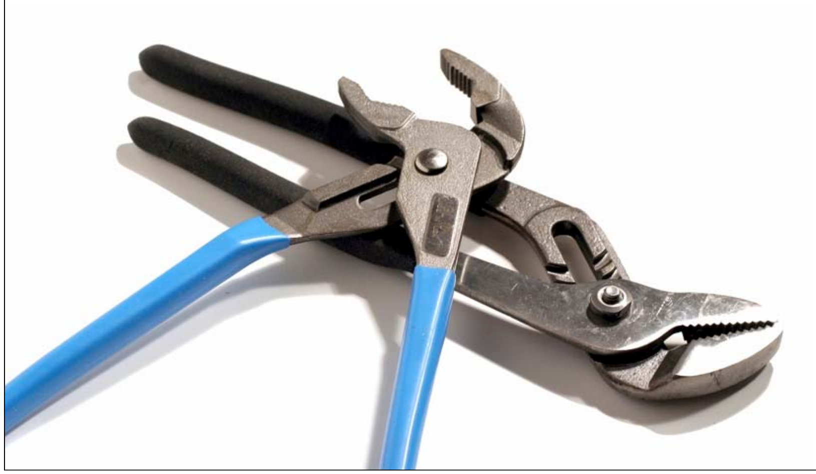
  Bioinformatics: A practical guide to the analysis of genes and proteins

- Questions? Please contact:

Dr. Joanne Fox
Michael Smith Laboratories
joanne@msl.ubc.ca

BLAST

GUIDED TOUR: Advanced Tips & Tricks for Using BLAST



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

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

Learn more about how to use the new BLAST design

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

find human sequences in a database that can be amplified with a particular primer pair.

A frequent use of nucleotide-nucleotide BLAST is to check the specificity of oligonucleotides for hybridization in PCR. The goal is usually to make sure that the primers will give a unique product from the target genome or cDNA population. Because BLAST is local and

News

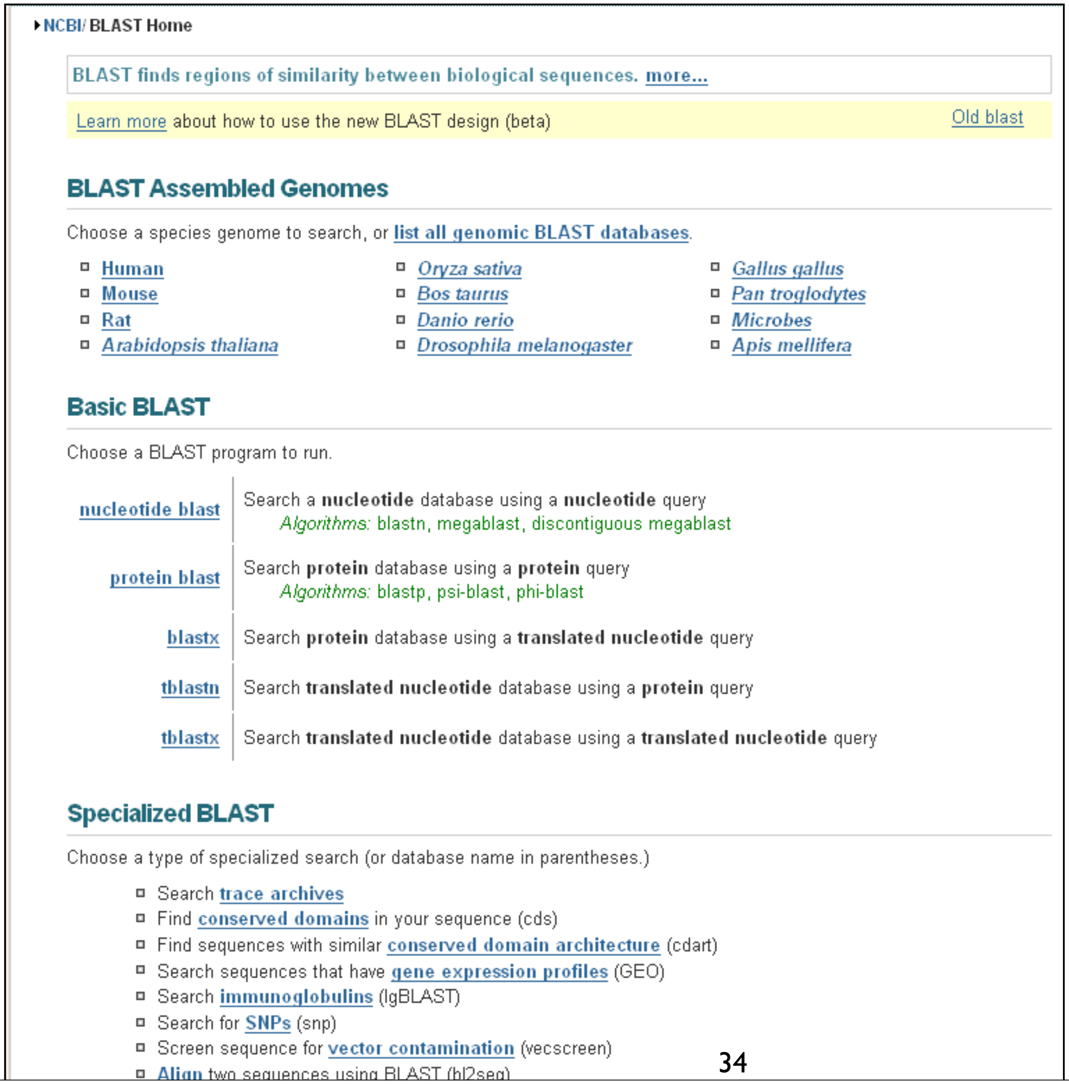
Old BLAST Web Pages to be deleted June 11th 2007

As previously announced access to the old pages will be removed on June 11, 2007.

2007-06-01 12:15:00

More BLAST news...

New BLAST homepage



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

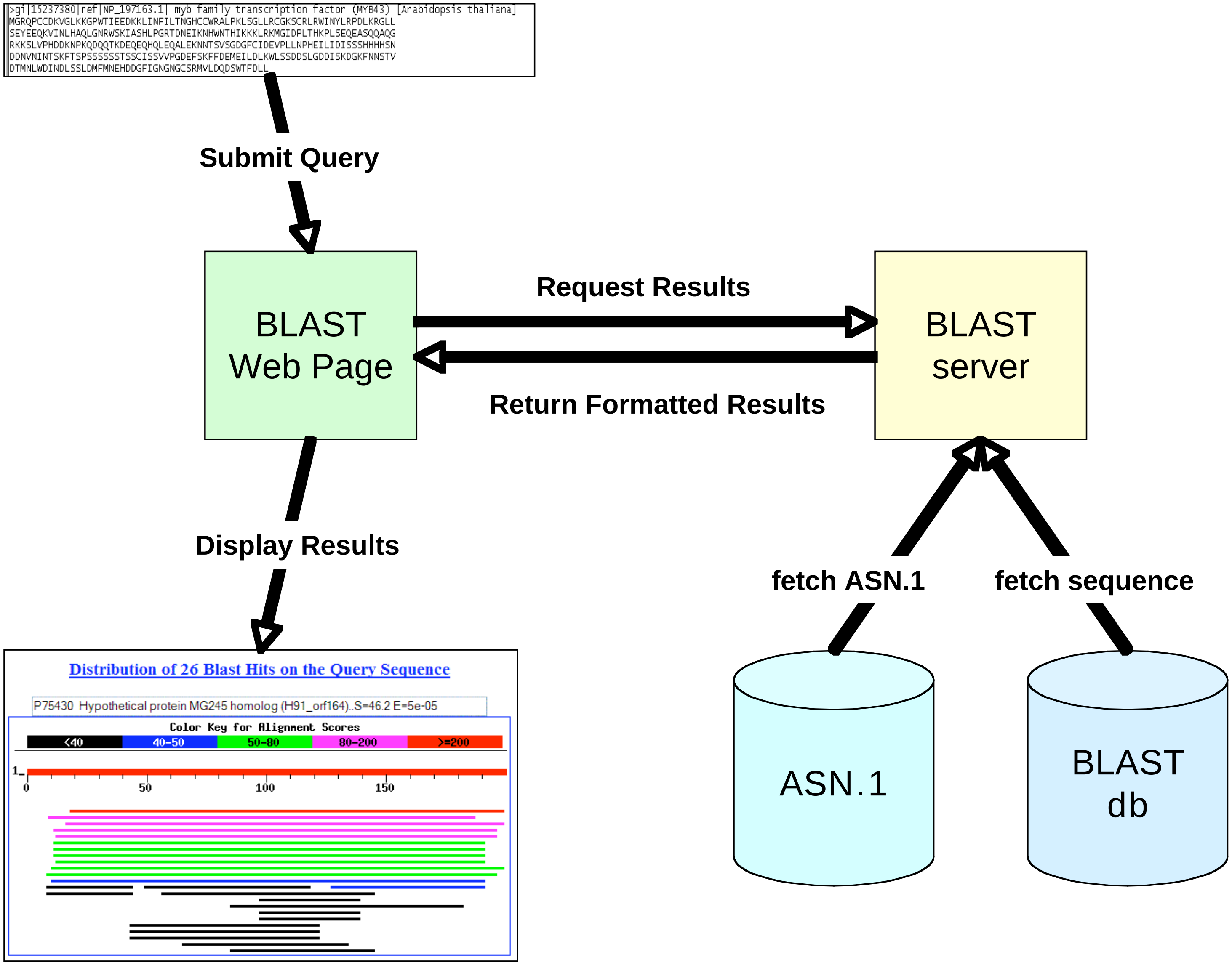
Choose a BLAST program to run.

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)



```
>gi|15237380|ref|NP_197163.1| myb family transcription factor (MYB43) [Arabidopsis thaliana]
MGRQPCCDKVGLKKGPWTIEEDKKLINFILTNGHCCWRALPKLSGLLRCGKSCRLRWINYLRPDLKRGLL
SEYEEQKVINLHAQLGNRWSKIASHLPGRTDNEIKNHWNTHIKKKLRKMGIDPLTHKPLSEQEASQQAQG
RKKSLVPHDDKNPKQDQQTKDEQEQHQLEQALEKNNTSVSGDGFCIDEVPLLNPHEILIDISSSHHHHSN
DDNVNINTSKFTSPSSSSSSTSSCISSVVPGDEFSKFFDEMEILDLKWLSSDDSLGDDISKDGKFNNSTV
DTMNLWDINDLSSLDMFMNEHDDGFIGNGNGCSRMVLDQDSWTFDLL
```

Consider your research question ...

- Are you looking for an particular gene in a particular species?
- Are you looking for additional members of a protein family across all species?
- Are you looking to annotate genes in your species of interest?

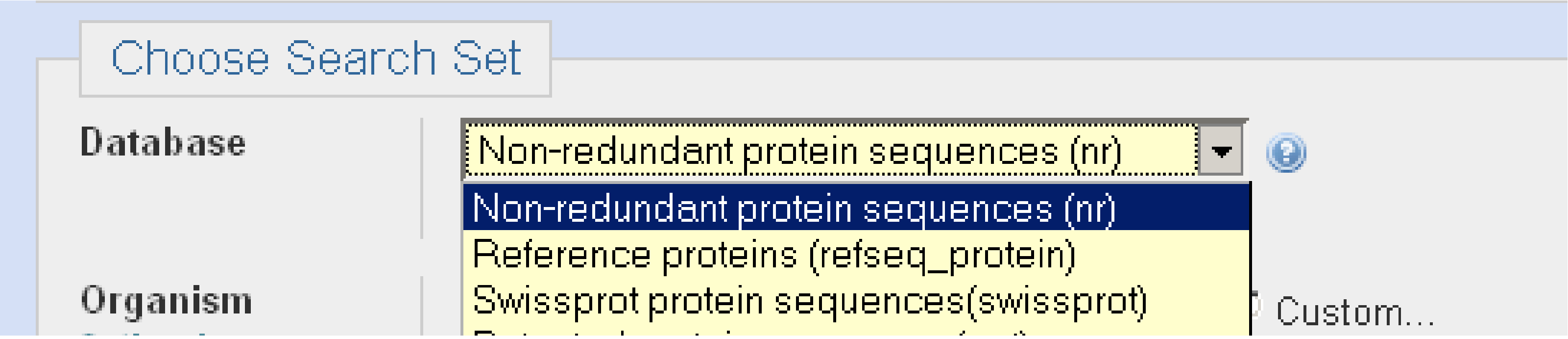
Know your reagents

- Changing your choice of database is changing your search space

- Database size affects the BLAST statistics

- Databases change rapidly and are updated frequently

Protein Databases: nr



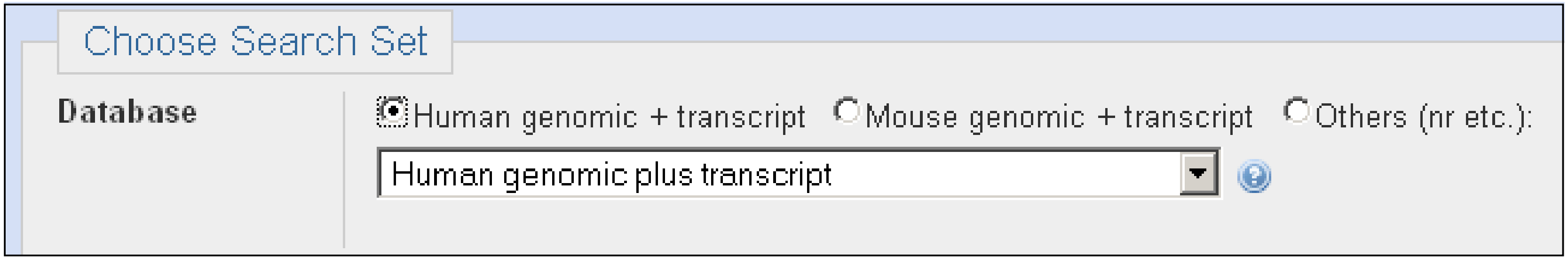
Choose Search Set

Database: Non-redundant protein sequences (nr)

- nr (non-redundant protein sequences)
  - GenBank CDS translations
  - NP_ RefSeqs
  - Outside Protein
    - PIR, Swiss-Prot, PRF
    - PDB (sequences from structures)
- pat protein patents
- env_nr environmental samples

<u>Services</u>
blastp
blastx

Nucleotide Databases: Human and Mouse



Choose Search Set

Database

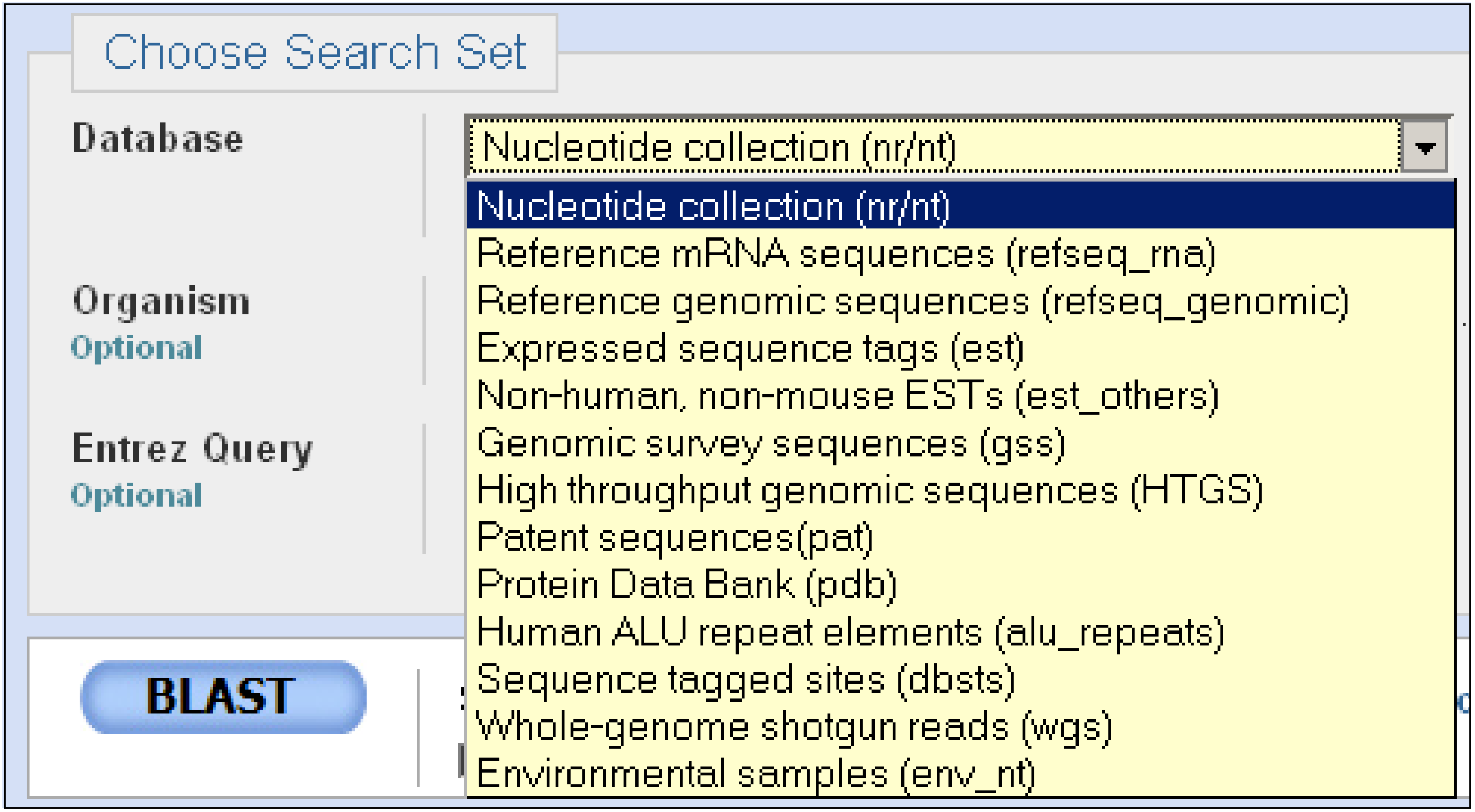
● Human genomic + transcript ○ Mouse genomic + transcript ○ Others (nr etc.):

Human genomic plus transcript

- Human and mouse genomic + transcript default
- Separate sections in output for mRNA and genomic
- Direct links to Map Viewer for genomic sequences

Megablast, blastn service

Nucleotide Databases: Traditional



Database: Nucleotide collection (nr/nt)

<u>Services</u>

blastn

tblastn

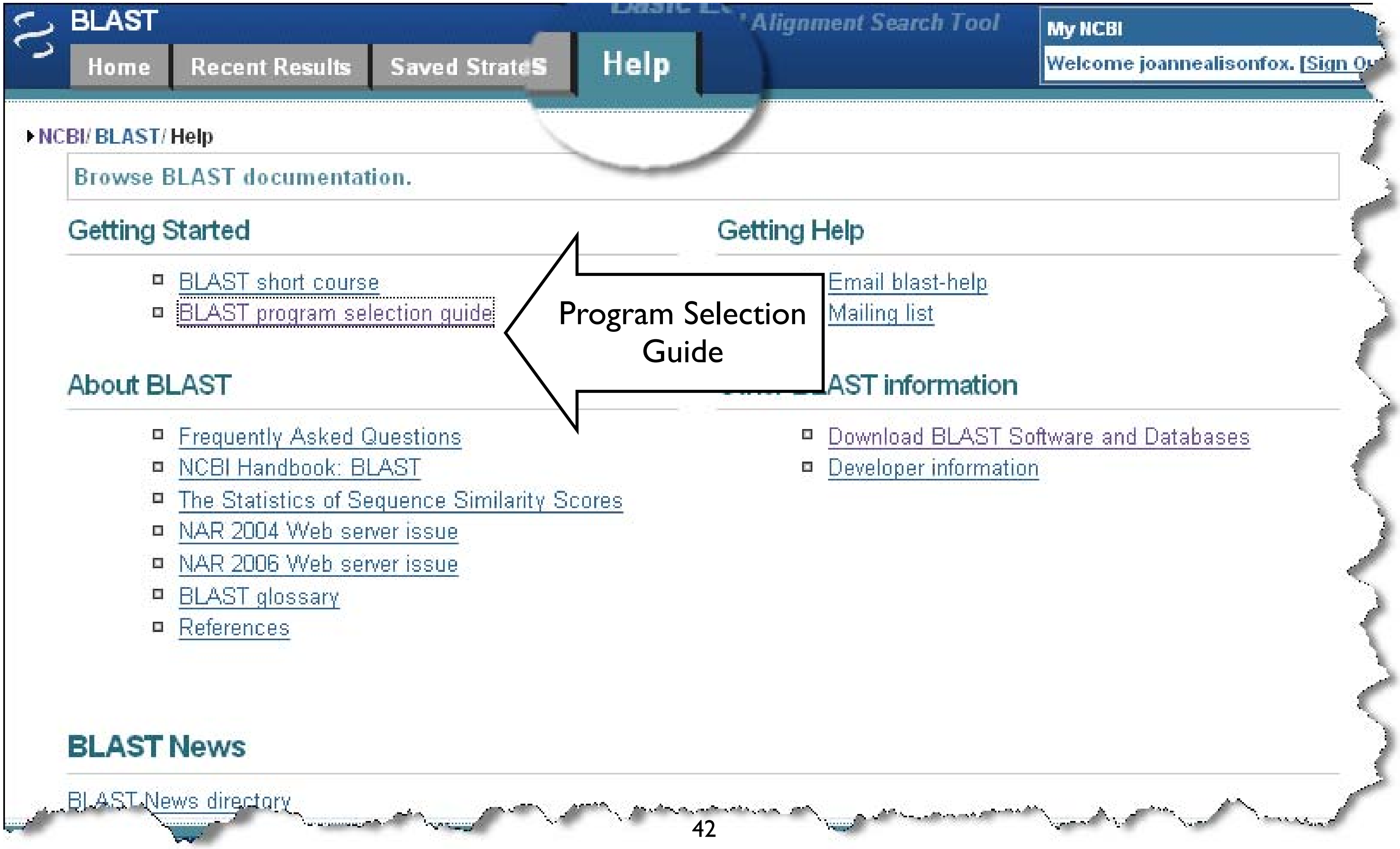
tblastx

Nucleotide Databases: Traditional

- **nr (nt)**
  - Traditional GenBank
  - NM_ and XM_ RefSeqs
    - refseq_rna
- **refseq_genomic**
  - NC_ RefSeqs
- **dbest**
  - EST Division
    - est_human, mouse, others
- **htgs**
  - HTG division
- **gss**
  - GSS division
- **wgs**
  - whole genome shotgun
- **env_nt**
  - environmental samples

Databases are mostly non-overlapping

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



NCBI/ BLAST/ Help

Browse BLAST documentation.

Getting Started

- BLAST short course
- BLAST program selection guide

Program Selection Guide

Getting Help

- Email blast-help
- Mailing list

About BLAST

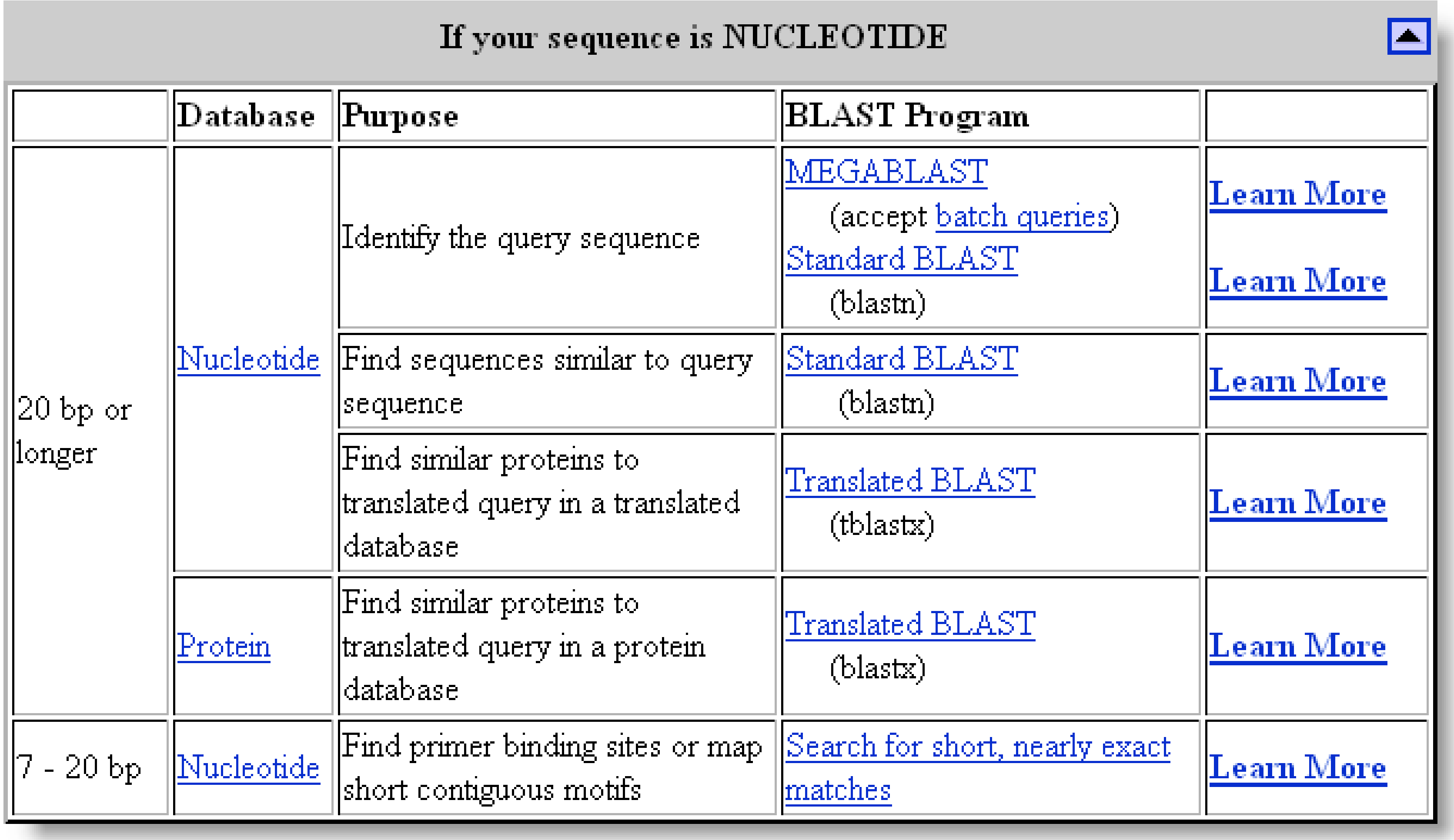
- Frequently Asked Questions
- NCBI Handbook: BLAST
- The Statistics of Sequence Similarity Scores
- NAR 2004 Web server issue
- NAR 2006 Web server issue
- BLAST glossary
- References

AST information

- Download BLAST Software and Databases
- Developer information

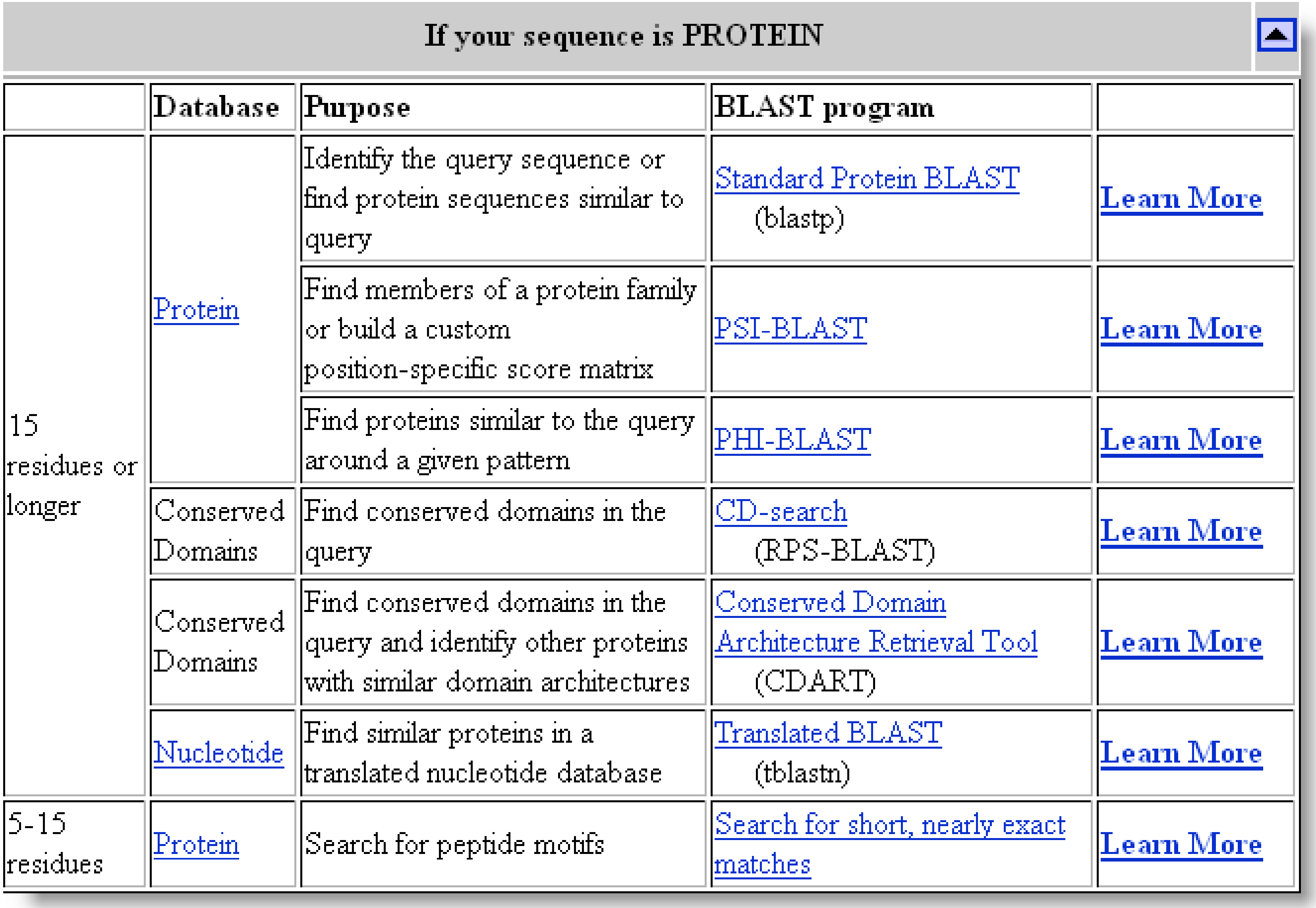
BLAST News

BLAST News directory



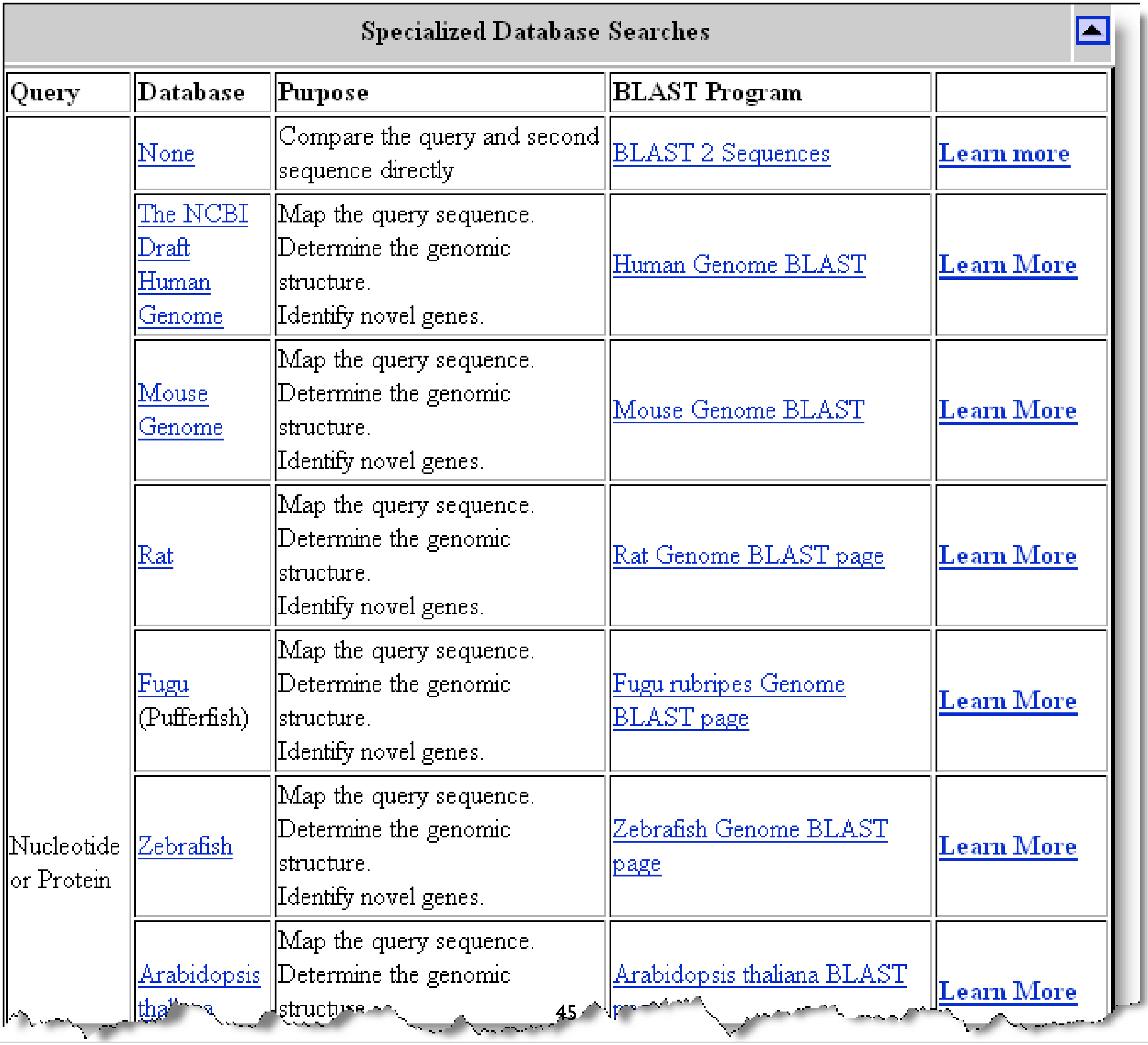
If your sequence is NUCLEOTIDE

	Database	Purpose	BLAST Program	
20 bp or longer	Nucleotide	Identify the query sequence	MEGABLAST (accept batch queries) Standard BLAST (blastn)	Learn More Learn More
		Find sequences similar to query sequence	Standard BLAST (blastn)	Learn More
		Find similar proteins to translated query in a translated database	Translated BLAST (tblastx)	Learn More
	Protein	Find similar proteins to translated query in a protein database	Translated BLAST (blastx)	Learn More
7 - 20 bp	Nucleotide	Find primer binding sites or map short contiguous motifs	Search for short, nearly exact matches	Learn More



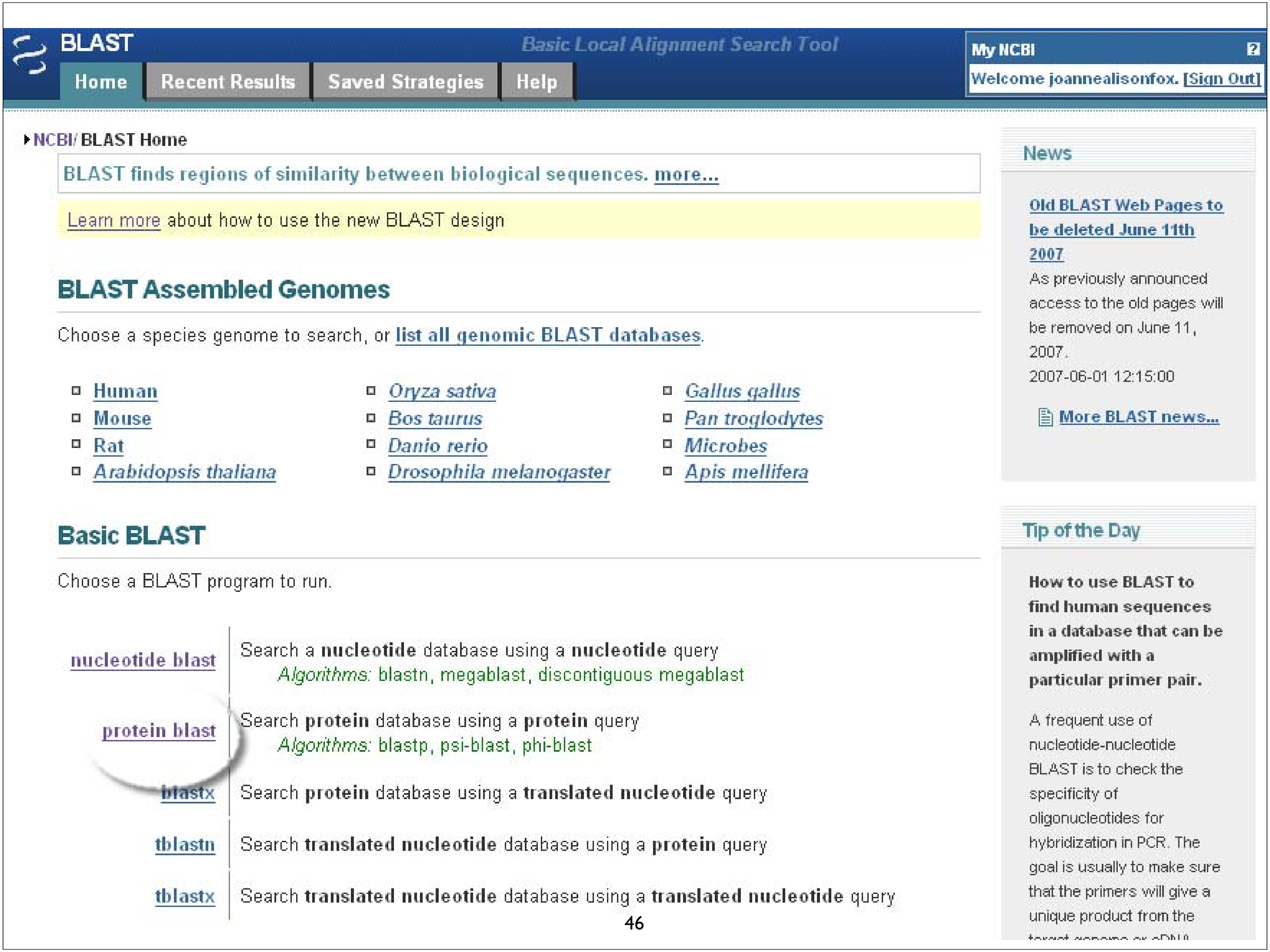
If your sequence is PROTEIN

	Database	Purpose	BLAST program	
15 residues or longer	Protein	Identify the query sequence or find protein sequences similar to query	Standard Protein BLAST (blastp)	**Learn More**
		Find members of a protein family or build a custom position-specific score matrix	PSI-BLAST	**Learn More**
		Find proteins similar to the query around a given pattern	PHI-BLAST	**Learn More**
	Conserved Domains	Find conserved domains in the query	CD-search (RPS-BLAST)	**Learn More**
	Conserved Domains	Find conserved domains in the query and identify other proteins with similar domain architectures	Conserved Domain Architecture Retrieval Tool (CDART)	**Learn More**
	Nucleotide	Find similar proteins in a translated nucleotide database	Translated BLAST (tblastn)	**Learn More**
5-15 residues	Protein	Search for peptide motifs	Search for short, nearly exact matches	**Learn More**



Specialized Database Searches

Query	Database	Purpose	BLAST Program	
Nucleotide or Protein	None	Compare the query and second sequence directly	BLAST 2 Sequences	Learn more
	The NCBI Draft Human Genome	Map the query sequence. Determine the genomic structure. Identify novel genes.	Human Genome BLAST	Learn More
	Mouse Genome	Map the query sequence. Determine the genomic structure. Identify novel genes.	Mouse Genome BLAST	Learn More
	Rat	Map the query sequence. Determine the genomic structure. Identify novel genes.	Rat Genome BLAST page	Learn More
	Fugu (Pufferfish)	Map the query sequence. Determine the genomic structure. Identify novel genes.	Fugu rubripes Genome BLAST page	Learn More
	Zebrafish	Map the query sequence. Determine the genomic structure. Identify novel genes.	Zebrafish Genome BLAST page	Learn More
	Arabidopsis tha	Map the query sequence. Determine the genomic structu	Arabidopsis thaliana BLAST	Learn More



Welcome joannealisonfox. [Sign Out]

NCBI/ BLAST Home

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

Learn more about how to use the new BLAST design

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

News

Old BLAST Web Pages to be deleted June 11th 2007

As previously announced access to the old pages will be removed on June 11, 2007.

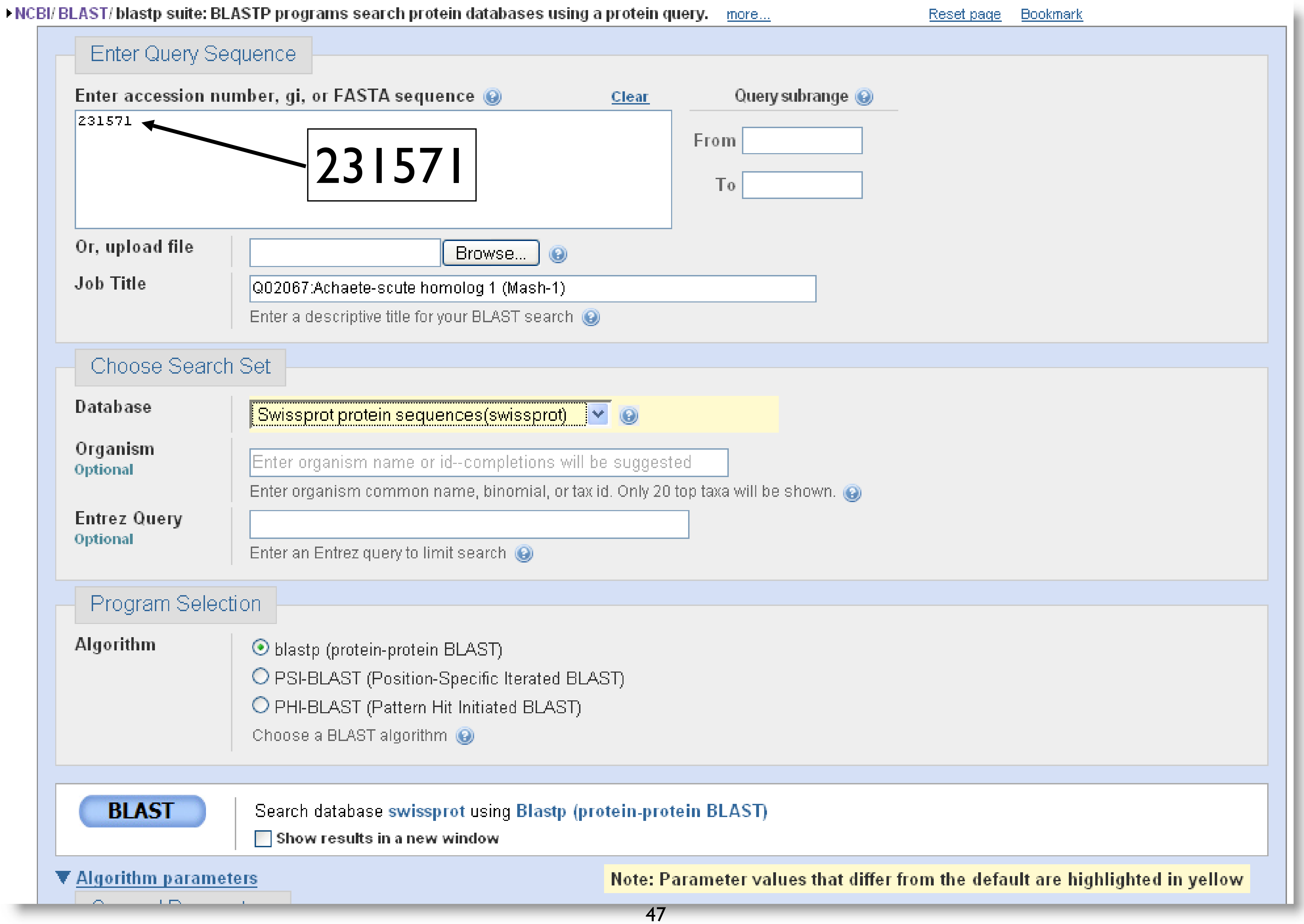
2007-06-01 12:15:00

More BLAST news...

Tip of the Day

How to use BLAST to find human sequences in a database that can be amplified with a particular primer pair.

A frequent use of nucleotide-nucleotide BLAST is to check the specificity of oligonucleotides for hybridization in PCR. The goal is usually to make sure that the primers will give a unique product from the



NCBI/ BLAST/ blastp suite: BLASTP programs search protein databases using a protein query. more...

Reset page Bookmark

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
231571
```

231571

Query subrange

From

To

Or, upload file

Job Title: Q02067:Achaete-scute homolog 1 (Mash-1)

Enter a descriptive title for your BLAST search

Choose Search Set

Database: Swissprot protein sequences(swissprot)

Organism
Optional

Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Entrez Query
Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

● blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)

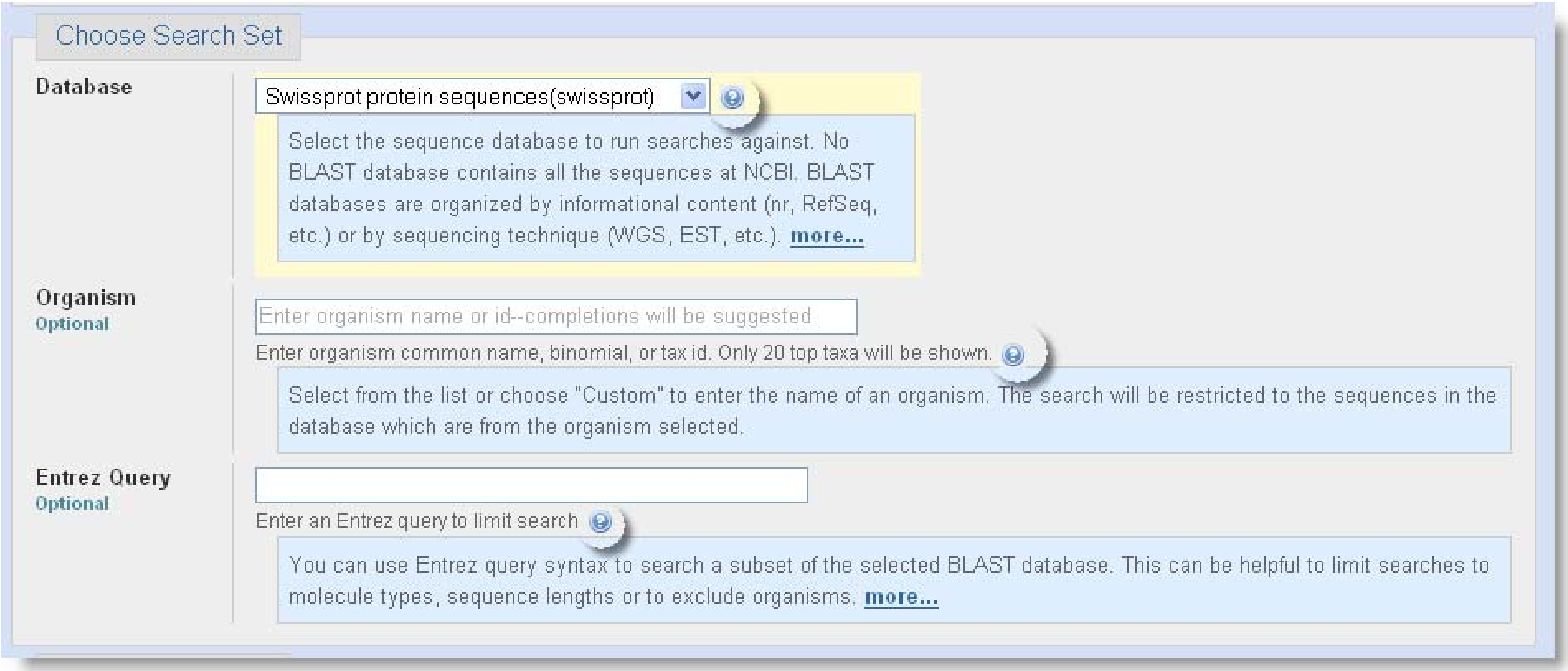
Choose a BLAST algorithm

Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow

Context Specific Help



Select the sequence database to run searches against. No BLAST database contains all the sequences at NCBI. BLAST databases are organized by informational content (nr, RefSeq, etc.) or by sequencing technique (WGS, EST, etc.). more...

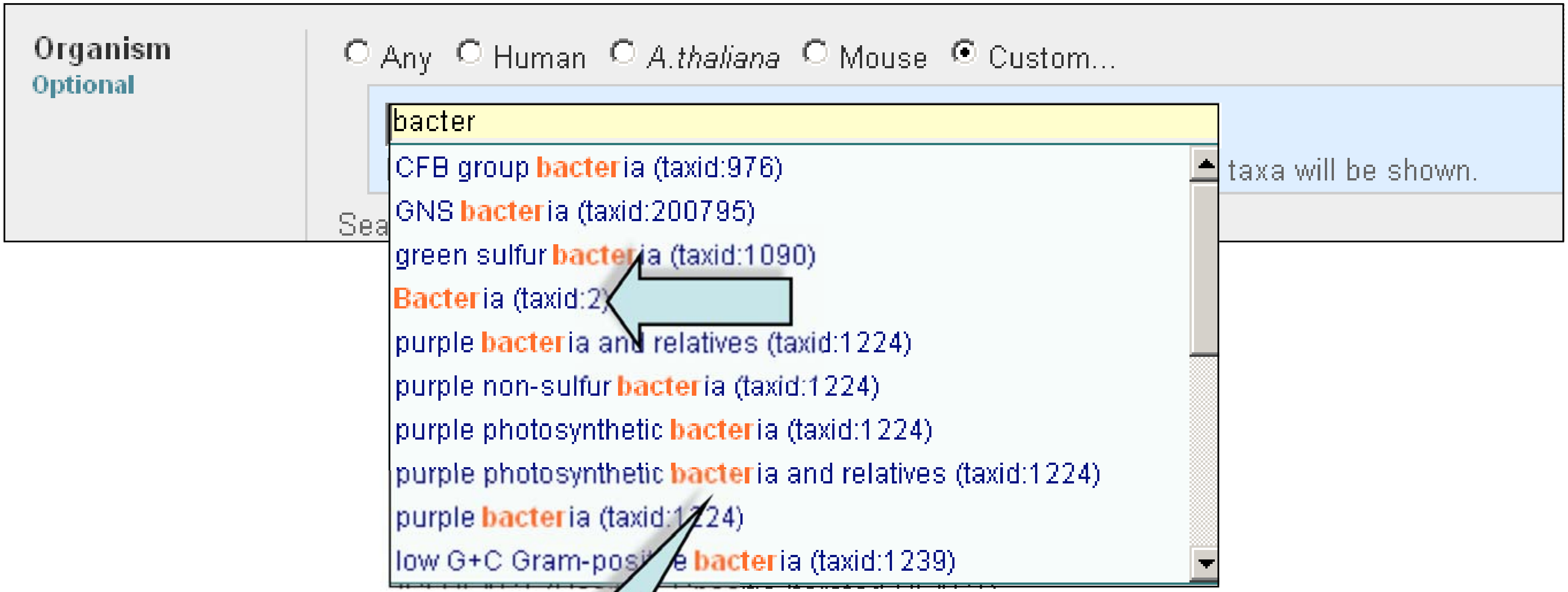
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Select from the list or choose "Custom" to enter the name of an organism. The search will be restricted to the sequences in the database which are from the organism selected.

Enter an Entrez query to limit search

You can use Entrez query syntax to search a subset of the selected BLAST database. This can be helpful to limit searches to molecule types, sequence lengths or to exclude organisms. more...

Limiting Database: Organism



Organism
Optional

○ Any ○ Human ○ *A.thaliana* ○ Mouse ● Custom...

bacter

- CFB group bacteria (taxid:976)
- GNS bacteria (taxid:200795)
- green sulfur bacteria (taxid:1090)
- Bacteria (taxid:2)
- purple bacteria and relatives (taxid:1224)
- purple non-sulfur bacteria (taxid:1224)
- purple photosynthetic bacteria (taxid:1224)
- purple photosynthetic bacteria and relatives (taxid:1224)
- purple bacteria (taxid:1224)
- low G+C Gram-positive bacteria (taxid:1239)

taxa will be shown.

Sea

Organism autocomplete

Limiting Database: Entrez Query

Entrez Query
Optional

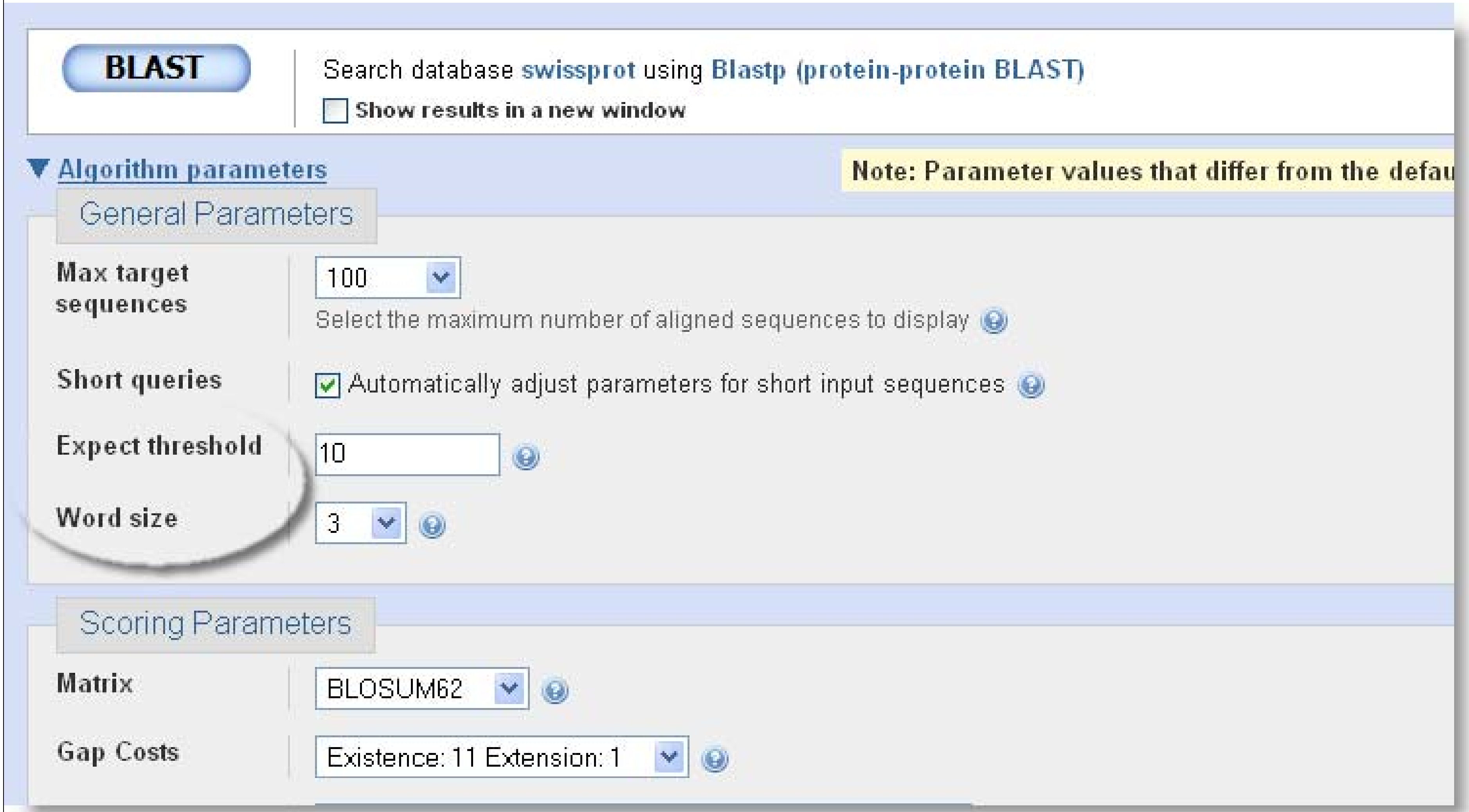
all[filter] NOT mammals[organism]

Enter an Entrez query to limit search

all[filter] NOT mammals[organism]

gene_in_mitochondrion[Properties]
2006:2007 [Modification Date]

Nucleotide
biomol_mrna[Properties]
biomol_genomic[Properties]



Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window

Note: Parameter values that differ from the defau

General Parameters

Max target sequences: 100

Select the maximum number of aligned sequences to display

Short queries: ☑ Automatically adjust parameters for short input sequences

Expect threshold: 10

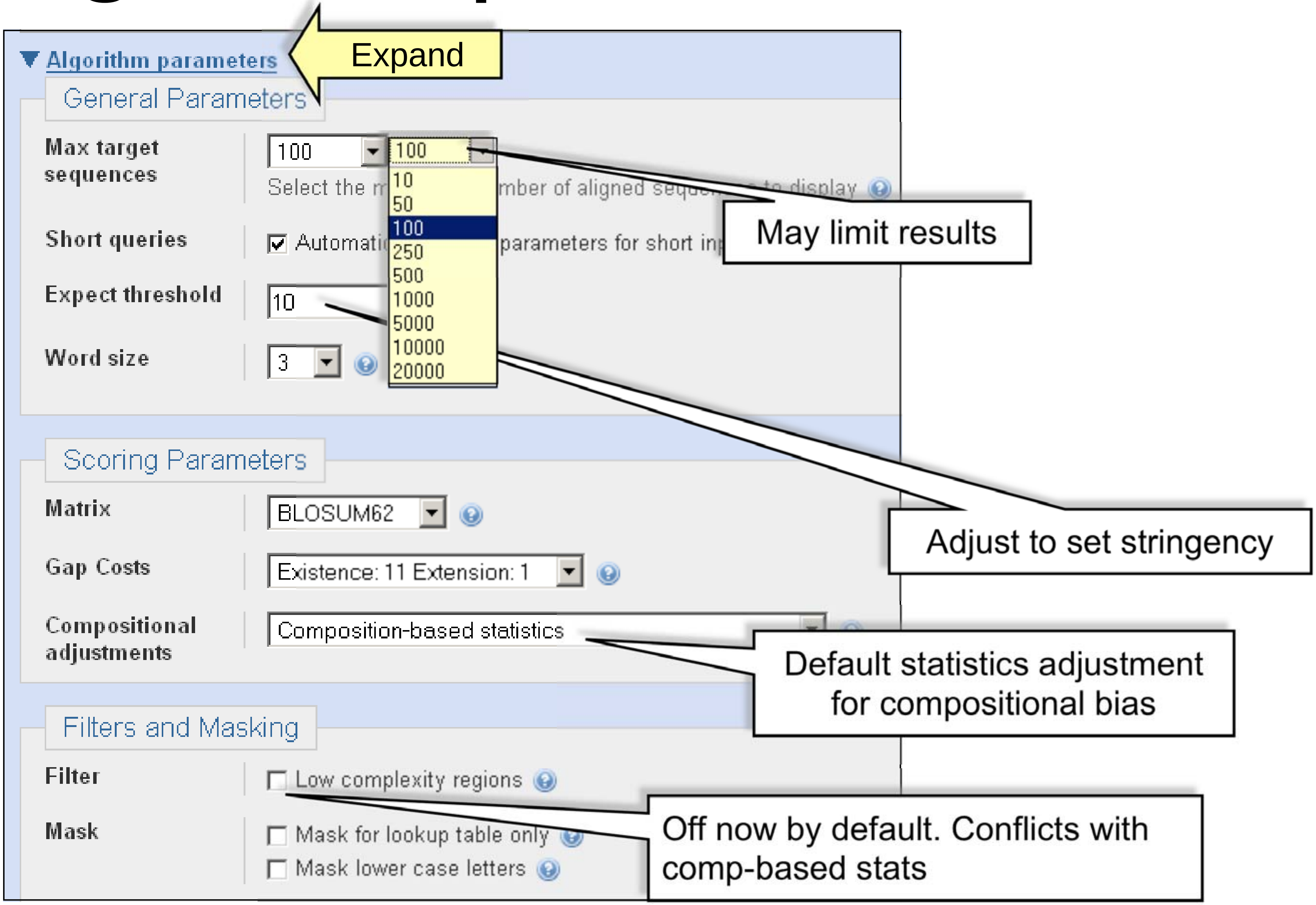
Word size: 3

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Algorithm parameters: Protein



▼ Algorithm parameters

Expand

General Parameters

Max target sequences: 100

10
50
100
250
500
1000
5000
10000
20000

Select the m mber of aligned seque to display

May limit results

Short queries: ☑ Automati parameters for short inp

Expect threshold: 10

Word size: 3

Adjust to set stringency

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Composition-based statistics

Default statistics adjustment for compositional bias

Filters and Masking

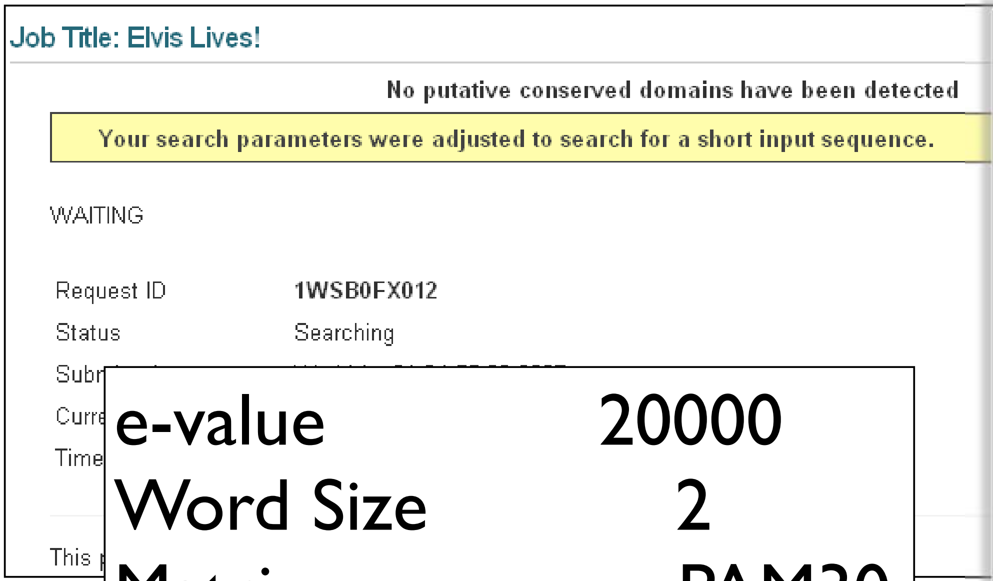
Filter: ☐ Low complexity regions

Mask: ☐ Mask for lookup table only

☐ Mask lower case letters

Off now by default. Conflicts with comp-based stats

Automatic Short Sequence Adjustment



Job Title: Elvis Lives!

No putative conserved domains have been detected

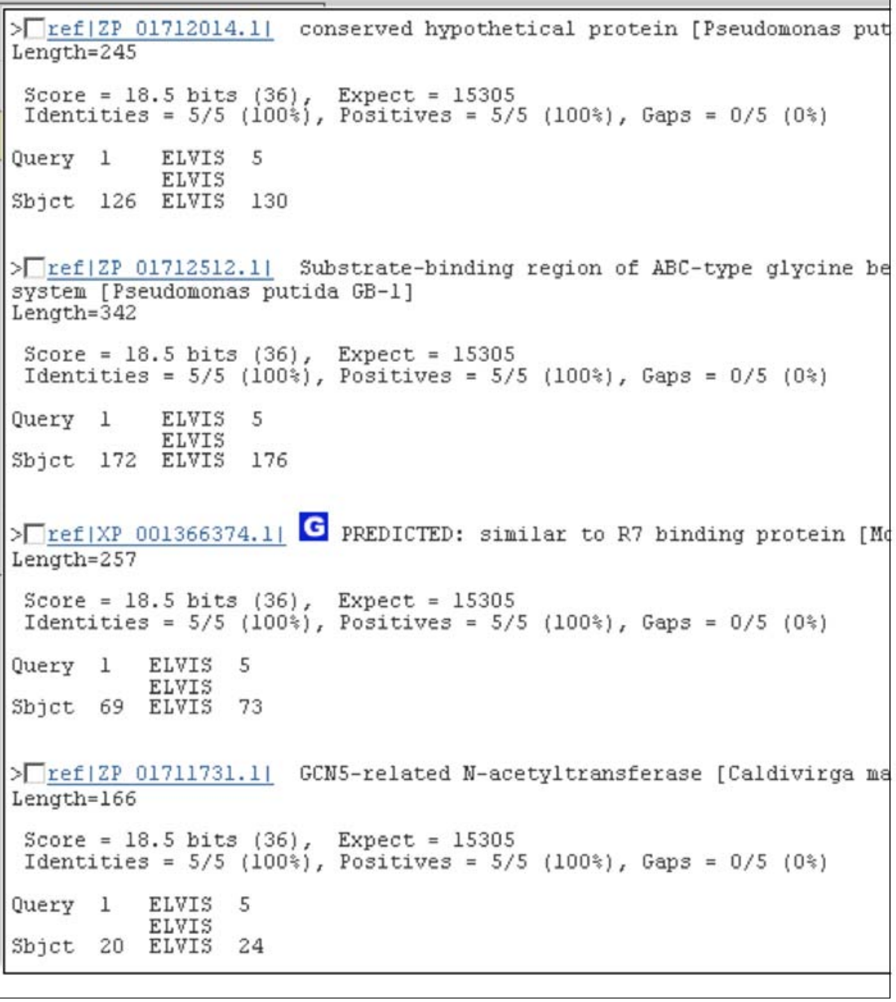
Your search parameters were adjusted to search for a short input sequence.

WAITING

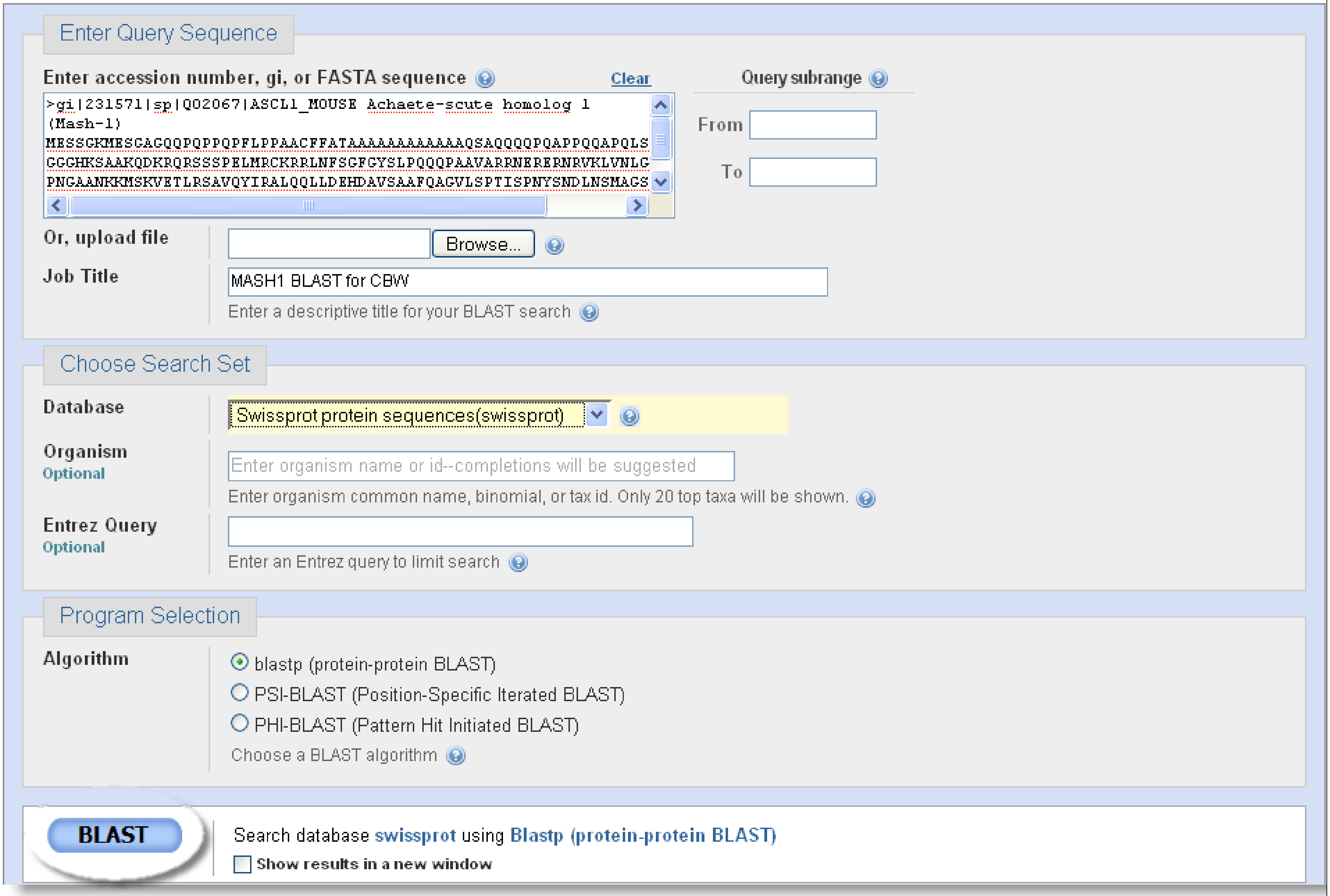
Request ID	1WSB0FX012
Status	Searching
Subr	
Curre	
Time	

This

e-value	20000
Word Size	2
Matrix	PAM30
Comp Stats	Off
Low Comp Filter	Off



```
>☐ref|ZP_01712014.1|  conserved hypothetical protein [Pseudomonas put
Length=245

 Score = 18.5 bits (36),  Expect = 15305
 Identities = 5/5 (100%), Positives = 5/5 (100%), Gaps = 0/5 (0%)

Query  1    ELVIS  5
            ELVIS
Sbjct  126  ELVIS  130


>☐ref|ZP_01712512.1|  Substrate-binding region of ABC-type glycine be
system [Pseudomonas putida GB-1]
Length=342

 Score = 18.5 bits (36),  Expect = 15305
 Identities = 5/5 (100%), Positives = 5/5 (100%), Gaps = 0/5 (0%)

Query  1    ELVIS  5
            ELVIS
Sbjct  172  ELVIS  176


>☐ref|XP_001366374.1| G PREDICTED: similar to R7 binding protein [Mo
Length=257

 Score = 18.5 bits (36),  Expect = 15305
 Identities = 5/5 (100%), Positives = 5/5 (100%), Gaps = 0/5 (0%)

Query  1   ELVIS  5
           ELVIS
Sbjct  69  ELVIS  73


>☐ref|ZP_01711731.1|  GCN5-related N-acetyltransferase [Caldivirga ma
Length=166

 Score = 18.5 bits (36),  Expect = 15305
 Identities = 5/5 (100%), Positives = 5/5 (100%), Gaps = 0/5 (0%)

Query  1   ELVIS  5
           ELVIS
Sbjct  20  ELVIS  24
```



Enter Query Sequence

Enter accession number, gi, or FASTA sequence

```
>gi|231571|sp|Q02067|ASCL1_MOUSE Achaete-scute homolog 1
(Mash-1)
MESSGKMESGAGQQPQPPQPFLPPAACFFATAAAAAAAAAAAAQSAQQQQPQAPPQQAPQLS
GGGHKSAAKQDKRQRSSSPELMRCKRRLNFSGFGYSLPQQQPAAVARRNERERNRVKLVNLG
PNGAANKKMSKVETLRSAVQYIRALQQLLDEHDAVSAAFQAGVLSPTISPNYSNDLNSMAGS
```

Query subrange

From:

To:

Or, upload file

Job Title: MASH1 BLAST for CBW

Enter a descriptive title for your BLAST search

Choose Search Set

Database: Swissprot protein sequences(swissprot)

Organism Optional

Enter organism name or id--completions will be suggested

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

Entrez Query Optional

Enter an Entrez query to limit search

Program Selection

Algorithm

● blastp (protein-protein BLAST)

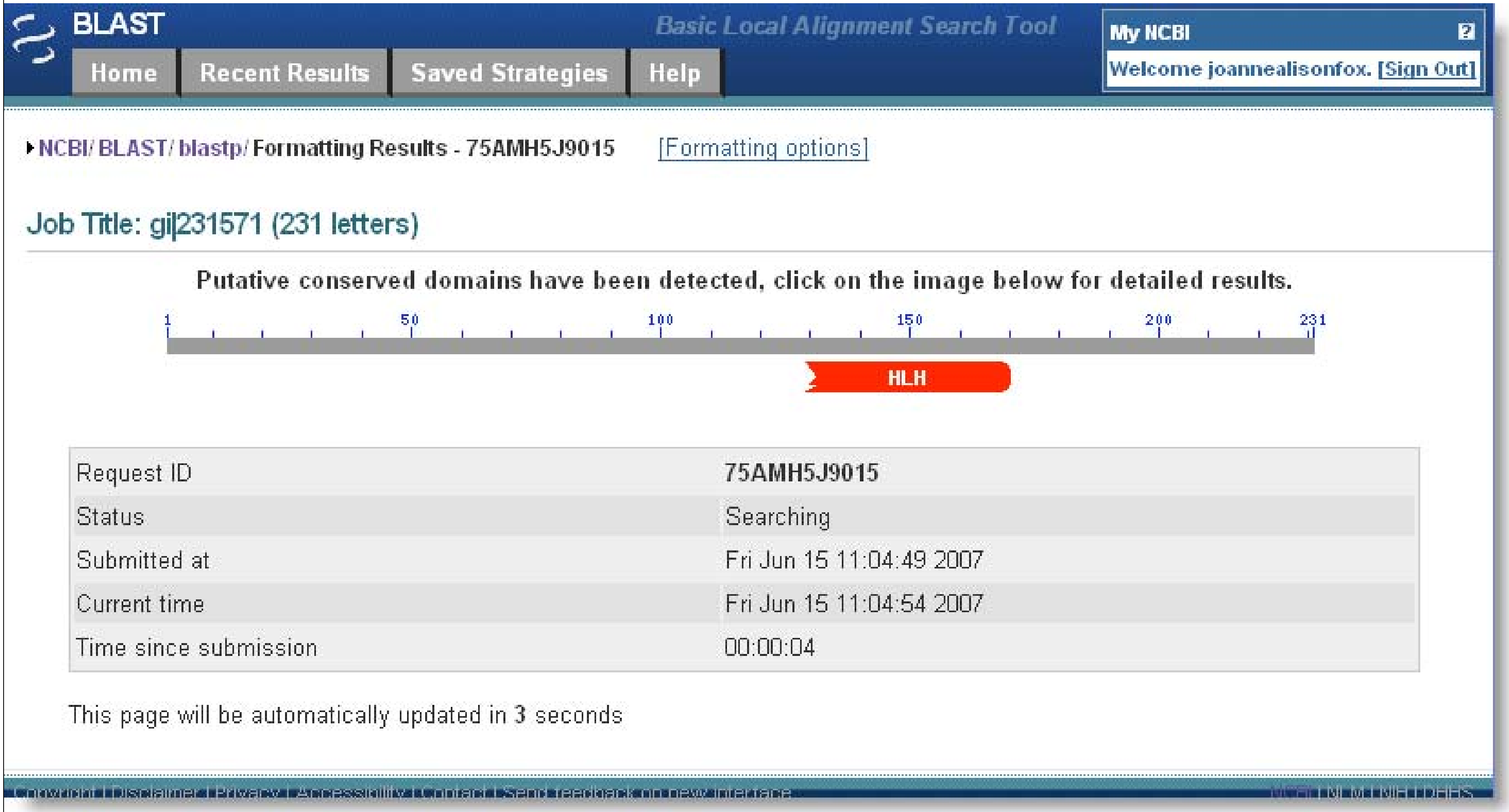
○ PSI-BLAST (Position-Specific Iterated BLAST)

○ PHI-BLAST (Pattern Hit Initiated BLAST)

Choose a BLAST algorithm

Search database swissprot using Blastp (protein-protein BLAST)

☐ Show results in a new window



NCBI/ BLAST/ blastp/ **Formatting Results - 75AMH5J9015** [Formatting options]

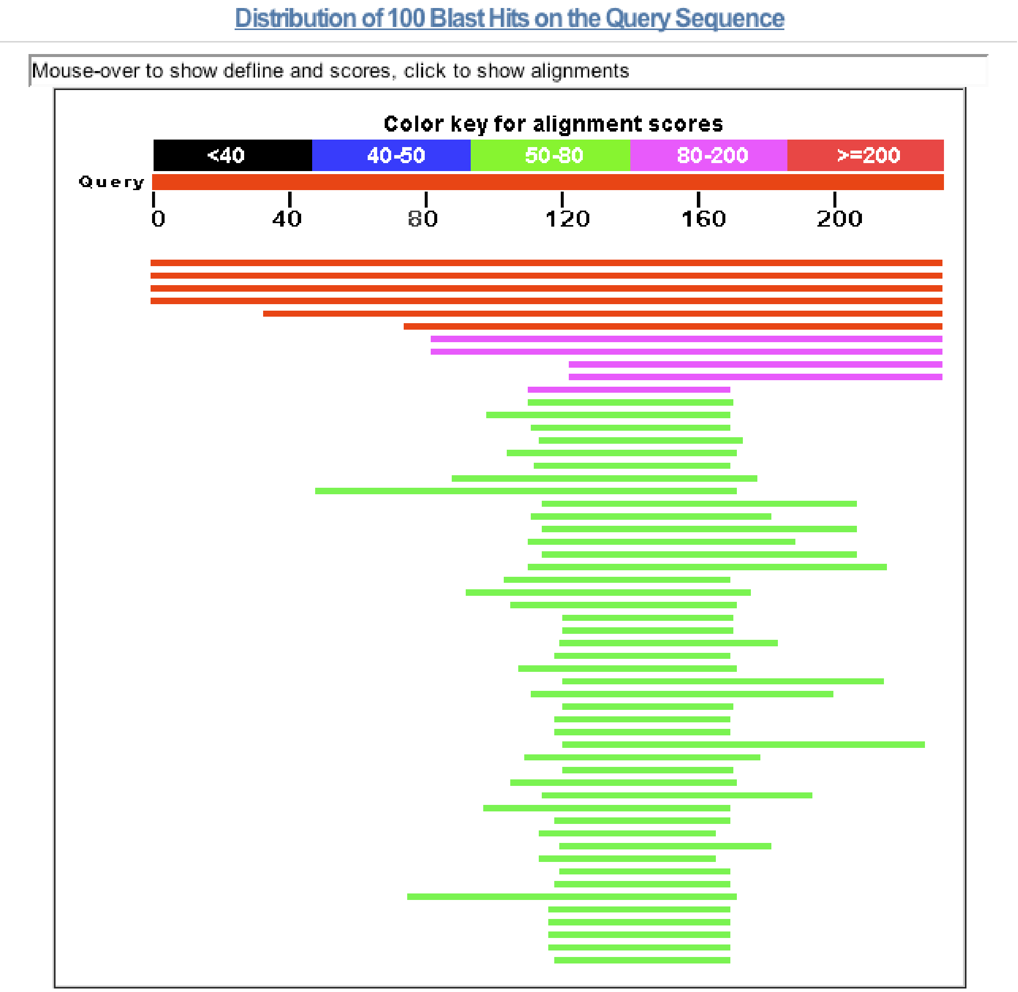
Job Title: gi|231571 (231 letters)

Putative conserved domains have been detected, click on the image below for detailed results.

Request ID	**75AMH5J9015**
Status	Searching
Submitted at	Fri Jun 15 11:04:49 2007
Current time	Fri Jun 15 11:04:54 2007
Time since submission	00:00:04

This page will be automatically updated in **3** seconds

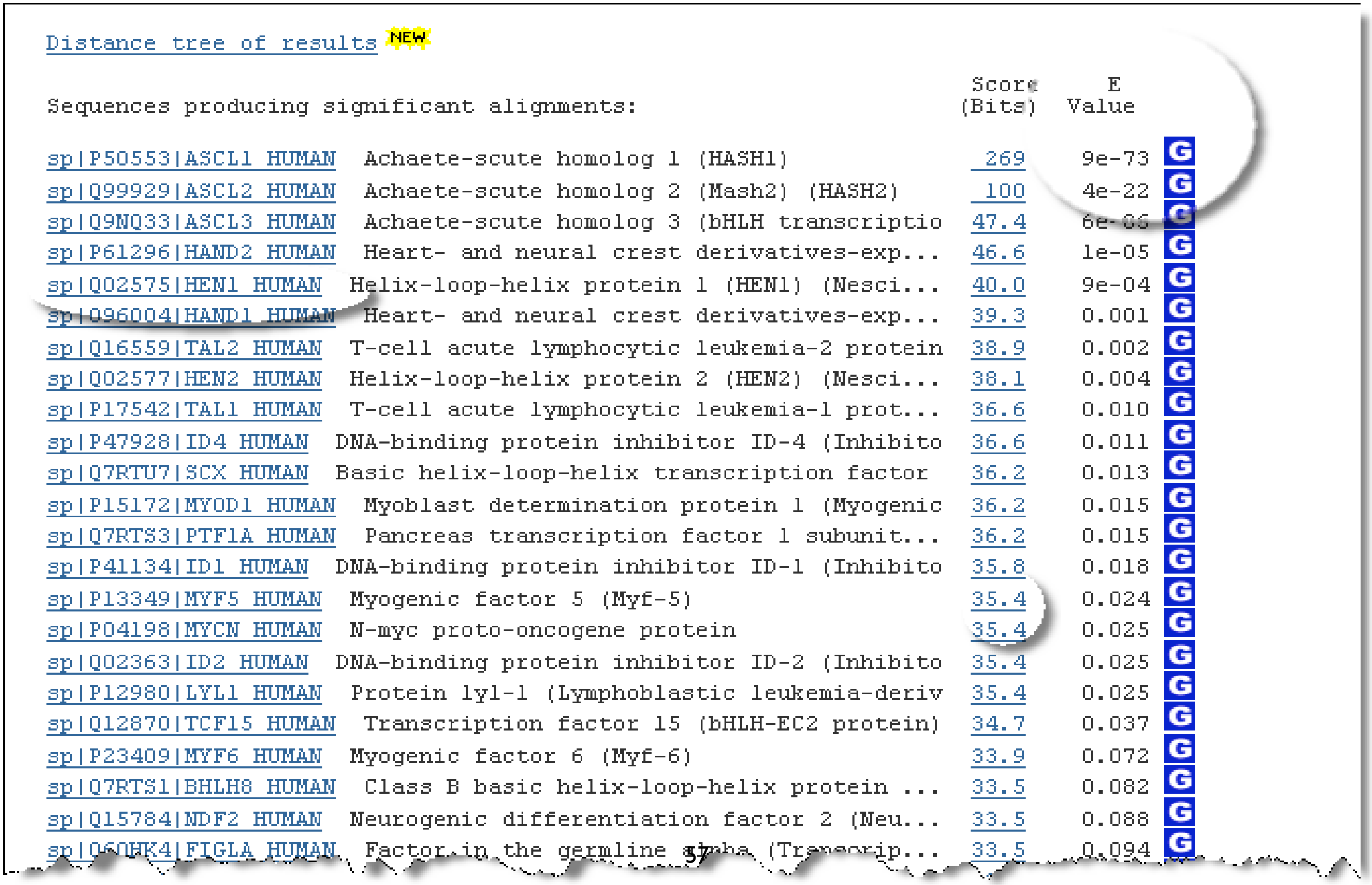
A graphical view



Distribution of 100 Blast Hits on the Query Sequence

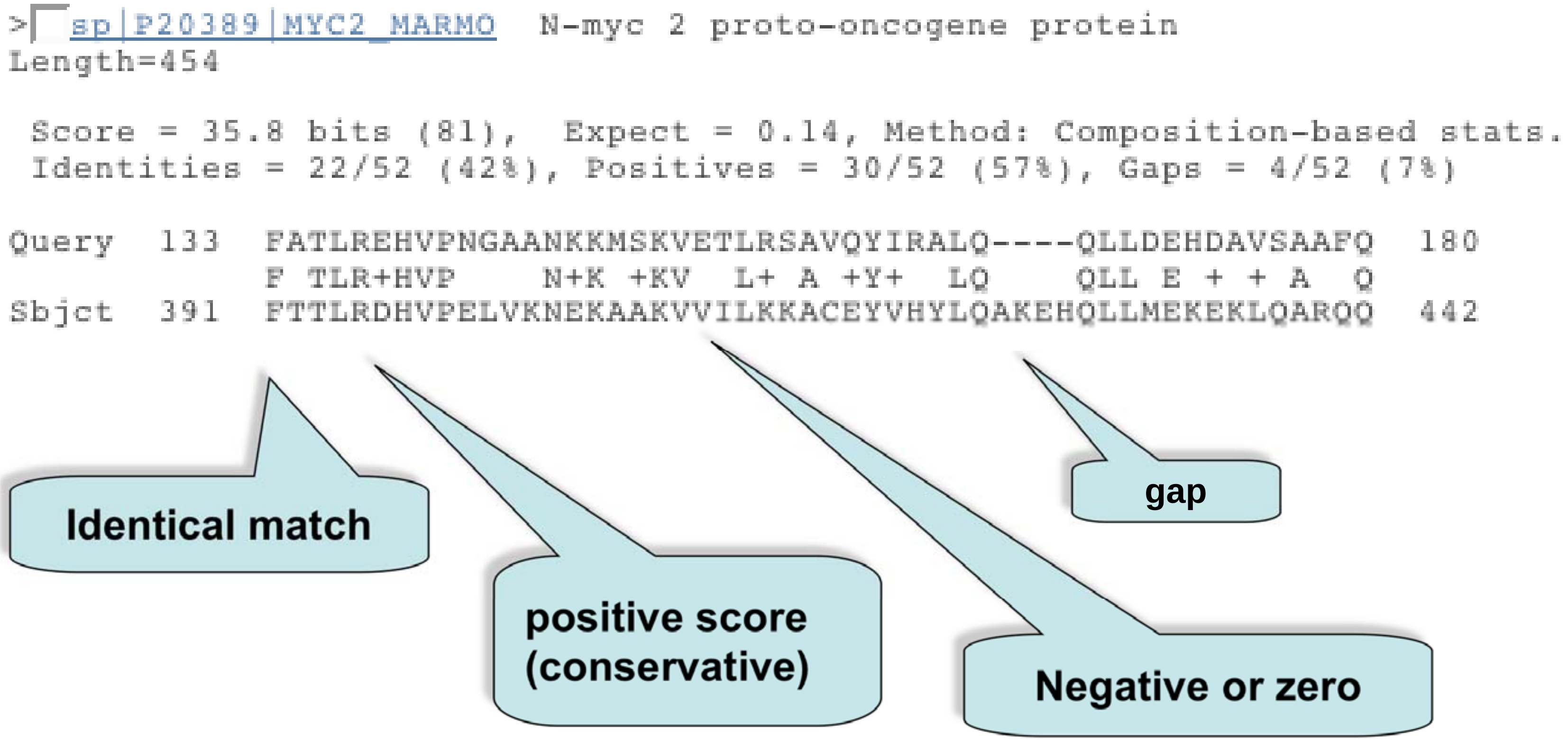
Mouse-over to show defline and scores, click to show alignments

The BLAST hit list



```
Distance tree of results NEW

                                                                  Score    E
Sequences producing significant alignments:                      (Bits)  Value

sp|P50553|ASCL1_HUMAN  Achaete-scute homolog 1 (HASH1)              269   9e-73  G
sp|Q99929|ASCL2_HUMAN  Achaete-scute homolog 2 (Mash2) (HASH2)      100   4e-22  G
sp|Q9NQ33|ASCL3_HUMAN  Achaete-scute homolog 3 (bHLH transcriptio  47.4   6e-06  G
sp|P61296|HAND2_HUMAN  Heart- and neural crest derivatives-exp...  46.6   1e-05  G
sp|Q02575|HEN1_HUMAN  Helix-loop-helix protein 1 (HEN1) (Nesci...  40.0   9e-04  G
sp|O96004|HAND1_HUMAN  Heart- and neural crest derivatives-exp...  39.3   0.001  G
sp|Q16559|TAL2_HUMAN  T-cell acute lymphocytic leukemia-2 protein  38.9   0.002  G
sp|Q02577|HEN2_HUMAN  Helix-loop-helix protein 2 (HEN2) (Nesci...  38.1   0.004  G
sp|P17542|TAL1_HUMAN  T-cell acute lymphocytic leukemia-1 prot...  36.6   0.010  G
sp|P47928|ID4_HUMAN  DNA-binding protein inhibitor ID-4 (Inhibito  36.6   0.011  G
sp|Q7RTU7|SCX_HUMAN  Basic helix-loop-helix transcription factor   36.2   0.013  G
sp|P15172|MYOD1_HUMAN  Myoblast determination protein 1 (Myogenic  36.2   0.015  G
sp|Q7RTS3|PTF1A_HUMAN  Pancreas transcription factor 1 subunit...  36.2   0.015  G
sp|P41134|ID1_HUMAN  DNA-binding protein inhibitor ID-1 (Inhibito  35.8   0.018  G
sp|P13349|MYF5_HUMAN  Myogenic factor 5 (Myf-5)                    35.4   0.024  G
sp|P04198|MYCN_HUMAN  N-myc proto-oncogene protein                 35.4   0.025  G
sp|Q02363|ID2_HUMAN  DNA-binding protein inhibitor ID-2 (Inhibito  35.4   0.025  G
sp|P12980|LYL1_HUMAN  Protein lyl-1 (Lymphoblastic leukemia-deriv  35.4   0.025  G
sp|Q12870|TCF15_HUMAN  Transcription factor 15 (bHLH-EC2 protein)  34.7   0.037  G
sp|P23409|MYF6_HUMAN  Myogenic factor 6 (Myf-6)                    33.9   0.072  G
sp|Q7RTS1|BHLH8_HUMAN  Class B basic helix-loop-helix protein ...  33.5   0.082  G
sp|Q15784|NDF2_HUMAN  Neurogenic differentiation factor 2 (Neu...  33.5   0.088  G
sp|Q6QHK4|FIGLA_HUMAN  Factor in the germline alpha (Transcrip...  33.5   0.094  G
```

BLAST Alignments



```
>sp|P20389|MYC2_MARMO  N-myc 2 proto-oncogene protein
Length=454

 Score = 35.8 bits (81),  Expect = 0.14, Method: Composition-based stats.
 Identities = 22/52 (42%), Positives = 30/52 (57%), Gaps = 4/52 (7%)

Query  133  FATLREHVPNGAANKKMSKVETLRSAVQYIRALQ----QLLDEHDAVSAAFQ  180
            F TLR+HVP    N+K +KV  L+ A +Y+  LQ    QLL E + + A  Q
Sbjct  391  FTTLRDHVPELVKNEKAAKVVILKKACEYVHYLQAKEHQLLMEKEKLQARQQ  442
```

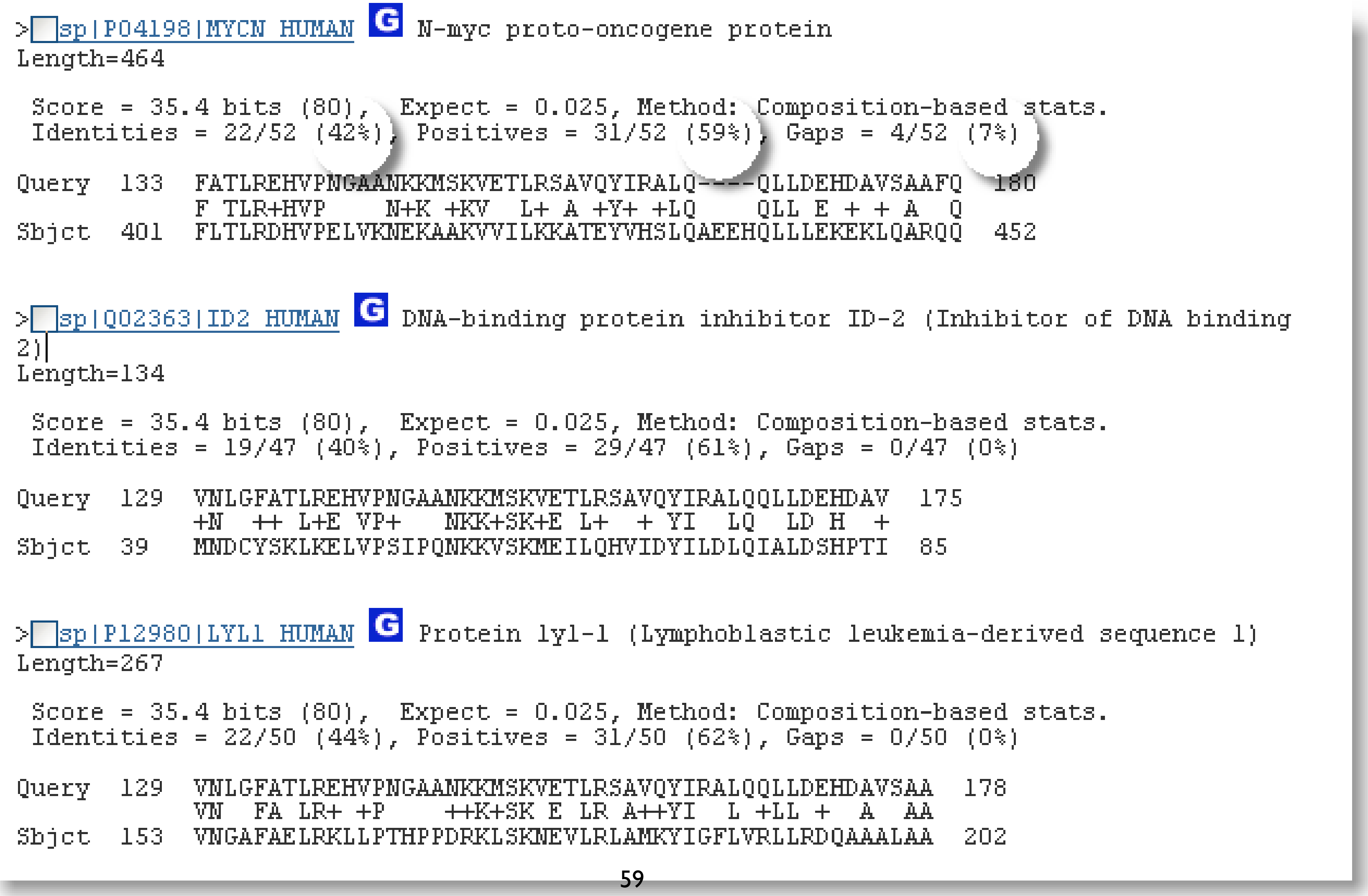
Identical match

positive score
(conservative)

Negative or zero

gap

BLAST Alignments



```
>☐sp|P04198|MYCN_HUMAN G N-myc proto-oncogene protein
Length=464

 Score = 35.4 bits (80),  Expect = 0.025, Method: Composition-based stats.
 Identities = 22/52 (42%), Positives = 31/52 (59%), Gaps = 4/52 (7%)

Query  133  FATLREHVPNGAANKKMSKVETLRSAVQYIRALQ----QLLDEHDAVSAAFQ  180
            F TLR+HVP    N+K +KV  L+ A +Y+ +LQ    QLL E + + A  Q
Sbjct  401  FLTLRDHVPELVKNEKAAKVVILKKATEYVHSLQAEEHQLLLEKEKLQARQQ  452


>☐sp|Q02363|ID2_HUMAN G DNA-binding protein inhibitor ID-2 (Inhibitor of DNA binding
2)
Length=134

 Score = 35.4 bits (80),  Expect = 0.025, Method: Composition-based stats.
 Identities = 19/47 (40%), Positives = 29/47 (61%), Gaps = 0/47 (0%)

Query  129  VNLGFATLREHVPNGAANKKMSKVETLRSAVQYIRALQQLLDEHDAV  175
            +N  ++ L+E VP+   NKK+SK+E L+  + YI  LQ  LD H  +
Sbjct  39   MNDCYSKLKELVPSIPQNKKVSKMEILQHVIDYILDLQIALDSHPTI  85


>☐sp|P12980|LYL1_HUMAN G Protein lyl-1 (Lymphoblastic leukemia-derived sequence 1)
Length=267

 Score = 35.4 bits (80),  Expect = 0.025, Method: Composition-based stats.
 Identities = 22/50 (44%), Positives = 31/50 (62%), Gaps = 0/50 (0%)

Query  129  VNLGFATLREHVPNGAANKKMSKVETLRSAVQYIRALQQLLDEHDAVSAA  178
            VN  FA LR+ +P    ++K+SK E LR A++YI  L +LL +  A  AA
Sbjct  153  VNGAFAELRKLLPTHPPDRKLSKNEVLRLAMKYIGFLVRLLRDQAAALAA  202
```

- Similarity

  The extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. In BLAST similarity refers to a positive matrix score.

- Identity

  The extent to which two (nucleotide or amino acid) sequences are invariant.

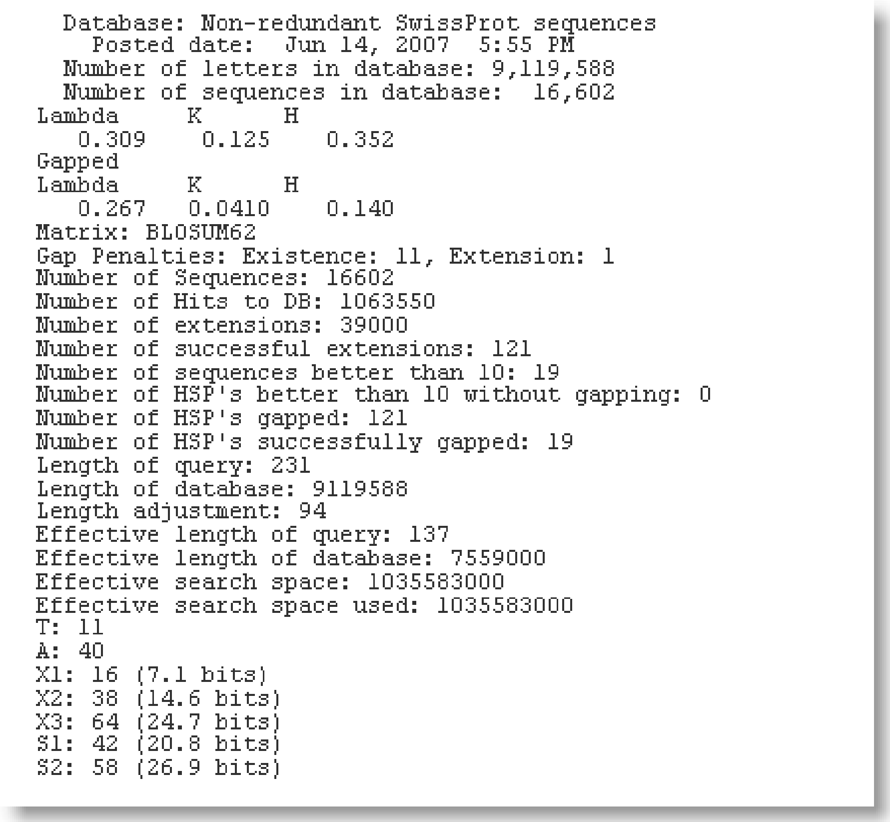
- Homology

  Similarity attributed to descent from a common ancestor.

  It is your responsibility as an informed bioinformatician to use these terms correctly: A sequence is either homologous or not. Don't use % with this term!

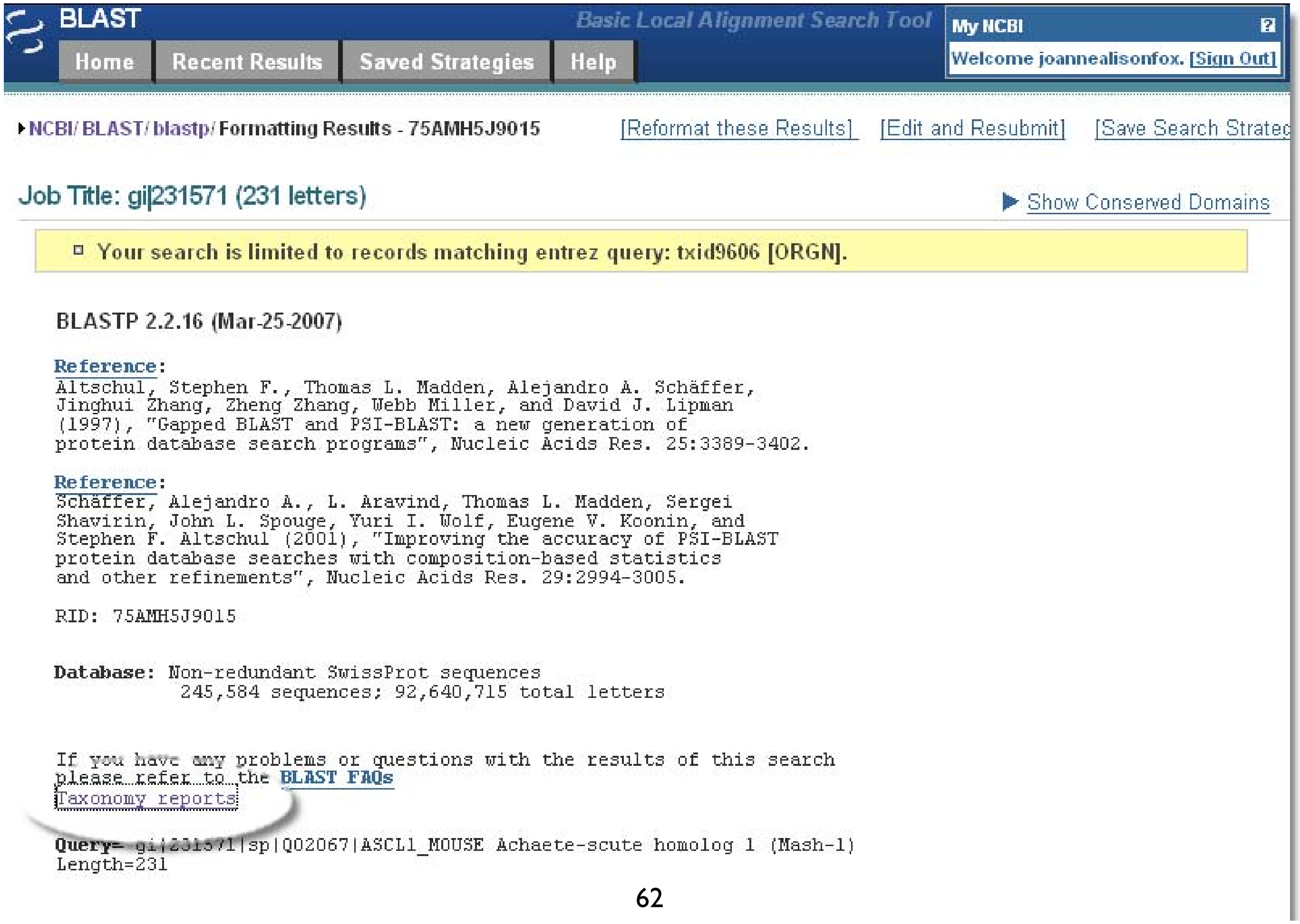
BLAST statistics to record in your bioinformatics labbook

It can be helpful to record the statistics that are found at bottom of your BLAST results



```
  Database: Non-redundant SwissProt sequences
    Posted date:  Jun 14, 2007  5:55 PM
  Number of letters in database: 9,119,588
  Number of sequences in database:  16,602
Lambda     K      H
   0.309    0.125    0.352
Gapped
Lambda     K      H
   0.267   0.0410    0.140
Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 16602
Number of Hits to DB: 1063550
Number of extensions: 39000
Number of successful extensions: 121
Number of sequences better than 10: 19
Number of HSP's better than 10 without gapping: 0
Number of HSP's gapped: 121
Number of HSP's successfully gapped: 19
Length of query: 231
Length of database: 9119588
Length adjustment: 94
Effective length of query: 137
Effective length of database: 7559000
Effective search space: 1035583000
Effective search space used: 1035583000
T: 11
A: 40
X1: 16 (7.1 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 42 (20.8 bits)
S2: 58 (26.9 bits)
```

Sorting BLAST by Taxonomy



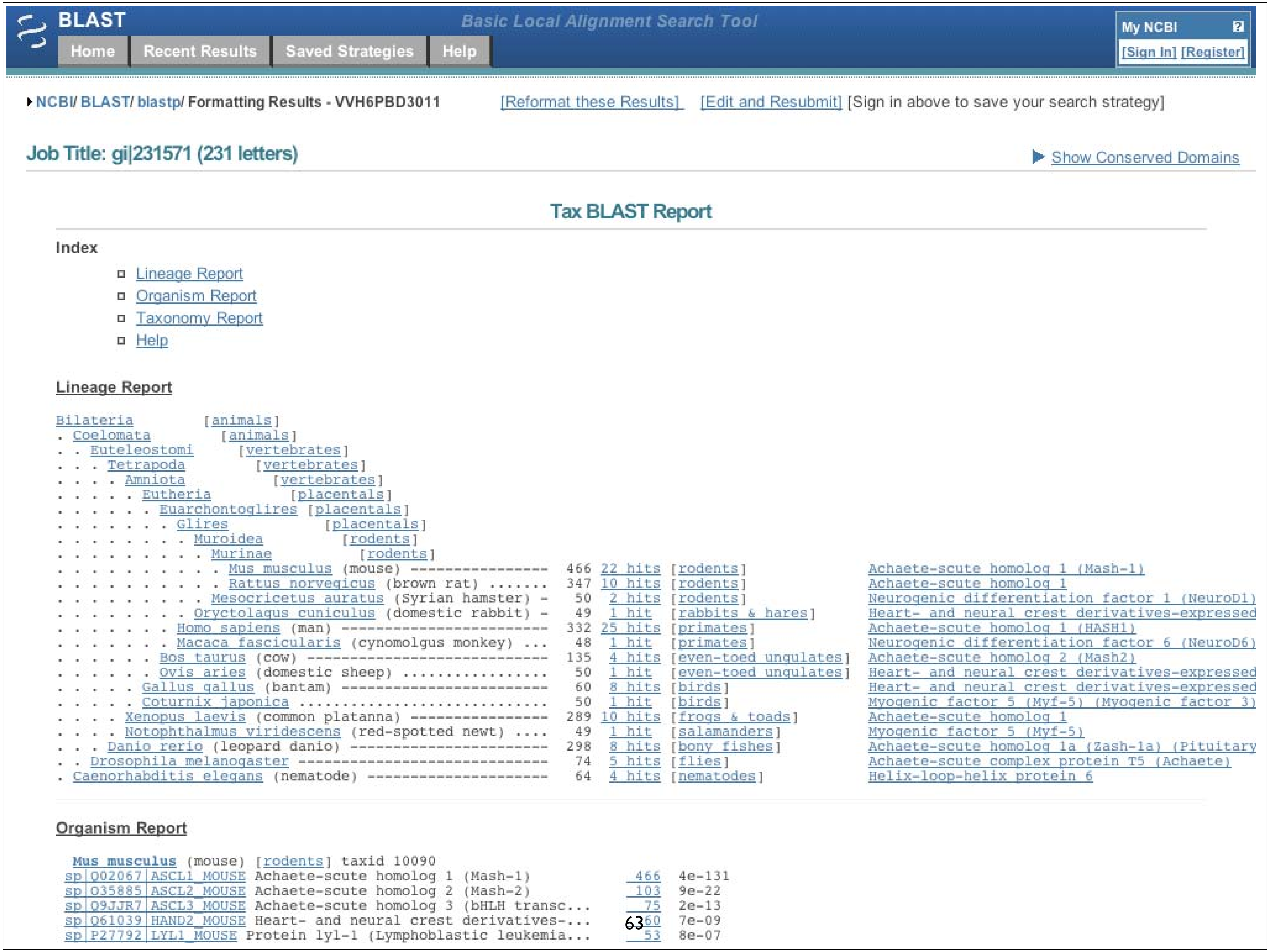
NCBI/ BLAST/ blastp/ Formatting Results - 75AMH5J9015

Job Title: gi|231571 (231 letters)

Your search is limited to records matching entrez query: txid9606 [ORGN].

BLASTP 2.2.16 (Mar-25-2007)

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

Reference:
Schäffer, Alejandro A., L. Aravind, Thomas L. Madden, Sergei
Shavirin, John L. Spouge, Yuri I. Wolf, Eugene V. Koonin, and
Stephen F. Altschul (2001), "Improving the accuracy of PSI-BLAST
protein database searches with composition-based statistics
and other refinements", Nucleic Acids Res. 29:2994-3005.

RID: 75AMH5J9015


Database: Non-redundant SwissProt sequences
           245,584 sequences; 92,640,715 total letters


If you have any problems or questions with the results of this search
please refer to the BLAST FAQs
Taxonomy reports

Query= gi|231571|sp|Q02067|ASCL1_MOUSE Achaete-scute homolog 1 (Mash-1)
Length=231
```



Job Title: gi|231571 (231 letters)

Tax BLAST Report

Index

- Lineage Report
- Organism Report
- Taxonomy Report
- Help

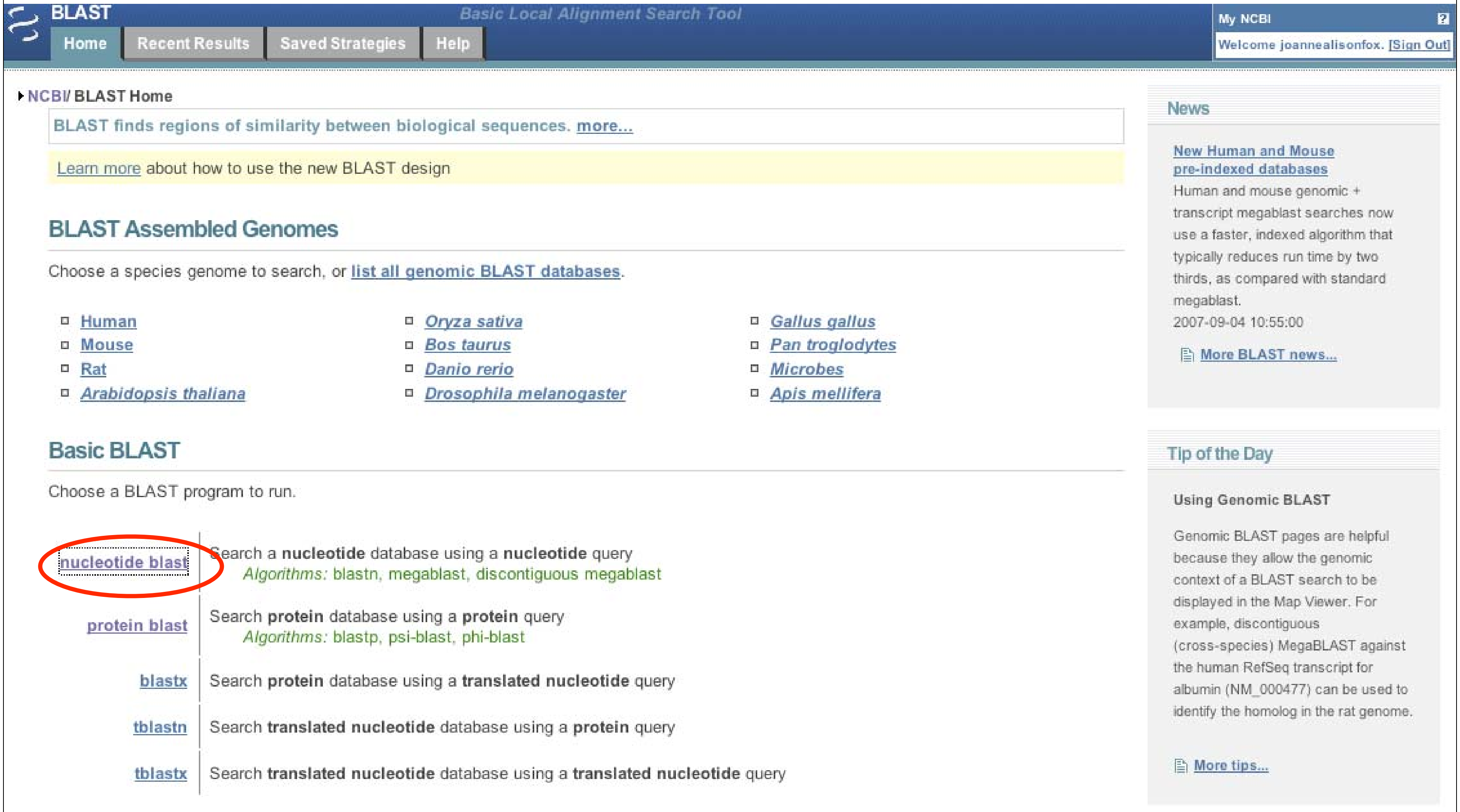
Lineage Report

```
Bilateria           [animals]
. Coelomata           [animals]
. . Euteleostomi        [vertebrates]
. . . Tetrapoda           [vertebrates]
. . . . Amniota             [vertebrates]
. . . . . Eutheria            [placentals]
. . . . . . Euarchontoglires [placentals]
. . . . . . . Glires             [placentals]
. . . . . . . . Muroidea            [rodents]
. . . . . . . . . Murinae             [rodents]
. . . . . . . . . . Mus musculus (mouse) -----------------  466 22 hits [rodents]              Achaete-scute homolog 1 (Mash-1)
. . . . . . . . . . Rattus norvegicus (brown rat) .......  347 10 hits [rodents]              Achaete-scute homolog 1
. . . . . . . . . Mesocricetus auratus (Syrian hamster) -   50  2 hits [rodents]              Neurogenic differentiation factor 1 (NeuroD1)
. . . . . . . . Oryctolagus cuniculus (domestic rabbit) -   49  1 hit  [rabbits & hares]      Heart- and neural crest derivatives-expressed
. . . . . . . Homo sapiens (man) ------------------------  332 25 hits [primates]             Achaete-scute homolog 1 (HASH1)
. . . . . . . Macaca fascicularis (cynomolgus monkey) ...   48  1 hit  [primates]             Neurogenic differentiation factor 6 (NeuroD6)
. . . . . . Bos taurus (cow) ----------------------------  135  4 hits [even-toed ungulates]  Achaete-scute homolog 2 (Mash2)
. . . . . . Ovis aries (domestic sheep) .................   50  1 hit  [even-toed ungulates]  Heart- and neural crest derivatives-expressed
. . . . . Gallus gallus (bantam) ------------------------   60  8 hits [birds]                Heart- and neural crest derivatives-expressed
. . . . . Coturnix japonica .............................   50  1 hit  [birds]                Myogenic factor 5 (Myf-5) (Myogenic factor 3)
. . . . Xenopus laevis (common platanna) ----------------  289 10 hits [frogs & toads]        Achaete-scute homolog 1
. . . . Notophthalmus viridescens (red-spotted newt) ....   49  1 hit  [salamanders]          Myogenic factor 5 (Myf-5)
. . . Danio rerio (leopard danio) -----------------------  298  8 hits [bony fishes]          Achaete-scute homolog 1a (Zash-1a) (Pituitary
. . Drosophila melanogaster -----------------------------   74  5 hits [flies]                Achaete-scute complex protein T5 (Achaete)
. Caenorhabditis elegans (nematode) ---------------------   64  4 hits [nematodes]            Helix-loop-helix protein 6
```

Organism Report

```
 Mus musculus (mouse) [rodents] taxid 10090
sp|Q02067|ASCL1_MOUSE Achaete-scute homolog 1 (Mash-1)            466  4e-131
sp|O35885|ASCL2_MOUSE Achaete-scute homolog 2 (Mash-2)            103  9e-22
sp|Q9JJR7|ASCL3_MOUSE Achaete-scute homolog 3 (bHLH transc...      75  2e-13
sp|Q61039|HAND2_MOUSE Heart- and neural crest derivatives-...      60  7e-09
sp|P27792|LYL1_MOUSE Protein lyl-1 (Lymphoblastic leukemia...      53  8e-07
```

Nucleotide BLAST



NCBI/ BLAST Home

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

Learn more about how to use the new BLAST design

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

News

New Human and Mouse pre-indexed databases

Human and mouse genomic + transcript megablast searches now use a faster, indexed algorithm that typically reduces run time by two thirds, as compared with standard megablast.

2007-09-04 10:55:00

More BLAST news...

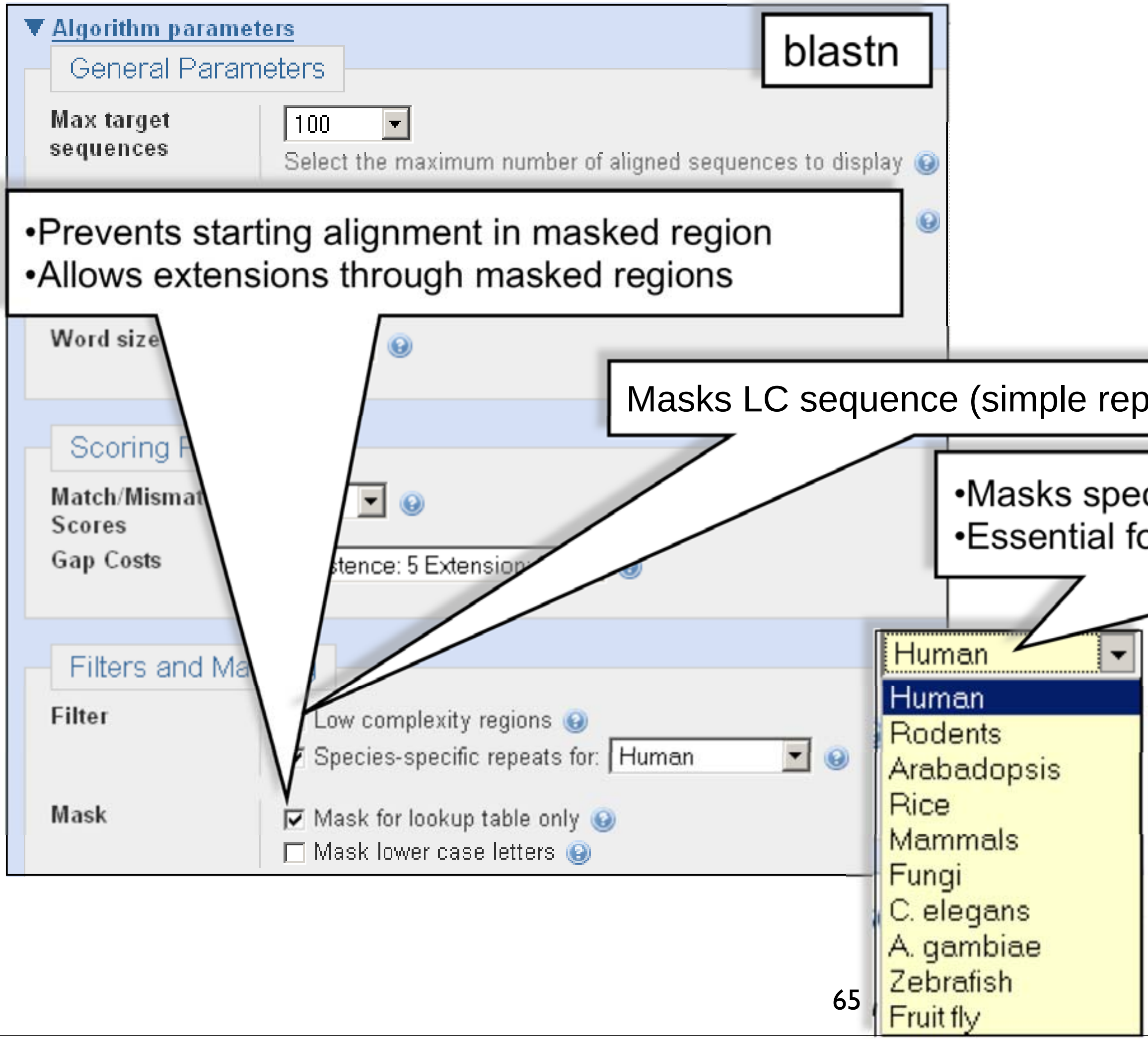
Tip of the Day

Using Genomic BLAST

Genomic BLAST pages are helpful because they allow the genomic context of a BLAST search to be displayed in the Map Viewer. For example, discontiguous (cross-species) MegaBLAST against the human RefSeq transcript for albumin (NM_000477) can be used to identify the homolog in the rat genome.

More tips...

Algorithm parameters: Nucleotide



blastn

Max target sequences: 100

Select the maximum number of aligned sequences to display

Word size

Match/Mismat Scores

Gap Costs: tence: 5 Extension

Filter: Low complexity regions

Species-specific repeats for: Human

Mask: ☑ Mask for lookup table only

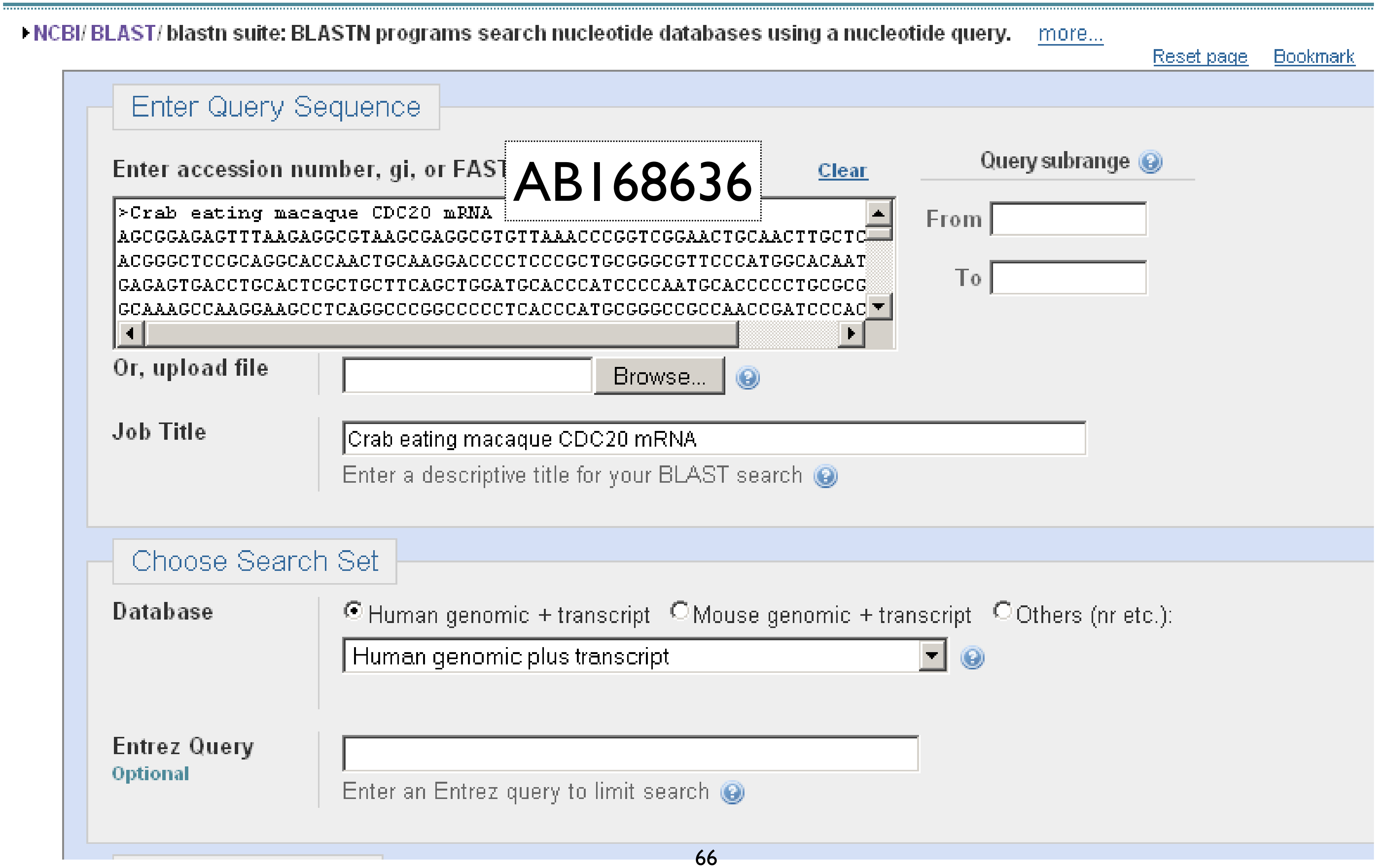
☐ Mask lower case letters

- Prevents starting alignment in masked region
- Allows extensions through masked regions

Masks LC sequence (simple repeats)

- Masks species-specific interspersed repeats
- Essential for genomic query sequences

nt BLAST: New Output



NCBI/ BLAST/ blastn suite: BLASTN programs search nucleotide databases using a nucleotide query. more...

Reset page Bookmark

Enter Query Sequence

Enter accession number, gi, or FAST

AB168636

Clear

```
>Crab eating macaque CDC20 mRNA
AGCGGAGAGTTTAAGAGGCGTAAGCGAGGCGTGTTAAACCCGGTCGGAACTGCAACTTGCTC
ACGGGCTCCGCAGGCACCAACTGCAAGGACCCCTCCCGCTGCGGGCGTTCCCATGGCACAAT
GAGAGTGACCTGCACTCGCTGCTTCAGCTGGATGCACCCATCCCCAATGCACCCCCTGCGCG
GCAAAGCCAAGGAAGCCTCAGGCCCGGCCCCCTCACCCATGCGGGCCGCCAACCGATCCCAC
```

Query subrange

From

To

Or, upload file

Browse...

Job Title: Crab eating macaque CDC20 mRNA

Enter a descriptive title for your BLAST search

Choose Search Set

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

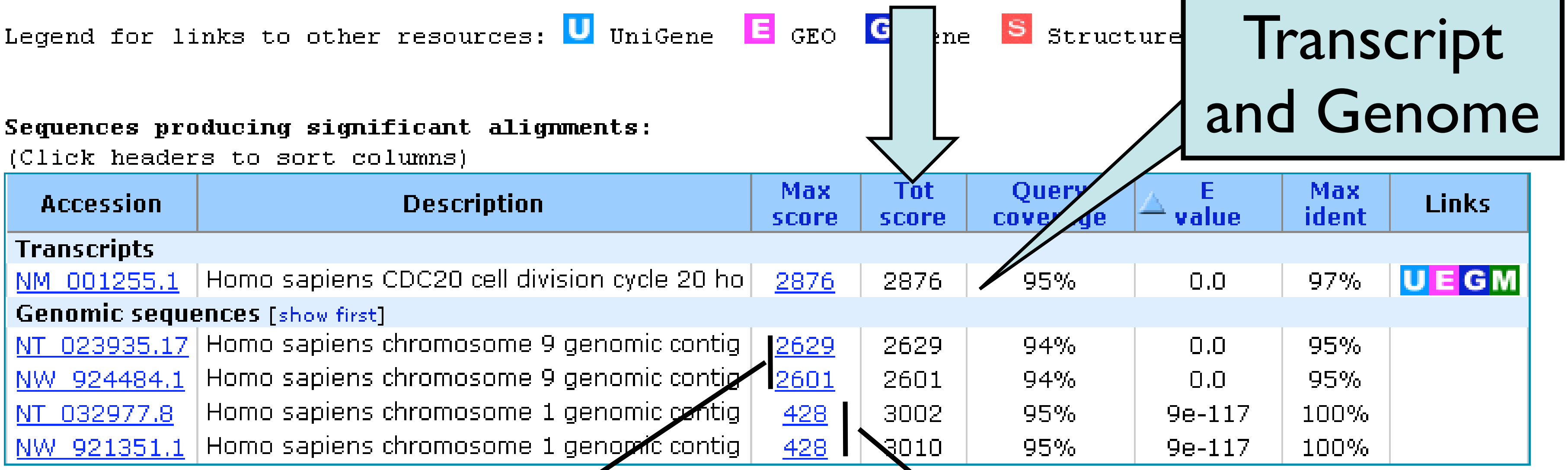
Human genomic plus transcript

Entrez Query

Optional

Enter an Entrez query to limit search

Sortable Results



Legend for links to other resources: U UniGene E GEO G Gene S Structure

Sequences producing significant alignments:
(Click headers to sort columns)

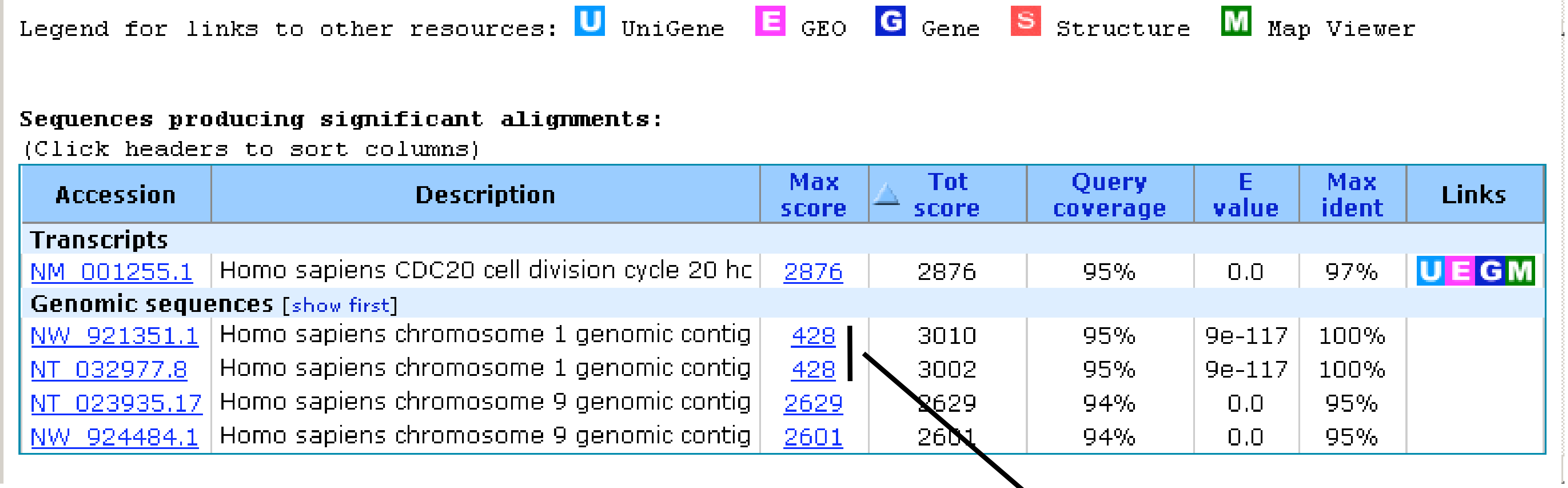
Accession	Description	Max score	Tot score	Query coverage	▲ E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell division cycle 20 ho	2876	2876	95%	0.0	97%	U E G M
Genomic sequences [show first]							
NT_023935.17	Homo sapiens chromosome 9 genomic contig	2629	2629	94%	0.0	95%	
NW_924484.1	Homo sapiens chromosome 9 genomic contig	2601	2601	94%	0.0	95%	
NT_032977.8	Homo sapiens chromosome 1 genomic contig	428	3002	95%	9e-117	100%	
NW_921351.1	Homo sapiens chromosome 1 genomic contig	428	3010	95%	9e-117	100%	

Separate Sections for Transcript and Genome

Pseudogene on Chromosome 9

Functional Gene on Chromosome 1

Total Score: All Segments



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	Tot score	Query coverage	E value	Max ident	Links
Transcripts							
NM_001255.1	Homo sapiens CDC20 cell division cycle 20 hc	2876	2876	95%	0.0	97%	U E G M
Genomic sequences [show first]							
NW_921351.1	Homo sapiens chromosome 1 genomic contig	428	3010	95%	9e-117	100%	
NT_032977.8	Homo sapiens chromosome 1 genomic contig	428	3002	95%	9e-117	100%	
NT_023935.17	Homo sapiens chromosome 9 genomic contig	2629	2629	94%	0.0	95%	
NW_924484.1	Homo sapiens chromosome 9 genomic contig	2601	2601	94%	0.0	95%	

Functional Gene Now First

Sorting in Exon Order

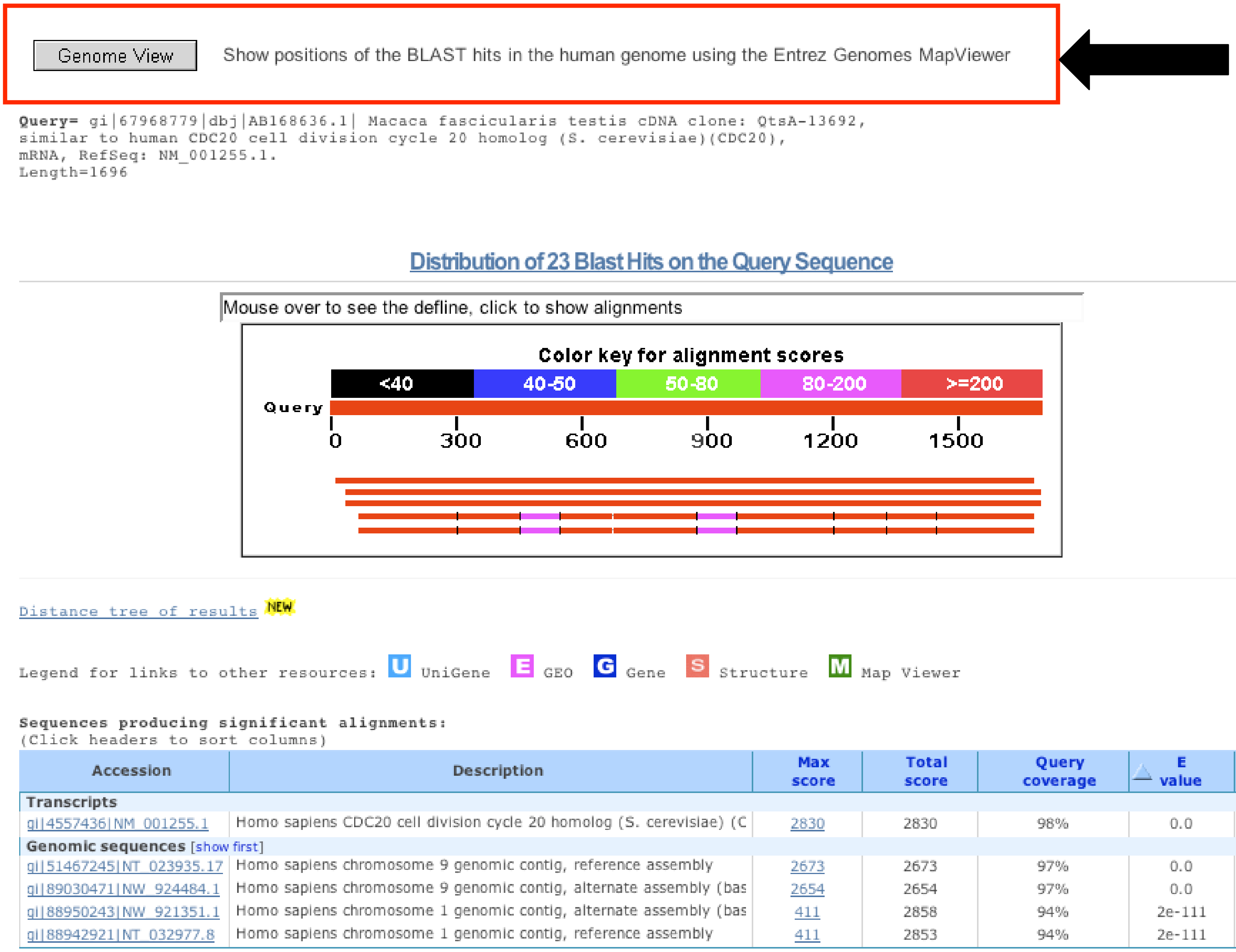
```
> ☐ ref|NT_032977.8|Hs1_33153 D Homo sapiens chromosome 1 genomic contig, reference assembly
Length=73835825

 Features i
  cell div

 Score =  41
 Identities
 Strand=Plus

Query  965

Sbjct  1379

Query  1025

Sbjct  1379
```

Defau
Lon

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

 Features flanking this part of subject sequence:
   6169 bp at 5' side: myeloproliferative leukemia virus oncogene
   223 bp at 3' side: cell division cycle 20


 Score = 89.7 bits (45),  Expect = 1e-14
 Identities = 51/53 (96%), Gaps = 0/53 (0%)
 Strand=Plus/Plus

Query  1         AGCGGAGAGTTTAAGAGGCGTAAGCGAGGCGTGTTAAACCCGGTCGGAACTGC  53
                 ||||||||||||||||||||||||| |||||||||||| ||||||||||||||
Sbjct  13796530  AGCGGAGAGTTTAAGAGGCGTAAGCCAGGCGTGTTAAAGCCGGTCGGAACTGC  13796582


 Features in this part of subject sequence:
   cell division cycle 20


 Score =  412 bits (208),  Expect = 5e-112
 Identities = 226/232 (97%), Gaps = 0/232 (0%)
 Strand=Plus/Plus

Query  73        GGGCTCCGCAGGCACCAACTGCAAGGACCCCTCCCGCTGCGGGCGTTCCCATGGCACAAT  132
                 |||||||| |||||||||||||||||||||||||| ||||||||| |||||||||||| |
Sbjct  13796755  GGGCTCCGTAGGCACCAACTGCAAGGACCCCTCCCCCTGCGGGCGCTCCCATGGCACAGT  13796814

Query  133       TCGCGTTCGAGAGTGACCTGCACTCGCTGCTTCAGCTGGATGCACCCATCCCCAATGCAC  192
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  13796815  TCGCGTTCGAGAGTGACCTGCACTCGCTGCTTCAGCTGGATGCACCCATCCCCAATGCAC  13796874
```

Query start position

Exon order



Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

```
Query= gi|67968779|dbj|AB168636.1| Macaca fascicularis testis cDNA clone: QtsA-13692,
similar to human CDC20 cell division cycle 20 homolog (S. cerevisiae)(CDC20),
mRNA, RefSeq: NM_001255.1.
Length=1696
```

Distribution of 23 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments

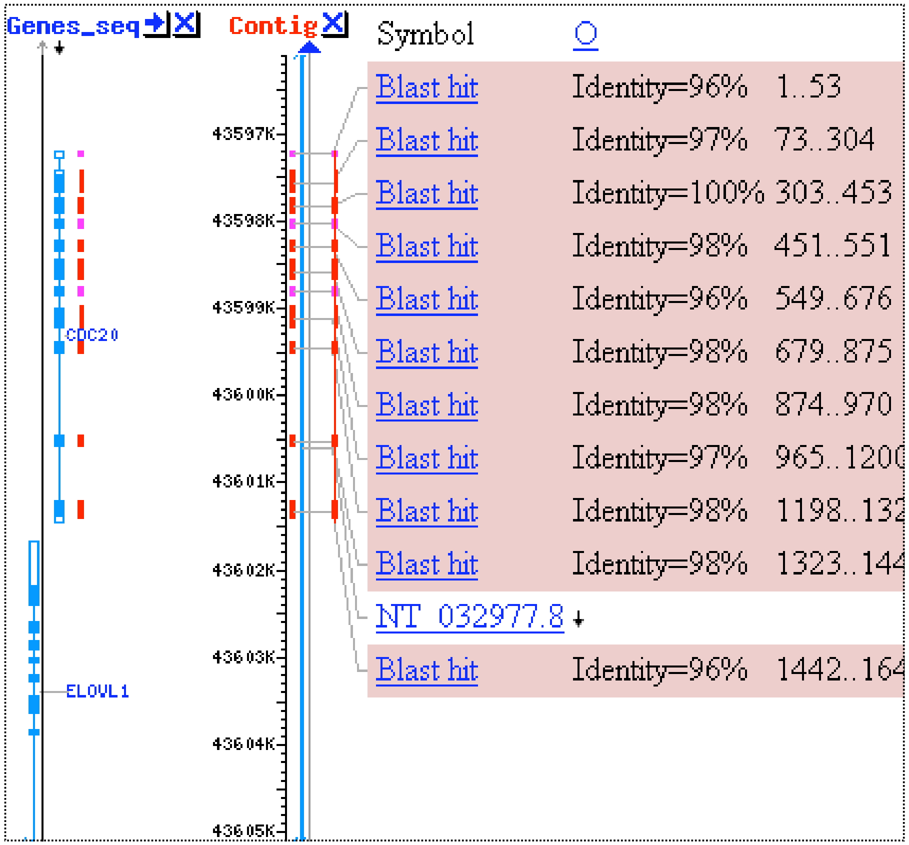
Distance tree of results NEW

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)

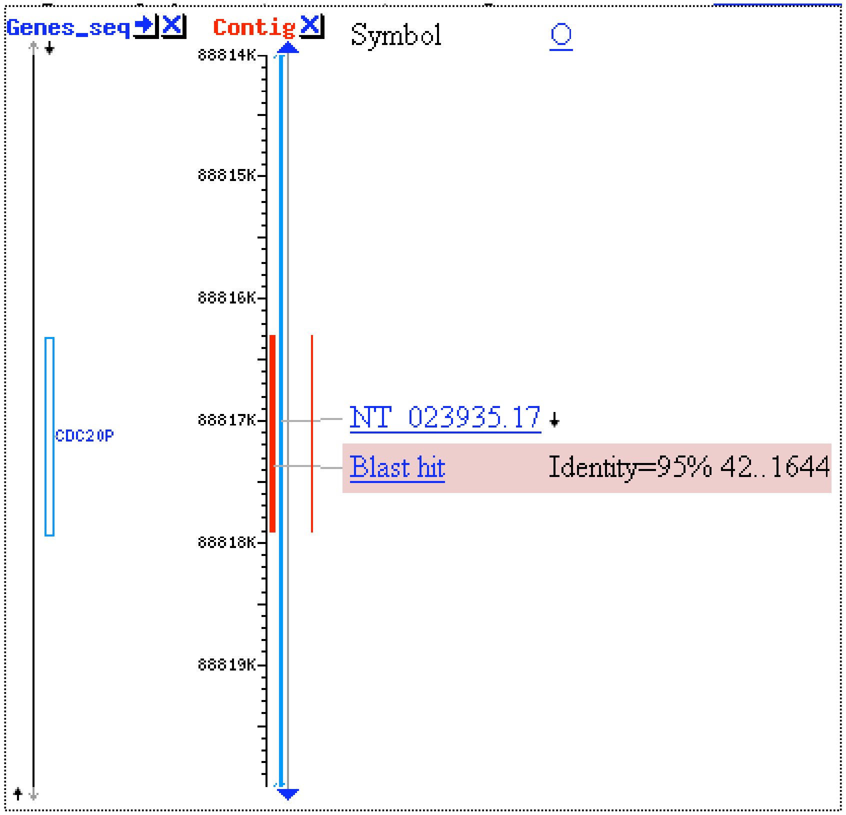
Accession	Description	Max score	Total score	Query coverage	E value	Max ident
Transcripts						
gi|4557436|NM_001255.1	Homo sapiens CDC20 cell division cycle 20 homolog (S. cerevisiae) (C	2830	2830	98%	0.0	97%
Genomic sequences [show first]						
gi|51467245|NT_023935.17	Homo sapiens chromosome 9 genomic contig, reference assembly	2673	2673	97%	0.0	95%
gi|89030471|NW_924484.1	Homo sapiens chromosome 9 genomic contig, alternate assembly (bas	2654	2654	97%	0.0	95%
gi|88950243|NW_921351.1	Homo sapiens chromosome 1 genomic contig, alternate assembly (bas	411	2858	94%	2e-111	100%
gi|88942921|NT_032977.8	Homo sapiens chromosome 1 genomic contig, reference assembly	411	2853	94%	2e-111	100%

Links to Map Viewer



Symbol	O	
Blast hit	Identity=96%	1..53
Blast hit	Identity=97%	73..304
Blast hit	Identity=100%	303..453
Blast hit	Identity=98%	451..551
Blast hit	Identity=96%	549..676
Blast hit	Identity=98%	679..875
Blast hit	Identity=98%	874..970
Blast hit	Identity=97%	965..120
Blast hit	Identity=98%	1198..132
Blast hit	Identity=98%	1323..144
NT_032977.8		
Blast hit	Identity=96%	1442..164

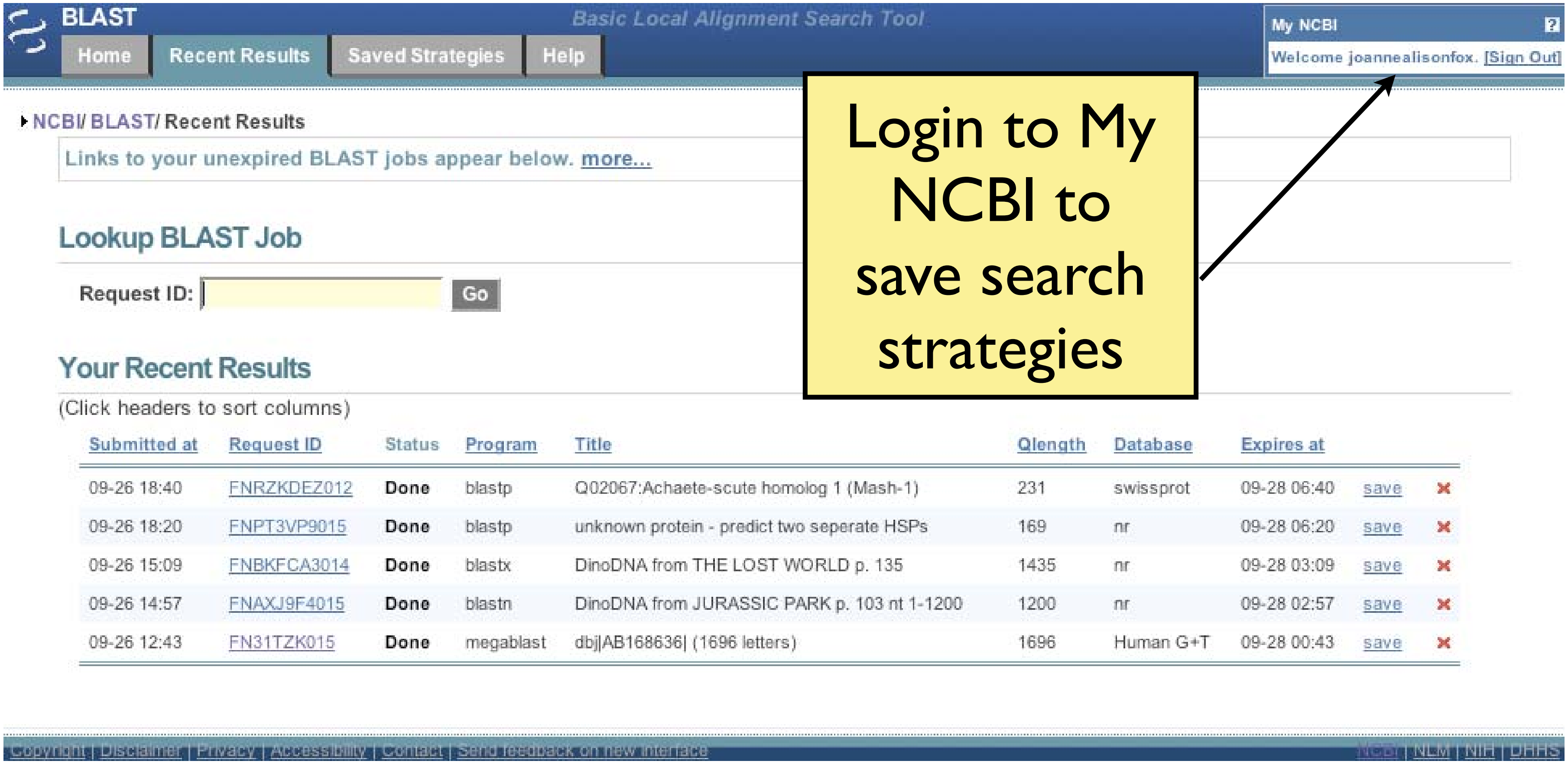
Chromosome 1



Symbol	O	
NT_023935.17		
Blast hit	Identity=95%	42..1644

Chromosome 9

Recent and Saved Strategies



Welcome joannealisonfox. [Sign Out]

NCBI/ BLAST/ Recent Results

Links to your unexpired BLAST jobs appear below. more...

Lookup BLAST Job

Request ID:

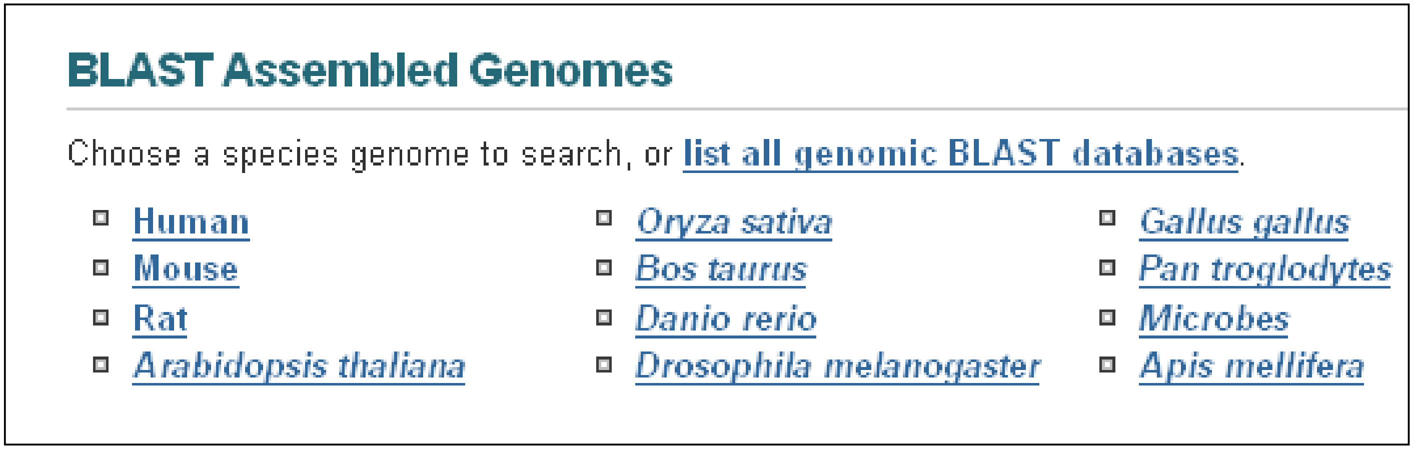
Your Recent Results

(Click headers to sort columns)

Submitted at	Request ID	Status	Program	Title	Qlength	Database	Expires at		
09-26 18:40	FNRZKDEZ012	**Done**	blastp	Q02067:Achaete-scute homolog 1 (Mash-1)	231	swissprot	09-28 06:40	save	✕
09-26 18:20	FNPT3VP9015	**Done**	blastp	unknown protein - predict two seperate HSPs	169	nr	09-28 06:20	save	✕
09-26 15:09	FNBKFCA3014	**Done**	blastx	DinoDNA from THE LOST WORLD p. 135	1435	nr	09-28 03:09	save	✕
09-26 14:57	FNAXJ9F4015	**Done**	blastn	DinoDNA from JURASSIC PARK p. 103 nt 1-1200	1200	nr	09-28 02:57	save	✕
09-26 12:43	FN31TZK015	**Done**	megablast	dbj|AB168636| (1696 letters)	1696	Human G+T	09-28 00:43	save	✕

Login to My NCBI to save search strategies

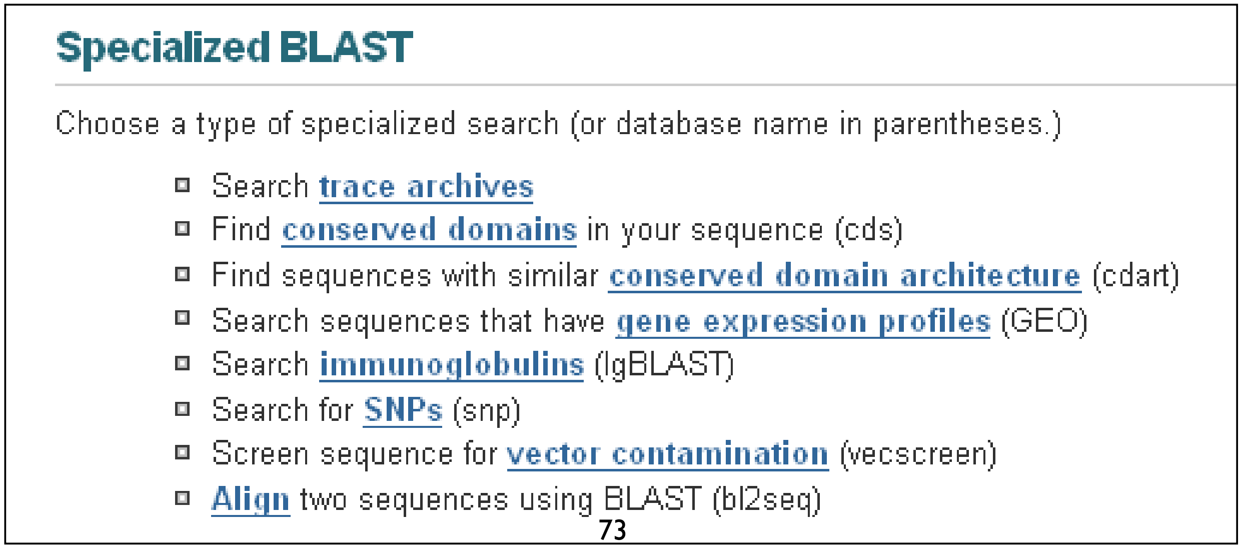
Genomic and Specialized BLAST pages



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*



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)

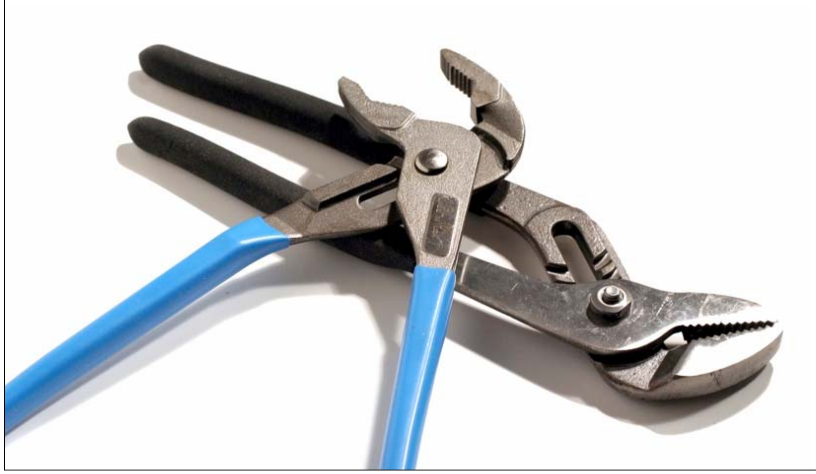
Service Addresses

•***General Help***	`info@ncbi.nlm.nih.gov`
•***BLAST***	`blast-help@ncbi.nlm.nih.gov`

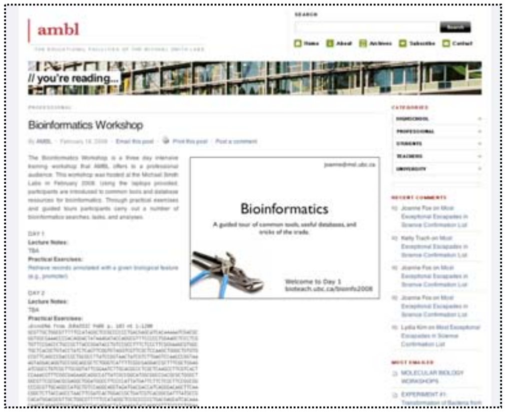
Telephone support: 301- 496- 2475

BLAST

PRACTICAL EXERCISE: The Jurassic Park Detective Story



navigate to:
bioteach.ubc.ca/bioinfo2008



ambl

// you're reading...

Bioinformatics Workshop

DAY 1
Lecture Notes:
TBA
Practical Exercises:

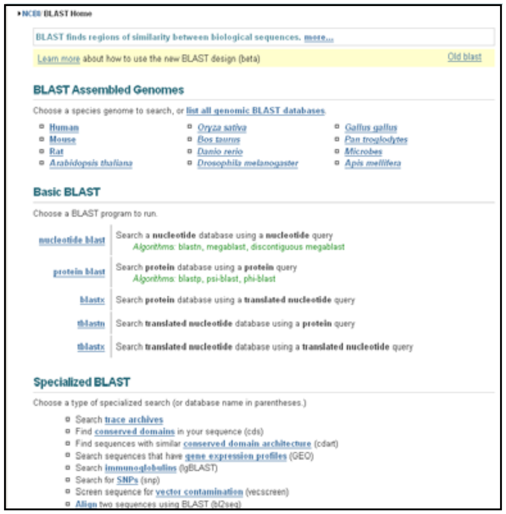
DAY 2
Lecture Notes:
TBA
Practical Exercises:

Bioinformatics

Welcome to Day 1
bioteach.ubc.ca/bioinfo2008

MOLECULAR BIOLOGY WORKSHOPS

EXPERIMENT #1



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

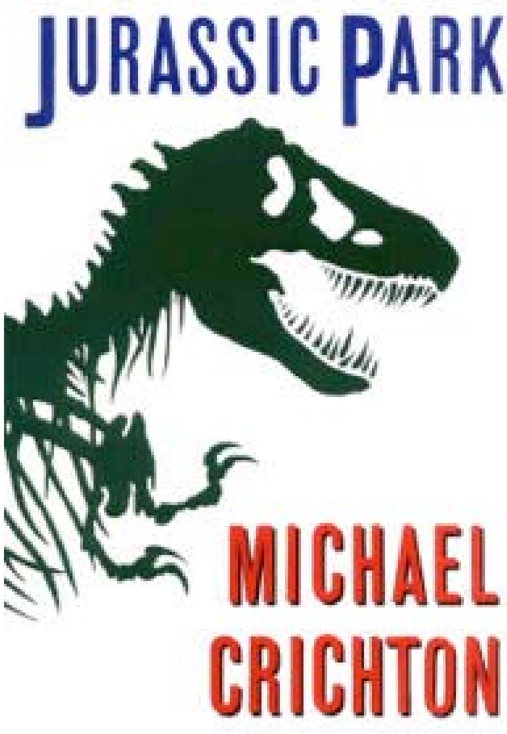
Search #2:
The Lost World
sequence

use blastx

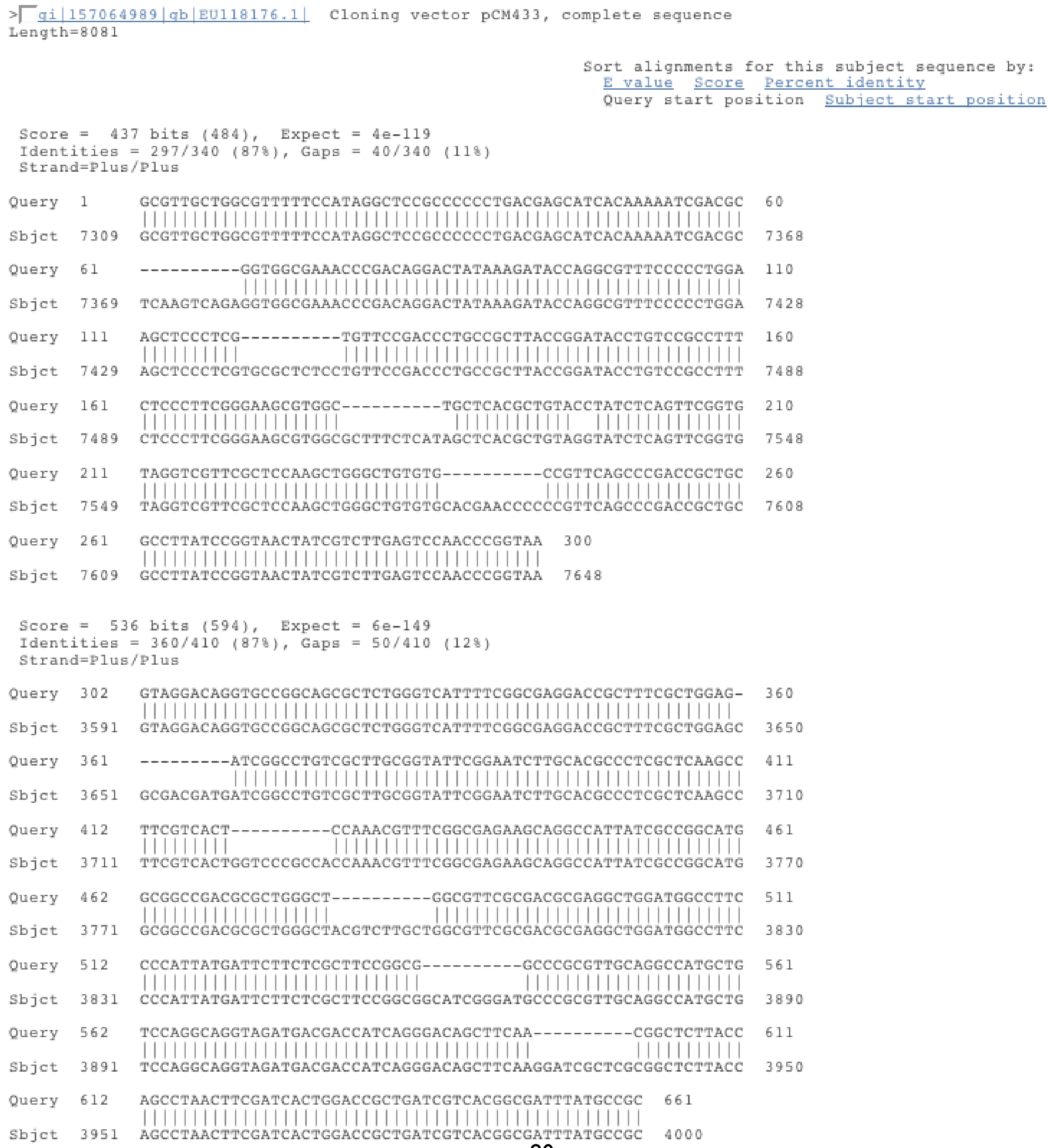
Try some BLAST searches with your own sequence of interest…

Explore what happens when you change advanced parameters…

Search #1 - blastn against nr



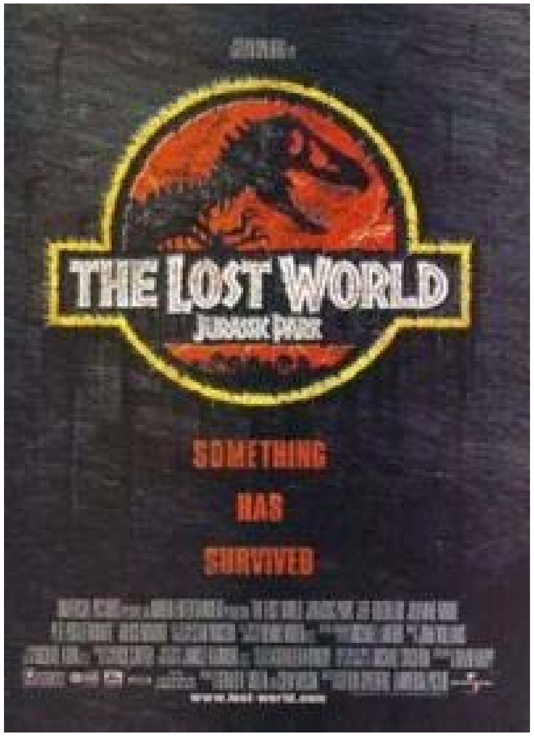
- Most common use of blastn
  - ✓ Sequence identification
  - ✓ Establish whether an exact match for a sequence is already present in the database

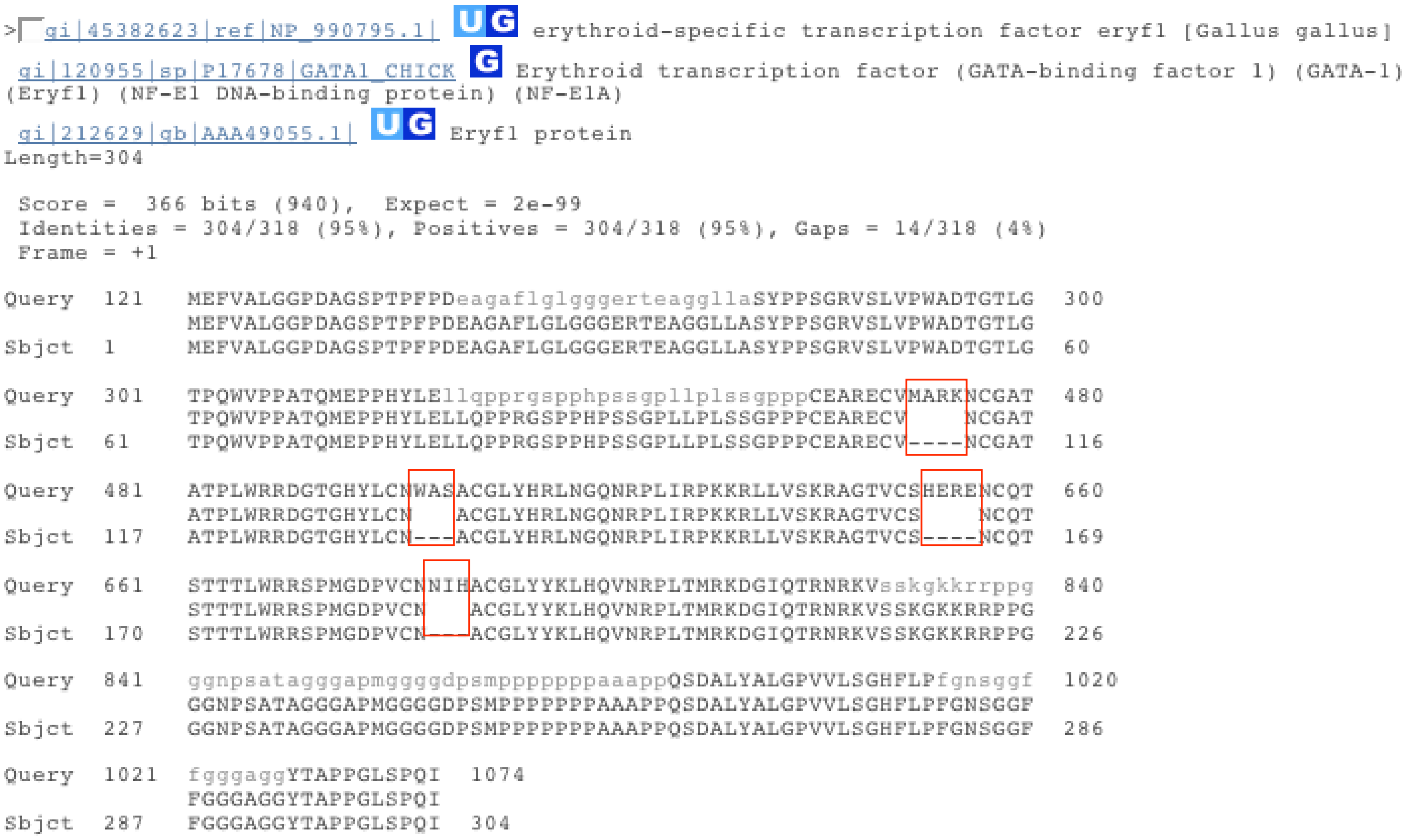


```
>gi|157064989|gb|EU118176.1|  Cloning vector pCM433, complete sequence
Length=8081

                                                         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

Credits

- Materials for this presentation have been adapted with permission from the following NCBI HelpDesk course materials:

  Field Guide Course Materials

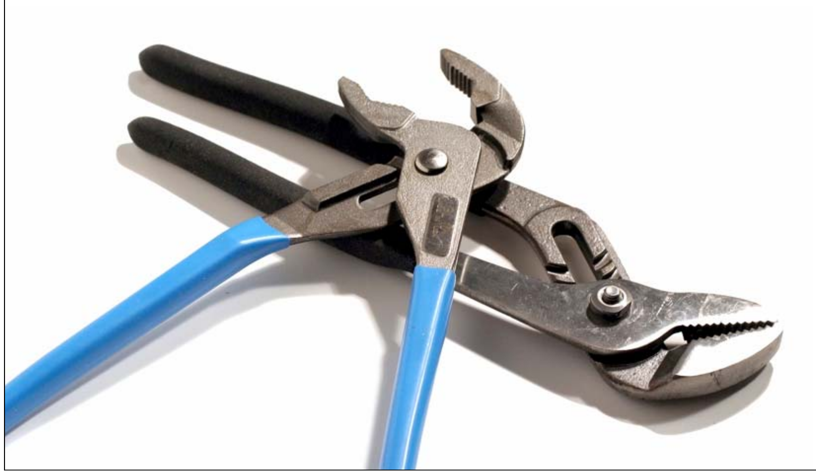
  Advanced Workshop for Bioinformatics Information Specialists

- NCBI BLAST

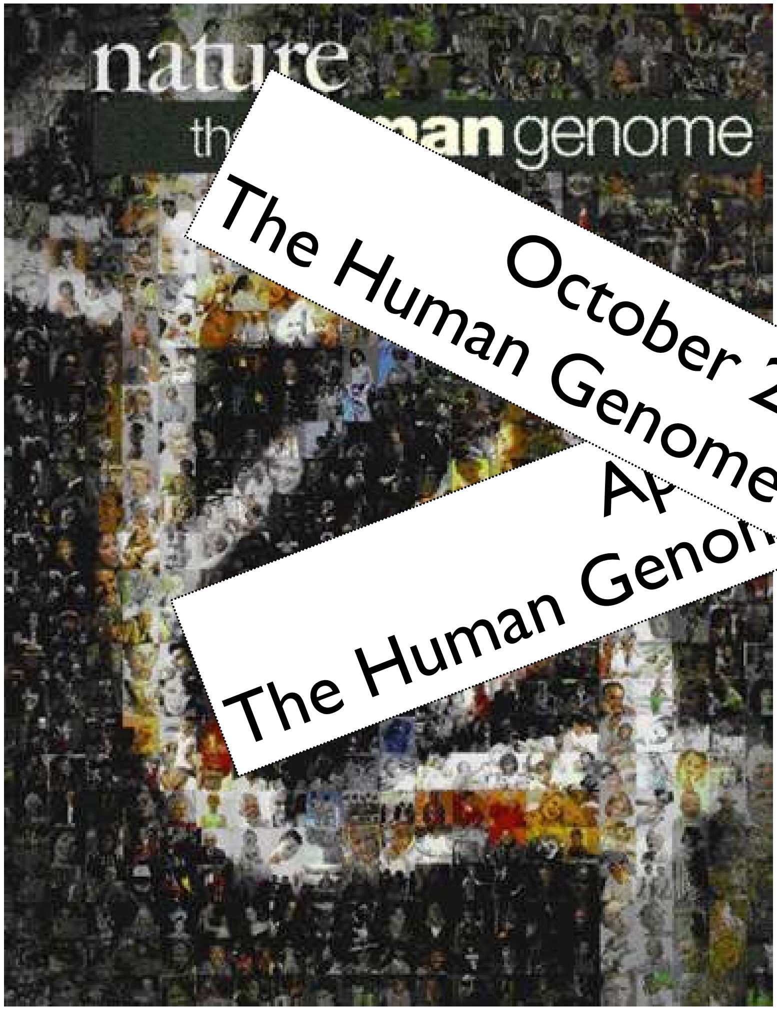
  http://www.ncbi.nlm.nih.gov/blast/Blast.cgi

Genome Browsers

Accessing Genome Annotations &
PRACTICAL EXERCISE: Three Different
Views of the BRCA1 Gene



The Human Genome Project



003:
A
The Human Genom
mpleted – again!

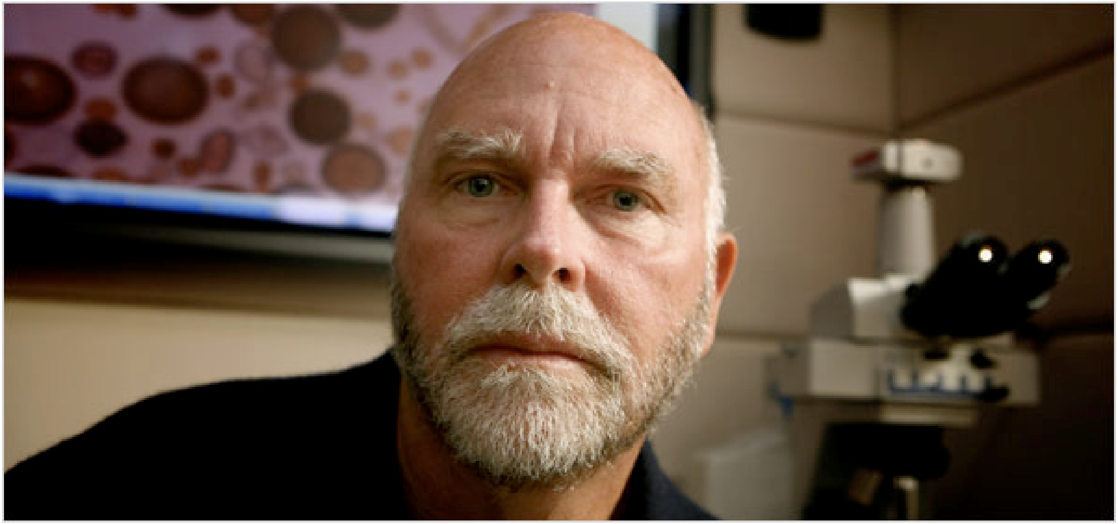
October 2004 - present:
The Human Genome is now really finished!

Public HGP

Celera Genomics

February 2001: Completion of the Draft Human Genome

In the Genome Race, the Sequel Is Personal



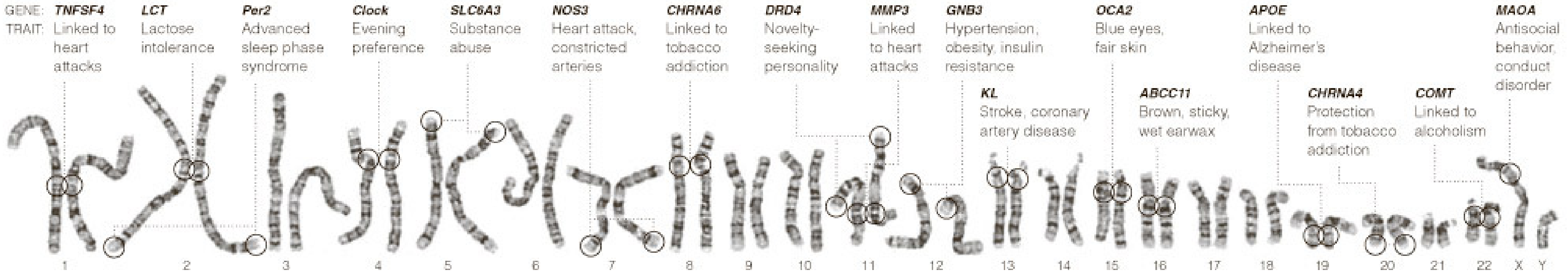
Thor Swift for The New York Times

A team led by J. Craig Venter, above, has finished the first mapping of a full, or diploid, genome, made up of DNA inherited from both parents. The genome is Dr. Venter's own.

The New York Times

September 3, 2007

DECODING HIMSELF A team led by J. Craig Venter, above, has finished the first mapping of a full, or diploid, genome, made up of DNA inherited from both parents. The genome is Dr. Venter's own.



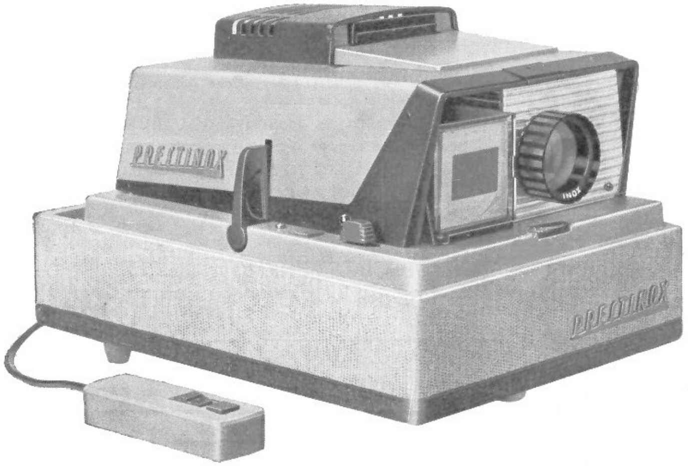
GENE:	TRAIT:
TNFSF4	Linked to heart attacks
LCT	Lactose intolerance
Per2	Advanced sleep phase syndrome
Clock	Evening preference
SLC6A3	Substance abuse
NOS3	Heart attack, constricted arteries
CHRNA6	Linked to tobacco addiction
DRD4	Novelty-seeking personality
MMP3	Linked to heart attacks
GNB3	Hypertension, obesity, insulin resistance
KL	Stroke, coronary artery disease
OCA2	Blue eyes, fair skin
ABCC11	Brown, sticky, wet earwax
APOE	Linked to Alzheimer's disease
CHRNA4	Protection from tobacco addiction
COMT	Linked to alcoholism
MAOA	Antisocial behavior, conduct disorder

Sources: PLoS Biology; GeneCards

Genes highlighted above are linked to specific traits but do not mean the conditions are inevitable.

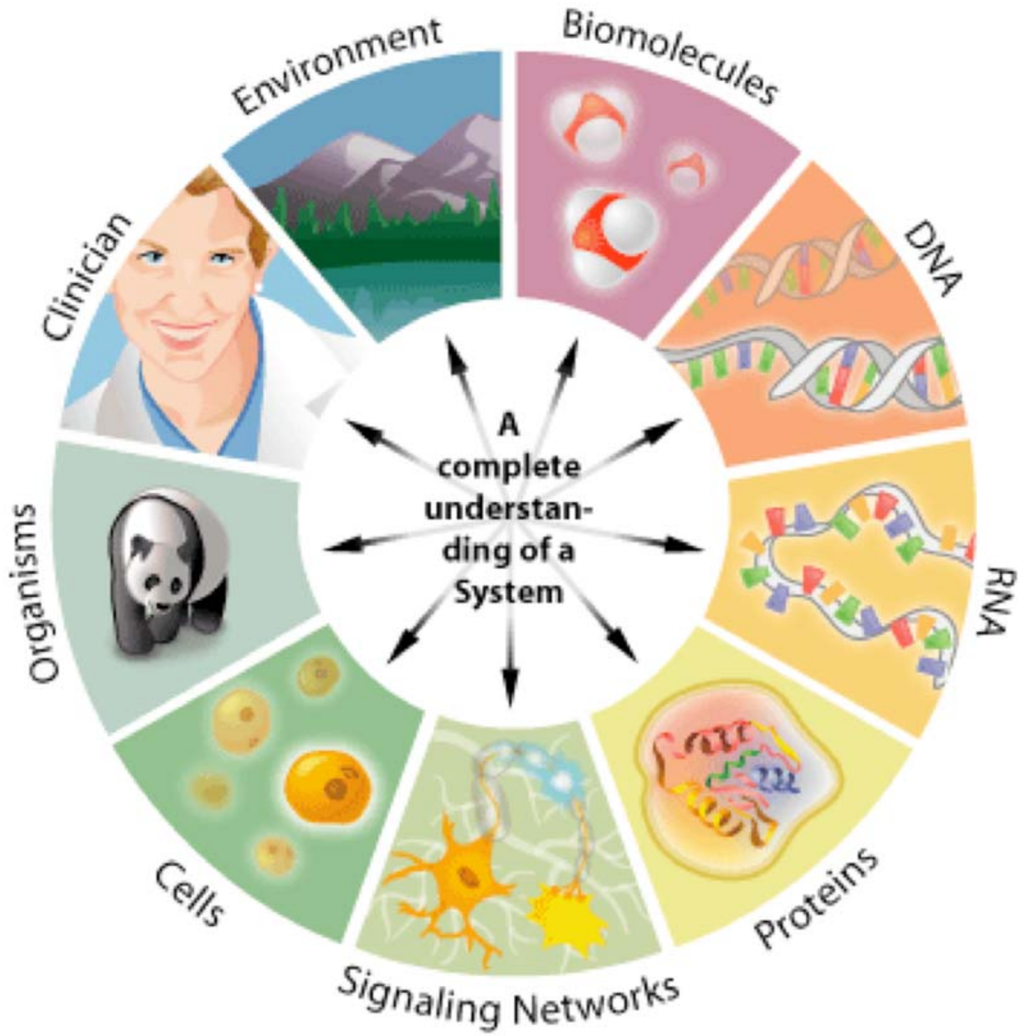
THE NEW YORK TIMES

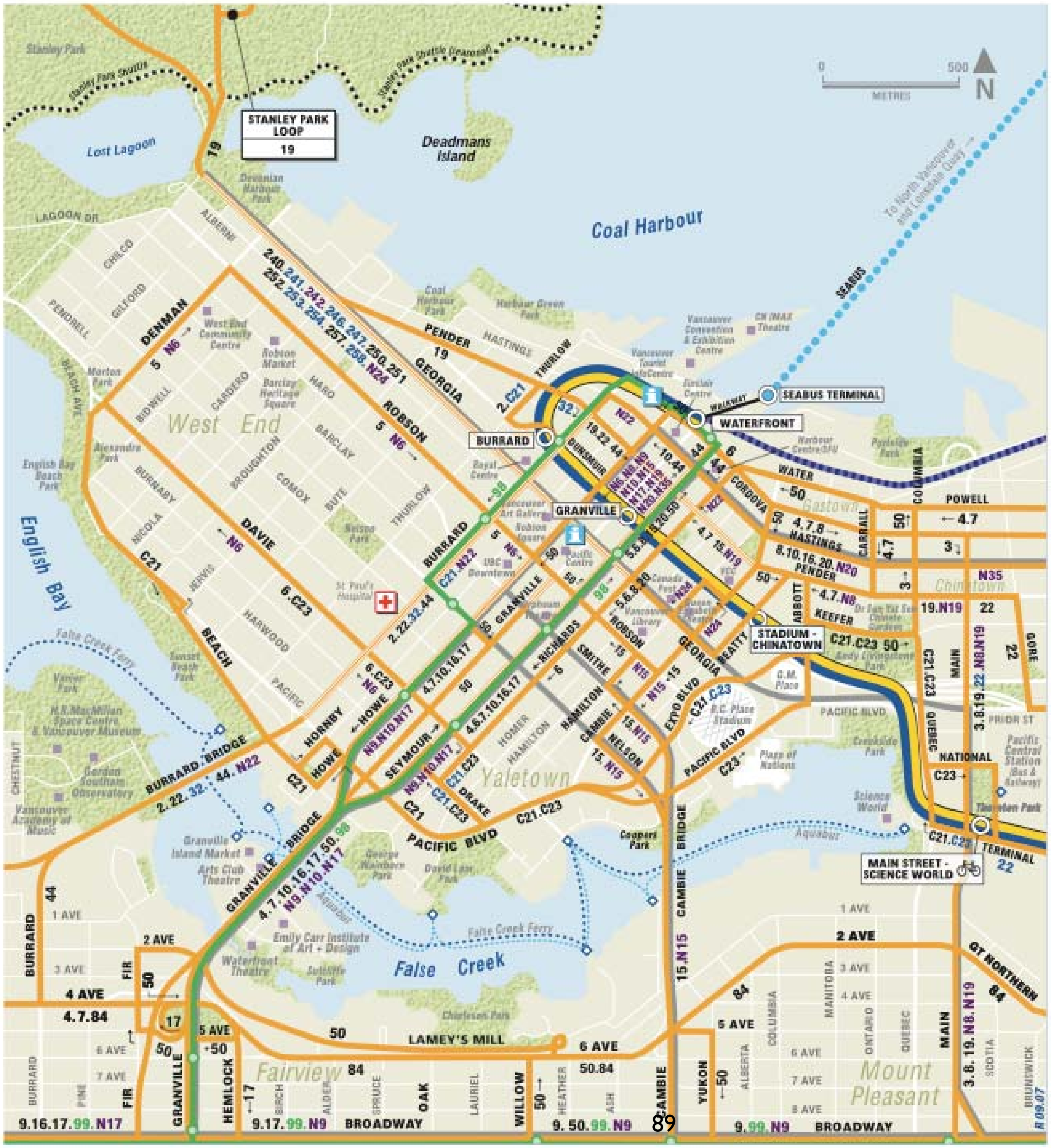
Technology


PRESTINOX
PRESTINOX

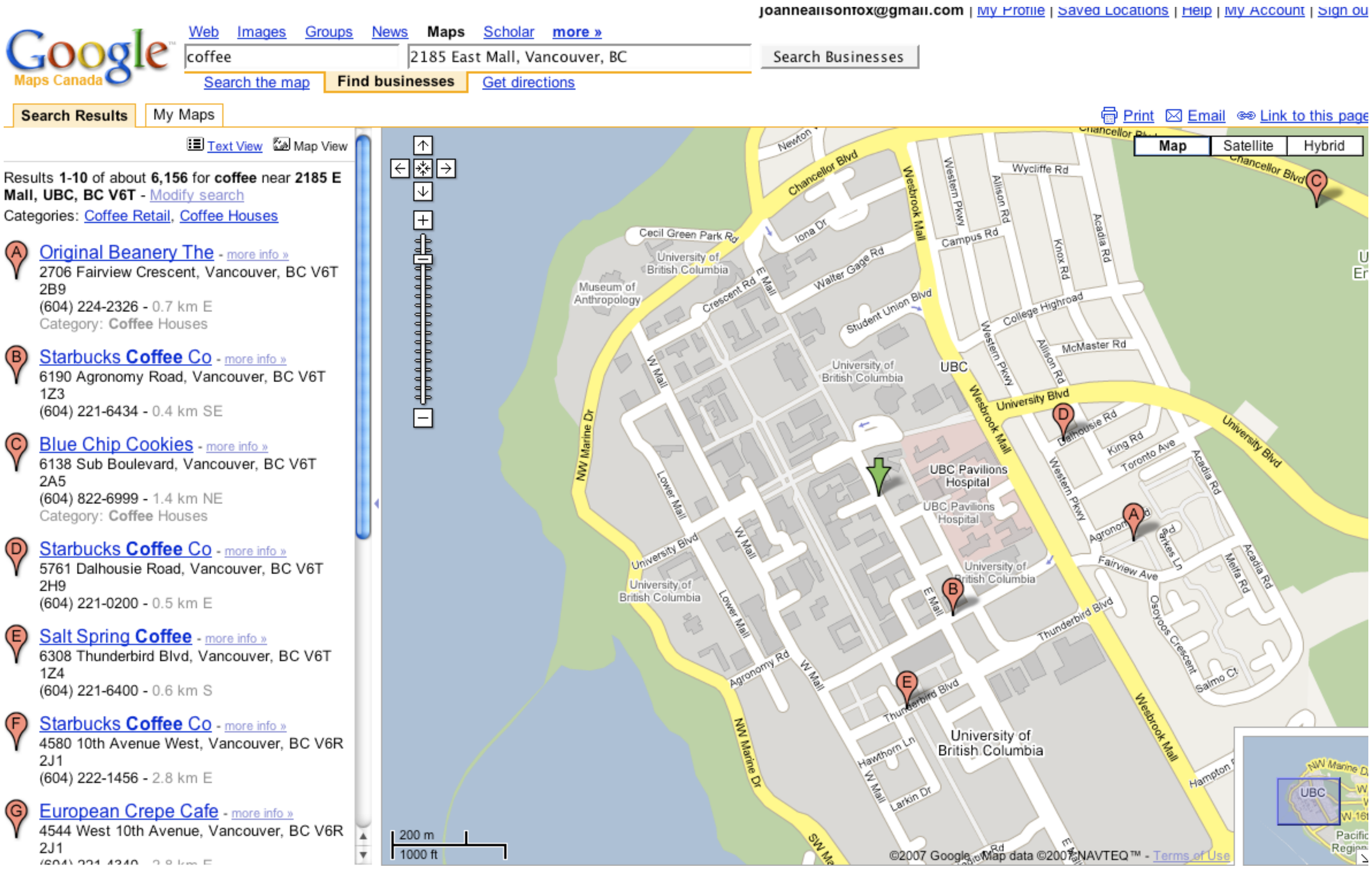

MENU
MENU
MENU
MENU
MENU
FOREVER YOUNG!!

What is Bioinformatics?




Stanley Park
Stanley Park Shuttle
Stanley Park Shuttle (seasonal)
STANLEY PARK LOOP
19
Lost Lagoon
Deadmans Island
Devonian Harbour Park
Coal Harbour
LAGOON DR
ALBERNI
CHILCO
GILFORD
PENDRELL
DENMAN
5 N6
West End Community Centre
Robson Market
Barclay Heritage Square
HARO
ROBSON
5 N6
240.241.242.246.247.250.251
252.253.254.257.258.N24
PENDER
19
GEORGIA
HASTINGS
THURLOW
Coal Harbour Park
Harbour Green Park
Vancouver Convention & Exhibition Centre
CN IMAX Theatre
Vancouver Tourist InfoCentre
Sinclair Centre
SEABUS TERMINAL
WATERFRONT
Walkway
SEABUS
To North Vancouver and Lonsdale Quay
0
500
METRES
N
BURRARD
GRANVILLE
DUNSMUIR
10.22.44
N22
N6.N8.N9 N10.N15 N17.N19 N20.N35
10.44
44
6
Harbour Centre/SFU
Portside Park
WATER
CORDOVA
50
Gastown
CARRALL
COLUMBIA
POWELL
4.7
4.7.8
HASTINGS
8.10.16.20.N20
PENDER
3
N35
Chinatown
4.7.N8
KEEFER
Dr Sun Yat Sen Classical Gardens
19.N19
22
ABBOTT
STADIUM - CHINATOWN
C21.C23
50
Andy Livingstone Park
C21.C23
MAIN
3.8.19.22.N8.N19
22
GORE
PRIOR ST
QUEBEC
NATIONAL
C23
Pacific Central Station (Bus & Railway)
Thornton Park
TERMINAL
22
C21.C23
MAIN STREET - SCIENCE WORLD
Science World
Aquabus
Creekside Park
PACIFIC BLVD
G.M. Place
B.C. Place Stadium
Plaza of Nations
EXPO BLVD
West End
BIDWELL
CARDERO
BARCLAY
BROUGHTON
COMOX
BUTE
Nelson Park
THURLOW
Alexandra Park
English Bay Beach Park
English Bay
Morton Park
BEACH AVE
BURNABY
NICOLA
DAVIE
N6
6.C23
C21
JERVIS
HARWOOD
St. Paul's Hospital
2.22.32.44
BEACH
Sunset Beach Park
PACIFIC
False Creek Ferry
Vanier Park
H.R. MacMillan Space Centre & Vancouver Museum
Gordon Southam Observatory
Vancouver Academy of Music
CHESTNUT
BURRARD BRIDGE
2.22.32.44.N22
HORNBY
HOWE
C21
6.C23 N6
HOWE
N9.N10.N17
4.7.10.16.17
SEYMOUR
N9.N10.N17
4.6.7.10.16.17
RICHARDS
C21.N22
98
UBC Downtown
Vancouver Art Gallery
Robson Square
Pacific Centre
Royal Centre
Orpheum Theatre
Vancouver Public Library
Canada Post
Queen Elizabeth Theatre
ROBSON
SMITHE
HOMER
HAMILTON
CAMBIE
NELSON
15.N15
DRAKE
C21.C23
Yaletown
PACIFIC BLVD
C21.C23
Coopers Park
David Lam Park
George Wainborn Park
Granville Island Market
Arts Club Theatre
GRANVILLE BRIDGE
4.7.10.16.17.50.98
N9.N10.N17
Emily Carr Institute of Art + Design
Waterfront Theatre
Sutcliffe Park
False Creek
Charleson Park
CAMBIE BRIDGE
15.N15
1 AVE
2 AVE
3 AVE
4 AVE
4.7.84
5 AVE
6 AVE
7 AVE
8 AVE
BURRARD
44
FIR
50
17
PINE
GRANVILLE
HEMLOCK
17
BIRCH
ALDER
SPRUCE
OAK
LAUREL
WILLOW
HEATHER
ASH
Fairview
84
LAMEY'S MILL
50.84
YUKON
ALBERTA
COLUMBIA
MANITOBA
ONTARIO
QUEBEC
MAIN
SCOTIA
BRUNSWICK
Mount Pleasant
GT NORTHERN
84
9.16.17.99.N17
9.17.99.N9
BROADWAY
9.50.99.N9
9.99.N9
BROADWAY
R 09.07

maps.google.ca



joannealisonfox@gmail.com | My Profile | Saved Locations | Help | My Account | Sign ou

Web Images Groups News **Maps** Scholar **more »**

coffee | 2185 East Mall, Vancouver, BC | Search Businesses

Search the map | **Find businesses** | Get directions

Search Results | My Maps

Print | Email | Link to this page

Text View | Map View

Results **1-10** of about **6,156** for **coffee** near **2185 E Mall, UBC, BC V6T** - Modify search

Categories: Coffee Retail, Coffee Houses

A **Original Beanery The** - more info »
2706 Fairview Crescent, Vancouver, BC V6T 2B9
(604) 224-2326 - 0.7 km E
Category: **Coffee** Houses

B **Starbucks Coffee Co** - more info »
6190 Agronomy Road, Vancouver, BC V6T 1Z3
(604) 221-6434 - 0.4 km SE

C **Blue Chip Cookies** - more info »
6138 Sub Boulevard, Vancouver, BC V6T 2A5
(604) 822-6999 - 1.4 km NE
Category: **Coffee** Houses

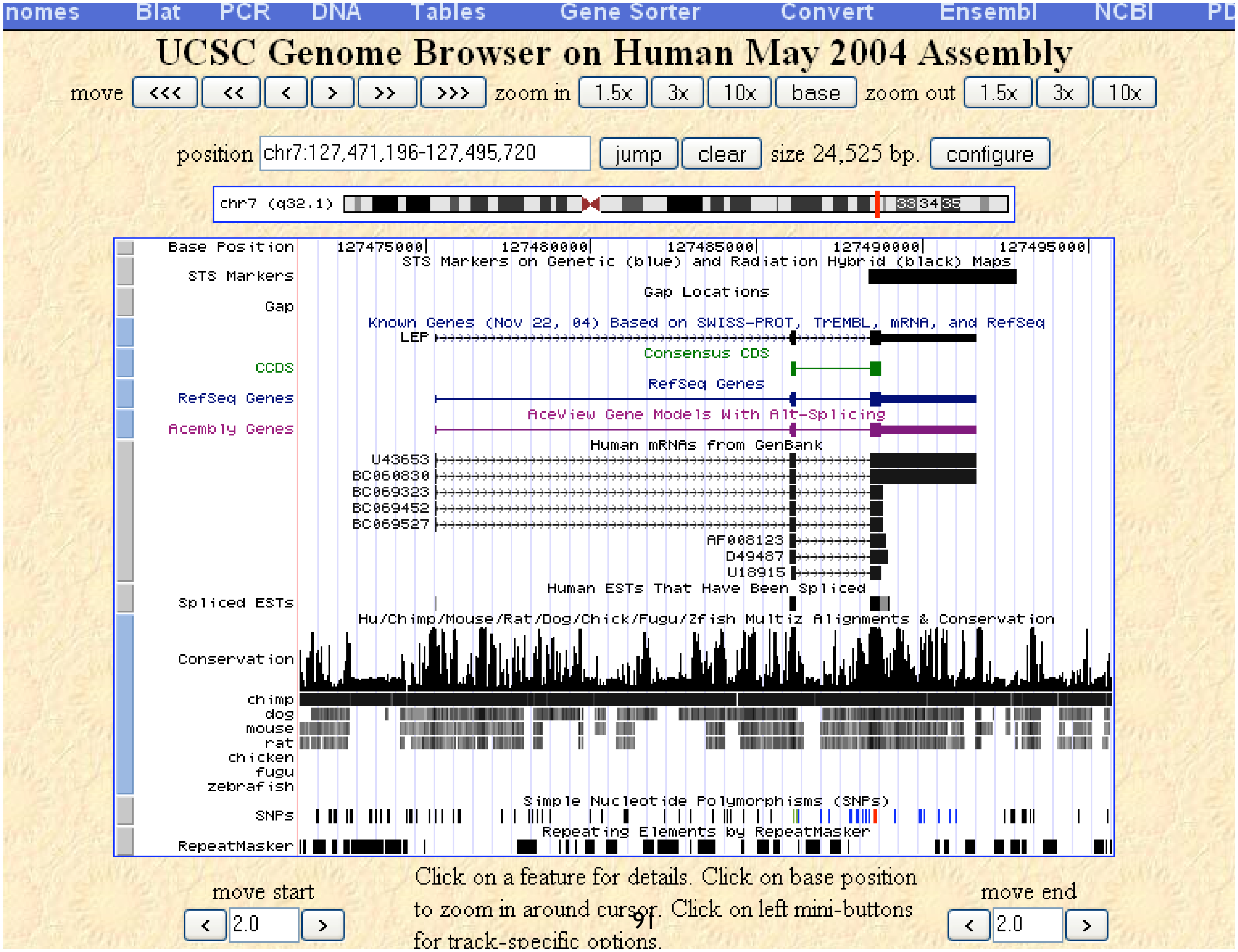
D **Starbucks Coffee Co** - more info »
5761 Dalhousie Road, Vancouver, BC V6T 2H9
(604) 221-0200 - 0.5 km E

E **Salt Spring Coffee** - more info »
6308 Thunderbird Blvd, Vancouver, BC V6T 1Z4
(604) 221-6400 - 0.6 km S

F **Starbucks Coffee Co** - more info »
4580 10th Avenue West, Vancouver, BC V6R 2J1
(604) 222-1456 - 2.8 km E

G **European Crepe Cafe** - more info »
4544 West 10th Avenue, Vancouver, BC V6R 2J1

Let’s Look at the Human Genome...



UCSC Genome Browser on Human May 2004 Assembly

position chr7:127,471,196-127,495,720 size 24,525 bp.

Click on a feature for details. Click on base position to zoom in around cursor. Click on left mini-buttons for track-specific options.

Objectives

- By the end of this module:

✓ You will be able to describe the following concepts: genome annotation, genome builds, and genome browsers.

✓ You will view the genomic location that contains the BRCA1 gene in the human genome using three different genome browsers.

✓ You will be able to compare and contrast the UCSC, Ensembl and MapViewer systems for visualizing genome information.

Genome Browsers

- What is a Genome Browser?
  - System for displaying, viewing, and accessing genome annotation data
- Genome annotations = knowledge attached to raw genome sequence.
  - Annotation information comes from many different sources
    - ✓ Computational pipelines
    - ✓ Research groups
    - ✓ Databases

The “Neopolitan Ice Cream” World of Genome Browsing:

- UCSC Genome Browser

  http://genome.cse.ucsc.edu/

- Ensembl

  http://www.ensembl.org/

- NCBI Map Viewer

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

The underlying data is common for all three “flavors” of Genome Browsers.

- NCBI, UCSC and Ensembl use the same human genome assembly that is generated by NCBI
  - release timing is different between sites.
- Note the version of genome assembly to which you are referring
  - available precomputed info and locations of features will be different between different assemblies.

Let’s compare the view of the BRCA1 gene in all three genome browsers.

Viewing the genomic region containing BRCA1

- Common features:

✓ Coordinate system is based on the build

✓ Zoom in and out

✓ Annotations displayed – ie. Gene features

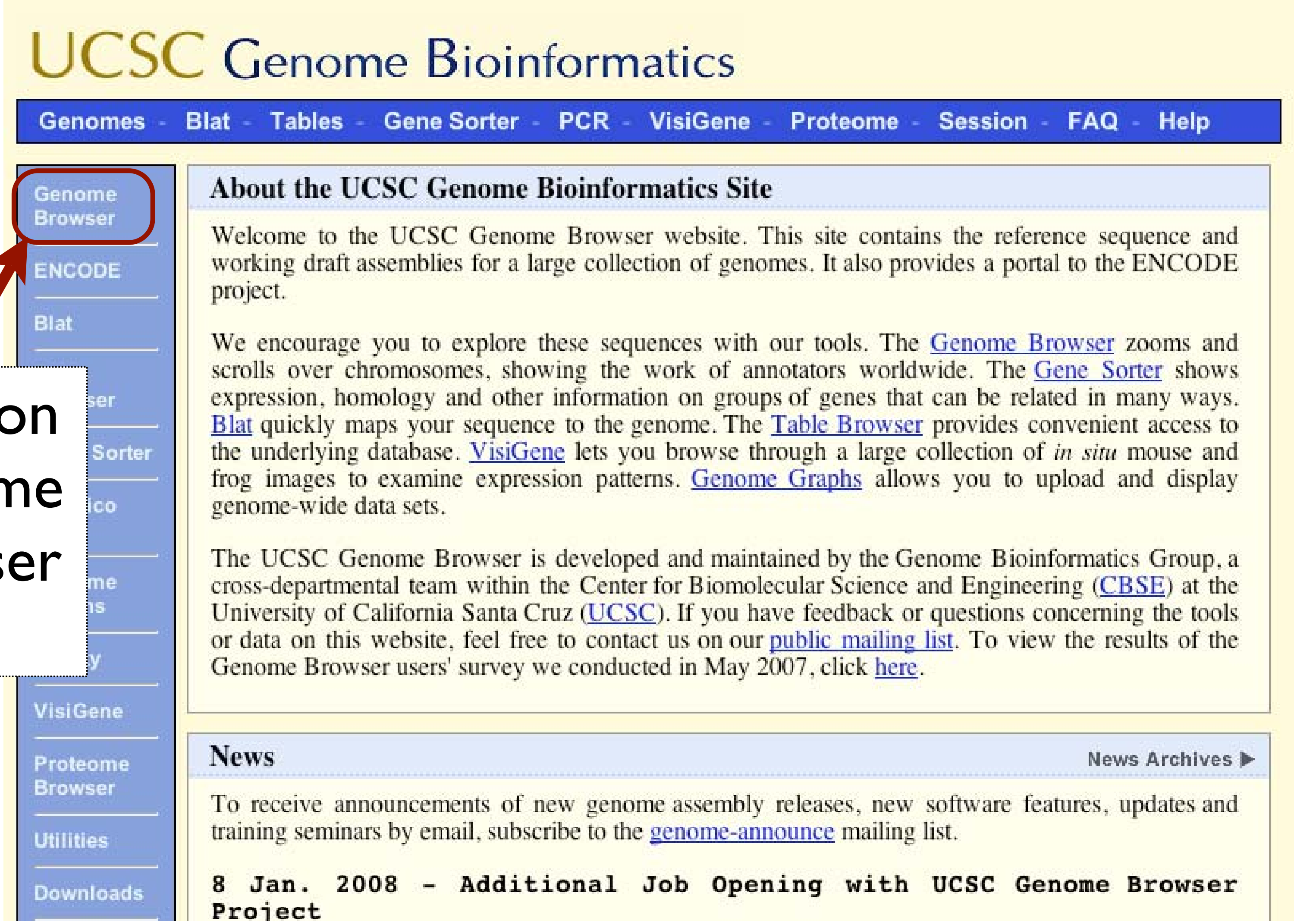
- Major Differences:

✓ Each Browser has a very different look and feel

✓ Annotation information displayed differently

✓ Different ways to navigate through the information

http://genome.cse.ucsc.edu/



About the UCSC Genome Bioinformatics Site

Welcome to the UCSC Genome Browser website. This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.

We encourage you to explore these sequences with our tools. The Genome Browser zooms and scrolls over chromosomes, showing the work of annotators worldwide. The Gene Sorter shows expression, homology and other information on groups of genes that can be related in many ways. Blat quickly maps your sequence to the genome. The Table Browser provides convenient access to the underlying database. VisiGene lets you browse through a large collection of *in situ* mouse and frog images to examine expression patterns. Genome Graphs allows you to upload and display genome-wide data sets.

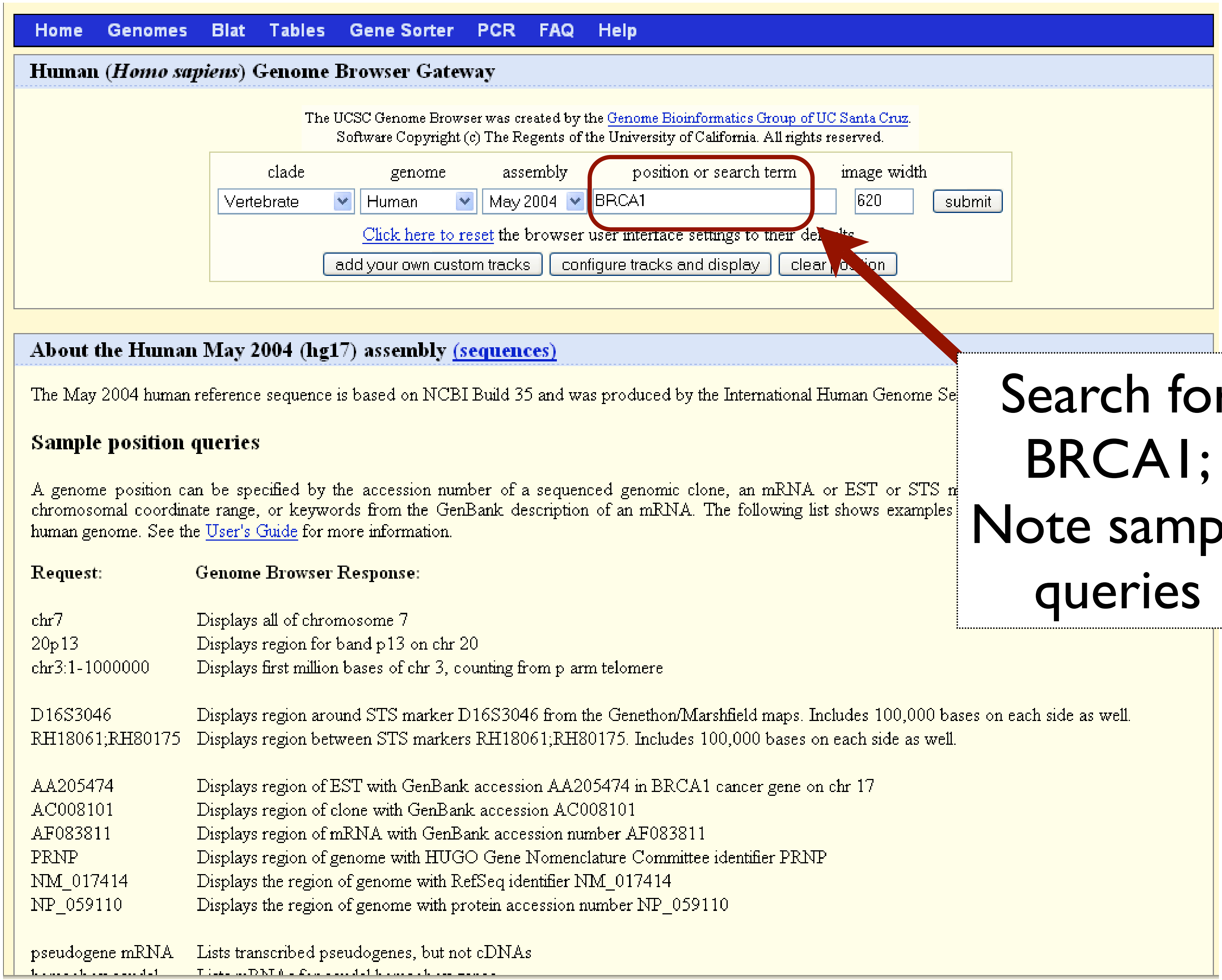
The UCSC Genome Browser is developed and maintained by the Genome Bioinformatics Group, a cross-departmental team within the Center for Biomolecular Science and Engineering (CBSE) at the University of California Santa Cruz (UCSC). If you have feedback or questions concerning the tools or data on this website, feel free to contact us on our public mailing list. To view the results of the Genome Browser users' survey we conducted in May 2007, click here.

News

To receive announcements of new genome assembly releases, new software features, updates and training seminars by email, subscribe to the genome-announce mailing list.

8 Jan. 2008 - Additional Job Opening with UCSC Genome Browser Project

Click on Genome Browser link



Human (*Homo sapiens*) Genome Browser Gateway

The UCSC Genome Browser was created by the Genome Bioinformatics Group of UC Santa Cruz.
Software Copyright (c) The Regents of the University of California. All rights reserved.

clade	genome	assembly	position or search term	image width
Vertebrate	Human	May 2004	BRCA1	620

Click here to reset the browser user interface settings to their defaults.

About the Human May 2004 (hg17) assembly (sequences)

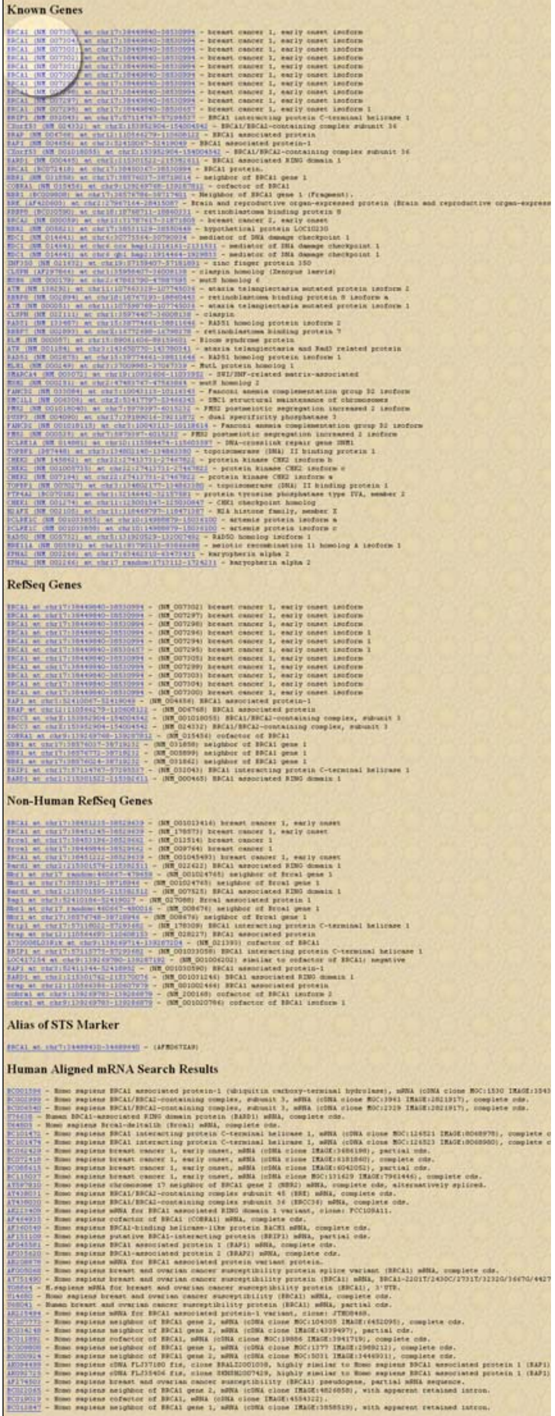
The May 2004 human reference sequence is based on NCBI Build 35 and was produced by the International Human Genome Se

Sample position queries

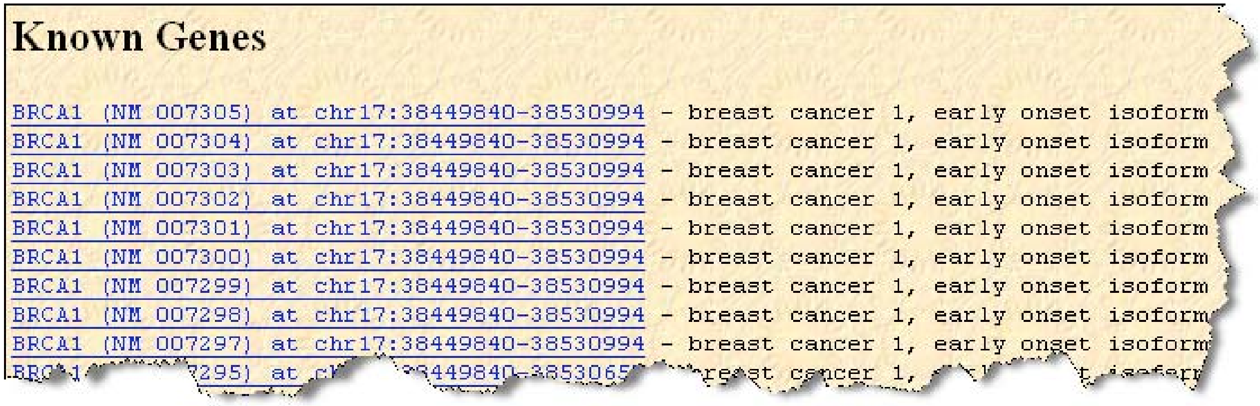
A genome position can be specified by the accession number of a sequenced genomic clone, an mRNA or EST or STS m chromosomal coordinate range, or keywords from the GenBank description of an mRNA. The following list shows examples human genome. See the User's Guide for more information.

Request:	Genome Browser Response:
chr7	Displays all of chromosome 7
20p13	Displays region for band p13 on chr 20
chr3:1-1000000	Displays first million bases of chr 3, counting from p arm telomere
D16S3046	Displays region around STS marker D16S3046 from the Genethon/Marshfield maps. Includes 100,000 bases on each side as well.
RH18061;RH80175	Displays region between STS markers RH18061;RH80175. Includes 100,000 bases on each side as well.
AA205474	Displays region of EST with GenBank accession AA205474 in BRCA1 cancer gene on chr 17
AC008101	Displays region of clone with GenBank accession AC008101
AF083811	Displays region of mRNA with GenBank accession number AF083811
PRNP	Displays region of genome with HUGO Gene Nomenclature Committee identifier PRNP
NM_017414	Displays the region of genome with RefSeq identifier NM_017414
NP_059110	Displays the region of genome with protein accession number NP_059110
pseudogene mRNA	Lists transcribed pseudogenes, but not cDNAs

Search for BRCA1;
Note sample queries



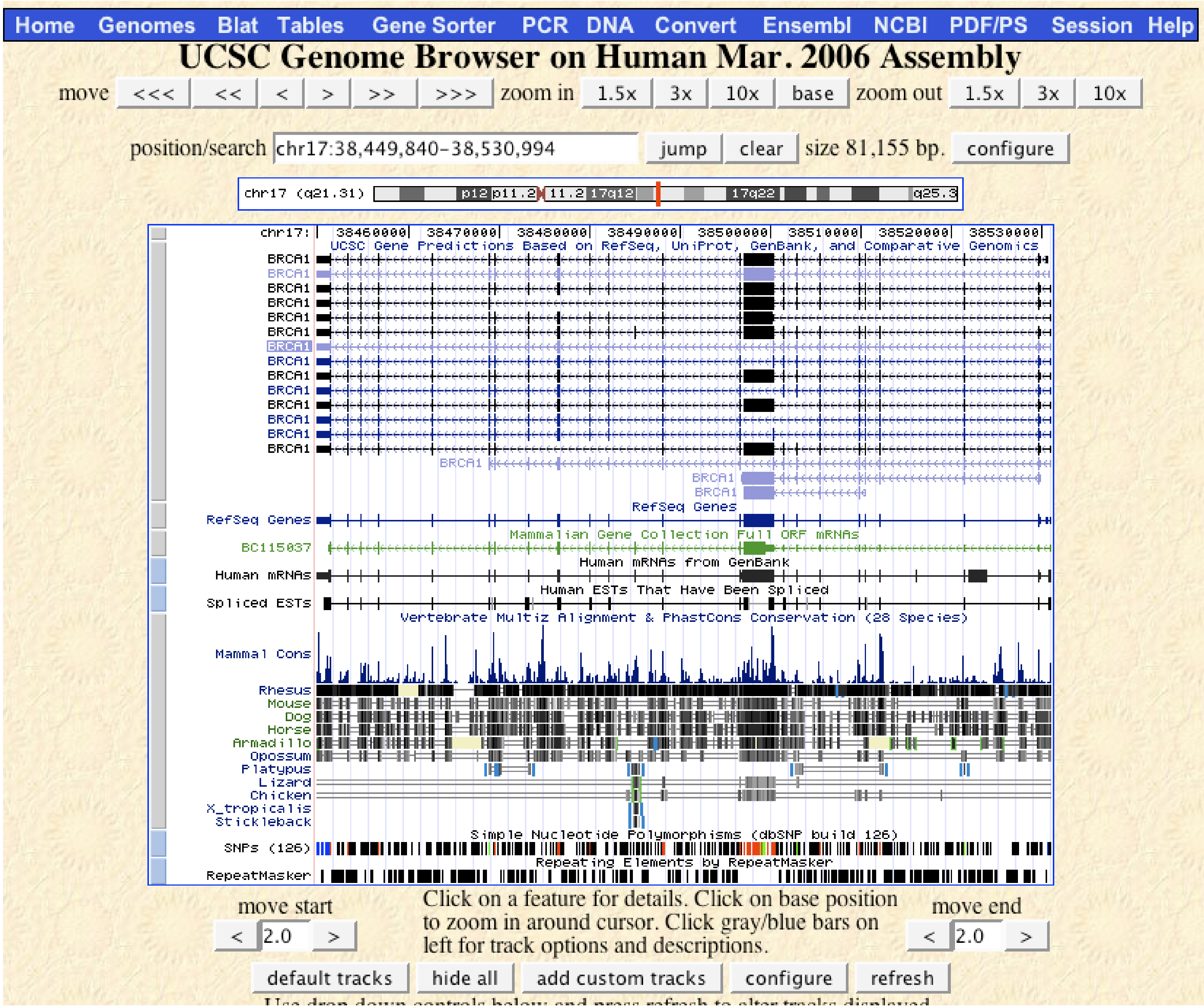
The Search Results



Known Genes

BRCA1 (NM 007305) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007304) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007303) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007302) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007301) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007300) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007299) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007298) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRCA1 (NM 007297) at chr17:38449840-38530994 - breast cancer 1, early onset isoform
BRC 1 7295) at c 3449840-38530657 reast cancer 1,

- Many BRCA1 isoforms
  - ✓ All located on chr 17
  - ✓ same chr coordinates
  - ✓ different gene structures



UCSC Genome Browser on Human Mar. 2006 Assembly

position/search chr17:38,449,840-38,530,994

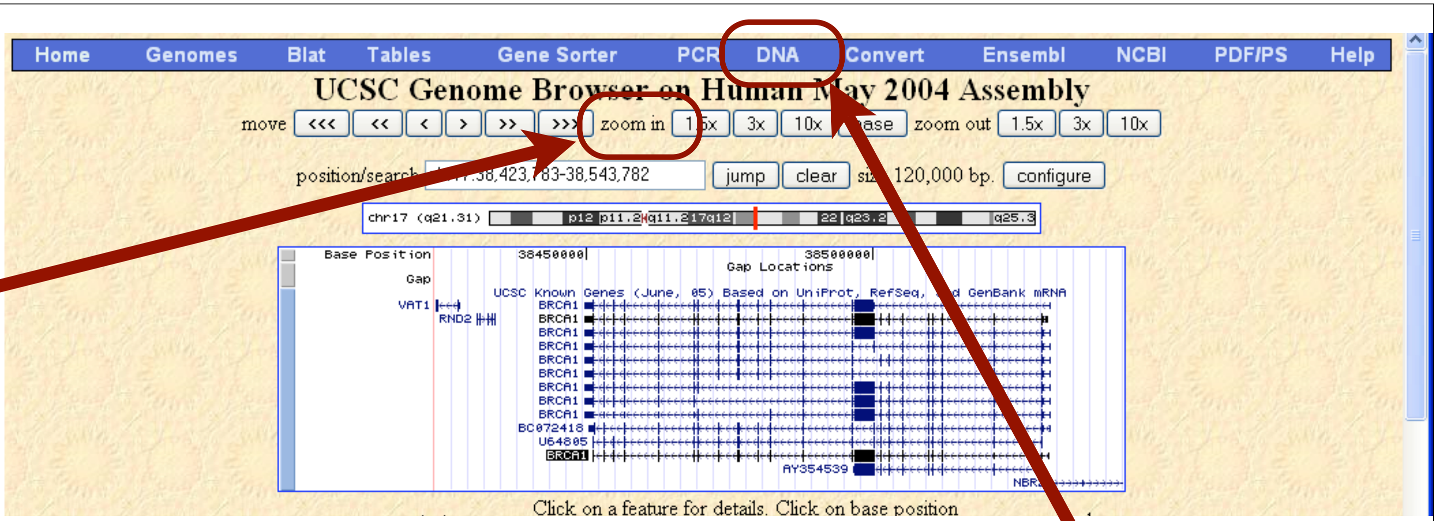
size 81,155 bp.

Click on a feature for details. Click on base position to zoom in around cursor. Click gray/blue bars on left for track options and descriptions.

Use drop down controls below and press refresh to alter tracks displayed.

Two tasks

- What genes are on either side of BRCA1 on chr 17?
- Can you figure out how to download the genomic sequence for the BRCA1 region?



Zoom in
Zoom out

DNA link
Download
Sequence

Get DNA in Window

Get DNA for

Position chr17:38,451,220-38,530,831

Note: if you would prefer to get DNA for features of a particular track or table, try the Table Browser using the output format sequence.

Sequence Retrieval Region Options:

Add 0 extra bases upstream (5') and 0 extra downstream (3')

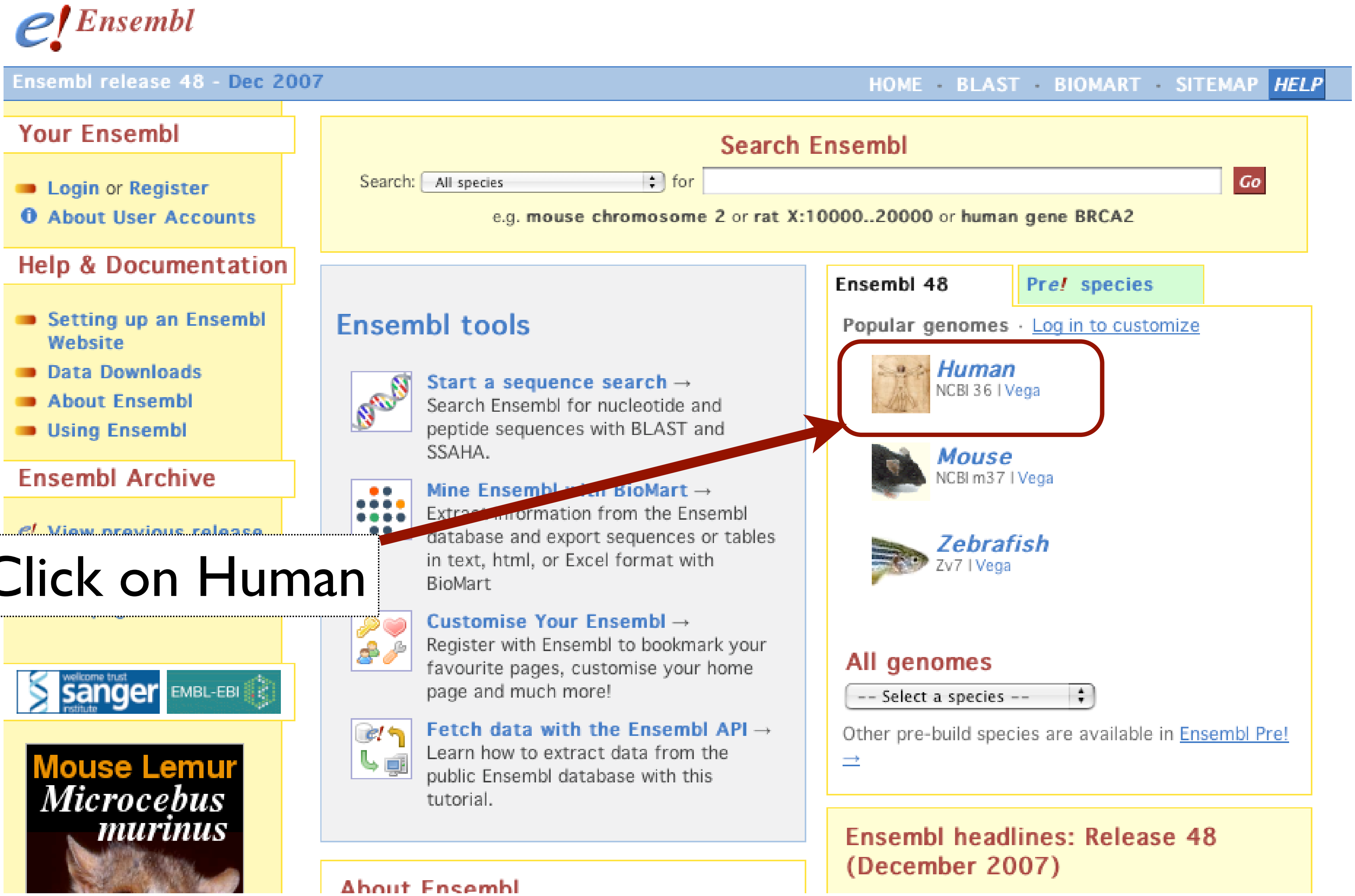
Note: if a feature is close to the beginning or end of a chromosome and upstream/downstream bases are added, they may be truncated in order to avoid extending past the edge of the chromosome.

Sequence Formatting Options:

● All upper case.
○ All lower case.
☐ Mask repeats: ● to lower case ○ to N
☐ Reverse complement (get '-' strand sequence)

Note: The "Mask repeats" option applies only to "get DNA", not to "extended case/color options".

http://www.ensembl.org/



e! Ensembl

Ensembl release 48 - Dec 2007

HOME · BLAST · BIOMART · SITEMAP HELP

Your Ensembl

- Login or Register
- About User Accounts

Help & Documentation

- Setting up an Ensembl Website
- Data Downloads
- About Ensembl
- Using Ensembl

Ensembl Archive

- View previous release

Search Ensembl

Search: All species for Go

e.g. **mouse chromosome 2** or **rat X:10000..20000** or **human gene BRCA2**

Ensembl tools

Start a sequence search →
Search Ensembl for nucleotide and peptide sequences with BLAST and SSAHA.

Mine Ensembl with BioMart →
Extract information from the Ensembl database and export sequences or tables in text, html, or Excel format with BioMart

Customise Your Ensembl →
Register with Ensembl to bookmark your favourite pages, customise your home page and much more!

Fetch data with the Ensembl API →
Learn how to extract data from the public Ensembl database with this tutorial.

About Ensembl

Ensembl 48 | Pre! species

Popular genomes · Log in to customize

Human NCBI 36 | Vega

Mouse NCBI m37 | Vega

Zebrafish Zv7 | Vega

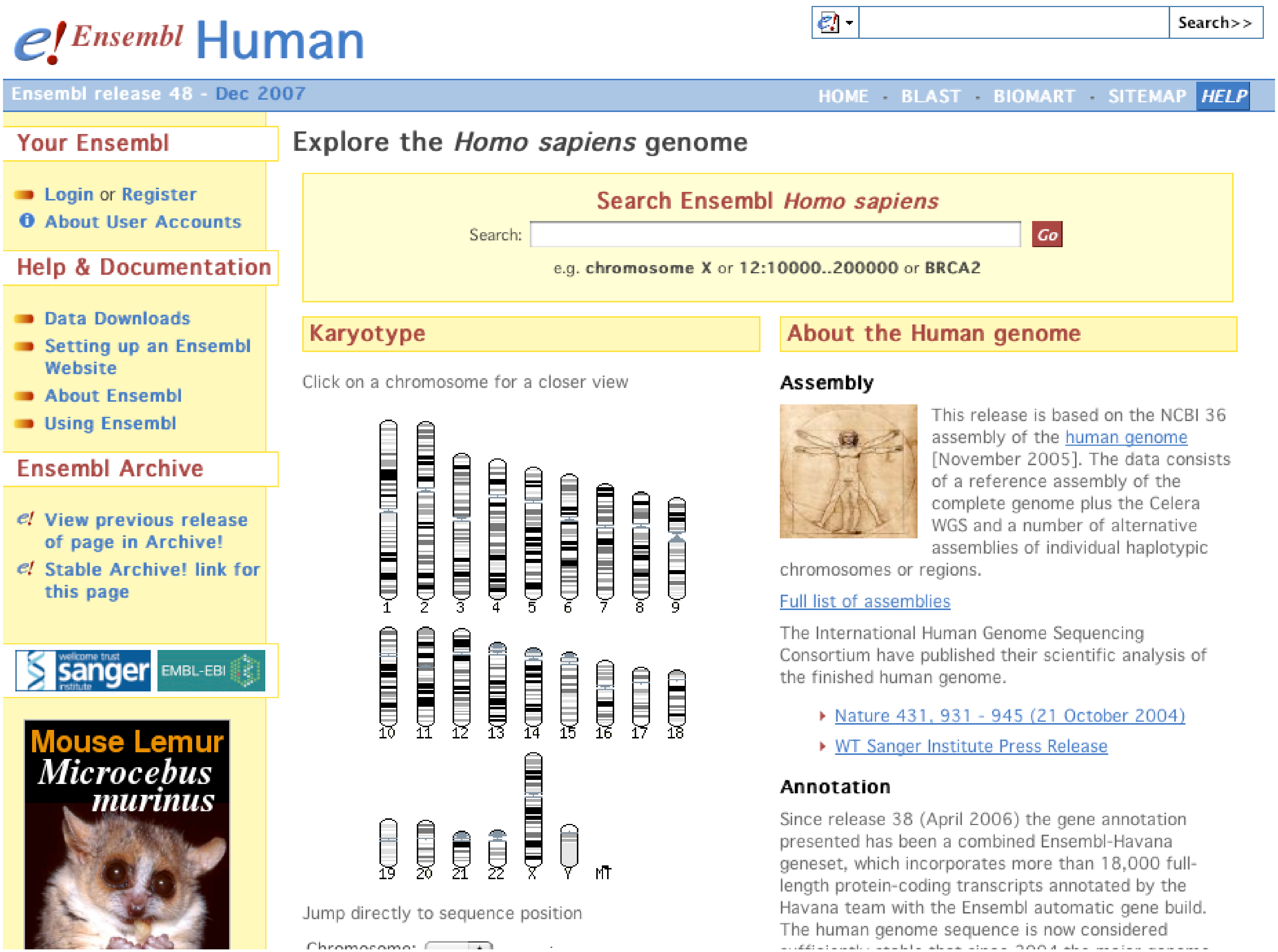
All genomes

-- Select a species --

Other pre-build species are available in Ensembl Pre! →

Ensembl headlines: Release 48 (December 2007)

Click on Human



Ensembl Human

Ensembl release 48 - Dec 2007

Your Ensembl

- Login or Register
- About User Accounts

Help & Documentation

- Data Downloads
- Setting up an Ensembl Website
- About Ensembl
- Using Ensembl

Ensembl Archive

- View previous release of page in Archive!
- Stable Archive! link for this page

Explore the *Homo sapiens* genome

Search Ensembl *Homo sapiens*

Search:

e.g. **chromosome X** or **12:10000..200000** or **BRCA2**

Karyotype

Click on a chromosome for a closer view

Jump directly to sequence position

About the Human genome

Assembly

This release is based on the NCBI 36 assembly of the human genome [November 2005]. The data consists of a reference assembly of the complete genome plus the Celera WGS and a number of alternative assemblies of individual haplotypic chromosomes or regions.

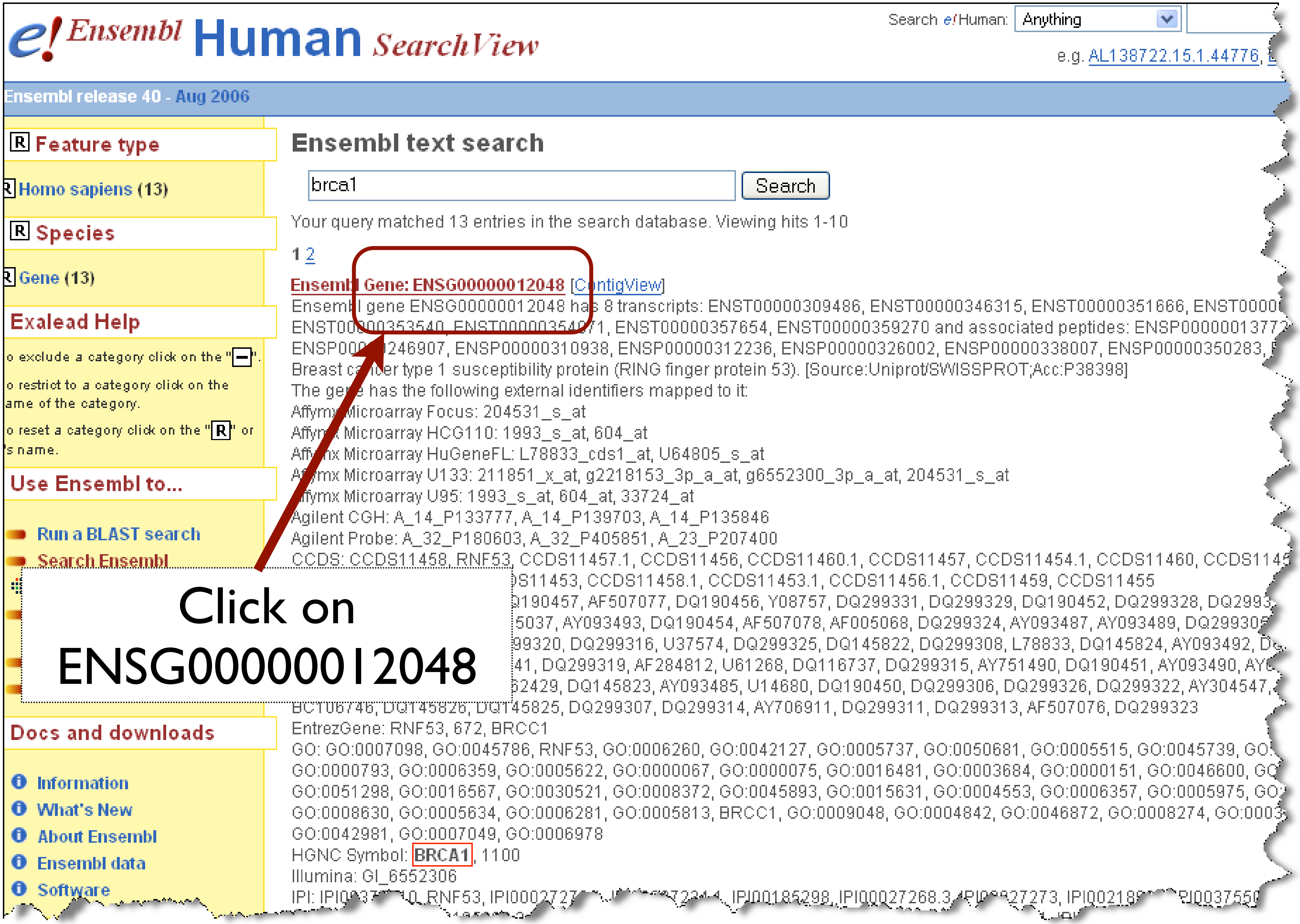
Full list of assemblies

The International Human Genome Sequencing Consortium have published their scientific analysis of the finished human genome.

- Nature 431, 931 - 945 (21 October 2004)
- WT Sanger Institute Press Release

Annotation

Since release 38 (April 2006) the gene annotation presented has been a combined Ensembl-Havana geneset, which incorporates more than 18,000 full-length protein-coding transcripts annotated by the Havana team with the Ensembl automatic gene build. The human genome sequence is now considered



Ensembl text search

brca1

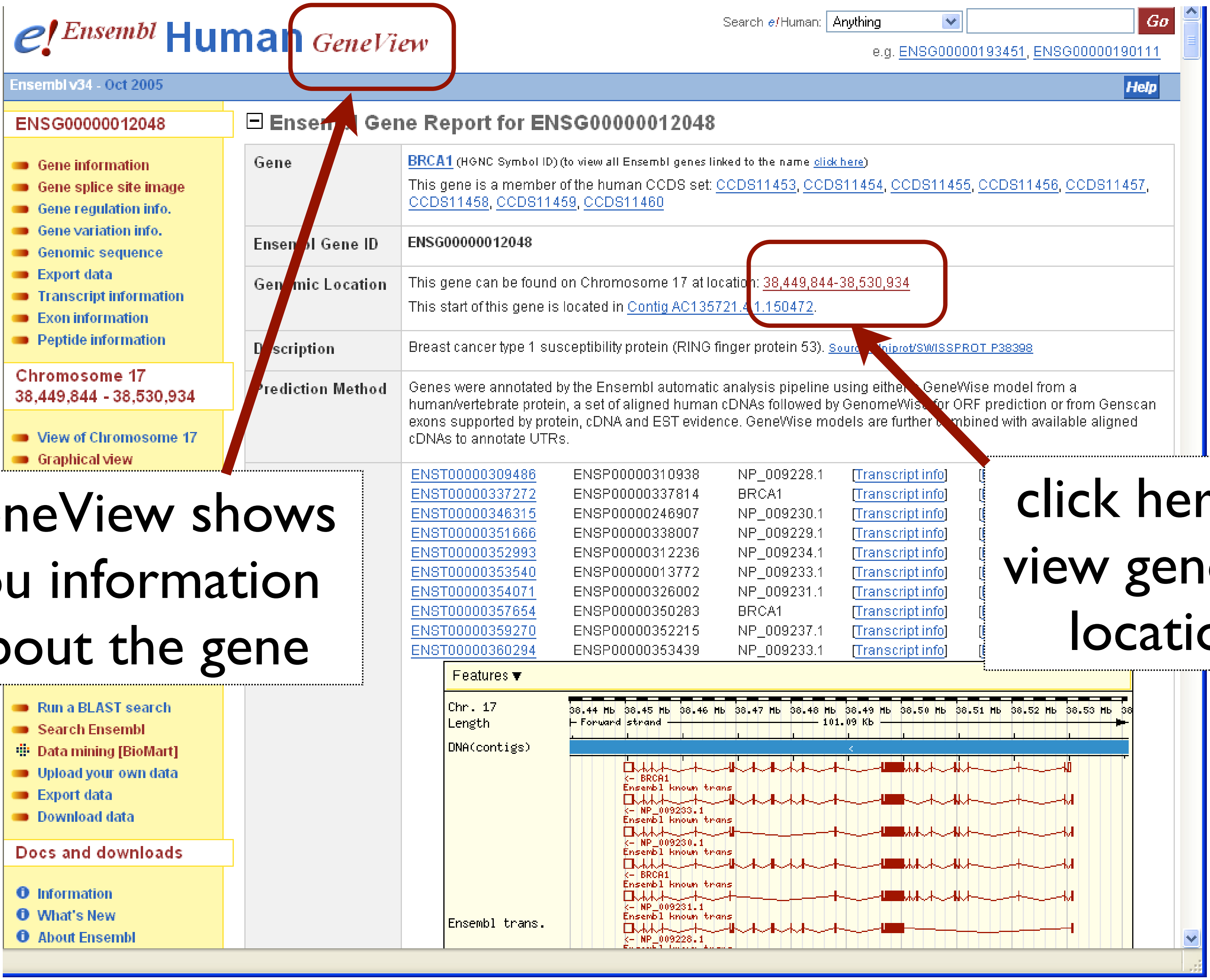
Your query matched 13 entries in the search database. Viewing hits 1-10

1 2

Ensembl Gene: ENSG00000012048 [ContigView]

Ensembl gene ENSG00000012048 has 8 transcripts: ENST00000309486, ENST00000346315, ENST00000351666, ENST0000
ENST00000353540, ENST00000354071, ENST00000357654, ENST00000359270 and associated peptides: ENSP0000013777
ENSP00000246907, ENSP00000310938, ENSP00000312236, ENSP00000326002, ENSP00000338007, ENSP00000350283,
Breast cancer type 1 susceptibility protein (RING finger protein 53). [Source:Uniprot/SWISSPROT;Acc:P38398]
The gene has the following external identifiers mapped to it:
Affymx Microarray Focus: 204531_s_at
Affymx Microarray HCG110: 1993_s_at, 604_at
Affymx Microarray HuGeneFL: L78833_cds1_at, U64805_s_at
Affymx Microarray U133: 211851_x_at, g2218153_3p_a_at, g6552300_3p_a_at, 204531_s_at
Affymx Microarray U95: 1993_s_at, 604_at, 33724_at
Agilent CGH: A_14_P133777, A_14_P139703, A_14_P135846
Agilent Probe: A_32_P180603, A_32_P405851, A_23_P207400
CCDS: CCDS11458, RNF53, CCDS11457.1, CCDS11456, CCDS11460.1, CCDS11457, CCDS11454.1, CCDS11460, CCDS114
DS11453, CCDS11458.1, CCDS11453.1, CCDS11456.1, CCDS11459, CCDS11455
Q190457, AF507077, DQ190456, Y08757, DQ299331, DQ299329, DQ190452, DQ299328, DQ2993
5037, AY093493, DQ190454, AF507078, AF005068, DQ299324, AY093487, AY093489, DQ29930
99320, DQ299316, U37574, DQ299325, DQ145822, DQ299308, L78833, DQ145824, AY093492, D
41, DQ299319, AF284812, U61268, DQ116737, DQ299315, AY751490, DQ190451, AY093490, AY
52429, DQ145823, AY093485, U14680, DQ190450, DQ299306, DQ299326, DQ299322, AY304547,
BC106746, DQ145826, DQ145825, DQ299307, DQ299314, AY706911, DQ299311, DQ299313, AF507076, DQ299323
EntrezGene: RNF53, 672, BRCC1
GO: GO:0007098, GO:0045786, RNF53, GO:0006260, GO:0042127, GO:0005737, GO:0050681, GO:0005515, GO:0045739, GO
GO:0000793, GO:0006359, GO:0005622, GO:0000067, GO:0000075, GO:0016481, GO:0003684, GO:0000151, GO:0046600, GO
GO:0051298, GO:0016567, GO:0030521, GO:0008372, GO:0045893, GO:0015631, GO:0004553, GO:0006357, GO:0005975, GO
GO:0008630, GO:0005634, GO:0006281, GO:0005813, BRCC1, GO:0009048, GO:0004842, GO:0046872, GO:0008274, GO:0003
GO:0042981, GO:0007049, GO:0006978
HGNC Symbol: BRCA1, 1100
Illumina: GI_6552306
IPI: IPI0037 10, RNF53, IPI00027 , IPI00185298, IPI00027268.3, IPI 27273, IPI0021 IPI003755

Click on ENSG00000012048



Ensembl Gene Report for ENSG00000012048

Gene	BRCA1 (HGNC Symbol ID) (to view all Ensembl genes linked to the name click here) This gene is a member of the human CCDS set: CCDS11453, CCDS11454, CCDS11455, CCDS11456, CCDS11457, CCDS11458, CCDS11459, CCDS11460
Ensembl Gene ID	ENSG00000012048
Genomic Location	This gene can be found on Chromosome 17 at location: 38,449,844-38,530,934 This start of this gene is located in Contig AC135721.4.1.150472.
Description	Breast cancer type 1 susceptibility protein (RING finger protein 53). Source: Uniprot/SWISSPROT P38398
Prediction Method	Genes were annotated by the Ensembl automatic analysis pipeline using either a GeneWise model from a human/vertebrate protein, a set of aligned human cDNAs followed by GenomeWise for ORF prediction or from Genscan exons supported by protein, cDNA and EST evidence. GeneWise models are further combined with available aligned cDNAs to annotate UTRs.

ENST00000309486	ENSP00000310938	NP_009228.1	[Transcript info]
ENST00000337272	ENSP00000337814	BRCA1	[Transcript info]
ENST00000346315	ENSP00000246907	NP_009230.1	[Transcript info]
ENST00000351666	ENSP00000338007	NP_009229.1	[Transcript info]
ENST00000352993	ENSP00000312236	NP_009234.1	[Transcript info]
ENST00000353540	ENSP00000013772	NP_009233.1	[Transcript info]
ENST00000354071	ENSP00000326002	NP_009231.1	[Transcript info]
ENST00000357654	ENSP00000350283	BRCA1	[Transcript info]
ENST00000359270	ENSP00000352215	NP_009237.1	[Transcript info]
ENST00000360294	ENSP00000353439	NP_009233.1	[Transcript info]

GeneView shows you information about the gene

click here to view genomic location

Two tasks

- Using GeneView, can you figure out how many different alternatively spliced isoforms exist for BRCA1?
- Using ContigView, can you figure out how to download the genomic sequence for the BRCA1 region?

GeneView shows you information about the transcripts

Transcripts

ENST00000309486	ENSP00000310938	NP_009228.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000337272	ENSP00000337814	BRCA1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000346315	ENSP00000246907	NP_009230.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000351666	ENSP00000338007	NP_009229.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000352993	ENSP00000312236	NP_009234.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000353540	ENSP00000013772	NP_009233.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000354071	ENSP00000326002	NP_009231.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000357654	ENSP00000350283	BRCA1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000359270	ENSP00000352215	NP_009237.1	[Transcript info]	[Exon info]	[Peptide info]
ENST00000360294	ENSP00000353439	NP_009233.1	[Transcript info]	[Exon info]	[Peptide info]

Select region/feature to Export

Choose at least one feature to export. Features must map to the current Ensembl Golden tile path. *Please note we will not export more than 5Mb.*

Region

Chromosome name/fragment: 17

From (type): Base pair 38449840 *

To (type): Base pair 38530994

Context

Bp upstream (to the left):

Bp downstream (to the right):

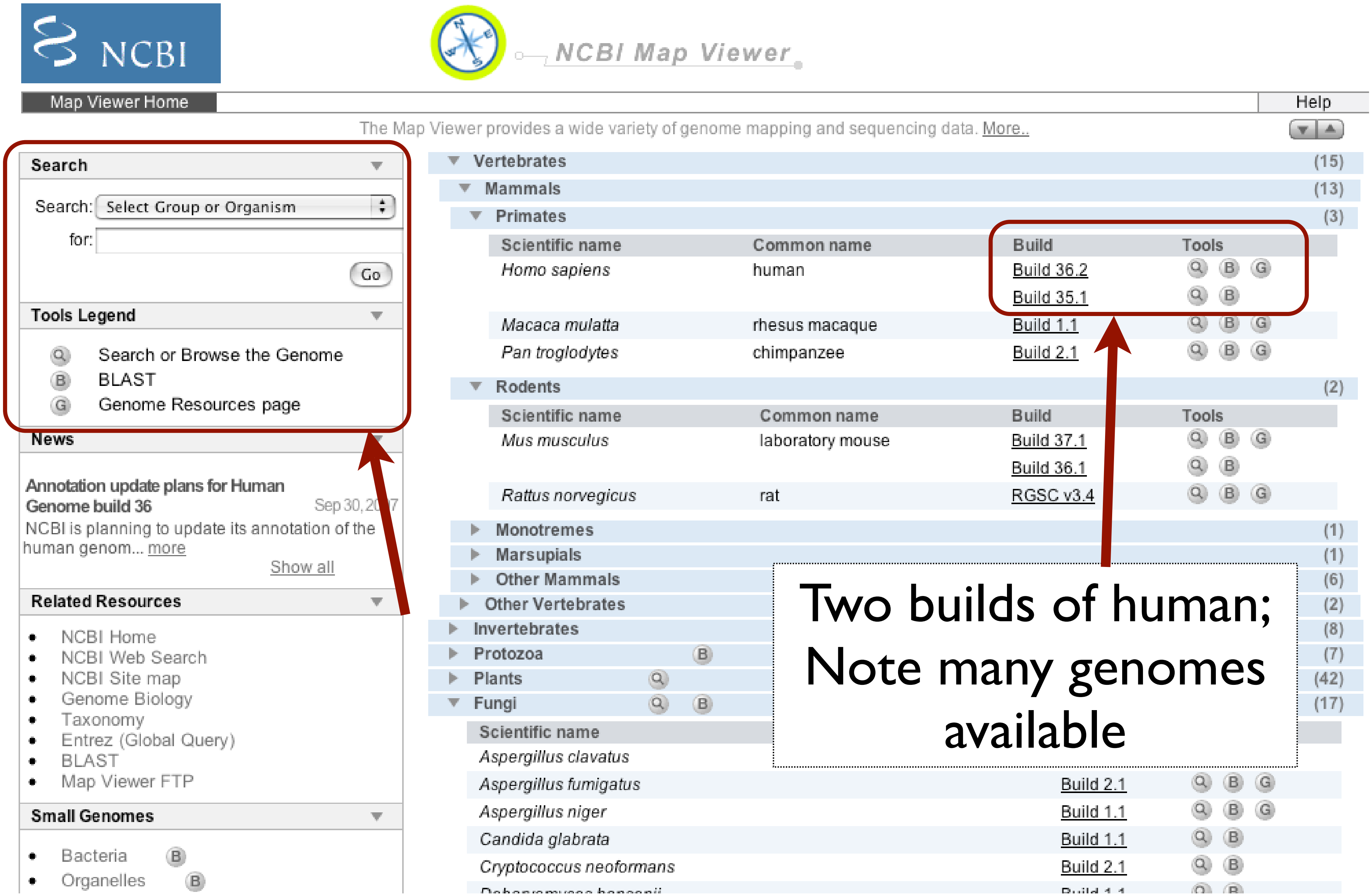
Output

Output: FASTA sequence

Fields marked with * are required

ExportView gives you access to sequence data

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



The Map Viewer provides a wide variety of genome mapping and sequencing data. More..

Search

Search: Select Group or Organism

for:

Go

Tools Legend

🔍 Search or Browse the Genome

B BLAST

G Genome Resources page

News

Annotation update plans for Human Genome build 36 Sep 30, 2007

NCBI is planning to update its annotation of the human genom... more

Show all

Related Resources

- NCBI Home
- NCBI Web Search
- NCBI Site map
- Genome Biology
- Taxonomy
- Entrez (Global Query)
- BLAST
- Map Viewer FTP

Small Genomes

- Bacteria B
- Organelles B

Vertebrates (15)

Mammals (13)

Primates (3)

Scientific name	Common name	Build	Tools
Homo sapiens	human	Build 36.2	🔍 B G
		Build 35.1	🔍 B
Macaca mulatta	rhesus macaque	Build 1.1	🔍 B G
Pan troglodytes	chimpanzee	Build 2.1	🔍 B G

Rodents (2)

Scientific name	Common name	Build	Tools
Mus musculus	laboratory mouse	Build 37.1	🔍 B G
		Build 36.1	🔍 B
Rattus norvegicus	rat	RGSC v3.4	🔍 B G

Monotremes (1)

Marsupials (1)

Other Mammals (6)

Other Vertebrates (2)

Invertebrates (8)

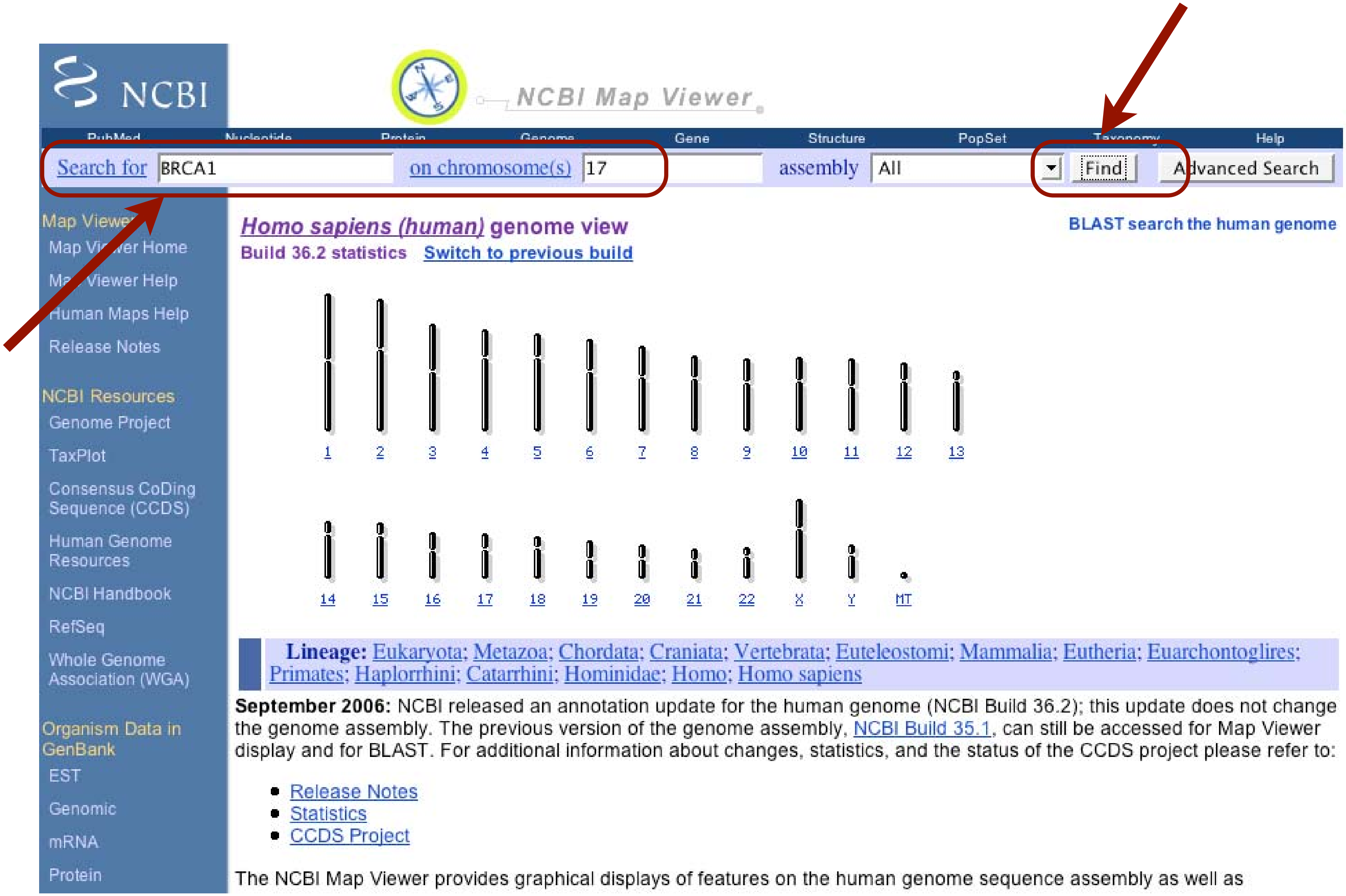
Protozoa B (7)

Plants 🔍 (42)

Fungi 🔍 B (17)

Scientific name	Build	Tools
Aspergillus clavatus		
Aspergillus fumigatus	Build 2.1	🔍 B G
Aspergillus niger	Build 1.1	🔍 B G
Candida glabrata	Build 1.1	🔍 B
Cryptococcus neoformans	Build 2.1	🔍 B

Two builds of human;
Note many genomes available



Search for BRCA1 on chromosome(s) 17 assembly All

Homo sapiens (human) genome view

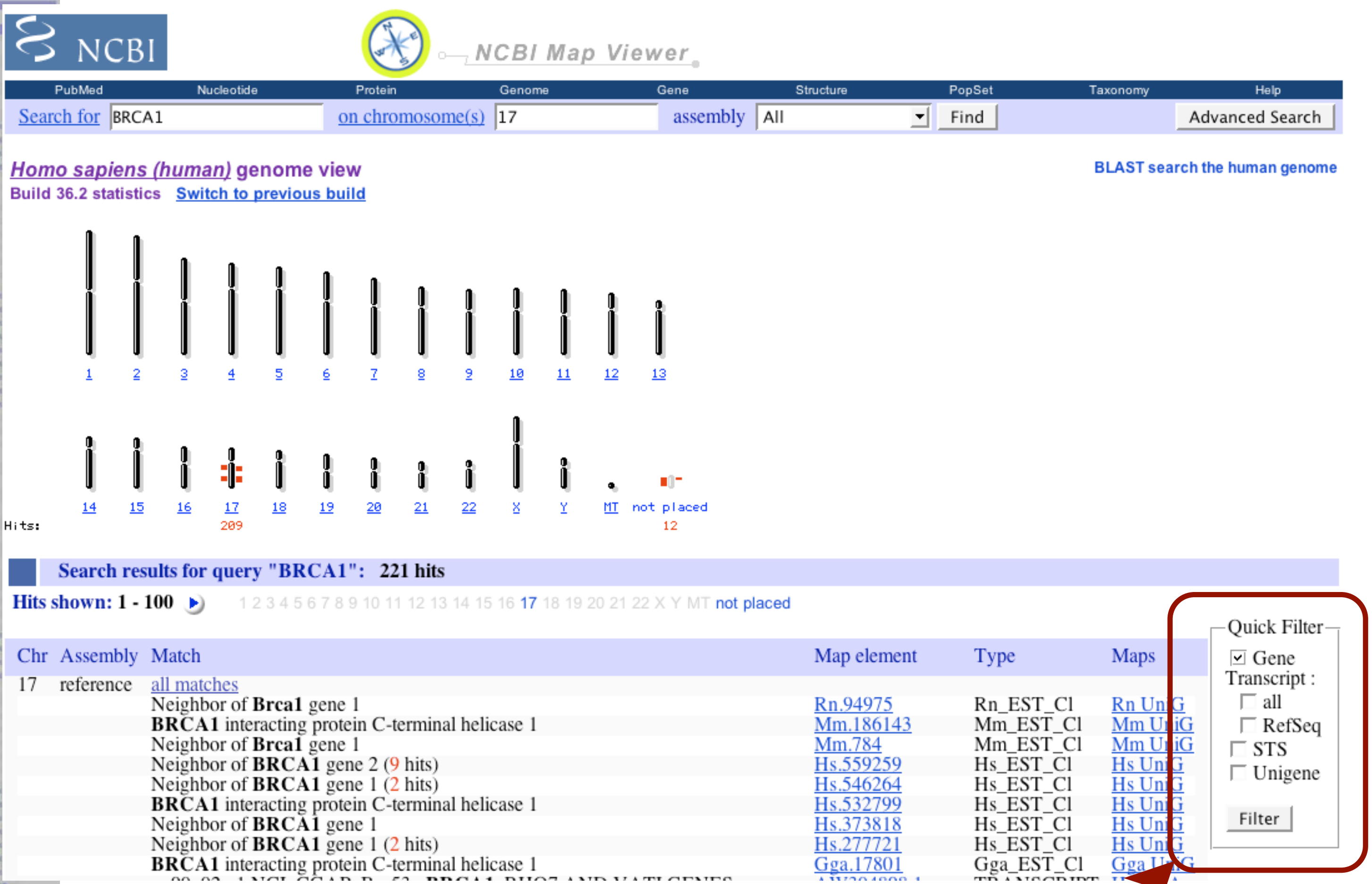
Build 36.2 statistics Switch to previous build

Lineage: Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo; Homo sapiens

September 2006: NCBI released an annotation update for the human genome (NCBI Build 36.2); this update does not change the genome assembly. The previous version of the genome assembly, NCBI Build 35.1, can still be accessed for Map Viewer display and for BLAST. For additional information about changes, statistics, and the status of the CCDS project please refer to:

- Release Notes
- Statistics
- CCDS Project

The NCBI Map Viewer provides graphical displays of features on the human genome sequence assembly as well as



Search for BRCA1 on chromosome(s) 17 assembly All Find Advanced Search

Homo sapiens (human) genome view

Build 36.2 statistics Switch to previous build

BLAST search the human genome

Search results for query "BRCA1": 221 hits

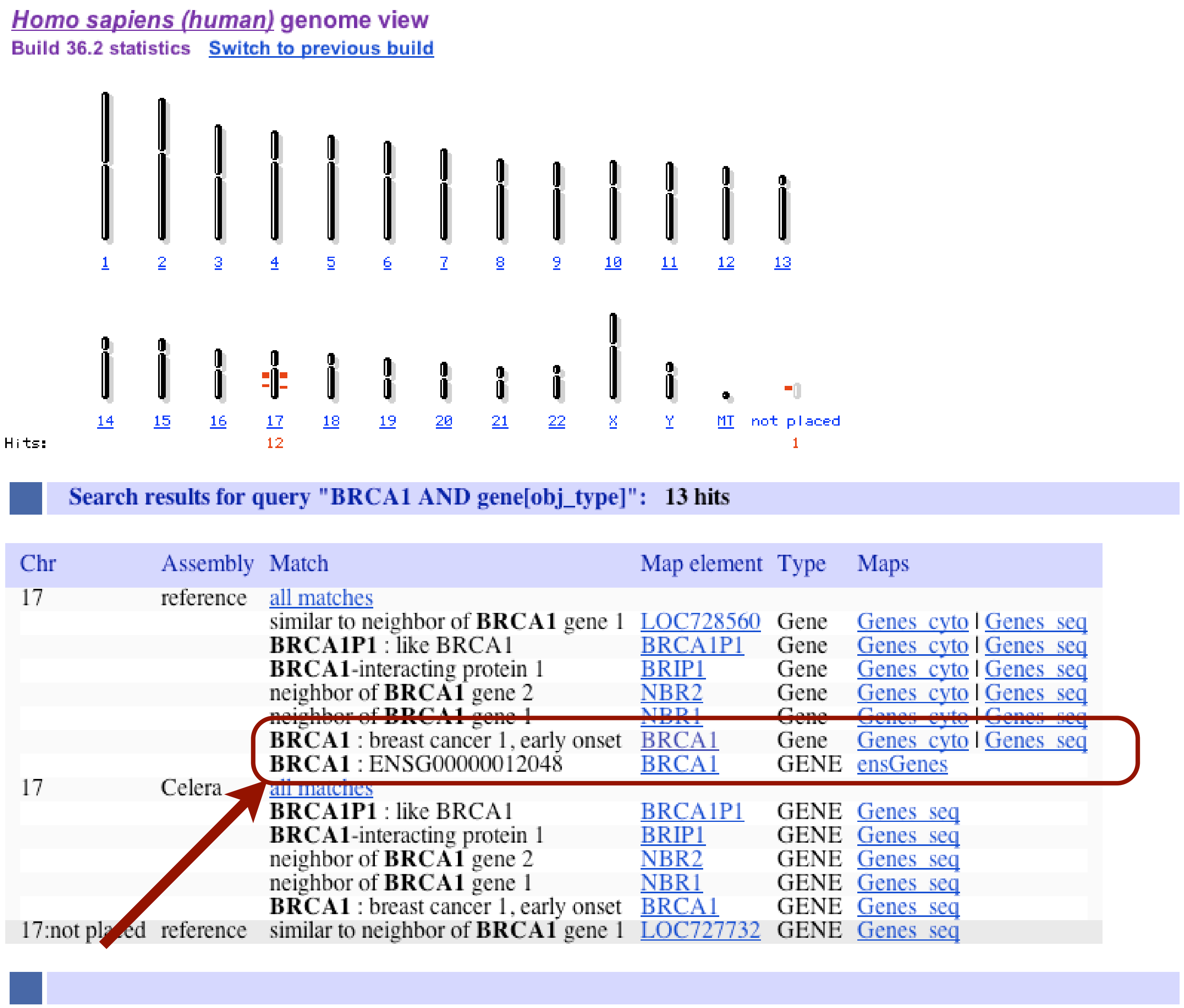
Hits shown: 1 - 100

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y MT not placed

Chr	Assembly	Match	Map element	Type	Maps
17	reference	all matches			
		Neighbor of **Brca1** gene 1	Rn.94975	Rn_EST_Cl	Rn UniG
		BRCA1 interacting protein C-terminal helicase 1	Mm.186143	Mm_EST_Cl	Mm UniG
		Neighbor of **Brca1** gene 1	Mm.784	Mm_EST_Cl	Mm UniG
		Neighbor of **BRCA1** gene 2 (9 hits)	Hs.559259	Hs_EST_Cl	Hs UniG
		Neighbor of **BRCA1** gene 1 (2 hits)	Hs.546264	Hs_EST_Cl	Hs UniG
		BRCA1 interacting protein C-terminal helicase 1	Hs.532799	Hs_EST_Cl	Hs UniG
		Neighbor of **BRCA1** gene 1	Hs.373818	Hs_EST_Cl	Hs UniG
		Neighbor of **BRCA1** gene 1 (2 hits)	Hs.277721	Hs_EST_Cl	Hs UniG
		BRCA1 interacting protein C-terminal helicase 1	Gga.17801	Gga_EST_Cl	Gga UniG

Quick Filter
☑ Gene
Transcript :
☐ all
☐ RefSeq
☐ STS
☐ Unigene
Filter

Quick Filter
✓ Gene

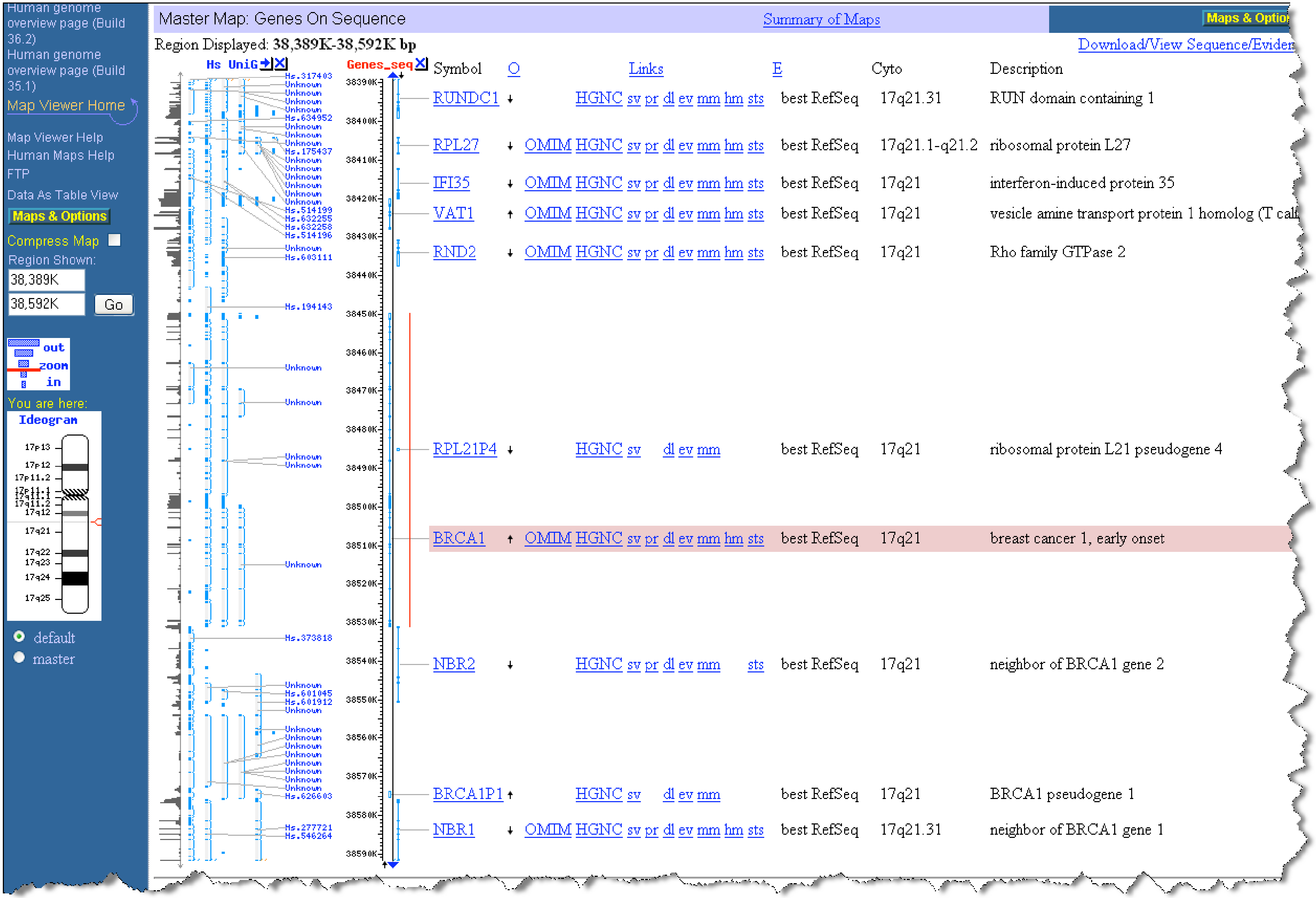


Homo sapiens (human) genome view

Build 36.2 statistics   Switch to previous build

Search results for query "BRCA1 AND gene[obj_type]":   13 hits

Chr	Assembly	Match	Map element	Type	Maps
17	reference	all matches			
		similar to neighbor of **BRCA1** gene 1	LOC728560	Gene	Genes_cyto | Genes_seq
		BRCA1P1 : like BRCA1	BRCA1P1	Gene	Genes_cyto | Genes_seq
		BRCA1-interacting protein 1	BRIP1	Gene	Genes_cyto | Genes_seq
		neighbor of **BRCA1** gene 2	NBR2	Gene	Genes_cyto | Genes_seq
		neighbor of **BRCA1** gene 1	NBR1	Gene	Genes_cyto | Genes_seq
		BRCA1 : breast cancer 1, early onset	BRCA1	Gene	Genes_cyto | Genes_seq
		BRCA1 : ENSG00000012048	BRCA1	GENE	ensGenes
17	Celera	all matches			
		BRCA1P1 : like BRCA1	BRCA1P1	GENE	Genes_seq
		BRCA1-interacting protein 1	BRIP1	GENE	Genes_seq
		neighbor of **BRCA1** gene 2	NBR2	GENE	Genes_seq
		neighbor of **BRCA1** gene 1	NBR1	GENE	Genes_seq
		BRCA1 : breast cancer 1, early onset	BRCA1	GENE	Genes_seq
17:not placed	reference	similar to neighbor of **BRCA1** gene 1	LOC727732	GENE	Genes_seq



Master Map: Genes On Sequence

Summary of Maps

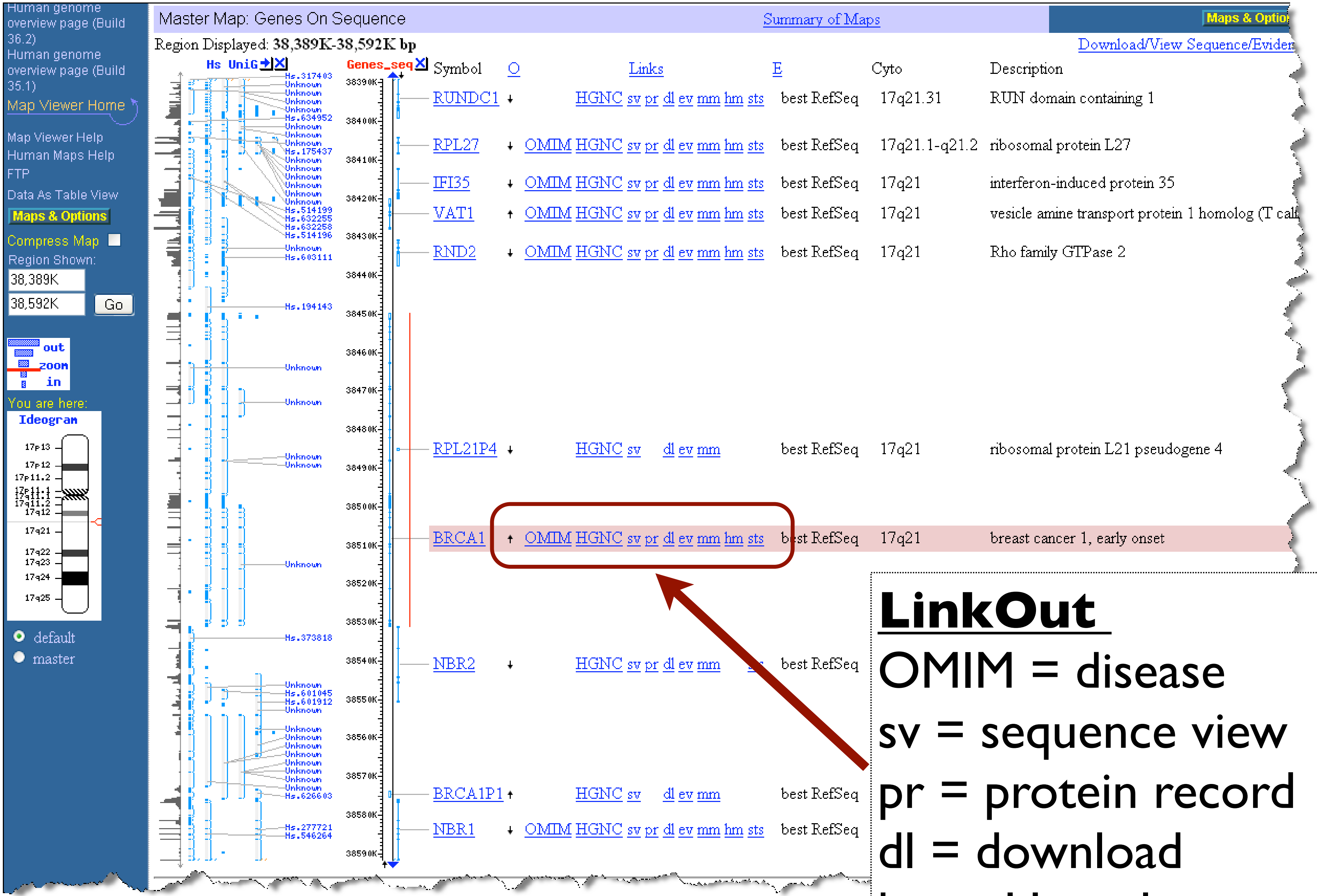
Region Displayed: **38,389K-38,592K bp**

Download/View Sequence/Eviden

Symbol	O	Links	E	Cyto	Description
RUNDC1	↓	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.31	RUN domain containing 1
RPL27	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.1-q21.2	ribosomal protein L27
IFI35	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	interferon-induced protein 35
VAT1	↑	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	vesicle amine transport protein 1 homolog (T cal
RND2	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	Rho family GTPase 2
RPL21P4	↓	HGNC sv dl ev mm	best RefSeq	17q21	ribosomal protein L21 pseudogene 4
BRCA1	↑	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	breast cancer 1, early onset
NBR2	↓	HGNC sv pr dl ev mm sts	best RefSeq	17q21	neighbor of BRCA1 gene 2
BRCA1P1	↑	HGNC sv dl ev mm	best RefSeq	17q21	BRCA1 pseudogene 1
NBR1	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.31	neighbor of BRCA1 gene 1

Two tasks

- Can you figure out how to LinkOut to the OMIM and/or Homologene entries for BRCA1?
- Can you figure out how to download the genomic sequence for the BRCA1 region?



Human genome overview page (Build 36.2)
Human genome overview page (Build 35.1)
Map Viewer Home
Map Viewer Help
Human Maps Help
FTP
Data As Table View
Maps & Options
Compress Map ☐
Region Shown:
38,389K
38,592K
Go

Master Map: Genes On Sequence

Summary of Maps

Maps & Optio

Region Displayed: 38,389K-38,592K bp

Download/View Sequence/Eviden

Symbol	O	Links	E	Cyto	Description
RUNDC1	↓	HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.31	RUN domain containing 1
RPL27	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21.1-q21.2	ribosomal protein L27
IFI35	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	interferon-induced protein 35
VAT1	↑	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	vesicle amine transport protein 1 homolog (T cali
RND2	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	Rho family GTPase 2
RPL21P4	↓	HGNC sv dl ev mm	best RefSeq	17q21	ribosomal protein L21 pseudogene 4
BRCA1	↑	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq	17q21	breast cancer 1, early onset
NBR2	↓	HGNC sv pr dl ev mm	best RefSeq		
BRCA1P1	↑	HGNC sv dl ev mm	best RefSeq		
NBR1	↓	OMIM HGNC sv pr dl ev mm hm sts	best RefSeq		

○ default
○ master

LinkOut

OMIM = disease

sv = sequence view

pr = protein record

dl = download

hm = Homologene

Credits

- UCSC Genome Browser

  http://genome.cse.ucsc.edu/

- Ensembl Genome Browser

  http://www.ensembl.org/index.html

- NCBI MapViewer

  http://www.ncbi.nlm.nih.gov/mapview/index.html