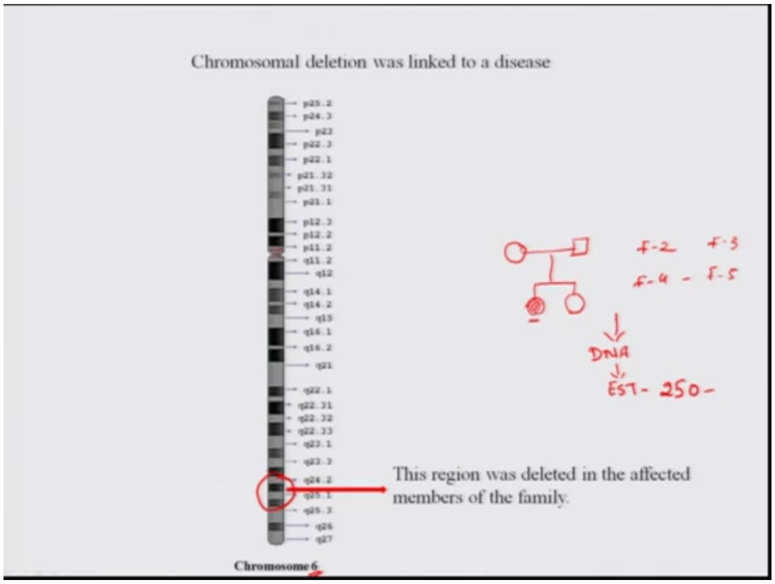
Functional Genomics
Professor S Ganesh
Department of Biological Sciences & Bioengineering
Indian Institute of Technology Kanpur
Lecture No 13
Tutorial Part 1

Hello every one and welcome to the tutorial session of functional genomics myself Anupama and as already introduced I am one of the teaching assistance in this course that is functional genomics today we will be learning about whatever information can be arrived at or fetched using few hundred base pairs of DNA sequence.

(Refer Slide Time: 0:50)



Chromosomal deletion was linked to a disease

This region was deleted in the affected members of the family.

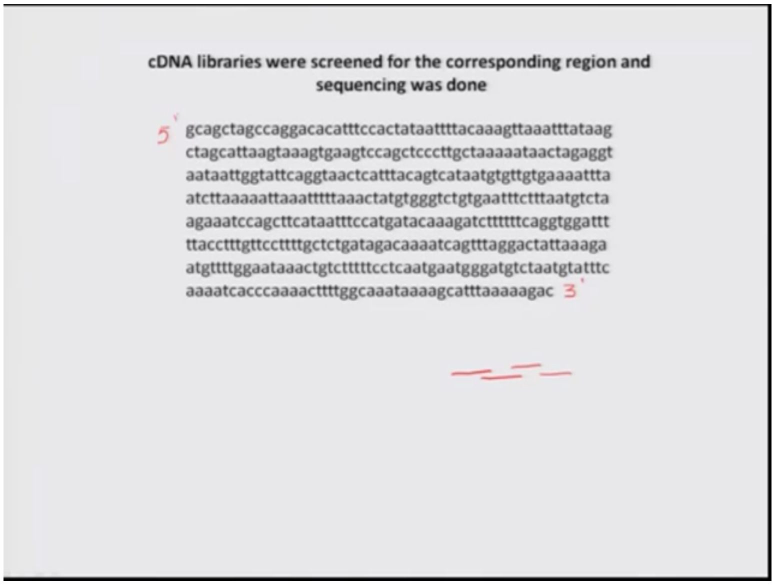
So now let us look into that so this is a part of chromosome which was deleted in a disease and linked to that disease. This was identified by a team of scientists so few of you must be wondering that how did they arrived at the region of the chromosomes. So that has been covered already in one of the NPTEL course that is Human Molecular Genetics and also we are rerunning that course but since you are here very briefly I will like to tell you that how do we arrive at such sequences, so what is done that there are a families which come for counseling to the geneticists and they have members who are affected suppose this female was affected and she has a another sister and their parents were unaffected.

Similarly there was a another family, family 2, family 3 with family 4 so when you once you have a certain amount of family for which you can screen, screen for a cause of the disease. So what is done that DNA is extracted from all the family members which are present and then they are EST markers. Now what are EST markers these are small expressed sequence tag, which are unique for particular region of the chromosome.

These are unique for the regions so what do the scientists do they screen for these EST sequence, so suppose approximately around 250 EST sequence can cover almost all the autosomes of the chromosomes and from there on they link, they try to see that what EST region or what EST is getting segregated with the disease by that I mean that which one are always present in the individuals which are affected.

For example in this female and then the other person of the other family which is affected if that EST marker is present or not. So from that they narrowed down to the region of the chromosome and in this disease the case was the chromosom6 and this region, so from here what using this EST and a small probe they designed and then they screened then they screened a cDNA libraries for that.

(Refer Slide Time: 3:30)

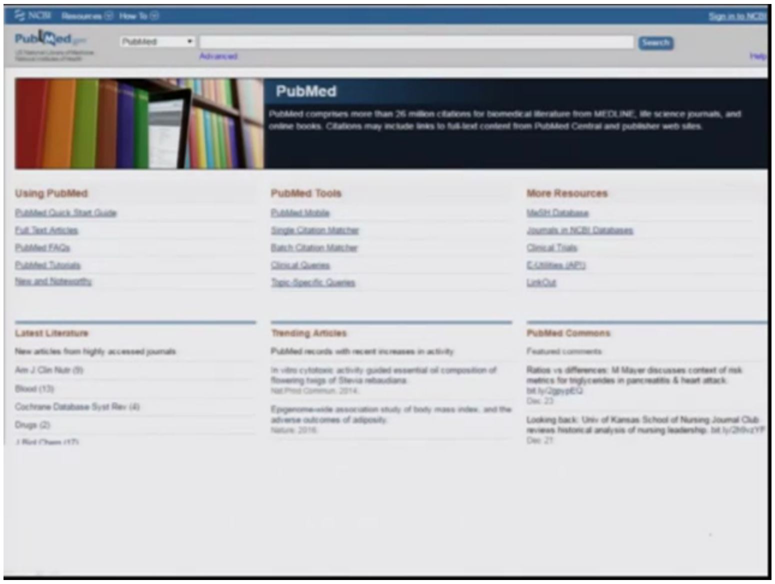


cDNA libraries were screened for the corresponding region and sequencing was done

5' gcagctagccaggacacatttccactataattttacaaagttaaatttataag
ctagcattaagtaaagtgaagtccagctcccttgctaaaaataactagaggt
aataattggtattcaggtaactcatttacagtcataatgtgttgtgaaaattta
atcttaaaaattaaatttttaaactatgtgggtctgtgaatttctttaatgtcta
agaaatccagcttcataatttccatgatacaaagatctttttcaggtggattt
ttacctttgttccttttgctctgatagacaaaatcagtttaggactattaaaga
atgttttggaataaactgtctttttcctcaatgaatgggatgtctaatgtatttc
aaaatcacccaaaacttttggcaaataaaagcatttaaaaagac 3'

They try to align them as they screen for many libraries sequence them as sequencing has already being taught and there are various ways by which you can sequence and then you arrive at a sequence, in this session we will start with the sequence which has been arrived and they have few hundred base pair of the sequence, this is the 5 prime end and this is the 3 prime end of the sequence, so using the current search tools whatever information we can take is we can arrive at or which can be fetched using this few hundred base pair of the sequence that is what we will be learning today.

(Refer Slide Time: 4:14)



PubMed

PubMed comprises more than 26 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Using PubMed

- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities (API)
- LinkOut

Latest Literature

New articles from highly accessed journals

- Am J Clin Nutr (9)
- Blood (13)
- Cochrane Database Syst Rev (4)
- Drugs (2)
- J Biol Chem (17)

Trending Articles

PubMed records with recent increases in activity

In vitro cytotoxic activity guided essential oil composition of flowering twigs of Stevia rebaudiana.
Nat Prod Commun. 2014.

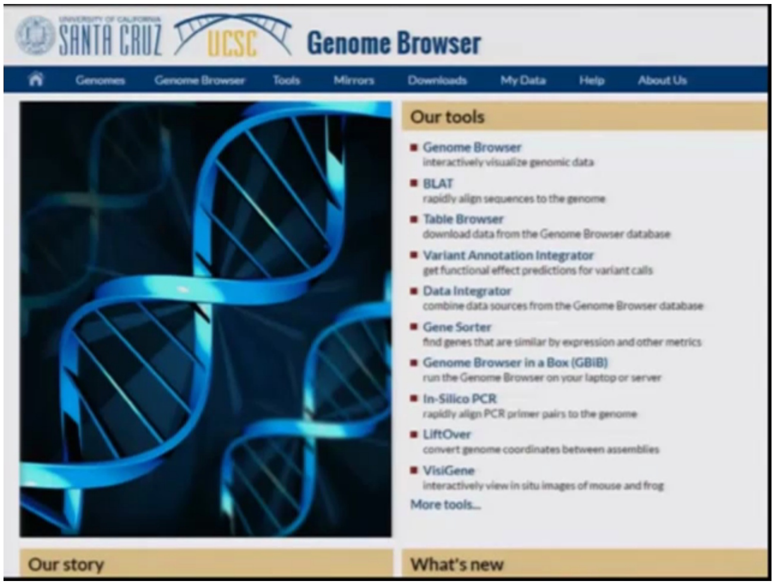
Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity.
Nature. 2016.

PubMed Commons

Featured comments

Ratios vs differences: M Mayer discusses context of risk metrics for triglycerides in pancreatitis & heart attack. bit.ly/2gpypEQ
Dec 23

Looking back: Univ of Kansas School of Nursing Journal Club reviews historical analysis of nursing leadership. bit.ly/2h9vzYF
Dec 21



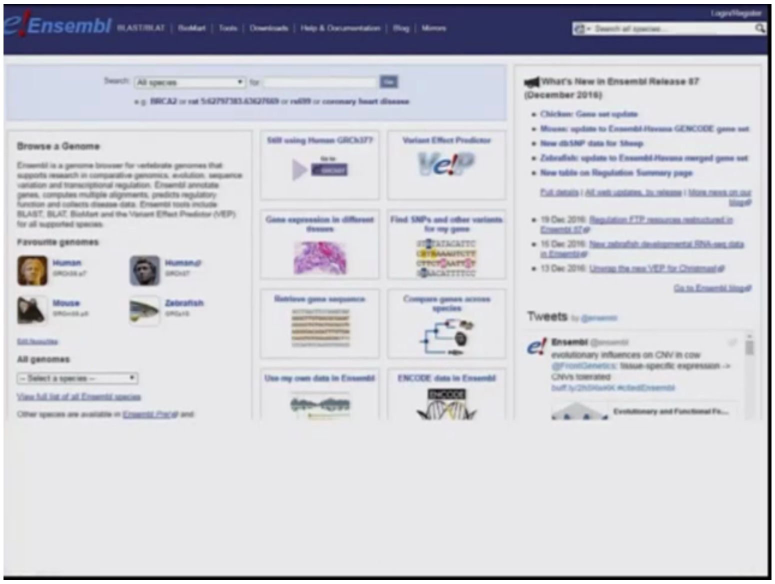
Our tools

- Genome Browser
  interactively visualize genomic data
- BLAT
  rapidly align sequences to the genome
- Table Browser
  download data from the Genome Browser database
- Variant Annotation Integrator
  get functional effect predictions for variant calls
- Data Integrator
  combine data sources from the Genome Browser database
- Gene Sorter
  find genes that are similar by expression and other metrics
- Genome Browser in a Box (GBiB)
  run the Genome Browser on your laptop or server
- In-Silico PCR
  rapidly align PCR primer pairs to the genome
- LiftOver
  convert genome coordinates between assemblies
- VisiGene
  interactively view in situ images of mouse and frog

More tools...

Our story

What's new


Ensembl
BLAST/BLAT | BioMart | Tools | Downloads | Help & Documentation | Blog | Mirrors
Login/Register
Search all species...
Search: All species for Go
Still using Human GRCh37?
Variant Effect Predictor
Gene expression in different tissues
Find SNPs and other variants for my gene
Retrieve gene sequence
Compare genes across species
Use my own data in Ensembl
ENCODE data in Ensembl
Favourite genomes
Human
Human
Mouse
Zebrafish
All genomes
-- Select a species --
Tweets by @ensembl

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

View full list of all Ensembl species

Other species are available in Ensembl Pre! and

What's New in Ensembl Release 87 (December 2016)

- Chicken: Gene set update
- Mouse: update to Ensembl-Havana GENCODE gene set
- New dbSNP data for Sheep
- Zebrafish: update to Ensembl-Havana merged gene set
- New table on Regulation Summary page

Full details | All web updates, by release | More news on our blog

- 19 Dec 2016: Regulation FTP resources restructured in Ensembl 87
- 16 Dec 2016: New zebrafish developmental RNA-seq data in Ensembl
- 13 Dec 2016: Unwrap the new VEP for Christmas!

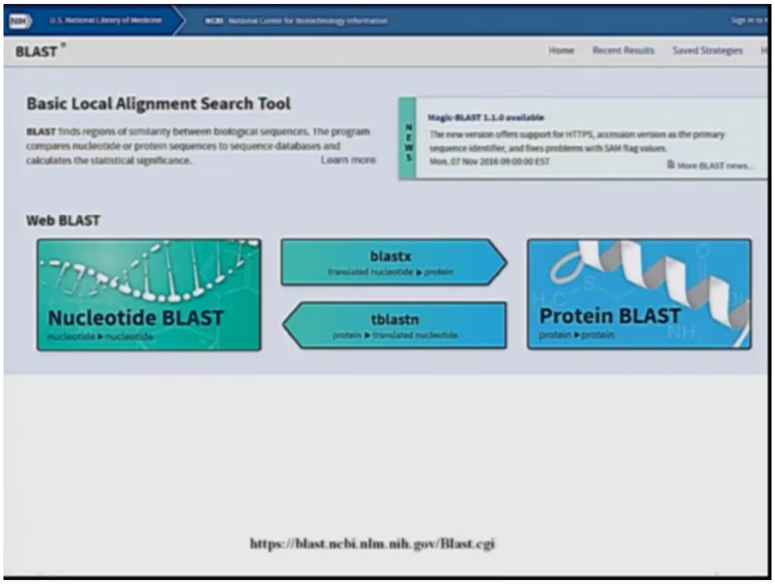
Go to Ensembl blog

Ensembl @ensembl
evolutionary influences on CNV in cow @FrontGenetics: tissue-specific expression -> CNVs tolerated

Evolutionary and Functional Fe...

So using there are few widely used search engines one of those is NCBI another one is UCSC genome browser another one is Ensembl, so today we will be learning about more about NCBI and then we will look online on one of the another search engine that is Ensembl. So since we had a few hundred base pair of sequence and we did not knew from which gene it is or whether this is a coding region or not coding region or this is just a intron of a gene or this a exon of a gene all those information we do not have.

(Refer Slide Time: 5:05)


NIH U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
Sign in to
BLAST
Home
Recent Results
Saved Strategies
NEWS
Web BLAST
Nucleotide BLAST
nucleotide ▶ nucleotide
blastx
translated nucleotide ▶ protein
tblastn
protein ▶ translated nucleotide
Protein BLAST
protein ▶ protein

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

Magic-BLAST 1.1.0 available

The new version offers support for HTTPS, accession version as the primary sequence identifier, and fixes problems with SAM flag values.

Mon, 07 Nov 2016 09:00:00 EST

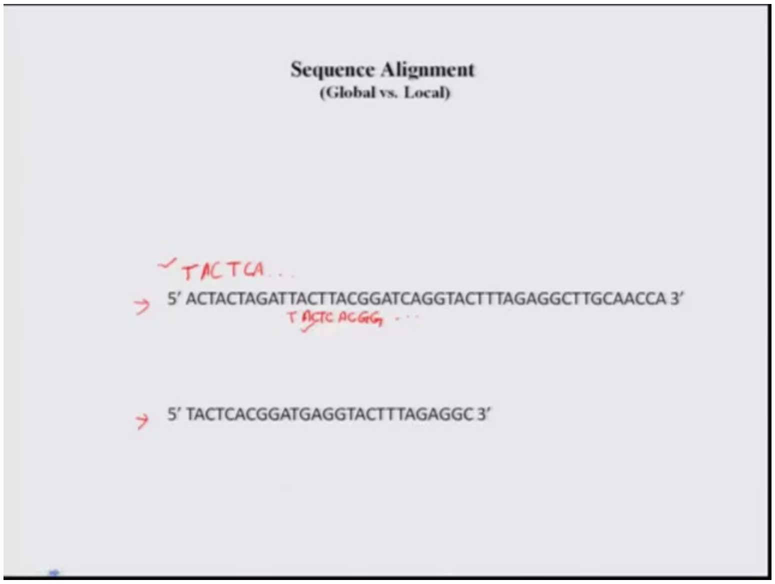
More BLAST news...

https://blast.ncbi.nlm.nih.gov/Blast.cgi

So one such way is to use blast and here is the page of blast that is that is Basic Local Alignment Search Tool. So what does blast does it finds the region of similarity between biological sequences, this programme it compares the nucleotides or protein sequences to sequence data base and calculate the statistical significance, so they are various blast programmes like one is nuclear type blast, the other one is protein blast. Now you have a stretch of a sequence that can be converted into protein and then again it can be blast to the protein data base that is called blastx.

Similarly the protein can be translated to nucleotide and then it can be blast to the other nucleotide data base and that is called tblastn, so as we have seen that we had a DNA sequence so we will go for nucleotide blast and try to look that what region it is aligning to but before we go into what region it is aligning to very briefly we will look into what does sequence alignment means and what are the various strategies used to align the sequence.

(Refer Slide Time: 6:15)



Sequence Alignment
(Global vs. Local)

5′ ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3′

5′ TACTCACGGATGAGGTACTTTAGAGGC 3′

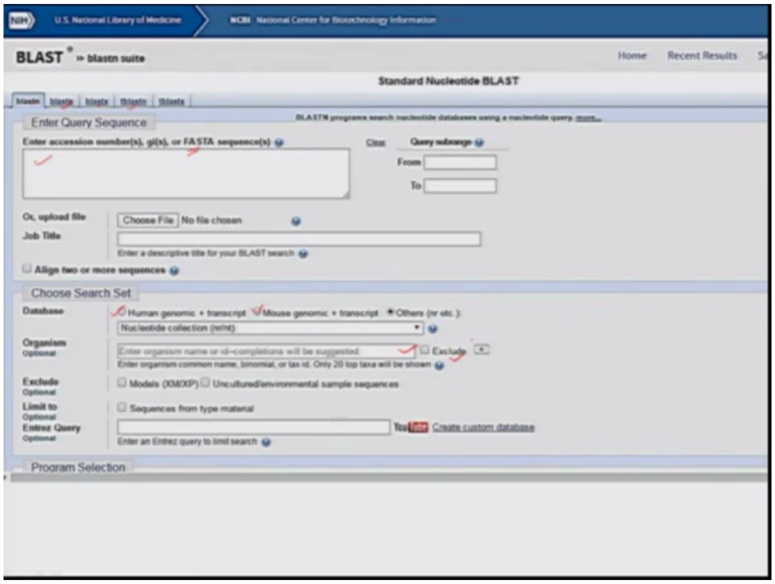
So there are two key terms which you will encounter more often that is global versus local alignment so very briefly what is global and what is local alignment. Let us look into these two sequences now one you have to align these two sequences one is the one this sequence and the other one is this. Now there are two ways by two ways by which you can align this sequence one is that you just go one by one and try and align for example here if I align this one with this so this would be TACTCA and so on.

And you will try and calculate this core which is based on the various algorithms that what is this core of this alignment, now another way is to that you leave you do not go one by one but you will try and look for small stretch which are aligning the best for example here it would be if you align here with this TACTCACGG and so on you will see that the local stretches are aligning more, so this is called global alignment the first method and the other method is called the local alignment.

Similarly if you have to align these two sequences which are almost of same lengths then it is best to align them globally. So when there is a small stretch of sequence its best to align them local and if there are equal length of equal length of DNA sequences then that can be aligned globally. So on these alignments using different algorithms scores are calculated, few of those are like from where you are starting, what is the gap between the first not aligned and the one where the alignment starts there is a gap penalty then in between there are gap penalty where there are mismatches or you skin few bases and then you align to another stretch.

For example suppose you this is one of this sequence this is quite frank 3 prime and then you have another sequence so it aligns here but then there are some mismatches so you skip that and then again you align. So there would be a certain penalty for this and then that penalty would increase at these regions so on basis of the score is calculated and that is also represented in a blast data.

(Refer Slide Time: 9:13)


NIH U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
BLAST ® » blastn suite
Home
Recent Results
Sa
blastn
blastp
blastx
tblastn
tblastx

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

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

Clear

Query subrange

From

To

Or, upload file: Choose File No file chosen

Job Title

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

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

Nucleotide collection (nr/nt)

Organism Optional: Enter organism name or id--completions will be suggested □ Exclude

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

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

Limit to Optional: □ Sequences from type material

Entrez Query Optional: You Tube Create custom database

Enter an Entrez query to limit search

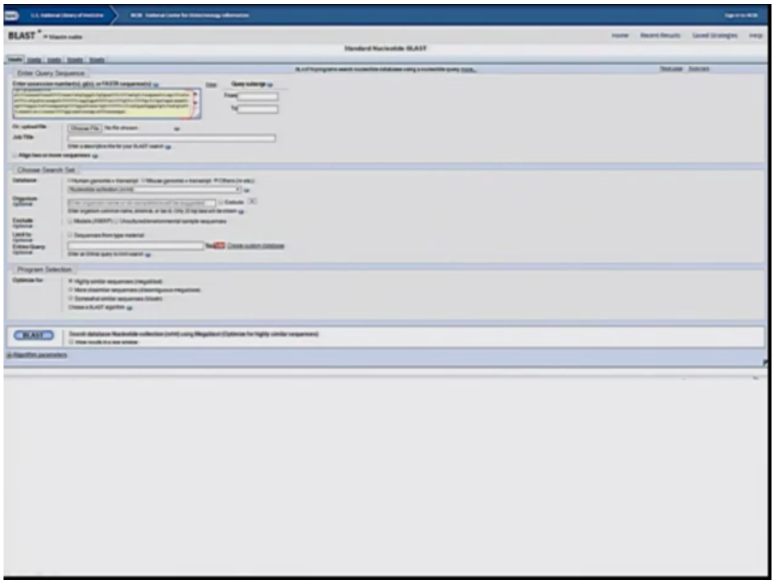
Program Selection

So now let us look into how do we go about it, so what happens like this is the blast home page where you have different ways as I have already described blastn, blastp for protein blastx where nucleotide is converted to the protein and then blast to the whole protein data base.

Similarly the nucleotide is translated and then it is blasted to or aligned to to the sequences in the protein data base and so on, so here you can enter your sequence which has that sequence can be entered in a different formats which has accretion numbers which are given by NCBI and that is at the unique suppose you have a gene for which you want to blast, you just need to put that accession number and the tool would take it or you can paste faster sequence.

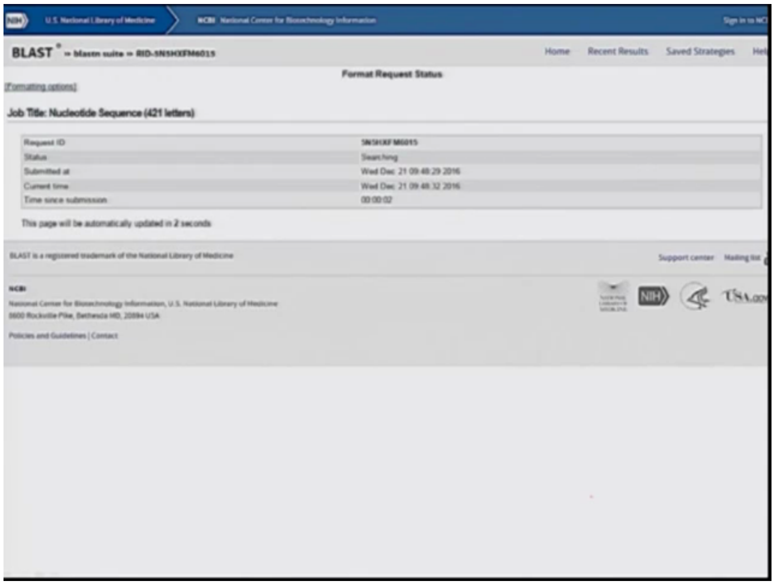
Now faster sequence is a format of writing the nucleotide or the DNA sequences then on the basis of your of your requirement you can make the... You can use a particular data base suppose you just want to use human the genomics or the transcript or mouse genomics transcript or you can choose from other organism or exclude them add them whatever is your criteria you can do that and then you can just click to blast.

(Refer Slide Time: 10:28)



So in our case this was the query sequence which I have pasted here and then you click on blast just this is the default setting. I have not changed anything and you click on blast.

(Refer Slide Time: 10:45)


U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
Sign in to NCBI
BLAST® » blastn suite » RID-5N5HXFM6015
Home
Recent Results
Saved Strategies
Help
[Formatting options]
Support center
Mailing list

Format Request Status

Job Title: Nucleotide Sequence (421 letters)

Request ID	5N5HXFM6015
Status	Searching
Submitted at	Wed Dec 21 09:48:29 2016
Current time	Wed Dec 21 09:48:32 2016
Time since submission	00:00:02

This page will be automatically updated in 2 seconds

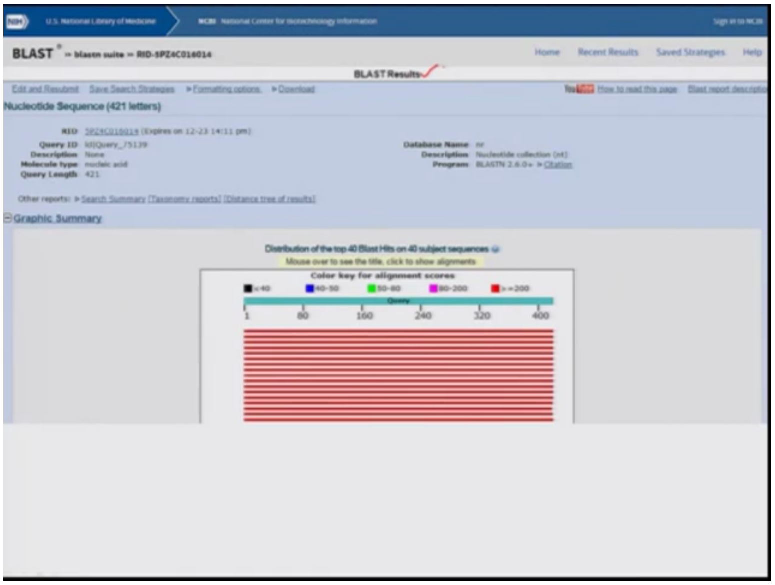
BLAST is a registered trademark of the National Library of Medicine

NCBI
National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA

Policies and Guidelines | Contact

So what do you get the results sometimes the software takes time and doing that and this is how the page it comes. This is the job title and this is your query sequence and this is the time you submitted and how much time duration has passed on.

(Refer Slide Time: 11:07)


U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
Sign in to NCBI
BLAST® » blastn suite » RID-5PZ4C016014
Home
Recent Results
Saved Strategies
Help
BLAST Results
Edit and Resubmit
Save Search Strategies
Formatting options
Download
How to read this page
Blast report description
Graphic Summary
Distribution of the top 40 Blast Hits on 40 subject sequences
Mouse over to see the title, click to show alignments
Color key for alignment scores
<40
40-50
50-80
80-200
>=200
Query
1
80
160
240
320
400

Nucleotide Sequence (421 letters)

RID 5PZ4C016014 (Expires on 12-23 14:11 pm)

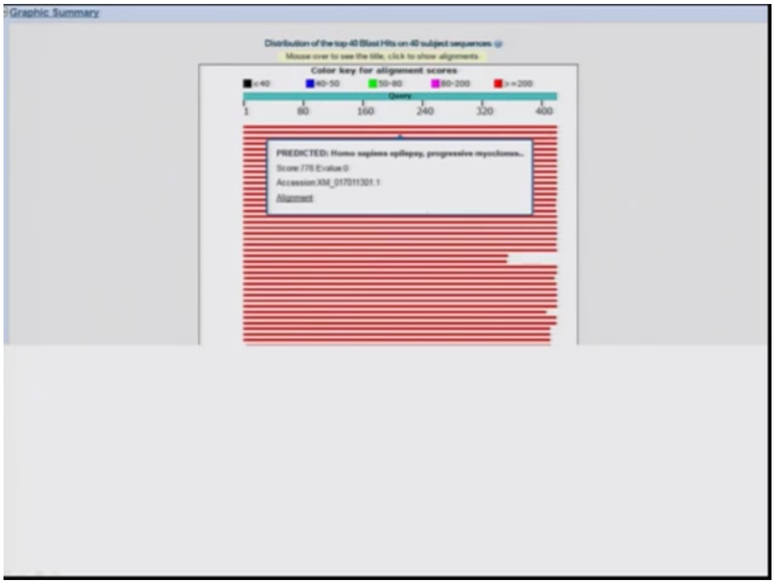
Query ID lcl|Query_75139
Description None
Molecule type nucleic acid
Query Length 421

Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.6.0+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Then you have a result place this is the blast result page so here are the this is the query the description you had put in nucleic acid which was 421 base pair long and this is the data base through which it has done the search and that is non-redundant data base. Non redundant data base then it has the description as the nucleotide collections then this is blastn and then you get a distribution of the top 40 blast on the 40 subject sequence, so these are the 40 sequence subjects, subjects sequences against which these are the best scored ones and the score has been here shown in the format of colour. So red shows the highest one and the black one black shows the lowest one and this is your query sequence which was 421 base pair long.

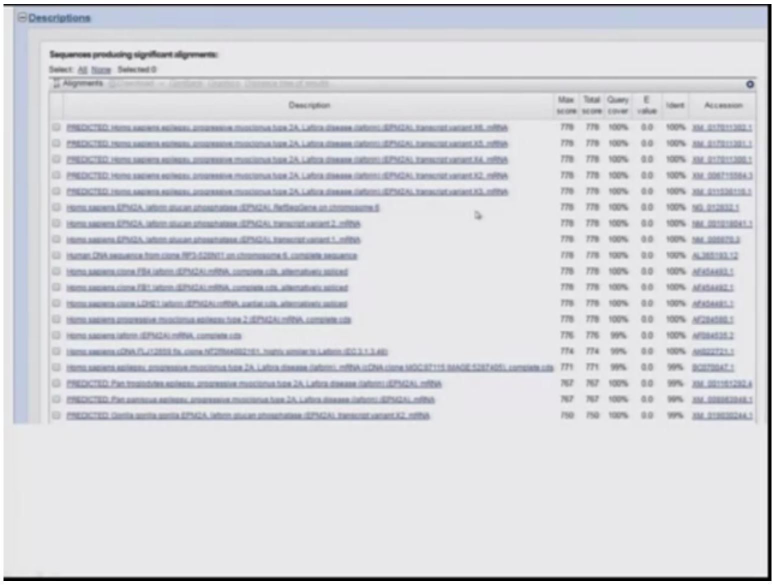
(Refer Slide Time: 12:05)


Graphic Summary
Distribution of the top 40 Blast Hits on 40 subject sequences
Mouse over to see the title, click to show alignments
Color key for alignment scores
■ <40
■ 40-50
■ 50-80
■ 80-200
■ >=200
Query
1
80
160
240
320
400

PREDICTED: Homo sapiens epilepsy, progressive myoclonus..
Score:776 Evalue:0
Accession:XM_017011301.1
Alignment

Now in this you can always get your mouse there and you can find that what that sequence is what is the alignment, what is the accession number and what and what is the E values score which I will be explaining in a short duration.

(Refer Slide Time: 12:24)



Descriptions

Sequences producing significant alignments:

Select: All None Selected 0

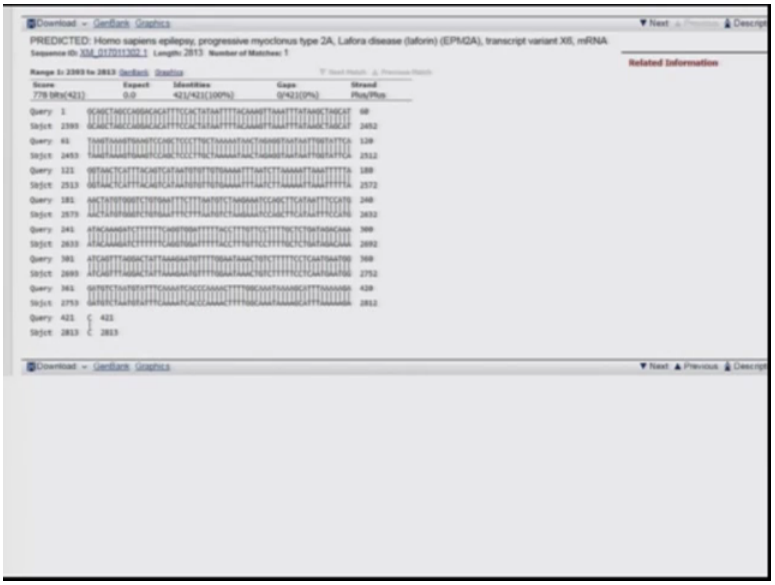
Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Accession
□	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA	778	778	100%	0.0	100%	XM_017011302.1
□	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X5, mRNA	778	778	100%	0.0	100%	XM_017011301.1
□	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X4, mRNA	778	778	100%	0.0	100%	XM_017011300.1
□	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X2, mRNA	778	778	100%	0.0	100%	XM_006715584.3
□	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X3, mRNA	778	778	100%	0.0	100%	XM_011536116.1
□	Homo sapiens EPM2A, laforin glucan phosphatase (EPM2A), RefSeqGene on chromosome 6	778	778	100%	0.0	100%	NG_012832.1
□	Homo sapiens EPM2A, laforin glucan phosphatase (EPM2A), transcript variant 2, mRNA	778	778	100%	0.0	100%	NM_001018041.1
□	Homo sapiens EPM2A, laforin glucan phosphatase (EPM2A), transcript variant 1, mRNA	778	778	100%	0.0	100%	NM_005670.3
□	Human DNA sequence from clone RP3-526N11 on chromosome 6, complete sequence	778	778	100%	0.0	100%	AL365193.12
□	Homo sapiens clone FB4 laforin (EPM2A) mRNA, complete cds, alternatively spliced	778	778	100%	0.0	100%	AF454493.1
□	Homo sapiens clone FB1 laforin (EPM2A) mRNA, complete cds, alternatively spliced	778	778	100%	0.0	100%	AF454492.1
□	Homo sapiens clone LDH21 laforin (EPM2A) mRNA, partial cds, alternatively spliced	778	778	100%	0.0	100%	AF454491.1
□	Homo sapiens progressive myoclonus epilepsy type 2 (EPM2A) mRNA, complete cds	778	778	100%	0.0	100%	AF284580.1
□	Homo sapiens laforin (EPM2A) mRNA, complete cds	776	776	99%	0.0	100%	AF084535.2
□	Homo sapiens cDNA FLJ12659 fis, clone NT2RM4002161, highly similar to Laforin (EC 3.1.3.48)	774	774	99%	0.0	100%	AK022721.1
□	Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin), mRNA (cDNA clone MGC:87115 IMAGE:5267405), complete cds	771	771	99%	0.0	99%	BC070047.1
□	PREDICTED: Pan troglodytes epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), mRNA	767	767	100%	0.0	99%	XM_001161292.4
□	PREDICTED: Pan paniscus epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), mRNA	767	767	100%	0.0	99%	XM_008963848.1
□	PREDICTED: Gorilla gorilla gorilla EPM2A, laforin glucan phosphatase (EPM2A), transcript variant X2, mRNA	750	750	100%	0.0	99%	XM_019030244.1

So this is our description page where we have seen that our query sequence has aligned to these many sequences and there were many more however through the print screen I have got only this one this much. But when you will be doing it online you can scroll down and see that there were many more sequences from various species to which your queries sequence has aligned so what could be the maximum score the total score and how much the query has been covered is given here.

Now what is E value so E value defines the number of hits which you can expect by change, so that here it is zero which means that all the base pairs of your query were aligning hundred percent and it was not a chance event and here you have accession numbers by which you can access this particular transcript which is here it is like EPM2A which is associated with I suppose lafora disease and the protein is laforan and it is a transcript variant X6 mRNA and you can access these by this accession ID.

(Refer Slide Time: 13:48)



PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA
Sequence ID: XM_017011302.1 Length: 2813 Number of Matches: 1

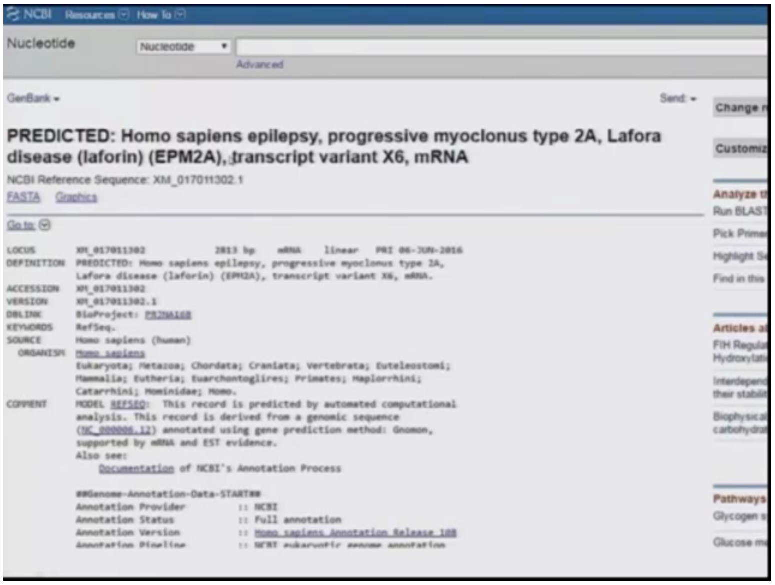
Range 1: 2393 to 2813 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
778 bits(421)	0.0	421/421(100%)	0/421(0%)	Plus/Plus

```
Query  1     GCAGCTAGCCAGGACACATTTCCACTATAATTTTACAAAGTTAAATTTATAAGCTAGCAT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2393  GCAGCTAGCCAGGACACATTTCCACTATAATTTTACAAAGTTAAATTTATAAGCTAGCAT  2452

Query  61    TAAGTAAAGTGAAGTCCAGCTCCCTTGCTAAAAATAACTAGAGGTAATAATTGGTATTCA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2453  TAAGTAAAGTGAAGTCCAGCTCCCTTGCTAAAAATAACTAGAGGTAATAATTGGTATTCA  2512

Query  121   GGTAACTCATTTACAGTCATAATGTGTTGTGAAAATTTAATCTTAAAAATTAAATTTTTA  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2513  GGTAACTCATTTACAGTCATAATGTGTTGTGAAAATTTAATCTTAAAAATTAAATTTTTA  2572

Query  181   AACTATGTGGGTCTGTGAATTTCTTTAATGTCTAAGAAATCCAGCTTCATAATTTCCATG  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2573  AACTATGTGGGTCTGTGAATTTCTTTAATGTCTAAGAAATCCAGCTTCATAATTTCCATG  2632

Query  241   ATACAAAGATCTTTTTTCAGGTGGATTTTTACCTTTGTTCCTTTTGCTCTGATAGACAAA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2633  ATACAAAGATCTTTTTTCAGGTGGATTTTTACCTTTGTTCCTTTTGCTCTGATAGACAAA  2692

Query  301   ATCAGTTTAGGACTATTAAAGAATGTTTTGGAATAAACTGTCTTTTTCCTCAATGAATGG  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2693  ATCAGTTTAGGACTATTAAAGAATGTTTTGGAATAAACTGTCTTTTTCCTCAATGAATGG  2752

Query  361   GATGTCTAATGTATTTCAAAATCACCCAAAACTTTTGGCAAATAAAAGCATTTAAAAAGA  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2753  GATGTCTAATGTATTTCAAAATCACCCAAAACTTTTGGCAAATAAAAGCATTTAAAAAGA  2812

Query  421   C  421
             |
Sbjct  2813  C  2813
```

So once you have that this accession ID then you can also look into how your query sequence was aligning to the subject sequence or to this sequence which the search engine has searched for. So here you see that your this is the query and this is the subject sequence. Your query first nucleotide is aligning with with the 2393 or 2393 base pair of the subject sequence and from there on it is aligning to till 2813 base pair.

(Refer Slide Time: 14:28)



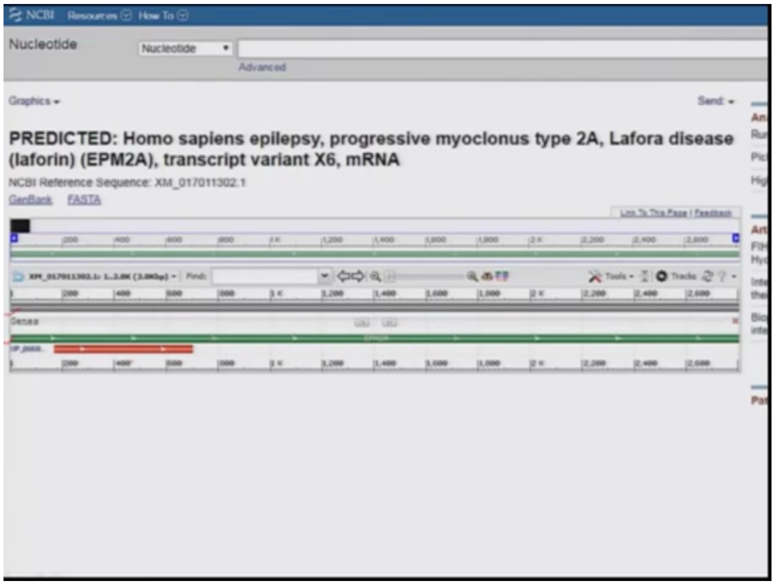
PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA

NCBI Reference Sequence: XM_017011302.1

```
LOCUS       XM_017011302            2813 bp    mRNA    linear   PRI 06-JUN-2016
DEFINITION  PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A,
            Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA.
ACCESSION   XM_017011302
VERSION     XM_017011302.1
DBLINK      BioProject: PRJNA168
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
COMMENT     MODEL REFSEQ:  This record is predicted by automated computational
            analysis. This record is derived from a genomic sequence
            (NC_000006.12) annotated using gene prediction method: Gnomon,
            supported by mRNA and EST evidence.
            Also see:
                Documentation of NCBI's Annotation Process

            ##Genome-Annotation-Data-START##
            Annotation Provider         :: NCBI
            Annotation Status           :: Full annotation
            Annotation Version          :: Homo sapiens Annotation Release 108
            Annotation Pipeline         :: NCBI eukaryotic genome annotation
```

Now from here you can access the sequence what do you find that this the first hit which the search engine has given, it is a protein EPM2A or laforan EPM2A or Laforan and it is associated with a progressive myoclonus type of epilepsy. The below you have all the information about this entry in whatever organisms it is found it is found in the (())(14:49) cod data then there are commons about it. What genome annotation data it has been provided for example from NCBI the annotation status and then as you will scroll down you will get lots of information about that.

(Refer Slide Time: 15:23)

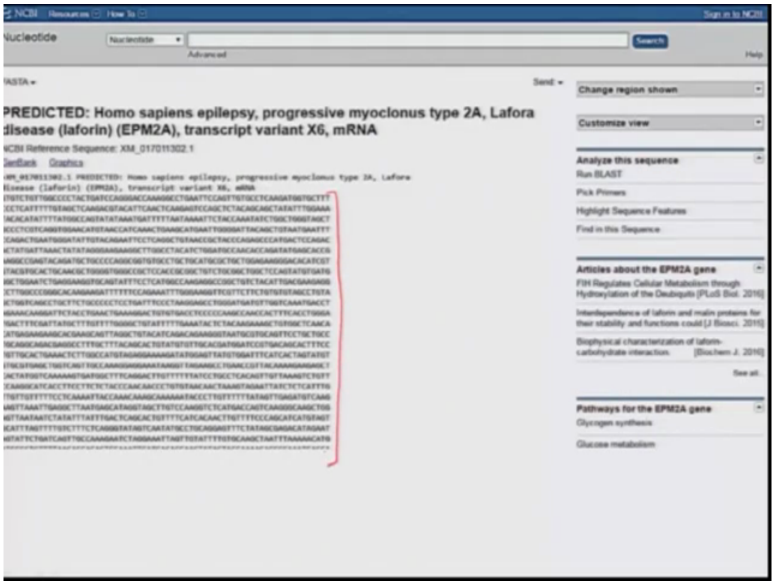


PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA

NCBI Reference Sequence: XM_017011302.1

If you click on the graphics of this you can see the alignment in the form of a graph here this is your query sequence and this is EPM2A and how they are aligning. You can always explore going to these arrows or you can use the tools and tracks and whatever you can explore all this online.

(Refer Slide Time: 15:41)



PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA

NCBI Reference Sequence: XM_017011302.1

```
>XM_017011302.1 PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora
disease (laforin) (EPM2A), transcript variant X6, mRNA
[illegible]
```

Analyze this sequence

Articles about the EPM2A gene

FIH Regulates Cellular Metabolism through Hydroxylation of the Deubiquiti [PLoS Biol. 2016]

Interdependence of laforin and malin proteins for their stability and functions could [J Biosci. 2015]

Biophysical characterization of laforin-carbohydrate interaction. [Biochem J. 2016]

Pathways for the EPM2A gene

Glycogen synthesis

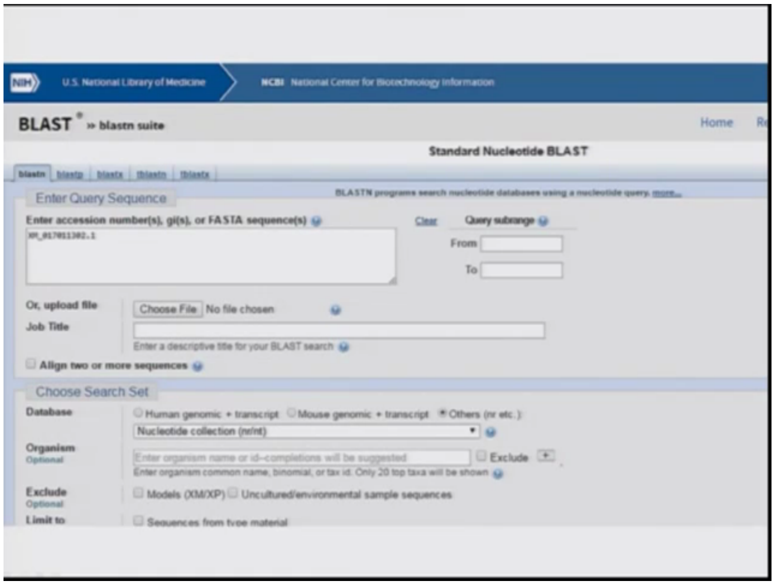
Glucose metabolism

Similarly then there you can go for the FASTA sequence and you will find the sequence of the gene to which your query has aligned at the most, from here as you can see on the right hand side there are various ways to analyse these sequences for example you can run a blast in which this now this sequence of the subject which you have arrived at that is EPM2A gene or lafora can be used to used to run a blast against all the nucleotide sequence.

Similarly you can use primers to amplify this particular gene you can highlight the sequence speeches what are the characteristics sequence speeches and you can also find something in this sequence suppose you are looking for a particular stretch of a sequence you can use this tool to identify, similarly there are articles about EPM2A gene as you are trying to find out the gene for the linked disease so you have come so far so you can always look into what are the recent articles related to with gene, what are the pathways in which this genes is involved like for example glycogen , synthesis or glucose metabolism.

You can always from here link to the phenotype of the disease is this are theses pathways defective in those patience in which this chromosomal region is deleted. Do you see something abrupt with glycogen metabolism in those patients? So with this you can link the information in the tool.

(Refer Slide Time: 17:20)


NIH
U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
BLAST ® » blastn suite
Home
blastn
blastp
blastx
tblastn
tblastx
Choose File

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s): XM_017011302.1

Clear

Query subrange

From

To

Or, upload file: No file chosen

Job Title

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

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

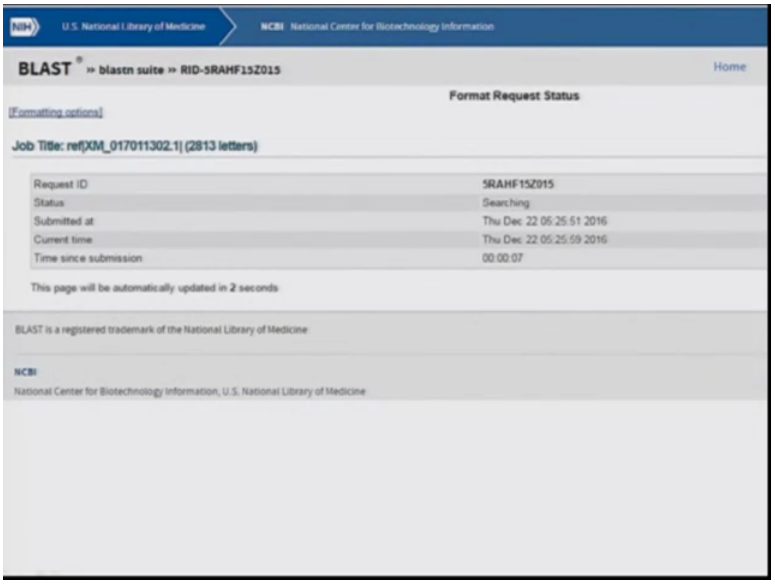
Nucleotide collection (nr/nt)

Organism Optional: Enter organism name or id--completions will be suggested □ Exclude

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

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

Limit to: □ Sequences from type material


NIH
U.S. National Library of Medicine
NCBI National Center for Biotechnology Information
BLAST ® » blastn suite » RID-5RAHF15Z015
Home
[Formatting options]

Format Request Status

Job Title: ref|XM_017011302.1| (2813 letters)

Request ID	5RAHF15Z015
Status	Searching
Submitted at	Thu Dec 22 05:25:51 2016
Current time	Thu Dec 22 05:25:59 2016
Time since submission	00:00:07

This page will be automatically updated in 2 seconds

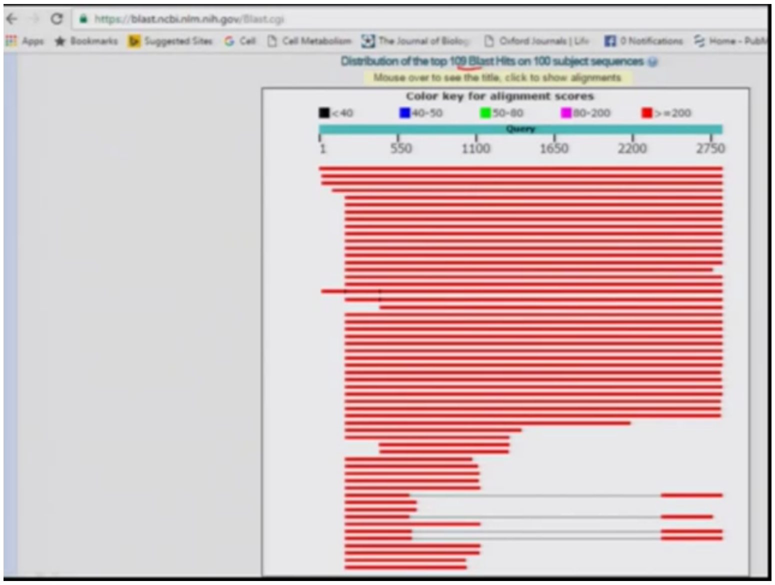
BLAST is a registered trademark of the National Library of Medicine

NCBI

National Center for Biotechnology Information, U.S. National Library of Medicine

When you run blast using that query now again this is just taken the accession number. As I have already described that anything you can use as a query sequence in the NCBI blast and then you can again using the default as I have done here you can run blast again the search engine takes time and thus the page appears on your screen and then this is the way the result is again.

(Refer Slide Time: 17:54)

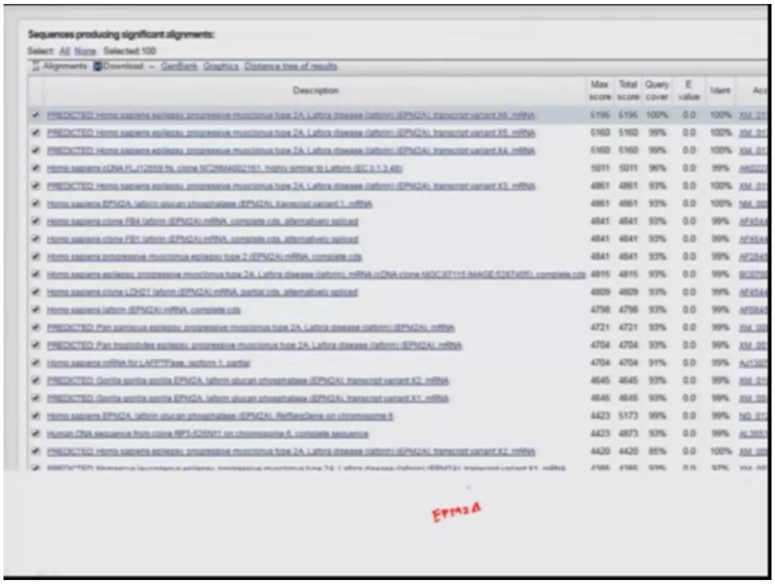

https://blast.ncbi.nlm.nih.gov/Blast.cgi

Distribution of the top 109 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

So this is the similar result as I have described so here these are top 109 blast hits on the hundred subject sequences and this is how it is aligning mostly are red almost all are red and these are the stretch of sequences now these are the stretch of sequences which were not aligning and the last part was aligning.

(Refer Slide Time: 18:14)


Sequences producing significant alignments:

Select: All None Selected: 100

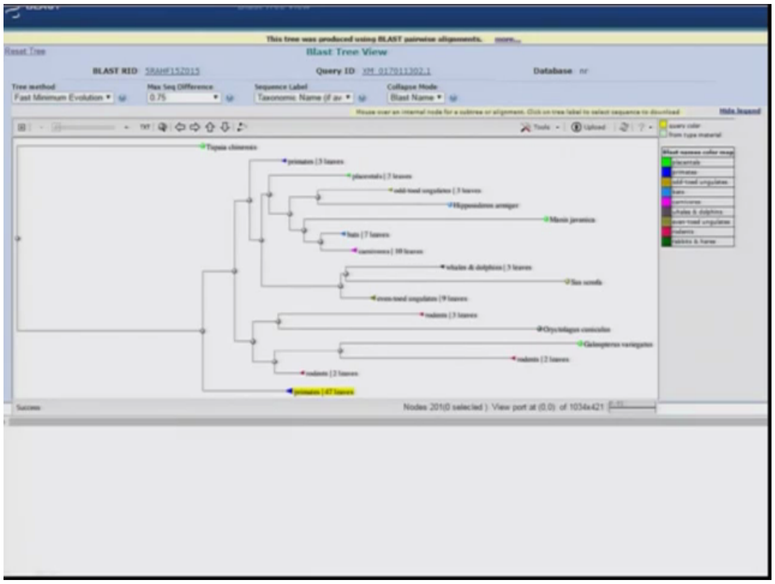
Alignments Download GenBank Graphics Distance tree of results

	Description	Max score	Total score	Query cover	E value	Ident	Acc
✓	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X6, mRNA	5196	5196	100%	0.0	100%	XM_01
✓	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X5, mRNA	5160	5160	99%	0.0	100%	XM_01
✓	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X4, mRNA	5160	5160	99%	0.0	100%	XM_01
✓	Homo sapiens cDNA FLJ12659 fis, clone NT2RM4002161, highly similar to Laforin (EC 3.1.3.48)	5011	5011	96%	0.0	99%	AK022
✓	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X3, mRNA	4861	4861	93%	0.0	100%	XM_01
✓	Homo sapiens EPM2A, laforin glucan phosphatase (EPM2A), transcript variant 1, mRNA	4861	4861	93%	0.0	100%	NM_00
✓	Homo sapiens clone FB4 laforin (EPM2A) mRNA, complete cds, alternatively spliced	4841	4841	93%	0.0	99%	AF454
✓	Homo sapiens clone FB1 laforin (EPM2A) mRNA, complete cds, alternatively spliced	4841	4841	93%	0.0	99%	AF454
✓	Homo sapiens progressive myoclonus epilepsy type 2 (EPM2A) mRNA, complete cds	4841	4841	93%	0.0	99%	AF284
✓	Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin), mRNA (cDNA clone MGC:87115 IMAGE:5287405), complete cds	4815	4815	93%	0.0	99%	BC070
✓	Homo sapiens clone LDH21 laforin (EPM2A) mRNA, partial cds, alternatively spliced	4809	4809	93%	0.0	99%	AF454
✓	Homo sapiens laforin (EPM2A) mRNA, complete cds	4798	4798	93%	0.0	99%	AF084
✓	PREDICTED: Pan paniscus epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), mRNA	4721	4721	93%	0.0	99%	XM_00
✓	PREDICTED: Pan troglodytes epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), mRNA	4704	4704	93%	0.0	99%	XM_00
✓	Homo sapiens mRNA for LAFPTPase, isoform 1, partial	4704	4704	91%	0.0	99%	AJ130
✓	PREDICTED: Gorilla gorilla gorilla EPM2A, laforin glucan phosphatase (EPM2A), transcript variant X2, mRNA	4645	4645	93%	0.0	99%	XM_01
✓	PREDICTED: Gorilla gorilla gorilla EPM2A, laforin glucan phosphatase (EPM2A), transcript variant X1, mRNA	4646	4646	93%	0.0	99%	XM_00
✓	Homo sapiens EPM2A, laforin glucan phosphatase (EPM2A), RefSeqGene on chromosome 6	4423	5173	99%	0.0	99%	NG_01
✓	Human DNA sequence from clone RP3-526N11 on chromosome 6, complete sequence	4423	4873	93%	0.0	99%	AL355
✓	PREDICTED: Homo sapiens epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X2, mRNA	4420	4420	85%	0.0	100%	XM_00
✓	PREDICTED: Nomascus leucogenys epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) (EPM2A), transcript variant X1, mRNA	4386	4386	93%	0.0	97%	XM_00

EPM2A

So once you have the sequence or the all the sequence which are aligning to that gene that was EPM2A, EPM2A gene and these are the genes which are now aligning to that gene. You can select all clicking on all and then you can go for distance tree results so from here you can get the phylogenetic tree of the gene of your interest.

(Refer Slide Time: 18:46)



This tree was produced using BLAST pairwise alignments. more...

Blast Tree View

BLAST RID 5RAHF15Z015 Query ID XM_017011302.1 Database nr

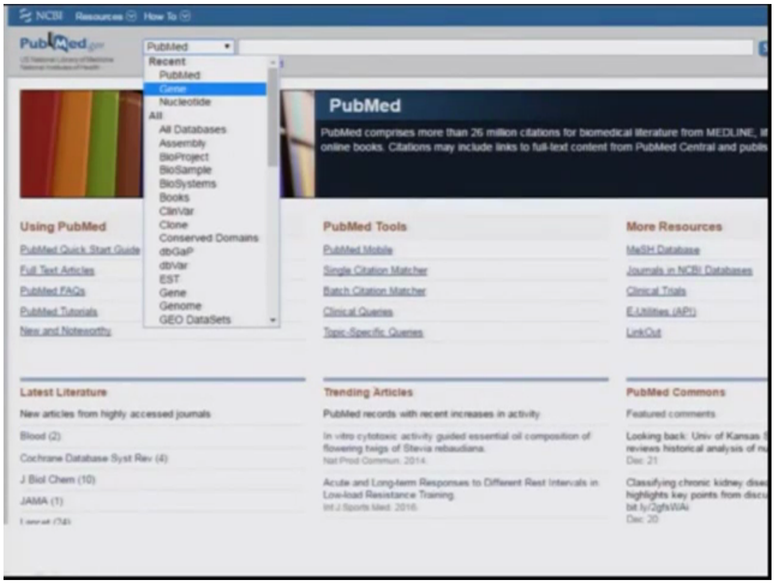
Tree method: Fast Minimum Evolution
Max Seq Difference: 0.75
Sequence Label: Taxonomic Name (if av
Collapse Mode: Blast Name

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Success Nodes 201(0 selected) View port at (0,0) of 1034x421

So this tree shows how many how the gene has evolved and how close to one or the other organisms genes is how it is. So for example primates are more closer to rodents and so on, so once you have the sequence.

(Refer Slide Time: 19:03)



PubMed

PubMed comprises more than 26 million citations for biomedical literature from MEDLINE, li
online books. Citations may include links to full-text content from PubMed Central and publis

Using PubMed
- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools
- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources
- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities (API)
- LinkOut

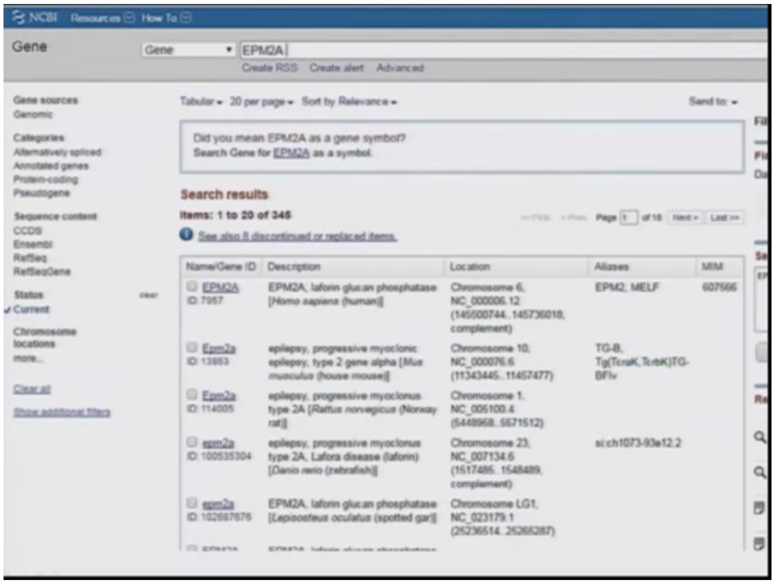
Latest Literature
New articles from highly accessed journals
- Blood (2)
- Cochrane Database Syst Rev (4)
- J Biol Chem (10)
- JAMA (1)
- Lancet (24)

Trending Articles
PubMed records with recent increases in activity
- In vitro cytotoxic activity guided essential oil composition of flowering twigs of Stevia rebaudiana.
  Nat Prod Commun. 2014.
- Acute and Long-term Responses to Different Rest Intervals in Low-load Resistance Training.
  Int J Sports Med. 2016.

PubMed Commons
Featured comments
- Looking back: Univ of Kansas $
  reviews historical analysis of nu
  Dec 21
- Classifying chronic kidney dise
  highlights key points from discu
  bit.ly/2gfsWAi
  Dec 20

So you can also access to the protein of the sequence you can go to PubMed and you can go on gene.

(Refer Slide Time: 19:16)



Gene: EPM2A

Did you mean EPM2A as a gene symbol?
Search Gene for EPM2A as a symbol.

Search results

Items: 1 to 20 of 345

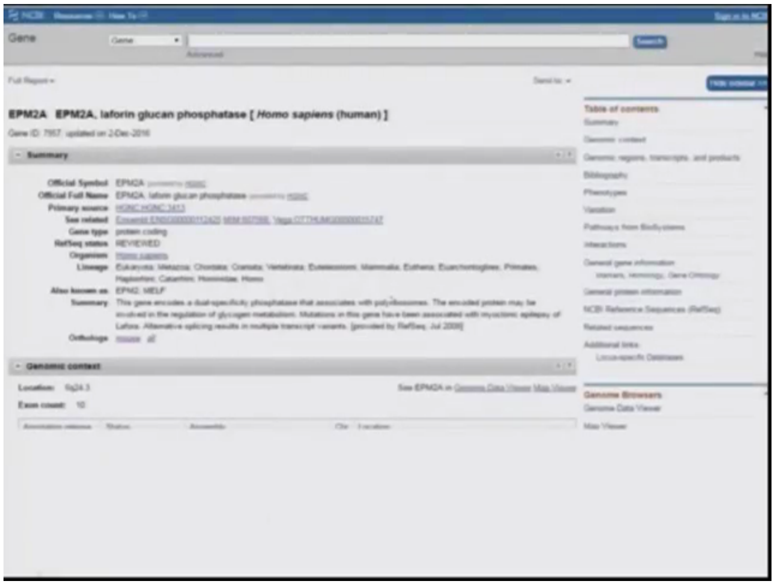
Page 1 of 18

See also 8 discontinued or replaced items.

Name/Gene ID	Description	Location	Aliases	MIM
□ EPM2A ID: 7957	EPM2A, laforin glucan phosphatase [*Homo sapiens* (human)]	Chromosome 6, NC_000006.12 (145500744..145736018, complement)	EPM2, MELF	607566
□ Epm2a ID: 13853	epilepsy, progressive myoclonic epilepsy, type 2 gene alpha [*Mus musculus* (house mouse)]	Chromosome 10, NC_000076.6 (11343445..11457477)	TG-B, Tg(TcraK,TcrbK)TG-BFlv	
□ Epm2a ID: 114005	epilepsy, progressive myoclonus type 2A [*Rattus norvegicus* (Norway rat)]	Chromosome 1, NC_005100.4 (5448968..5671512)		
□ epm2a ID: 100535304	epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) [*Danio rerio* (zebrafish)]	Chromosome 23, NC_007134.6 (1517485..1548489, complement)	si:ch1073-93e12.2	
□ epm2a ID: 102687676	EPM2A, laforin glucan phosphatase [*Lepisosteus oculatus* (spotted gar)]	Chromosome LG1, NC_023179.1 (25236514..25266287)		

From here you will write the gene which you have identified that is EPM2A and then you can select whatever organisms organisms you want for example you want the gene which is there sequenced in the humans or in the mouse or in the rat here we our query was about a sequence which we have identified from a part of a chromosome which was deleted in the family of of the affected people.

(Refer Slide Time: 19:45)



Gene

EPM2A EPM2A, laforin glucan phosphatase [*Homo sapiens* (human)]

Gene ID: 7957, updated on 2-Dec-2016

Summary

Official Symbol	EPM2A provided by HGNC
Official Full Name	EPM2A, laforin glucan phosphatase provided by HGNC
Primary source	HGNC:HGNC:3413
See related	Ensembl:ENSG00000112425 MIM:607566; Vega:OTTHUMG00000015747
Gene type	protein coding
RefSeq status	REVIEWED
Organism	Homo sapiens
Lineage	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo
Also known as	EPM2; MELF
Summary	This gene encodes a dual-specificity phosphatase that associates with polyribosomes. The encoded protein may be involved in the regulation of glycogen metabolism. Mutations in this gene have been associated with myoclonic epilepsy of Lafora. Alternative splicing results in multiple transcript variants. [provided by RefSeq, Jul 2008]
Orthologs	mouse all

Genomic context

Location: 6q24.3

Exon count: 10

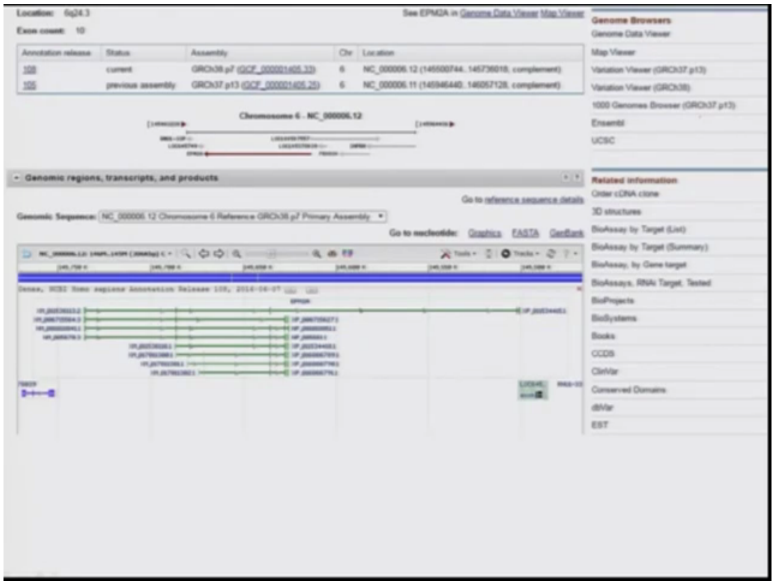
See EPM2A in Genome Data Viewer Map Viewer

Table of contents
Summary
Genomic context
Genomic regions, transcripts, and products
Bibliography
Phenotypes
Variation
Pathways from BioSystems
Interactions
General gene information
Markers, Homology, Gene Ontology
General protein information
NCBI Reference Sequences (RefSeq)
Related sequences
Additional links
Locus-specific Databases

Genome Browsers
Genome Data Viewer
Map Viewer

So we will go for humans and then you can have all the information about this gene in the humans from what is the lineage what is the summary what does this gene does what are the orthologs present.

(Refer Slide Time: 20:02)



Location: 6q24.3

Exon count: 10

See EPM2A in Genome Data Viewer Map Viewer

Annotation release	Status	Assembly	Chr	Location
108	current	GRCh38.p7 (GCF_000001405.33)	6	NC_000006.12 (145500744..145736018, complement)
105	previous assembly	GRCh37.p13 (GCF_000001405.25)	6	NC_000006.11 (145946440..146057128, complement)

Genomic regions, transcripts, and products

Genomic Sequence: NC_000006.12 Chromosome 6 Reference GRCh38.p7 Primary Assembly

Go to reference sequence details

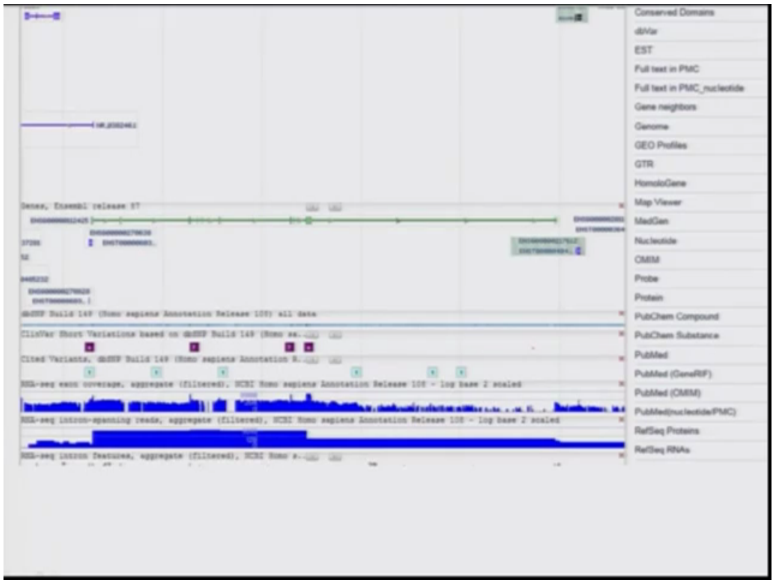
Go to nucleotide: Graphics FASTA GenBank

Genome Browsers
Genome Data Viewer
Map Viewer
Variation Viewer (GRCh37.p13)
Variation Viewer (GRCh38)
1000 Genomes Browser (GRCh37.p13)
Ensembl
UCSC

Related information
Order cDNA clone
3D structures
BioAssay by Target (List)
BioAssay by Target (Summary)
BioAssay, by Gene target
BioAssays, RNAi Target, Tested
BioProjects
BioSystems
Books
CCDS
ClinVar
Conserved Domains
dbVar
EST

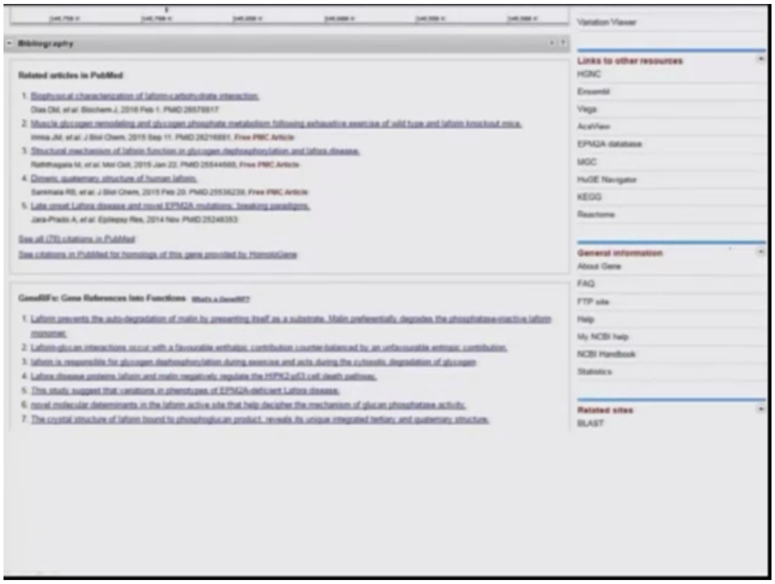
Similarly from the on the right hand side you have all the other information then you can scroll down and look into what are the other cDNA clones, 3D structures and everything. What are the EST associated with this gene and so on, now here you see on the screen that this is showing what are the transcripts present so there are I, 2 3, 4, 5, 6, 7 and 8, 8 transcripts for this gene that is EPM2A.

(Refer Slide Time: 20:38)


Genes, Ensembl release 87
dbSNP Build 149 (Homo sapiens Annotation Release 108) all data
ClinVar Short Variations based on dbSNP Build 149 (Homo sa...
Cited Variants, dbSNP Build 149 (Homo sapiens Annotation R...
RNA-seq exon coverage, aggregate (filtered), NCBI Homo sapiens Annotation Release 108 - log base 2 scaled
RNA-seq intron-spanning reads, aggregate (filtered), NCBI Homo sapiens Annotation Release 108 - log base 2 scaled
RNA-seq intron features, aggregate (filtered), NCBI Homo s...
Conserved Domains
dbVar
EST
Full text in PMC
Full text in PMC_nucleotide
Gene neighbors
Genome
GEO Profiles
GTR
HomoloGene
Map Viewer
MedGen
Nucleotide
OMIM
Probe
Protein
PubChem Compound
PubChem Substance
PubMed
PubMed (GeneRIF)
PubMed (OMIM)
PubMed(nucleotide/PMC)
RefSeq Proteins
RefSeq RNAs

Then on again you can look for OMIM which is Online Mendelian Inheritance of Man that is this is another search tool by you can search for EPM2A and you can see that to what all disease it is associated with.

(Refer Slide Time: 21:10)


Bibliography

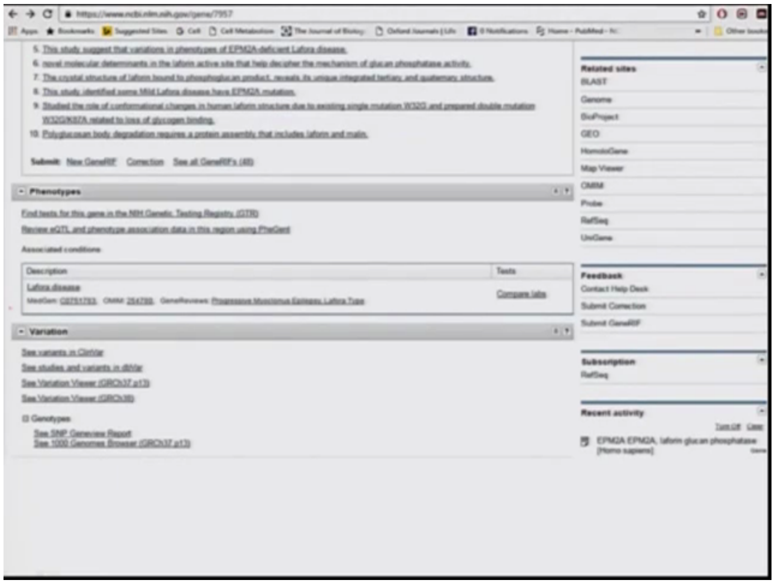
Related articles in PubMed

1. Biophysical characterization of laforin-carbohydrate interaction.
   Dias DM, et al. Biochem J. 2016 Feb 1. PMID 26578817
2. Muscle glycogen remodeling and glycogen phosphate metabolism following exhaustive exercise of wild type and laforin knockout mice.
   Irimia JM, et al. J Biol Chem. 2015 Sep 11. PMID 26216881. Free PMC Article
3. Structural mechanism of laforin function in glycogen dephosphorylation and lafora disease.
   Raththagala M, et al. Mol Cell. 2015 Jan 22. PMID 25544560. Free PMC Article
4. Dimeric quaternary structure of human laforin.
   Sankhala RS, et al. J Biol Chem. 2015 Feb 20. PMID 25538239. Free PMC Article
5. Late onset Lafora disease and novel EPM2A mutations: breaking paradigms.
   Jara-Prado A, et al. Epilepsy Res. 2014 Nov. PMID 25246353

See all (79) citations in PubMed
See citations in PubMed for homologs of this gene provided by HomoloGene

GeneRIFs: Gene References Into Functions What's a GeneRIF?

1. Laforin prevents the auto-degradation of malin by presenting itself as a substrate. Malin preferentially degrades the phosphatase-inactive laforin monomer.
2. Laforin-glucan interactions occur with a favourable enthalpic contribution counter-balanced by an unfavourable entropic contribution.
3. laforin is responsible for glycogen dephosphorylation during exercise and acts during the cytosolic degradation of glycogen
4. Lafora disease proteins laforin and malin negatively regulate the HIPK2-p53 cell death pathway.
5. This study suggest that variations in phenotypes of EPM2A-deficient Lafora disease.
6. novel molecular determinants in the laforin active site that help decipher the mechanism of glucan phosphatase activity.
7. The crystal structure of laforin bound to phosphoglucan product, reveals its unique integrated tertiary and quaternary structure.



5. This study suggest that variations in phenotypes of EPM2A-deficient Lafora disease.
6. novel molecular determinants in the laforin active site that help decipher the mechanism of glucan phosphatase activity.
7. The crystal structure of laforin bound to phosphoglucan product, reveals its unique integrated tertiary and quaternary structure.
8. This study identified some Mild Lafora disease have EPM2A mutation.
9. Studied the role of conformational changes in human laforin structure due to existing single mutation W32G and prepared double mutation W32G/K87A related to loss of glycogen binding.
10. Polyglucosan body degradation requires a protein assembly that includes laforin and malin.

Submit: New GeneRIF Correction See all GeneRIF's (48)

Phenotypes

Find tests for this gene in the NIH Genetic Testing Registry (GTR)
Review eQTL and phenotype association data in this region using PheGenI

Associated conditions

Description	Tests
Lafora disease MedGen: C0751783, OMIM: 254780, GeneReviews: Progressive Myoclonus Epilepsy, Lafora Type	Compare labs

Variation

See variants in ClinVar
See studies and variants in dbVar
See Variation Viewer (GRCh37.p13)
See Variation Viewer (GRCh38)

Genotypes
See SNP Geneview Report
See 1000 Genomes Browser (GRCh37.p13)

Related sites
BLAST
Genome
BioProject
GEO
HomoloGene
Map Viewer
OMIM
Probe
RefSeq
UniGene

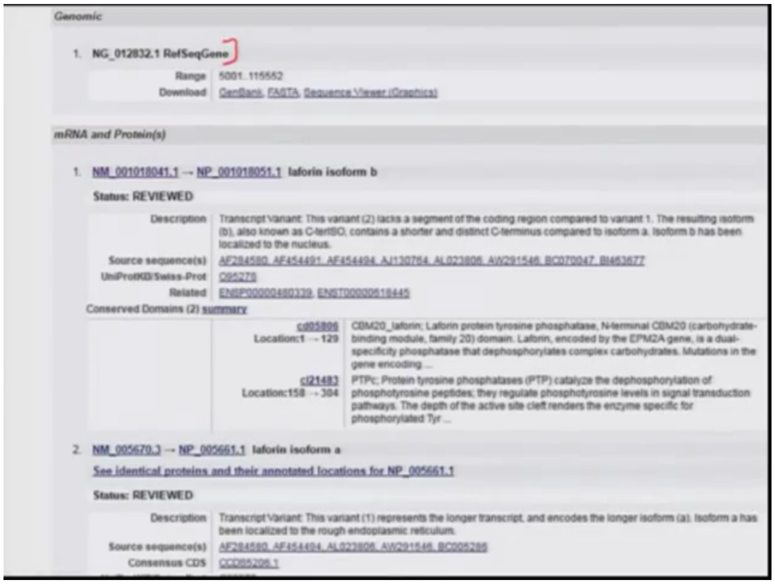
Feedback
Contact Help Desk
Submit Correction
Submit GeneRIF

Subscription
RefSeq

Recent activity
Turn Off Clear
EPM2A EPM2A, laforin glucan phosphatase [Homo sapiens]

Then again you have all the related articles and then gene reference and there is a huge amount of information already available on the internet which can help you arrive if all kind of information you are looking at but you need to know what you are what do you what exactly you want to search.

(Refer Slide Time: 21:21)



Genomic

1. NG_012832.1 RefSeqGene

Range 5001..115552
Download GenBank, FASTA, Sequence Viewer (Graphics)

mRNA and Protein(s)

1. NM_001018041.1 → NP_001018051.1 laforin isoform b

Status: REVIEWED

Description: Transcript Variant: This variant (2) lacks a segment of the coding region compared to variant 1. The resulting isoform (b), also known as C-terISO, contains a shorter and distinct C-terminus compared to isoform a. Isoform b has been localized to the nucleus.
Source sequence(s): AF284580, AF454491, AF454494, AJ130764, AL023806, AW291546, BC070047, BI463677
UniProtKB/Swiss-Prot: O95278
Related: ENSP00000480339, ENST00000618445
Conserved Domains (2) summary

cd05806 Location:1 → 129: CBM20_laforin; Laforin protein tyrosine phosphatase, N-terminal CBM20 (carbohydrate-binding module, family 20) domain. Laforin, encoded by the EPM2A gene, is a dual-specificity phosphatase that dephosphorylates complex carbohydrates. Mutations in the gene encoding ...

cl21483 Location:158 → 304: PTPc; Protein tyrosine phosphatases (PTP) catalyze the dephosphorylation of phosphotyrosine peptides; they regulate phosphotyrosine levels in signal transduction pathways. The depth of the active site cleft renders the enzyme specific for phosphorylated Tyr ...

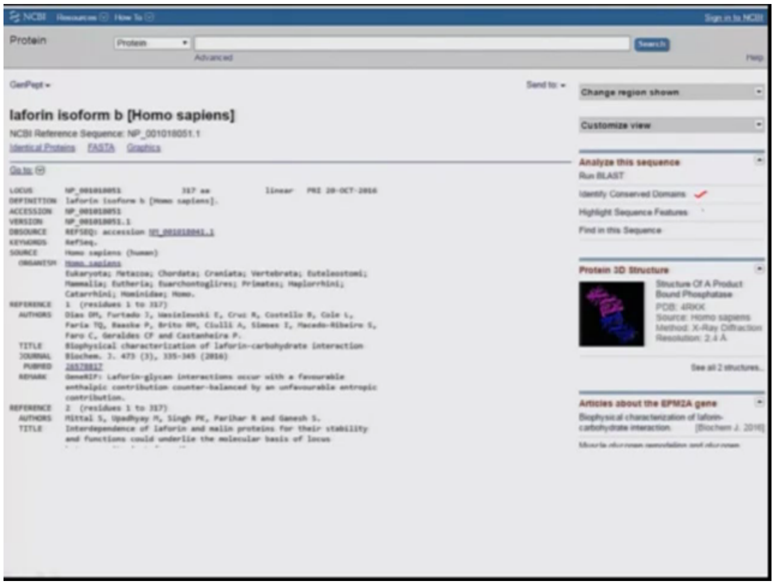
2. NM_005670.3 → NP_005661.1 laforin isoform a
See identical proteins and their annotated locations for NP_005661.1

Status: REVIEWED

Description: Transcript Variant: This variant (1) represents the longer transcript, and encodes the longer isoform (a). Isoform a has been localized to the rough endoplasmic reticulum.
Source sequence(s): AF284580, AF454494, AL023806, AW291546, BC005286
Consensus CDS: CCDS5206.1

Similarly when you will scroll down you wills see that there are the reference sequences of the gene you can go to the chromosomal region from which this gene was coded and there you can have all the information about introns, exons what are the sequences as the human chromosome or the human genome has been sequenced aligned and deposited on internet for our ease. Similarly there is a protein sequence and mRNA sequence NM generally is for the mRNA sequence and NP is mainly for proteins but sometimes you will also see accession IDs with xp and all, these are for the proteins.

(Refer Slide Time: 22:17)



laforin isoform b [Homo sapiens]

NCBI Reference Sequence: NP_001018051.1

```
LOCUS       NP_001018051             317 aa            linear   PRI 20-OCT-2016
DEFINITION  laforin isoform b [Homo sapiens].
ACCESSION   NP_001018051
VERSION     NP_001018051.1
DBSOURCE    REFSEQ: accession NM_001018041.1
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 317)
  AUTHORS   Dias DM, Furtado J, Wasielewski E, Cruz R, Costello B, Cole L,
            Faria TQ, Baaske P, Brito RM, Ciulli A, Simoes I, Macedo-Ribeiro S,
            Faro C, Geraldes CF and Castanheira P.
  TITLE     Biophysical characterization of laforin-carbohydrate interaction
  JOURNAL   Biochem. J. 473 (3), 335-345 (2016)
   PUBMED   26578817
  REMARK    GeneRIF: Laforin-glycan interactions occur with a favourable
            enthalpic contribution counter-balanced by an unfavourable entropic
            contribution.
REFERENCE   2  (residues 1 to 317)
  AUTHORS   Mittal S, Upadhyay M, Singh PK, Parihar R and Ganesh S.
  TITLE     Interdependence of laforin and malin proteins for their stability
            and functions could underlie the molecular basis of locus
```

Analyze this sequence

Run BLAST

Identify Conserved Domains ✓

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Structure Of A Product Bound Phosphatase
PDB: 4RKK
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 2.4 Å

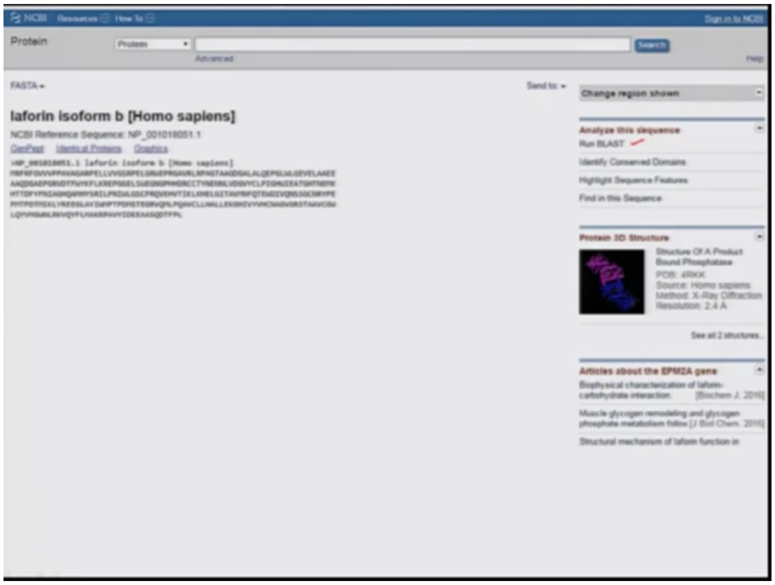
See all 2 structures...

Articles about the EPM2A gene

Biophysical characterization of laforin-carbohydrate interaction. [Biochem J. 2016]

So from here you can go to the protein sequence which this gene is coding for and then you have all the information about the protein data base that is you can have the conserve domain known you can highlight some sequence feature you can use the FASTA sequence and you can run a blast as in against all the protein data base and so on.

(Refer Slide Time: 22:38)



laforin isoform b [Homo sapiens]

NCBI Reference Sequence: NP_001018051.1

GenPept Identical Proteins Graphics

```
>NP_001018051.1 laforin isoform b [Homo sapiens]
[illegible]
```

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Structure Of A Product Bound Phosphatase

PDB: 4RKK

Source: Homo sapiens

Method: X-Ray Diffraction

Resolution: 2.4 Å

See all 2 structures...

Articles about the EPM2A gene

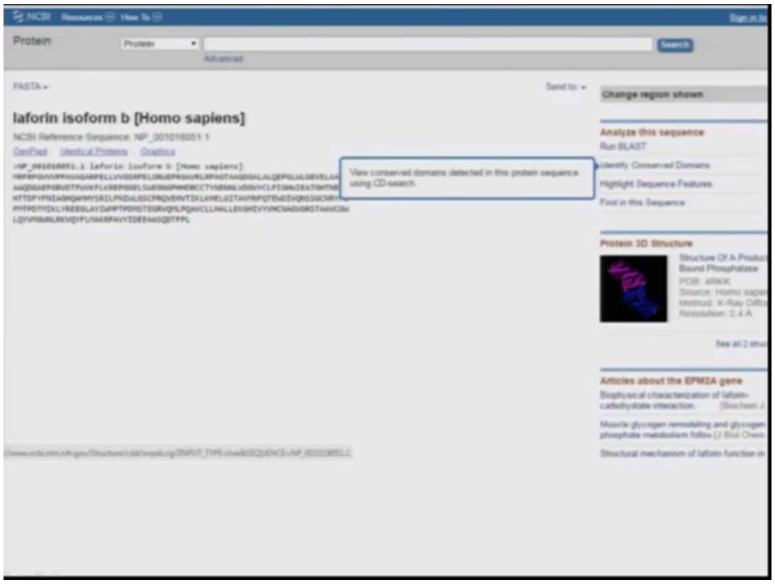
Biophysical characterization of laforin-carbohydrate interaction. [Biochem J. 2016]

Muscle glycogen remodeling and glycogen phosphate metabolism follow [J Biol Chem. 2016]

Structural mechanism of laforin function in

So with this we have the FASTA sequence and we can run blast this protein is aligning to what all species of the protein sequences it is aligning to the protein sequences of what all species.

(Refer Slide Time: 22:47)



laforin isoform b [Homo sapiens]

NCBI Reference Sequence: NP_001018051.1

GenPept Identical Proteins Graphics

```
>NP_001018051.1 laforin isoform b [Homo sapiens]
[illegible]
```

View conserved domains detected in this protein sequence using CD-search.

Analyze this sequence

Run BLAST

Identify Conserved Domains

Highlight Sequence Features

Find in this Sequence

Protein 3D Structure

Structure Of A Product Bound Phosphatase

PDB: 4RKK

Source: Homo sapie

Method: X-Ray Diffra

Resolution: 2.4 Å

See all 2 stru

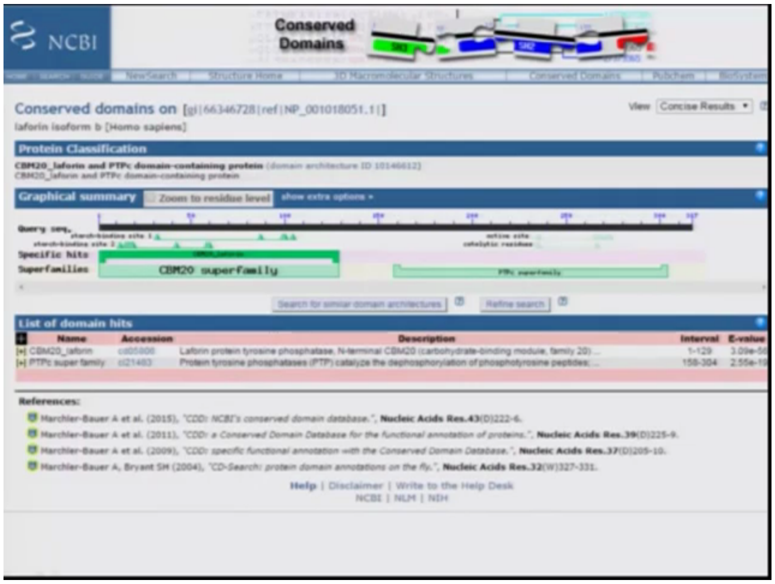
Articles about the EPM2A gene

Biophysical characterization of laforin-carbohydrate interaction. [Biochem J

Muscle glycogen remodeling and glycogen phosphate metabolism follow [J Biol Chem

Structural mechanism of laforin function in

//www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi?INPUT_TYPE=live&SEQUENCE=NP_001018051.1



NCBI
Conserved Domains
NewSearch | Structure Home | 3D Macromolecular Structures | Conserved Domains | Pubchem | BioSystems

Conserved domains on [gi|66346728|ref|NP_001018051.1|]

laforin isoform b [Homo sapiens]

View Concise Results

Protein Classification

CBM20_laforin and PTPc domain-containing protein (domain architecture ID 10146612)

CBM20_laforin and PTPc domain-containing protein

Graphical summary □ Zoom to residue level show extra options »

Search for similar domain architectures | Refine search

List of domain hits

Name	Accession	Description	Interval	E-value
[+] CBM20_laforin	cd05806	Laforin protein tyrosine phosphatase, N-terminal CBM20 (carbohydrate-binding module, family 20) ...	1-129	3.09e-5
[+] PTPc super family	cl21483	Protein tyrosine phosphatases (PTP) catalyze the dephosphorylation of phosphotyrosine peptides; ...	158-304	2.55e-1

References:

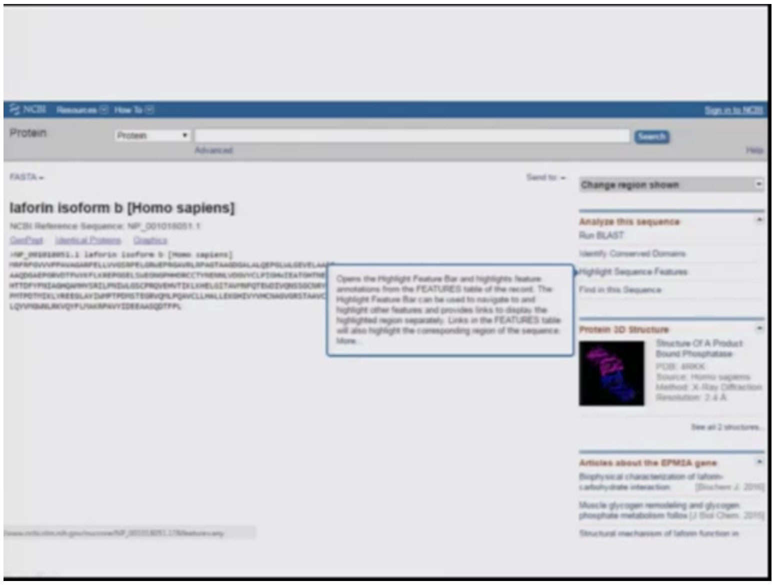
- Marchler-Bauer A et al. (2015), *"CDD: NCBI's conserved domain database."*, **Nucleic Acids Res.43**(D)222-6.
- Marchler-Bauer A et al. (2011), *"CDD: a Conserved Domain Database for the functional annotation of proteins."*, **Nucleic Acids Res.39**(D)225-9.
- Marchler-Bauer A et al. (2009), *"CDD: specific functional annotation with the Conserved Domain Database."*, **Nucleic Acids Res.37**(D)205-10.
- Marchler-Bauer A, Bryant SH (2004), *"CD-Search: protein domain annotations on the fly."*, **Nucleic Acids Res.32**(W)327-331.

Help | Disclaimer | Write to the Help Desk
NCBI | NLM | NIH

Similarly you can identify conserve domains. So if you click here you will get to know what are the conserved domains or motives in this proteins sequence for example in this protein which was being coded by EPM2A it has CBM20, it is carbohydrate binding domain. So as the function of EPM2A was in glycogen metabolism and in glycogen synthesis path way so it will goes with the the functions which it does that it would bind to the starch or the glycogen, similarly there is another region that is PTPC super family. What is PTPC super family? It is a protein tyrosine phosphatase which catalyses the de-phosphorylation of phospho tyrosine peptide.

So this protein which was being encoded by EPM2A it has two domains one is CBM20 which has ability to bind two glycogen like molecules and similarly it has an another domain that is PTPC which is for the which is a phosphatase domain by that we mean that it can de-phosphorylate the phosphate group from the peptides.

(Refer Slide Time: 24:08)



laforin isoform b [Homo sapiens]

NCBI Reference Sequence: NP_001018051.1

>NP_001018051.1 laforin isoform b [Homo sapiens]

Opens the Highlight Feature Bar and highlights feature annotations from the FEATURES table of the record. The Highlight Feature Bar can be used to navigate to and highlight other features and provides links to display the highlighted region separately. Links in the FEATURES table will also highlight the corresponding region of the sequence. More...

Structure Of A Product Bound Phosphatase
PDB: 4RKK
Source: Homo sapiens
Method: X-Ray Diffraction
Resolution: 2.4 Å

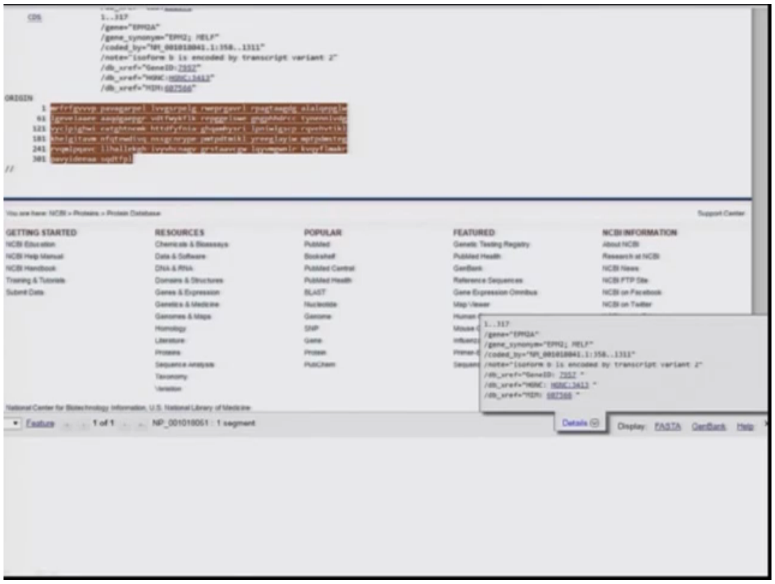
Biophysical characterization of laforin-carbohydrate interaction [Biochem J. 2016]

Muscle glycogen remodeling and glycogen phosphate metabolism follow [J Biol Chem. 2015]

Structural mechanism of laforin function in

Similarly you can highlight the sequence features of this peptide sequence.

(Refer Slide Time: 24:14)

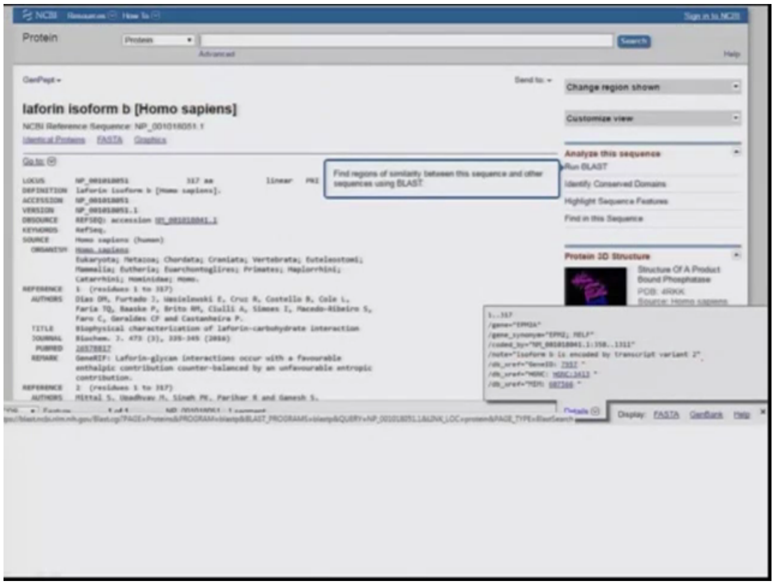


```
     CDS             1..317
                     /gene="EPM2A"
                     /gene_synonym="EPM2; MELF"
                     /coded_by="NM_001018041.1:358..1311"
                     /note="isoform b is encoded by transcript variant 2"
                     /db_xref="GeneID:7957"
                     /db_xref="HGNC:HGNC:3413"
                     /db_xref="MIM:607566"
ORIGIN
        1
       61
      121
      181
      241
      301
//
```

```
1..317
/gene="EPM2A"
/gene_synonym="EPM2; MELF"
/coded_by="NM_001018041.1:358..1311"
/note="isoform b is encoded by transcript variant 2"
/db_xref="GeneID: 7957 "
/db_xref="HGNC: HGNC:3413 "
/db_xref="MIM: 607566 "
```

And this is the sequence feature which is highlighted which shows that it is the iso form B is encoded by transcript variant 2 and so on.

(Refer Slide Time: 24:26)



laforin isoform b [Homo sapiens]

NCBI Reference Sequence: NP_001018051.1

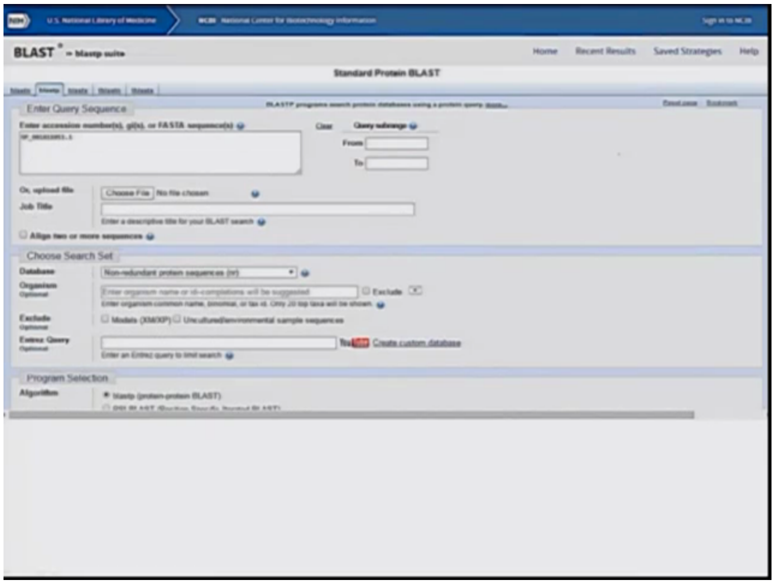
Find regions of similarity between this sequence and other sequences using BLAST.

```
LOCUS       NP_001018051             317 aa            linear   PRI
DEFINITION  laforin isoform b [Homo sapiens].
ACCESSION   NP_001018051
VERSION     NP_001018051.1
DBSOURCE    REFSEQ: accession NM_001018041.1
KEYWORDS    RefSeq.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 317)
  AUTHORS   Dias DM, Furtado J, Wasielewski E, Cruz R, Costello B, Cole L,
            Faria TQ, Baaske P, Brito RM, Ciulli A, Simoes I, Macedo-Ribeiro S,
            Faro C, Geraldes CF and Castanheira P.
  TITLE     Biophysical characterization of laforin-carbohydrate interaction
  JOURNAL   Biochem. J. 473 (3), 335-345 (2016)
   PUBMED   26578817
  REMARK    GeneRIF: Laforin-glycan interactions occur with a favourable
            enthalpic contribution counter-balanced by an unfavourable entropic
            contribution.
REFERENCE   2  (residues 1 to 317)
  AUTHORS   Mittal S, Upadhyay M, Singh PK, Parihar R and Ganesh S.
```

Structure Of A Product Bound Phosphatase
PDB: 4RKK
Source: Homo sapiens

```
1..317
/gene="EPM2A"
/gene_synonym="EPM2; MELF"
/coded_by="NM_001018041.1:158..1111"
/note="isoform b is encoded by transcript variant 2"
/db_xref="GeneID: 7957 "
/db_xref="HGNC: HGNC:3413 "
/db_xref="MIM: 607566 "
```

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&QUERY=NP_001018051.1&LINK_LOC=protein&PAGE_TYPE=BlastSearch



Standard Protein BLAST

BLASTP programs search protein databases using a protein query. more...

Enter Query Sequence

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

NP_001018051.1

Query subrange

From

To

Or, upload file: No file chosen

Job Title

Enter a descriptive title for your BLAST search

□ Align two or more sequences

Choose Search Set

Database: Non-redundant protein sequences (nr)

Organism (Optional): Enter organism name or id--completions will be suggested □ Exclude

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

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

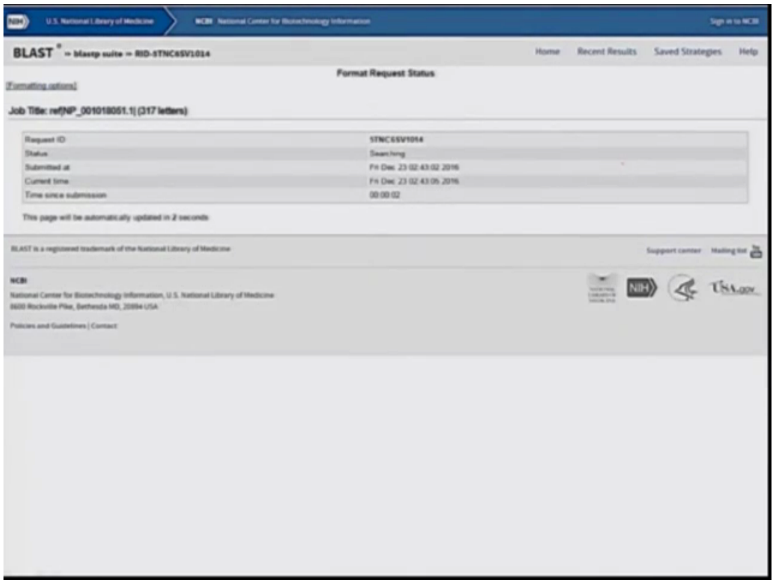
Entrez Query (Optional)

Enter an Entrez query to limit search

Program Selection

Algorithm: ● blastp (protein-protein BLAST)

○ PSI-BLAST (Position-Specific Iterated BLAST)



BLAST ® » blastp suite » RID-5TNC6SV1014

Format Request Status

Job Title: ref|NP_001018051.1| (317 letters)

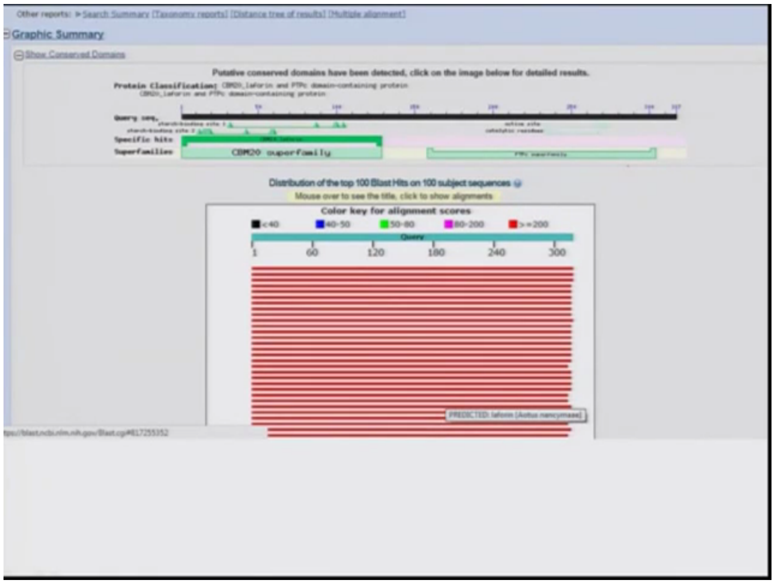
Request ID	5TNC6SV1014
Status	Searching
Submitted at	Fri Dec 23 02:43:02 2016
Current time	Fri Dec 23 02:43:05 2016
Time since submission	00:00:02

This page will be automatically updated in 2 seconds

BLAST is a registered trademark of the National Library of Medicine

NCBI
National Center for Biotechnology Information, U.S. National Library of Medicine
8600 Rockville Pike, Bethesda MD, 20894 USA

Policies and Guidelines | Contact



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

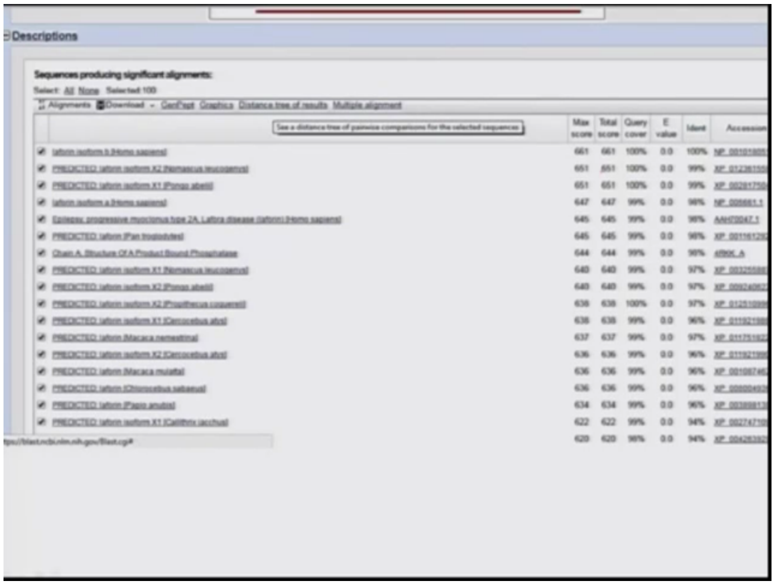
Protein Classification: CBM20_laforin and PTPc domain-containing protein
CBM20_laforin and PTPc domain-containing protein

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Similarly you can run a blast and what do you find that you find that again this blast would take NP as a query sequence and using the default settings of the search engine when we run blast. This page appears showing that it is running the the job which we have assigned to it and then you get the results again in a similar fashion. How well your query sequence is aligning to the subjects sequence what are the top hits what are organisms it is aligning all those information we can get from here.

(Refer Slide Time: 25:05)



Descriptions

Sequences producing significant alignments:

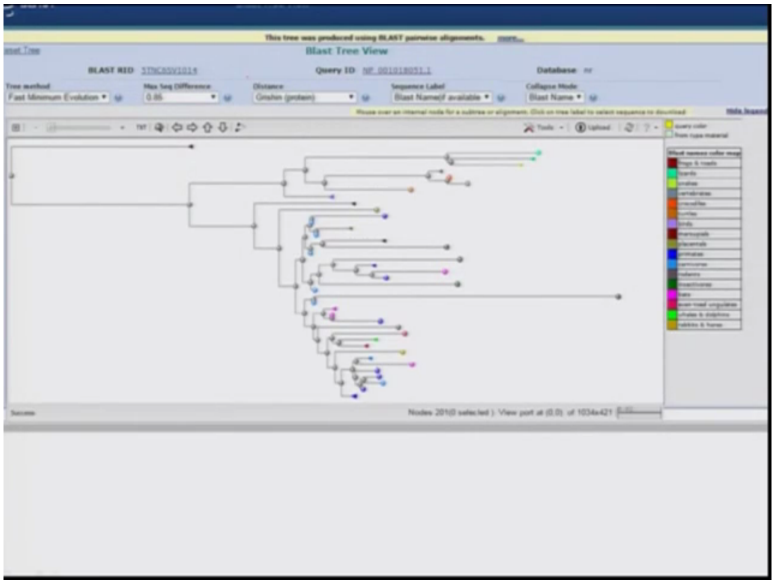
Select: All None Selected 100

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

See a distance tree of pairwise comparisons for the selected sequences

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☑	laforin isoform b [Homo sapiens]	661	661	100%	0.0	100%	NP_00101805
☑	PREDICTED: laforin isoform X2 [Nomascus leucogenys]	651	651	100%	0.0	99%	XP_0123615
☑	PREDICTED: laforin isoform X1 [Pongo abelii]	651	651	100%	0.0	99%	XP_0028175
☑	laforin isoform a [Homo sapiens]	647	647	99%	0.0	98%	NP_005661.1
☑	Epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) [Homo sapiens]	645	645	99%	0.0	98%	AAH70047.1
☑	PREDICTED: laforin [Pan troglodytes]	645	645	99%	0.0	98%	XP_0011612
☑	Chain A, Structure Of A Product Bound Phosphatase	644	644	99%	0.0	98%	4RKK_A
☑	PREDICTED: laforin isoform X1 [Nomascus leucogenys]	640	640	99%	0.0	97%	XP_0032558
☑	PREDICTED: laforin isoform X2 [Pongo abelii]	640	640	99%	0.0	97%	XP_0092406
☑	PREDICTED: laforin isoform X2 [Propithecus coquereli]	638	638	100%	0.0	97%	XP_0125109
☑	PREDICTED: laforin isoform X1 [Cercocebus atys]	638	638	99%	0.0	96%	XP_0119219
☑	PREDICTED: laforin [Macaca nemestrina]	637	637	99%	0.0	97%	XP_0117516
☑	PREDICTED: laforin isoform X2 [Cercocebus atys]	636	636	99%	0.0	96%	XP_0119219
☑	PREDICTED: laforin [Macaca mulatta]	636	636	99%	0.0	96%	XP_0010874
☑	PREDICTED: laforin [Chlorocebus sabaeus]	636	636	99%	0.0	96%	XP_0080049
☑	PREDICTED: laforin [Papio anubis]	634	634	99%	0.0	96%	XP_0038981
☑	PREDICTED: laforin isoform X1 [Callithrix jacchus]	622	622	99%	0.0	94%	XP_0027471
		620	620	98%	0.0	94%	XP_0042839

tps://blast.ncbi.nlm.nih.gov/Blast.cgi#



This tree was produced using BLAST pairwise alignments. more...

Blast Tree View

BLAST RID: 5TNCA5V1014 Query ID: NP_001018051.1 Database: nr

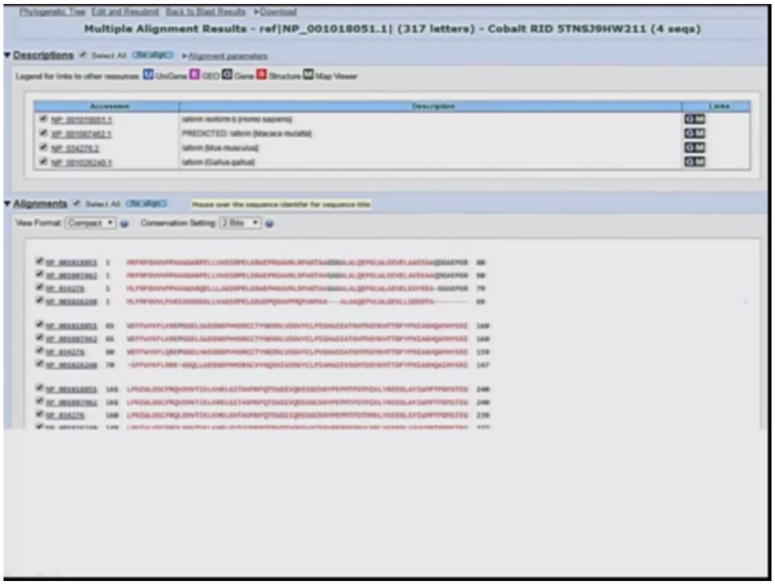
Tree method: Fast Minimum Evolution
Max Seq Difference: 0.85
Distance: Grishin (protein)
Sequence Label: Blast Name(if available)
Collapse Mode: Blast Name

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download

Nodes: 201(0 selected). View port at (0,0) of 1034x421

Now you have all the sequences which has aligned to your query sequence and from there you can select all and then again you can go to see the distance tree results which will bring the phylogenetic tree for this protein along the various phylums of the various texonomic groups along the various texonomic groups. For example frog, toads or turtle or insectivorous rabbits and so on, so you can see that how your protein of interest has evolved through this stages of evolution and how conserved it is.

(Refer Slide Time: 25:50)



Phylogenetic Tree Edit and Resubmit Back to Blast Results ▸Download

Multiple Alignment Results - ref|NP_001018051.1| (317 letters) - Cobalt RID 5TNSJ9HW211 (4 seqs)

▾ Descriptions ☑ Select All (Re-align) ▸Alignment parameters

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

Accession	Description	Links
☑ NP_001018051.1	laforin isoform b [Homo sapiens]	G M
☑ XP_001087462.1	PREDICTED: laforin [Macaca mulatta]	G M
☑ NP_034276.2	laforin [Mus musculus]	G M
☑ NP_001026240.1	laforin [Gallus gallus]	G M

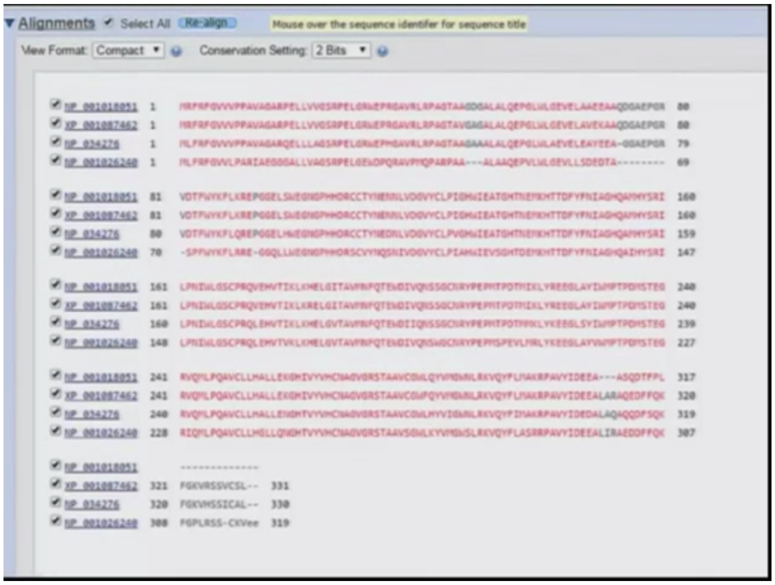
▾ Alignments ☑ Select All (Re-align) Mouse over the sequence identifier for sequence title

View Format: Compact ▾ Conservation Setting: 2 Bits ▾

```
☑ NP_001018051   1   MRFRFGVVVPPAVAGARPELLVVGSRPELGRWEPRGAVRLRPAGTAAGDGALALQEPGLWLGEVELAAEEAAQDGAEPGR  80
☑ XP_001087462   1   MRFRFGVVVPPAVAGARPELLVVGSRPELGRWEPRGAVRLRPAGTAVGAGALALQEPGLWLGEVELAVEKAAQDGAEPGR  80
☑ NP_034276      1   MLFRFGVVVPPAVAGARQELLLAGSRPELGRWEPHGAVRLRPAGTAAGAAALALQEPGLWLAEVELEAYEEA-GGAEPGR  79
☑ NP_001026240   1   MLFRFGVVLPARIAEGGGALLVAGSRPELGEWDPQRAVPMQPARPAA---ALAAQEPVLWLGEVLLSDEDTA--------  69

☑ NP_001018051   81  VDTFWYKFLKREPGGELSWEGNGPHHDRCCTYNENNLVDGVYCLPIGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI 160
☑ XP_001087462   81  VDTFWYKFLKREPGGELSWEGNGPHHDRCCTYNENNLVDGVYCLPIGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI 160
☑ NP_034276      80  VDTFWYKFLQREPGGELHWEGNGPHHDRCCTYNEDNLVDGVYCLPVGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI 159
☑ NP_001026240   70  -SPFWYKFLRRE-GGQLLWEGNGPHHDRSCVYNQSNIVDGVYCLPIAHWIEVSGHTDEMKHTTDFYFNIAGHQAIHYSRI 147

☑ NP_001018051   161 LPNIWLGSCPRQVEHVTIKLKHELGITAVMNFQTEWDIVQNSSGCNRYPEPMTPDTMIKLYREEGLAYIWMPTPDMSTEG 240
☑ XP_001087462   161 LPNIWLGSCPRQVEHVTIKLKRELGITAVMNFQTEWDIVQNSSGCNRYPEPMTPDTMIKLYREEGLAYIWMPTPDMSTEG 240
☑ NP_034276      160 LPNIWLGSCPRQLEHVTIKLKHELGVTAVMNFQTEWDIIQNSSGCNRYPEPMTPDTMRKLYKEEGLSYIWMPTPDMSTEG 239
```

You can also select few sequences for example here I selected the proteins sequence from the humans then monkey then mouse and chicken and then you can see how well it is aligning and you can also identify what all regions are what all regions are more conserved.

(Refer Slide Time: 26:09)



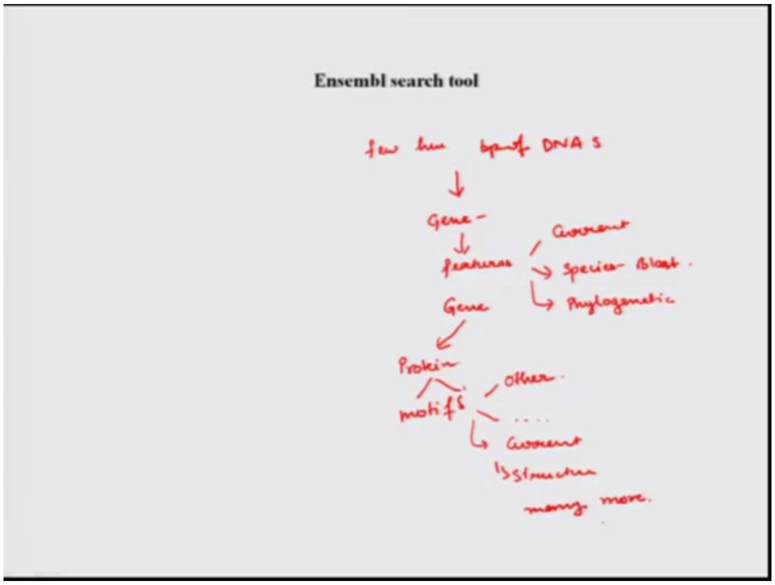
▾ Alignments ☑ Select All (Re-align) Mouse over the sequence identifer for sequence title

View Format: Compact ▾ Conservation Setting: 2 Bits ▾

```
☑ NP_001018051   1   MRFRFGVVVPPAVAGARPELLVVGSRPELGRWEPRGAVRLRPAGTAAGDGALALQEPGLWLGEVELAAEEAAQDGAEPGR  80
☑ XP_001087462   1   MRFRFGVVVPPAVAGARPELLVVGSRPELGRWEPRGAVRLRPAGTAVGAGALALQEPGLWLGEVELAVEKAAQDGAEPGR  80
☑ NP_034276      1   MLFRFGVVVPPAVAGARQELLLAGSRPELGRWEPHGAVRLRPAGTAAGAAALALQEPGLWLAEVELEAYEEA-GGAEPGR  79
☑ NP_001026240   1   MLFRFGVVLPARIAEGGGALLVAGSRPELGEWDPQRAVPMQPARPAA---ALAAQEPVLWLGEVLLSDEDTA--------  69

☑ NP_001018051   81  VDTFWYKFLKREPGGELSWEGNGPHHDRCCTYNENNLVDGVYCLPIGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI 160
☑ XP_001087462   81  VDTFWYKFLKREPGGELSWEGNGPHHDRCCTYNENNLVDGVYCLPIGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI 160
☑ NP_034276      80  VDTFWYKFLQREPGGELHWEGNGPHHDRCCTYNEDNLVDGVYCLPVGHWIEATGHTNEMKHTTDFYFNIAGHQAMHYSRI 159
☑ NP_001026240   70  -SPFWYKFLRRE-GGQLLWEGNGPHHDRSCVYNQSNIVDGVYCLPIAHWIEVSGHTDEMKHTTDFYFNIAGHQAIHYSRI 147

☑ NP_001018051   161 LPNIWLGSCPRQVEHVTIKLKHELGITAVMNFQTEWDIVQNSSGCNRYPEPMTPDTMIKLYREEGLAYIWMPTPDMSTEG 240
☑ XP_001087462   161 LPNIWLGSCPRQVEHVTIKLKRELGITAVMNFQTEWDIVQNSSGCNRYPEPMTPDTMIKLYREEGLAYIWMPTPDMSTEG 240
☑ NP_034276      160 LPNIWLGSCPRQLEHVTIKLKHELGVTAVMNFQTEWDIIQNSSGCNRYPEPMTPDTMRKLYKEEGLSYIWMPTPDMSTEG 239
☑ NP_001026240   148 LPNIWLGSCPRQLEHVTVKLKHELGVTAVMNFQTEWDIVQNSWGCNRYPEPMSPEVLMRLYKEEGLAYVWMPTPDMSTEG 227

☑ NP_001018051   241 RVQMLPQAVCLLHALLEKGHIVYVHCNAGVGRSTAAVCGWLQYVMGWNLRKVQYFLMAKRPAVYIDEEA---ASQDTFPL 317
☑ XP_001087462   241 RVQMLPQAVCLLHALLEKGHIVYVHCNAGVGRSTAAVCGWFQYVMGWNLRKVQYFLMAKRPAVYIDEEALARAQEDFFQK 320
☑ NP_034276      240 RVQMLPQAVCLLHALLENGHTVYVHCNAGVGRSTAAVCGWLHYVIGWNLRKVQYFIMAKRPAVYIDEDALAQAQQDFSQK 319
☑ NP_001026240   228 RIQMLPQAVCLLHGLLQNGHTVYVHCNAGVGRSTAAVSGWLKYVMGWSLRKVQYFLASRRPAVYIDEEALIRAEDDFFQK 307

☑ NP_001018051       -------------
☑ XP_001087462   321 FGKVRSSVCSL--   331
☑ NP_034276      320 FGKVHSSICAL--   330
☑ NP_001026240   308 FGPLRSS-CKVee   319
```

So you see that in these four groups which I which I have selected mainly most of it is aligning and only like suppose the one of the humans is not aligning to the other three that is monkey, mouse or and the chicken. The few last base pairs are not aligning that well.

(Refer Slide Time: 26:32)

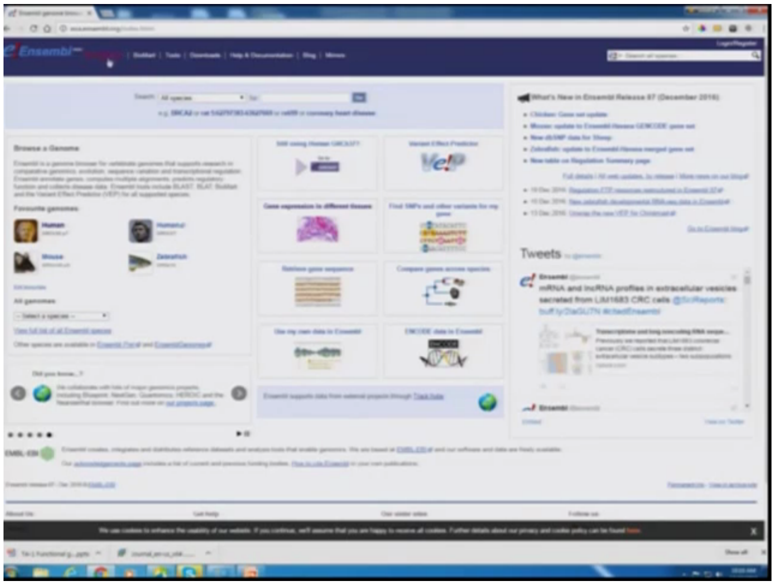


Ensembl search tool

This is how this huge amount of information we just got from few hundred base pair of the DNA sequence so what all we arrived at that we had few hundred base pairs few hundred base pair of the DNA sequence. From there we arrived to a gene then we looked into what are the features of that gene, we can look into that. What is the current literature about that we looked into in what all species it is present that is by using blast. We can also look into the how conserved that is that is by looking into the phylogenetic tree in and what all organisms it is there. From that gene you can look go to the protein sequence, you can see into what all motives or conserved motives it has how well it is conserved in the other species and so on. What is the current literature about that protein you can look into the structure of that protein and many more.

So this is the power of Bio-informatics and computer science which has been given to the biologists that when you have a small sequence this all you can search it against all the data base present. Now we will look into another search engine that is Ensembl search engine and very quickly we will go we have almost understood what all we can do but let us look into what we can do. So now we will look into another search engine that is Ensembl and we will go online and look into what all information can be fished from that search engine.

(Refer Slide Time: 28:32)


Ensembl
Human
Mouse
Zebrafish
Gene expression in different tissues
Compare genes across species
VEP
Tweets

Ensembl

mRNA and lncRNA profiles in extracellular vesicles secreted from LIM1683 CRC cells @SciReports: buff.ly/2laGLTN #citedEnsembl

So this is the Ensembl home page and you can do here blast or what are other tools are available similarly you can have your favourite genome selected if you want the gene to be only looked into humans or in zebra-fish or mouse information, similarly you can go for gene expression in different tissues how you gene expresses in different tissues of your organism of your interest, you can compare the gene across the species which we have also done with the help of NCBI using the phylogenetic tree which we created selecting all the sequences.

You can use your own data or you can also upload the data which you have created so depending on your requirement you can search the available search engines for the information and you can explore whatever you wished to. There are various tutorials already available on the YouTube which we can use to do so that is all thank you I hope this session has been somewhat informative for you and which will help you in using these search engines that is all. Thank you.