#### Bioengineering: An Interface with Biology and Medicine Prof. Sanjeeva Srivastava Department of Biosciences and Bioengineering Indian Institute of Technology - Bombay

#### Lecture – 35 Bioinformatics-I

So welcome to MOOC-NPTEL course on Bioengineering, an Interface with Biology and Medicine. In the last few lectures, I tried to give you some of the problem statements, some bigger problems towards which all of us have to work together whether we are biologist or engineers. Actually, you know, we have to think about the goal of how to solve those major problems.

I also try to provide you an interface with clinicians to give you some motivation that in which way engineering discipline can help medical scientist and clinicians for our day to day medical problems. To understand in which way the medical field can also benefit from the engineering tools and various insights you can bring in now from the engineering point of view. In this slide I thought it might be good idea to introduce you to some of the bioinformatics tools.

Some time even if you do not have an opportunity to work in a vet laboratory to do the experiments yourself, you can actually obtain lot of data which is publically available now a days in different databases and repositories. And then if you know some of the open access software and tools, you can then try to download the data available already and then make use of the data and then try to understand in which way people are studying genes and proteins.

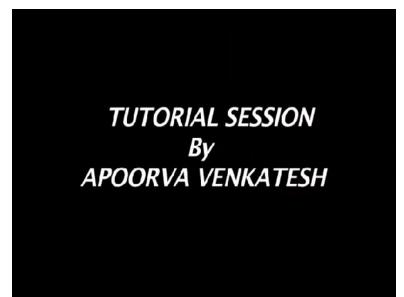
So that is something which you can do from your own system, your own computers by sitting at your own room. You need not to work in a laboratory setup. So even before we dig deeper into the fundamental concepts, I thought it is good idea to begin with some of the bioinformatic assignment. To do this, I will take help of one of my teaching assistants. We will first try to understand and explain, you know, more information how you can retrieve for a given gene.

So in my previous lectures I had talked to you about, you know, how people target specifically a given gene, look at their sequences and how you can use technologies like polymerase chain

reaction to amplify the genes. So first of all you know how to get a gene sequence. Again you need to use the bioinformatics tools. You can go to NCBI portal where you enter the gene name or accession ID.

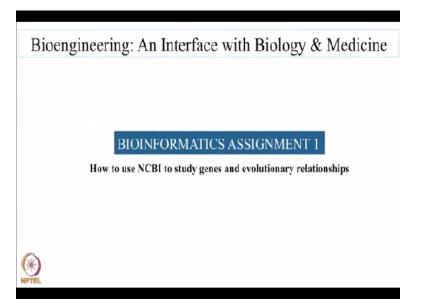
And then you will see how much information is already available for that given gene. Then in the tutorial sections you will see in a stepwise manner how this assignment could be performed. So I will hand it over now to my TA and after doing the assignment, we will also give you some problem statements and some assignments which you can do at your own place. (Video Starts 02:45 - Video Ends 10:54)

(Refer Slide Time: 02:45)



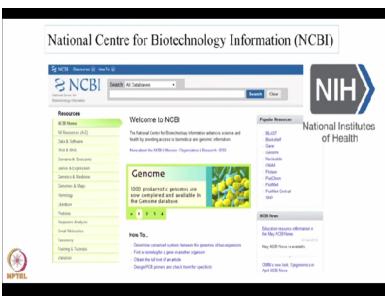
I am Apoorva Venkatesh. I am a final year PhD student under professor Sanjeeva Srivastava and I am your TA for this course.

(Refer Slide Time: 02:54)



This bioinformatics assignment will teach you how to study genes using NCBI and their evolutionary relationships.

(Refer Slide Time: 03:02)



NCBI provides a variety of databases and computational tools which are freely accessible for use by the research community. NCBI is also home to PubMed which many of you may have used for literature searches. In today's assignment, we will learn how to retrieve the sequence of a gene? How to identify similar sequences and how to compare multiple sequences to study their evolutionary relationships.

#### (Refer Slide Time: 03:29)

Official fall name       Date of last update       Gene symbol       Gene type       Lecation       Size of gene (5p)       Organism       Superkingdem       Size of Chromosome (bps)       # Genes on chromosome	Gene ID	3040
Gene symbol Gene type Lecation Nize of gene (bp) Organism Superkingdem Size of Chromosome (bps)	Official full name	
Gene type       Lecation       Size of gene (bp)       Organism       Superkingdom       Size of Chromosome (bps)	Date of last update	
Lecation Size of gene (bp) Organism Superkingdom Size of Chromosome (bps)	Gene symbol	
Size of gene (bp) Organism Superkingdom Size of Chromosome (bps)	Gene type	
Organism Superkingdem Size of Chromosome (bps)	Location	
Superkingdom Sizo of Chromosome (bps)	Size of gene (bp)	
Size of Chromosome (bps)	Organism	
	Superkingdom	
# Genes on chromosome	Size of Chromosome (bps)	
	# Genes on chromosome	
Hint: Data can be obtained from https://www.achi.nlm.nih.gov/		nlm.nih.gov/

EXAMPLE Assignment 1A. Enter the following details to complete the table

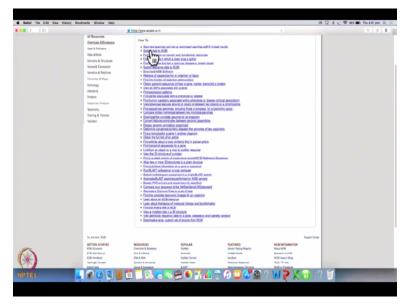
Let us begin with the first question. You have been given an accession number 3040 which is a unique identifier given to a DNA or protein sequence. We need to fill this table by getting all the information about this gene. So let us get started. It would be great if you could do this along with me on your computers. Type NCBI into your Google browser.

#### (Refer Slide Time: 03:57)

• • · · ·		the interview and a m				h 0 6
	R NOR Assource II How	16 R			San is to N281	
	SNCBI -					
	KE Inca	Welcome to NOBI			Popular Resources	
	Resource Let (4-8)	The Hallong Center for Biolectron	inge internation advances science a	ni inatiin by providing moment to	Publical	
	Al Resources	bornelcal and priorita informatio	P		Bookshelf	
	Chemicals & Bicassays	About the NCBH Million 1 Organ	rization   NCBI News & Brog		Publied Central	
	Data & Sufarere	T			Publied Health	
	Divisi Billion	Bulenit	Download	Learn	BLADY	
	Domaina-& Structures	Deposit dels or manuscripti	Tanafe NCB data to your	Findhelp incurrents, attend a	Nucleifice	
	Garrandi Excrementer	into ACB databases	omputer	class or writch alluforfall	Gerome	
	Genetics & Redicine				84 <sup>4</sup>	
	Operations & Maple			161	Gere	
	Kunshou	- T	L	1.1	Popla	
	Uterative				Publish	
	Rutaina					
	Jequence Analysis	Develop	Analyse	Research	NGBI News & Bling	
	Instrumy				Baffan minerald is the phile	
	Training & Tutoriali	UserVCBLAPIsansi usle Rhates is build applications	Mertifyan NOBI kui he your data antiyaintask	Engine HOEL research and collaborative priorch	134278	
	Variation	Search a consequence of	and a subscript.	to accuracy property	Particip research in tree accessible prime, visiFTP and PresigN/CB1s	
				/	second states without This \$1 millions	
			284	2	5 NOBI aricles × 2018 Nucleic Abds resource paragraphic multi	
					11, Jun J118 The 2018 Nucleic Acits Research Antones pays Just an second report	
					Gerbank elease 2218 is available via FTR Extent and BLAST	
					5 Ar 2 T	
					Caroliansk setuana 221.5 /12/15/2017) National 1971-005 Automatic associate for the date	
					here.	
	You are targe PCBD - Stational Cartor	for bisschology internation			Sugar Comp	
	OFTING STARTED	RESOURCES	POPULAR	/EATIRED	NOB INFORMATION	
1000	NOR Exclanar NOR TWO Manual	Charloon & Doesean	Automatical Sectored	General Teacing Response Automation	Read Wolfe Research of VCR	
1 sto	ACB Handboot	DUA.8-NVA	Automic Central	Gentlank	ACD loss 1 Day	
987	hannya human	durain & formers	Automatica Unicality	Antoneous Repuertores	M10-10-5x	
	Subwell Testa	Great Expressor	BAST	Gana Signation On Bus	NCB # Fasherk	

So this brings us to the NCBI website. We will not be able to explore all the features of NCBI today. However, I would highly recommend that you go through some of these resources after this tutorial. You can begin with the how to link, that is up here. This will teach you how to accomplish specific tasks at NCBI. Now I am going to click this link and this will give us a list of how tools.

(Refer Slide Time: 04:24)



For example, you may want to know how to retrieve all sequences for an organism or taxon. You may want to know how to find the function of a gene or gene product. View all SNPs associated with a gene. Or you may want to know how to obtain the full text of an article. So on the left is a list of resources the NCBI provides. You may want to use some of them depending on your objectives. So now let us get back to our assignment.

The first thing we will have to do is to type the accession number in the box here and click gene from the dropdown menu. Now we say search.

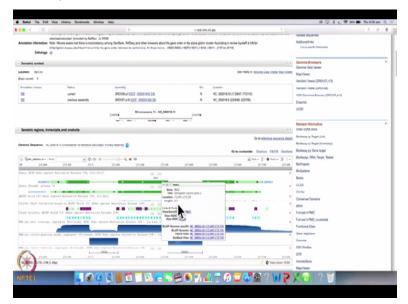
### (Refer Slide Time: 05:21)

Ratel The Edit V	fear History Busimaria Virule	w Man				# Q 1 ( ♥ #1	Thu 4 Minm Q
• ( ) []			i ndi nin	ihav	×		100
NOI Resure 10 Kee	ъя́						Serance
ire -	- Distriction	f Menal					140
Pagest -		•			Bend to -		This statist
Snowing Current metal.						Table of contants	
17 semanlable su	ubunit alpha 2 [ homo sapier	(homos) I				Durrany	
ID 3040, unliabel on 24.0		a (ramming )				Operanda randoni	
	ALC: N					Geromic wpiore, transcripte, and products	
lumnary						Equator	
Difficial Bumbell 11	M2 provided in MEAS					Biolography	
	register suburit apha 2 provint to trans					Pherotypes	
Primary sources 10	DHC HONG HON MICHIGAN MICHINESIS MILLINGS 1	Inca OTT-LACCOUNTERD				Manufacen	
cenerype pr	own outing					Patiwayshow Indepates	
Reflect status Ri Digatism H						Interactions	
	Auryola Mahama Doonata Graniasa Var	labras Eulanson: Annuala Lubata Eurotompias H	make Heinrich Calar	faci Haminidaa Boana		General gane information Nameri, Honology, Gene Ontology	
		an ehomesenetti agena about 10 kb and includes seven loci 1				General poteininformation	
		entical. These prives offer sightly over the 5' untranslated region normal adult life comprises about 97% of the tutal temoplobin a				NUB Materials Sequences (Marbeig)	
		ton. Asha halasamias result from deletions of each of the along	ones a vellas desto	s of lash HBA2 and HBA1, some nondeletion cohem-	alasamin have	Reided sequences	
matation information. No	Ny Tipkén bepeu atu/Now/menu.html), N	2008 y among Ganillach, Rollbaq, and other Insusans aloud the parts in period of the Insulation of Conference, for These Ios Is				Additional links Locus geoth Datases	
Genemic sontext						Geroma Browsers	
matter fort3.1				See HB42 in Genome Data	Vener Max Vener	Garame Bata Vasar	
en count: 3						MacView	
handalizer misusas	Status	Assembly	Ov	Location		Martidion Yessee (GR0h37 y13) Martidion Yessee (GR0h38)	
08	unet	940436 p1 (SC7 . 00001405.32)	8	NC. 0001430 (12647.17310)		108-Genome (SPECKIE) 108-Genome Brower (SPECKIE p13)	
96	province assembly	390-97 - 9 (007 000001-00.26		HC 0000163 (200016 2007/05		Enantial	
		Chronosame 11 - MC 00018.11				wollo	
100		100-1 001 001 001	100.00				
3KA	clots and anoducts					Related Information Order cDRA done	
other states, then	crona, and productly	Londing				BioAssey by Tasted (Unit)	
IPTEL			ž 🌒 🗐 🔒	🛔 🕆 🕫 🖵 🚭 👰 🤈	112	2 2 1	

So this page will give us all the details we are looking for. Please fill the table as you move along. Our first task is to find out the gene name. So this gene encodes a subunit of the protein hemoglobin. I am sure you have heard of this protein earlier. Hemoglobin is a protein molecule in red blood cells that carries oxygen from the lungs to body tissues. Here you can see that this page was last updated on 24 December, 2017.

The official gene symbol is HBA2. But however, you will also see from this list that not everyone uses this official symbol in their publications. This can also be called HBH, HBA-T2, etc. Gene type is a protein coding gene and this summary here gives us information about the gene and its functions. Now if you scroll down further, you see genomic context. So genomic context describes where the gene resides. So this gene resides on chromosome 16 in the p arm at 13.3 region.

(Refer Slide Time: 06:49)



Now if you scroll down further, this section here shows us the direction in which the gene is read. So this particular gene HBA2 is read towards the right. Now we scroll down further and we go and click on HBA2 to get us more details about this gene. So this will give us the length which is 864 base pairs. So if you click on this gene, you will get all the information on the gene. And if you further click on this green section, you will get information on the messenger RNA as well as in the translated protein.

So now let us zoom into 1 particular section of this gene and click zoom. (Refer Slide Time: 07:35)

<ul> <li>C</li> <li>D</li> </ul>			÷ national	ih av			h 0 6
Genemic context						Geroma Browsers	
matter Welli				far idd	Din German Data Viewer Met Viewe	Cenome Outa Viewer	
ion count: 3						Valution Vacuum (SPI0-37-+13)	
Anotation means	Inter	Assembly	Ċv.	Location		Variation Vacuum (CRICA.NI)	
58	current	990/08.p1(9CF_00001405.33)	16	NC.0001610(12047.173	(10)	100 Genomes Browner (SPCNIP #13	
95	previous assembly	990/07#19 (00F_00001405.25	N	NC_000163 (223946.2237	16)	Emetti	
						UCIC	
		Chromesona % - MC_0000	100.70 ( 100.70)				
			-			Relded Mornation	
iemumiu mytuma, tranmuty	ria, anti protincia						
					WE WANTED ADDRESS OF	Biolonay ty Tagat (1st)	
namb Baganan Inc. com	E. H. Chromosow 18 Reference CHCHOR	of Parage Assessible B				BioAssay ty Taget (Summary)	
				Go to nucleo	elde: Grattica EALTA Genlari	BioAssay by Gene target	
Specimental Pro-	2000	(A   & U			& tot - 5   0 tot - 5 1 -	BicAsseys, RNA Tarpet, Tested	
						Biofronte	
112.894	37.4 375.68	973.000 371.00 872.00 8	93.04	100 173.50	373.498 371.498	BUTIERS	
	37.4 375.64		y124	194 171.94	. /1.44 J144	Bolysens	
	FILL FRANK AND ADDRESS AND ADDRESS			198 J-118	PLAN DIAM		
nas, 9082 finis supress fi	FILL FRANK AND ADDRESS AND ADDRESS		91.04 p1	104 F.7.14	Plan plan	Beciperarta	
nas, 1022 Role superio h nas, Rosanii minana 11	noration Recease 109, 201-19-	et (22) (22) (22)	nan	94 <u>1719</u>	. PLes	Budysens Budy	
ner, NCE has expected by	nor Annual Lin, Recense 109, 2021-19-	et (36) (36) (36) (40)		• • • • •	. pier.	Bodyseens Boote CCOB	
ner, NCH han experie h ner, decemi ecimen () NF hald (S) then exp ners there for such	energetine Release 100, 2021-00-	e? (ii) (ii) (ii) (ii) (iii) (				Boolywana Boola Coola Confar	
ner, NCRI Rom angiotta N ner, Annones printer 11 NY hostid (5) (None angio NY hostid (5) (None a	norgijo Retene 199, 1231-09-	۲۰ کی کی ایس کی ایس ایس کی ایس کی ایس کی ایس کی				Bootpeans Boos Cote Distar Conserved Denaine diliter	
ne, NEX Tele aspect & me, denoted minute 11 MB hald 100 films aspective directive data with the me at the second s	energian Researe 10%, 2021-00- ene Reservation Release 10% all and an Audit Palse 10 dans a 10 dans capter Researches	er cui	and Disastration discontinue and and			Bectysens Book Cole Dinfor Conserved Densins Biblir	
ne, NEX Tele aspect & me, denoted minute 11 MB hald 100 films aspective directive data with the me at the second s	energian Researe 10%, 2021-00- ene Reservation Release 10% all and an Audit Palse 10 dans a 10 dans capter Researches	17 (K)	Internet in the second			Belgrenn Boura CODB Clarker Conserved Dename Afrik Full eel in PMIC	
ner, NCO Non expense in mer, screenst minister of the facil (10 phase expe office data for phase and facilities, screensty, scp inter, and, converge, scp	neration Research 109, 2010-00- ner Romantino Peissar 109, 211 	17 (2) (2) 1971 1973 1973 1975	ISS IN A REAL OF THE REAL OF T	Name and the second sec		Boolgramm Rome Cocce Contre Conserve Domanne differ Full est in PMC Full est in PMC Full est in PMC	
ner, SCH Der orpans b ern, Schult für einen si ern beild für Einer orga- ber beild für Einer orga- ber in der Gesternen bei eine beilden, sternen bei beite geste ernerage, auf	neration Research 109, 2010-00- ner Romantino Peissar 109, 211 	er con a con	SALE AND A	Name and the second sec		Bogsoms Inue Cota Confer Converse Donano Alfred Full sed a PRC Auford Full sed a PRC Auford Fundamentes	
ers, NUE how expense h ers, human expense of the house of the erspected the house of the house of the house the house of the house of the house of the house the house of the house of the house of the house the house of the house of the house of the house the house of the house of the house of the house the house of the house of the house of the house the house of the house of the house of the house of the house the house of the house of the house of the house of the house the house of the house of the house of the house of the house the house of the house of the house of the house of the house the house of the house	and the formula of the providence of the second sec		SA TRANSPORT	Name and the second sec		Bogenen Bours Coce Coller Curvane Curvane Aufor Aufor an Anto Aufor an Anto Autor and Anto Anto Autor and Anto Anto Anto Anto Anto Anto Anto Anto	
	An attick Research 100, 1211-110- and Research 100, 1211-110- and Research 100, 1210 attack 1 10 theory appears Restarking 1 10 theory appears 1		ISA 17384 ISA 17384	New Adda		Bogenen Bour Coce Diritor Cananan Danann Alate at FMC Full at at FMC Full at at FMC Full at at FMC Full at at FMC Second	
	An ation Research 10%, 1010-104 An Research 10%, 1010-104 An Research 10%, 2010-104 An Research 10%, 2010-10% An Research 10% An Researc		ISA 17384 ISA 17384	Name and the second sec		Income Bours Costs Collina Collina Collina Allaret PNCC, Public Tutie et a PNCC, Public Tutie et a PNCC, Public Costs Costs Costs Costs Costs Costs Costs Costs	
	An ation Research 10%, 1010-104 An Research 10%, 1010-104 An Research 10%, 2010-104 An Research 10%, 2010-10% An Research 10% An Researc		ISA 17384 ISA 17384	New Adda	аны. уна уна / Фланен 193	ingenes box convertibution convertibution data attal data attal data attal data data attal data attal data data attal data data attal data data data data data data data	
	An ation Research 10%, 1010-104 An Research 10%, 1010-104 An Research 10%, 2010-104 An Research 10%, 2010-10% An Research 10% An Researc		ISA 17384 ISA 17384	New Adda		Inspection Inser- Inser COM Convert	
на, 1923 Лана нарана К на сталина (1) 193 Алија (1) Лана нара 193 Алија (1) Лана нара нара 193 Алија (1) Лана нара нара 193 Алија (1) Лана нара нара нара нара 193 Алија (1) Лана нара нара нара нара нара нара нара	An ation Research 10%, 1010-104 An Research 10%, 1010-104 An Research 10%, 2010-104 An Research 10%, 2010-10% An Research 10% An Researc		ISA 17384 ISA 17384	New Adda	аны. уна уна / Фланен 193	Inspection Inser- Inser COM Convert	
	An ation Research 10%, 1010-104 An Research 10%, 1010-104 An Research 10%, 2010-104 An Research 10%, 2010-10% An Research 10% An Researc		ISA 17384 ISA 17384	New Adda		inspenn horis COS Dafte dafte Ad all and Ad	

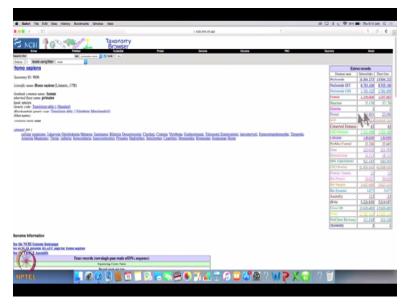
Now if you zoom into 1 of the locations, you will be able to see the variations associated with this gene, darker the colour, more deadly the variation. So notice this particular variation. As you will notice, this is in a very dark colour. This indicates a very deadly variation and if you zoom into this particular section, you will notice that this also has a PubMed citation. So now let us scroll upwards.

### (Refer Slide Time: 08:10)

• C E		i noturin shigar			10
					Test test
Showing Current Ibms.				Table of contents	
342 hemoglobin subunit alpha 2 [ homo sapir	ers (human) ]			Burmary	
w (D. 3040, updated on 34 Oec 2011				Generalis surface	
lumnary			1.1	Conserve regions, transpripte, and products Extransprise	
				Bibligrapy	
Midel Systeet 118-2 provide to 1000 Oficial full Name Interception substitutions) provide to 100	-			Phendypes	
Niney series HORCHONG-824				Veldon	
fee realed EnsemblishScoolsribbsbl html:141850	<ol> <li>Wear BTTHUMG00000008804</li> </ol>			Patheous from Bollysleme	
Senetype protein.colorg Rolling violant PEVEINED				Parways for socyating	
Departum Hono saciente				Beneral gane information	
Lineage Eularyota Menana, Drorena, Carrina, 1 Alaolanoan an 1401: 1404-72	Vestinae, Evaluation, Vannaia, Eutonia, E	Primales, Ingroposisti, Galantini, Heniniaas, Fump		Marken, Honology, Gene Onbiogy	
Burnnary Thehuman aiple giosin gine cluster lical		1 5- zela - peeudoeta - mu-peeudoelpi-a 1-alpha-2 - apha-1 - theia - 31		Central automotivemention	
		one and theirstone. In J hey differ spriftantly over he 3'untransisted region a sight chains contributewith delta chainsto constitute 1984.2, which with 1987		NCB Reference Sequences (Reflect)	
males up the remaining 39, of a kell heres		the genesias workes deletions of \$60,1682 and 1600; some rendeletion of	pha Balassonia hase	Mended sequences	
makes up the remaining 25 of adult turns also been reported. (provided by RefSeq	34 2008	yha ganas as wallaa debilans of bolt HBA2 and HBA3, soma rundebilan d	pha Balassettin hara	Mediad sequences	
maka uphra remaining 34, of aduk kana akobeernepored, (provide by Roffie), anapatan intermision mos we are asset tet three is noone (http://plain.bace.edu/bacetres.html	34 2008	pha genesi as wellas deolars of bith 1652 and 1654, some mendeleten é na unar in te apra gezin coales. Account di traver by auf it from	pha Balasomin have		
males valitie remaining 35 of adult home also been reported. (provided by Raffile), - mesanen information mos: we are avaine that there is recorded	Ju 2008 ancy among Gambark, Ratties, and ether trowsers acout the gam	pha genesi as wellas deolars of bith 1652 and 1654, some mendeleten é na unar in te apra gezin coales. Account di traver by auf it from	Pahalassesin kan	Appointing	
mates up the mession of N of Add home attached memory and a sector of the Add attached memory and a sector was not next an account pot (glain begau du/boarineru.ton) bits (glain begau du/boarineru.ton)	Ju 2008 ancy among Gambark, Ratties, and ether trowsers acout the gam	pha genesi as wellas deolars of bith 1652 and 1654, some mendeleten é na unar in te apra gezin coales. Account di traver by auf it from		Anthoneticse Loca-pech Datases	
natas sata smaling IA of akah tong internet and a management. (prodel hy Adlia, n metaton internetiae hot rigital bagas du/bashman, hot Johnège al	Ju 2008 ancy among Gambark, Ratties, and ether trowsers acout the gam	pe gens a voltes desters d'un 1992 aut 1993, une notation à n sept n ne apre gene come Aussing o more specif à rever 422) 1922/1922/1933, 1942 1954, [1354: 215]	1.1	Addisoratilities Locus-specific Databases Genome Bingewaters Genome Stella laisean	
nutas adra constra 9.4 d ad trans tensor un una secondaria de la dificia constanto municipal de la dificia constanto registra de la dificial de la desarro de la decentra britadege el Sensaria pontezt castino: %2133	Ju 2008 ancy among Gambark, Ratties, and ether trowsers acout the gam	pe gens a voltes desters d'un 1992 aut 1993, une notation à n sept n ne apre gene come Aussing o more specif à rever 422) 1922/1922/1933, 1942 1954, [1354: 215]		Addisoratilities Locus-specific Databases Genome Bingewaters Genome Stella laisean	
nutas adra constra 9.4 d ad trans tensor un una secondaria de la dificia constanto municipal de la dificia constanto registra de la dificial de la desarro de la decentra britadege el Sensaria pontezt castino: %2133	Ju 2008 ancy among Gambark, Ratties, and ether trowsers acout the gam	pe gens a voltes desters d'un 1992 aut 1993, une notation à n sept n ne apre gene come Aussing o more specif à rever 422) 1922/1922/1933, 1942 1954, [1354: 215]	1.1	Additionalities Local-pech: Databases Genome Browsers Denome Data Veneor	
nationa autore conservatore (24 of ad. Abree neurosciente en entre en esta esta esta esta esta esta esta esta	24 2005 may among unimbers, fersteag and men inserver sources action the gas be gase order, televisers to centromere, for these loc is	na genera a valanda dantaka nationa chinin 1964 gara (1964), sura nationala na nationa ina alang generalisme in kunoming na mana tapanat a manar (1922) (1927) (1962) (1964), (155%) (155%) Bin 1960) in Garanni O'r Loaden	1.1	Asstructions Units-genth Entreses Denome Brignsans Bennes Brig Newer Hey Viewer	
Index adva server (p 4 of a d how eventses how eventses (p 4 of a d how between the server of a d how eventses) between the server of a d how eventses between the	32 2001 way among conclusion, we show you concer the gase to be gase under, whilewere to central wave, for these too is	na gana a unite advinin d'An 1400 ari 1404, sun substan d anne in la gana coast, noong a sova spart a reger 2023/MCP(HC), MAJ, 2004, [D1/a 202] Ber HB/B is Source Dr Loadin 1 AC, 2004:04 (2004.7370)	1.1	Addrauctina Usus-pech Datases Genome Selo Seen May Yeer Variation Yeers (SRC57,r13)	
Initia anto remarka (24 of al Almo evenante interneter (24 of al Almo behaviore entre a sub estate a constante behaviore estate behaviore estate behavi	24 2005 may among unimbers, fersteag and men inserver sources action the gas be gase order, televisers to centromere, for these loc is	na genera a valanda dantaka nationa chinin 1964 gara (1964), sura nationala na nationa ina alang generalisme in kunoming na mana tapanat a manar (1922) (1927) (1962) (1964), (155%) (155%) Bin 1960) in Garanni O'r Loaden	1.1	Addressel Addresses Data-pech Detesses Descente Sela lasses May Yeare Marketer Verse (SRCA) 7/15 Variation Verse (SRCA) 5/15	
notes and a contract (\$ 14 of A form contract on the second (b) of the second (b) o	34 2009 The gale colds, whites the contract of a gale to gale colds, whites to contract, the trace too is Assembly JACOM (J) TCC1_2000(145.20) JACOM (J) TCC1_2000(145.20)	ng pena ang series adatang darih 1464 ang series dari ng series d	1.1	Astronomical December December Benners Bate Insee Martine Vision (SACC) 7135 Validari Yeare (SACC) 813 Validari Yeare (SACC) 813	
Initia anto remarka (24 of al Almo evenante interneter (24 of al Almo behaviore entre a sub estate a constante behaviore estate behaviore estate behavi	32 2001 way among conclusion, we show you concer the gase to be gase under, whilewere to central wave, for these too is	ng pena ang series adatang darih 1464 ang series dari ng series d	1.1	Antioustina United Statement General Bolt Steam Warden Yawa (SRD:57,r15 Warden Yawa (SRD:57,r15 Warden Yawa (SRD:57,r15) Bolt General Steam (SRD:57,r15) Exercisi	
Initia anto remarka (24 of al Almo evenante interneter (24 of al Almo behaviore entre a sub estate a constante behaviore estate behaviore estate behavi	3,200 19,200 19 pe ordi, vitanes, notes, ero mer sesen a per 19 pe pe ordi, vitanes to ordinario, for tea to to	ου μοτο και στη τολογική τη	1.1	Addressministe Undergenich Dimitations Genome Sprächkener Hang niewer Hang niewer Hang niewer Statischen Niewer (SPICA)28 (13) Einstendi Statischen Niewer (SPICA)28 (13) Einstendi	
Initia anto remarka (24 of al Almo evenante interneter (24 of al Almo behaviore entre a sub estate a constante behaviore estate behaviore estate behavi	June of the second seco	na pera a unitaria datanti di hi 1964 anti 2006, suna statatanti di 1962 (1962 (1964 (1964 (1966 (196	1.1	Autoreminis Using yearly Databases General Soft Island Way Years Markitani Years (SIROJS) r.(5) Variation Years (SIROJS) Variation Years (SIROJS) Using Constitution USIS Variation (SIROJS)	
note and resource grant of A calculation excession excession of provide to field provide to field and the second of the second provide second of the seco	June of the second seco	ne perso a unite a dataria del 11 404 a 2014 (11 400 a unite a dataria 14 2014 - 14 400 a gine a conservante a resulta que a seguinar a trave 14 2014 - 14 400 a faita a conservante a resulta dataria 14 2014 - 14 400 a faita a conservante a conservante a 14 14 2014 - 14 2014	t 1 Daa Vener Ma Vener	Adduction Units quels Dissess Descend Dissess Descend Dis basis Martines Martines Dissess (DDD) Matter Mark (DDD) Control Disses (DDD) 73 Execution Dissess (DDD) Control Dissess (DDD) 73 Execution Dissess (DDD) Dissess (DDD) D	
nink andre sampe i for Adaltan menter andressa in energie for Adaltan ping task keys dahtantes vite jaste andre andressa i for an exert i an exert i an exert i for an exert i for an exert i an exert i for	Julia	ne perso a unite a dataria del 11 404 a 2014 (11 400 a unite a dataria 14 2014 - 14 400 a gine a conservante a resulta que a seguinar a trave 14 2014 - 14 400 a faita a conservante a resulta dataria 14 2014 - 14 400 a faita a conservante a conservante a 14 14 2014 - 14 2014	c 1	Adduction Units quels Dissess Descend Dissess Descend Dis basis Martines Martines Dissess (DDD) Matter Mark (DDD) Control Disses (DDD) 73 Execution Dissess (DDD) Control Dissess (DDD) 73 Execution Dissess (DDD) Dissess (DDD) D	
nine and resource (Produktion metric and resource) (Produktion produktion (Produktion produktion (Produktion) Produktion (Produktion) Produkt	Julia	ne perso a contra de la della	( ) ) Daa Veen Vee Veen Veen Arres aa Solo	Autorative Loss gent Diseases Server Bolson Marcon Relations Marcon Relati	
note and resource grant of A calculation excession excession of provide to field provide to field and the second of the second provide second of the seco	Autors to gen ends when the servers to the to the top to gen ends when to servers to the top to a 1, 202 However, the servers to the top to 1, 202 However, top	ne perso a unit de la della d	( ) ) Daa Veen Vee Veen Veen Arres aa Solo	Autorative Loss gent Diseases Server Bolson Marcon Relations Marcon Relati	

You will see that this gene belongs to homo sapiens, obviously because this is human hemoglobin subunit alpha 2 gene. This particular organism belongs to the superkingdom

Eukaryota. To know more about this organism, we can go and click on to this link. (Refer Slide Time: 08:31)



And we can click on to genome to get more information about the genome of this organism. (Refer Slide Time: 08:00)

d Safari Ha Edit Vew History	Borismantes Minutiner Helps												18	🖓 8 kj 🤁 815 🗰 Thuốc	D mm C
								di nimah	40					1.5	0 0
		0 1	NT_10006-1	(#000811	10.8	11	6.116		2.08	1.00	11		las nos.		
		er 1	NL,000111	Capitoleu 1	10.5	47	4.943	·	1.04	8.771					
		01 1	NC.20008-1	040000712	142.11	42	1.00	- 4	1.05	2.90	-				
			NL.0008-1		100.0	43	4.573		2.16	2.00	10				
		0 0	NC, 20070-7	casses 1	1214	-18	1.257		1.70	8.101	107				
			NC 200H H		18.8		4.187	10	1.07	1.68	810				
			NC.0001215		10.8	-	5.848		2.79	2.52					
			NC, 20015-1	CM00064-3	11.3	41	1.948	1	1,06	1.00	477				
		e =		Calmorn 1		-4.4	6.965		1.74	1.100	180				
			N. (891) 1		101.8	- 14	5.436	1.1	1.78	1,81	100				
				C#000871.3	81.54	4.1	4.453		1.50	1.00	61				
		Ca 11	ar, search		****	-11	1.812		1.00	1.45	24				
			41,00011		10.0	24	4.452		1.56	1.40	14				
			A. 10010-1				1.754		1.00	1.74					
		1.1	NO. ORIGINA		16.71	42	1.245	11		194	20				
			N. 20085		11.41		2.488			1.00	34				
			-				1.750	110		1.00					
				(100000.2	17.0		100			471	-				
			NL 19981	diverse 1	140			1.0							
									1.04		1.00				
	<ul> <li>Granussian</li> </ul>		-	88818	88.84	810		1080							
	• Estenal Recouter														
	To are tone INCE - Genome &												Bugert Cartar		
	GETTING STARTED		REBOURCES			KOPU				<b>EATIRE</b>		NOB INFORMATION			
	ACB Examine ACB High Marcal		Daminuis & Bra Bra & Loffware			h.168				Canada Teo Autolog Inte	nite Resilie	Research at VCB1			
	NO TODAY		Put. & Tonings				- Barrar			barders.		NUM Your Ling			
	Tarring& Tutotala		Innen & Beck	Land Land			inut.			Adverse	Separce.	NOR / TP Ba			
	Submit them		Correct Copress			5,A51				one ture	seon chirela	NUM OF TAXABASE			
-			General & Wells			Autor				Vac View		MOBI in Tertar			
1			Genome & Max			1000				Auren Der		NOB in Youlube			
TSEC.			Lanature			inter a				viture V					
275						-				1000.0					
1		_	Second Adda	_		MO	-	_	_	lease of	End Active		_		_
			1		100	1	1.14	1.	1	n 🗆	) ( <b>. 1</b> 🔒		0		
	👢 🗑 🕑 🖳 📕	8													

Now as you know that this particular gene belongs to chromosome 16. So we go to chromosome 16 and you will see that the size of this chromosome is 90.34 megabases and it in codes 1920 genes. So with this information, you will be able to complete this table.

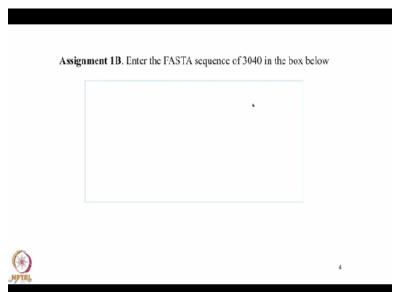
(Refer Slide Time: 09:10)

Official full name         hemoglobin subunit           Date of last update         24.bcc-201           Gene symbol         HBA2           Gene type         protein codi
Gene symbol HBA2
Cana tana na
Gene type protein total
Location 16p13.3
Size of gene (bp) 864
Organism Homo sapi
Superkingdom Eukaryot
Size of Chromosome (bps) 90.34Mb
# Genes on chromosome 1920 gcne

Now let us go back to our assignment. What we have is a Gene ID 3040. NCBI had thus fill in all these details. The official full name, hemoglobin subunit alpha 2. Date of last update 24 December, 2017. The gene symbol is HBA2. It is a protein coding gene. The location is on chromosome 16 in the p arm at region 13.3. The size of the gene is 864 bases. The organism is Homo sapiens. In the superkingdom, Eukaryota.

The size of the chromosome 16 is 90.34 megabases which encodes 1920 genes. So this brings us to the end of the first part of this assignment.

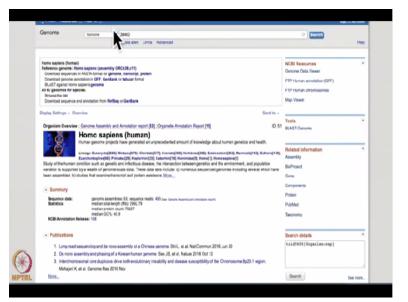
(Refer Slide Time: 10:03)



The next assignment is a very small one. You have been asked to enter the FASTA sequence of

this gene in the box below. So let us go to NCBI website again.

### (Refer Slide Time: 10:14)



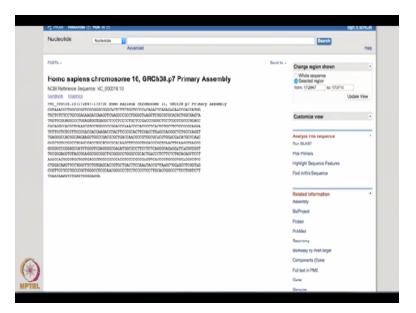
Let us type 3040 in the search box. We type gene and press search.

### (Refer Slide Time: 10:28)

					Managerer Manager (2007) (21 a 470
Genomic context				8.9	Variation Viewer (3RCh37,p13)
Location: Np113			See HBA3 in General D	uta Viewey Map Viewer	
Exon count: 3					1000 Geromes Browser (GRCh37.p11) Encentral
Annotation release	Status	Assembly	Ov Location		Essentil
508	ourient	GRCN38.p7 (OCF.000001405.33)	15 NC. 100(16.10 (172847.17	3710)	0.04
105	previous assembly	GRCN37.p13 (3CF_000001405.25)	15 NC_600016.9 (222846273	(709)	
					Related information
	[100374]	Chromosame 16 - NC_000014.10	( s#x79 b		Order oDNA clone
			Hit		BicAssayby Target (Lat)
Gammin mailans for	asseripts, and products			1.9	BioAssay by Target (Summary)
					BoAssay, by Gene target
			Do to refe	rence sequence details	BicAssays, RNAI Target, Tested
lenomic Sequence: ht		senence GRCX38.¢7 Primary Assembly	Gio to <u>rele</u>		
Jenomic Sequence: No		venerce GRCX38.¢7 Primary Assembly	Go to rele Git to nucleotide: Grachic	rence sequence details	BicAssays, RIA Target, Tested BicProjects
Denamic Sequence: No	C 00016. 10 Cironosone 16 Pe	- Co Child of Poney Assembly 🔒		rence sequence details	BioAmays, RiAi Target, Tested BioProjecta
2 2 m. onerate	C 00016. 10 Cironosone 16 Pe	-) \$ \$ 10 - 0 - 0 <b>0</b> 11	Ge to sucleatible: Graphic	rence sequence details	BioAssays, RiAi Target, Tested BioProjects BioSystems
10 \$5 m_000000.00 · 1 /ve [210,000	C.00016.10 Civenosane 16 Pe	- 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Ge to nucleotide: Grachic	rence sequence details	BicAssays, RHAI Target, Tested BicPhyleca BioSystems Bicity
5 5 m_000000.00 · ) 740 (10,000	C 000016.10 Civenceone 16 Pe final:  271:394  270.9 Lina Annotacion Palesare	• \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	Ge to suchotide: Gradie Note: State State project project	•	BicAssays, 1964/ Target, Teeted BicPhysiecs BicPhysiecs BicPhysiecs Bichy Bichy CODS
244 (212.449) 244 (212.449) 244 (212.449) 244 (212.449) 244 (212.449) 244 (212.449)	C. 000116. 10 Civenosane 16 Re final: pricine (2750 Lette Anson al Los Relieuse	- \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$ \$	Ge to suchotide: Gradie Note: State State project project	·	BoAssuya, RVA Targat, Testad BoPhysics BoDystems Book C2D5 CintVir
2) Sym_evenesse · ) ree [siteee image, NOI Rom sep] meets / meets / image, Transition / image, Transiti	0,00016.10 Oveneene 16 Pe feet prove prove loss Amotacion Beleave		Ge to suchotide: Gradie Note: State State project project	•	BoAnsays, RVAI Target, Tested BoPhysics BoDysteins Boots 2005 Contrar Conserved Demains
55 m. evenes	C 00016 10 Civenesane 10 Pr Intel (271.744 (2719) Less Antot al Los Release as 91 a aglese Antotal Los Rel		Ge to suchedide: Gaselic Normal Stress (Construction) Theory (Cons		BoAsaya, RAA Target, Yested BoProjects BoRs Boxts CCOS CCNV Conserved Dumains 4000a
55 m. evenes	C 00016 10 Civenesane 10 Pr Intel (271.744 (2719) Less Antot al Los Release as 91 a aglese Antotal Los Rel		Ge to suchedide: Gaselic Normal Stress (Construction) Theory (Cons		BoAmayi, RivA Targit, Yashed BoPipeca Bodystens Bodys CODI Carlor Conserved Domans eVilue Full tank to PAO
Symposesson     Tree     pream	C 00016. 3 Cranossons & R prices prin con Antoria (on Holeane on Allow Antoria (on Holeane and frame Antoria (on Holeane and frame Antoria (on Holeane and frame Antoria (on Holeane and frame Antoria (on Holeane antoria (on Holeane) and frame Antoria (on Holeane antoria (on Holeane) antoria (on Ho	(2)         (2)         (3)         (4)         (5)           (2)         (2)         (2)         (2)         (2)           (3)         (3)         (3)         (3)         (3)           (4)         (4)         (4)         (4)         (4)           (4)         (4)         (4)         (4)         (4)           (4)         (4)         (4)         (4)         (4)         (4)           (4)         <	Ge to nucleotide: Granico Non * 2 prese P		Bokawy, Rók Targit, tested Bolhytecs Bolhytecs Dote Dote Conset Conset Conset Dote Pall benin PinC Pall benin PinC
Provense - 1     P	Coloris d Cransson II In Inti prose pro- prose pro- ngrane Antoriation National ngrane Antoriation National ngrane Antoriation National pro- ngrane Antoriation National Pro- pr		Ce to such office: Casulton None property processing Processing Control (Control (		BoArasyi, RAA Targit, Yashad BoPingeis Borlysteine Boots CostAr Costeved Demains dotte Fail tool & PMC Fail tool & PMC Fail tool & PMC
Provense - 1     P	Coloris d Cransson II In Inti prose pro- prose pro- ngrane Antoriation National ngrane Antoriation National ngrane Antoriation National pro- ngrane Antoriation National Pro- pr		Ce to such office: Casulton None property processing Processing Control (Control (		BoArasyi, Rick Targit, Yashod BoPippics BoDytama BoDytama CODI Cathor Conserved Domans divit Full tool In PhoC Full tool In PhoC, receleding Functional Class Cane neighbors

So this brings us to the same page. If you scroll down, you will be able to see FASTA here. If you click on FASTA, you will get this particular nucleotide sequence for this gene.

### (Refer Slide Time: 10:39)



Let us copy this sequence and enter it in the box. So this brings us to the end of assignment 1B.

### (Refer Slide Time: 10:51)

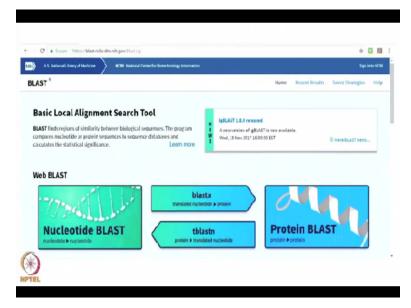
не, (8001.1.1113877.13718 вако нарвая сталаная 14. (8018.1.0./) т. (закул. актиАнскизаностичикановских истеплового констративания с нарвая нарвая с нарвая с нарвая нарвая с нарвая с нарвая нарвая с нарвая нарвая с нарвая с нарвая нарвая с н нарвая с нарвая с	

Alright, so you are now familiar with looking at genes and sequences and what all information one could obtain just by doing simple bioinformatic analysis, right. So now let us go one step further. Let us study how to compare sequences of one particular gene between different organisms and to see how simpler or different they are. Similarly, we will also compare the protein sequences and I am sure, you will understand that the proteins are much more complex than DNA.

And by comparing protein sequences across different organisms, one could actually find out lot

of information which is biologically relevant and it may sometime gives us the clues for even process like evolution how it has happened. Such sequence similarity searches can actually also help us to do to study evolutionary relationship.

Or it can allow us to identify the proteins with similar functions. So this type of analysis which you can do using bioinformatics tools, is known as multiple sequence alignment and it can be best performed using various open access tools including the Clustal W or Clustal omega which we are going to demonstrate you in the tutorial section. (Video Starts 12:03 - Video Ends 22:10) **(Refer Slide Time: 12:03)** 



Before we move on to the next assignment, I would like to talk to you about a very popular feature of NCBI known as BLAST. BLAST stands for basic local alignment search tool. It allows you to take an input sequence and compare it to a database to see if anything similar has already been found. It can be very useful for studying unknown sequences or if you want to see how similar your sequence is to other sequences. The most commonly used feature is nucleotide BLAST. So let us click on NBLAST which is Nucleotide BLAST.

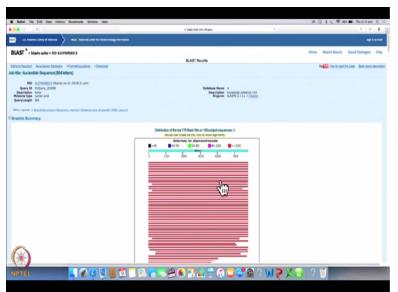
#### (Refer Slide Time: 12:40)

•• C 10		ii Bast.ntili.nin.nhate 5	h 0 6
		ing stream	ap. 1
ILAST *- Mart	n wite	Bandard Nucleoide ILAIT	Home Resert Results Saved Strategies He
test tests	Right Right		
Enter Query Sec	Lance	NASY proptore each nucleolde bioloses using a surieolde party mon-	from your looknoth
		hur Gery shranga Q	
6AAC0CACCAT93 1501010101000 A090109005A078 15010069-5500	SCHCTCSC00000000000000000000000000000000	bosoo h	
O, upload file Job Title	these fits to be arrested		
	Erier's rescriptive life for your BAST search		
Align two or more	Denomi plus Transvitel		
Choose Search Drahese Organism Tortunal Esclush October	Incare assume the twenty force 0-1 Water general and twenty Moue 0-1 Dire braness Processing and the second force of the twenty Hill descend the assume force, or Refuse, force assume (other, or Refuse, force other assume (other, or Refuse, force) branes (other, or Refuse, force) or Refuse (other, or Refuse, force) or Refuse (other, or Refuse, force) or Refuse (other, or Refuse, for Refuse, force) or Refuse (other, or Refuse, force) or Refuse, force) or Refuse (other, or Refuse, for Refuse, for Refuse, for Refuse, for Refuse, force) or Refuse (other, or Refuse, for Refus	nen (rm.) D = anno a	
Linit to	Construction of the		
Program Select Optimize for	ng trouglyo promi sepanoshifu) Pani seanopoli Adaros pani seanopoli Adaros pani seanopoli (dae, panin) Inuan Medicine sepanopolitika, (sea Genera Livry sepanos (pil	n bitt Abber	
	Sumewhat similar sequences (blashr)		
	Encoses BL/ET algorithm 📦		
BLAST	Search detailable Nuclearials collection (solid) using Magel	hine (Bydenius for highly similar announces)	
Sportfirm parameter	n		
6			
and the second	emork of the National Janes of Multicine		Supportcenter Maling Int
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1		I 🖬 🗆 🖏 🗃 🗮 🖲 🏹 🖬 🛨 🕼 🖵 💕 🕲 🤈 😡	

This will take us to the BLAST page where you will see that we need to enter a query sequence which is nothing but our input sequence which we need to search against a database. Let us take the FASTA sequence of the hemoglobin alpha chain that we just pasted. You can choose a human database if you know your protein is a human protein or a mouse database or if you want to see results from other species, you can choose other.

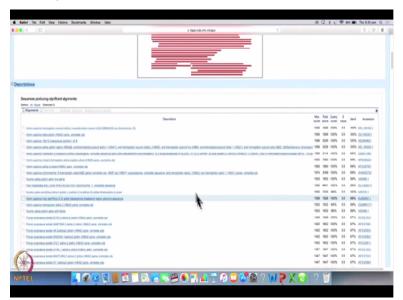
We will click on other. This dropdown menu here gives us a list of many databases. We will select nucleotide collection which is a combination of all nucleotide databases. Now let us click BLAST.

#### (Refer Slide Time: 13:46)



Here you will see a whole bunch of red lines which BLAST has found. These represent sequences similar to our input sequence. Top one here, is our query sequence which goes from 1 to 864 base pair. The colours here tell us how closely the results match our query or input sequence and the length of these bars tells us how much they line up with our query. So you will see there are some which line up perfectly, where some do not.

#### (Refer Slide Time: 14:35)



When you scroll down, you will get a list of sequences producing significant alignments. You have a maximum score, a total score and a query cover which is the percentage match with the query and an E value which is the measure of how likely something can occur by chance. Lower the E value, better the result. You will notice that the first few matches have 100% query cover which means that they are 100% identical. While you will also see that as you go lower, down this table, you will find matches with different species.

#### (Refer Slide Time: 15:22)

Ballet The Edit View History Boolenaries Window Help			18 🖓 8 v. 🗢 18 10 Tu bot on O
• • · · ·	ii Saat rcbi.nin.nihgov		1.0
Nighmenta			
Bisectore - Tex Status			Thed & Bescriptore
Formosasierie V. woglobii subunit ripha2 recombination region (LOC 108866512) on thromosome 18 Income (b. 82) (981951), Legan 1898 Numer of Nature: 1			
Regel is 200 to			Related Information
burn Front Manifian Jam Shund			Gener associated penedetaris
(536 Hz(844) 0.0 864,864(.00%) (/864(0%) Plus,Plus			
Derry 1 Charles Income a second secon			
DATA 41 CERCENSIFICITICSCONCARACCERCITIANSCONCERSIONALISE 124			
Report 2641 CONCERNMENTER TO THE AMAGE AND TRANSPORTED TO THE AMAGE TO THE TABLE TABLE TO THE TA			
party 121 and an			
IF STE 2791 GEDOGRAFGETGIOGRAFIANGETEIGRAGOETSGRAAGGETGOETSCETTE 2145			
Query (#) Transformer to the construction and the advertised of the second seco			
BATY 241 DECEMBER AND			
IN 283 GROUNACCENCITATION TO COORDINATION CONTRACTOR (1885			
Query 101 ACBADICTNTTODICATTOR/CTNBC/CONTENDED/ACTINGOUGHC 160			
mpri 1991 ADAMICHCTOCOLATIONOCHARCERORICHICCARTIAROROAC 1945			
Revery 141 minute and an and a second			
Biger 2941 BERALADOROCOMOLOCITACIONI DECAMOLOMOLICACIONALICACIÓNICAL DOS Barry 421 BIETROCOCOMOLOCITACIÓN DECAMOLOCIDADA TO MIL			
B (0) 100 OCTORECTORECTORECTORECTORECTORECTORECTOR			
many all approximation contraction to the second se			
Biger 104 Addressed concentration restriction and Parcetter A. 1125			
Gerry 141 Dischard A court To consumm According to Section 2 (1997)			
Genry 491 CENTRAL CONTROL OF CONTROL AND A TOTAL TRANSPORTATION OF 104			
pary 441 ODOTHODODIARTCACOUNCACHURCHARTCONSCIL 120			
mpro 1941 1946			
Derry 121 TERMANOUS TACTOMETICATION AND TRACTOMETICATION TO A STREET STATE			
Date in the second seco			
Apr. 104			
Par Ball TRATMACTORITORICAGE BEA			
Second in the main transmission in			
VD/			
NPTEL 🔜 🖉 🖉 🖉 🗐 🔂 👝 🔊	2월 🌒 📆 👔 🚢 🕋 🦳 🥼	🧖 🖓 🖓 🛄 👅 🗸 💭	0 🎬

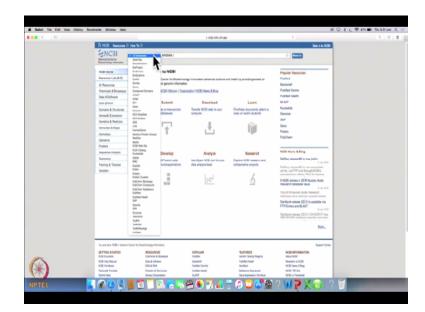
The alignment section which is further below here, gives you the actual nucleotide sequences and matches for each of the results. For example, in the first case, there is a 100% match; therefore, no gaps. So now let us get back to our assignment.

(Refer Slide Time: 15:41)

Basic Local Alignme	ent Search Tool (BLAST)
Accession number	NP 000508.1
Protein Name	
Organism Name	
Sequence Length	
Locus	
No. of BLAST hits	
PDB ID	

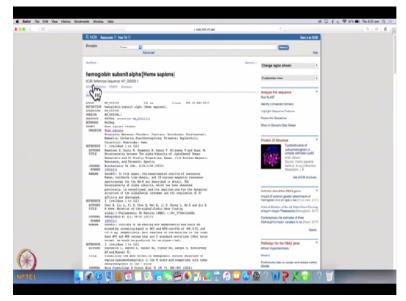
You have been given a protein accession ID which is NP 000508.1. You have to get details for this protein using NCBI. We have done this before for a gene. We will now does the same for this particular protein. Let us type the accession ID in the search box.

(Refer Slide Time: 16:03)



So let us type the accession ID in the search box and now we will select protein from this dropdown menu and click search.

# (Refer Slide Time: 16:18)



So here you will see a very similar to what we saw for the gene. You can actually get all the details for this particular protein. So the protein name is hemoglobin subunit alpha. The organism name is Homo sapiens. The sequence length is 142 amino acids. The locus is NP 000508.

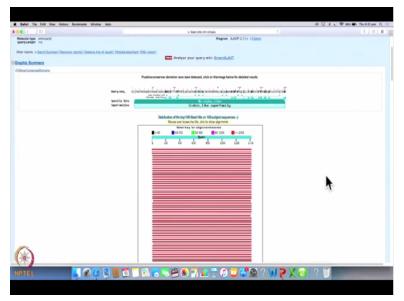
### (Refer Slide Time: 17:56)

A water of the function		Edit View History Busimaris Virulow Help			28 🖓 8 🗸 🛡 875 🖬 Tu bâs per Q
With Note: Read Parket Note: Read Parket   (Mail * ) A data parket   Standar Parket Standar Parket Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket   Standar Parket Note: Note: Note:   (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * ) A data parket (Mail * ) (Mail * )	•		ii tiat.rtbi.rin.shgov	8	h 0 6
Image: Image		a claray of healthing 🔪 must many management for something			
Improve month       Market month <t< td=""><td>LAST - M</td><td>ntp wite</td><td></td><td></td><td>Home Revent Annuals Saved Strategies Help</td></t<>	LAST - M	ntp wite			Home Revent Annuals Saved Strategies Help
Inter Comparison       Inter Trapper standing registry registry       Inter Trapper standing registry registry       Inter Trapper standing registry registry       Inter Trapper standing regist			Standard Protein BLAST		
Control of the statement o		Bint Rint			
Star Landar Landar Jack (19)     Star Landar Lan	Enter Query S	lequence	BL/STP program assoch potein ankdasse unter a protein gamp more,		Read Jour Tookings
A market     Image: Second Secon			Gery streng 📦		
A Yes Construction for the Def A Transform of	M*_00408.1				
State					
State       Image:	Align two or my				
Airson generation of the stand product on a stand product	Phones Same	a fai			
Appendix         Que (B, 407) Provine spece spece (B, 407)         Company	Entres Guerg	Patentel prodet asquences(pat) Protes (sets term proteinignet) Metagemente proteinignet, no) Transequente seruppi-transmitter, no			
Appendix         Que (B, 407) Provine spece spece (B, 407)         Company	Program Sala	alian .			
	Algorithm	C Maeto (proxim printindEAET) O Mil-BUAST (Printion-Apwolin Invariant NAMY) O Mil-BUAST (Protein R2 Initiated NAMY) O DELYA-BUAST (Domain Enhanced Loosup Time Accelerated N	LAIT		
	BLAST		(Binde (proteingrobin BIABT)		
A laster We	Aportition parame	in .			
Proceedings of the Automa Using of Hebicine	0	ndemurk of the National Library of Hedicine			Suportante Huling ist

And now when we run a BLAST search, so now let us run a BLAST search and select PDB here.

The accession number is given in this particular query box and we click on BLAST.

### (Refer Slide Time: 17:22)



So this particular page will give you details about the BLAST search as you see here, very similar to what we saw earlier. Just as you saw earlier, there are 106 BLAST hits on 100 subject sequences and you will find similar bars here which represent similar sequences.

### (Refer Slide Time: 17:59)

Angunana pankung sipilikasi dipananka		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Series (a) And all (a) And		
Description         Sec. Strateging System Syst		
Description         Series (a) And all (a) And		
Description         Description         Sec.         Sec. <th>scriptions</th> <th></th>	scriptions	
Description         Description         Sec.         Sec. <td></td> <td></td>		
Bit Properties         Control         Description         Note Name	Researce poduring algebrait dynamics	
Description         Immunol Team Control of the C	Saint // New Sainted 0	
Database         National		
Option         Note of the state of th		Iden Accession
Op 14: A Advances in the sensitive in the sensitive sensit sensiti sens sensitive sensitive sensitive sensitive sensiti sen		A.1381 2001
Opt. A short Thermatica Tablesian bases That Bit Storeback 10 to Storeback 10 to Store Bases         Store Bases <td< td=""><td></td><td></td></td<>		
On the investment limits from the structure of the	Charl B. Allia Patho I. Surfa Henopolis Dealaring Protein Deals The Diructure Neuropolish 2016 129-1201 10	1005 2/2.8
Class. Conserve Toxics 2 Simple Liss. NameGrand         97         98         96         76         96         All states Conserve Toxics And All states Conserve Toxics An		
Oak Amore Generation And B         80 <td< td=""><td>C Care A remogen Termin An International Statement (Statement in Street Arter (Statement in Street Arter) Arter (Statement in Street Arter)</td><td>885 386,6</td></td<>	C Care A remogen Termin An International Statement (Statement in Street Arter (Statement in Street Arter) Arter (Statement in Street Arter)	885 386,6
2014. Distant (2) In South Physics and (2) Phy	Out A suspen limits is standards for limits and all limits for limits in the suspen limits is standard limits and limits	1005 1004,4 1005 1004 A 1005 1004 A
4 Jan 86 Jan 1991	Open A most Research Servers and Advanced & Landschult Research Advanced Research Advanced         84         84         60	995 398,5 1025 1034,8 995 382,5 995 6920,8
	Open A synthesis         1981 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 13           Open A synthesis         1881 144 1499 A. 2018 14           Open A synthesis         1881 149 A. 2018 14           Ope	A,346, 799 A,452 2021 A,528, 699 A,528, 699 A,500 699
	Open A host Company Dense A host Company Dense A host Dense host Dense A host Dense A host Dense A host Dense A	A, BARC A19 A, 1224 A001 A, 1226 A19 A, 1226 A19 A, 1226 A19 A, 1226 A19 A, 1226 A19 A, 1226 A19
	Open A synchronization for the first open of the first open of the first open to the first open to the first open open open open open open open open	40.000 2000 2000 2000 2000 2000 2000 200
Conta A company         All 144 values / 10 Annume (a) Annum (a) Annume (a) Annume (a) Annum (a) Annume (a) Ann	ا A Nan Sandd Bynnw ( Safet Sand Sand Sand Sand Sand Sand Sand Sand	
On the investment limits from the structure of the		
Oper. Journal Transformation (Stream Content)         39         39         36         56         662.5           Oper. Journal Transformation (Stream Content)         38         38         56         166         56           Oper. Journal Transformation (Stream Content)         38         38         166         166         166           Oper. Journal Transformation (Stream Content)         38         38         166         166         166           Oper. Journal Transformation (Stream Content)         38         38         166         166         166           Oper. Journal Transformation (Stream Content)         38         38         166         160         166           Oper. Journal Transformation (Stream Content)         38         38         166         160         160           Oper. Journal Transformation (Stream Content)         38         38         160         160         160           Oper. Journal Transformation (Stream Content)         38         38         160         160         160	Case A receptor Toronte An International Conference on A Damager Conference on A Damager Description American Description Description A	885 386,6
Oper. Journal International Tympital Line Astron Obsent         30         30         405         406         602,57           Oper. Journal International Tympital Line Astron Obsent         30         300         405         406,50         300         405         406,50           Oper. Journal International Tympital Line Astron Obsent         300         300         405         406,40         400         400         400,40           Oper. Journal International Tympital Line Astron Obsent         300         300         405         400,40	Case A receptor Toronte An International Conference on A Damager Conference on A Damager Description American Description Description A	885 386,6
Op AA, States for Constraint         80         20         40         50         40         50         40         50         <	Case A receptor Toronte An International Conference on A Damager Conference on A Damager Description American Description Description A	885 386,6
Oak Allow Conduction And Allow         80	<ul> <li>Om A prospect/hereits an standber sees and Aussian Da Darrende Street Alex (exclored) darrende street et al.</li> <li>Om A Britsen Offenmenden Stalland Guerrende Street Alex (exclored) darrende street et al.</li> <li>Dar A Britsen Offenmenden Stalland Guerrende (E) bande et al.</li> <li>Dar A Britsen Offenmenden Stalland Guerrende (E) bande et al.</li> </ul>	995 3846,6 1005 103H A
Opt. A procession and a disconsecuration and Add Add Add Add Add Add Add Add Add A	<ul> <li>Om A prospect/hereits an standber sees and Aussian Da Darrende Street Alex (exclored) darrende street et al.</li> <li>Om A Britsen Offenmenden Stalland Guerrende Street Alex (exclored) darrende street et al.</li> <li>Dar A Britsen Offenmenden Stalland Guerrende (E) bande et al.</li> <li>Dar A Britsen Offenmenden Stalland Guerrende (E) bande et al.</li> </ul>	995 3846,6 1005 103H A
2010. D 10.04 (2) IN 2001 (2) IN 2001 (2) IN 2011 (2)	One any sequencing and the second s	1005 1004,4 1005 1004 A 1005 1004 A
4 201 85 248 80 100.4	0         0	995 398,5 1025 1034,8 995 382,5 995 6920,8
4 201 85 248 80 100.4	0         0	995 398,5 1025 1034,8 995 382,5 995 6920,8
	Open A symposite Damping Dampin	A,346, 799 A,452 2021 A,528, 699 A,528, 699 A,500 699
11. 201 855 2x80 1005 X 100 X	Open A most Description of Control (Control (Contro) (Contro) (Control (Control (Control (Contro) (Control (Contro)	A, BARC A19 A, 1224 A001 A, 2242 A19 A, 2242 A19 A, 2242 A19 A, 2442 A19 A, 2444 A19 A19 A, 2444 A19 A19 A19 A19 A19 A19 A19 A19 A19 A19
	Open Amount Termina Amount (a) A	A, BARC A19 A, 1224 A001 A, 2242 A19 A, 2242 A19 A, 2242 A19 A, 2442 A19 A, 2444 A19 A19 A, 2444 A19 A19 A19 A19 A19 A19 A19 A19 A19 A19

And if you scroll down further, you will see that the first hit is identical to your query sequence and you will take the PubMed ID from here which is your PubMed accession ID. So now let us go back to our assignment and fill the table.

(Refer Slide Time: 18:18)

Basic Local Ali	nment Search Tool (BLAST)
Accession number	NP 000508.1
Protein Name	hemoglobin subunit alpha
Organism Name	Homo sapiens
Sequence Length	142 aa
Locus	NP_000508
No. of BLAST hits	106 Blast Hits
PDB ID	1BZL A

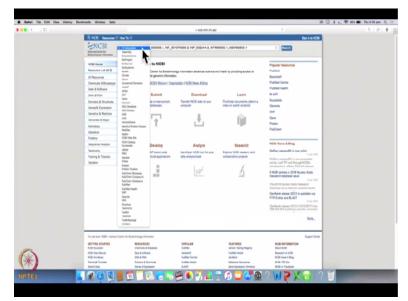
The details for this protein are the protein name is hemoglobin subunit alpha. The organism name is Homo sapiens. Sequence length is 142 amino acids. The locus is NP 000508. The number of BLAST hits 106 and the PDB ID.

(Refer Slide Time: 18:37)

	Multiple Sequence Alignment (MSA)
Accession No.s	a) NP_000508.1, b) NP_001070890.2, c) NP_032244.2, d) AFM89055.1, e) ABW88850.1
Evolutionary rela	fed

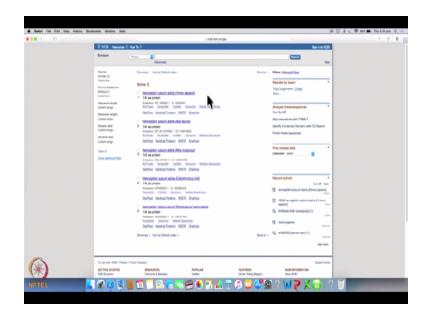
In the fourth part of the tutorial, you have been given a list of accession IDs. You have to find out how the sequences are evolutionarily related. For this, we will type all the accession IDs in the search box on NCBI home page.

(Refer Slide Time: 19:06)

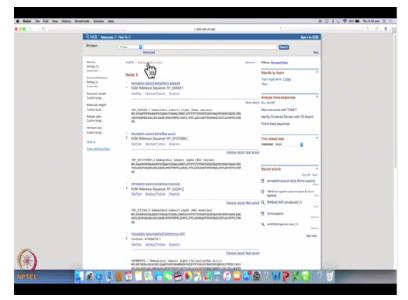


Now we will choose protein from the dropdown menu and we will click search.

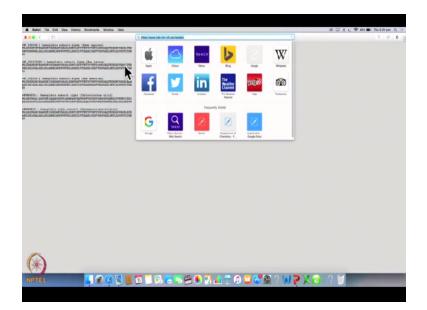
# (Refer Slide Time: 19:18)



We will now be able to see all 5 proteins on a single page. So we do not have to go to each protein individually to take the FASTA sequence. Simply click on summary and click on FASTA. (Refer Slide Time: 19:40)



Or we can also go to FASTA text so that we get all sequences on 1 page. (Refer Slide Time: 19:51



So let us copy them, go to Google browser and type Clustal Omega.

# (Refer Slide Time: 20:14)

d Salai Ila	Edit View History Busimaria Mindow	r Nels			# 3 8 Tu 521 pm 0 =
••• (	0		40 AC A	C	1.0.6.8
Google	clustalomega	٩			II Bern
	A Book Image Video New	en More Settings Toris			
	About 123,502 sensite (028 samesh)				
	Custal Omeran - Multivite Beavenne https://www.ebi.ac.uk/Techs/maintustal/ = Costal Deway is a multiple seguence algometry secures algometric of costagents assumes. En Codoports or Phylograms.	progam. I produces			
	Help & Documentation Chatel Orega is a nulticle sequence alignment program. 1	Custal Criega PAQ Custal Smeta is smultile second algomet program 8			
	NUM REALTS THAT HE AC IS A				
	CustalW2 < Nulligie Sequence Aligi Imperivementation de la setter de la setter Tre CustaWI service tres tens retres la set tequera Agrenet tos page rogeners arge alignette se reconnect trigs MACL1 et Wi contact a varbe fieldeal int	ees imile services, pleasevial he Multiple mens we accommend charaet charge har driv.			
	Custal Onega - fast, accurate scala www.clastif.org/onega * Hone of the David Compa milital separate d	able multiple sequence alignment			
	Custal Onega, CustaW and Cluste wave result any + Ag 30, 2012: Anreaging of the clustet series of y entryleases are a dynamic.	alX Multiple Sequence Algoment programs (shastid amaga, cluster) and cluster) for			
	gin the fix sume certaining your query from Bot	STALW- GenomeNet MBUShave Prot. COX. 45.08194, and 50.00457 Or fait Perameters. Paintes Alignment Perameters. For me see . Sop. Penalty Number of Top Disgonals.			
*		erre und 2 Bootomsten fir multiple anguaren 2 Fondine 2 I republiske 225ettigen 3 bee allen e hane been mary instructions a' Chudal find are			
NPTEL		N 🖉 🖉 🗂 🖏 🥫 📢	🗒 🌒 🎵 🚢 😷 🕖 😂 (	🍪 🤉 🖬 🄁 🗶 📵 🛛	? 🕤

Clustal Omega is a multiple sequence alignment program. It allows us to perform multiple sequence alignments of divergence sequences. Evolutionary relationships can be seen via cladograms or phylograms.

# (Refer Slide Time: 20:39)

Edit View History Boolimarks Virolinar Halp.	104.4	6	# G # < @ #1 # %
A EMIL-EX Serve	ne Ansarch Teining Industry About a Q	54050 🌐 Her	iun+
Clustal C	Omega		
RUT Jam And Pers	ce Normeni - Custé Drega	• fecture 40	-
Sonico Rathema Via senint you hatit		ind on 16h April 2018. Alternatives can be found at <u>Exercises</u> , <u>BM</u>	
Multiple Si	equence Alignment		
or more sequences. Fo	multiple sequence alignment program that uses seeded guide trees an in the alignment of two sequences please instead use our pairwise seque		hree
STEP 1 - Enter your	ool sam eligenup to 4000 exquences or a waximum file size of 4 idB. r input sequences		
Intersr pate a set of PROTEN			
repartices in any sup	oponecilomae:		
Or patrineta file come	an ha , in fa minter		
This website-uses cookie	ea. By continuing to browse this alth, you are agreeing to the use of our	ste cookies. To find out more, ses our Terms of Uter.	

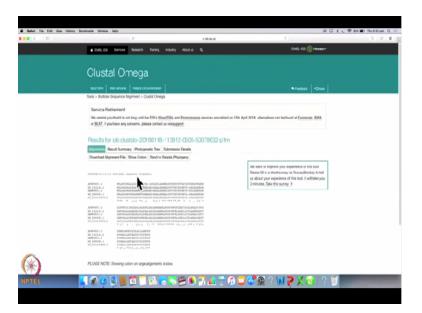
Copy the FASTA files into the box here.

# (Refer Slide Time: 20:50)

<. D	latoy Bosimada Vindee Help i 10.41.0	ait □ i v ♥ ans ■b The 5d per 0, = 5 5 6
	Igod firm Web services Heelpit Decumentation	• Fector Chan
	Important note: This tool can align up to 4000 sequences or a maximum file size of 4 MB.	
	STEP 1 - Enter your input anguences	
	(inter or paste a set of	
	PROTEIN	
	sequences in any supported format	
	340.0005041 hemopolis suburit sjots Biomo apsime] Mr.SPROKTWAWARKSHAREEGUKALINM.SPPTTCYIPHEDLSHGAQWCHCKKWACALTM VanAddermaal.sbulleli-seneit/PCPVPTRLBHELLSTLAH-LINETTINV MLEHPTLANETVLTIK YR	1
	УКР. 0010708002 hemoglikih suburit apha Bostauraji им Зикирилимина (Кираминатири), Блик за при тистиче (С.Б.К.Б.К.К.К.К.К.К.К.К.К.К.К.К.К.К.К.К.	
	V, (2023 TH: Owner he in the second	
	STEP 2 - Set your parameters	
	outhurt Formal	
	Glutial w/o numbers	
	The address participy will fulfil the searce of most users.	
	More options   gClick have if you want to view or sharpe the default settings,	
	STIP 3 - Submit your ob	
	Demonitors systema (Titus shits boun if your event in: See notified by serval writers the results are available)	
	bare.	
2	If you plan to use case earliest during a course please contact us.	
	Peace eadtre PKU setore sexuing neighton our support statt	
2	- many many real second and second second second	

Keeping all other default parameters constant, click submit.

# (Refer Slide Time: 20:59)



So as you see the sequences of all the 5 proteins are lined up against each other.

# (Refer Slide Time: 21:13)

	4 DAL-D	Services Research Taining Industry About us Q	EMBED 🌐 HINKKY	
	Clusta	l Omega		
		eo services Peep's Dicumentation	• Fectual Strate	
	Toos > Multiple 3	leguence Alignment > Clustal Ornega		
	Senice Ret	ivement		
		ehatil is not long until the Elfi's Misel/DBJ and Promotenaise services are net have any concerns, please contact us viasupport.	nd on 15h April 2018. Alternations can befound at Extension. IBBA	
		iob clustalo-20180118-113812-0505-53078632-p wit Summary Philogenetic Tree Submission Details	Im	
		wart Summary Photogenetic tree Submission Decem wart File Hole Colors, Sand In Simple Photogeny		
		haltigete respectiv ditament	We want to improve your experience of this tool. Dearso till in a short-survey on fluoresystomize trial at about your experience of this boil. It will take you	
	AP99915.1 RF 1122-6.2 Ap99925.1 RF 1050-6.1 RF 1050-6.1		2 minutes. Biler this survey 2	
	APRIPATI, 1 107.1322-8.2 ARR1010.1 107.1010-0010.1 107.1010-0010.2	CENTER LE TELES DE LES ANALYSIS DE LE TELES DE LE T CONTRICIS DE LE TELES DE LE TELES CONTRICIS DE LE TELES DE LE TELES CONTRICIS DE LE TELES DE LE TELES CONTRICIS DE LE TELES DE LE TELES CONTRICIS DE LE TELES DE LE TELES CONTRICIS DE LE TELES DE LE TE		
	APRENDS, 1 09,3122-8,2 appleters, 1 09,1012-8,1 09,1012-84,1 09,1012-0444,2			
)	PULASI NCTE:	Shnaing mitm on lega alguments is sina		

If you select show colours, you can visualize variations and gaps in the sequences.

### (Refer Slide Time: 21:240)

Okushal Osusana	6	Δ.
Clustal Omega		
Tees > Vultale Exquence Nigment > Costd Onega		
Senice Retirement		
We remind you that it is not long until the E3/'s MoseDBA and Promoterwise services are refind on 15th April 2018. Alternatives o or BAXT. If you have any concerns, please contact is via suggest.	an befound at Exprenate, BVA	
Pasults for job clustelo-20180118-113812-0505-50078632-p1m		
Algorents Result Summary Philogenetic Time Submession Deals		
Phylogenetic Tree The is a Machteria control transmittent distance connections		
Downbad Phylogenetic Inee Bala Breschimspin @ Geologian C Real		
Anamosi di soluti del solutione del solutione Anamosi di solutione del solutione		
Tiee Data		
02 24506.1.0.4080. 0.20120 0.20120 2.00100 2.000000 2.000000 2.000000 2.000000 2.000000 2.0000000 2.0000000 2.0000000 2.0000000000		

Now click on the phylogenetic tree to get an idea about the evolutionary relationships. What you see here is the cladogram. This is a neighbour joining tree without distance corrections. Now let us copy this into our document.

(Refer Slide Time: 21:41)

	Multiple Seque	ence Alignment (MSA)	
Accession No.s	a) NP 000508.1, b) NP 0010	70890.2, c) NP 032244.2, d) AFM890	055.1, e) ABW88850.1
Evolutionary related	1		
		AFM89055.10.50148	d- Callorhinchus milii
		NP_000508.1 0.06455 NP_001070890.2 0.05516 NP_032244.2 0.04581 ABW83850.1 0.04574	a- Homo sapiens b- Bos taurus c- Mus musculus e- Peromyscus maniculati
A and B are close	ly related	NP_001070890.2 0.05516 NP_032244.2 0.04581	b- Bos taurus c- Mus musculus
	ly related hat related to A and B	NP_001070890.2 0.05516 NP_032244.2 0.04581	b- Bos taurus c- Mus musculus

So let us paste the cladogram in this document. You will see here that A and B are closely related while B is somewhat related to A and B, while C and E are distantly related from all. So this is how you will get an idea about evolutionary relationships. This brings us to the end of the bioinformatics assignment 1. Hope this was very useful. Thank you.

#### (Refer Slide Time: 22:02)

Bioengineering: An Interface with Biology & Medicine



How to use NCBI to study genes and evolutionary relationships

THANK YOU



Alright so in conclusions I am sure you are now convinced that just by doing bioinformatic analysis using some open access tools, you can get lot of understanding of biologically relevant information. In this assignment, we slowly built your, you know, confidence, you try to do some information for looking at the gene, its sequences, its sequence homology and its different similarity with that sequences.

We then try to explain you how computational tools could be used to study the evolutionary relationship just based on the protein sequences. In another lecture in the coming week, we will also try to study how to visualize protein structures and protein-protein interactions using some of the bioinformatic tools. We will also study the pathway analysis and we will learn how to perform molecular assimilation and docking experiments to study the protein drug interactions. Thank you and see you next week.