



Databases



Dept. of Computer Science and Engineering University of Rajshahi www.ru.ac.bd

Dr. Shamim Ahmad

What is a database?

Which databases are important for molecular cell biology research?

How is information processed in databases?



Biological databases use different organizing principles Hyperlinks connect records in different databases

Databases are organized collections of information

Information is stored in records



Bioinformaticians work with computer scientists to set up the database structure

Curators review and link records within and between databases

The information in databases ultimately derives from experimental data



Curators will process the submissions and link entries in different databases

Database Examples in Bioinformatics

- Primary (archival)
 - GenBank/EMBL/DDBJ (seqs) - PDB
 - (protein structures) Medline
 - (literature)

7

- IMEx databases (protein interactions)
- Secondary (curated)

 - RefSeq (seqs)
 - UniProt SwissProt (seqs)
 - Taxon (taxonomy)
 - PROSITE (binding sites)
 - OMIM (genetics literature/reviews)
 - IMEx databases (protein interactions)

Sequence Databases

- > DNA
 - NCBI: GenBank -> RefSeq
 - ♦EBI: EMBL

NCBI: GenPept

➢ Protein

8

TrEM BL= "translated EM BL"

K EBI: UniProt: TrEMBL -> UniProt: Swiss-Prot

National Center for Biotechnology Information www.ncbi.nlm.nih.gov

European Bioinformatics Institute www.ebi.ac.uk



GenBank

- · Direct submissions from individual laboratories
- Bulk submissions from large-scale sequencing centers.

Direct submissions to GenBank

- BankIt, which is a Web-based form
- Sequin stand-alone submission program •

Upon receipt of a sequence submission

- GenBank staff examines the originality of • the data
- Assigns an accession number to the • sequence
- Performs quality assurance checks

GenBank

- GenBank has become an important database for research in biological fields
- Ii has grown in recent years at an exponential rate by doubling <u>roughly every</u> 18 months
- Entries are retrievable by Entrez
- Downloadable by FTP

Bulk submissions

- Expressed Sequence Tag (EST)
- Sequence-tagged site (STS)
- Genome Survey Sequence (GSS)
 High-Throughput Genome Sequence (HTGS)



Growth in GenBank base pairs, 1982 to 2007, on a semi-log scale

GenBank

An expressed sequence tag (EST) is a short subsequence of a cDNA sequence

A sequence-tagged site (or STS) is a short (200 to 500 base pair) DNA sequence that has a single occurrence in the genome

Genome Survey Sequences (GSS) are nucleotide sequences similar to EST's that the only difference is that most of them are genomic in origin, rather than mRNA

Complementary DNA

- Complementary DNA (cDNA) is DNA synthesized from a single stranded RNA (e.g., messenger RNA (mRNA)
- DNA is derived from mRNA, so it contains only exons, with no introns.
- cDNA is also produced naturally by retroviruses (such as HIV-1, HIV-2, simian immunodeficiency virus, etc

Complementary DNA

- cDNA is often used to clone eukaryotic genes in prokaryotes.
- When scientists want to express a specific protein in a cell that does not normally express that protein, they will transfer the cDNA that codes for the protein to the recipient cell.
- cDNA is also produced naturally by retroviruses (such as HIV-1, HIV-2, simian immunodeficiency virus, etc

Top organisms inGenBank (Release 191)[9]	
Organism	base pairs
Homo sapiens	16,310,774,187
Mus musculus	9,974,977,889
Rattus norvegicus	6,521,253,272
Bos taurus	5,386,258,455
Zea mays	5,062,731,057
Sus scrofa	4,887,861,860
Danio rerio	3,120,857,462
Strongylocentrotus purpuratus	1,435,236,534
Macaca mulatta	1,256,203,101
Oryza sativa Japonica Group	1,255,686,573
Nicotiana tabacum	1,197,357,811
Xenopus (Silurana) tropicalis	1,249,938,611
Drosophila melanogaster	1,119,965,220
Pan troglodytes	1,008,323,292
Arabidopsis thaliana	1,144,226,616
Canis lupus familiaris	951,238,343
Vitis vinifera	999,010,073
Gallus gallus	899,631,338
Glycine max	906,638,854
Triticum aestivum	898,689,329

ID AF115338 standard: DBA PRO: 551 80. AC AF15339; SF AF11539; SF AF11539; DF AF1539; DF AF1539;	GenBank Flat	t File	10 AV115330 standardy EMDA 1907 591 B9. AC AV153393 50 O-2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 - 2000 50 O-2000 - 2000	EMBL Fla	t File
<pre>0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0</pre>	Header	• Title • Taxonomy • Citation	<pre>0 investment flowteness include and the set of the</pre>	Header	•Title •Taxonomy •Citation
<pre>m res 1</pre>	Features (A	A seq)	Margine Production Margine Production Margine Production <td>Features</td> <td>(AA seq)</td>	Features	(AA seq)
ogt gctettag ataaattgog tgagaaattt goaggogaga etgaaaett a $g \qquad 591$ //	DNA Sequer	nce	optgetetag atalattgoj tgagalatt gelggogaga etgalaetta g 591 //	DNA Seq	uence

UniProt

The UniProt consortium

European Bioinformatics Institute (EBI), UK Swiss Institute of Bioinformatics (SIB) , Switzerland Protein Information Resource (PIR), USA

• EBI and SIB

- Swiss-Prot and TrEMBL databases
- (European Molecular Biology Laboratory)
- PIR
 - Protein Sequence Database (PIR-PSD)

UniProt

- UniProtKB
 - Swiss-Prot
 - TrEMBL)
- UniParc
- UniRef.

UniProt

- UniProtKB
 - Swiss-Prot (reviewed, manually annotated entries)
 1.93 B Amino acid, 1914.3
 - TrEMBL (unreviewed, automatically annotated entries) 14 M Amino Acid, 1914.3

UniProt

- UniProtKB (UniProt Knowledgebase)
 - Information extracted from scientific literature and <u>biocurator</u>evaluated computational analysi
 - To provide all known relevant information about a particular protein

UniProtKB (Annotation)

- Protein and gene names
- Function
- Enzyme-specific
- Subcellular location
- Protein-protein interactions
- Pattern of expression
- Ion-, substrate- and cofactorbinding sites
- Protein variant forms produced by natural genetic variation, RNA editing,

UniProt: Swiss-Prot – An example of curated, reviewed annotation

> Incorporates:

24

- ✓ Function of the protein
- ✓ Subcellular localization of protein
- $\checkmark {\sf Post-translational}\ {\sf modification}$
- ✓ Domains and sites
- ✓ Secondary structure
- ✓Quaternarystructure
- ✓ Similarities to other proteins
- \checkmark Diseases associated with deficiencies in the protein
- $\checkmark {\tt Sequence \ conflicts, variants, etc.}$



Swiss-Prot

- Reliable protein sequences associated
 - With a high level of annotation
 - Description of the function of a protein
 - Its domain structure
 - Post-translational modifications
 - Variants
 - Minimal level of redundancy
 - High level of integration with other databases.

TrEMBL (Translated EMBL Nucleotide Sequence Data Library)

was created to provide automated annotations for those proteins not in Swiss-Prot

UniProt Archive (UniParc)

- It is a comprehensive and nonredundant database,
- Proteins may exist in several different source databases, and in multiple copies in the same database
- Information extracted from scientific literature and <u>biocurator</u>-evaluated computational analysi
- Each sequence is given a stable and unique identifier (UPI)

UniProt Archive (UniParc)

- INSDC EMBL-Bank/DDBJ/GenBank nucleotide sequence databases
- Ensembl
- European Patent Office (EPO)
- FlyBase: the primary repository of genetic and molecular data for the insect family Drosophilidae(FlyBase)
- H-Invitational Database (H-Inv)
- International Protein Index (IPI)
- Japan Patent Office (JPO)

UniRef

The UniProt Reference Clusters (UniRef) consist of three databases

- Protein sequences from UniProtKB
- Selected UniParc records

The Protein Data Bank(PDB)

- Crystallographic database for the threedimensional structural data of large biological molecules,
 - Proteins
 - Nucleic acids.
- The data, typically obtained by
 - X-ray Crystallography
 - NMR spectroscopy,
 - · Cryo-electron microscopy

The Protein Data Bank(PDB)

Experimental Method	<u>Proteins</u>	<u>Nucleic</u> <u>Acids</u>	Protein/Nudeic Acid complexes	Other	Total
<u>X-ray</u> diffraction	106595	1820	5471	4	113890
<u>NMR</u>	10296	1190	241	8	11735
<u>Electron</u> microscopy	1021	30	367	0	1418
Hybrid	99	3	2	1	105
Other	181	4	6	13	204
Total:	118192	3047	6087	26	127352

SNCBI Search All Databases All Databases All Databases HomoloGene MeSH NCBI Web Site NLM Catalog OMIA OMIM PMC Popse Protein Clusters Probene Toropound PubChem SiloAssay PubChem SiloAssay PubChem SiloAssay PubChem SiloAssay RubChem SiloAssay Taxonomy ToolKit ToolKit ToolKit All Databa PubMed National Center for Biotechnology Information Protein Nucleotide Nucleotide Cos ST ST ST STructure Cenome BioSpraiet BioSystems Books CancerChromosomes Conserved Domains dbGaP dbVar Epigenomics CertST CEVD Ext Sets Conserved Domains dbGaP dbVar CertST NCBI Home Site Map (A-Z) All Resources Chemicals & Bioassavs Data & Software DNA & RNA Domains & Structures Genes & Expression UniGene UniSTS 0---

NCBI maintains both primary and derivative databases We'll look at three of them

PubMed is the premier literature database in the world

Curators are responsible for data flow between the NCBI databases

Questions for today:

What is a database?

Which databases are important for molecular cell biology research?

How is information processed in databases?

Annotated nucleic acid sequences are submitted to GenBank from many sources, including genome projects, individual investigators, and other databases GenBank Nucleotide there is considerable REDUNDANCY in the sequences information Sequences are compiled to Automated RefSeq generate non-redundant translations of Non-redundant reference sequences nucleotide nucleotide and sequences otein sequenc Protein Amino acid sequences Experimentally determined amino acid sequences and information from other protein databases

Most records in the Protein database have been derived by automated translation of nucleotide sequences



On a larger scale: Genome projects have produced the reference sequences in nucleotide databases

(robots and computers do much of the work)

1. Pieces of chromosomal DNA are sequenced, each ~1000 bp long

<u>S cerevisiae</u> genome is ~12 Mbp - how many reads would be necessary to cover each base pair in the genome once? 3. Chromosomal sequences are analyzed for the presence of potential transcripts (open reading frames; ORFs)

Open reading frame



ORF-finding computer algorithms look for sequences that

begin with a methionine
methionine is separated from a stop codon in the same reading frame by a large number of <u>amino acids</u> (often 100, equiv. to 300bp)

GenBank NM_####### records are predicted ORFs

4. Protein sequences are computationally predicted from $\ensuremath{\mathsf{ORF}}$ sequences

GenBank NP_###### records

X-ray crystallography For determining the atomic and molecular structure of a

For determining the atomic and molecular structure of a crystal, in which the crystalline atoms cause a beam of incident X-rays to diffract into many specific directions



Nuclear magnetic resonance spectroscopy of proteins (NMR)

- It is a field of structural biology
- NMR spectroscopy is used
- To obtain information about the structure and dynamics of
 - proteins, and also nucleic acids,

Entrez Global Query Cross-Database Search System <u>CDS</u> Coding sequence; The locus name: Tthe first three characters usually designated the organism Accession Number [ACCN] Sequence Length Sequence Length [SLEN] 2500:2600[SLEN] <u>Molecule Type</u> The type of molecule that was sequenced. In this example, the molecule type is <u>DNA</u> Properties [PROP] biomol_genomic, biomol_mRNA {PROP} <u>GenBank Division</u> gbdiv_pri, gbdiv_est, Properties [PROP] human[ORGN] NOT gbdiv_est[PROP]

- 1) To retrieve records within a range of lengths, use the colon as the range operator, e.g., 2500:2600[SLEN].
- 2) To retrieve all sequences shorter than a certain number, use 2 as the lower bound, e.g., 2:100[SLEN].
- 3)To retrieve all sequences longer than a certain number, use a series of 9's as the upper bound, e.g., 325000:999999999[SLEN].

Molecule Type

The type of molecule that was sequenced. In this example, the molecule type is <u>DNA</u>

- genomic DNA
- genomic RNA
- precursor RNA
- mRNA (cDNA)
- ribosomal RNA
- transfer RNA,
- small nuclear RNA
- small cytoplasmic RNA.

 Modification Date

 Modification Date [MDAT]

 To retrieve records modified between two dates,

 1999/07/25:1999/07/31[MDAT].

 DEFINITION

 Brief description of sequence

 Coding region (CDS),

 Word [TITL]

 ACCESSION

 The unique identifier for a sequence record

 VERSION

 If there is any change to the sequence data (even a single base),

 the version number will be increased, e.g., U12345.1 ? U12345.2,

 KEYWORDS

 Word or phrase describing the sequence

 SOURCE

 Free-format information including an abbreviated form of the

Free-format information including an abbreviated form of the organism name, sometimes followed by a molecule type.

<u>Organism</u>

The formal scientific name for the source organism **REFERENCE** Publications by the authors **AUTHORS** List of authors in the order in which they appear in the cited article **TITLE** Title of the published work or tentative title of an unpublished work. < symbol indicates partial on the 5' end. Example: <1..206

symbol indicates partial on the 3' end. Example: 4821..>5028

The GenBank database is divided into 18 divisions:

- PRI primate sequences 1. 2.
- ROD rodent sequences MAM other mammalian sequences
- 3. 4. 5. VRT - other vertebrate sequences INV - invertebrate sequences
- 6. 7. PLN - plant, fungal, and algal sequences BCT - bacterial sequences
- 8. VRL - viral sequences
- PHG bacteriophage sequences SYN- synthetic sequences
- UNA unannotated sequences EST EST sequences (expressed sequence tags)
- PAT patent sequences
- FAI parent sequences STS STS sequences (sequence tagged sites) GSS GSS sequences (genome survey sequences) HTG HTG sequences (high-throughput genomic sequences) HTC unfinished high-throughput cDNA sequencing
- 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. ENV - environmental sampling sequences

FASTA format

- It is a text-based format for representing
 - Nucleotide sequences
 - Peptide sequences
- Nucleotides or amino acids are represented using single-letter codes
- The format also allows for sequence names and comments to precede the sequences.

- □ The simplicity of FASTA format
- Makes it easy to manipulate and parse sequences using textprocessing tools and scripting languages
 - □ R programming language
 - Python
 - □ Ruby,
 - 🗆 Perl

- A sequence in FASTA format is represented as a series of lines
- · Each of which should be no longer than 120 characters and usually do not exceed 80 characters.
- ٠ The first line in a FASTA file starts either with a ">" (greater-than) symbol or, less frequently, a ";" (semicolon) and was taken as a comment.

:LCBO - Prolactin precursor - Bovine : a sample sequence in FASTA format MDSKSSSQKESPLLLLUXQRVVSTPVCPNGPGNCQVSLRDLFDRAVMVSHYTHDLSS EMFNEFDKRYAQGKGFITMALNSCHTSSLPTPEDKEQAQQTHHEVLMSLILGLLRS WNDPLYHL VTEVRGMKGAPDAILSRAIEIEEENKRLLEGMEMIFGQVIPGAKETEPYPVWSGLPSLQTKDED ARYSAFYNLLHCLRRDSSKIDTYL KLLNCRIIYNNNC*

>MCHU - Calmodulin - Human, rabbit, bovine, rat, and chicken ADQLTEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGNGTID FPEFL TMMARKMKD TD SEEEI REA FRVF DKDGNGYI SA AEL RHVM TNLGEKLT DE EVDEMI REA DIDGDGQVNYEEFVQMMTAK*

ygi[5524211]gb]AAD44166.1] cytochrome b [Elephas maximus maximus] LQ.YTHGRNITYGSYLSYSETWN TGIMLLITMATAFMG YULPWGQMSFWGATVITNLFSAIPYIGTNLV EWIWGGFSVDKATLNRFFAFHFILPFTMVALG6VH_ITE, HETGSNNPLGLTSDSDKIPFHPYYTKDFLG LIIILILLLILSPDMLG6PDNHWPADPLNTPLHKPEWYELFAYAILBSVPNKLG6VLALFLSIVIL GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFLPIAGX TENV IENY

>SEQUENCE 1

- ADEQUENCE_1 MTEITAAMVKELRES TGAGMMDCKNA LSETNØDFDKA VQLLREKGLGKAAKKADRLAAEG LVS VKVSDDFTTAAMRPSVLSV EDLDMTTVENEV KALVAELEKENEERRRLKDPNKPEHK IPQFASRKQLSDATLKEAEEKIKEELKAQGKPEKIWDNITPGKMNSFIADNSQLDSKL TL MØQFV VMDDKK TVEQUTAEKEKEFGGKLKIVEFICF EVGEGLEKKTEDFAAEVAAQL >SEQUENCE_2 SATVSEINSETDFVAKNDQFIALTKDTTAHIQSNSLQSVEELHSSTINGVKFEEYLKSQI
- ATIGENLVVRRFATLKA GANGV VNGY IHTNGRV GVV IAAACD SAEVASKSRDLLRQIC MH

| Nucleic Acid
Code | Meaning | Mnemonic |
|----------------------|---------------------------------|-------------------------|
| A | А | Adenine |
| С | с | <u>Cytosine</u> |
| G | G | Guanine |
| т | т | Thymine |
| U | U | Uracil |
| R | A or G | puRine |
| Y | C, T or U | <u>pYrimidines</u> |
| к | G, T or U | bases which are Ketones |
| м | A or C | bases with aMino groups |
| S | C or G | Strong interaction |
| w | A, Tor U | Weak interaction |
| В | not A (i.e. C, G, T or U) | B comes after A |
| D | not C (i.e. A, G, T or U) | D comes after C |
| н | not G (i.e., A, C, T or U) | H comes after G |
| v | neither Tnor U (i.e. A, C or G) | V comes after U |
| N | ACGTU | Nucleic acid |
| - | gap of indeterminate length | |

| Amino Acid Code | Meaning |
|-----------------|-------------------------------------|
| A | <u>Manina</u> |
| x | any |
| 8 | Argining |
| N | Assassing |
| D | An provide - and M |
| В | Aspartic acid (D) or Asparagine (N) |
| с | Cysteine |
| | gap of indeterminate length |
| E | Glutamic acid |
| z | Glutamic acid (E) or Glutamine (Q) |
| Q | Glutamine |
| G | Glycine |
| н | Histidine |
| 1 | Isoleucine |
| L | Leucine |
| 1 | Leucine (L) or Isoleucine (I) |
| к | Lysine |
| м | Methionine |
| F | Phenylalanine |
| Р | Proline |
| 0 | Pyrrolysine |
| U | Selenocysteine |
| s | Serine |
| T | Threonine |
| • | translation stop |
| w | Tryptophan |
| Y | Tyrosing |
| v | Valine |

| vucleotide Nucleotide • | | Contraction of the second s |
|-------------------------|---|---|
| Advanced | | Reining |
| ACCCARACACATA | T, Nucleotide | |
| TGTAGCTTACCACACCG | The Nucleotide database is a collection of PDB Genome, gene and transcript sequ | of sequences from several sources, including GenBank, RefSeq, TPA and
ence data provide the foundation for biomedical research and discovery |
| Ising Nucleotide | Nucleotide Tools | Other Resources |
| luisk Start Guide | Submit to GenBank | GenBank Home |
| <u>A0</u> | LinkOut | Ref5eg Home |
| icio | E-Utilities | Gene Home |
| ienBack FTP | BLAST | SRA Home |
| leffica FTP | Batch Entres | INSOG |
| | | |





| < → C | | w.ncbi.nlm.nih.gov/nucco | e/advanced | | \$ |
|------------|---------------------------------|-------------------------------------|--|---|----------|
| 🗄 Apps 🚺 | Suggested Sites 🛄 মান্সচন্দ নাই | te - Gurgan | | | |
| NCBI | Resources 🕑 How To 🕤 | | | shamm_cst_My.NCE | Sign Out |
| Nucleotide | Home Help | | | | |
| Nucleotid | e Advanced Search Buil | lder | | | |
| | (Coding region (CDS) Al | ND 2500 2600[SLEN].) | | | |
| | Edd | | | Clear | |
| | Builder | | | | |
| | All Fields | Coding region (| DS) AND 2500 2600[SLEN] | Show index list | |
| | AND All Fields | • | | © O Show.ndex.list | |
| | Search or Add to histor | α | | | |
| | History | | | Download history Clear history | |
| | Search Add to builder | | Query | Items found Time | |
| | e4 Add Se | earch 2500:2600[SLEN]. | | 1449603 08.45.43 | |
| | | D- D- | Contraction of the local division of the loc | The second se | 1. 1848 |











| Use th | e builder below | o create your search | | | |
|---------|-----------------|--|-------------------------------------|---------------|--|
| Edit | | | | Clear | |
| Builder | | | | | |
| | All Fields | * | Show index list | | |
| AND 1 | Al Fields | | O O Show index list | | |
| History | | | Download history C | licar history | |
| Searc | h Add to builde | Query | Items found | Time | |
| 1 | Add | Search ("2017/1/1"(Date - Entrea) : "2017/1/1"(Date - Entrea)) | 2304 | 08:18:33 | |
| 1 | El Add | Search ("2018/04/01"[Date - Entrez] : "3000"[Date - Entrez]) | 74783 | 08:12:31 | |
| | | | | | |

| s 🚺 Suggested Sit | es 🔲 🕾 | নারপ সন্মিত | - Gurgan | | | | | | |
|--------------------------|----------------|-------------|--|-----|--|---------------|-------------------|-------------------|---|
| Hirt: cikking the "Searc | n" butter with | tout any te | arms letted in the search box will transport you | | | | Literator | Saved Searches + | |
| to that database's home | 2000 | | | | Collections | | | | * |
| b Bibliography | | | | | Colecter Name | . darte | Settings/Sharing | 7/2# | |
| ny ananographiy | | | | 24 | Emortes | R 0 | O Ethate | Standard | |
| | Your be | loanty | contains no Herrs. | | thy Bitkopracty e | g 0 | O Ethate | Standard | |
| | | 0.000 | the second second | - 1 | Citras Castors e | . 0 | O Entrate | Standard | |
| | | | Internation and Annual Contraction | _ | | | t.ta | tage Collectors a | |
| Recent Activity | | | • | × | Filters | | | | * |
| Tre | Database | Type | Term | | and state of the second se | | | | |
| 06:15-AM | Published | search | 12017/1/11Date - Entrez) 12017/1 | | Filters for PubMed | | | | |
| 05:13:AM | Published | record | And champes in drowing resided to | | You do not have an | active titler | for this delabase | | |
| 00:12 AM | Publied | reach | 12010/04/01/Tone : Entred . 10000 | | ALLINAS | the period to | L CADICIANA. | | |
| 06:07.AM | Books | record | Publied Help - Publied Help | | | | | Hanage Ellers | |
| 06:04 A85 | Books | record | Entrez Hela - Entrez Hela | | | | | | = |
| Visiterday 3.55 PM | NUCKION DR | record | Saccharotryces cerevisiae TCP1-beta | | SciENcv | | | | |
| Yesterday 11.11 AM | PMC | record | The SIMSS-FROT protein sequence dat | | Citck here to create a new CV | | | | |

Entrez Global Query page

- Literature
- Health
- Genomes
- Genes
- Proteins
- Chemicals.

Nucleotide

The Nucleotide database is a collection of sequences from several sources, GenBank RefSeq TPA PDB Genome Gene transcript sequence data

RefSeq: NCBI Reference Sequence Database

- A comprehensive
- Integrated
- non-redundant
- well-annotated set of reference sequences including genomic, transcript, and protein.

BioSystems

The NCBI BioSystems Database provides integrated access to biological systems and their component

- Genes
- Proteins
- Small molecules
- Literature describing those biosystems
- Other related data throughout Entrez.

Gene

- Gene integrates information from a wide range of species. A record may include
- Nomenclature
- Reference Sequences (RefSeqs)
- Maps
- Pathways
- Variations
- Phenotypes
- · Links to genome-, phenotype-



Nucleotide is selected

| C @ | © 🖗 https://www.ncti.nim.nih.gow/nuccore/?term=500CA 1 (110%) | 🗟 🖉 🖉 |
|--|---|---|
| S NCBI Resource | 😌 How To 🕤 | shamim_cat My NCBI Sign C |
| Nucleotide | Nucleotide - BRCA 1
Create alert Advanced | C Search H |
| The Nucleotide da
Species
Animals (165) | abase will include EST and GSS sequences in early 2019. <u>Read more</u> .
Summary + 20 per page + Sort by Default order + Send to: - | Filters: Manage Filters |
| Rants (254)
lungi (233)
holists (101)
lacteria (237)
lustomize | See BRCA1 (BRCA 1) BRCA1, DNA repair associated in the Gene database
brea 1 reference sequences Genomic (1) Transcript.(6) Protein (5) | Results by taxon
Top Organisms [Tree]
synthetic construct (#14)
Escherichie coli (225) |
| Nolecule types
penomic
DNA/RNA (1,305)
mRNA (533)
Customize | Pope 1 of the local sequence and the local sequence Anabidoass mains chromosome 4 sequence Sequence | Homo sapers (751)
Eotheromys melanogaster (62)
Crecitura beccarli (23)
All other taxa (1338)
More |
| | Accessor COODERT OF TODALL | Find related data |

Genbank is selected

| Search NCBI d | en 🗙 🔛 BRCA 1 | - Nucleon X | Arabidopsis their | × 🗧 Arabidopsis the | 🗙 🛃 DNA & RNA - | s= 🗙 🔛 🛚 | lational Center : 🗙 🕂 | | I X |
|--------------------------------|--|------------------|------------------------------------|-------------------------|---------------------------|------------|---|---------------|-------------------|
| (←) → C | ŵ | 00 | https://www.ncbi.nlm | .nih.gov/nuccore/CP00 | 2687.1?report=genbar | (90%) | 🖂 🖒 | III\. | ⊡ ≡ |
| S NCBI Re | sources 🖻 How To | | | | | | shamim_cst | My NCBI SI | an Out |
| Nucleotide | Nucleo | otide • | | | | | Searc | 3 | |
| | | | Advanced | | | | | | Help |
| O The Nucleo | 6de database will in | iclude EST a | ind GSS sequences in e | arly 2019. Read more. | | | | | |
| | | | | | | | | | |
| GenBank - | | | | | | Send to: + | Change region show | m | |
| Å Due to the s | ize of this record, a | nnotated fea | tures are not shown by o | lefault. Use "Customize | view" section to change t | he display | | | |
| Arabidor | sis thalian | a chro | mosome / se | nuence | | | Customize view | | |
| GenBank: CPN
FASTA Graph | 002687.1
105 | | 110301112 4 32 | quence | | | Basic Features
Default features
Gene, RNA, and CD | S features on | εy |
| LOCUS
DEFINITION | CP002687
Arabidopsis th | 185
aliana ch | 05056 bp DNA
romosome 4 sequenc | linear PLN 20- | JUL-2017 | | Features added by NO | BI
mains | |
| ACCESSION
VERSION
DBLINK | CP002687
CP002687.1
BioProject: PR | JNA10719 | | | | | Display options
Show sequence
Show reverse compl | loment | |
| RETWORDS
SOURCE | Arabidopsis th | aliana (t | hale cress) | | | | | Update \ | law |
| * Downloading L | arge Sequence: 1.618 | 18 | | | | | () P | ause 🔀 🤇 | Cancel |
| (2) | 6 🤉 | | 🍯 📐 💽 | 🚽 🕊 | | | DN = 50 | n - 0 | 00.22
18-12-18 |

FASTA format is selected

| C III | 000 | https://www.ncbi.nlm.nih.gov/nuccore/CP002687.3 | Preport=fasta (11 | 0% 🕑 😰 | 11/ |
|--|---|--|-------------------|------------------------------------|-----|
| O The Nacieobide data | daase will include EST an | d GSS sequences in early 2019. <u>Beat more</u> | | | |
| PASTA + | | | Send to + | Change region shown | |
| Arabidopsis t | haliana chron | nosome 4 sequence | | Continents viter | 100 |
| GenBark: CP002687 | 1 | | | Customer view | - |
| Gerflack Graphics
>CD002487.1 Arabia
HUNCOBRERENEEDER | depais thaliana ch | COLORIDA 4 RADIASCH
DESIDERESIESTESTESTESTESTESTESTESTESTESTESTESTESTE | | Analyze this sequence
Run ELACT | - |
| MINISTRATION OF THE OWNER OWNE OWNER OWNE OWNER | CONTRACTOR OF CONT | Acceleration of the second | | Pick Primers | _ |
| | | | | Highlight Sequence Peakares | |
| | | 103080300000000000000000000000000000000 | | Related information
Asserbly | |
| 1010110101010101010101010101010101010101 | 004000000000000000000000000000000000000 | 0010000334100000341000001010030000000000 | | BaProject | _ |
| 10010403030303030303 | 030302030303030303030 | 000000000000000000000000000000000000000 | | BieSample | |
| INTERPORT OF CONTRACTOR OFICACIONO OFICIA OFICON OFICIA OFICON OFICONO OFICIA OFICON O | INTERCOMPANY AND ADDRESS OF ADDRESS ADDRESS OF ADDRESS | MORDEN DEPENDENCING STORES | | Protein | |
| 000TTC3400ATTCSTC | BACCROSTOTTOGRAFCO | CORCORNTCT IN SKATTOSTASACCESSACSIC | | PubMed | |
| TOGAATCCTCGACCAGG | CAAAGAATTOSTOSACC | ACCOUNT OF ATTOTIC CONTENTS AND C | | Teconomy | |
| STATATCHITACHARATCH | SACECTESTGAAAAAAA
MATOSSCOATGASSAT | CADSSCRATCSTSSATASTATCAASSCRACOSAGE | | Component Of | |
| AGAGCCATGAGCODOS | ADDCATADOADOADOATO | LOCIALIZATION CONTRACTOR CONTRACTOR | | Full text in PWC | |
| and the second state of th | | | | (Contra) | |

Search for gip3 (G1P3, an interferon inducible gene)



| Search NCIII 🛃 BRCA 1 - No | Arabidopsin 1 | 🛿 Arabidopes : 🔛 DNA 8 | RMA E glp3-N × | Synthetic con | 🕝 g1p3 gene | | Received. | X |
|----------------------------|--------------------------------|-------------------------------------|--------------------|---------------|-------------|--------------|----------------------|------------------|
| e) → C @ | 0 0 A https://w | www.ncbi.nim. nih.gov /nucce | re/?term=g1p3 | P (150%) | 🖂 🕁 | | In E | D ≡ |
| Viruses (1) | | | | | | Result | ts by ta | ixon |
| Customize | See IFI6 (G | 1P3) interferon al | pha inducible pro | tein 6 in the | | Top O | rganisn
no sanie | |
| Molecule | Gene datab | ase | | | | synt | hetic co | instru |
| types
genomic | g1p3 referen | nce sequences Tran | script (4) Protein | (4) | | Pan
Mac | troglody
aca fasc | ytes (
cicula |
| DNA/RNA (12)
mRNA (17) | Items: 1 to 2 | 20 of 30 | | | | Pan
All o | paniscu
ther tax | us (1)
ta (4) |
| Customize | | | t Priny Page 1 | of 2 Next | > Last >> | More | | |
| Source
databases | Found 99 n | ucleotide sequence | s. Nucleotide (30 | 0) EST (67) | GSS (2) | Tind a | -Interal - | data |
| INSDC | Synthetic | construct Homo: | sapiens clone FL | H141188.0 | 1L | rind n | elated | data |
| (GenBank) (15) | interferon | alpha-inducible p | protein (G1P3) m | RNA, partia | al cds | Databa | ISC: | |
| RefSeq (15)
Customize | 393 bp lin
Accession: | ear mRNA
AY892940.1 GI: 60 | 554332 | | | Select | | |
| Sequence | Protein 1 | faxonomy | | | | | | |
| length | GenBank | FASTA Graphics | | | | | | |
| Custom ronan | | Independent of the other states | | | | - | | |

| e ⇔ e ⊕ | (i) (i) A https:// | ewenchi nim nih | | econstion/ | | 105 | 00 | IN D | |
|---|--------------------|--|---------------------|----------------------|--------------------------|------------------------------|--------|--|--|
| SNCBI Resources C How | то 9 | | | | | | shamin | n cst My NCBI Sign O | |
| SNCBI All
Natored Center for
Bolechickogy Information | Databases + | | | | | | | Search | |
| NCBI Home | Genes & | Expression | | | | | | Quick Links | |
| Resource List (A-Z) | Al | Databases | Downloads | Submissions | Tools | How To | | BioProject (formerly
Genome Project) | |
| All Resources | Databa | - | | | | | | BinSyclams | |
| Chemicals & Bioassays | Datatod | 585 | | | | | | Duogseus | |
| Data & Software | BioProj | ect (formerly G | nome Project) | | | | | and Phenotypes (dbGaP) | |
| DNA & RNA | A conect | ion of genomics,
datasets. This r | functional genon | scs, and genetics s | tucies and
terial and | sinks to their objectives an | | Cana | |
| Domains & Structures | provides | provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent | | | | | 0 | | |
| Genes & Expression | annotati | n, multiple inde | pendent submissi | ons, and the varied | nature of a | diverse data t | ypes | Gene Expression
Omnibus (GEO) | |
| Genetics & Medicine | which ar | e often stored in | different databas | 65. | | | | Database | |
| Genomes & Maps | BioSyst | ems. | | | | | | Gene Expression | |
| Homology | Databas | e that groups bio | medical Iterature | , small molecules, | and seque | nce data in te | ms | Omnthus (GEO) Datasets | |
| Literature | or biolog | nue renascrisnipi | <u> </u> | | | | | Gene Expression
Oranibus (GEO) Profiles | |
| Proteins | ClinVar | | | | | | | Online Mandaller | |
| Samianra Analysis | A resour | ce to provide a p | sublic, tracked rec | ord of reported rela | Bolated i | setween hum | an | Inheritance in Man | |

| Search Num catabases - NUM 2 | SUNARY | ne - 5% | Usice IniCi X | Home - Gene - | aron X | N Home - | avoyateris - NGB | · • | Contraction of the | |
|--|-----------|-----------|--|---|--|-------------------------|----------------------------|--------------------|--------------------------|--|
| -) -) C @ | 00 🔒 | https://v | www.ncbi.nim. nih | gov/guidé/dna-m | a/ | | 10% 🕞 | ☆ | IN ED | |
| NCBI Resources 🕑 Ho | e To 😒 | | | | | | sha | mim_cst_M | NCB Sign Out | |
| SINCEI Al
sultonal Center for
Indechrology Information | Databases | • [| | | | | | See | sch) | |
| NCBI Home | DN | A & RI | NA | | | | | Quick | Links | |
| Resource List (A-Z) | | All | Databases | Downloads | Submissions | Tools | How To | BioProje
Genome | et (formerly
Project) | |
| All Resources | _ [| antak a | | | | | | Databas | in of Chort | |
| Chemicals & Bloassays | _ | Jacaba | 363 | | | | | Genete | Genetic Variations | |
| Data & Software | | kssemb | Y | | | | | (dbSNP) | 1 | |
| DNA & RNA | | ames a | ase providing info
and other meta-d | trmation on the size at a statistical rec | ructure of assemble
orts, and links to or | ed genome
enomic sed | is, assembly
uence data | GenBan | R | |
| Domains & Structures | | | | | | | | Nucleoti | de Database | |
| Genes & Expression | | ScCale | ctions
d sat of matachet | tor sulture colle | tions museums b | otheria an | f other natural | PopSet | | |
| Genetics & Medicine | 1 | history c | oliections. The n | cords display co | lection codes, infor | mation abo | ut the collections' | RetSeq | Gene | |
| Genomes & Maps | | nome int | stitutions, and lin | ks to relevant dat | a at NCBI. | | | Referen | ce Sequence | |
| Homology | | BioProle | ect (formerly Ge | nome Project) | | | | (RefSeq | 1 | |
| Literature | | A collect | tion of genomics. | functional genom | nics, and genetics s | tudies and | links to their | Sequence | ce Read Archive | |
| Proteins | | esuiting | datasets. This r | esource describe | s project scope, ma | terial, and | objectives and | (SRA) | | |
| | | muni005 | on methololism to | pandant submissi | ons, and the varied | nature of a | tiverse data tunes | Trace Ar | ,chive | |



Select Blastx



Gip3 sequence found from FASTA

A TGCGGCAGAAGGCGG TA TCGC TT TTC TTG TGC TACCTGCTGC TCTTCACTTGCA GTG GGGTGGAGGCAG GTAAGAAAAAGTGCTCGG AGA GCTC GGACAGC GGC TCCGGGG TTC TGGAA GGCCC TGA CCTTCA TGGCCGT CGGAGGAGGACTCGCAG TCGCC GGGC TGCCC GCGC TGGGC TTCACCGGCGCC GGCA TCG CGGCCAACTCG GTGGCTGCCTCGCTGATGAGC TGG TCTGCGA TCC TGAA TGGGGGGCGGCG TGCCCCGCC GGGGGCTAGTGG CCACGCTGCAGAGCCTCGGGGCTGG TGGCA GCAGC GTC GTCA TA GG TAA TATTGGTGC CCTGCTGCAAGCCTCCGGGGCTGG TGGAGCAGCAGC GTC GTCA TA GG TAA TATTGGTGC CCCCACCCCAAGTATCTCGATAGTGAGGA GGA TGA GGAGTTG

| Search NCIII | BRCA 1 - No | Arabidopsis | 1 🛃 Arabidopeis 1 | DNA & RMA | g1p3 - Nucle | Synthetic core | S NCBLER: X | |
|--|--|-------------------|-----------------------|------------------------|--|---|---|------------------|
| ← → @ | ŵ | 🛈 🕑 🔒 http | ps://blast.ncbi.nlm.r | ihgev/Blast.cgi | | P 50% | 🖂 🏠 | IN ED E |
| Query
Description
Molecule by
Query Leng
Other reports | ID IR(Coory_324479
on None
pe microic acid
ph 303
> Search Summary (Ta | aconomy response) | | | Database Name in
Description 2
Program 0 | er
All non-redundant GenB
excluding environmental
BLASTX 2.8.1+ (= Ctute | ank CDS transations+PDB+Swi
samples from WGS projects
II | sainst-Piik-Piik |
| Graphic Su | mmary | | | | | | | |
| Show Com | erved Domains | | | | | | | |
| | | | Distribution of th | e top 100 Elast Hits o | n 100 subject sequer | xes @ | | |
| | | | HILLING OV | Color key for aligan | eat scores | | | |
| | | | -640 40-5 | io 5 5-80 | BD-200 | >-200 | | |
| | | 5 | 1 70 | 140 | 10 280 | 350 | | |
| | | | | | | | | |
| | | | | | | | - | |
| toc//blactechin | Non trail and a second second | | | | | | III Quest | ions/comments |
| | and an | | | | | | and the second se | |



| | BRCA 1 - No | Arabidopsin | Arabidopeis | DNA & RMA | glp3 - Nucle | Synthetic core | S NOR BIAN X | + 1018 | | × |
|---|---|--|--|--|---------------------------|----------------|---|--|---|---|
| → C | ŵ | 🛈 🖸 🔒 http | s://blast.ncbi.nlm.r | ih.gov/blast.cgi#al | nindr_30584479 | E 50% | 🖂 🏠 | III/ | 0 | Ξ |
| Down | Abad + GenPept Ge | utika. | | | | | ▼ Next = | Descriptions | | |
| Homo | sepiens interferon, a
ce ID: AAP36432.1 La | lpha-inducible pro
righ: 131 Number o | tein (clone IF1-6-16
Matches 1 |), partial (synthetic) | construct) | | | | | |
| P CARD | Tautile General Control | | | Thursday, a la | | | Related information | | | |
| 224 bit | (STT) 24-73 Composit | anal restric adjust | Decideo Positivo
D V 131(100%) 13 V 13 | a Daga 55
(100%) D(131(0%) + | arsa
I | | ANP36452.1 | stitical proteins to | | |
| Query
Roles | 1 MAGRAVILILOT
MAGRAVILILOT | | E SOGOGO PARALITIKA
E ISOGOGO PARALITIKA
E SOGOGO PARALITIKA | Conteventeria 18
Conteventeria 18
Conteventeria 60 | | | | | | |
| Query | 101 PENASTLANSVA
PESASTLANSVA | ULTERREATIONOOVPA
LITERREATIONOOVPA | OSLVATIONICAROSSIV | TONTINIZATE >+ | 6 | | | | | |
| Sejet | 41 PTGAGIAANEVA
141 EVLOHERDEEL | STWERE STREET | DOEVATEQUE ABOSSIV | 10815ALNEYATE 12 | | | | | | |
| Sejes | 121 KYLDSERDERL
121 KYLDSERDERL | 131 | | | | | | | | |
| | | | | | | | | | | |
| Cow | koad - GenPapt On | iphica. | | | | | V Next A Previou | us 🛓 Descriptions | | |
| Court
Interfe | koad - <u>GenPapt Or</u>
ron alpha-inducible p | i <u>chica</u>
irolein 6 isoform a | precursor (Homo s | apiens) | | | Vient A Previou | us 🛓 Descriptions | | |
| EDown
Interfe
Segree
P Sco | Koet - <u>GenPapt</u> On
ron alpha-inducible p
ce to <u>MP (02020c)</u> L
12 more Ifficia) | iphica
irolein 6 isoform a
orgth: 130 Number | precursor [Homo s
of Matches: 1 | apiens] | | | ▼ Next ▲ Previou | us 🛓 Descriptions | | |
| EDown
Interfe
Sequen
P Seq.
Targe f | Koad – <u>GenPupt Gr</u>
ron alphe-inducible p
ce ito <u>HP 002020 (</u>
12 more 1865s)
14 133 Genter Station | edica
rolein 6 isoform a
orgili: 130 Number | precursor (Homo s
of Hatches: 1 | apiens)
Thursdame (2.5 | | | Neat Previou | us 🛓 Descriptions | | |
| B Down
interfe
Sequen
P Seq
Name 1
Secon
222 bits | Koad - <u>GenPupt Gr</u>
ron alphe-inducible p
ce (0. <u>hP_003010_1_1</u>
12 more titlets)
14 133 <u>Gentus</u>
<u>Capent Michool</u>
(566) 34-73 Composit | ephica
instein 6 isoform a
origiti: 130 Number
istal matrix adjust. | precursor [Homo s
of Matches: 1
mentice Paulies
(39130(10%) 13913 | apiens)
7 too toos & to
8 Gape 77
(100%) 0138(2%) * | ernal Valde | | Related information | en
In Stats
Igressed seguence | | |
| Interfe
Sequent
Integer fr
Securit
Securit
222 bits
Durry | Alexed - GenePaipt Gen
rom adpha-inductible
cer to MP_0020000 (1
12 more Miccia)
1 to 113 General Miccia)
1 to 113 General Micciae
(200) 38-73 Composit
1 Mispanit/CFCC
Mispanit/CFCC | enten 6 soform a
regiti 130 Norder
anal rubic adjust. | precursor (Homo s
of Matchen 1
De100(10%) 13F10
Executive Technology
Executive Technology
Exe | T suctain & P
a Gaps P
(100%) 01380%) +
Goolariostato | erraal bards
anna
1 | | Related information
Gamy - associated per
(Inform - dustriation
Tags | n
n
pro Solada
gressied Sequence
algrest generat | | |

| Search NCBI databases - NC | E X 🔛 DNA & ENA | - Site Galde - N.C. X 😌 Home - Gene - NCBI | X 🛃 Home - BioSystems - NCBI 🛛 X 📩 | (C)(C) |
|----------------------------|-----------------|--|---|------------------------|
| ← → ሮ ŵ | 3 🗑 🊔 http | as://www.ncbi.nlm. nlh.gav /gene | 110% … 🗟 🟠 | III (D |
| S NCBI Resources 🖻 | How To 🕤 | | <u>shamim_cst_k</u> | y NCBI Sign Out |
| Gene | Gene + | | S | earch |
| | | Advanced | | Help |
| - | 1 | Gene | | |
| | 11 | Gene integrates information from a v
Reference Sequences (RefSeqs), m
genome-, phenotype-, and locus-spe | vide range of species. A record may include no
aps, pathways, variations, phenotypes, and lin
crific resources worldwide. | omenclature,
ks to |
| Using Gene | | Gene Tools | Other Resources | |
| Gene Quick Start | | Submit GeneRIFs | HamoloGene | |
| FAQ | | Submit Correction | OMIM | |
| Download FTP | | Statistics | RefSeq | |
| RefSeq Mailing List | | BLAST | RefSegGene | |
| Gene News 🔯 | | Genome Workbench | UniGene | |
| S D 6 | Q 📋 🚳 | E 🖪 🗃 🖷 | Profess Charless | R = (20.29
17-12-1 |

| Search NCEI databases - NCEI X | 🚼 DNA & RNA - Ste-Gakle - N 🗆 🗙 🛃 Home - Gene - NCBI | X S Home - Budgeterns - NCH X + |
|--------------------------------|--|--|
| ⊖ → ଫ ŵ | (i) (ii) (iii) https://www.ncbi.nlm.nit.gov/biosystems/ | 1288 … 回☆ ■ □ |
| S NCBI Resources 🕑 He | w To 🖸 | shamim_cst My NCBI Sign Out |
| BioSystems | BioSystems • | Search |
| 37 | Limits Advanced | Holp |
| इन | BioSystems | |
| 6-lactori 8H4 | The NCBI BioSystems Database provinces genes, proteins, and small biosystems and other related data thro | des integrated access to biological systems and their
molecules, as well as iterature describing those
sughout Entrez. |
| Using BioSystems | BioSystem Tools | Other Resources |
| About Biosystems | Tools Overview | Entrez Gene |
| How To (Quick Start) Guides | FLink Frequency-weighted Links | Entrez Protein |
| Help | E-Utilities (Entrez programming utilities) | PubChem |
| ETP | | GEO Gene Expression |
| 🗿 🖸 🤌 🌍 | 📋 🧕 📐 🛃 🔊 🗶 | 65 - 🗞 🖻 🖉 1733 |

| ← → ♂ @ | 🛈 🛈 🛔 https://www.ncbi.nim.nih.gow/guide/dna-ma/ 🛛 🛛 🔝 🗰 🔂 🗘 | IN ED E |
|---------|---|---------|
| | | |
| | Downloads | |
| | BLAST (Stand-alone) | |
| | BLAST executables for local use are provided for Solaris, LINUX, Windows, and | |
| | MacUSX systems, See the reLAUME tile in the tip directory for more information.
Pre-formatted databases for BLAST nucleotide, protein, and translated searches | |
| | also are available for downloading under the db subdirectory. | |
| | FTP: BLAST Databases | |
| | Sequence databases for use with the stand-alone BLAST programs. The files in this | |
| | directory are pre-formatted databases that are ready to use with BLAST. | |
| | FTP: FASTA BLAST Databases | |
| | Sequence databases in FASTA format for use with the stand-alone BLAST | |
| | used with BLAST. | |
| | | |
| | FLP: Gendark
This site contains files for all sequence records in GenBank in the default flat file | |
| | format. The files are organized by GenBank division, and the full contents are | |
| | described in the README.genbank file. | |

| Search NCIII data 🗙 🤮 | DNA & RNA - S⊨ 🗙 | 🗧 Home - Nucleoti 🗙 🔛 RefSeq: NCIII Ret 🗙 | 🔛 Home - Gene - II: X 🛛 🔛 Home - BliaSy | ste X + I-IUI-X |
|---|--|--|--|--|
| - → @ @ | 0 🗊 🔒 | https://www.ncbi.nlm. nih.gov /huccore | (110%) 🖂 | ☆ IN D |
| S NCBI Resources 🛛 | How To 🕤 | | shi | amim_cst_My_NCBI_Sign_Out |
| Nucleotide | Nucleotide | • Advanced | | Search Help |
| The Nucleotide datability of the Nucleotid | ase will include ES | 3T and GSS sequences in early 2019. <u>Read mo</u> | g. | |
| | | | | |
| ACCCACE | ACATT | CCT) Nucleotide | | |
| ACCCAGCA
TGTAGCTT
GGTTTGCT | CACATT,
CACACACO
ACCTCO | ATT The Nucleotide
ATT The Nucleotide database is a c
ATT RefSer, TPA and PDB. Genom
biomedical research and disco | ollection of sequences from several sou
e, gene and transcript sequence data p
reny. | rces, including GenBank,
rovide the foundation for |
| ACCCAGCA
TGTAGCTT
GTTTCCT | CACATT,
CACATT,
CACACC
ACCTCC | ATT Revision for the second se | olection of sequences from several sou
e, gene and transcript sequence data p
wy.
Other Resource | rces, including GenBank,
rovide the foundation for
S |
| ACCCAGCA
TGTAGCTT
GTTTCCT
Using Nucleotide
Duck Start Gide | | Nucleotide The Nucleotide database is a s The Nucleotide database is a s The Nucleotide database is a s The Nucleotide Tools Submit is GenBark | olaction of sequences from several sous
e, gene and transcript sequence data p
evg.
Other Resource
GenBank Home | rces, including GenBank,
rovide the foundation for
S |
| ACCCAGCA
TGTAGCTT
GGTTTCCT
Using Nucleotide
Quick Start Gide
EAQ | | Nucleotide The Nucleotide distances is a c CCC Reflex, The and POB Genom bornesical research and deco Nucleotide Tools Stant to Centers ListOut | oliscion of sequences from several soci
e, gene and transcript sequence data p
erg.
Other Resource
GenBick Home
RefSeq Home | rces, including GenBank,
rovide the foundation for
S |

| ← → C @ | 🖲 🗑 🏔 https:// | www.ncbi.nim. nih.go wirefseq/ | □ 110% ···· □ ☆ | ₩ 00 Ξ |
|----------------------------|----------------|--|--|---------------|
| S NCBI Resources 🕑 Ho | w To 🕤 | | shamim_cst_M | NCBI Sign Out |
| RefSeq R | efSeq • | | Se | rch |
| | () | RefSeq: NCBI Reference | e Sequence Database | |
| | (| A comprehensive, integrated, non-redundar
including genomic, transcript, and protein. | nt, well-annotated set of reference sequen | ces . |
| Using RefSeq | | RefSeq Access | RefSeq projects | |
| About RefSeg | | Human Genome Resources and Download | Consensus CDS (CCDS) | |
| Human Reference Genome | | RefSeq FTP | RefSeq Functional Elements | |
| Prokaryotic RefSeq Genomes | | RefSeq genomes FTP | RefSeqGene | |
| | | New RefSeq genomic (last 30 days) | Targeted Loci | |
| EAQ | | New Bellins insertion (and 20 down) | Marie Mariation | |
| FAQ
NCBI Handbook | | New Merced Equiprible (rear on onlia) | 1150 V80180511 | |

8/5/2019

8/5/2019