

**AFRICAN CENTERS OF EXCELLENCE  
IN BIOINFORMATICS**

KAMPALA, UGANDA

Genome Databases  
Mariam Quiñones, PhD

1

---

---

---


---

---

---

---

---



### Today's Instructor

- Bioinformatics and Computational Biosciences Branch (BCBB), NIAID
- National Institutes of Health, Bethesda, MD USA.
- Contact our team via email:
  - Email: [ace@icermali.org](mailto:ace@icermali.org)
  - Listserv: [ACE-GLOBAL-ANNOUNCE@list.nih.gov](mailto:ACE-GLOBAL-ANNOUNCE@list.nih.gov)
  - Instructor: [mariam.quinones@nih.gov](mailto:mariam.quinones@nih.gov)

**Dr. Mariam Quiñones,**  
Ph.D. in Molecular Microbiology

Ongoing Computational Biology projects:

- 16S microbiome
- Ion Torrent Zika virus genome assembly

2

---

---

---

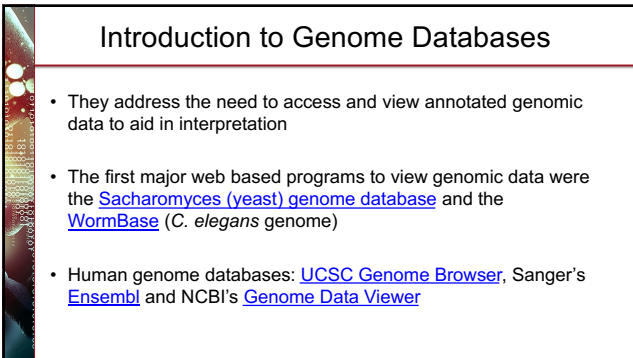
---

---

---

---

---



### Introduction to Genome Databases

- They address the need to access and view annotated genomic data to aid in interpretation
- The first major web based programs to view genomic data were the [Saccharomyces \(yeast\) genome database](#) and the [WormBase](#) (*C. elegans* genome)
- Human genome databases: [UCSC Genome Browser](#), Sanger's [Ensembl](#) and NCBI's [Genome Data Viewer](#)

---

---

---

---

---

---

---

---

## Other Biological Databases

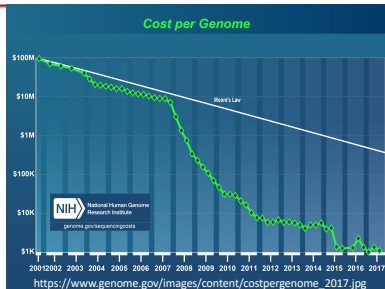
DNA Variants Genomes  
RNA Plants Gene  
Protein Immune Expression  
Pathways Metabolic ...

### NAR Database Summary Paper Category List

Nucleotide Sequence Databases  
RNA sequence databases  
Protein sequence databases  
Structure Databases  
Genomics Databases (non-vertebrate)  
Metabolic and Signaling Pathways  
Human and other Vertebrate Genomes  
Human Genes and Diseases  
Microarray Data and other Gene Expression Databases  
Proteomics Resources  
Other Molecular Biology Databases  
Organelle databases  
Plant databases  
Immunological databases  
Cell biology

View list: <http://www.oxfordjournals.org/nar/database/c/>

The reduction in sequencing costs allow increase in genome sequencing



- Whole Genomes
- Targeted regions
- Deep sequencing



- Comparative genomics
- Gene expression
- Epigenetics / Regulation
- Metagenomics

<https://www.genome.gov/27541954/dna-sequencing-costs-data/>

## What can you do with genome databases?

- Search for genomic features (e.g. genes, regulatory elements, transcripts)
- Explore region of interest for annotations such as gene expression, histone modification, conservation across species, mutations
- Use browser to upload custom tracks of data and compare with published tracks
- Download sequence data in multiple

## What else is in a Genome Browser?

With a genome browser, you display of genomic information from biological databases

Besides genome sequence, they provide additional data:

- known and predicted genes
- ESTs
- mRNAs
- CpG islands
- assembly gaps and coverage
- chromosomal bands
- homology to other organisms
- RNA-seq data
- Transcription factor binding sites
- GC percent
- Splicing variants
- Known SNPs
- Associated publications
- Sequence repeats

---

---

---

---

---

---

---

---



---

---

---

---

---

---

---

---

## Select which tracks to display or add custom

---

---

---

---

---

---

---

---

## Download data using the Table Browser

# Download data using the Table Browser

Genomes
Genome Browser
Tools
Mirrors
Downloads
My Data
Help
About Us

## Table Browser

Use this program to retrieve the data associated with a track in a text format, to calculate intersections between tracks, and to retrieve DNA sequence covered by a track. For help in using this application see [Using the Table Browser](#) for a description of the controls in this form, the [User's Guide](#) for general information and sample queries, and the [Copeptella Table Browser tutorial](#) for a formatted presentation of the software features and usage. For more complex queries, you may want to use [Galaxy](#) or our public [MGI-SQL server](#). To examine the biological function of your set through annotation enrichments, send data to [Gene2Path](#). Send data to [Gene2Path](#) for analysis with diverse computational tools. Refer to the [Credits](#) page for the list of contributors and usage agreements associated with these tools. All tables can be downloaded in their entirety from the [Sequence and Annotation Downloads](#) page.

[illegible]

NCBI databases - <https://www.ncbi.nlm.nih.gov/>

The screenshot displays the NCBI database interface for the *Mycobacterium tuberculosis* H37Rv complete genome. The top navigation bar includes the NCBI logo and the text "NCBI databases - https://www.ncbi.nlm.nih.gov/". Below this, a search bar is visible with the text "Nucleotide" and a "Search" button. The main content area shows the "Mycobacterium tuberculosis H37Rv, complete genome" entry, with a link to "NCBI Reference Sequence NC\_009682.3". The genome is visualized as a circular map with various annotations, including gene names, features, and links to related resources. On the right side, there are two panels: "Access to related tools" and "Links to related data". The "Access to related tools" panel includes links to "Analyze this sequence", "Run BLAST", "Find Proteins", and "Highlight Sequence Features". The "Links to related data" panel includes links to "Related Information", "Assembly", "Protein", "Pubmed", "Taxonomy", "Comments (Gene)", "Full text in PMC", "Map", "Genome", and "Virtual GenBank Sequence".

[illegible]

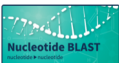
NCBI BLAST to DNA and Protein sequences  
<https://blast.ncbi.nlm.nih.gov/Blast.cgi>

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

# NCBI BLAST to DNA and Protein sequences

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


## Basic Local Alignment Search Tool



**Nucleotide BLAST**

**tblastn**  
protein vs translated nucleotide

**blastx**  
translated nucleotide vs protein



**Protein BLAST**

## Other Search Tools

Specialized searches			
<b>SearchBLAST</b> Find matches highly similar to your query	<b>Primer-BLAST</b> Design primers specific to your PCR template	<b>Global Align</b> Compare two sequences across their full length (Needleman-Wu)	<b>CD-search</b> Find conserved domains in your sequence
<b>SSO</b> Find matches to gene expression profiles	<b>SPR-BLAST</b> Search microarrays and find fold-up genes	<b>VecScreen</b> Search sequences for vector contamination	<b>CDART</b> Find sequences with similar conserved domain architecture
<b>Targeted Loci</b> Search exons for phylogenetic analysis	<b>Multiple alignment</b> Align sequences using multiple alignment	<b>Rikemap</b> Search protein or nucleotide ligands to Rikemap	<b>WUOL-BLAST</b> Establish taxonomy for classification

---

---

---

---

---

---

## NCBI repository makes publishes your data

**Submit**

NCBI collects submissions of data for the world's largest public repository of biological and scientific information.

**Check the Status of Your Submissions**

**Submission Wizard**

Need help figuring out where to start? Try this!

**Dysbiotic microbiome triggers Th17 cells to mediate oral mucosal immunopathology in mice and humans**

This submission was on behalf of Dr. Nik Moudapour

Accession: PRJNA484972

Data Type: Metagenome

Scope: Environment

Submission: Registration date: 7-Aug-2018  
NAID: Bioinformatics and Computational Biosciences Branch

Related Resources: Website name

**Project Data:**

Resource Name	Number of Links
Sequence sets	
SRA Experiments	35
Other samplers	
BioRx	35

**SRA Data Details**

Parameter	Value
Data volume, Gbases	2
Data volume, Mbytes	1558

---

---

---

---

---

---

---

---

## European Nucleotide Database

- It mirrors NCBI SRA and allows easy download of fastq files



Study accession	Sample accession	Secondary sample accession	Experiment accession	Run accession	Tax ID	Scientific name	Instrument model	Library layout	FASTQ files (FTP)	FASTQ files (Galaxy)
PRJNA480459	SAMN09636795	SRS3531855	SRX4374315	SRR7504914	410661	mouse gut metagenome	Illumina MiSeq	PAIRED	File 1 File 2	File 1 File 2

---

---

---

---

---

---

---

---

## Other Specialized Genome Databases

- Prokaryotes
  - Integrated Microbial Genomes (IMG) <https://img.jgi.doe.gov>
- Eukaryotic Pathogens (EuPathD) <https://eupathdb.org/eupathdb>
- Influenza:
  - GSAID <https://www.gisaid.org>
  - IRD <https://www.fludb.org>
- Viruses
  - VIPR <https://www.viprbrc.org>

---

---

---

---

---

---

---

---

**Let's practice!**

- Please follow exercises provided

---

---

---

---

---

---

---