

## **WEBLEM 9**

### **Introduction of Genome Database**

- **Genome Database:**

The Genome Database (GDB) is a public repository of data on human genes, clones, STSs, polymorphisms and maps. GDB entries are highly cross-linked to each other, to literature citations and to entries in other databases, including the sequence databases, OMIM, and the Mouse Genome Database. Mapping data from large genome centers and smaller mapping efforts are added to GDB on an ongoing basis. The database can be searched by a variety of methods, ranging from keyword searches to complex queries. Major functionality extensions in the last year include the ongoing computation of integrated human genome maps, called Comprehensive Maps, and the use of those maps to support positional queries and graphic displays. The capabilities of the GDB map viewer (Mapview) have been extended to include map printing and the graphical display of ad hoc query results. The HUGO Nomenclature Committee continues to curate the proposed and official gene symbols and related data in collaboration with GDB. As genome research shifts its emphasis from mapping to sequencing and functional analysis, the scope of the GDB schema is being extended. We are in the process of adding representations of gene function and expression, and improving our representation of human polymorphism and mutation.

1. **Genome Databases in NCBI:**

The Genome database contains sequence and map data from the whole genomes of over 1000 species or strains. The genomes represent both completely sequenced genomes and those with sequencing in-progress. All three main domains of life (bacteria, archaea, and eukaryota) are represented, as well as many viruses, phages, viroids, plasmids, and organelles. This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

2. **GOLD:**

Genomes OnLine Database (GOLD) is an open-access repository of genome and metagenome sequencing projects with their associated metadata. Login-free access is provided to a growing catalogue of manually curated public projects from all over the world. Starting from whole genome sequencing of cultured bacteria, uncultured single cells, complex eukaryotes or metagenome sequencing of environmental samples, the world of sequencing is advancing at a rapid pace. In overcoming the limitations of culturing microbial isolates, improved sequencing and analysis methods have broadened our understanding of the microbial world. Metagenome- assembled genomes (MAGs) and single-amplified genomes (SAGs) (1,2) are two such examples of genomes of uncultivated organisms that have recently expanded our knowledge about microorganisms. In any sequencing project, metadata or accessory information about the sample being sequenced plays a key role. In order to interpret raw sequence data and make accurate scientific predictions, the presence of a wide variety of metadata is extremely important. For example, comparative analyses of the microbial community of soil samples from multiple locations would be very difficult if the geographic location information or soil type is not recorded. On the other hand, the scientific value of the same sequence will likely increase several-fold if detailed environmental and physicochemical properties of the soil are collected.

## WEBLEM 9/a

(URL: <https://www.ncbi.nlm.nih.gov/genome/>)**Aim:**

To study the *Staphylococcus epidermidis* in NCBI genome browser

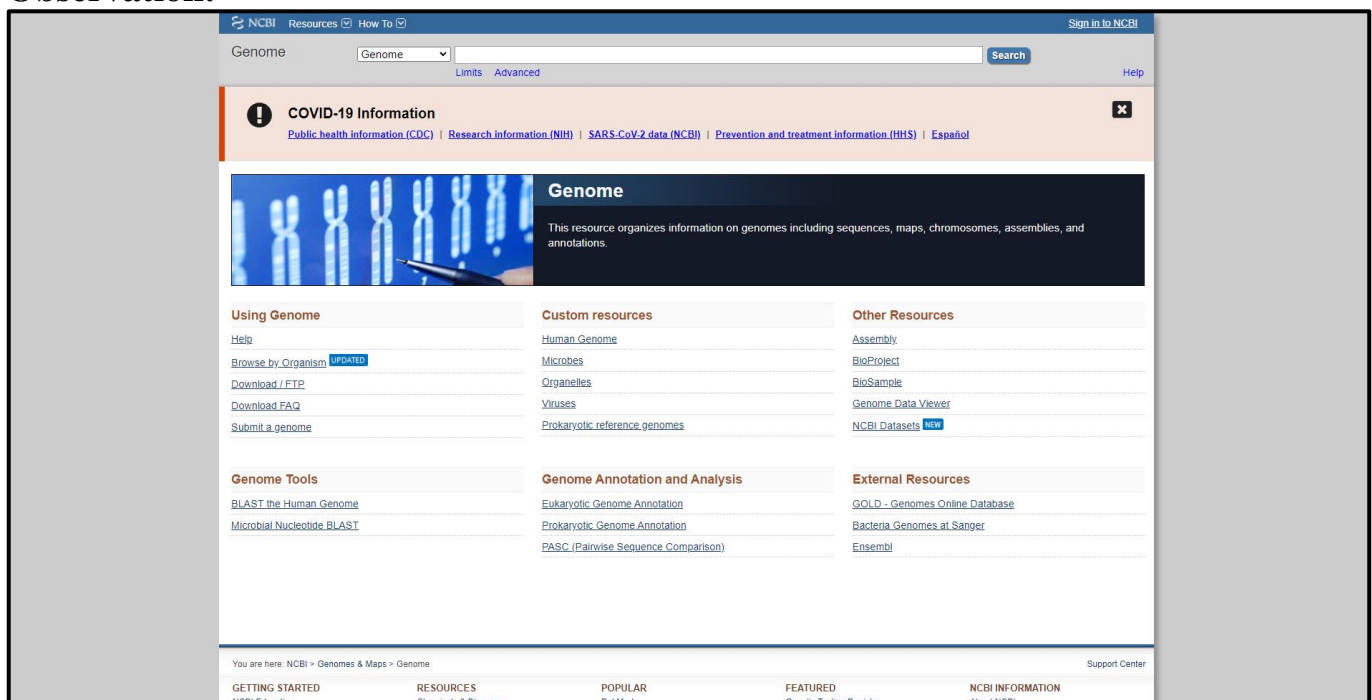
**Introduction:**

The Genome database contains sequence and map data from the whole genomes of over 1000 species or strains. The genomes represent both completely sequenced genomes and those with sequencing in-progress. All three main domains of life (bacteria, archaea, and eukaryota) are represented, as well as many viruses, phages, viroids, plasmids, and organelles. This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

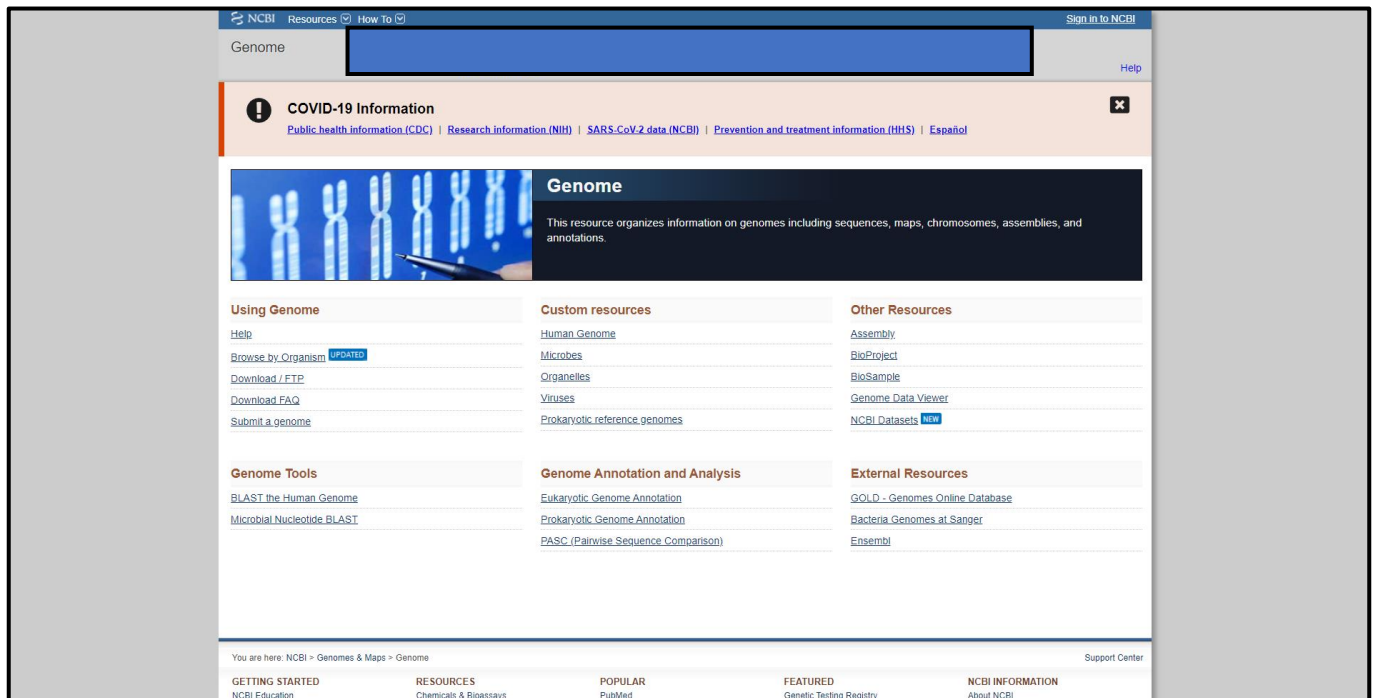
Staphylococci are common bacterial colonizers of the skin and mucous membranes of humans and other mammals. *S. epidermidis* in particular is the most frequently isolated species from human epithelia. It colonizes predominantly the axillae, head, and nares. Analysis of the *S. epidermidis* genome indicated that the species is well equipped with genes assumed to provide protection from the harsh conditions encountered in its natural habitat. For example, to cope with extremes of salt concentration and osmotic pressure, *S. epidermidis* has eight sodium ion/proton exchangers and six transport systems for osmoprotectants.

**Methodology:**

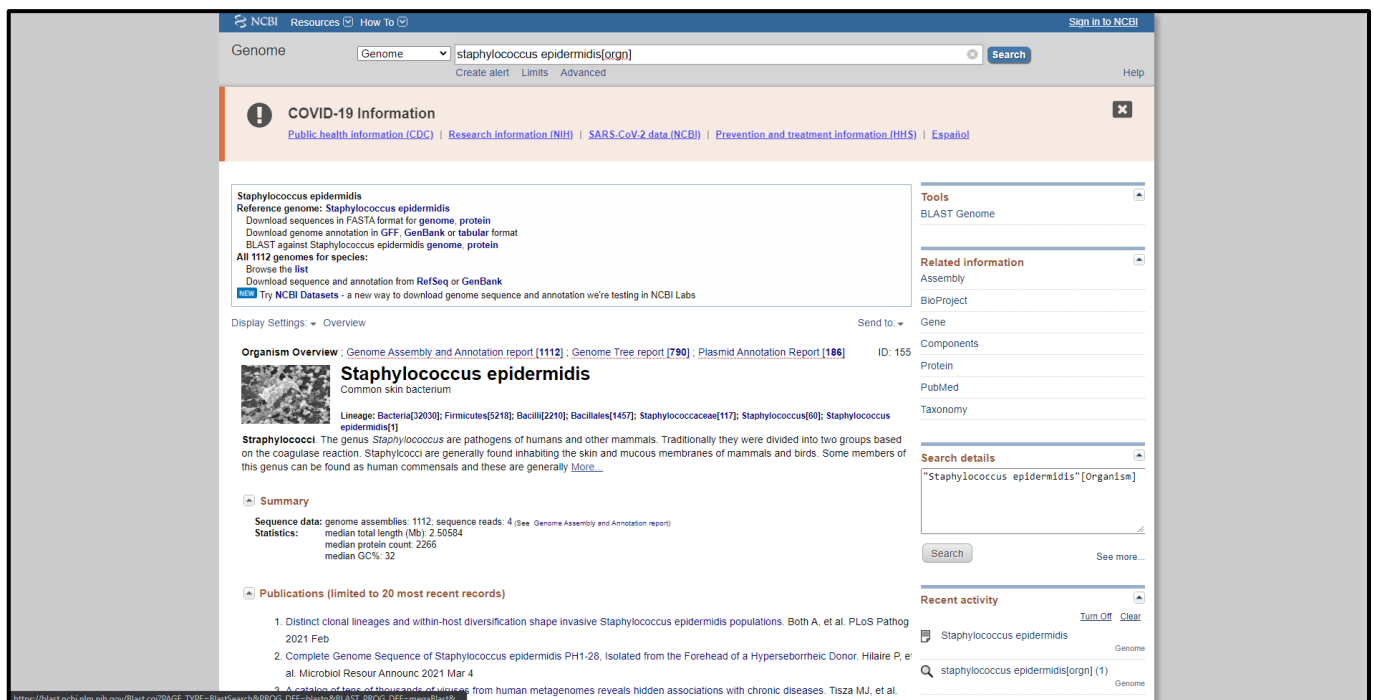
1. Open The homepage of Genome in NBCI
2. Enter the query in Search bar.
3. Open any one of the result page.
4. Interpret the Result.

**Observation:**

**Fig1. Homepage of Genome in NCBI**



**Fig2. Search page of Genome in NCBI with query Staphylococcus epidermidis**



**Fig3. Result page of Genome in NCBI for query Staphylococcus epidermidis.**

NIH

U.S. National Library of Medicine

NCBI

National Center for Biotechnology Information

Log in

New Genome Table

Try our new [Genome page](#) and use the feedback button to let us know what you think

Genome > Genome Information by Organism > Protein Table for Staphylococcus epidermidis

Protein name, accession, locus, locus\_tag ...

Q Search

Filters

Name

☐ chromosome (2,192)

☐ plasmid unnamed1 (21)

☐ plasmid unnamed2 (3)

Download

#	Name	Accession	Start	Stop	Strand	GeneID	Locus	Locus tag	Protein product	Length	Protein Name
1	chromosome	NZ_CP035288.1	137354	139417	+	50017543		EQW00_RS00700	WP_002437601.1	687	copper-translocating P-type ATPase
2	chromosome	NZ_CP035288.1	139432	140865	+	50017544		EQW00_RS00705	WP_002437604.1	477	MULTISPECIES: multicopper oxidase domain-containing protein
3	chromosome	NZ_CP035288.1	140885	141367	+	50017545		EQW00_RS00710	WP_001830525.1	160	MULTISPECIES: YdhK family protein
4	chromosome	NZ_CP035288.1	141839	143167	-	50017546	csfB	EQW00_RS00715	WP_002485357.1	442	MULTISPECIES: persulfide dioxygenase-sulfurtransferase CsfB
5	chromosome	NZ_CP035288.1	143186	144256	-	50017547		EQW00_RS00720	WP_001830526.1	356	MULTISPECIES: DsrE/DsrF/DsrH-like family protein
6	chromosome	NZ_CP035288.1	144395	144655	+	50017548	csfR	EQW00_RS00725	WP_000466194.1	86	MULTISPECIES: persulfide-sensing transcriptional repressor CsfR
7	chromosome	NZ_CP035288.1	144670	145410	+	50017549		EQW00_RS00730	WP_002455948.1	246	MULTISPECIES: sulfite exporter TsaE/SatE family protein
8	chromosome	NZ_CP035288.1	145471	145866	-	50017550	arsC	EQW00_RS00735	WP_001830465.1	131	MULTISPECIES: arsenate reductase (thioredoxin)
9	chromosome	NZ_CP035288.1	145885	147174	-	50017551	arsB	EQW00_RS00740	WP_002455949.1	429	MULTISPECIES: arsenite efflux transporter membrane subunit ArsB
10	chromosome	NZ_CP035288.1	147177	147491	-	50017552		EQW00_RS00745	WP_002437624.1	104	MULTISPECIES: metalloregulator ArsR/SmtB family transcription factor
11	chromosome	NZ_CP035288.1	147641	149365	-	50017553	arsA	EQW00_RS00750	WP_002494543.1	574	MULTISPECIES: arsenical pump-driving ATPase
12	chromosome	NZ_CP035288.1	149352	149699	-	50017554	arsD	EQW00_RS00755	WP_002455950.1	115	MULTISPECIES: arsenite efflux transporter metallochaperone ArsD
13	chromosome	NZ_CP035288.1	150220	150540	+	50017555		EQW00_RS00760	WP_002437630.1	106	MULTISPECIES: metalloregulator ArsR/SmtB family transcription factor
14	chromosome	NZ_CP035288.1	150628	151512	+	50017556		EQW00_RS00765	WP_002455951.1	294	MULTISPECIES: permease
15	chromosome	NZ_CP035288.1	152007	152792	+	50017557		EQW00_RS00775	WP_002437636.1	261	tandem-type lipoprotein
16	chromosome	NZ_CP035288.1	153110	153865	+	50017558		EQW00_RS00780	WP_002455952.1	251	tandem-type lipoprotein
17	chromosome	NZ_CP035288.1	154289	154970	+	50017559		EQW00_RS00785	WP_002437638.1	253	MULTISPECIES: tandem-type lipoprotein
18	chromosome	NZ_CP035288.1	155327	156214	+	50017560		EQW00_RS00790	WP_002455963.1	295	MULTISPECIES: mechanosensitive ion channel family protein
19	chromosome	NZ_CP035288.1	156579	157391	+	50017561		EQW00_RS00795	WP_002437642.1	270	MULTISPECIES: HAD-IIB family hydrolase
20	chromosome	NZ_CP035288.1	157555	158238	+	50017562		EQW00_RS00800	WP_002437638.1	184	hypothetical protein

Choose Columns

Page 1 of 45

View 1 - 50 of 2,216

Fig4. Genome annotation Table with query Staphylococcus epidermidis

Staphylococcus epidermidis

Common skin bacterium

Lineage: Bacteria[32030]; Firmicutes[5218]; Bacilli[2210]; Bacillales[1457]; Staphylococcales[117]; Staphylococcus[90]; Staphylococcus epidermidis[1]

**Staphylococci.** The genus *Staphylococcus* are pathogens of humans and other mammals. Traditionally they were divided into two groups based on the coagulase reaction. *Staphylococci* are generally found inhabiting the skin and mucous membranes of mammals and birds. Some members of this genus can be found as human commensals and these are generally [MORC...](#)

Dendrogram (based on genomic BLAST)

Find:

all

Tools

Upload

PubMed

Taxonomy

Search details

"Staphylococcus epidermidis"[Organism]

Search

See more...

Recent activity

Staphylococcus epidermidis

Genome

staphylococcus epidermidis[orgn] (1)

Genome

Staphylococcus (70)

Genome

Staphylococcus epidermidis -- the 'accidental' pathogen

Genome

Staphylococcus epidermidis strain ATCC 14990 chromosome, complete geno nucleotide

Genome

See more...

Fig5. Result page for query “Staphylococcus epidermidis” showing publications and genome information for reference and representative genomes.

GenBank

Due to the large size of this record, sequence and annotated features are not shown. Use the "Customize view" panel to change the display.

### Staphylococcus epidermidis strain ATCC 14990 chromosome, complete genome

NCBI Reference Sequence: NZ\_CP035288.1

[FASTA](#) [Graphics](#)

[Go to:](#)

LOCUS NZ\_CP035288 2466502 bp DNA circular CON 31-MAR-2021

DEFINITION Staphylococcus epidermidis strain ATCC 14990 chromosome, complete genome.

ACCESSION NZ\_CP035288

VERSION NZ\_CP035288.1

DBLINK BioProject: PRJNA224116  
BioSample: SAMN10738425  
Assembly: GCF\_0066094375.1

KEYWORDS RefSeq.

SOURCE Staphylococcus epidermidis

ORGANISM Staphylococcus epidermidis  
Bacteria; Firmicutes; Bacilli; Bacillales; Staphylococcaceae;

REFERENCE 1 (bases 1 to 2466502)  
AUTHORS Chen,H. and Liang,Q.  
TITLE Staphylococcus epidermidis Genome sequence and assembly  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2466502)  
AUTHORS Chen,H. and Liang,Q.  
TITLE Direct Submission  
JOURNAL Submitted (16-JAN-2019) Key Laboratory of Microbial Technology and Bioinformatics of Zhejiang Province, Zhejiang Tianke High Technology Development Co.Ltd., Huanggushan Road, Hangzhou, Zhejiang 310000, China

COMMENT **REFSEQ INFORMATION:** The reference sequence is identical to CP035288.1.  
The annotation was added by the NCBI Prokaryotic Genome Annotation Pipeline (PGAP). Information about PGAP can be found here: [https://www.ncbi.nlm.nih.gov/genome/annotation\\_prok/](https://www.ncbi.nlm.nih.gov/genome/annotation_prok/)

**##Genome-Assembly-Data-START##**  
Assembly Method :: Unicycler v. v0.4.5  
Genome Representation :: Full  
Expected Final Version :: Yes  
Genome Coverage :: 240.0x  
Sequencing Technology :: Oxford Nanopore GridION  
**##Genome-Assembly-Data-END##**

**##Genome-Annotation-Data-START##**  
Annotation Provider :: NCBI RefSeq  
Annotation Date :: 03/31/2021 12:01:08

Change region shown

Customize view

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Related information

Assembly

BioProject

BioSample

Protein

Taxonomy

BioCollections

Components (Core)

Gene

Genome

Identical GenBank Sequence

PubMed (Weighted)

LinkOut to external resources

Staphylococcus epidermidis

[BacDrive]

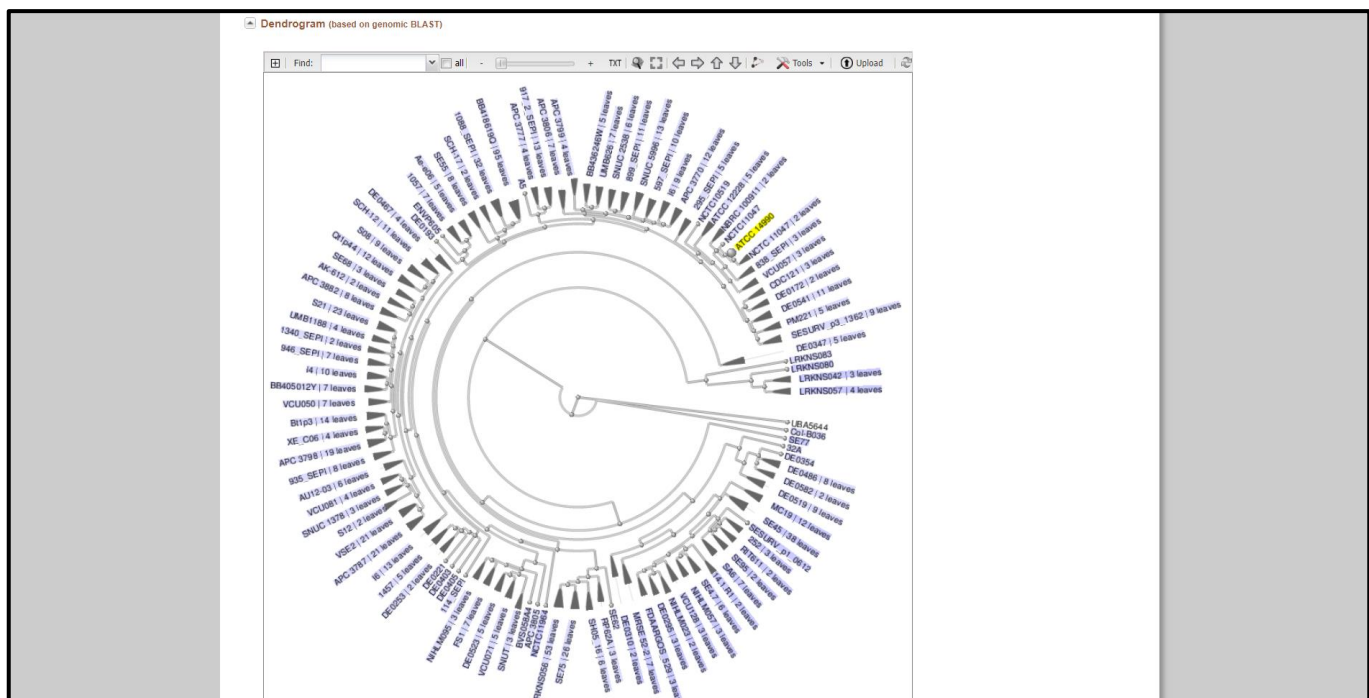
Order ARSA cDNA clone/Protein/Antibody/RNAi [ORFgene]

Order nuvB cDNA clone/Protein/Antibody/RNAi [ORFgene]

Order dnaJ cDNA clone/Protein/Antibody/RNAi [ORFgene]

Recent activity

**Fig6. Reference sequence information for query “Staphylococcus epidermidis” in NCBI GenBank database.**



**Fig7. Result page for query “Staphylococcus epidermidis” showing dendrogram on genomic BLAST**

U.S. National Library of Medicine

National Center for Biotechnology Information

Log in

New Genome Table

Try our new [Genome page](#) and use the feedback button to let us know what you think

Genome >

Genome Information by Organism

Staphylococcus epidermidis

Organism name (common or scientific) or Accession (Assembly, BioProject or replicon) ...

Q Search

Download Reports from FTP site

Organism Overview

Plasmids (186)

Filters

Download

Choose Columns

Page 1 of 23

View 1 - 50 of 1,112

#	Organism Name	Organism Groups	Strain	Bio Sample	BioProject	Assembly	Levi	Size	GC%	Replicons	WG	Scaffok	CDS	Release Da	FTP
1	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	ATCC 14990	SAMN10738425	PRJNA51533	GCA_006094375.1		2.49	32.17	chromosome: NZ_CP035288.1/CP035288.1 plasmid unname1: NZ_CP035289.1/CP035289.1 plasmid unname2: NZ_CP035290.1/CP035290.1	3	2.216	06-Jun-2019	R	G
2	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	NBRC 100911	SAMD00169832	PRJDB88042	GCA_006742205.1		2.43	32.30	chromosome: NZ_AP019721.1/AP019721.1 plasmid pSEP1: NZ_AP019722.1/AP019722.1	2	2.154	11-Jun-2019	R	G
3	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	B1276912	SAMN16699880	PRJNA72109	GCA_019329745.1		2.83	31.97	chromosome: NZ_CP073865.1/CP073865.1 plasmid p1B1276912: NZ_CP073866.1/CP073866.1 plasmid p2B1276912: NZ_CP073867.1/CP073867.1	3	2.530	23-Jul-2021	R	G
4	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	B1272014	SAMN16699884	PRJNA72109	GCA_019329665.1		2.82	31.98	chromosome: NZ_CP073857.1/CP073857.1 plasmid p1B1272014: NZ_CP073858.1/CP073858.1	2	2.612	23-Jul-2021	R	G
5	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	BPH0662	SAMEA3982825	PRJEB13975	GCA_900088615.2		2.85	31.95	chromosome 1: NZ_LT571450.1/LT571450.1 plasmid 2: NZ_LT571451.1/LT571451.1 plasmid 3: NZ_LT571451.1/LT571451.1 Show all 4 replicons	4	2.636	12-May-2016	R	G
6	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	B1230143	SAMN16699875	PRJNA72109	GCA_019329425.1		2.78	31.96	chromosome: NZ_CP073835.1/CP073835.1 plasmid p1B1230143: NZ_CP073833.1/CP073833.1 plasmid p2B1230143: NZ_CP073834.1/CP073834.1	3	2.577	23-Jul-2021	R	G
7	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	NCTC13924	SAMEA104318201	PRJEB6403	GCA_900638695.1		2.75	32.00	chromosome 1: NZ_LR134536.1/LR134536.1	1	2.538	19-Dec-2018	R	G
8	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	IRL01	SAMN09103952	PRJNA53248	GCA_009685135.1		2.75	31.98	chromosome: NZ_CP045648.1/CP045648.1 plasmid pBPH0747-01: NZ_CP045647.1/CP045647.1	2	2.516	21-Nov-2019	R	G
9	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	DAR1907	SAMN0744319	PRJNA30832	GCA_002850315.1		2.73	32.10	chromosome: NZ_CP013943.1/CP013943.1	1	2.493	03-Jan-2018	R	G
10	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	B1266875	SAMN16699880	PRJNA72109	GCA_019329585.1		2.73	31.88	chromosome: NZ_CP073847.1/CP073847.1 plasmid p1B1266875: NZ_CP073848.1/CP073848.1	2	2.500	23-Jul-2021	R	G
11	Staphylococcus epidermidis	Bacteria, Terrabacteria group, Firmicutes	B1262351	SAMN16699876	PRJNA72109	GCA_019329505.1		2.74	31.87	chromosome: NZ_CP073836.1/CP073836.1 plasmid p1B1262351: NZ_CP073837.1/CP073837.1 plasmid p2B1262351: NZ_CP073838.1/CP073838.1	3	2.501	23-Jul-2021	R	G
12	Staphylococcus	Bacteria, Terrabacteria	B1287475	SAMN18696894	PRJNA72109	GCA_019330105.1		2.74	31.85	chromosome: NZ_CP073883.1/CP073883.1 plasmid p1B1287475: NZ_CP073884.1/CP073884.1	3	2.512	23-Jul-2021	R	G

### Result:

## Conclusion:

## References:



DATE: 1-11-21

**WEBLEM 9/b**  
**GOLD**  
(URL: <https://gold.jgi.doe.gov/>)

Genomes OnLine Database (GOLD) is an open-access repository of genome and metagenome sequencing projects with their associated metadata. Login-free access is provided to a growing catalogue of manually curated public projects from all over the world. Starting from whole genome sequencing of cultured bacteria, uncultured single cells, complex eukaryotes or metagenome sequencing of environmental samples, the world of sequencing is advancing at a rapid pace. In overcoming the limitations of culturing microbial isolates, improved sequencing and analysis methods have broadened our understanding of the microbial world. Metagenome-assembled genomes (MAGs) and single-amplified genomes (SAGs) are two such examples of genomes of uncultivated organisms that have recently expanded our knowledge about microorganisms.

There are three different ways in which projects and their associated metadata are entered into GOLD:

- Projects that are added by individual GOLD users.
- Projects for samples sequenced at the Joint Genome Institute (JGI) as part of the JGI User Programs and
- Projects sourced from public databases such as the NCBI GenBank and Sequence Read Archive (SRA).

Well-defined genome and metagenome projects in GOLD are a prerequisite for submitting sequence data to the Integrated Microbial Genomes with Microbiome (IMG/M) system for annotation and analysis. GOLD and IMG work closely to provide a seamless experience to our broad user base of scientists, students, and policy makers as well as novice science enthusiasts. Users can browse through several metadata fields in GOLD and take advantage of the multiple filters and advanced search tools to identify a subset of genomes or metagenomes of interest. They can then use the IMG identifiers from GOLD and directly go to the IMG user interface and conduct further research and comparative analyses.

**GOLD Organization and Current Status:**

GOLD is organized in a four-level hierarchical system to describe the overall proposal, samples or organisms studied, sequencing projects undertaken and their analysis process. These levels include Study, Organism or Biosample, Sequencing Project (SP), and Analysis Project (AP). All four levels, along with their complex list of metadata fields and controlled vocabulary (CV) terms, are connected to each other in a lucid framework to enhance scientific discovery.

A brief description of each of these four levels is provided below:

**Study:** The overall research objectives and goals are captured in a Study, which lies at the helm of the four-level organization structure. Studies can vary in the type of samples collected. Subsequently, a single Study may have several Sequencing and Analysis Projects that differ in their methodology and application such as Whole Genome Sequencing (WGS) and analysis, metagenome analysis or a combination of both.

**Organism/Biosample:** A GOLD Organism or Biosample contains descriptive information about the biological or physical material that is being sequenced. Any living entity such as bacteria, archaea, fungi, virus, plant or animal may constitute an Organism in GOLD. An Organism may be cultured, uncultured

(such as single cells) or even bioinformatically predicted (such as Metagenome-Assembled Genomes or MAGs). The most important metadata feature of a GOLD Organism is its Taxonomic assignment including phylum, genus, species, strain etc. To promote interconnectedness with other databases, all GOLD Organisms have the NCBI Taxonomy ID.

**Sequencing Project:** The process of generating sequencing data from a Biosample or Organism is described in a Sequencing Project (SP). GOLD currently has 15 different types of SPs, from which whole-genome sequencing (WGS) and metagenome are most commonly used. The input material for an SP can either be DNA or RNA corresponding to a genome or transcriptome project, respectively. This material can come from either an organism, in the case of WGS and transcriptomes, or from a Biosample, in the case of metagenomes and metatranscriptomes.

**Analysis project:** The information about the processing of sequence data and its analysis is captured in a GOLD Analysis Project (AP). A GOLD AP is required in order to submit the corresponding sequence data to IMG for annotation. There are 13 different types of Analysis Projects that can be defined in the current version, namely: Genome, Metagenome, Metagenome—Cell Enrichment, Metagenome—Single Particle Sort, Metatranscriptome, Single Cell (Unscreened), Single Cell (Screened), Metagenome-Assembled Genome, as well as five different types of Combined Assembly.

## **BROWSING AND SEARCHING GOLD:**

Public data in GOLD are freely accessible through the website (<https://gold.jgi.doe.gov>). The homepage provides key statistics on different entities, along with links to related metadata.

**Search:** The Search tab has two options: Advanced Search and Metadata Search. While there are different ways to search the database, one of the most popular and commonly used features is the ‘Advanced Search.’ It provides a quick and easy way to search across all the different levels using several metadata fields.

**Distribution Graphs:** The Distribution Graphs section gives a top-down view of different fields in which the user receives a snapshot of the metadata scale and their diversity. A brief summary of the different types of Sequencing Projects for genome and metagenome projects and their sequencing statuses are displayed as pie charts in the Distribution Graphs section. Additionally, a user can look at the phylogenetic distribution of projects both as pie charts and in hyperlinked expandable table formats. Finally, it also shows the breakdown of Biosamples by each of the five different Ecosystem levels.

**Biogeographical Metadata:** The Biogeographical Metadata section shows the isolation location of Biosamples and Organisms as placeholders on interactive maps.

**Ecosystem classification:** GOLD Ecosystem Classification paths describe the surroundings from which an environmental sample or an organism is collected. This five-level hierarchical classification system shown below, was originally described by Ivanova et. al. in a paper titled “A call for standardized classification of metagenome projects.”

To facilitate the annotation and classification of a large number of diverse environmental samples and organisms we need a simple, versatile and adaptable classification system. GOLD ecosystem classification provides a simple but powerful framework to describe samples from environmental, host associated as well as engineered or built environments.

This classification system is not a comprehensive list of all possible paths from any given environment,



host or engineered setting. Instead it is primarily driven by the samples we curate. Within this framework we periodically review and revise these paths based on existing as well as the new samples we encounter.

The Five Levels of the classification system are as follows:

Ecosystem -> Ecosystem Category -> Ecosystem Type -> Ecosystem Subtype -> Specific Ecosystem

**SRA Explorer:** NCBI Sequence Read Archive (SRA) stores a collection of sequence data from all forms of life including metagenomic samples. The NCBI SRA Explorer provides a way to search SRA data. Using a combination of free text fields (such as organism name or SRA Experiment Title) and fields with predefined or controlled vocabulary based values (such as sequencing technologies or library strategies), the different types of data available in SRA can be reviewed. The SRA Explorer also has a filter where users can see which SRA data are currently present in GOLD.

**Statistics:** The Statistics tab contains graphical descriptions of several Sequencing Project metadata fields. This ranges from the phylogenetic distribution and relevance of bacterial genome projects, breakdown of projects by sequencing centers or the yearly growth of sequencing projects by organism domain.

**Downloads:** The Downloads page gives users the ability to access different types of data, ranging from a list of several controlled vocabulary terms, public metagenome and metatranscriptome projects from SRA that were annotated in IMG or a data dump of GOLD-containing select fields from various database entities. These files can be downloaded in Excel format and the underlying data is updated daily.

GOLD database thus can be used to retrieve genome and metagenome sequencing projects with their associated metadata information. It consists of organized four-level hierarchical system which includes study, organism or biosample, sequencing project, and analysis project and it provides various browsing and searching resources that makes it easier for the users to access information.

#### REFERENCES:

1. Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Sundaramurthi, J., Lee, J., . . . Reddy, T. a. (2020). Genomes OnLine Database (GOLD) v.8: Overview and updates. *Nucleic Acids Research*, 49(D1). doi:10.1093/nar/gkaa983
2. JGI gold: Home. JGI GOLD | Home. (n.d.). Retrieved October 20, 2021, from <https://gold.jgi.doe.gov/>