

ATCC Genome Portal

Getting Started

The ATCC Genome Portal is a reference database for high-quality bacterial genomes originating from authenticated ATCC materials. Before diving in to the genomic data that the portal has to offer, all new users will need to create an account.

ATCC GENOMES SEQUENCE SEARCH ATCCPREVIEW@ONECODEX.COM

Welcome to the ATCC Genome Portal

A comprehensive collection of high-quality microbial genomics reference data

[VIEW ALL GENOMES >](#)

Search for a genome

Search for genomes and/or metadata

Recently published

- Acinetobacter johnsonii (ATCC® 17909™)**
Added 05/13/2019
- Lactobacillus jensenii (ATCC® 25258™)**
Added 05/13/2019
- Veillonella parvula (ATCC® 17745™)**
Added 05/13/2019

To create your account, click on the **LOG IN** icon located on the home page and select “Register” from the drop-down menu. On the registration page, simply enter some basic contact information and create a password to set up your account.

ATCC

Create your account

Name

Organization

Phone

Email

Password

Confirm Password

I agree with the [Terms of Use](#)

Powered by ONE CODEX

After logging in to the portal, you can explore the variety of genomes published on the portal by clicking on the “Genomes” link, search for a genome using your own nucleotide sequence by clicking on the “Sequence Search” link, or search for a genome using the “Search for a genome” search bar.

Search for genomes and/or metadata

Search by species
species:"escherichia coli"

Search by genus
genus:escherichia

Search by isolation source
isolation:"blood"

Search by catalog number
catalog_number:19089

When using the search bar, you can search for genomes by full taxonomic name, species, genus, family, order, NCBI taxonomy ID, the isolation source, ATCC catalog number, type strain status, or bio-safety level number.

Published Genomes

The “Genomes” table lists all of the genomes available on the portal. From this table, you can link out to the ATCC product page if interested in purchasing the source material for a particular genome by clicking on the  icon, view genome size, download the assemblies and annotations, or take a deeper dive into a genome by clicking on “View”.

Genomes

[All Genomes](#)

Sort Taxonomic Name  

Taxonomic name	ATCC Product Name	Date Published	Length	Download	Genomic Data
<i>Caulobacter vibrioides</i>	ATCC® 19089™ 	May 14, 2019	4.1 Mb	 Download	<input type="button" value="View"/>
<i>Cellulomonas denverensis</i>	ATCC® BAA-788™ 	August 27, 2019	4.0 Mb	 Download	<input type="button" value="View"/>
<i>Chromobacterium violaceum</i>	ATCC® 12472™ 	August 27, 2019	4.8 Mb	 Download	<input type="button" value="View"/>
<i>Chromohalobacter salexigens</i>	ATCC® BAA-138™ 	August 27, 2019	3.7 Mb	 Download	<input type="button" value="View"/>

In addition to filtering the table by the previously mentioned search criteria, the table is filterable by the taxonomic name and the date genomes were published to the portal.

The portal provides assemblies in FASTA format and annotations in GBK format. To download either file from the “Genomes” page, simply click  Download and make your selection from the drop-down menu.

Genome Overview

Clicking on “View” from the “Genomes” page will direct you to the “Overview” tab for a specific genome.

Streptococcus mutans (ATCC® 700610™)

[Overview](#) [Genome Browser](#) [Related Genomes](#) [Quality Control](#)

Assembly Summary		Organism Summary	
Date Published	June 17, 2019	Name	ATCC® 700610™
Length	2,031,444 bp	Isolation	Child with active caries, 1982
Sequencing Technology 	Illumina + Oxford Nanopore Hybrid Assembly	Biosafety Level	1
Number of Contigs	1 (All Circularized)	Type Strain	No
N50	2,031,444 bp	Antigenic Properties	Serotype c
%GC	36.83%	Antibiotic Resistance	Sensitive to erythromycin Sensitive to rifampin rifampicin, rifamycin AMP Sensitive to spectinomycin actinospectacin Sensitive to streptomycin
Annotations Summary		Input Reads Summary	
Number of CDS	1,325	Oxford Nanopore Read Count	45,801
Number of Hypothetical Proteins	611	Oxford Nanopore Median Q Score 	14

All of the genomes published on the portal were assembled using a hybrid assembly method that leverages highly accurate Illumina short-reads with the scaffolding ability of ultra-long Oxford Nanopore reads. This approach allows for the generation of a genome assembly that is more complete and accurate than what could be generated with each individual technology alone. The “Overview” tab provides a summary of the genome assembly, ATCC source organism, annotations, and the sequencing reads that went into the assembly.

Genome Browser

Annotations can be visualized using the linear genome browser on the “Genome Browser” tab. Each annotation is color coded by protein family (see [Annotation Legend](#)). If you are interested in downloading the nucleotide and amino acid sequences for the entire annotations table, simply click on [Download Table CSV](#).

Streptococcus mutans (ATCC® 700610™)

Overview [Genome Browser](#) Related Genomes Quality Control [Annotation Legend](#)

Include Hypothetical Proteins [Download Table CSV](#)

Contig	Start	End	Name	Protein Product	EC Number	Type	Uniprot ID	Jump
1	1	1360	<i>dnaA</i>	Chromosomal replication initiator protein DnaA		CDS	P05648	📍 ...
1	1515	2652	<i>dnaN</i>	Beta sliding clamp		CDS	O06672	📍 ...
1	2913	3105		hypothetical protein		CDS		📍 ...

To exclude hypothetical proteins from the annotations table simply uncheck the “Include Hypothetical Proteins” checkbox. If you are interested in exploring which genes neighbor a particular gene listed in the table, click on the 📍 icon to jump to the region of the linear viewer where the gene is located.

If you are interested in filtering the table for smaller subset of annotations, the “Filter annotations” search bar can be used. Using the genome for ATCC® 700610™ as an example, we can search for *purB*, a gene that encodes for adenylosuccinate lyase, and then search for other genomes on the portal that contain a similar gene sequence.

Streptococcus mutans (ATCC® 700610™)

Overview [Genome Browser](#) Related Genomes Quality Control [Annotation Legend](#)

Include Hypothetical Proteins [Download Table CSV](#)

Contig	Start	End	Name	Protein Product	EC Number	Type	Uniprot ID	Jump
1	59078	60350	<i>purB</i>	Adenylosuccinate lyase	4.3.2.2	CDS	P12047	📍 ...

Clicking on the ... icon allows users to immediately copy the nucleotide or amino acid sequence of a given annotation to their clipboard. If you are interested in discovering what other genomes on the portal share a similar nucleotide sequence, simply click on “Search for nucleotide sequence” and you will be redirected to the “Sequence Search” page to view your results.

Streptococcus mutans (ATCC® 700610™)

Overview [Genome Browser](#) Related Genomes Quality Control [Annotation Legend](#)

Include Hypothetical Proteins [Download Table CSV](#)

Contig	Start	End	Name	Protein Product	EC Number	Type	Uniprot ID	Jump
1	59078	60350	<i>purB</i>	Adenylosuccinate lyase	4.3.2.2	CDS	P12047	📍 ...

- Copy nucleotide sequence
- Copy amino acid sequence
- Search for nucleotide sequence

Related Genomes

If you are interested in discovering what other genomes on the portal are related to your genome of interest, you can do so by navigating to the “Related Genomes” tab.

Streptococcus mutans (ATCC® 700610™) [↗](#)

Overview [Genome Browser](#) [Related Genomes](#) [Quality Control](#)

Most similar Genomes

The following genomes have the greatest genomic similarity to this one (>95% average nucleotide identity).

***Streptococcus mutans* (ATCC® 25175™)** 99.3% similar

1 contig
2.0 Mb

[View Genome](#)

Other members of this genus

The following genomes share the same genus according to the NCBI taxonomy.

***Streptococcus pyogenes* (ATCC® 700294™)**

1 contig
1.9 Mb

[View Genome](#)

***Streptococcus pneumoniae* (ATCC® BAA-334™)**

1 contig
2.2 Mb

[View Genome](#)

***Streptococcus pyogenes* (ATCC® 12344™)**

1 contig
1.9 Mb

[View Genome](#)

On the ATCC Genome Portal, there are two ways of classifying genomes as related. The first is based on average nucleotide identity (ANI) and lists all genomes that have greater than a 95% genomic similarity. The second is based on NCBI taxonomy and lists all genomes that fall within the same genus as your species of interest. If you are interested in taking a deeper dive into one of these related genomes, you can do so by clicking on “View Genome”.

Quality Control

In order to be published on the portal, genomes must go through an extensive quality control (QC) process that includes both sequencing and assembly QC.

Streptococcus mutans (ATCC® 700610™) [↗](#)

Overview [Genome Browser](#) [Related Genomes](#) [Quality Control](#)

Sequencing Quality Control

Quality control statistics on Illumina sequencing data.

4/4 4 out of 4 passed

Passed	Number of trimmed reads	2,939,778
Passed	Median Q score, all bases	38
Passed	Percent of median Q scores per cycle greater than 25	100%
Passed	Ambiguous content (% N bases)	0

Assembly Quality Control

Metrics assessing the assembly quality (from CheckM). **i**

3/3 3 out of 3 passed

Passed	Estimated genome completeness	100%
Passed	Estimated genome contamination	0%
Passed	Average depth of coverage	330.059x

In order to pass sequencing QC, we require a minimum of 1,000,000 Illumina reads with a median Q score of 30 or greater for all bases, and a median Q score of 25 or greater per base. Additionally, there must be less 5% ambiguous content or “N” bases.

For assembly QC, we utilize CheckM, which is a tool that uses a set of Hidden Markov Models (HMMs) from phylogenetically close reference genomes, to determine if the query assembly contains all expected HMMs as predicted by the reference genomes (a percentage called “CheckM completeness”) and what percent of the query’s HMMs differ in copy number or come from reference genomes that are phylogenetically distant (called “CheckM contamination”). We require final assemblies to have completeness values $\geq 95\%$ and contamination values $\leq 5\%$. Additionally, all assemblies have an average of 100X Illumina coverage across the entire span of the genome.

Sequence Search

The final feature on the ATCC Genome Portal is the *k*-mer based nucleotide search. The sequence search matches all *k*-mers ($k=31$) in the query against all available ATCC reference genomes and highlights portions of the sequence that match. The minimum requirement is matching 40 *k*-mers and 80% of the sequence to call a hit.

Search for a genome

Enter a nucleotide sequence (at least 40 bases) to find genomes that match >80% of the sequence.

```
ATGATCAACCGTTATTACGCCCTGAGATGGCGAACATTTGGAGTGAAGAAAACAAGTACCGTGCTTGGTTGGAAGTGGAAATTTGGCTGAC
GAGGCTTGGGCTGAGTGGGTGAGATTCCTAAGGAAGATGTGGCAAGATTTCGCGAAGGCGGATTTTGATATGACCGCATTTGGAGATT
GAGCAGCAACCGCTCAGATGGTGGTGGCTTACCCTGCGGTTCTGAGACGTTGGGTGAGGAGCGCAAGTGGGTACACATTTGGGCTGACC
TCAACAGAGCTGGTGGACACGGCTACGGCTACCTATACAAGCAGGCTAACGCCATCATTGTAAGGACTTGGATAATTTCTCAGTATCATC
GCTGATAAGGCCAAGAAGCACCAAGTTACCATCATGATGGGGCGCACGCGGTCCATGCGGAGCCAAACAATTTGGTCTCAAGCTGGCG
ACTTGGTACAGTGAATGAAGCGTAATATTGAGCGTTTGAACATGCTGCCGCTGGTGTGGAACTGGGAAAATCTTGGTCTGTTGGGAAAC
TTTGGCAACATTCGCCCTTTTGGAAAAATATGCTGTGACAAATAGGCATTCGTGCTCAGGAAATTTCAACTCAAGTCTTACCGCTGAT
TTGATGCTGAATATTTTGGCTGCTAGCCAGCATTTGCCACATCAATCGAGCGCATGGGCACTGAAATCCGTTGGTCTGCAAAAATCCGAACAG
CGTGAAGTAGAAGATTCTTTGCCAAGGGCAAAAAGGACGCTCTGCTATGCCCAATAAAGCAACCCATCGCTGTAATAATGACGGGG
CTTGGCGTGTGATTCGGGACACATGGTGGCGCTATGAAAATGTGCCCTCTGGCACGAACGGATATTTCCCACTCATCAGCTGAGCGC
ATTATCGCCGCTGACGACCATTCATGACATGCTCAATCGCTTTGGCAATCGTGAAAAATTTGACCGCTTCCCAAAAATATG
AAGCGCAACATGGGCTCACCCTTGGTCTCATTTTCAECAGCGTCTATGCTGACCTTAATGAAAAAGCATGACCGCGAGCAGGCTTAC
GACCTTGTTCAGCTAAGACTGCGAGCTTGGGACAAACAAGTGGACTTCAAAACGTTGCTGGAAGCAGATCCAGAGATCAGCTCTCGCTT
AGCCAAGAGAAATCGACGAAATCTTAAATCAGTCTACTACCAACGCGTGGATGAAATCTTTGACCGGATGGATTGGGAGACTAA
```

[Search](#)

Results on 1299 bases

Streptococcus mutans (ATCC® 700610™) 1299 bases matched (100%)

1 contig
2.0 Mb [View Genome](#)

Streptococcus mutans (ATCC® 25175™) 996 bases matched (76%)

1 contig
2.0 Mb [View Genome](#)

To use the search, simply insert a nucleotide sequence into the text box on the left and click “Search”. If any portion of your sequence matches the sequence of a published genome, it will become highlighted. To see what portion of your sequence matches the search results on the right, simply hover over a genome to highlight your sequence in the color associated with that search result.

Search for a genome

Enter a nucleotide sequence (at least 40 bases) to find genomes that match >80% of the sequence.

```
ATGATCAACCGTTATTACGCCCTGAGATGGCGAACATTTGGAGTGAAGAAAACAAGTACCGTGCTTGGTTGGAAGTGGAAATTTGGCTGAC
GAGGCTTGGGCTGAGTGGGTGAGATTCCTAAGGAAGATGTGGCAAGATTTCGCGAAGGCGGATTTTGATATGACCGCATTTGGAGATT
GAGCAGCAACCGCTCAGATGGTGGTGGCTTACCCTGCGGTTCTGAGACGTTGGGTGAGGAGCGCAAGTGGGTACACATTTGGGCTGACC
TCAACAGAGCTGGTGGACACGGCTACGGCTACCTATACAAGCAGGCTAACGCCATCATTGTAAGGACTTGGATAATTTCTCAGTATCATC
GCTGATAAGGCCAAGAAGCACCAAGTTACCATCATGATGGGGCGCACGCGGTCCATGCGGAGCCAAACAATTTGGTCTCAAGCTGGCG
ACTTGGTACAGTGAATGAAGCGTAATATTGAGCGTTTGAACATGCTGCCGCTGGTGTGGAACTGGGAAAATCTTGGTCTGTTGGGAAAC
TTTGGCAACATTCGCCCTTTTGGAAAAATATGCTGTGACAAATAGGCATTCGTGCTCAGGAAATTTCAACTCAAGTCTTACCGCTGAT
TTGATGCTGAATATTTTGGCTGCTAGCCAGCATTTGCCACATCAATCGAGCGCATGGGCACTGAAATCCGTTGGTCTGCAAAAATCCGAACAG
CGTGAAGTAGAAGATTCTTTGCCAAGGGCAAAAAGGACGCTCTGCTATGCCCAATAAAGCAACCCATCGCTGTAATAATGACGGGG
CTTGGCGTGTGATTCGGGACACATGGTGGCGCTATGAAAATGTGCCCTCTGGCACGAACGGATATTTCCCACTCATCAGCTGAGCGC
ATTATCGCCGCTGACGACCATTCATGACATGCTCAATCGCTTTGGCAATCGTGAAAAATTTGACCGCTTCCCAAAAATATG
AAGCGCAACATGGGCTCACCCTTGGTCTCATTTTCAECAGCGTCTATGCTGACCTTAATGAAAAAGCATGACCGCGAGCAGGCTTAC
GACCTTGTTCAGCTAAGACTGCGAGCTTGGGACAAACAAGTGGACTTCAAAACGTTGCTGGAAGCAGATCCAGAGATCAGCTCTCGCTT
AGCCAAGAGAAATCGACGAAATCTTAAATCAGTCTACTACCAACGCGTGGATGAAATCTTTGACCGGATGGATTGGGAGACTAA
```

[Search](#)

Results on 1299 bases

Streptococcus mutans (ATCC® 700610™) 1299 bases matched (100%)

1 contig
2.0 Mb [View Genome](#)

Streptococcus mutans (ATCC® 25175™) 996 bases matched (76%)

1 contig
2.0 Mb [View Genome](#)

Using the nucleotide sequence of *purB* from ATCC® 700610™ as an example, you can see the variation between the *purB* gene in two different *Streptococcus mutans* strains in the ATCC collection.