

Intro to NCBI/BLAST

5/14/24

National Center for Biotechnology Information (NCBI)

- Governmental institute; branch of the National Institute for Health (NIH).
- “NCBI” is also slang for their website and databases.
 - e.g. “Look up the *E. coli* genome in NCBI.”
- Extremely trustworthy! The industry standard for finding information/papers and sequence data for **any** organism. Not just bacteria!
- It’s good practice to upload any sequences you generate to NCBI when you publish papers on them.



Log in

All Databases

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and code libraries to build applications



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Navigate to
<https://www.ncbi.nlm.nih.gov/>

Popular Resources

PubMed

Bookshelf

PubMed Central

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

NCBI News & Blog

NCBI SARS-CoV-2 Resources Page Will Redirect Soon

21 Jun 2023

End of the COVID-19 Public Health Emergency During the COVID-19

Now Available! Access Data from the Human Pangenome Research Consortium (HPRC) at NCBI

15 Jun 2023

Have you ever wondered how your Making Discoveries in Canine & Human Oncology using the NIH Comparative Genomics Resource (CGR)

14 Jun 2023

Do you use model organisms to study

More...

FOLLOW NCBI



Log in

All Databases

Search

NCBI Home

Resource List (A-Z)

All Resources

Chemicals & Bioassays

Data & Software

DNA & RNA

Domains & Structures

Genes & Expression

Genetics & Medicine

Genomes & Maps

Homology

Literature

Proteins

Sequence Analysis

Taxonomy

Training & Tutorials

Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Download

Transfer NCBI data to your computer



Learn

Find help documents, attend a class or watch a tutorial



Develop

Use NCBI APIs and code libraries to build applications



Analyze

Identify an NCBI tool for your data analysis task



Research

Explore NCBI research and collaborative projects



COVID-19 Information

[Public health information \(CDC\)](#) | [Research information \(NIH\)](#) | [SARS-CoV-2 data \(NCBI\)](#) | [Prevention and treatment information \(HHS\)](#) | [Español](#)

Popular Resources

PubMed

Bookshelf

PubMed Central

BLAST

Nucleotide

Genome

SNP

Gene

Protein

PubChem

NCBI News & Blog

NCBI SARS-CoV-2 Resources Page Will Redirect Soon

21 Jun 2023

End of the COVID-19 Public Health Emergency During the COVID-19

Now Available! Access Data from the Human Pangenome Research Consortium (HPRC) at NCBI

15 Jun 2023

Have you ever wondered how your Making Discoveries in Canine & Human Oncology using the NIH Comparative Genomics Resource (CGR)

14 Jun 2023

Do you use model organisms to study

More...

Tells you about each of the NCBI databases.

FOLLOW NCBI



Log in

All Databases

Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

All Resources

All Databases Downloads Submissions Tools How To

Databases

Assembly

A database providing information on the structure of assembled genomes, assembly names and other meta-data, statistical reports, and links to genomic sequence data.

BioCollections

A curated set of metadata for culture collections, museums, herbaria and other natural history collections. The records display collection codes, information about the collections' home institutions, and links to relevant data at NCBI.

BioProject (formerly Genome Project)

A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

BioSample

The BioSample database contains descriptions of biological source materials used in experimental assays.

Bookshelf

A collection of biomedical books that can be searched directly or from linked data in other NCBI databases. The collection includes biomedical textbooks, other scientific titles, genetic resources such as *GeneReviews*, and NCBI help manuals.

ClinVar

A resource to provide a public, tracked record of reported relationships between human variation and observed health status with supporting evidence. Related information in the [NIH Genetic Testing Registry \(GTR\)](#), [MedGen](#), [Gene](#), [OMIM](#), [PubMed](#) and other sources is accessible through hyperlinks on the records.

ClinicalTrials.gov

A registry and results database of publicly- and privately-supported clinical studies of human participants conducted around the world.

Computational Resources from NCBI's Structure Group

A centralized page providing access and links to resources developed by the Structure Group of the NCBI Computational Biology Branch (CBB). These resources cover databases and tools to help in the study of macromolecular structures, conserved domains and protein classification, small molecules and their biological activity, and biological pathways and systems.

Consensus CDS (CCDS)

A collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

Tells you about each of the NCBI databases.

Includes databases for sequencing data, sample collection, general species information, publications, etc.

Let's find the reference genome for *Bacillus cereus*!



Log in

All Databases

Search

All Resources

All Databases Downloads Submissions Tools How To

Databases

Assembly

A database providing information on the structure of assembled genomes, assembly names and other meta-data, statistical reports, and links to genomic sequence data.

BioCollections

A curated set of metadata for culture collections, museums, herbaria and other natural history collections. The records display collection codes, information about the collections' home institutions, and links to relevant data at NCBI.

BioProject (formerly Genome Project)

A collection of genomics, functional genomics, and genetics studies and links to their resulting datasets. This resource describes project scope, material, and objectives and provides a mechanism to retrieve datasets that are often difficult to find due to inconsistent annotation, multiple independent submissions, and the varied nature of diverse data types which are often stored in different databases.

BioSample

The BioSample database contains descriptions of biological source materials used in experimental assays.

Bookshelf

A collection of biomedical books that can be searched directly or from linked data in other NCBI databases. The collection includes biomedical textbooks, other scientific titles, genetic resources such as *GeneReviews*, and NCBI help manuals.

ClinVar

A resource to provide a public, tracked record of reported relationships between human variation and observed health status with supporting evidence. Related information in the [NIH Genetic Testing Registry \(GTR\)](#), [MedGen](#), [Gene](#), [OMIM](#), [PubMed](#) and other sources is accessible through hyperlinks on the records.

ClinicalTrials.gov

A registry and results database of publicly- and privately-supported clinical studies of human participants conducted around the world.

Computational Resources from NCBI's Structure Group

A centralized page providing access and links to resources developed by the Structure Group of the NCBI Computational Biology Branch (CBB). These resources cover databases and tools to help in the study of macromolecular structures, conserved domains and protein classification, small molecules and their biological activity, and biological pathways and systems.

Consensus CDS (CCDS)

A collaborative effort to identify a core set of human and mouse protein coding regions that are consistently annotated and of high quality.

Search for “Bacillus cereus”

Search NCBI

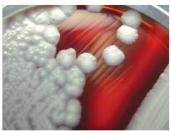
bacillus cereus



Search

Results found in 27 databases (1 error)

TAXONOMY



Bacillus cereus

Bacillus cereus is a species of firmicute in the family *Bacillaceae*.

Taxonomy ID: 1396

Was this helpful?



Genomes

Browse all *Bacillus cereus* genomes

Literature

Bookshelf	205
MeSH	12
NLM Catalog	27
PubMed	12,376
PubMed Central	35,135

Genes

Gene	94,862
GEO DataSets	1,242
GEO Profiles	0
HomoloGene	error
PopSet	2,226

Proteins

Conserved Domains	107
Identical Protein Groups	2,045,921
Protein	7,954,560
Protein Family Models	423
Structure	617

The output will show you results from all the databases! Scroll down and click “Genome”.

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa —

Bacillus cereus × Enter one or more taxonomic names

Filters ←

Download ▼

Select columns ←

2,523 Genomes

Rows per page

20 ▼

1-20 of 2,523

<input type="checkbox"/> Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation
<input type="checkbox"/> ASM222028v1 	GCA_002220285.1	GCF_002220285.1	Bacillus cereus	FORC_047 (strain)	NCBI Refs Submitter
<input type="checkbox"/> ASM609429v1	GCA_006094295.1	GCF_006094295.1	Bacillus cereus ATCC 14579	ATCC 14579 (strain)	NCBI Refs Submitter
<input type="checkbox"/> ASM782v1	GCA_000007825.1	GCF_000007825.1	Bacillus cereus ATCC 14579	ATCC 14579 (strain)	NCBI Refs Submitter
<input type="checkbox"/> ASM2165533v1	GCA_021655335.1	GCF_021655335.1	Bacillus cereus	30075 (strain)	NCBI Refs Submitter
<input type="checkbox"/> ASM2920126v1	GCA_029201265.1	GCF_029201265.1	Bacillus cereus	M72-4 (strain)	NCBI Refs

Here you can see a list of all *B. cereus* genomes in NCBI. You could...

- Download genomes by selecting them.
- Filter for a specific genome assembly level (contigs, scaffolds, full genome)
- See more statistics about the genome assembly (gene number, N50, completeness, etc.)

Genome

Download a genome data package including genome, transcript and protein sequence, annotation and a data report

Selected taxa —

Bacillus cereus  Enter one or more taxonomic names FiltersDownload 

Select columns

2,523 Genomes

Rows per page

20

1-20 of 2,523

<input type="checkbox"/> Assembly	GenBank	RefSeq	Scientific name	Modifier	Annotation	
<input type="checkbox"/> ASM222028v1 	GCA_002220285.1	GCF_002220285.1	Bacillus cereus	FORC_047 (strain)	NCBI RefSeq Submitter	
<input type="checkbox"/> ASM609429v1	GCA_006094295.1	GCF_006094295.1	Bacillus cereus ATCC 14579	ATCC 14579 (strain)	NCBI RefSeq Submitter	
<input type="checkbox"/> ASM782v1	GCA_000007825.1	GCF_000007825.1	Bacillus cereus ATCC 14579	ATCC 14579 (strain)	NCBI RefSeq Submitter	
<input type="checkbox"/> ASM2165533v1	GCA_021655335.1	GCF_021655335.1	Bacillus cereus	30075 (strain)	NCBI RefSeq Submitter	
<input type="checkbox"/> ASM2920126v1	GCA_029201265.1	GCF_029201265.1	Bacillus cereus	M72-4 (strain)	NCBI RefSeq	

The green check denotes the NCBI designated “reference genome,” or a high-quality genome that has genes that are representative of other strains in the species.

Click here for more info.

Genome assembly ASM222028v1

reference

Download

datasets

curl

		Actions
NCBI RefSeq assembly	GCF_002220285.1	⋮
Submitted GenBank assembly	GCA_002220285.1	⋮
Taxon	<i>Bacillus cereus</i>	
Strain	FORC_047	
Submitter	Food-borne Pathogen Omics Research Center, FORC	
Date	Jul 17, 2017	

[View the legacy Assembly page](#)



[View annotated genes](#)

Additional genomes

[Browse all *Bacillus cereus* genomes \(2503\)](#)

BioProject

[PRJNA336246](#)

Bacillus cereus strain:FORC_047 Genome sequencing and assembly

Pathogen Detection Resource

[Isolate Browser](#)

[Genotypes identified by AMRFinderPlus](#)

Publications

Showing 5 of 87

[Int J Mol Sci 2023](#)

[Saline-Alkali Soil Property Improved by the](#)

General info about the assembly, including assembly accession numbers.

Scroll for more info.

Annotation details

	RefSeq	GenBank
Provider	NCBI RefSeq	Food-borne Pathogen Omics Research Center, FORC
Name	NCBI Prokaryotic Genome Annotation Pipeline (PGAP)	Annotation submitted by Food-borne Pathogen Omics Research Center, FORC
Date	May 4, 2023	Jul 17, 2017
Genes	5,910	6,045
Protein-coding	5,610	5,896
Software version	6.5	
	View RefSeq annotation	View GenBank annotation

About PGAP

The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) uses multiple approaches to predict protein-coding and RNA genes and other functional elements directly from sequence.

[Continue reading](#)

Throughout the page, you'll notice that there is "RefSeq" and "GenBank" data.

RefSeq is a manually curated database that contains reference sequences for genomes.

GenBank is a general DNA sequence database where anybody can submit any sequences. It is not manually curated, and thus is slightly less reliable.

Annotation details

	RefSeq	GenBank
Provider	NCBI RefSeq	Food-borne Pathogen Omics Research Center, FORC
Name	NCBI Prokaryotic Genome Annotation Pipeline (PGAP)	Annotation submitted by Food-borne Pathogen Omics Research Center, FORC
Date	May 4, 2023	Jul 17, 2017
Genes	5,910	6,045
Protein-coding	5,610	5,896
Software version	6.5	

[View RefSeq annotation](#)

[View GenBank annotation](#)



We won't do this now, but from here you can see all the annotated genes in the genome.

Once on the page, you can search for genes using the filter function.

About PGAP

The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) uses multiple approaches to predict protein-coding and RNA genes and other functional elements directly from sequence.

[Continue reading](#)

Chromosomes

Download

Chromosome	GenBank	RefSeq	Size (bp)	GC content (%)	Unlocalized count	Action
chromosome	CP017060.1	NZ_CP017060.1	5,365,433	35.5	0	⋮
pFORC47_1	CP018741.1	NZ_CP018741.1	257,708	32.5	0	⋮
pFORC47_2	CP018742.1	NZ_CP018742.1	213,830	32	0	⋮

Revision history

This record has not been revised

GenBank	RefSeq	Name	Level	Date	Action
GCA_002220285.1	GCF_002220285.1	ASM222028v1	Complete Genome	Jul 17, 2017	⋮

Click the **accession number** to view the genome.

Accession numbers are unique identifiers, like barcodes for a specific sequence.

Accession #s for reference genomes generally start with “NC_” or “NZ_”.

Keep track of accession #s (genome, gene, protein, etc.) if you use any sequences in your research! You’ll reference them in papers.

Bacillus cereus strain FORC_047 chromosome, complete genome

NCBI Reference Sequence: NZ_CP017060.1

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

Go to:

LOCUS NZ_CP017060 5365433 bp DNA circular CON 04-MAY-2023
DEFINITION Bacillus cereus strain FORC_047 chromosome, complete genome.

ACCESSION NZ_CP017060
VERSION NZ_CP017060.1
DBLINK BioProject: [PRJNA224116](#)
BioSample: [SAMN05506965](#)
Assembly: [GCF_002220285.1](#)

KEYWORDS RefSeq.

SOURCE Bacillus cereus

ORGANISM [Bacillus cereus](#)
Bacteria; Bacillota; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 5365433)
AUTHORS Lee,J.-H., Kim,Y.-T., Ku,H.-J., Kim,H., Kim,H., Choi,S.-H. and Ryu,S.

CONSRM Food-borne Pathogen Omics Research Center
TITLE complete genome sequence of *Bacillus cereus* FORC_047
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 5365433)
AUTHORS Lee,J.-H., Kim,Y.-T., Ku,H.-J., Kim,H., Kim,H., Choi,S.-H. and Ryu,S.

CONSRM Food-borne Pathogen Omics Research Center
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2016) Food Science and Biotechnology, Kyung Hee University, 1732, Deogyeong-daero, Giheung-gu, Yongin, Gyeonggi 17104, South Korea

COMMENT [REFSEQ INFORMATION](#): The reference sequence is identical to [CP017060.1](#).
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/
Bacteria and source DNA available from Food-borne Pathogen Omics Research Center.

##Genome-Assembly-Data-START##
Assembly Method :: PacBio SMRT Analysis v. 2.3.0
Genome Coverage :: 169
Sequencing Technology :: PacBio
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI RefSeq
Annotation Date :: 05/04/2023 03:42:51
Annotation Pipeline :: NCBI Prokaryotic Genome Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein set; GeneMarkS-2+

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

Related information

[Assembly](#)
[BioProject](#)
[BioSample](#)
[Protein](#)
[Taxonomy](#)
[Components \(Core\)](#)
[Gene](#)
[Genome](#)
[Identical GenBank Sequence](#)
[PubMed \(Weighted\)](#)

LinkOut to external resources

[UWBM:Mamm:84641](#) [Arctos Specimen Database]
[Bacillus cereus](#) [BacDive]
[Order ilvC cDNA clone/Protein/Antibody/RNAi](#) [OriGene]
[Order Ctsr cDNA clone/Protein/Antibody/RNAi](#) [OriGene]
[Order ackA cDNA clone/Protein/Antibody/RNAi](#) [OriGene]

Recent activity

[Turn Off](#) [Clear](#)
 [Bacillus cereus strain FORC_047 chromosome, complete genome](#) Nucleotide
 [ASM2309401v1 - Genome - Assembly - NCBI](#) Assembly

This is the “legacy” format for the genome page.

```
##Genome-Annotation-Data-START##  
Annotation Provider           :: NCBI RefSeq  
Annotation Date              :: 05/04/2023 03:42:51  
Annotation Pipeline          :: NCBI Prokaryotic Genome  
                               Annotation Pipeline (PGAP)  
Annotation Method            :: Best-placed reference protein  
                               set; GeneMarkS-2+  
Annotation Software revision :: 6.5  
Features Annotated           :: Gene; CDS; rRNA; tRNA; ncRNA  
Genes (total)                :: 5,910  
CDSs (total)                 :: 5,756  
Genes (coding)               :: 5,610  
CDSs (with protein)          :: 5,610  
Genes (RNA)                  :: 154  
rRNAs                        :: 14, 14, 14 (5S, 16S, 23S)  
complete rRNAs               :: 14, 14, 14 (5S, 16S, 23S)  
tRNAs                         :: 107  
ncRNAs                        :: 5  
Pseudo Genes (total)          :: 146  
CDSs (without protein)        :: 146  
Pseudo Genes (ambiguous residues) :: 0 of 146  
Pseudo Genes (frameshifted)    :: 65 of 146  
Pseudo Genes (incomplete)      :: 81 of 146  
Pseudo Genes (internal stop)   :: 39 of 146  
Pseudo Genes (multiple problems) :: 34 of 146  
##Genome-Annotation-Data-END##  
COMPLETENESS: full length.
```

This part is neat!

It tells what NCBI's annotation software found in the genome.

Bacillus cereus strain FORC_047 chromosome, complete genome

NCBI Reference Sequence: NZ_CP017060.1

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

Go to: ▾

LOCUS NZ_CP017060 5365433 bp DNA circular CON 04-MAY-2023
DEFINITION Bacillus cereus strain FORC_047 chromosome, complete genome.

ACCESSION NZ_CP017060
VERSION NZ_CP017060.1
DBLINK BioProject: [PRJNA224116](#)
BioSample: [SAMN05506965](#)
Assembly: [GCF_002220285.1](#)

KEYWORDS RefSeq.

SOURCE Bacillus cereus

ORGANISM [Bacillus cereus](#)

Bacteria; Bacillota; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.

REFERENCE 1 (bases 1 to 5365433)
AUTHORS Lee,J.-H., Kim,Y.-T., Ku,H.-J., Kim,H., Kim,H., Choi,S.-H. and
Ryu,S.

CONSRM Food-borne Pathogen Omics Research Center
TITLE complete genome sequence of *Bacillus cereus* FORC_047
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 5365433)
AUTHORS Lee,J.-H., Kim,Y.-T., Ku,H.-J., Kim,H., Kim,H., Choi,S.-H. and
Ryu,S.

CONSRM Food-borne Pathogen Omics Research Center
TITLE Direct Submission
JOURNAL Submitted (04-NOV-2016) Food Science and Biotechnology, Kyung Hee
University, 1732, Deogyeong-daero, Giheung-gu, Yongin, Gyeonggi
17104, South Korea

COMMENT [REFSEQ INFORMATION](#): The reference sequence is identical to
[CP017060.1](#).
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/
Bacteria and source DNA available from Food-borne Pathogen Omics
Research Center.

##Genome-Assembly-Data-START##
Assembly Method :: PacBio SMRT Analysis v. 2.3.0
Genome Coverage :: 169
Sequencing Technology :: PacBio
##Genome-Assembly-Data-END##

##Genome-Annotation-Data-START##
Annotation Provider :: NCBI RefSeq
Annotation Date :: 05/04/2023 03:42:51
Annotation Pipeline :: NCBI Prokaryotic Genome
Annotation Pipeline (PGAP)
Annotation Method :: Best-placed reference protein
set; GeneMarkS-2+

Customize view ▾

Analyze this sequence

Run BLAST

Pick Primers

Related information

Assembly

BioProject

BioSample

Protein

Taxonomy

Components (Core)

Gene

Genome

Identical GenBank Sequence

PubMed (Weighted)

LinkOut to external resources

UWBM:Mamm:84641

[Arctos Specimen Database]

Bacillus cereus

[BacDive]

Order ilvC cDNA clone/Protein/Antibody/RNAi

[OriGene]

Order Ctsr cDNA clone/Protein/Antibody/RNAi

[OriGene]

Order ackA cDNA clone/Protein/Antibody/RNAi

[OriGene]

Recent activity

[Turn Off](#) [Clear](#)

Bacillus cereus strain FORC_047
chromosome, complete genome Nucleotide

ASM2309401v1 - Genome - Assembly - NCBI Assembly

Click here to view the genome.

Clicking FASTA here will show the entire genome sequence. Don't do that now!

- The green bars are genes! Zoom in to see them up close.
- They'll turn red if they encode proteins and blue if they do not (tRNAs, rRNAs, etc.)

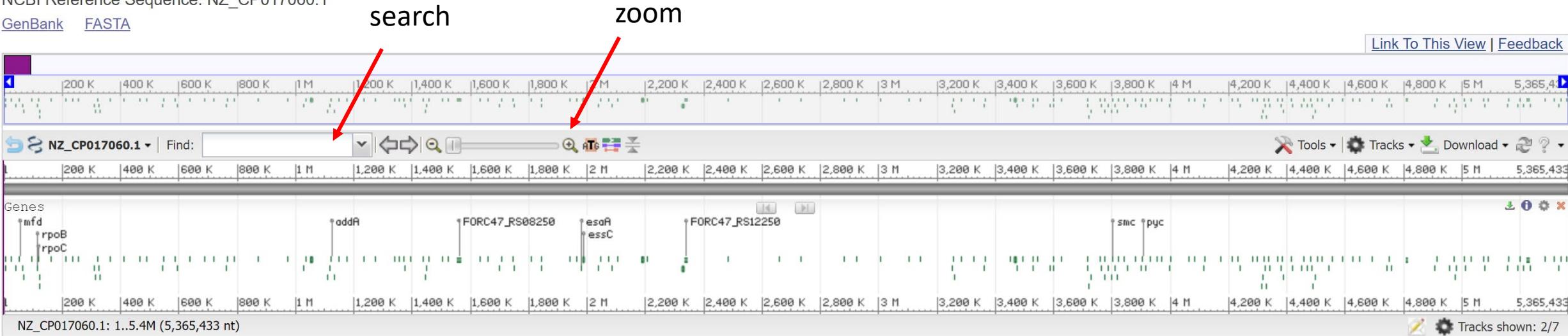
Graphics ▾

Send to: ▾

Bacillus cereus strain FORC_047 chromosome, complete genome

NCBI Reference Sequence: NZ_CP017060.1

[GenBank](#) [FASTA](#)



Let's find and download the sequence for the toxin gene *hbIC*.

Search “hblC”
Double-click the result in the pop-up.

Graphics ▾

Send to: ▾

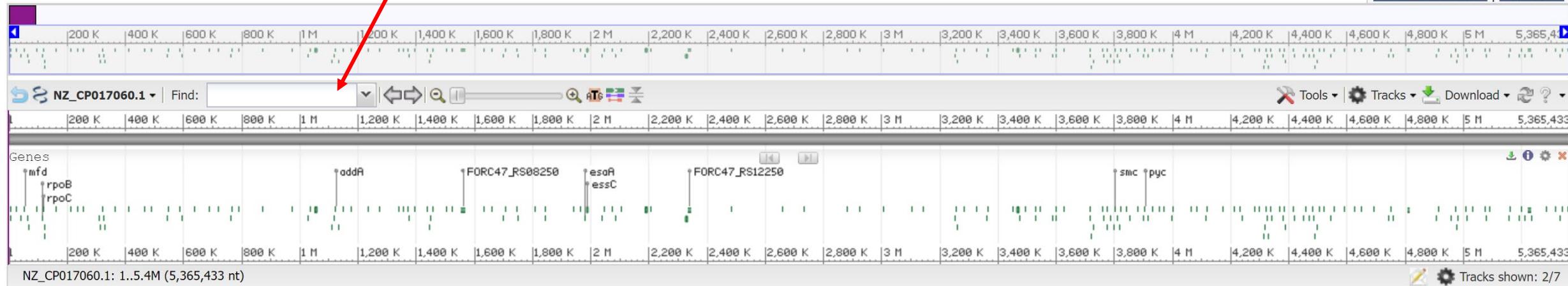
Bacillus cereus strain FORC_047 chromosome, complete genome

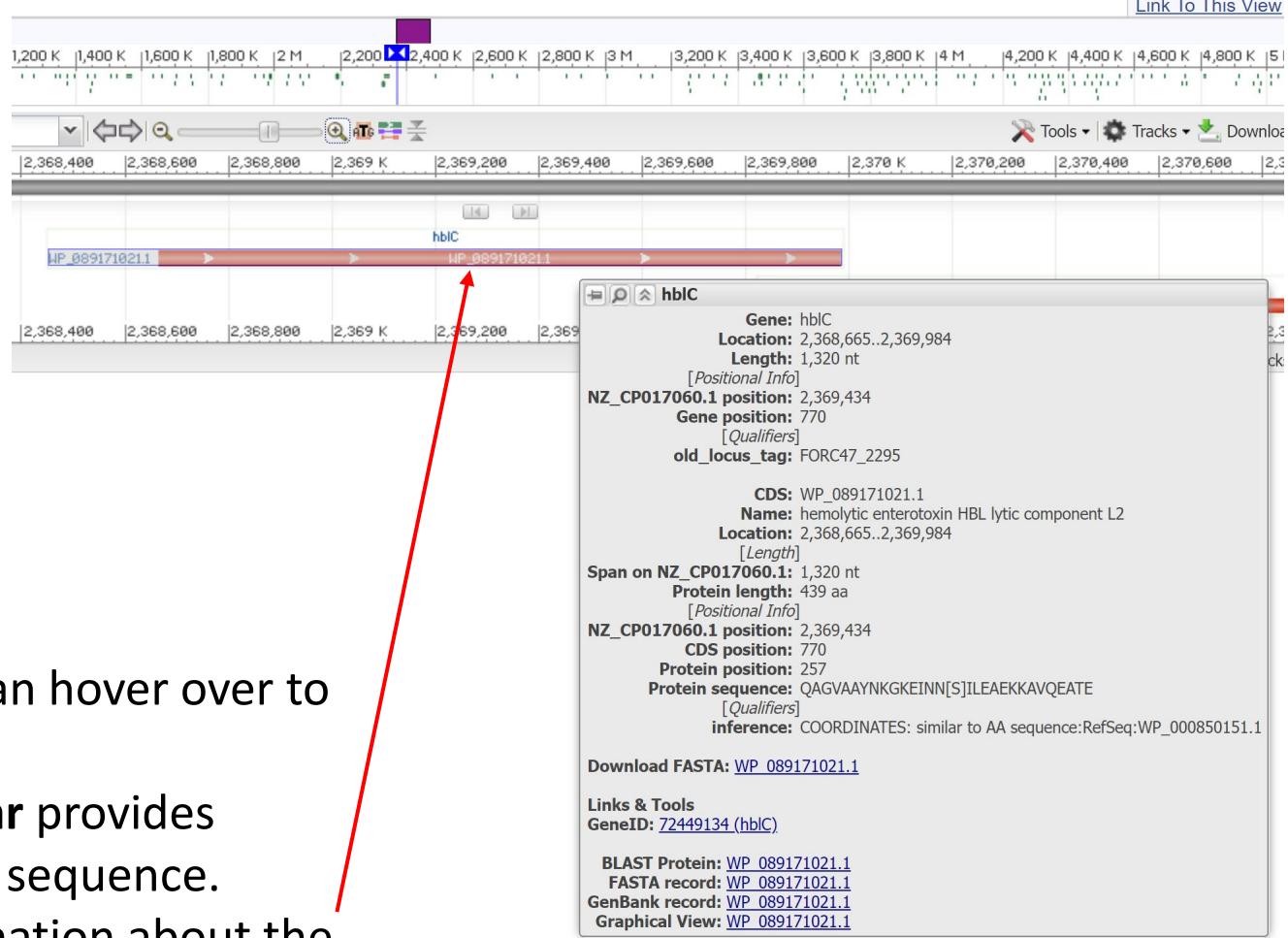
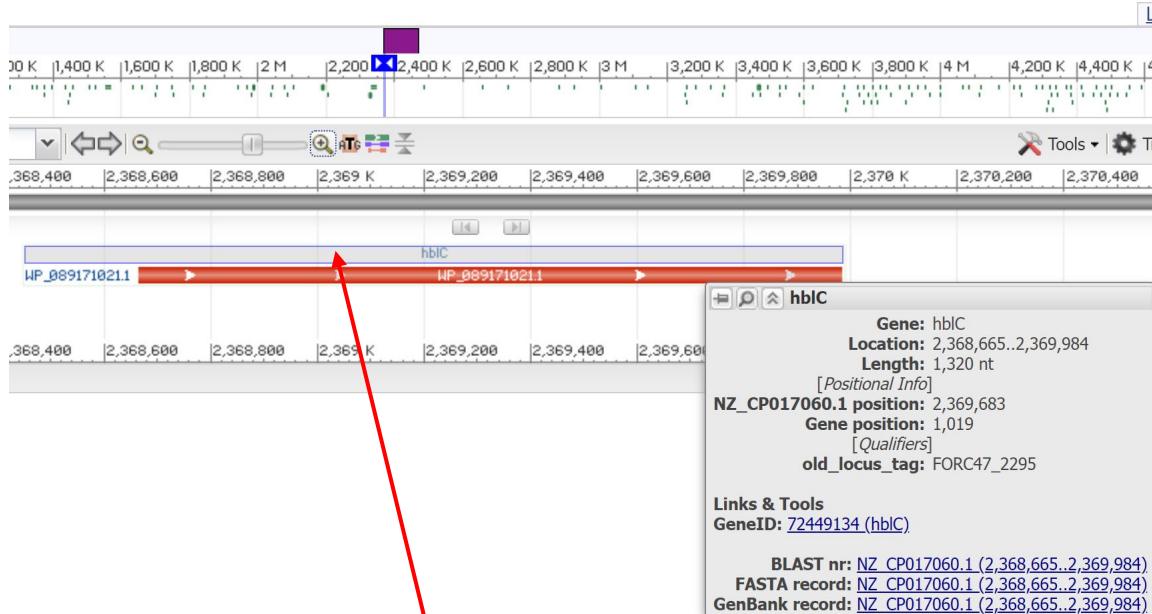
NCBI Reference Sequence: NZ_CP017060.1

[GenBank](#) [FASTA](#)

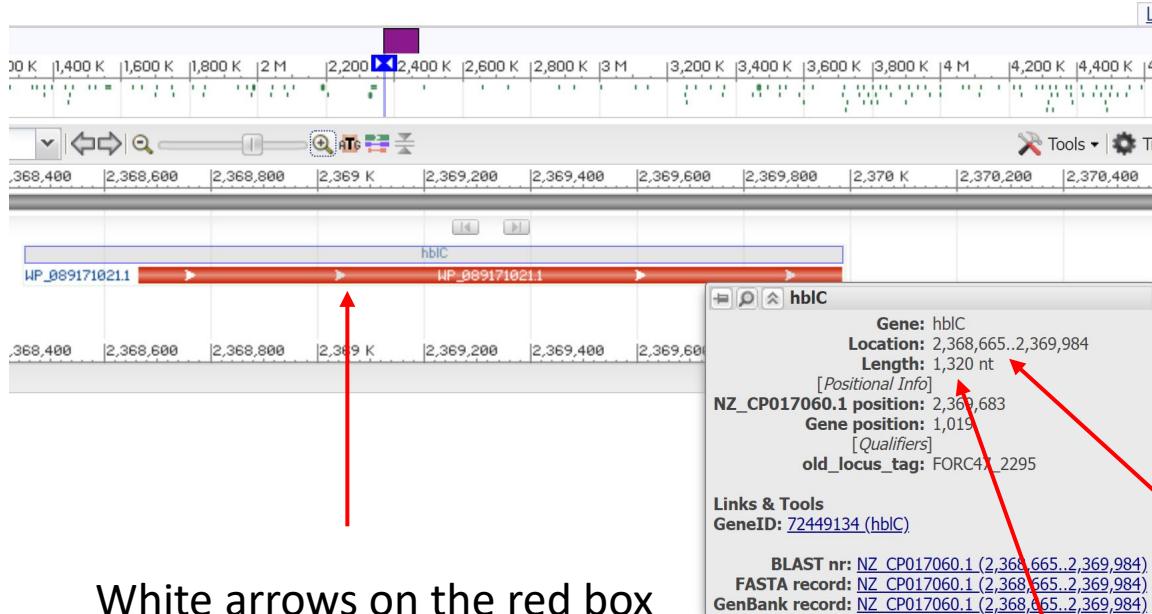
search

[Link To This View](#) | [Feedback](#)

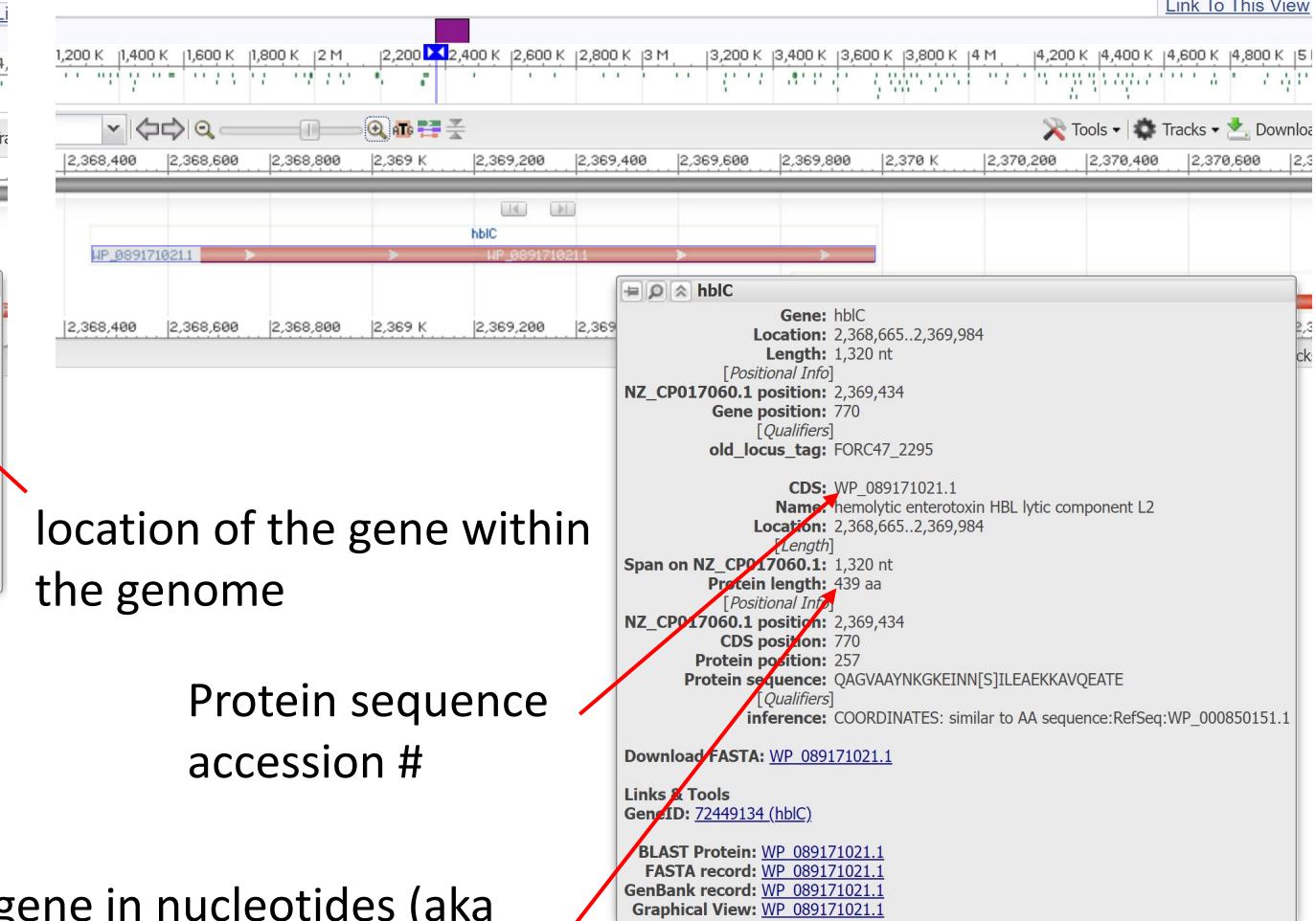




- There are now two things you can hover over to learn more about the gene:
 - The **name above the red bar** provides information about the **DNA sequence**.
 - The **red bar** provides information about the **protein sequence**.



White arrows on the red box show you which direction the gene is translated.

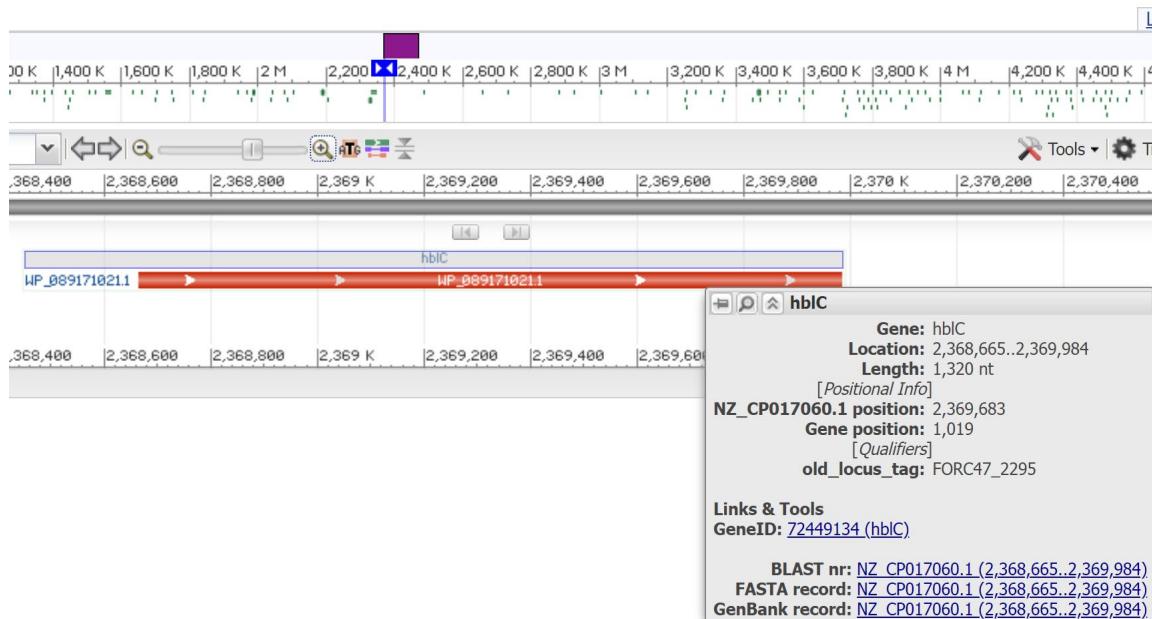


location of the gene within the genome

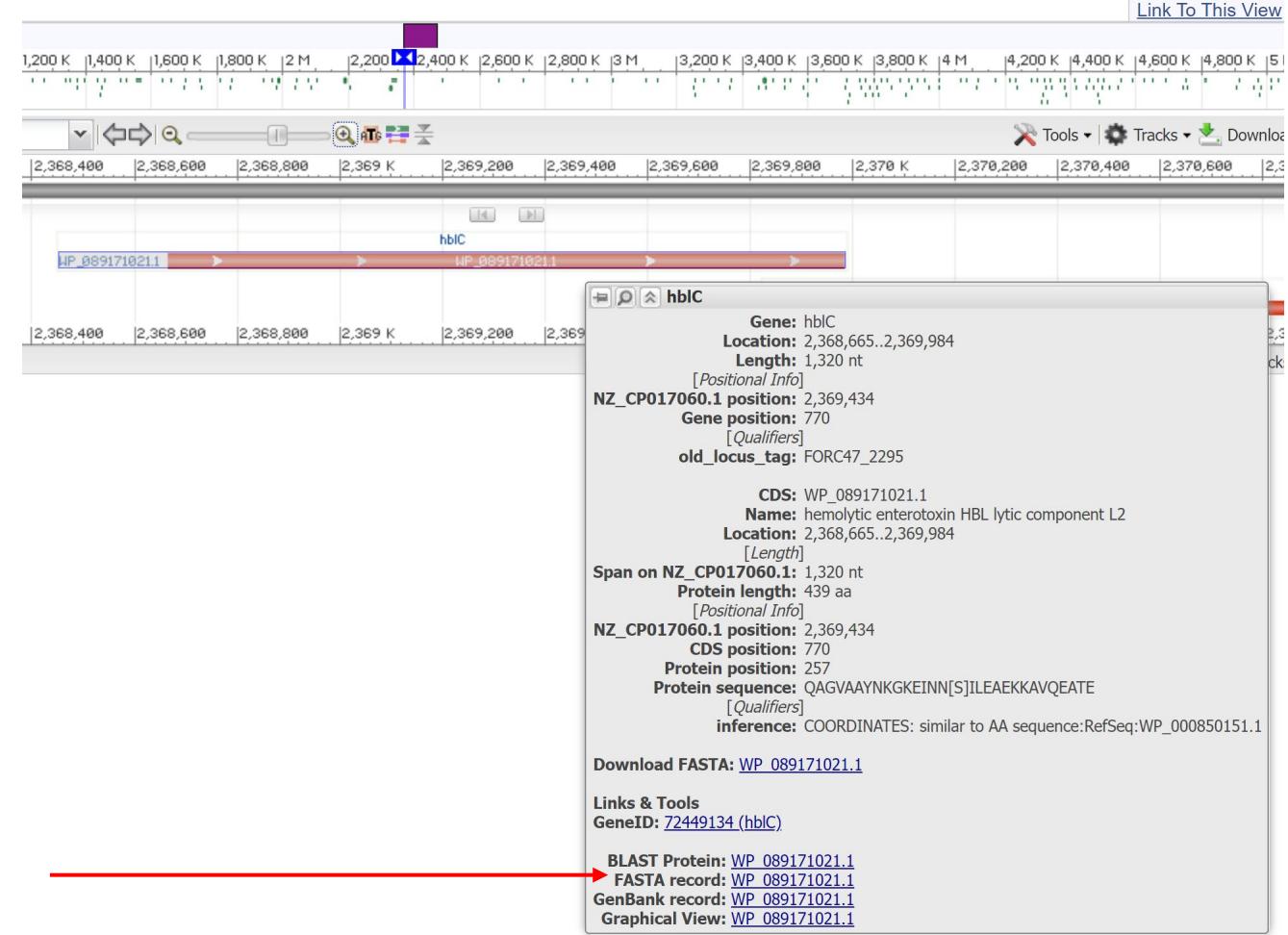
Protein sequence accession #

length of the gene in nucleotides (aka base pairs or “bp”)

length of the protein in amino acids



Click the link after “**FASTA record**”
to get the AA sequence.



GenPept

Send to:

This record is a non-redundant protein sequence. Please [read more here](#).

hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus]

 Download Datasets

NCBI Reference Sequence: WP_089171021.1

[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS WP_089171021 439 aa linear BCT 10-JUL-2019
DEFINITION hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus].
ACCESSION WP_089171021
VERSION WP_089171021.1
KEYWORDS RefSeq.
SOURCE Bacillus cereus
ORGANISM [Bacillus cereus](#)
Bacteria; Bacillota; Bacilli; Bacillales; Bacillaceae; Bacillus;
Bacillus cereus group.
REFERENCE 1 (residues 1 to 439)
AUTHORS Liu,B., Zheng,D., Jin,Q., Chen,L. and Yang,J.
TITLE VFDB 2019: a comparative pathogenomic platform with an interactive
web interface
JOURNAL Nucleic Acids Res. 47 (D1), D687-D692 (2019)
PUBMED [30395255](#)
REFERENCE 2 (residues 1 to 439)
AUTHORS Ryan,P.A., Macmillan,J.D. and Zilinskas,B.A.
TITLE Molecular cloning and characterization of the genes encoding the L1
and L2 components of hemolysin BL from Bacillus cereus
JOURNAL J. Bacteriol. 179 (8), 2551-2556 (1997)
PUBMED [9098052](#)
COMMENT REFSEQ: This record represents a single, non-redundant, protein
sequence which may be annotated on many different RefSeq genomes
from the same, or different, species.

##Evidence-For-Name-Assignment-START##
Evidence Category :: BlastRule
Evidence Accession :: [NBR008547](#)
Evidence Source :: NCBI

Switch this to FASTA
to see the sequence.

FASTA ▾

Send to: ▾

 This record is a non-redundant protein sequence. Please [read more here](#).

hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus]

NCBI Reference Sequence: WP_089171021.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>WP_089171021.1 hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus]
MKTKitGLLVTsIVTGGNIPINTLATPIVQAETQQEGMDISSLRKLGAQSKLIQTYIDQSLMSPNVQL
EEVTALNTNQFLIKQDMKEWSSELYPQLILLNSKSKGFVTKFNSYYPTLNSFVDNKEDREGFSDRLEVLQ
EMAMTNQENTQRQINELTDLKLQLDKKLKDFTNVATAQGILSTDGTGKIDQLKNEILNTKKAIQNDLQQ
IALIPGALNEQGFAIFKEVYSLSKEIIEPATQAGVAAYNKGKEINNSILEAEKKAVQEATEQGKTALEIE
SAKKAAREAIEKSQGEIAAAAAKTQEYDLMKVIDTEKIKKTFGVFAEVNLTAEQRAYLDDLEKQNQK
IYDLTTKLSIADLQKSMLLTQNDLHTFANQVDVELLLKRYKEDLNLIKNSITKLSTNVDTTSEQSQKD
TLRQLKNVISYLEEQVYKF
```

 [Download Datasets](#)

Click here to
download the fasta
file of this gene.

- Here you can see the protein sequence!
- “Fasta” is the file format that sequence data is stored in.
 - Anything on the same line as the “>” is a title/description for the sequence.
 - Everything below it is the sequence.

FASTA

This record is a non-redundant protein sequence. Please [read more here](#).

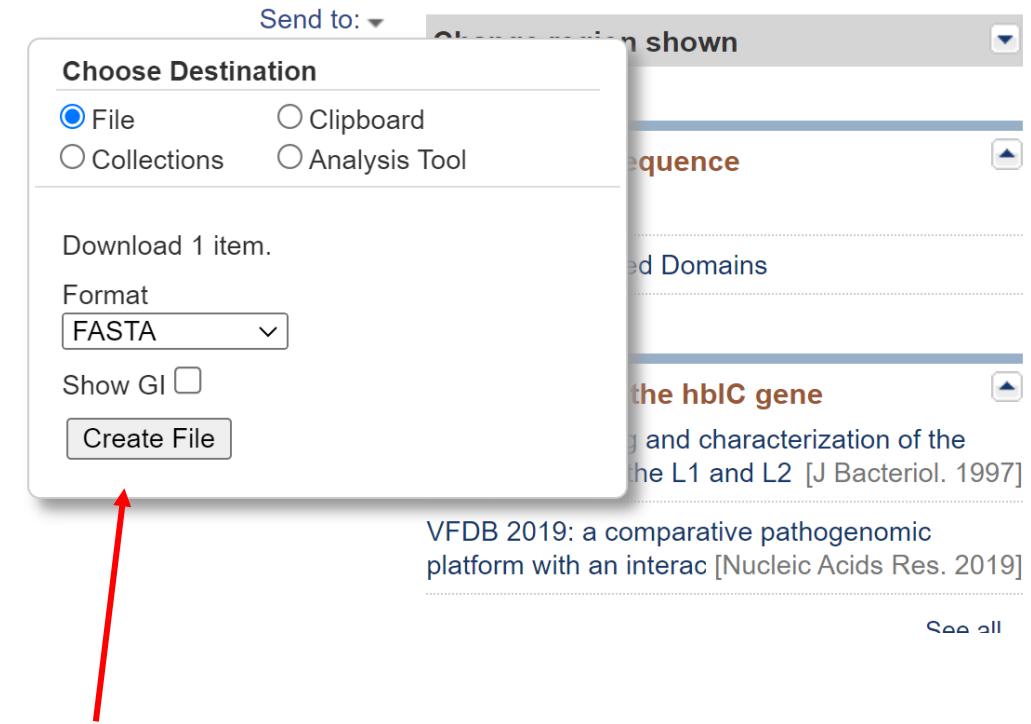
hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus]

NCBI Reference Sequence: WP_089171021.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

>WP_089171021.1 hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus]

```
MKTKIITGLLVTSLVTGGNIPINTLATPIVQAETQQEGMDISSSLRKLGQSKLIQTYIDQSLMSPNVQL  
EEVTALNTNQFLIKQDMKEWSSELYPQLILLNSKSKGFTVKFNSYYPTLNSFVDNKEDREGFSRLEVLQ  
EMAMTNQENTQRQINELTDLKLQLDKKLQDFDTNVATAQGILSTDGTGKIDQLKNEILNTKKAIQNDLQQ  
IALIPGALNEQGFAIFKEVYSLSKIEIPEPATQAGVAAYNKGKEINNSILEAEKKAVQEATEQGKTALEIE  
SAKKAAREAIEKSQGEIAAAAAAKTQEYDLMKVIDTEKIKKTFGVFAEVNLTAEQRAYLDDLEKQNQK  
IYDLTTKLSIADLKSMLLTQNDLHTFANQVDVELDLLKRYKEDLNLIKNSITKLSTNVDTTSEQSQKD  
TLRQLKNVISYLEEQQVYKF
```



Select these options and create file.

- You can now open your fasta file in your computer's Notepad app! (TextEdit on Mac)
- You can use this fasta file in many bioinformatics programs.

BLAST

- Link to webpage: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- Stands for “Basic Local Alignment Search Tool”
- Operated and owned by NCBI
- Rapidly finds a given query sequence in a database or specific sequence.
 - The “query” is the sequence given the program to find in other sequences.

Let's see what other organisms have genes similar
to *hbIC*!



Take the BLAST survey today

[Start survey](#)

Basic Local Alignment Search Tool

BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance.

[Learn more](#)

NEWS

BLAST+ 2.14.0 is here!

BLASTP, BLASTX, and TBLASTN are faster than before.

Fri, 28 Apr 2023

[More BLAST news...](#)

Web BLAST



blastx

translated nucleotide ▶ protein

tblastn

protein ▶ translated nucleotide



BLAST Genomes

Enter organism common name, scientific name, or tax id

Search

Human

Mouse

Rat

Microbes

We'll use protein BLAST today. Click here.

blastn

blastp

blastx

tblastn

tblastx

BLA

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

[Choose File](#) No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases

Standard databases (nr etc.): [New](#) Experimental databases

 Try experimental clustered nr database
For more info see [What is clustered nr?](#)

Standard

Database

Non-redundant protein sequences (nr) [?](#)

Organism
Optional

Enter organism name or id—completions will be suggested exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude
Optional

Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
 blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

+ Algorithm parameters

- Here are all the options for BLAST.
 - Blastn is for nucleotide queries against nucleotide databases.
 - Blastp is for protein queries against protein databases.

blastn

blastp

blastx

tblastn

tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From
To

Or, upload file No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases Standard databases (nr etc.) [New](#) Experimental databases [Try experimental clustered nr database](#)
For more info see [What is clustered nr?](#)

Standard

Database [?](#)

Organism Optional Enter organism name or id--completions will be suggested exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm Quick BLASTP (Accelerated protein-protein BLAST) blastp (protein-protein BLAST) PSI-BLAST (Position-Specific Iterated BLAST) PHI-BLAST (Pattern Hit Initiated BLAST) DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm [?](#)

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

You could:

Paste a sequence or accession #.

Choose this option to search for a specific sequence in another sequence of interest.

Select your database or organism.

Select a search algorithm.

I selected Quick BLASTP in the interest of time.

blastn

blastp

blastx

tblastn

tblastx

BLA

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

No file chosen

[?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases

Standard databases (nr etc.): [New](#) Experimental databases

 [Try experimental clustered nr database](#)
For more info see [What is clustered nr?](#)

Standard

Database

Non-redundant protein sequences (nr) [?](#)

Organism
Optional

Enter organism name or id--completions will be suggested exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude
Optional

Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
 blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

+ Algorithm parameters

Click here and select the fasta file we just downloaded.

blastn

blastp

blastx

tblastn

tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file Choose File No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases Standard databases (nr etc.): [New](#) Experimental databases [Try experimental clustered nr database](#) For more info see [What is clustered nr?](#)

Standard

Database

Organism Optional Enter organism name or id--completions will be suggested exclude [Add organism](#)

Exclude Optional Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm Quick BLASTP (Accelerated protein-protein BLAST) blastp (protein-protein BLAST) PSI-BLAST (Position-Specific Iterated BLAST) PHI-BLAST (Pattern Hit Initiated BLAST) DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)** Show results in a new window

+ Algorithm parameters

Type “**Bacillus cereus**” in the organism box and click “**exclude**”.

- We want to see organisms other than *B. cereus* that have hblC, so we use “**exclude**”.
- Make sure you pick the option that DOESN’T say “group” in it.

blastn

blastp

blastx

tblastn

tblastx

BLA

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

[Choose File](#) No file chosen [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Databases

Standard databases (nr etc.): [New](#) Experimental databases

 **Try experimental clustered nr database**
For more info see [What is clustered nr?](#)

Standard

Database

Non-redundant protein sequences (nr) [?](#)

Organism
Optional

Enter organism name or id--completions will be suggested exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude
Optional

Models (XM/XP) Non-redundant RefSeq proteins (WP) Uncultured/environmental sample sequences

Program Selection

Algorithm

- Quick BLASTP (Accelerated protein-protein BLAST)
 blastp (protein-protein BLAST)
 PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search **database nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

+ Algorithm parameters

BLAST!

How BLAST works

- Aligns the query sequence with a target sequence based on scoring.
 - Each matching nucleotide is given a score of +1
 - Every mismatch is given a score of -2
- Tallies up the scores for each attempt.
- Gives you information only on the results with the highest scores.

Local Alignment

Target Sequence

5' ACTACTAGATTACCTACGGATCAGGTACTTAGAGGCTTGCAACCA 3'

||||| ||||| ||||| ||||||||| |||||

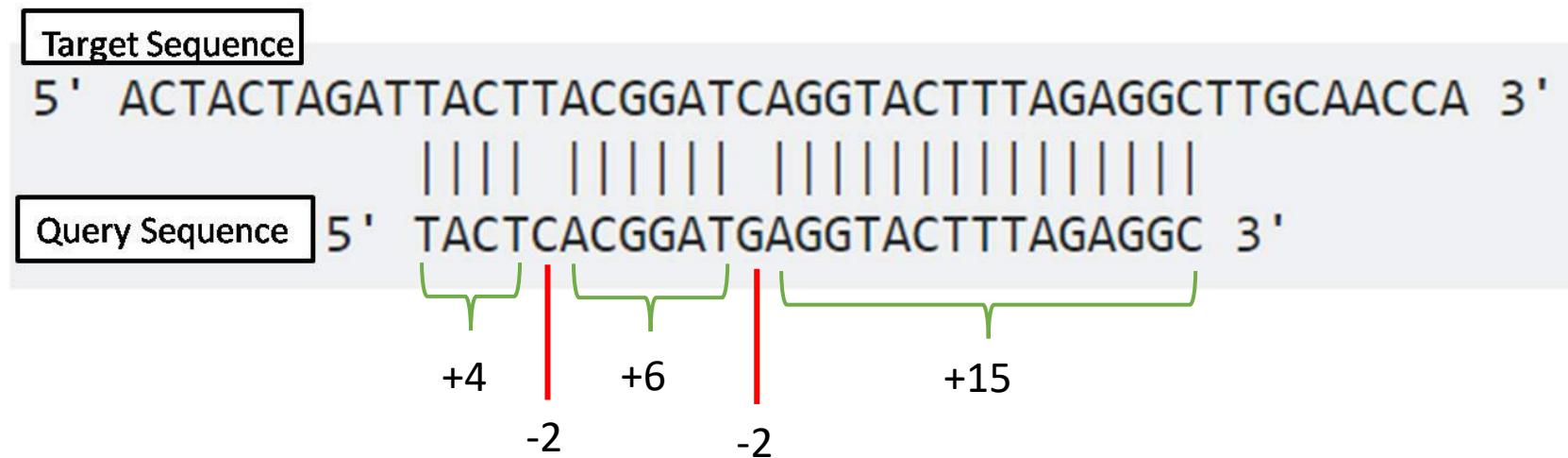
Query Sequence

5' TACTCACGGATGAGGTACTTAGAGGC 3'

How would BLAST score this sequence alignment?

1. Record the number of matching bases and mismatching bases.
2. Score the matches and mismatches and sum the scores.
(matches are +1, mismatches are -2)

Local Alignment



$$\text{Total score} = 4 + 6 + 15 - 2 - 2 = \mathbf{21}$$

AA alignments are a little different...

- BLOSUM (BLOcks Substitution Matrices)
 - Used to score amino acid alignments based on how likely the amino acid mutation is to occur
 - Originally calculated by assessing the substitutions of small portions (blocks) within protein sequence alignments
- PAM matrices are the alternative
 - Originally calculated by assessing the substitutions of entire protein sequence alignments
 - Developed by Margaret Dayhoff, “mother of bioinformatics” :)

C	S	T	A	G	P	D	E	Q	N	H	R	K	M	I	L	V	W	Y	F
C	9																		C
S	-1	4																S	
T	-1	1	5															T	
A	0	1	0	4														A	
G	-3	0	-2	0	6													G	
P	-3	-1	-1	-1	7													P	
D	-3	0	-1	-2	-1	-1	6											D	
E	-4	0	-1	-1	-2	-1	2	5										E	
Q	-3	0	-1	-1	-2	-1	0	2	5									Q	
N	-3	1	0	-2	0	-2	1	0	0	6								N	
H	-3	-1	-2	-2	-2	-2	-1	0	0	1	8							H	
R	-3	-1	-1	-1	-2	-2	-2	0	1	0	0	5						R	
K	-3	0	-1	-1	-2	-1	-1	1	1	0	-1	2	5					K	
M	-1	-1	-1	-1	-3	-2	-3	-2	0	-2	-2	-1	-1	5				M	
I	-1	-2	-1	-1	-4	-3	-3	-3	-3	-3	-3	-3	-3	1	4			I	
L	-1	-2	-1	-1	-4	-3	-4	-3	-2	-3	-3	-2	-2	2	2	4		L	
V	-1	-2	0	0	-3	-2	-3	-2	-2	-3	-3	-3	-2	1	3	1	4	V	
W	-2	-3	-2	-3	-2	-4	-4	-3	-2	-4	-2	-3	-3	-1	-3	-2	-3	11	W
Y	-2	-2	-2	-2	-3	-3	-3	-2	-1	-2	2	-2	-2	-1	-1	-1	-1	2	Y
F	-2	-2	-2	-2	-3	-4	-3	-3	-3	-3	-1	-3	-3	0	0	0	-1	1	F
C		S	T	A	G	P	D	E	Q	N	H	R	K	M	I	L	V	W	C

BLOSUM62

PSI-BLAST (Position-Specific Iterated BLAST)
 PHI-BLAST (Pattern Hit Initiated BLAST)
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm ?

BLAST | Search database nr using Blastp (protein-protein BLAST) | Show results in a new window

Algorithm parameters

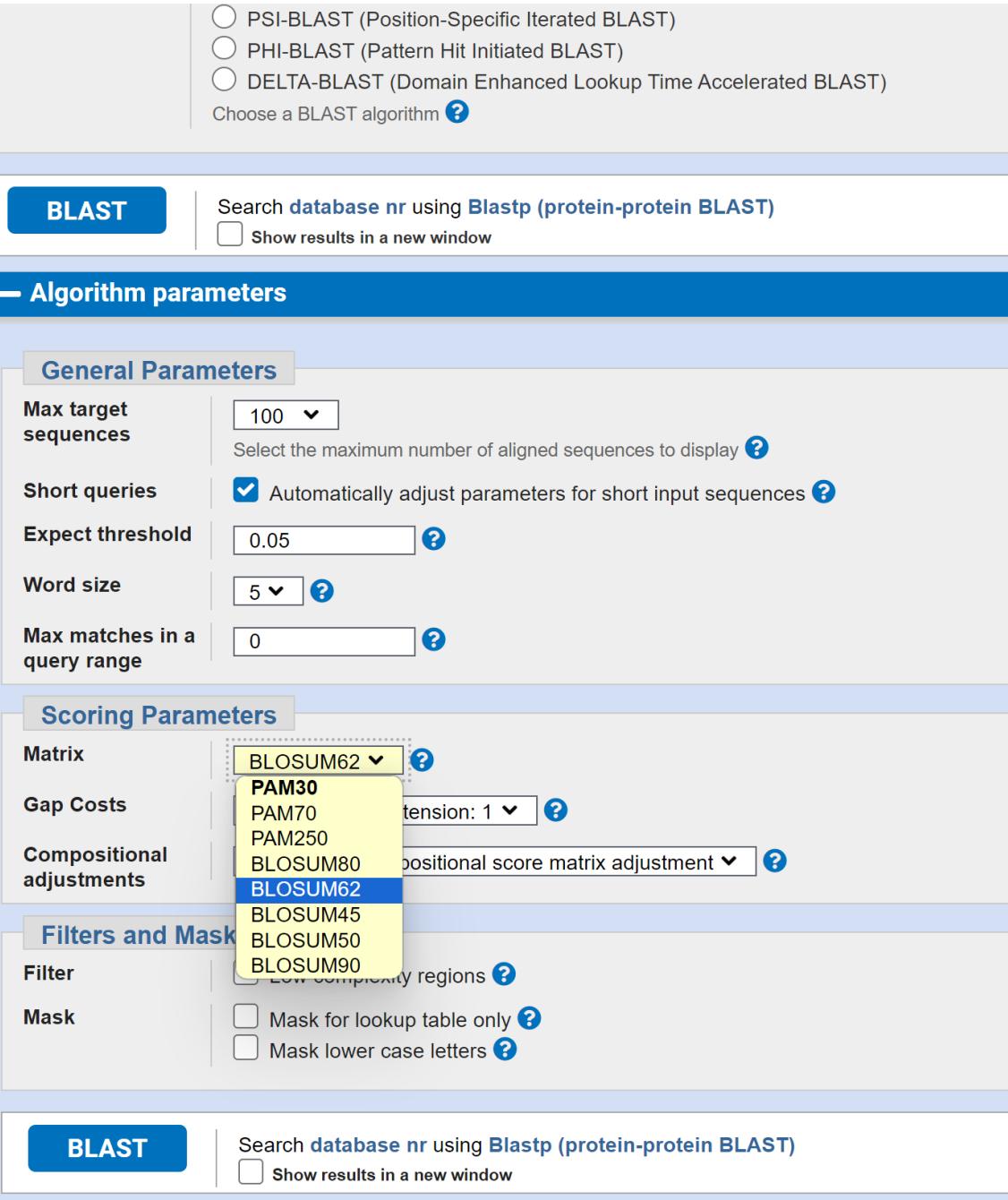
General Parameters

Max target sequences	100 <input type="button" value="▼"/>	Select the maximum number of aligned sequences to display ?
Short queries	<input checked="" type="checkbox"/> Automatically adjust parameters for short input sequences ?	
Expect threshold	0.05 <input type="button" value="?"/>	
Word size	5 <input type="button" value="▼"/> <input type="button" value="?"/>	
Max matches in a query range	0 <input type="button" value="?"/>	

Scoring Parameters

Matrix	BLOSUM62 <input type="button" value="?"/>
Gap Costs	PAM30 PAM70 PAM250 BLOSUM80 <input type="button" value="extension: 1 ▼"/> <input type="button" value="?"/>
Compositional adjustments	<input type="button" value="compositional score matrix adjustment"/> <input type="button" value="?"/>
Filters and Masks	
Filter	Low complexity regions <input type="button" value="?"/>
Mask	<input type="checkbox"/> Mask for lookup table only <input type="button" value="?"/> <input type="checkbox"/> Mask lower case letters <input type="button" value="?"/>

BLAST | Search database nr using Blastp (protein-protein BLAST) | Show results in a new window



If you scroll to the bottom of the blastp page, you could change your scoring matrix.

BLOSUM62 is the default because it is the most efficient!

[Edit Search](#)[Save Search](#)[Search Summary](#)[How to read this report?](#)[BLAST Help Videos](#)[Back to Traditional Results Page](#)

i Your search is limited to records that exclude: *Bacillus cereus* (taxid:1396)

Job Title [ref|WP_089171021.1|](#)RID [0BNBUCNJ016](#) Search expires on 03-30 04:39 am [Download All](#)Program Quick BLASTP [Citation](#)Database nr [See details](#)Query ID [WP_089171021.1](#)

Description hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus]

Molecule type amino acid

Query Length 439

Other reports [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)**Filter Results**Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity

E value

Query Coverage

[] to []

[] to []

[] to []

[Filter](#)[Reset](#)[Descriptions](#)[Graphic Summary](#)[Alignments](#)[Taxonomy](#)**Sequences producing significant alignments**[Download](#)[Select columns](#)

Show

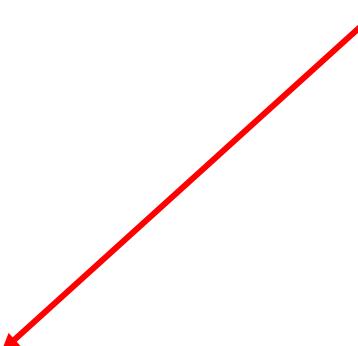
100

[?](#) select all 100 sequences selected[GenPept](#)[Graphics](#)[Distance tree of results](#)[Multiple alignment](#)[MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	870	870	100%	0.0	99.54%	439	WP_078401515.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	870	870	100%	0.0	99.32%	439	WP_000850151.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis	870	870	100%	0.0	99.32%	439	WP_144465172.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_078186771.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_053564472.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis	869	869	100%	0.0	99.32%	439	WP_078993245.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	869	869	100%	0.0	99.32%	439	WP_000850152.1

The results are in!
Here's how to read them.

- Right now, we have the “Descriptions” tab selected.
 - This is the best for a quick glance at what most similar hits are.
- We’ll investigate the other tabs later.



The screenshot shows a bioinformatics search interface. At the top, there are four tabs: 'Descriptions' (selected), 'Graphic Summary', 'Alignments', and 'Taxonomy'. Below the tabs, a section titled 'Sequences producing significant alignments' displays a table of results. The table has columns for selecting all sequences (checkbox), Description, Scientific Name, Max Score, Total Score, Query Cover, E value, Per. Ident, Acc. Len, and Accession. The table lists seven entries, each corresponding to a different strain of Bacillus, all identified as 'hemolytic enterotoxin HBL lytic component L2'. The 'E value' column is sorted in descending order, with values ranging from 0.0 to 870.

<input checked="" type="checkbox"/> select all 100 sequences selected	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	870	870	100%	0.0	99.54%	439	WP_078401515.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	870	870	100%	0.0	99.32%	439	WP_000850151.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis	870	870	100%	0.0	99.32%	439	WP_144465172.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_078186771.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_053564472.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis	869	869	100%	0.0	99.32%	439	WP_078993245.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	869	869	100%	0.0	99.32%	439	WP_000850152.1

In this tab, we have the following:

Description of the species and strain.

Accession # for the protein hit.

Measures of similarity.

Sequences producing significant alignments											Download	Select columns	Show	100	?
						GenPept	Graphics	Distance tree of results			Multiple alignment		MSA Viewer		
		Description		Scientific Name		Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession			
<input checked="" type="checkbox"/>	select all	100 sequences selected													
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus		870	870	100%	0.0	99.54%	439	WP_078401515.1			
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus		870	870	100%	0.0	99.32%	439	WP_000850151.1			
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis		870	870	100%	0.0	99.32%	439	WP_144465172.1			
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group		869	869	100%	0.0	99.32%	439	WP_078186771.1			
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group		869	869	100%	0.0	99.32%	439	WP_053564472.1			
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis		869	869	100%	0.0	99.32%	439	WP_078993245.1			
<input checked="" type="checkbox"/>			hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus		869	869	100%	0.0	99.32%	439	WP_000850152.1			

Breaking down the measures of similarity...

- **Score:** related to the number of exact nucleotide matches to the query. What we just learned to calculate!
 - **Query coverage:** the % of the query length that is found in the matching segment.
 - **E-value:** the number of times a hit with the calculated score would happen by chance.
 - **Percent identity:** the % of bases matching the query.

Descriptions	Graphic Summary	Alignments	Taxonomy										
Sequences producing significant alignments				Download	Select columns	Show	100	?					
<input checked="" type="checkbox"/> select all 100 sequences selected				GenPept	Graphics	Distance tree of results		Multiple alignment		MSA Viewer			
	Description			Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]			Bacillus	870	870	100%	0.0	99.54%	439	WP_078401515.1		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]			Bacillus	870	870	100%	0.0	99.32%	439	WP_000850151.1		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]			Bacillus thuringiensis	870	870	100%	0.0	99.32%	439	WP_144465172.1		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]			Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_078186771.1		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]			Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_053564472.1		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]			Bacillus thuringiensis	869	869	100%	0.0	99.32%	439	WP_078993245.1		
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]			Bacillus	869	869	100%	0.0	99.32%	439	WP_000850152.1		

Breaking down the measures of similarity...

- **Query coverage:** the % of the query length that is found in the matching segment.
- **Percent identity:** the % of bases matching the query.

Local Alignment

Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTAGAGGCTTGCAACCA 3'

||||| ||||| ||||||| ||||||||| |||||

Query Sequence

5' TACTCACGGATGAGGTACTTAGAGGC 3'

For a close match, you want the following to be:

- **Score** - high
 - **Query coverage** - high
 - **E-value** - low
 - **Percent identity** - high

Descriptions	Graphic Summary	Alignments	Taxonomy										
Sequences producing significant alignments				Download	Select columns	Show	100	?					
<input checked="" type="checkbox"/> select all 100 sequences selected				GenPept	Graphics	Distance tree of results		Multiple alignment		MSA Viewer			
	Description			Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len		Accession	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]			Bacillus	870	870	100%	0.0	99.54%	439		WP_078401515.1	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]			Bacillus	870	870	100%	0.0	99.32%	439		WP_000850151.1	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]			Bacillus thuringiensis	870	870	100%	0.0	99.32%	439		WP_144465172.1	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]			Bacillus cereus group	869	869	100%	0.0	99.32%	439		WP_078186771.1	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]			Bacillus cereus group	869	869	100%	0.0	99.32%	439		WP_053564472.1	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]			Bacillus thuringiensis	869	869	100%	0.0	99.32%	439		WP_078993245.1	
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]			Bacillus	869	869	100%	0.0	99.32%	439		WP_000850152.1	

We can view alignments of our query and target sequences by clicking on the “Alignments” tab!

(Make sure you have at least one sequence selected with a blue checkbox.)



Descriptions	Graphic Summary	Alignments	Taxonomy						
Sequences producing significant alignments									
<input checked="" type="checkbox"/> select all 100 sequences selected	Download	Select columns	Show 100 ?						
GenPept Graphics Distance tree of results Multiple alignment MSA Viewer									
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	870	870	100%	0.0	99.54%	439	WP_078401515.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	870	870	100%	0.0	99.32%	439	WP_000850151.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis	870	870	100%	0.0	99.32%	439	WP_144465172.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_078186771.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus cereus group]	Bacillus cereus group	869	869	100%	0.0	99.32%	439	WP_053564472.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus thuringiensis]	Bacillus thuringiensis	869	869	100%	0.0	99.32%	439	WP_078993245.1
<input checked="" type="checkbox"/>	hemolytic enterotoxin HBL lytic component L2 [Bacillus]	Bacillus	869	869	100%	0.0	99.32%	439	WP_000850152.1

[Descriptions](#)[Graphic Summary](#)[Alignments](#)[Taxonomy](#)

Alignment view

Pairwise

[Restore defaults](#)

100 sequences selected

[Download](#) [GenPept](#) [Graphics](#)[▼ Next](#) [▲ Previous](#) [◀ D](#)**MULTISPECIES: hemolytic enterotoxin HBL lytic component L2 [Bacillus]**Sequence ID: [WP_078401515.1](#) Length: 439 Number of Matches: 1[See 5 more title\(s\)](#) [▼](#) [See all Identical Proteins\(IPG\)](#)Range 1: 1 to 439 [GenPept](#) [Graphics](#)[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
870 bits(2249)	0.0	Compositional matrix adjust.	437/439(99%)	437/439(99%)	0/439(0%)

Query 1 MKTKIITGLLVTISVTGGNIPINTLATPIVQAETQQEGMDISSLRKLGAQSKLIQTYID 60

Sbjct 1 MKTKIITGLLVTISVTGGNIPINTLATPIVQAETQQEGMDISSLRKLGAQSKLIQTYID 60

Query 61 QSLMSPNVQLEEVTAALNTNQFLIKQDMKEWSSELYPQLILLNSKSKGKVTKFNSYYPTLN 120

Sbjct 61 QSLMSPNVQLEEVTAALNTNQFLIKQDMKEWSSELYPQLILLNSKSKGKVTKFNSYYPTLN 120

Query 121 SFVDNKEDREGFSDRLEVLQEMAMTNQENTQRQINELTDLKLQLDKKKLDFDTNVATAQG 180

Sbjct 121 SFVDNKEDREGFSDRLEVLQEMAMTNQENTQRQINELTDLKLQLDKKKLDFDTNVATAQG 180

Query 181 ILSTDGTGKIDQLKNEILNTKKAIQNDLQQQIALIPGALNEQGFAIFKEVYSLSKIEEPA 240

Sbjct 181 ILSTDGTGKIDQLKNEILNTKKAIQNDLQQQIALIPGALNEQGFAIFKEVYSLSKIEEPA 240

Query 241 TQAGVAAYNKGKEINNSILEAEKKAVQEATEQGKTALEIESAKKAAREAIEKSQGEIAA 300

Sbjct 241 TQAGVAAYNKGKEINNSILEAEKKAVQEATEQGKTALEIESAKKAAREAIEKSQGEIAA 300

Query 301 AAAAKTQEYDLMKVIDTEKIKKTFGVFAEVNLKTAEQRAYLDDLEKQNQKIYDLTTKLSI 360

Sbjct 301 AAAAKTQEYDLMKVIDTEKIKKTFGVFAEVNLKTAEQRAYLDDLEKQNQKIYDLTTKLSI 360

Query 361 ADLQKSMLLLTTQNDLHTFANQDVLEDLLKRYKEDLNLIKNSITKLSTNVDTTSEQSQKD 420

Sbjct 361 ADLQKSMLLLTTQNDLHTFANQDVLEDLLKRYKEDLNLIKNSITKLSTNVDTTSEQSQKD 420

Query 421 TLRQLKNVISYLEEQVYKF 439

Sbjct 421 TLRQLKNVISYLEEQVYKF 439

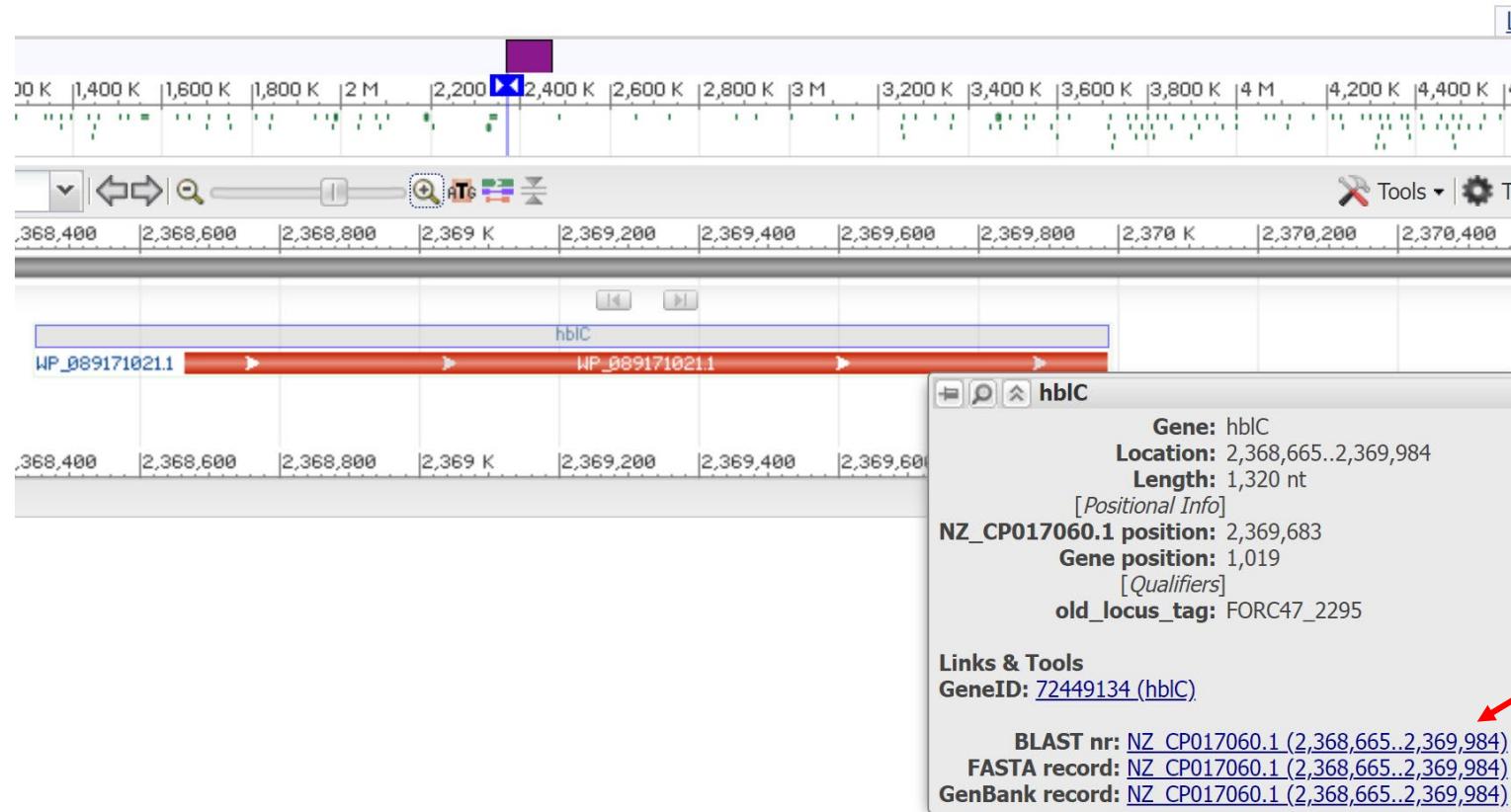
[Download](#) [▼](#)[FASTA \(complete sequence\)](#)[FASTA \(aligned sequences\)](#)[GenBank \(complete sequence\)](#)[Hit Table \(text\)](#)[Hit Table \(CSV\)](#)[Text](#)[XML](#)[ASN.1](#)**Related Information**[AlphaFold Structure - 3D](#)

structure displays

[Identical Proteins](#) - Identical proteins to WP_078401515.1

You can scroll down the page to see all the alignments.

You could click here to download the 100 sequences we got as hits.



Shortcut!

If you find a gene in a genome that you want to BLAST, you can use this link.

Questions?