

Université de JIJEL

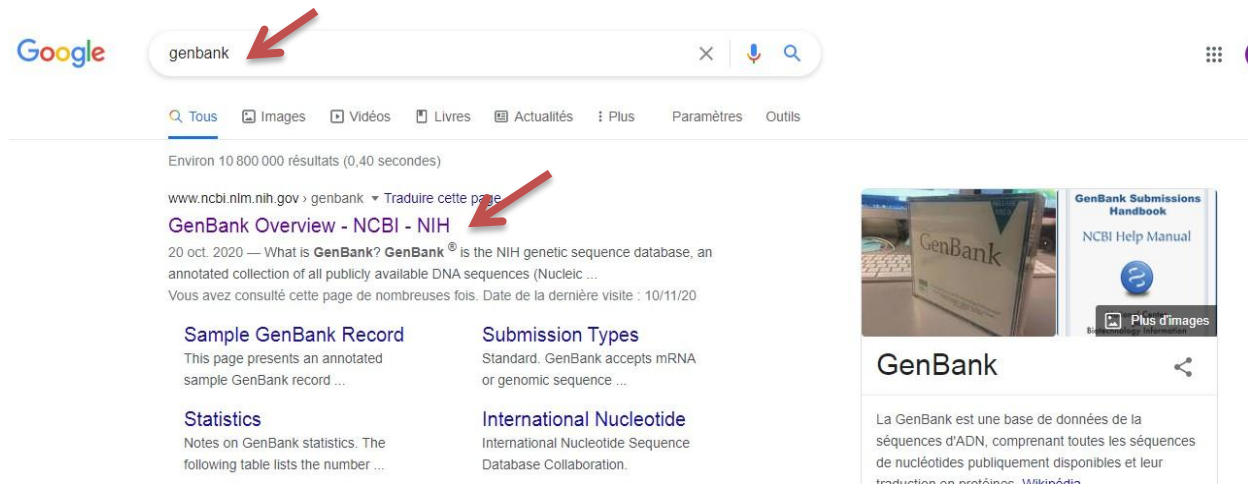
1^{ère} année Master Phytopharmacie

TP Bioinformatique

TP 2 : Recherche d'une séquence nucléique associée à une pathologie (Genbank)

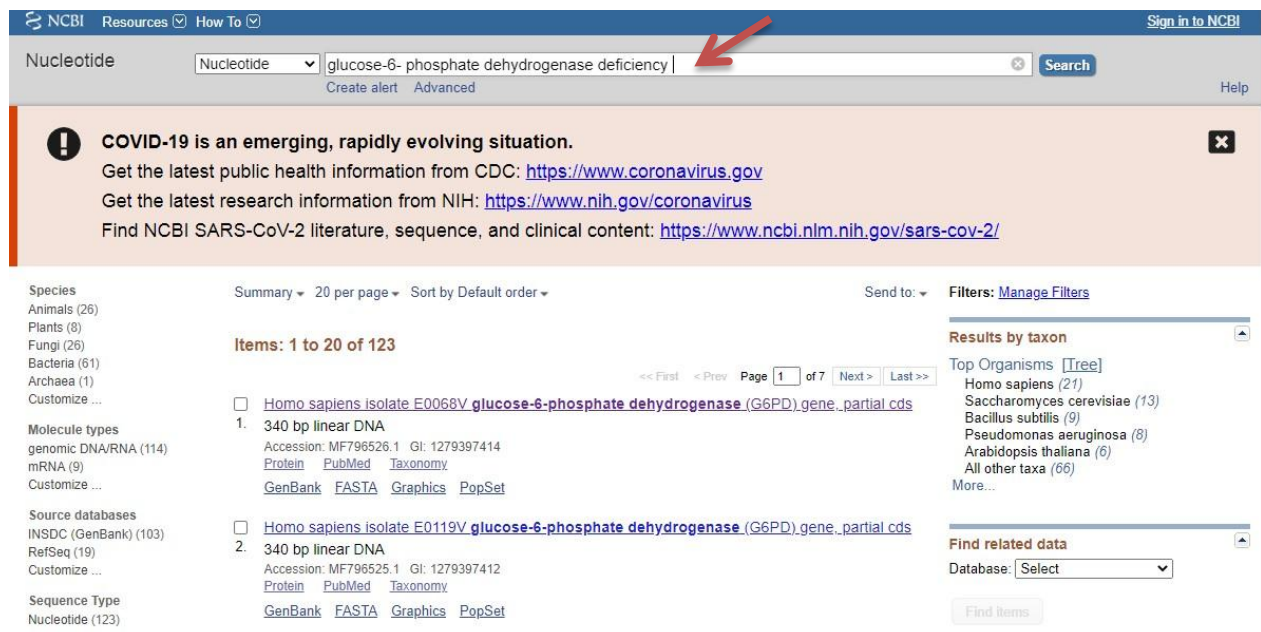
Les étapes :

Lancer une recherche de Genbank (base de données nucléotidiques) sur le moteur de recherche Google



Cliquer sur le premier résultat : Genbank Overview-NCBI- NIH

Entrer le mot clé convenable à votre séquence dans la zone de recherche, puis cliquer sur «Search».



Les résultats de l'information sont organisés sous forme de fichier, chaque fichier est appelé « une entrée »

Homo sapiens isolate E0068V glucose-6-phosphate dehydrogenase (G6PD) gene, partial cds

GenBank: MF796526.1

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

Go to: ☒

LOCUS MF796526 340 bp DNA linear PRI 26-NOV-2017
 DEFINITION Homo sapiens isolate E0068V glucose-6-phosphate dehydrogenase (G6PD) gene, partial cds.
 ACCESSION MF796526
 VERSION MF796526.1
 KEYWORDS .
 SOURCE Homo sapiens (human)
 ORGANISM [Homo sapiens](#)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 340)
 AUTHORS Carter,T.E., Mekonnen,S.K., Lopez,K., Bonnell,V., Damodaran,L., Aseffa,A. and Janies,D.A.
 TITLE Glucose-6-Phosphate Dehydrogenase Deficiency Genetic Variants in Malaria Patients in Southwestern Ethiopia
 JOURNAL Am. J. Trop. Med. Hyg. (2017) In press
 PUBMED [29141760](#)
 REMARK Publication Status: Available-Online prior to print
 REFERENCE 2 (bases 1 to 340)
 AUTHORS Carter,T.E., Mekonnen,S.K., Lopez,K., Bonnell,V., Damodaran,L., Aseffa,A. and Janies,D.A.
 TITLE Direct Submission
 JOURNAL Submitted (31-AUG-2017) Department of Bioinformatics and Genomics, University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA
 COMMENT ##Assembly-Data-START##
 Sequencing Technology :: Sanger dideoxy sequencing

Taille et type de molécule, division de l'organisme, date de dépôt de séquence

Numéro d'accès unique et numéro de version

Organisme +taxonomie

Références bibliographiques et nom des auteurs

ORIGIN


```
1 cacacatgga ctcaaagaga ggggctgaca tctgtctgtg tgtctgtctg tccgtgtctc
61 ccaggccacc ccagaggaga agctcaagct ggaggacttc ttgcccgcga actcctatgt
121 ggctggccag tacgatgatg cagcctccta ccagcgctc aacagccaca tgaatgccct
181 ccacctgggg tcacaggcca accgctctt ctacctggcc ttgccccga cgtctacga
241 ggccgtcacc aagaacattc acgagtcctg catgagccag atgtaaggct tgccgttgcc
301 ctcccttccc gcctgccagg ctggcccagg cagtgtctcc
```

//


Séquence nucléique

Base de données OMIM

OMIM est une base de données bibliographique qui donne des informations sur les maladies héréditaires chez l'être humain.

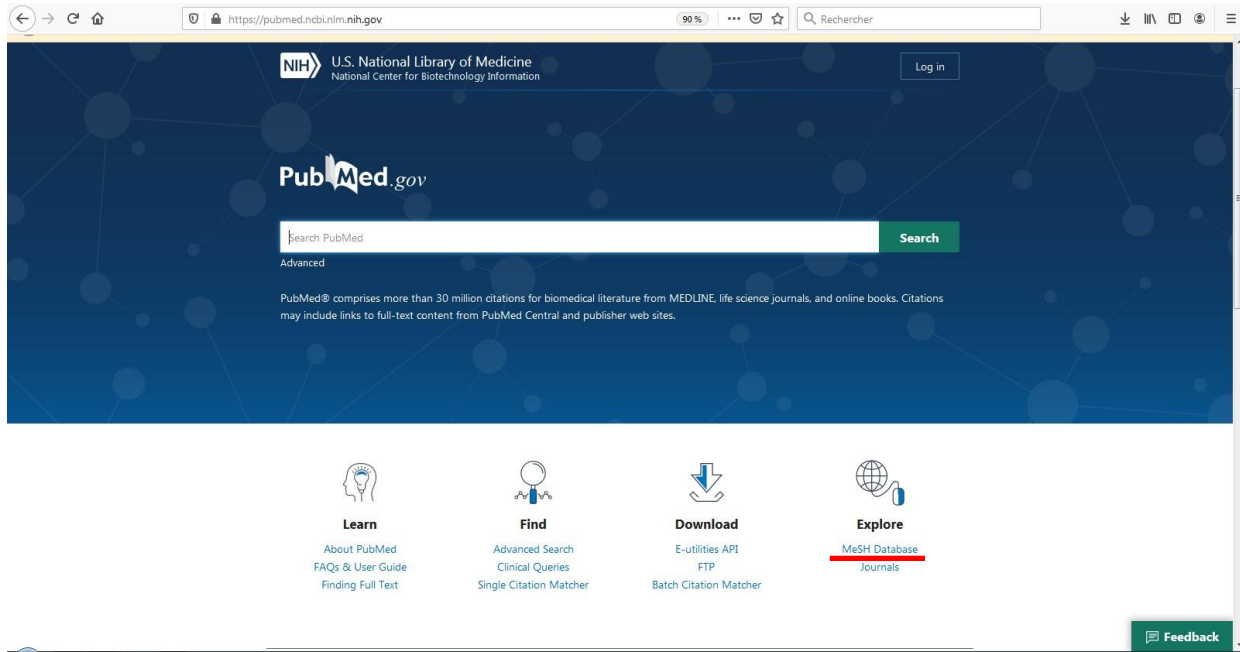


OMIM®
Online Mendelian Inheritance in Man®
An Online Catalog of Human Genes and Genetic Disorders
 Updated December 29, 2020

Advanced Search : OMIM, Clinical Synopses, Gene Map
 Need help? : Example Searches, OMIM Search Help,  OMIM Video Tutorials
 Mirror site : <https://mirror.omim.org>

OMIM is supported by a grant from NHGRI, licensing fees, and generous contributions from people like you.

TP3 : Bases de données bibliographiques (Pubmed) (TP réalisé)



TP 4 : Comparaison des séquences (Alignement)

1- Comparaison d'une séquence nucléique avec une base de données nucléique

- Le programme utilisé est Blast de Genbank sur le site du NCBI

- Il existe 5 types de BLAST (blastn, blastp, blastx, tblastn, tblastx)

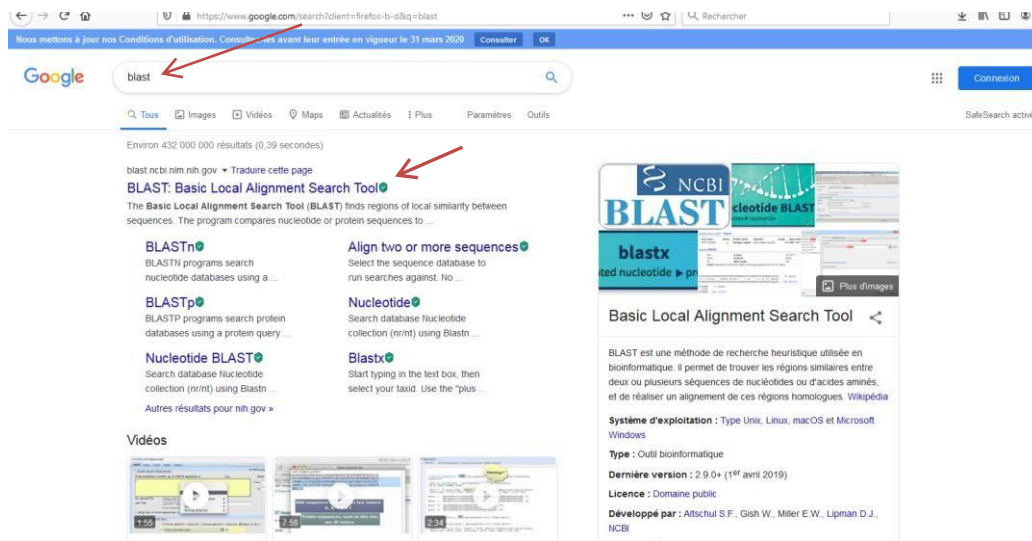
Pour comparer la séquence nucléotidique avec les séquences nucléotidiques de la base de données genbank on utilise le Blastn (Nucleotide BLAST).

Lancer une recherche BLAST pour la séquence suivante :

gggggggggtt	tccggattta	ttgggcgtaa	gcgagcgcag	gcggttttt	aagtctgatg	tgaaagccct
cggcttaacc	gaggaagcgc	atcggaact	gggaaacttg	agtgcagaag	aggacagtgg	aactccatgt
gtagcggtaga	aatgcgtaga	tatatggaag	aacaccagtg	gcgaaggcgc	ctgtctggtc	tgtaactgac
gctgaggctc	gaaagcatgg	gtagcgaaca	ggattagata	ccctggtagt	ccatgccgta	aacgatgaat
gctaggtgtt	ggagggttct	cgccttcag	tgccgcagct	aacgcattaa	gcattccgcc	tggggagtag
gaccgcaagg	ttgaaactca	aaggaattga	cggggggccc	cacaagcggg	ggagcatgtg	gtttaattcg
aagcaacgcg	aagaacctta	ccaggtcttg	acatcttttg	atcacctgag	agatcaggtt	tccccttcgg
gggcaaaatg	acaggtggtg	catggtgtc	gtcagctcgt	gtcgtgagat	ggtgggttaa	gtcccgaac
gagcgcgaacc	cttatgacta	gttgccagca	tttagttggg	cactctagta	agactgccgg	tgacaaaccg
gaggaagggtg	gggatgacgt	caaatcatca	tgccccttat	gacctgggct	acacacgtgc	tacaatggat
gggtacaacga	gttgcgagac	cgcgaggtca	agctaattct	ttaaagccat	tctcagttcg	gactgtaggc
tgcaactcgc	ctacacgaag	tcggaatcgc	tagtaatcgc	ggatcagcac	gccgcggtga	atagcttccc
gggccttgta	caccssca					

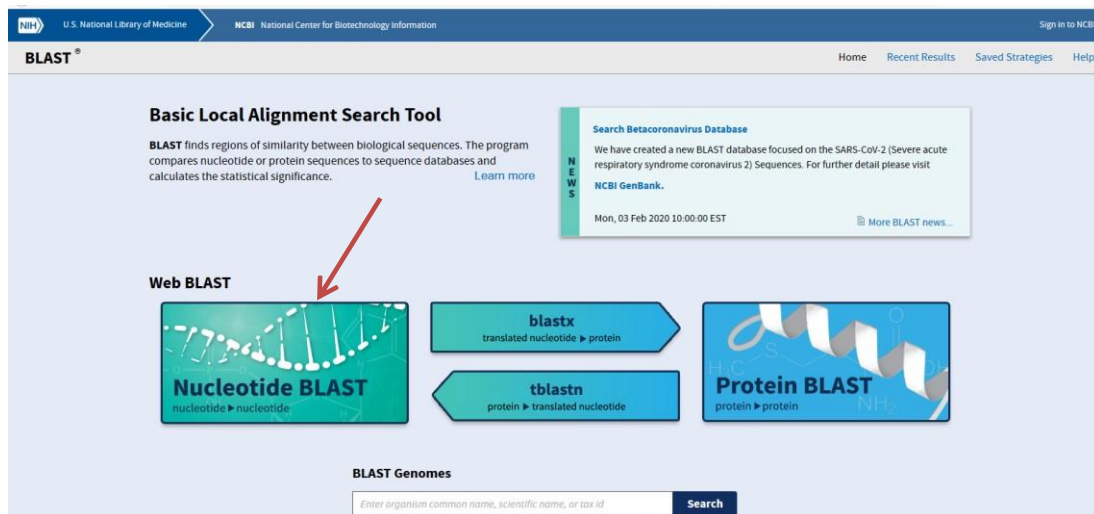
Les étapes :

Lancer une recherche de blast sur le moteur de recherche Google

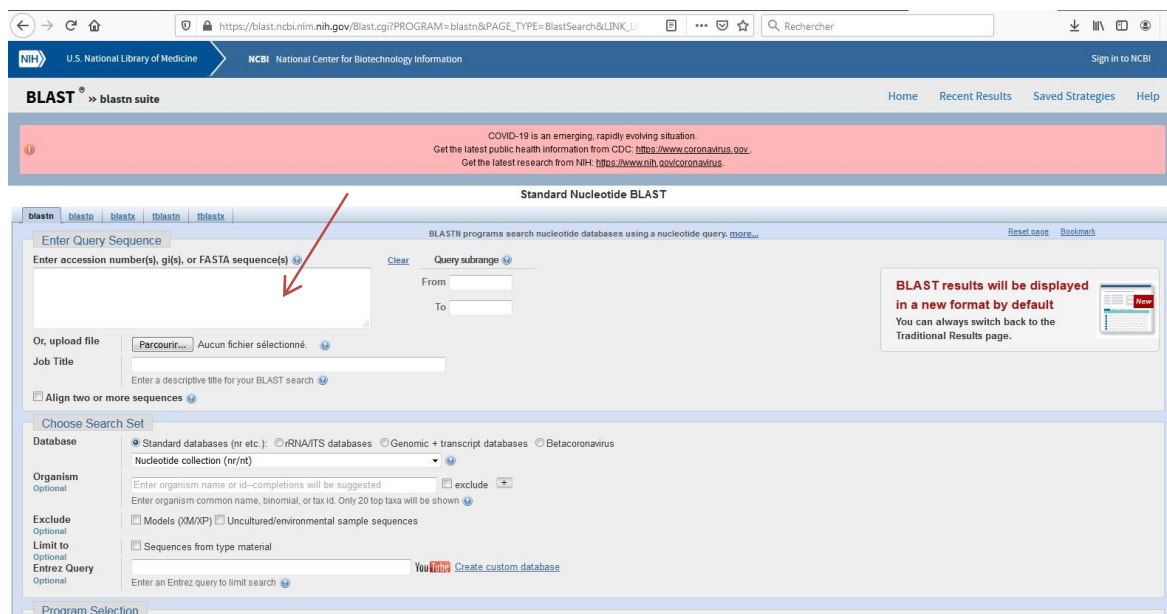


Cliquer sur le premier résultat : Blast : Basic local alignment Search Tool

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



Choisir le nucleotide BLAST



Entrer la séquence dans la fenêtre

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_...

Rechercher

U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® >> blastn suite Home Recent Results Saved Strategies Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

Standard Nucleotide BLAST

blastn | blasto | blastx | tblastn | tblastx

Enter Query Sequence

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

From

To

Or, upload file [Parcourir...](#) Aucun fichier sélectionné.

Job Title

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database ☒ Standard databases (nr etc.): ☐ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus

Nucleotide collection (nr/nt)

Organism

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

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude ☐ Models (MM/P) ☐ Uncultured/environmental sample sequences

Optional

Limit to ☐ Sequences from type material

Optional

Entrez Query [You Tube](#) [Create custom database](#)

Optional Enter an Entrez query to limit search

BLAST results will be displayed in a new format by default
You can always switch back to the Traditional Results page.

Cliquer sur Blast

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

Choose Search Set

Database
Standard databases (nr etc.): ☒ rRNA/ITS databases ☐ Genomic + transcript databases ☐ Betacoronavirus
Nucleotide collection (nr/nt)

Organism
Optional
Enter organism name or id-completions will be suggested
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown ☐ exclude

Exclude
Optional
☐ Models (MM/XP) ☐ Uncultured/environmental sample sequences

Limit to
Optional
☐ Sequences from type material

Entrez Query
Optional
Enter an Entrez query to limit search You [View](#) [Create custom database](#)

Program Selection

Optimize for
☒ Highly similar sequences (megablast)
☐ More dissimilar sequences (discontiguous megablast)
☐ Somewhat similar sequences (blastn)
Choose a BLAST algorithm

BLAST Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
☐ Show results in a new window

[Algorithm parameters](#)

BLAST is a registered trademark of the National Library of Medicine

[Support center](#) [Mailing list](#)

Le programme blast va comparer la séquence (la séquence requête ou query) avec les séquences de la base.

NIH U.S. National Library of Medicine
National Center for Biotechnology Information

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.cdc.gov/coronavirus>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>

BLAST® » blastn suite » results for RID-7PWP4M50016

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

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

Job Title Nucleotide Sequence
RID 7PWP4M50016 Search expires on 03-26 23:03 pm [Download All](#)
Program BLASTN [Citation](#)
Database nt [See details](#)
Query ID lcl|Query_62969
Description None
Molecule type dna
Query Length 859
Other reports [Distance tree of results](#) [MSA viewer](#)

Filter Results

Organism only top 20 will appear ☐ exclude
Type common name, binomial, taxid or group name
[Add organism](#)

Percent Identity to **E value** to **Query Coverage** to
[Filter](#) [Reset](#)

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

Sequences producing significant alignments [Download](#) [Manage Columns](#) [Show](#) 100 [?](#)

☒ select all 100 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per Ident	Accession
<input checked="" type="checkbox"/>	Lactobacillus casei strain B1 16S ribosomal RNA gene, partial sequence	1570	1570	99%	0.0	99.77%	KY764324.1
<input checked="" type="checkbox"/>	Lactobacillus casei strain B1M10 16S ribosomal RNA gene, partial sequence	1561	1561	99%	0.0	99.53%	KY764329.1
<input checked="" type="checkbox"/>	Lactobacillus casei strain 2SA3 16S ribosomal RNA gene, partial sequence	1557	1557	98%	0.0	99.88%	KC967210.1
<input checked="" type="checkbox"/>	Lactobacillus fermentum strain MBO003 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MT071603.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain ES91 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MT044263.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain NS16 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MT044251.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain NS12 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MT044244.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain ABK 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN994625.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain CACC 566 chromosome, complete genome	1554	7759	98%	0.0	99.76%	CP048003.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain TD 062 chromosome, complete genome	1554	7759	98%	0.0	99.76%	CP044361.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain 1490 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN750022.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain MG5191 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN749945.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain MG5190 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN749944.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain MG5189 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN749943.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain MG5179 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN749939.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain F3 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN744424.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain F2 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN744423.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain MG5310 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN720512.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain MG5245 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN704668.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain LMT13-95 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN689621.1
<input checked="" type="checkbox"/>	uncultured bacterium partial 16S rRNA gene	1554	1554	98%	0.0	99.76%	LR650498.1
<input checked="" type="checkbox"/>	Lactobacillus paracasei strain SHTR-1 16S ribosomal RNA gene, partial sequence	1554	1554	98%	0.0	99.76%	MN659715.1

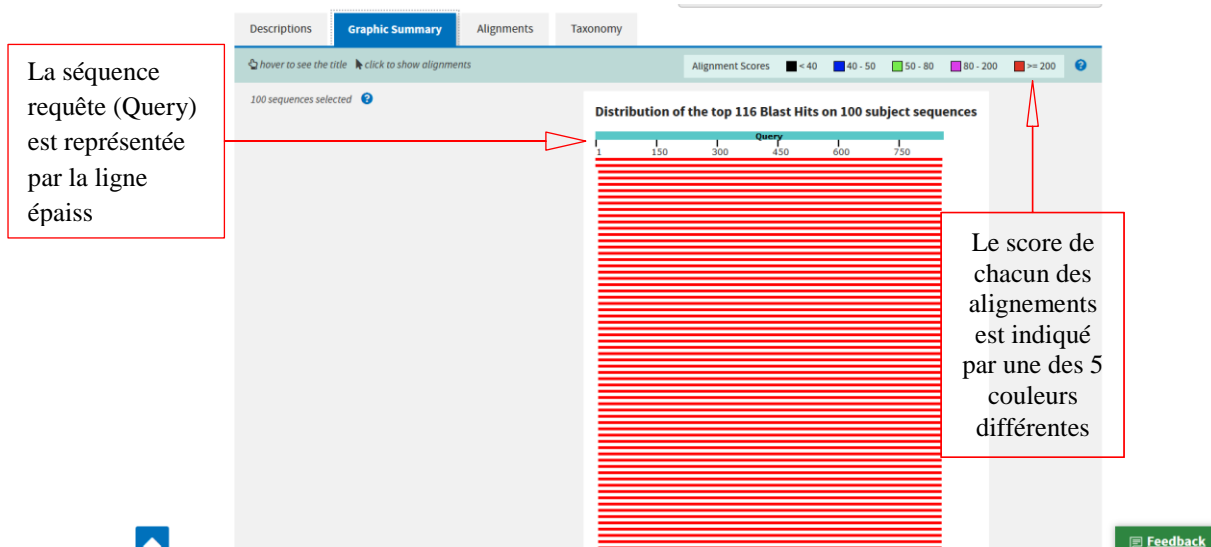
[Feedback](#)

Plus le score est grand, plus la qualité est bonne et plus le pourcentage d'identité est élevé

Les résultats de BLAST

La page résultat est divisée en 3 parties :

- La liste des séquences avec leur score et leur E-value
- Une vue graphique générale (Graphic summary) des séquences résultats avec différentes couleur



- Résultat détaillé, fournissant pour chaque séquence résultat (subject), l'alignement avec la séquence requête (query).

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

Descriptions Graphic Summary **Alignments** Taxonomy

Alignment view Pairwise CDS feature Download

100 sequences selected

Download GenBank Graphics

Lactobacillus casei strain B1 16S ribosomal RNA gene, partial sequence
Sequence ID: KY764324.1 Length: 860 Number of Matches: 1

Range 1: 1 to 856 GenBank Graphics

Score	Expect	Identities	Gaps	Strand
1570 bits(850)	0.0	854/856(99%)	1/856(0%)	Plus/Plus
Query 1	ggggggggTTCGGATTATTGGGCGTAAGCGAGCGCAGGCGGTTTTAAGTCTGATG	60		
Sbjct 1	ggggggggTTCGGATTATTGGGCGTAAGCGAGCGCAGGCGGTTTTAAGTCTGATG	60		
Query 61	TGAAAGGCGCTCGGCTTAACCGAGGAGCGCATCGGAACTGGGAACTGAGTGCAGAG	120		
Sbjct 61	TGAAAGGCGCTCGGCTTAACCGAGGAGCGCATCGGAACTGGGAACTGAGTGCAGAG	120		
Query 121	AGGACAGTGAACCTCCATGTGTAGCGGTGAATGCGTAGATATATGGAAGACACCAAGT	180		
Sbjct 121	AGGACAGTGAACCTCCATGTGTAGCGGTGAATGCGTAGATATATGGAAGACACCAAGT	180		
Query 181	GCGAAGGCGCGCTGTCTGGTCTGTAAGTACGCTGAGGCTCGAAGCATGGGTAGCGAACA	240		
Sbjct 181	GCGAAGGCGCGCTGTCTGGTCTGTAAGTACGCTGAGGCTCGAAGCATGGGTAGCGAACA	240		
Query 241	GGATTAGATACCGCTGGTAGTCCATGCCGTAAACGATGAATGCTAGGTGTGGAGGTTTC	300		
Sbjct 241	GGATTAGATACCGCTGGTAGTCCATGCCGTAAACGATGAATGCTAGGTGTGGAGGTTTC	300		
Query 301	CGCCCTTCAGTGCAGCGATTAACGATTTCGCGCTGGGGAGTACGACCGCAAGG	360		
Sbjct 301	CGCCCTTCAGTGCAGCGATTAACGATTTCGCGCTGGGGAGTACGACCGCAAGG	360		
Query 361	TTGAAACTCAAGGAATTGACGGGGGCGCCACAAAGCGGTGGAGCATGTGGTTAAATCG	420		
Sbjct 361	TTGAAACTCAAGGAATTGACGGGGGCGCCACAAAGCGGTGGAGCATGTGGTTAAATCG	420		

Feedback

TP5. Analyse structurale et modélisation moléculaire

La modélisation moléculaire permet la prédiction de la structure 3D d'une protéine à partir de sa structure primaire (l'enchaînement des acides aminés qui la composent), en prenant en compte les différentes propriétés physico-chimiques des acides aminés.

La Protein Data Bank (PDB) est la base de données dédiée au stockage de la structure des protéines obtenues expérimentalement (par RMN, cristallographie et diffraction de RX, et par microscopie électronique).

Base de données PDB : conçues pour la structure tridimensionnelle des protéines



Applications

- 1- Trouver dans la base Genbank la séquence dont le numéro d'accèsion est le suivant :
MW494397.
Quelles sont les informations accompagnant la séquence.
- 2- Trouver la structure tridimensionnelle de cette protéine dont le numéro identifiant est le suivant : 2GTL