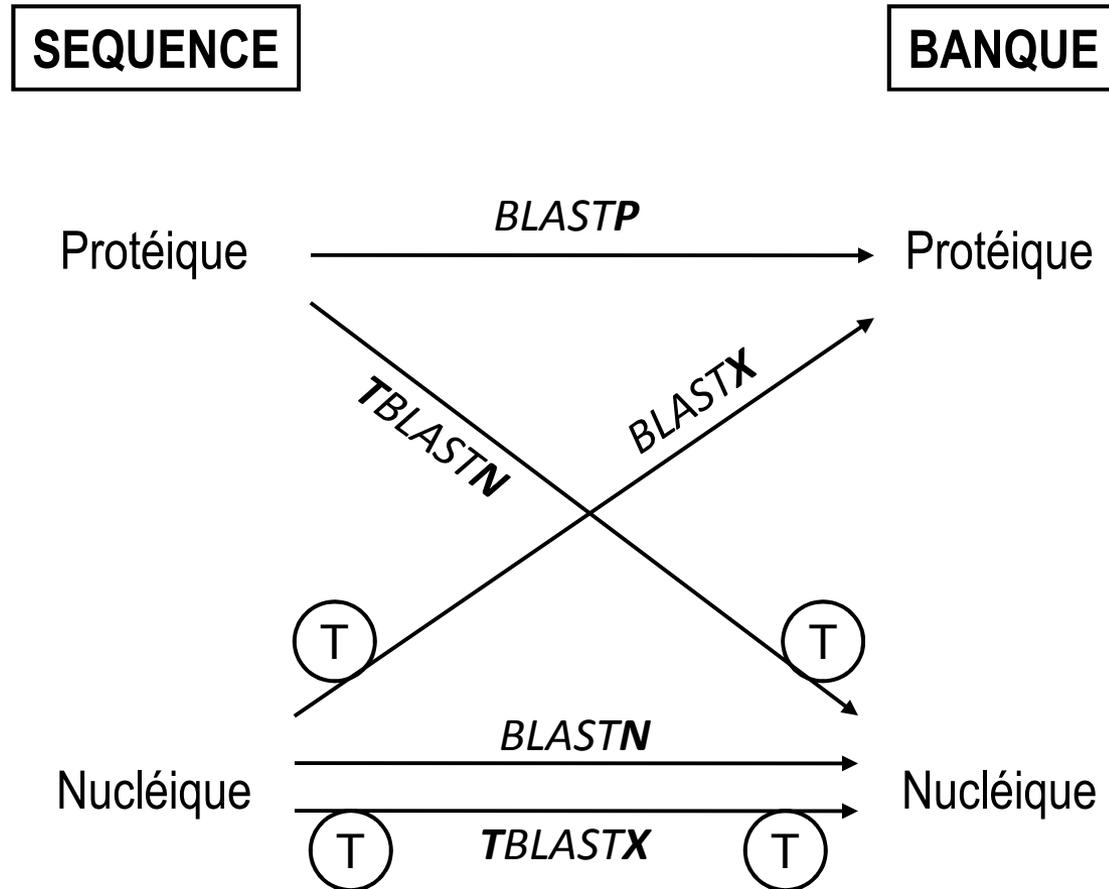


TP04 :BLAST

# BLAST: Choix du programme



- ❖ blastn, de nucléotides, séquence nucléotidique contre une base de données de séquences nucléotidiques.
- ❖ blastp, de protéines, séquence de protéine contre une base de données de séquences de protéines.
- ❖ blastx, séquence nucléotidique traduite en séquence de protéine contre une base de données de séquences de protéines.
- ❖ tblastn, séquence de protéine contre une base de données de séquences nucléotidiques traduites en séquences de protéines.
- ❖ tblastx, séquence nucléotidique traduite en séquence de protéine contre une base de données de séquences nucléotidiques traduites en séquences de protéines.

## 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](#)

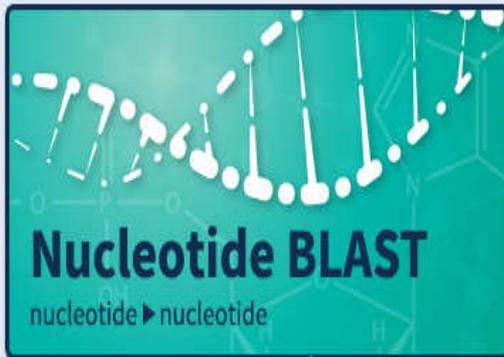
**NEWS**

[Search Betacoronavirus Database](#)

We have created a new BLAST database focused on the SARS-CoV-2 (Severe acute respiratory syndrome coronavirus 2) Sequences. For further detail please visit [NCBI GenBank](#).

Mon, 03 Feb 2020 10:00:00 EST [More BLAST news...](#)

## Web BLAST



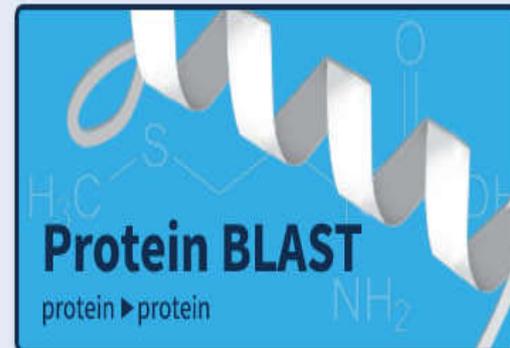
**Nucleotide BLAST**  
nucleotide ▶ nucleotide



**blastx**  
translated nucleotide ▶ protein



**tblastn**  
protein ▶ translated nucleotide



**Protein BLAST**  
protein ▶ protein



**Cliquer ici pour choisir le Blast nucléique**

**Cliquer ici pour choisir le Blast protéique**

# On entre la séquence à chercher

## Standard Nucleotide BLAST

[blastn](#) [blastp](#) [blastx](#) [tblastn](#) [tblastx](#)

### Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

[Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) ⓘ

[Clear](#) Query subrange ⓘ

From

To

```
GGCCATCCACAATGTTGTTTCATGCTATTATTCTGCATCAACAACAAAAACCACAACAACCATCGAGC
CAGGTCCTCTTCCAACAGCCTCTGCAACAATATCCATTAGGCCAGGGCTCCTCCGGCCATCTCAGCAA
ACCCACAGGCCCGGGCTCTGTCCAGCCTCAACAACCTGCCAGTTCGAGGAAATAAGGAACCTAGCGCT
ACAGACGCTACCCGCAATGTGCAATGTCTACATCCCTCCATATTGCACCATCGGCCATTGGCATCTTC
GGTACTAACTG
```

**BLAST results will be displayed in a new format by default**

You can always switch back to the Traditional Results page.



Or, upload file  Aucun fichier choisi ⓘ

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)** ⓘ

Limit by  Organism  BioProjectID  WGS Project

# Cliquer sur "Back to traditional page"

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>.

BLAST<sup>®</sup> » blastn suite » results for RID-8P3NW11G01R

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

[← Edit Search](#)

[Save Search](#)

[Search Summary](#) ▼

[? How to read this report?](#)

[▶ BLAST Help Video](#)

[↶ Back to Traditional Results Page](#)

Job Title	KC660359.1 Triticum aestivum clone pGli70...
RID	<a href="#">8P3NW11G01R</a> Search expires on 04-07 19:13 pm <a href="#">Download All</a> ▼
Program	BLASTN <a href="#">?</a> <a href="#">Citation</a> ▼
Database	nt <a href="#">See details</a> ▼
Query ID	lcl Query_64535
Description	KC660359.1 Triticum aestivum clone pGli70 gliadin (gli) gene ...
Molecule type	dna
Query Length	852

## 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

Query: KC660359.1 Triticum aestivum clone pGli70...

RID [8P3NW11G01R](#) (Expires on 04-07 19:13 pm)

Query ID [Id|Query\\_64535](#)

Description [KC660359.1 Triticum aestivum clone pGli70 gliadin \(gli\) gene, complete cds](#)

Sequence type [dna](#)

Sequence Length [852](#)

Database Name [nt](#)

Description [Nucleotide collection \(nt\)](#)

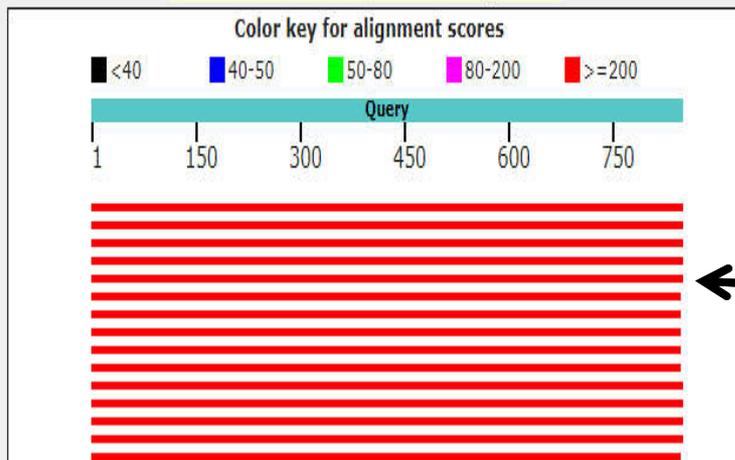
Program [BLASTN 2.10.0+](#) [Citation](#)

Reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

[Search Summary](#)

### Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



Nombres de hits  
(nombre de  
comparaison est =  
100)

Répartition des hits  
(comparaisons) en  
fonction du score

Job title: KC660359.1 Triticum aestivum clone pGli70...

RID [BP3NW11G01R](#) (Expires on 04-07 19:13 pm)

Query ID [ld|Query\\_64535](#)

Description KC660359.1 Triticum aestivum clone pGli70 gliadin (gli) gene, complete cds

Molecule type dna

Query Length 852

Database Name nt

Description Nucleotide collection (nt)

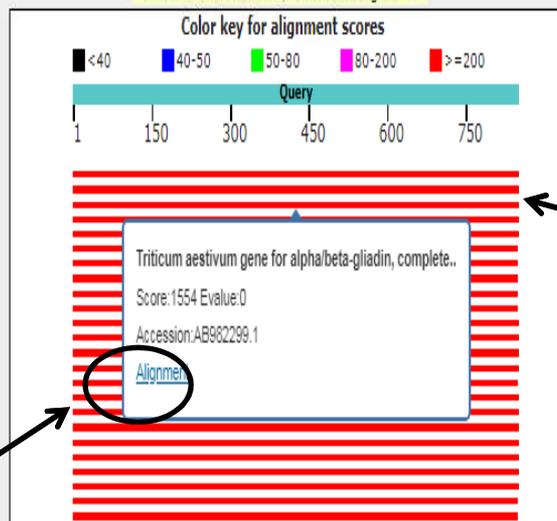
Program BLASTN 2.10.0+ [Citation](#)

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [MSA viewer](#)

## Graphic Summary

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments



2-Cliquez  
sur  
Alignment

1-Cliquez 2 fois sur  
la 3<sup>ème</sup> ligne rouge  
(chaque ligne  
correspond à une  
comparaison entre  
ma séquence et  
l'autre de la banque

E-value

% d'identité entre les 2 séquences  
(Requête et sujet)

Triticum aestivum gene for alpha/beta-gliadin, complete cds, cultivar: Chinese Spring, clone: TAC2  
Sequence ID: [AB982299.1](#) Length: 2161 Number of Matches: 1

Range 1: 596 to 1450 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1554 bits(841)	0.0	851/855(99%)	3/855(0%)	Plus/Plus

Query	1	ATGAAGACCTTTCTCATCCTTGCCTCCTTGTCTATCGTGCGACCACCGCCACAAC	60
Sbjct	596	ATGAAGACCTTTCTCATCCTTGCCTCCTTGTCTATCGTGCGACCACCGCCACAAC	655
Query	61	GTTAGAGTTCAGTGCCACAATTGCAGCCACAACATCCATCTCAGCAACAGCCACAAGAG	120
Sbjct	656	GTTAGAGTTCAGTGCCACAATTGCAGCCACAACATCCATCTCAGCAACAGCCACAAGAG	715
Query	121	CAAGTTCATTGGTACAACAACAACAATTTCTAGGGCAGCAACAACCATTTCCACCACAA	180
Sbjct	716	CAAGTTCATTGGTACAACAACAACAATTTCTAGGGCAGCAACAACCATTTCCACCACAA	775
Query	181	CAACCATATCCACAGCCGCAACCATTTCCATCACAACAACCATATCTGCAACTACAACCA	240
Sbjct	776	CAACCATATCCACAGCCGCAACCATTTCCATCACAACAACCATATCTGCAACTACAACCA	835
Query	241	TTTCGCGAGCCGCAACTACCATATTCGAGCCACAACCATTTGACCCACAACAACCATAT	300
Sbjct	836	TTTCGCGAGCCGCAACTACCATATTCGAGCCACAACCATTTGACCCACAACAACCATAT	895
Query	301	CCACAACCGCAACCACAGTATTCGcaaccacaacaaccaatcagcagcagcagcagcag	360
Sbjct	896	CCACAACCGCAACCACAGTATTCGCAACCACAACAACCAATTTACAGCAGCAGCAGCAG	955
Query	361	cag---caacaacaacaacaacaacaacaacaacaatccttcaacaatccttgaacaaca	417
Sbjct	956	CAGCAACAACAACAACAACAACAACAACAACAATCCTTCAACAATTTGCAACAACAA	1015
Query	418	cTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA	477
Sbjct	1016	CTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA	1075

Gaps: nombre de  
délétion dans la  
séquence requête

## Suite d'alignement (comparaison)

```
Query 418 cTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA 477
          |||
Sbjct 1016 CTGATTCCATGCATGGATGTTGTATTGCAGCAACACAACATAGCGCATGGAAGATCACAA 1075

Query 478 GTTTTGCAACAAAGTACTTACCAGCTGTTACAAGAATTGTGTTGTCAGCACCTATGGCAG 537
          |||
Sbjct 1076 GTTTTGCAACAAAGTACTTACCAGCTGTTACAAGAATTGTGTTGTCAGCACCTATGGCAG 1135

Query 538 ATCCCTGAGCAGTCGCAGTGCCAGGCCATCCACAATGTTGTTTCATGCTATTATTCTGCAT 597
          |||
Sbjct 1136 ATCCCTGAGCAGTCGCAGTGCCAGGCCATCCACAATGTTGTTTCATGCTATTATTCTGCAT 1195

Query 598 CAACAACAAAAACCACAACAACAACCATCGAGCCAGGTCTCCTTCCAACAGCCTCTGCAA 657
          |||
Sbjct 1196 CAACAACAAAAACCACAACAACAACCATCGAGCCAGGTCTCCTTCCAACAGCCTCTGCAA 1255

Query 658 CAATATCCATTAGGCCAGGGCTCCTTCCGGCCATCTCAGCAAAACCCACAGGCCCGGGGC 717
          |||
Sbjct 1256 CAATATCCATTAGGCCAGGGCTCCTTCCGGCCATCTCAGCAAAACCCACAGGCCCGGGGC 1315

Query 718 TCTGTCCAGCCTCAACAACCTGCCCCAGTTCGAGGAAATAAGGAACCTAGCGCTACAGACG 777
          |||
Sbjct 1316 TCTGTCCAGCCTCAACAACCTGCCCCAGTTCGAGGAAATAAGGAACCTAGCGCTACAGACG 1375

Query 778 CTACCCGCAATGTGCAATGTCTACATCCCTCCATATTGCACCATCGCGCCATTTGGCATC 837
          |||
Sbjct 1376 CTACCCGCAATGTGCAATGTCTACATCCCTCCATATTGCACCATCGCGCCATTTGGCATC 1435

Query 838 TTCGGTACTAACTGA 852
          |||
Sbjct 1436 TTCGGTACTAACTGA 1450
```

La taille de notre séquence (Query: 852) et la taille de la séquence de la banque (Sbjet: 1450).

-Pour le Blast protein, on va suivre les même étapes précédentes que pour le Blast Nucleotide.

-Le résultat apparait est presque identique à celui du Blast nucleotide mais à la place des nucleotides on utilise des acides aminés (AKWYTR.....) car les séquences à comparer sont des séquences protéiques.

alpha-gliadin protein [Triticum monococcum]

Sequence ID: [AGI15866.1](#) Length: 285 Number of Matches: 1

Range 1: 1 to 285 [GenPept](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
499 bits(1285)	3e-177	Compositional matrix adjust.	281/285(99%)	282/285(98%)	2/285(0%)

Query	1	MKTFLILALLAIVATTATTAVRVPVPLQVPLVQQQFLGQQPFPPQ	60
Sbjct	1	MKTFLILALLAIVATTATTAVRVPVPLQVPLVQQQFLGQQPFPPQ	60
Query	61	QYPYQPQPFPSQQPYLQLQPFQPQLPYSQPQFRPQQYPYQPQPYSPQQPISQRQQQ	120
Sbjct	61	QYPYQPQPFPSQQPYLQLQPFQPQLPYSQPQFRPQQYPYQPQPYSPQQPISQ+QQQ	120
Query	121	QQQQQQQQQQ - - ILQQILQQQLIPCMDVVLQQHNI AHGRSQV LQQSTYQLLQELCCQHLW	178
Sbjct	121	QQQQQQQQQQ ILQQILQQQLIPCMDVVLQQHNI AHGRSQV LQQSTYQLLQELCCQHLW	180
Query	179	QIPEQSQCQAIHNVVHAIILHQQQKPQQPSSQVSFQQPLQQYPLGQGSFRPSQQNPQAR	238
Sbjct	181	QIPEQSQCQAIHNVVHAIIPHQQQKPQQPSSQVSFQQPLQQYPLGQGSFRPSQQNPQAR	240
Query	239	GSVQPQQLPQFEEIRNLALQTL PAMCNVYIPPYCTIAPFGIFGTN	283
Sbjct	241	GSVQPQQLPQFEEIRNLALQTL PAMCNVYIPPYCTIAPFGIFGTN	285

← C'est une confirmation des acides aminés et non une 3<sup>ème</sup> séquence

# Interprétation des résultats

## Les résultats d'un Blast sont en 4 parties :

1. Un schéma graphique : les meilleurs résultats sont en rouge, suivis par les verts, les plus mauvais en bleu et noir)

2. La liste des meilleurs hits (liste de numéro d'accession cliquable), avec le E-value et le Bit score

- **Bit score** : mesure statistique de la validité de l'alignement : plus la valeur est élevée, plus les 2 séquences sont similaires. En-dessous de 50, le résultat n'est pas fiable.

- **E-value** : ?Expectation value? : estime la chance que vous auriez de trouver le même résultat par hasard. Plus la valeur est faible (proche de 0), meilleure est votre résultat. Au-dessus de 0.001, vos résultats ne sont pas bons.

3. **Alignements** : comparaison de votre séquence appelée (requête en français, query en anglais) avec la séquence la plus proche que Blast a trouvé appelée (Sujet, subject).