

USING NCBI to analyze DNA and protein Sequences for BoCAL and BobCAL Genes

1. Find the sequences you will be working with

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

The screenshot shows the NCBI homepage. At the top, there is a navigation bar with the NIH logo and the text "National Library of Medicine National Center for Biotechnology Information". A "Log in" button is located in the top right corner. Below the navigation bar is a search bar with a dropdown menu set to "All Databases" and a "Search" button. The main content area is divided into three columns. The left column contains a navigation menu with items such as "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", and "Variation". The middle column features a "Welcome to NCBI" message and six service tiles: "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), and "Research" (Explore NCBI research and collaborative projects). The right column contains "Popular Resources" (PubMed, Bookshelf, PubMed Central, BLAST, Nucleotide, Genome, SNP, Gene, Protein, PubChem) and "NCBI News & Blog" with two recent articles: "Introducing the New NCBI Datasets Genome Annotation Table" (10 Oct 2023) and "Comparing Yeast Species Used in Beer Brewing and Bread Making" (29 Sep 2023).

Select nucleotide from the dropdown menu and enter "BoCAL" or "BobCAL" in the dropdown menu

This screenshot shows the NCBI homepage with the search bar updated. The dropdown menu is now set to "Nucleotide" and the search input field contains the text "BoCAL". The rest of the page layout, including the navigation bar, main content area, and sidebars, remains the same as in the previous screenshot.

Results for BoCAL, make note of accession L36926

An official website of the United States government [Here's how you know](#) ▼

NIH National Library of Medicine
National Center for Biotechnology Information Log in

Nucleotide Help

Create alert Advanced

Species
Animals (2)
Plants (2)
Customize ...

Molecule types
mRNA (4)
Customize ...

Source databases
INSDC (GenBank) (2)
RefSeq (2)
Customize ...

Sequence Type
Nucleotide (4)

Sequence length
Custom range...

Release date
Custom range...

Summary ▼ Sort by Default order ▼ Send to: ▼ **Filters:** [Manage Filters](#)

See [LOC106320120 \(BOCAL\)](#) transcription factor CAULIFLOWER in the Gene database
bobcal reference sequences [Transcript \(2\)](#) [Protein \(2\)](#)

Items: 4

[Brassica oleracea \(clone pBS85\) BoCAL mRNA, complete cds](#)

1. **756 bp linear mRNA**
Accession: L36926.1 GI: 642590
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[Brassica oleracea var. capitata DNA binding protein \(CAL\) mRNA, partial cds](#)

2. **286 bp linear mRNA**
Accession: AY514056.1 GI: 45533881
[Protein](#) [Taxonomy](#)

Results by taxon

Top Organisms [\[Tree\]](#)
Danio rerio (2)
Brassica oleracea (2)

Analyze these sequences

Run BLAST

Find related data

Database:

Results for BobCAL, make note of accession L36927

An official website of the United States government [Here's how you know](#) ▼

NIH National Library of Medicine
National Center for Biotechnology Information Log in

Nucleotide Help

Create alert Advanced

Species
Plants (6)
Customize ...

Molecule types
genomic DNA/RNA (1)
mRNA (5)
Customize ...

Source databases
INSDC (GenBank) (6)
Customize ...

Sequence Type
Nucleotide (6)

Sequence length
Custom range...

Release date
Custom range...

Revision date
Custom range...

Summary ▼ 20 per page ▼ Sort by Default order ▼ Send to: ▼ **Filters:** [Manage Filters](#)

See [LOC106320120 \(BOBCAL\)](#) transcription factor CAULIFLOWER in the Gene database
bobcal reference sequences [Transcript \(2\)](#) [Protein \(2\)](#)

Items: 6

[Brassica oleracea \(subspecies botrytis\) BobCAL mRNA, complete cds](#)

1. **756 bp linear mRNA**
Accession: L36927.1 GI: 642592
[Protein](#) [PubMed](#) [Taxonomy](#)
[GenBank](#) [FASTA](#) [Graphics](#)

[Brassica oleracea var. botrytis DNA binding protein variant d \(CAL\) mRNA, complete cds, alternatively spliced](#)

2. **892 bp linear mRNA**
Accession: AY514055.1 GI: 45533879
[Protein](#) [Taxonomy](#)

Results by taxon

Top Organisms [\[Tree\]](#)
Brassica oleracea var. botrytis (5)
Punica granatum (1)

Analyze these sequences

Run BLAST

Find related data

Database:

Search details

You can click on the title (e.g. Brassica oleracea (subspecies botrytis) **BobCAL** mRNA, complete cds) in the results to see details including nucleotide and protein sequences, make note of the protein IDs for BoCAL and BobCAL (AAA64790 and AAA64791)

2. Compare two nucleotide sequences

Select BLAST under Popular Resources or go to <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

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

Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI News & Blog
Introducing the New NCBI Datasets Genome Annotation Table 10 Oct 2023
As part of our ongoing effort to modernize and improve your experience, we are...
Comparing Yeast Species Used in Beer Brewing and Bread Making 29 Sep 2023
Using the NIH Comparative Genomics

Click on nucleotide blast

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+ 2.12.0 is here!
We have made some improvements to how BLAST multi-threads and the amount of memory required by makeblastdb.
Tue, 13 Jul 2021 12:00:00 EST [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

Protein BLAST
protein ▶ protein

check box “align two or more sequences” and enter L36926 in the first box and L36927 in the second box then click on BLAST

Align Sequences Nucleotide BLAST

BLASTN programs search nucleotide subjects using a nucleotide query.

Enter Query Sequence

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

L36926

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 [?](#)

Enter Subject Sequence

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

L36927

Subject subrange [?](#)

From

To

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

Program Selection

Optimize for

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

Show results in a new window

Click on the Alignments tab in the results

BLAST® » blastn suite-2sequences » results for RID-N985A40211R [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

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

Job Title L36926:Brassica oleracea (clone pBS85) BoCAL...

RID [N985A40211R](#) Search expires on 10-01 01:21 am [Download All](#)

Program Blast 2 sequences [Citation](#)

Query ID [L36926.1](#) (nucleic acid)

Query Descr Brassica oleracea (clone pBS85) BoCAL mRNA, complete cds

Query Length 756

Subject ID [L36927.1](#) (nucleic acid)

Subject Descr Brassica oleracea (subspecies botrytis) BobCAL mRNA, cc ...

Subject Length 756

Other reports [MSA viewer](#)

Filter Results

Percent Identity to E value to Query Coverage to

Alignments | Descriptions | Graphic Summary | Dot Plot

Sequences producing significant alignments Download [New](#) Select columns Show [?](#)

select all 1 sequences selected [GenBank](#) [Graphics](#) [New MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Brassica oleracea (subspecies botrytis) BobCAL mRNA, complete cds	Brassica oleracea var. botrytis	1315	1315	100%	0.0	98.54%	756	L36927.1

The alignment of the two sequences is shown, the additional stop codon in the BobCAL (subject sequence) has been highlighted in the results below

Brassica oleracea (subspecies botrytis) BobCAL mRNA, complete cds

Sequence ID: [L36927.1](#) Length: 756 Number of Matches: 1

Range 1: 1 to 756 [GenBank](#) [Graphics](#)

▼ [Next Match](#) ▲ [F](#)

Score	Expect	Identities	Gaps	Strand
1315 bits(1457)	0.0	745/756(99%)	0/756(0%)	Plus/Plus
Query 1	ATGGGAAGGGGTAGGGTTGAAATGAAGAGGATAGAGAACAAGATCAACCGACAAGTGACG			60
Sbjct 1	ATGGGAAGGGGTAGGGTTGAAATGAAGAGGATAGAGAACAAGATCAACAGACAAGTGACG			60
Query 61	TTTTCGAAAAGAAGAGCTGGTCTTTTGAAGAAAGCCCATGAGATCTCGATCCTTTGTGAT			120
Sbjct 61	TTTTCGAAAAGAAGAGCTGGTCTTTTGAAGAAAGCCCATGAGATCTCGATTCTTTGTGAT			120
Query 121	GCTGAGGTTTCCCTTATTGTCTTCTCCATAAGGGGAAACTGTTGAGTACTCGTCTGAA			180
Sbjct 121	GCTGAGGTTTCCCTTATTGTCTTCTCCATAAGGGGAAACTGTTGAGTACTCGTCTGAA			180
Query 181	TCTTGCATGGAGAAGGTAAGTACTAGAACACTACGAGAGGTAAGTACTTACGCCGAGAAACAGCTA			240
Sbjct 181	TCTTGCATGGAGAAGGTAAGTACTAGAACCTACGAGAGGTAAGTACTTACGCCGAGAAACAGCTA			240
Query 241	AAAGTCCAGACTCTCACGTCATGCACAAACGAAGTGGTCAAGTGAATATAGCAGGCTT			300
Sbjct 241	AAAGTCCAGACTCTCACGTCATGCACAAACGAAGTGGTCAAGTGAATATAGCAGGCTT			300
Query 301	AAGGCTAAGATTGAGCTTTTGGAGAGAAACCAAAGGCATTATCTGGGCGAAGATTTAGAA			360
Sbjct 301	AAGGCTAAGATTGAGCTTTGGGAGAGGAACCAAAGGCATTATCTGGGAGAAAGATTTAGAA			360
Query 361	TCAATCAGCATAAAGGAGCTACAGAATCTGGAGCAGCAGCTTGACACTTCTTTAAACAT			420
Sbjct 361	TCAATCAGCATAAAGGAGCTACAGAATCTGGAGCAGCAGCTTGACACTTCTTTAAACAT			420
Query 421	ATTGCTCGAGAAAAAATCAACTAATGCACGAGTCCCTCAACCACCTCCAAGAAAGGAG			480
Sbjct 421	ATTGCTCGAGAAAAAATCAACTAATGCACAGTAGTCCCTCAACCACCTCCAAGAAAGGAG			480
Query 481	AAAGAAATACTGGAGGAAAACAGCATGCTTGCCAAACAGATAAGGGAGAGGGAGAGTATC			540
Sbjct 481	AAAGAAATACTGGAGGAAAACAGCATGCTTGCCAAACAGATAAAGGAGAGGGAGAGTATC			540
Query 541	CTAAGGACACATCAAACCAATCAGAGCAGCAAAACCGCAGCCACCATGTAGCTCCTCAG			600
Sbjct 541	CTAAGGACACATCAAACCAATCAGAGCAGCAAAACCGCAGCCACCATGTAGCTCCTCAG			600
Query 601	CCGCAACCGCAGTTAAATCCTTACATGGCATCATCTCCTTTCTAAATATGGGTGGCATG			660
Sbjct 601	CCGCAACCGCAGTTAAATCCTTACATGGCATCATCTCCTTTCTAAATATGGGTGGCATG			660
Query 661	TACCAAGGAGAATATCCAACGGCGGTGAGGAGGAACCGTCTCGATCTGACTCTTGAACCC			720
Sbjct 661	TACCAAGGAGAATATCCAACGGCGGTGAGGAGGAACCGTCTCGATCTGACTCTTGAACCC			720
Query 721	ATTTACAACCTGCAACCTTGGTTACTTTGCCGCATGA 756			
Sbjct 721	ATTTACAACCTGCAACCTTGGTTACTTTGCCGCATGA 756			

3. Compare two protein Sequences: <https://blast.ncbi.nlm.nih.gov/Blast.cgi>

Click on Protein BLAST

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+ 2.12.0 is here!
We have made some improvements to how BLAST multi-threads and the amount of memory required by makeblastdb.
Tue, 13 Jul 2021 12:00:00 EST [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

Protein BLAST
protein ▶ protein

check box “align two or more sequences”, enter AAA64790 in the first box and AAA64791 in the second box then click on BLAST (same as for nucleotide BLAST)

blastn **blastp** blastx tblastn tblastx Align Sequences Protein BLAST

BLASTP programs search protein subjects using a protein query. [more...](#)

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 [?](#)

Enter Subject Sequence

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

From

To

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

Program Selection

Algorithm blastp (protein-protein BLAST)
[Choose a BLAST algorithm \[?\]\(#\)](#)

In the results click on graphic summary, you can see that the subject protein (BobCAL AAA64791 in red) is much shorter than the Query protein (BoCAL AAA64790 in blue) because the translation stopped at the stop codon identified in the sequence above:

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

Job Title	AAA64790:amino acid feature: K-box, bp 283..480;...
RID	N93N5W8Z11R Search expires on 10-01 00:04 am Download All ▼
Program	Blast 2 sequences Citation ▼
Query ID	AAA64790.1 (amino acid)
Query Descr	amino acid feature: K-box, bp 283..480; amino acid feature ...
Query Length	251
Subject ID	AAA64791.1 (amino acid)
Subject Descr	amino acid feature: MADS box; codes for a putative DNA-b ...
Subject Length	150
Other reports	Multiple alignment MSA viewer ?

Filter Results

Percent Identity: to E value: to Query Coverage: to

[Filter](#) [Reset](#)

[Descriptions](#) [Graphic Summary](#) [Alignments](#) [Dot Plot](#)

hover to see the title click to show alignments Alignment Scores: ■ < 40 ■ 40 - 50 ■ 50 - 80 ■ 80 - 200 ■ >= 200 ?

1 sequences selected ?

Distribution of the top 1 Blast Hits on 1 subject sequences

4. Align multiple Brassica sequences and make a phylogenetic tree:

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

Click on nucleotide blast

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+ 2.12.0 is here!
We have made some improvements to how BLAST multi-threads and the amount of memory required by makeblastdb.
Tue, 13 Jul 2021 12:00:00 EST [More BLAST news...](#)

Web BLAST

Nucleotide BLAST
nucleotide ▶ nucleotide

blastx
translated nucleotide ▶ protein

tblastn
protein ▶ translated nucleotide

Protein BLAST
protein ▶ protein

check box “align two or more sequences” and enter AF241113 in the first box and the remaining 7 numbers (from page 7 section a in the Student Lab Handout) in the second box, then click on BLAST

Align Sequences Nucleotide BLAST

blastn blastp blastx tblastn tblastx

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

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 [?](#)

Enter Subject Sequence

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

From To

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

Program Selection

Optimize for

- Highly similar sequences (megablast)
- More dissimilar sequences (discontiguous megablast)
- Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

In the results, click on “distance tree of results”

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

Job Title AF241113:Brassica oleracea var. acephala isolate...

RID [N92E45HK11R](#) Search expires on 09-30 23:44 pm [Download All](#) [▼](#)

Program Blast 2 sequences [Citation](#) [▼](#)

Query ID [AF241113.1](#) (nucleic acid)

Query Descr Brassica oleracea var. acephala isolate HRI/CGN 6210 cau ...

Query Length 1976

Subject ID [AF241114.1](#) and 6 more subject(s) (nucleic acid)

Subject Descr [See details](#) [▼](#)

Subject Length 13903

Filter Results

Percent Identity E value Query Coverage

to to to

Descriptions Graphic Summary Alignments

Sequences producing significant alignments Download [▼](#) [New](#) Select columns [▼](#) Show [?](#)

select all 7 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Brassica oleracea var. oleracea isolate HRI/CGN 8707 cauliflower gene, partial cds	Brassica oleracea var. oleracea	3558	3558	99%	0.0	99.95%	1977	AF2411149.1
<input checked="" type="checkbox"/>	Brassica oleracea var. acephala isolate HRI/CGN 4036 cauliflower gene, partial cds	Brassica oleracea var. viridis	3545	3545	99%	0.0	99.85%	1974	AF2411114.1
<input checked="" type="checkbox"/>	Brassica oleracea var. oleracea isolate HRI/CGN 11125 cauliflower gene, partial cds	Brassica oleracea var. oleracea	3542	3542	99%	0.0	99.80%	1995	AF2411140.1
<input checked="" type="checkbox"/>	Brassica oleracea var. italica isolate HRI/CGN 10674 cauliflower gene, partial sequence	Brassica oleracea var. italica	3525	3525	99%	0.0	99.65%	1995	AF2411130.1
<input checked="" type="checkbox"/>	Brassica oleracea var. botrytis isolate HRI/CGN 7369 cauliflower gene, partial sequence	Brassica oleracea var. botrytis	3525	3525	99%	0.0	99.65%	1972	AF2411127.1
<input checked="" type="checkbox"/>	Brassica oleracea var. botrytis isolate HRI/CGN 8567 cauliflower gene, partial sequence	Brassica oleracea var. botrytis	3521	3521	99%	0.0	99.59%	1995	AF2411129.1
<input checked="" type="checkbox"/>	Brassica oleracea var. botrytis isolate HRI/CGN 4814 cauliflower gene, partial sequence	Brassica oleracea var. botrytis	3521	3521	99%	0.0	99.59%	1995	AF2411123.1

