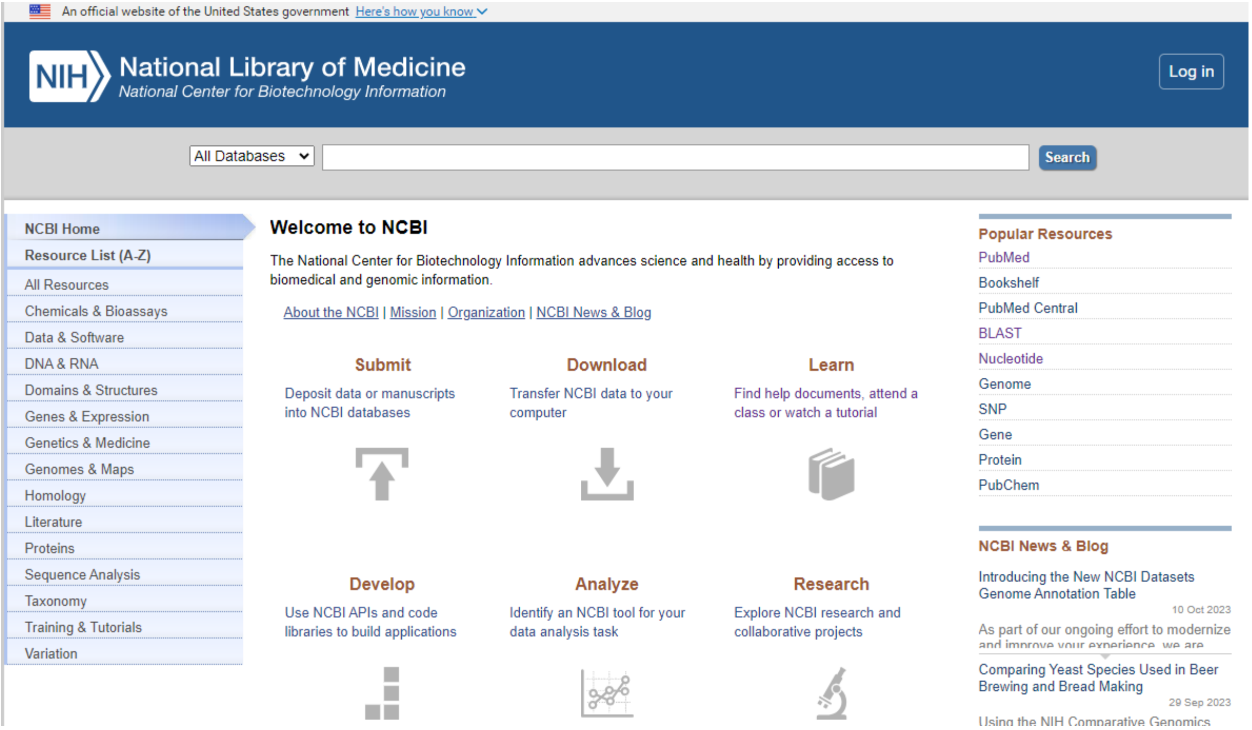
USING NCBI to analyze DNA and protein Sequences for BoCAL and BobCAL Genes

1. Find the sequences you will be working with

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



Welcome to NCBI

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

Deposit data or manuscripts into NCBI databases

Transfer NCBI data to your computer

Find help documents, attend a class or watch a tutorial

Use NCBI APIs and code libraries to build applications

Identify an NCBI tool for your data analysis task

Explore NCBI research and collaborative projects

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

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

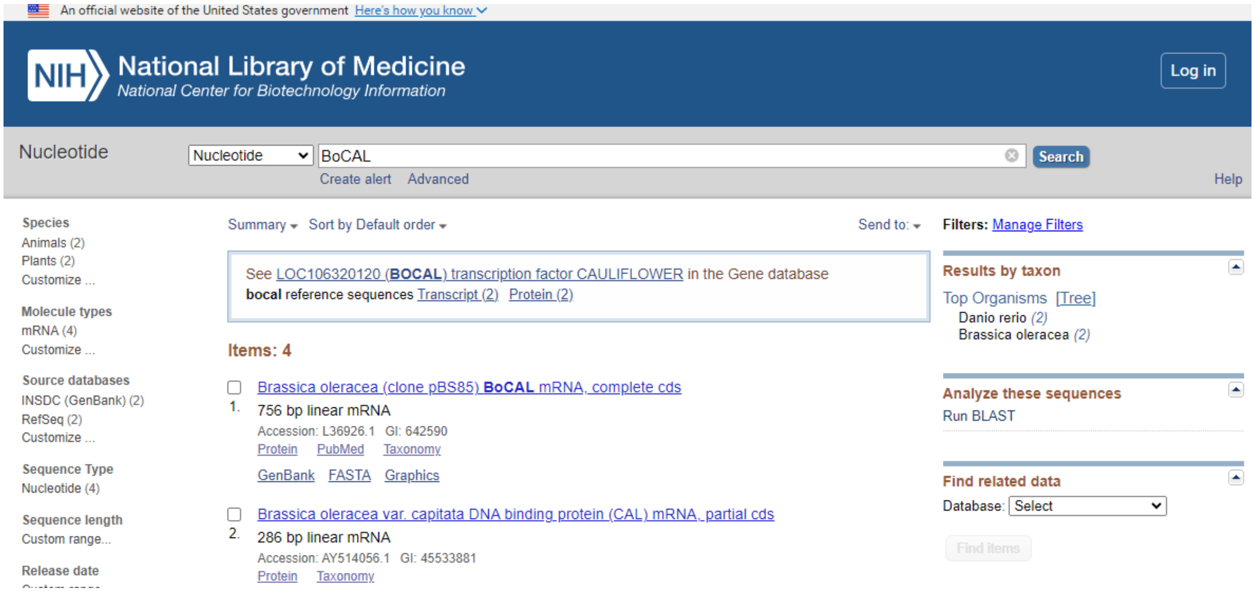
Nucleotide

BoCAL

Welcome to NCBI

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

Results for BoCAL, make note of accession L36926



See LOC106320120 (**BOCAL**) transcription factor CAULIFLOWER in the Gene database
bocal reference sequences Transcript (2) Protein (2)

Items: 4

☐ 1. Brassica oleracea (clone pBS85) **BoCAL** mRNA, complete cds
756 bp linear mRNA
Accession: L36926.1 GI: 642590
Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 2. Brassica oleracea var. capitata DNA binding protein (CAL) mRNA, partial cds
286 bp linear mRNA
Accession: AY514056.1 GI: 45533881
Protein Taxonomy

Results for BobCAL, make note of accession L36927

See LOC106320120 (**BOBCAL**) transcription factor CAULIFLOWER in the Gene database
bobcal reference sequences Transcript (2) Protein (2)

Items: 6

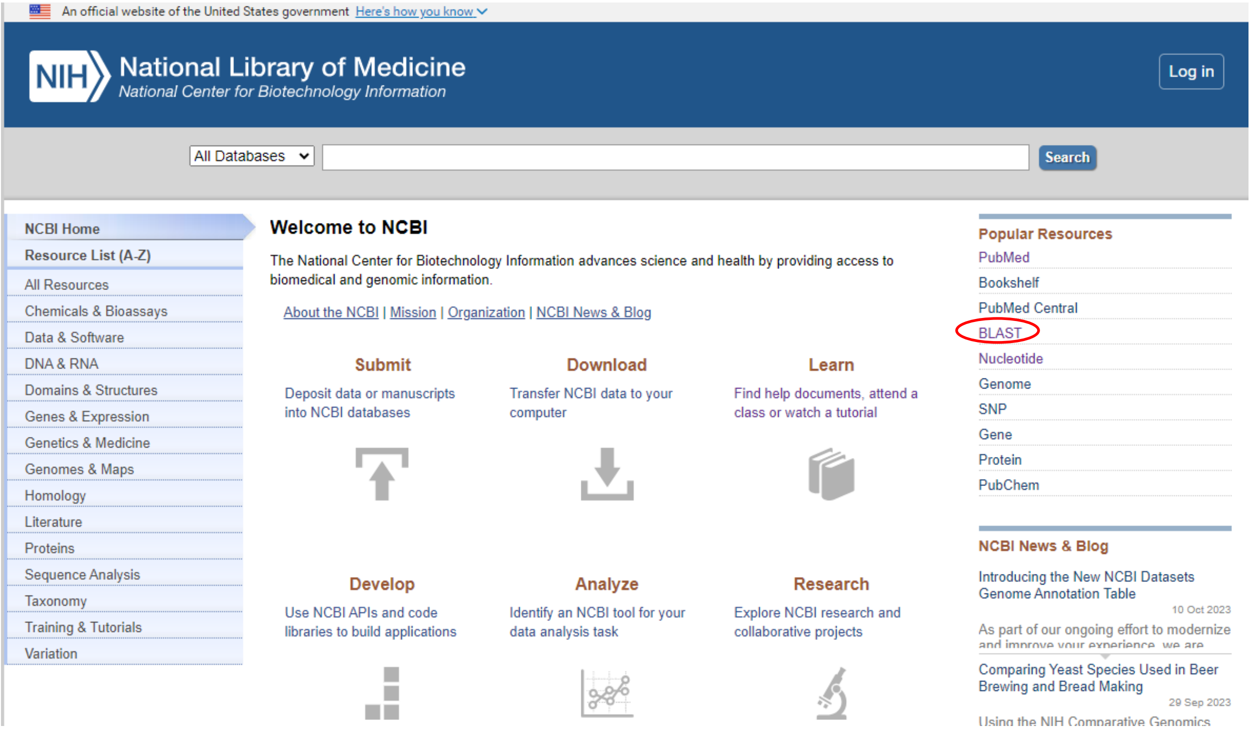
☐ 1. Brassica oleracea (subspecies botrytis) **BobCAL** mRNA, complete cds
756 bp linear mRNA
Accession: L36927.1 GI: 642592
Protein PubMed Taxonomy
GenBank FASTA Graphics

☐ 2. Brassica oleracea var. botrytis DNA binding protein variant d (CAL) mRNA, complete cds, alternatively spliced
892 bp linear mRNA
Accession: AY514055.1 GI: 45533879
Protein Taxonomy

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



Welcome to NCBI

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

Deposit data or manuscripts into NCBI databases

Transfer NCBI data to your computer

Find help documents, attend a class or watch a tutorial

Use NCBI APIs and code libraries to build applications

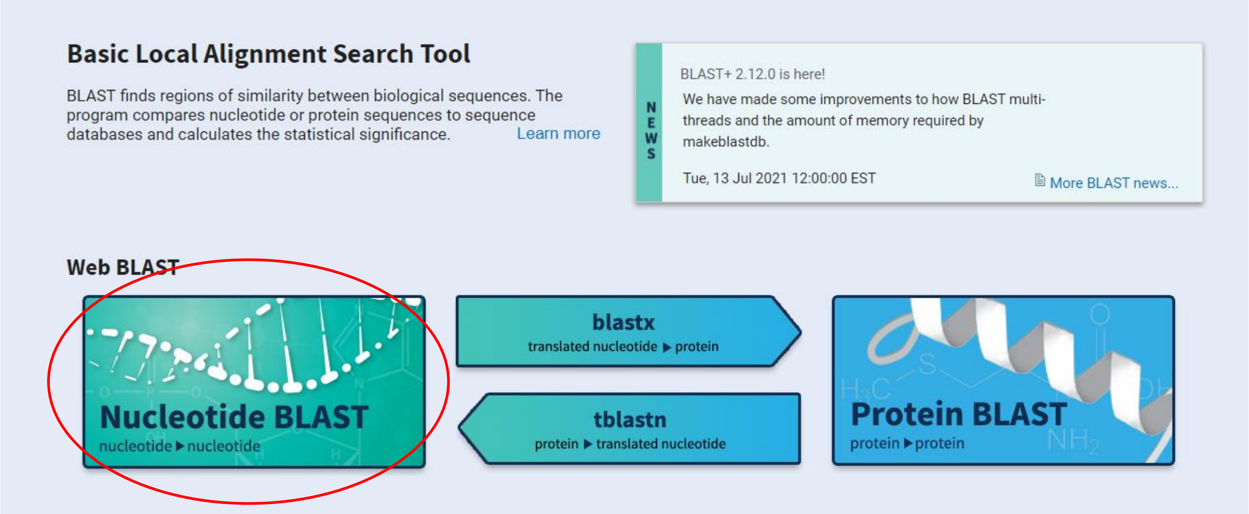
Identify an NCBI tool for your data analysis task

Explore NCBI research and collaborative projects

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

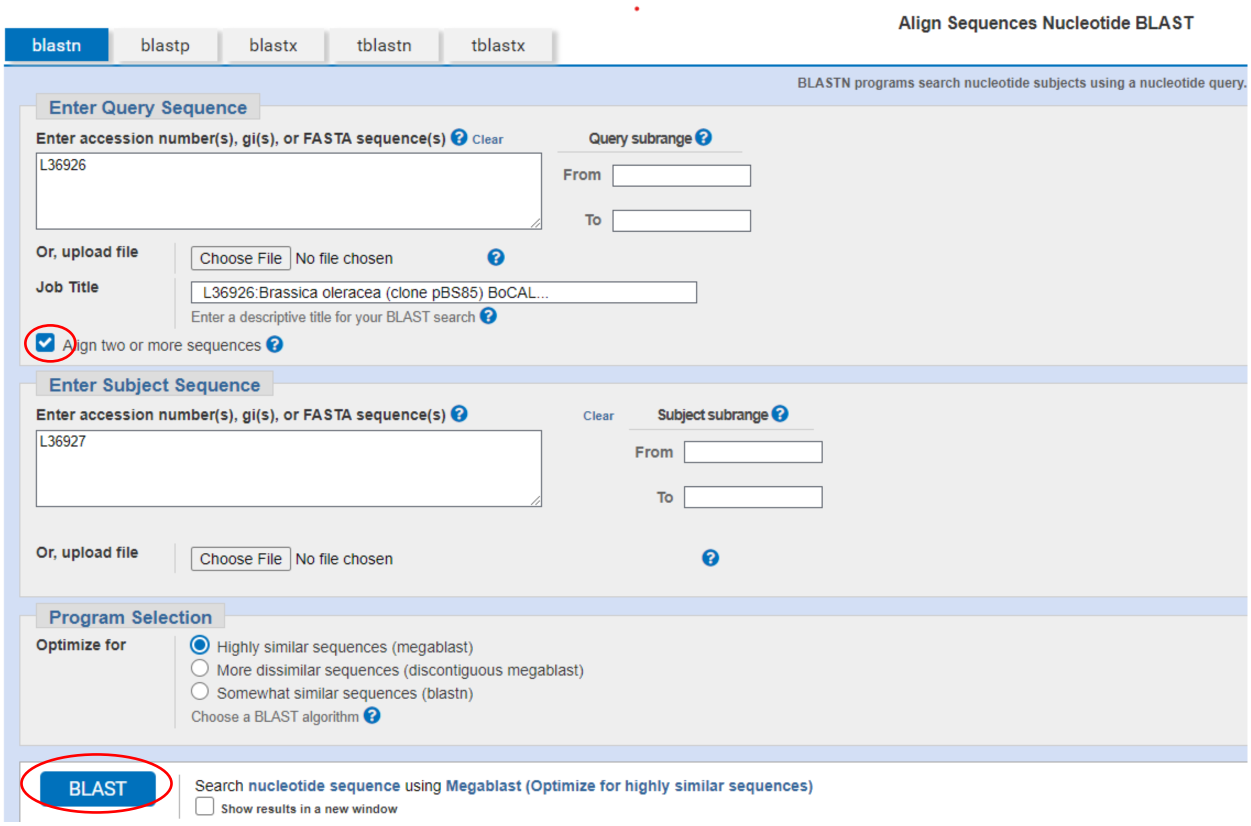
BLAST+ 2.12.0 is here!

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...

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



Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s): L36926

Query subrange: From ___ To ___

Or, upload file: No file chosen

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

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): L36927

Subject subrange: From ___ To ___

Or, upload file: No file chosen

Program Selection

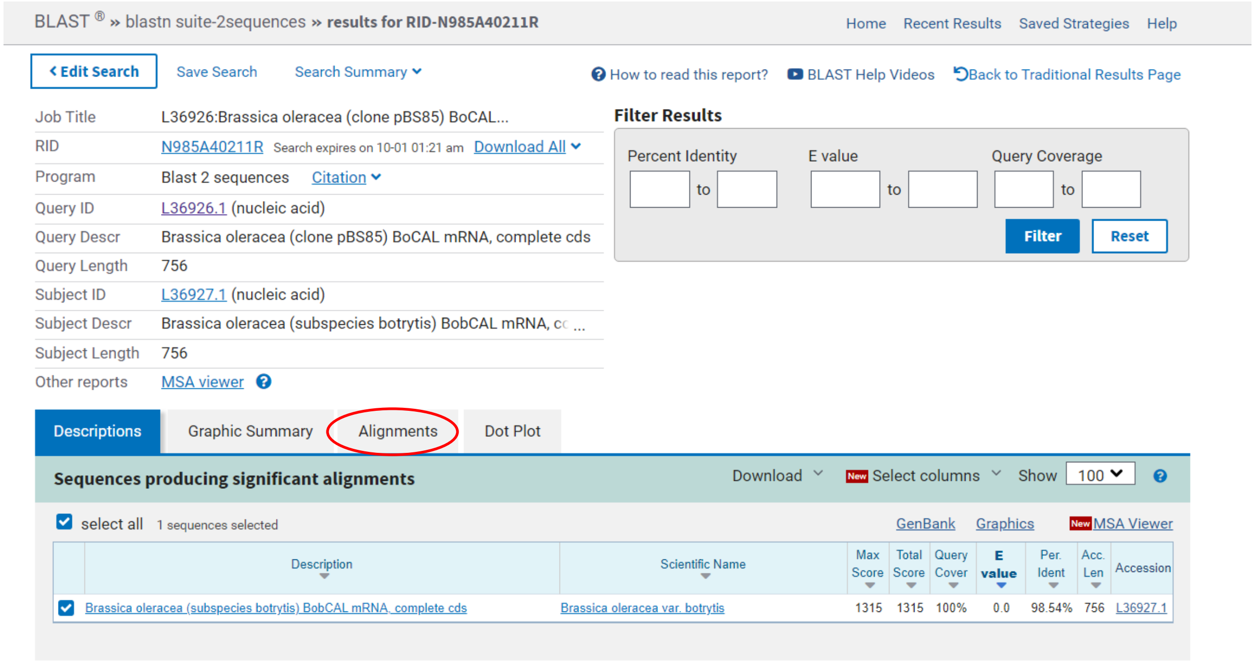
Optimize for:
- ● Highly similar sequences (megablast)
- ○ More dissimilar sequences (discontiguous megablast)
- ○ Somewhat similar sequences (blastn)

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



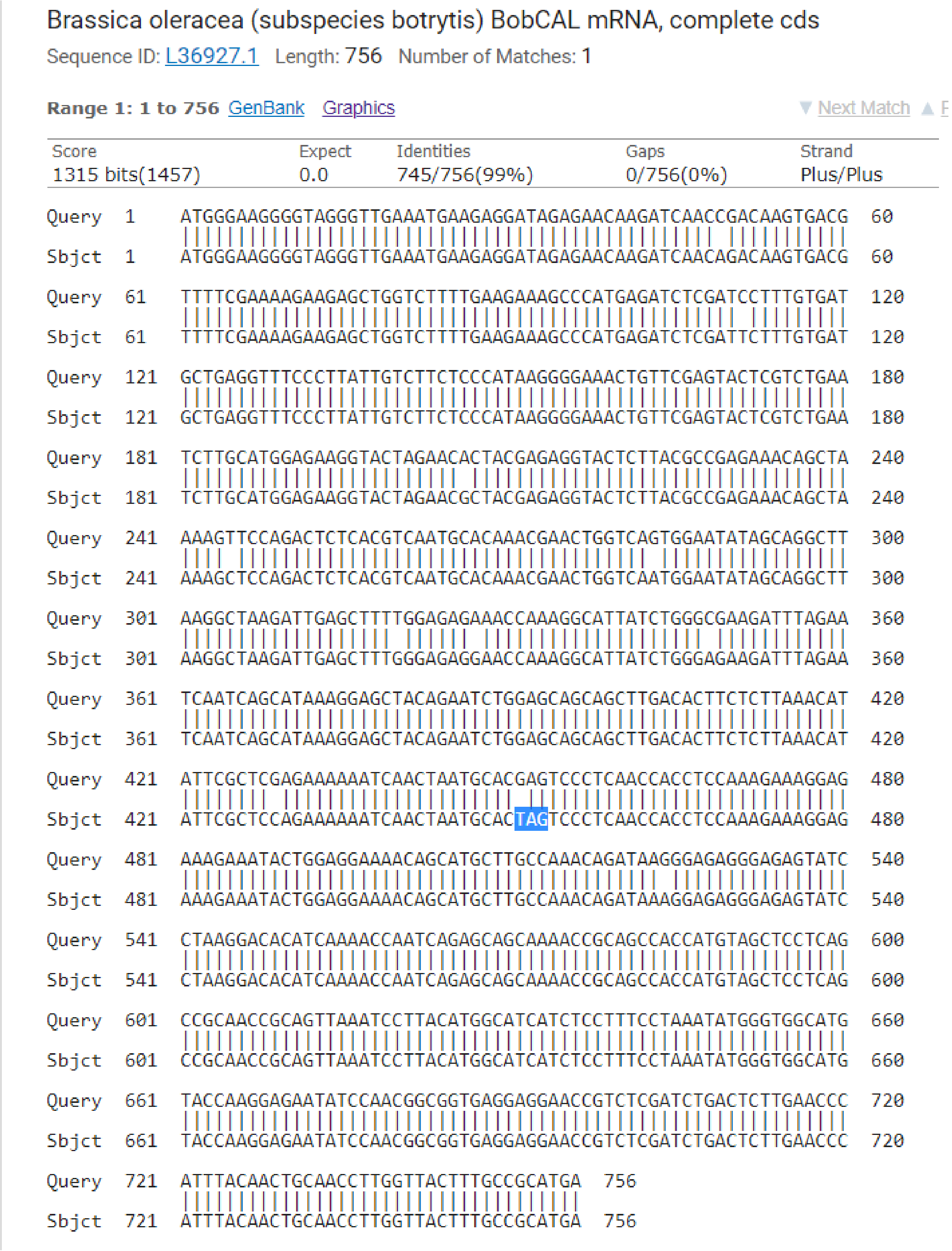
Job Title	L36926:Brassica oleracea (clone pBS85) BoCAL...
RID	N985A40211R Search expires on 10-01 01:21 am Download All
Program	Blast 2 sequences Citation
Query ID	L36926.1 (nucleic acid)
Query Descr	Brassica oleracea (clone pBS85) BoCAL mRNA, complete cds
Query Length	756
Subject ID	L36927.1 (nucleic acid)
Subject Descr	Brassica oleracea (subspecies botrytis) BobCAL mRNA, cc ...
Subject Length	756
Other reports	MSA viewer

Sequences producing significant alignments

☑ select all 1 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	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 (subjct 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

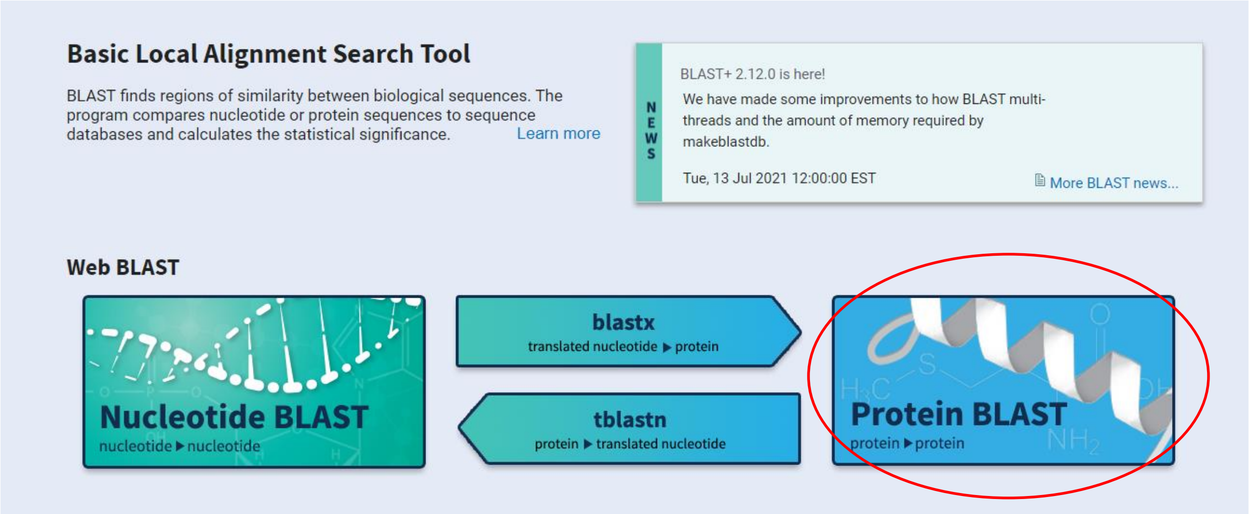
Range 1: 1 to 756 GenBank Graphics ▼ Next Match ▲ P

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  GCTGAGGTTTCCCTTATTGTCTTCTCCCATAAGGGGAAACTGTTCGAGTACTCGTCTGAA  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  121  GCTGAGGTTTCCCTTATTGTCTTCTCCCATAAGGGGAAACTGTTCGAGTACTCGTCTGAA  180

Query  181  TCTTGCATGGAGAAGGTACTAGAACACTACGAGAGGTACTCTTACGCCGAGAAACAGCTA  240
            ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Sbjct  181  TCTTGCATGGAGAAGGTACTAGAACGCTACGAGAGGTACTCTTACGCCGAGAAACAGCTA  240

Query  241  AAAGTTCCAGACTCTCACGTCAATGCACAAACGAACTGGTCAGTGGAATATAGCAGGCTT  300
            |||| ||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Sbjct  241  AAAGCTCCAGACTCTCACGTCAATGCACAAACGAACTGGTCAATGGAATATAGCAGGCTT  300

Query  301  AAGGCTAAGATTGAGCTTTTGGAGAGAAACCAAAGGCATTATCTGGGCGAAGATTTAGAA  360
            ||||||||||||||||||| |||||| |||||||||||||||||||| ||||||||||||
Sbjct  301  AAGGCTAAGATTGAGCTTTGGGAGAGGAACCAAAGGCATTATCTGGGAGAAGATTTAGAA  360

Query  361  TCAATCAGCATAAAGGAGCTACAGAATCTGGAGCAGCAGCTTGACACTTCTCTTAAACAT  420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  361  TCAATCAGCATAAAGGAGCTACAGAATCTGGAGCAGCAGCTTGACACTTCTCTTAAACAT  420

Query  421  ATTCGCTCGAGAAAAAATCAACTAATGCACGAGTCCCTCAACCACCTCCAAAGAAAGGAG  480
            |||||||| ||||||||||||||||||||| |||||||||||||||||||||||||||||
Sbjct  421  ATTCGCTCCAGAAAAAATCAACTAATGCACTAGTCCCTCAACCACCTCCAAAGAAAGGAG  480

Query  481  AAAGAAATACTGGAGGAAAACAGCATGCTTGCCAAACAGATAAGGGAGAGGGAGAGTATC  540
            ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Sbjct  481  AAAGAAATACTGGAGGAAAACAGCATGCTTGCCAAACAGATAAAGGAGAGGGAGAGTATC  540

Query  541  CTAAGGACACATCAAAACCAATCAGAGCAGCAAAACCGCAGCCACCATGTAGCTCCTCAG  600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541  CTAAGGACACATCAAAACCAATCAGAGCAGCAAAACCGCAGCCACCATGTAGCTCCTCAG  600

Query  601  CCGCAACCGCAGTTAAATCCTTACATGGCATCATCTCCTTTCCTAAATATGGGTGGCATG  660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601  CCGCAACCGCAGTTAAATCCTTACATGGCATCATCTCCTTTCCTAAATATGGGTGGCATG  660

Query  661  TACCAAGGAGAATATCCAACGGCGGTGAGGAGGAACCGTCTCGATCTGACTCTTGAACCC  720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661  TACCAAGGAGAATATCCAACGGCGGTGAGGAGGAACCGTCTCGATCTGACTCTTGAACCC  720

Query  721  ATTTACAACTGCAACCTTGGTTACTTTGCCGCATGA  756
            ||||||||||||||||||||||||||||||||||||
Sbjct  721  ATTTACAACTGCAACCTTGGTTACTTTGCCGCATGA  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!

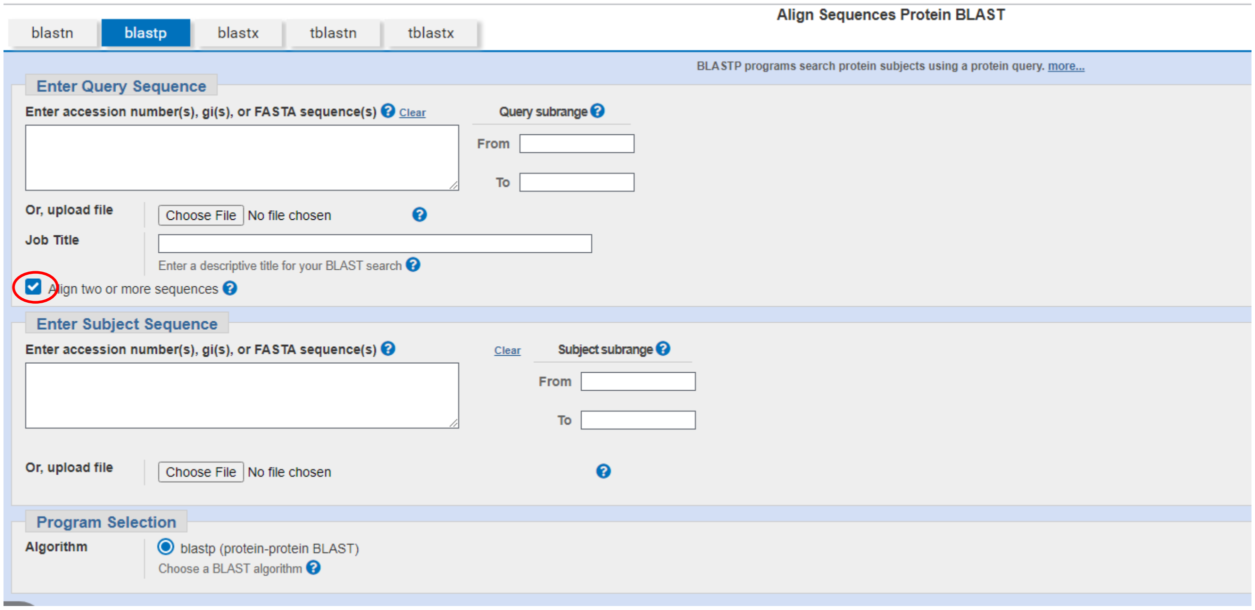
We have made some improvements to how BLAST multi-threads and the amount of memory required by makeblastdb.

Tue, 13 Jul 2021 12:00:00 EST

More BLAST news...

Web BLAST

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)



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: Choose 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

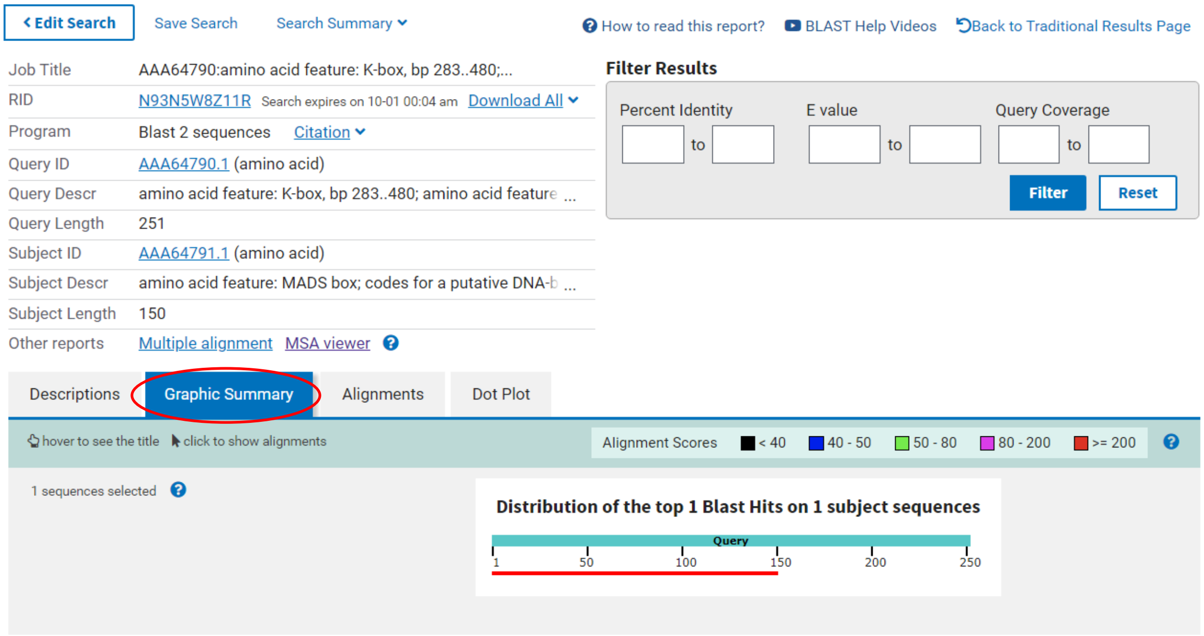
Or, upload file: Choose File No file chosen

Program Selection

Algorithm: ● blastp (protein-protein BLAST)

Choose a BLAST algorithm

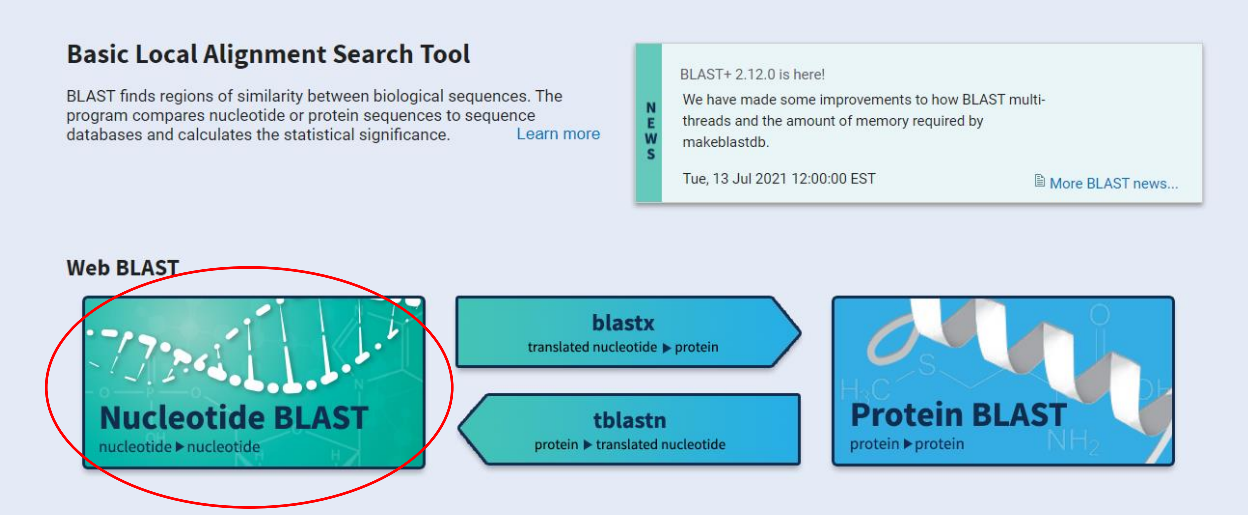
In the results click on graphic summary, you can see that the subject protein (BobCAL AAA64791 in red) is much shorter than the Query protein (BoCAL AAA64790 in blue) because the translation stopped at the stop codon identified in the sequence above:



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

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

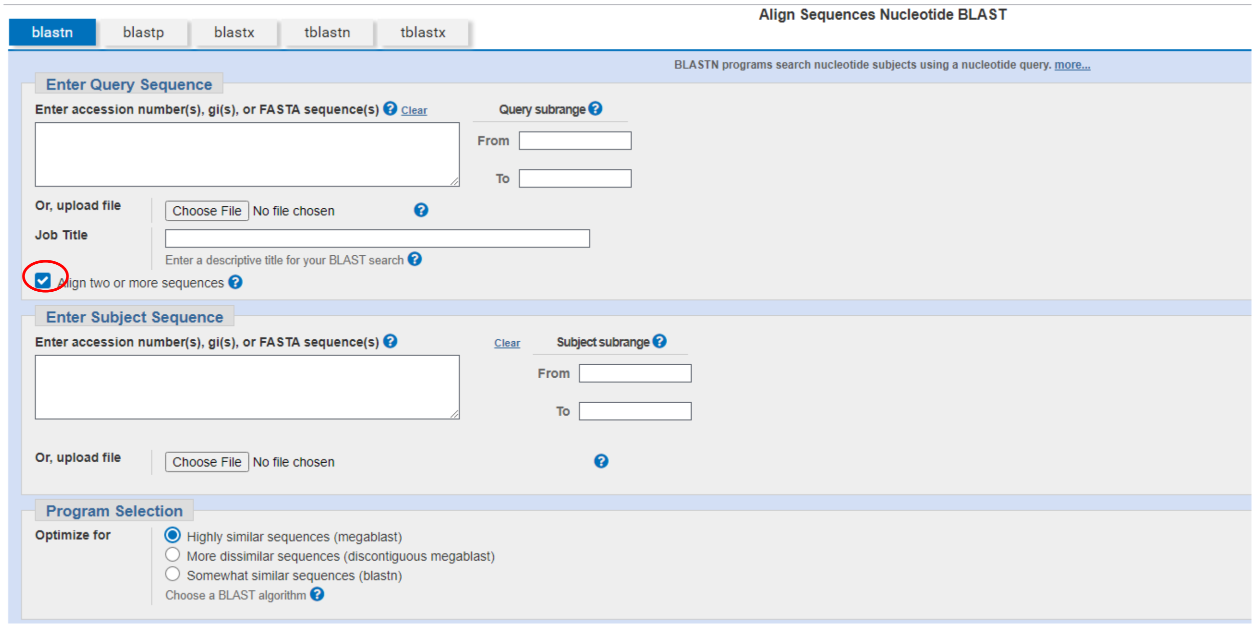
BLAST finds regions of similarity between biological sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance. Learn more

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

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

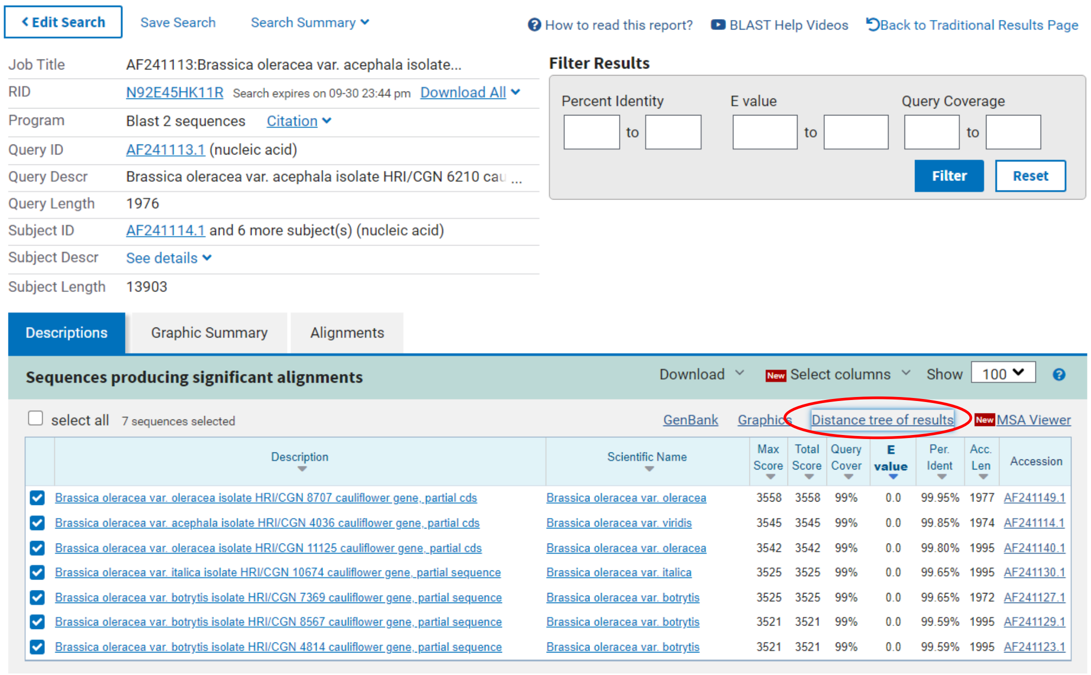


BLASTN programs search nucleotide subjects using a nucleotide query. more...

☑ Align two or more sequences

Optimize for
● Highly similar sequences (megablast)
○ More dissimilar sequences (discontiguous megablast)
○ Somewhat similar sequences (blastn)

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



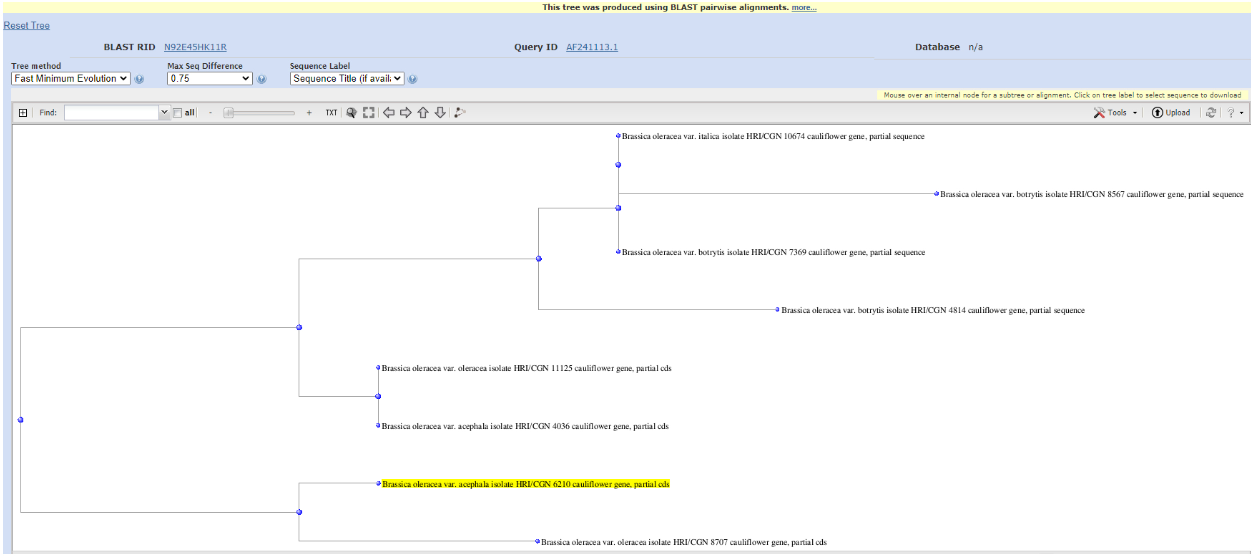
Job Title	AF241113:Brassica oleracea var. acephala isolate...
RID	N92E45HK11R Search expires on 09-30 23:44 pm Download All
Program	Blast 2 sequences Citation
Query ID	AF241113.1 (nucleic acid)
Query Descr	Brassica oleracea var. acephala isolate HRI/CGN 6210 cau ...
Query Length	1976
Subject ID	AF241114.1 and 6 more subject(s) (nucleic acid)
Subject Descr	See details
Subject Length	13903

Sequences producing significant alignments

☐ select all 7 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
☑	Brassica oleracea var. oleracea isolate HRI/CGN 8707 cauliflower gene, partial cds	Brassica oleracea var. oleracea	3558	3558	99%	0.0	99.95%	1977	AF241149.1
☑	Brassica oleracea var. acephala isolate HRI/CGN 4036 cauliflower gene, partial cds	Brassica oleracea var. viridis	3545	3545	99%	0.0	99.85%	1974	AF241114.1
☑	Brassica oleracea var. oleracea isolate HRI/CGN 11125 cauliflower gene, partial cds	Brassica oleracea var. oleracea	3542	3542	99%	0.0	99.80%	1995	AF241140.1
☑	Brassica oleracea var. italica isolate HRI/CGN 10674 cauliflower gene, partial sequence	Brassica oleracea var. italica	3525	3525	99%	0.0	99.65%	1995	AF241130.1
☑	Brassica oleracea var. botrytis isolate HRI/CGN 7369 cauliflower gene, partial sequence	Brassica oleracea var. botrytis	3525	3525	99%	0.0	99.65%	1972	AF241127.1
☑	Brassica oleracea var. botrytis isolate HRI/CGN 8567 cauliflower gene, partial sequence	Brassica oleracea var. botrytis	3521	3521	99%	0.0	99.59%	1995	AF241129.1
☑	Brassica oleracea var. botrytis isolate HRI/CGN 4814 cauliflower gene, partial sequence	Brassica oleracea var. botrytis	3521	3521	99%	0.0	99.59%	1995	AF241123.1

You will see the following phylogenetic tree:



This tree was produced using BLAST pairwise alignments. more...

BLAST RID N92E45HK11R

Query ID AF241113.1

Database n/a

Tree method: Fast Minimum Evolution

Max Seq Difference: 0.75

Sequence Label: Sequence Title (if avail

Mouse over an internal node for a subtree or alignment. Click on tree label to select sequence to download