

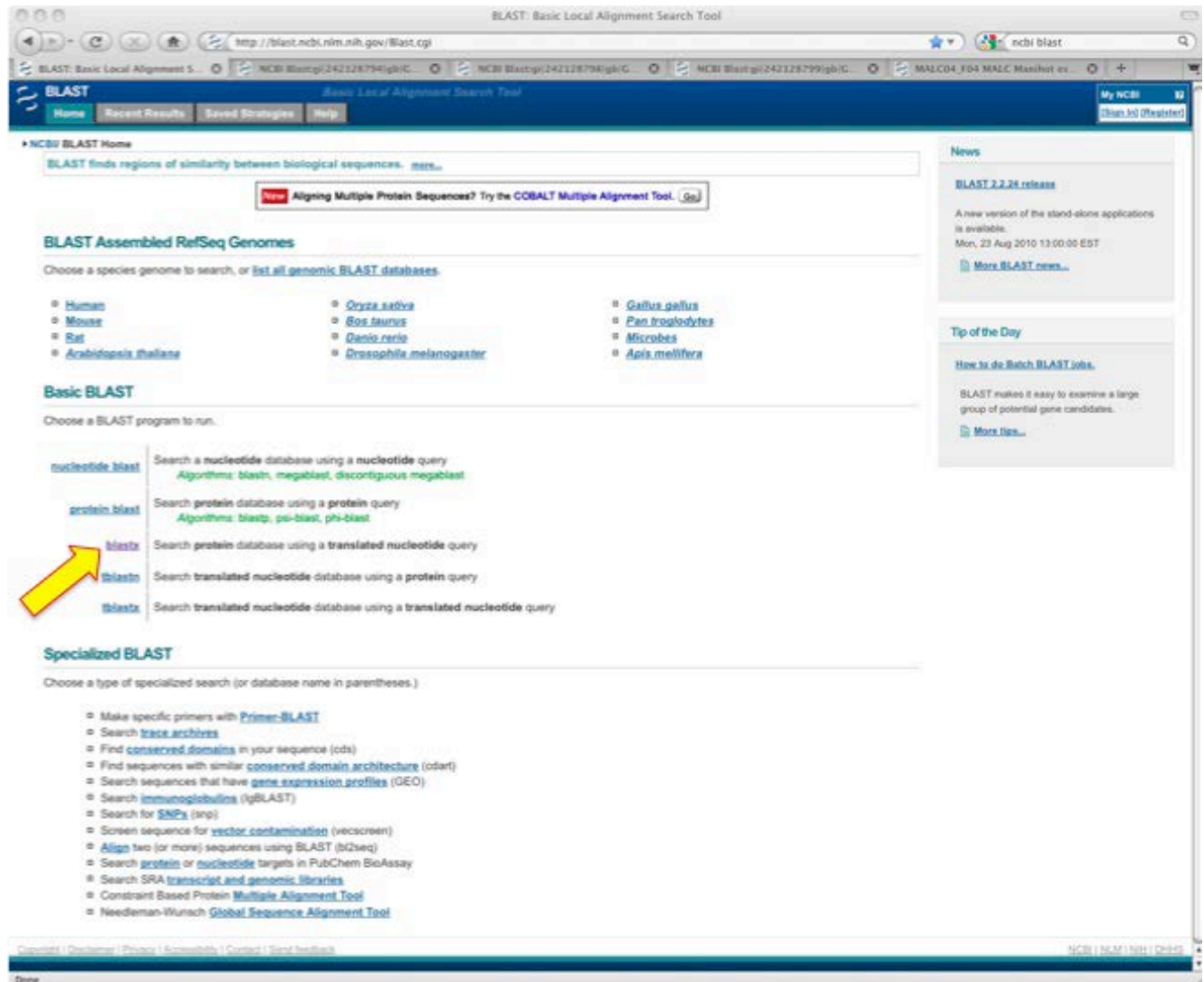
## Worked example - BLAST

Search Entrez Nucleotide for GR421296 and copy the sequence in FASTA format

Go to the NCBI BLAST page: <http://blast.ncbi.nlm.nih.gov/Blast.cgi>

See the different BLAST programs

Click on "blastx"



The screenshot shows the NCBI BLAST website interface. At the top, there is a navigation bar with "Home", "Recent Results", "Search Strategies", and "Help". Below this is a search bar with the text "BLAST finds regions of similarity between biological sequences." and a "New" alert for "Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool." The main content area is divided into sections: "BLAST Assembled RefSeq Genomes" with a list of species (Human, Mouse, Rat, Arabidopsis thaliana, Oryza sativa, Bos taurus, Danio rerio, Drosophila melanogaster, Gallus gallus, Pan troglodytes, Microbes, Apis mellifera); "Basic BLAST" with a list of programs (nucleotide\_blast, protein\_blast, blastx, tblastn, tblastx) and a yellow arrow pointing to the "blastx" option; and "Specialized BLAST" with a list of specialized search options (Primer-BLAST, trace.archives, conserved domains, conserved domain architecture, gene expression profiles, IgBLAST, SNPs, vector contamination, Align two or more sequences, protein or nucleotide targets, SRA transcript and genomic libraries, Constraint Based Protein Multiple Alignment Tool, Needleman-Wunsch Global Sequence Alignment Tool). The footer contains navigation links and the NCBI logo.

Paste the FASTA sequence in the input box, or paste just the accession number GR421296

Select "non-redundant" under "database" and check "Show results in a new window"

Click the BLAST button to run a blastx search trying to find out what protein that sequence may code for.

BLAST Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI (Sign In) (Register)

NCBI/ BLAST/ blastx

blastx search protein databases using a translated nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter Query Sequence

Enter accession number, gi, or FASTA sequence  Clear Query subrange

From

To

Or, upload file  Browse...

Genetic code Standard (L)

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database Non-redundant protein sequences (nr)

Organism Optional  Exclude

Exclude Optional  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query Optional

Enter an Entrez query to limit search

BLAST Search database Non-redundant protein sequences (nr) using Blastx (search protein databases using a translated nucleotide query)

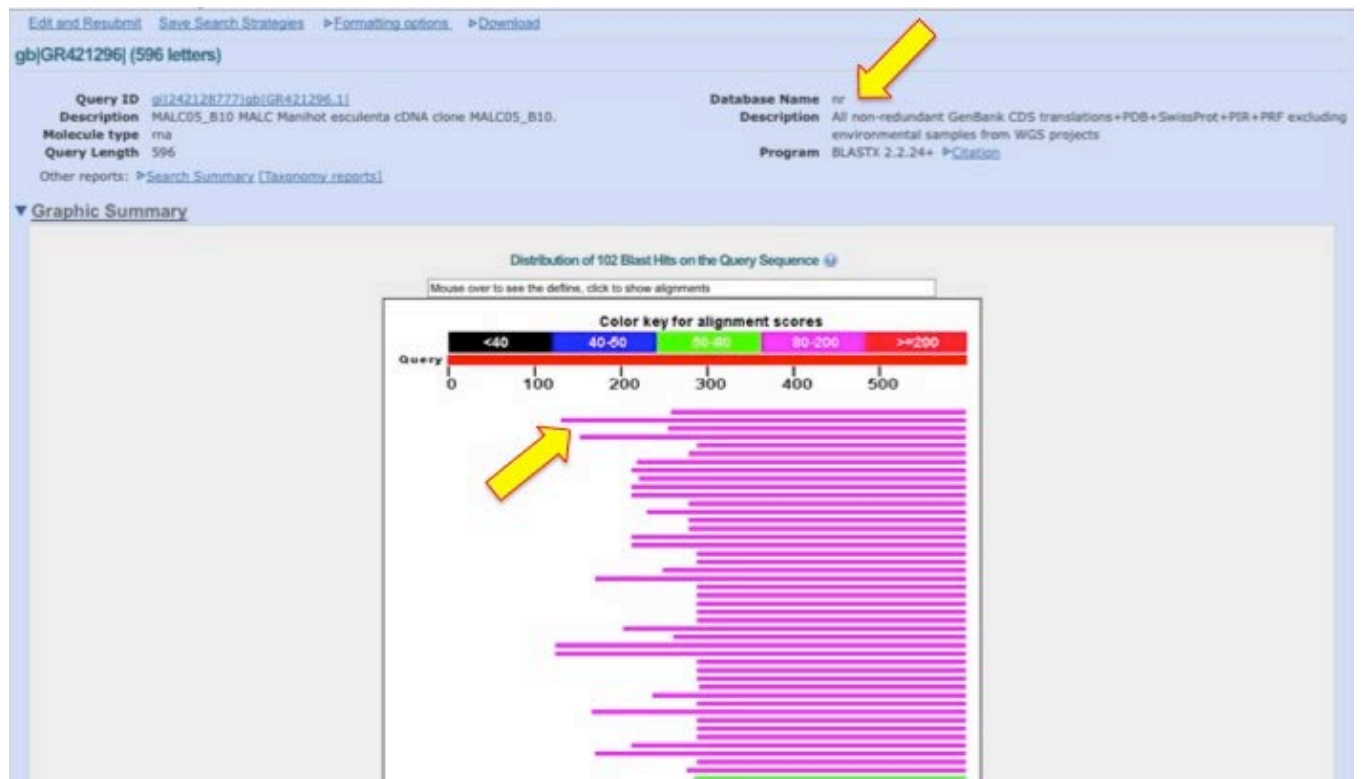
Show results in a new window

[blastx parameters](#)

See the database used (top right)

You can display or hide the different result sections

Mouse over the graphical display of the alignments to see the best hits



Click on the in the description links to see the GB entry for each hit. Click on the Max score to jump to the alignment

▼ Descriptions

Legend for links to other resources: UniGene GEO Gene Structure Map Viewer PubChem BioAssay

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
XP_002516048.1	serine-threonine protein kinase, plant-type, putative [Ricinus communis] >gb EF02448.1	88.6	88.6	56%	4e-16	46%	<a href="#">G</a>
XP_002516047.1	serine-threonine protein kinase, plant-type, putative [Ricinus communis] >gb EF02447.1	88.2	88.2	78%	5e-16	37%	<a href="#">G</a>
NP_199788.1	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis thaliana]	87.0	87.0	57%	1e-15	41%	<a href="#">UG</a>
XP_002864015.1	hypothetical protein ARALYDRAFT_495037 [Arabidopsis lyrata subsp. lyrata] >g	85.9	85.9	74%	3e-15	36%	<a href="#">G</a>
BAE46451.1	putative receptor protein kinase PERK1 [Glycine max] >gb ACM89502.1  PERK1	85.5	85.5	51%	3e-15	46%	<a href="#">G</a>
XP_002864016.1	predicted protein [Arabidopsis lyrata subsp. lyrata] >gb EFH40275.1  predicted	84.7	84.7	53%	6e-15	43%	<a href="#">G</a>
AC137407.1	leucine-rich repeat transmembrane protein kinase [Glycine max]	83.6	83.6	63%	1e-14	38%	
CB129612.3	unnamed protein product [Vitis vinifera]	83.2	83.2	64%	2e-14	39%	
XP_002441998.1	hypothetical protein SORBITRAFT_08g006710 [Sorghum bicolor] >gb EES1583	83.2	83.2	62%	2e-14	37%	<a href="#">UG</a>
XP_002269453.1	PREDICTED: hypothetical protein [Vitis vinifera]	83.2	83.2	64%	2e-14	38%	<a href="#">UG</a>
CAN68980.1	hypothetical protein [Vitis vinifera]	83.2	83.2	64%	2e-14	38%	
NP_199789.5	ATP binding / kinase / protein serine/threonine kinase [Arabidopsis thaliana]	82.8	82.8	53%	2e-14	42%	<a href="#">UG</a>
XP_002269419.1	PREDICTED: hypothetical protein [Vitis vinifera]	82.8	82.8	61%	2e-14	39%	<a href="#">UG</a>
BAD93860.1	receptor protein kinase-like [Arabidopsis thaliana]	82.8	82.8	53%	2e-14	42%	
RAA98166.1	receptor protein kinase-like [Arabidopsis thaliana]	82.8	82.8	53%	2e-14	42%	
NP_195120.2	protein kinase family protein [Arabidopsis thaliana] >dbj BAC43092.1  putative	82.8	82.8	64%	2e-14	38%	<a href="#">UG</a>
CAA18923.1	putative serine/threonine protein kinase [Arabidopsis thaliana] >emb CAB8016	82.8	82.8	64%	2e-14	38%	
RAF43699.1	PERK1-like protein kinase [Nicotiana tabacum]	82.4	82.4	51%	3e-14	44%	
AAK21965.1	receptor protein kinase PERK1 [Brassica napus]	82.4	82.4	51%	3e-14	44%	
XP_002539469.1	hypothetical protein SORBITRAFT_09g007240 [Sorghum bicolor] >gb EES1789	82.0	82.0	58%	4e-14	41%	<a href="#">UG</a>
XP_002560194.1	hypothetical protein SORBITRAFT_02g024390 [Sorghum bicolor] >gb EER9671	82.0	82.0	71%	4e-14	38%	<a href="#">UG</a>
XP_002514292.1	ATP binding protein, putative [Ricinus communis] >gb EEF48251.1  ATP binding	81.6	81.6	51%	5e-14	45%	<a href="#">G</a>
NP_001145267.1	hypothetical protein LOC100279274 [Zea mays] >gb ACL52550.1  unknown [Z	81.6	81.6	51%	5e-14	43%	<a href="#">UG</a>
NP_001148076.1	receptor protein kinase PERK1 [Zea mays] >gb ACG29662.1  receptor protein k	81.6	81.6	51%	5e-14	43%	<a href="#">UG</a>
ACG28793.1	receptor protein kinase PERK1 [Zea mays]	81.6	81.6	51%	5e-14	43%	
ACF86764.1	unknown [Zea mays]	81.6	81.6	51%	5e-14	43%	
XP_002315871.1	predicted protein [Populus trichocarpa] >gb EEF02042.1  predicted protein [Po	81.3	81.3	65%	6e-14	36%	
XP_002328823.1	predicted protein [Populus trichocarpa] >gb EEE77472.1  predicted protein [Po	81.3	81.3	56%	6e-14	41%	<a href="#">G</a>
EES1910.1	hypothetical protein Osl_33512 [Oryza sativa Japonica Group]	81.3	81.3	79%	6e-14	32%	
EEC62947.1	hypothetical protein Osl_35675 [Oryza sativa Indica Group]	81.3	81.3	79%	6e-14	32%	
XP_002885643.1	hypothetical protein ARALYDRAFT_319144 [Arabidopsis lyrata subsp. lyrata] >g	80.9	80.9	51%	8e-14	44%	<a href="#">G</a>
XP_002883531.1	hypothetical protein ARALYDRAFT_479967 [Arabidopsis lyrata subsp. lyrata] >g	80.9	80.9	51%	8e-14	44%	<a href="#">G</a>
BAH52220.1	AT3G24550 [Arabidopsis thaliana]	80.9	80.9	51%	8e-14	44%	<a href="#">UG</a>
XP_002277905.1	PREDICTED: hypothetical protein [Vitis vinifera] >emb CB127111.3  unnamed p	80.9	80.9	51%	8e-14	45%	<a href="#">G</a>
ACF83632.1	unknown [Zea mays] >gb ACN28727.1  unknown [Zea mays]	80.9	80.9	60%	8e-14	42%	<a href="#">UG</a>
AAK43886.1	protein kinase-like protein [Arabidopsis thaliana] >gb AAP37768.1  At3g24600	80.9	80.9	51%	8e-14	44%	
AAK92929.1	At5g49760 [Oryza sativa Japonica Group] >gb ABA92396.1  Protein kinase don	80.9	80.9	71%	8e-14	34%	
BAD02941.1	somatic embryogenesis receptor kinase-like protein [Arabidopsis thaliana]	80.9	80.9	51%	8e-14	44%	
NP_189098.1	ATPERK1 [PROLINE EXTENSIN-LIKE RECEPTOR KINASE 1]; ATP binding / protei	80.9	80.9	51%	8e-14	44%	<a href="#">UG</a>
XP_00288452.1	hypothetical protein ARALYDRAFT_475678 [Arabidopsis lyrata subsp. lyrata] >g	80.5	80.5	51%	1e-13	43%	<a href="#">G</a>
XP_002869145.1	predicted protein [Arabidopsis lyrata subsp. lyrata] >gb EFH45404.1  predicted	80.5	80.5	64%	1e-13	37%	<a href="#">G</a>
RAZ09065.1	hypothetical protein Osl_31326 [Oryza sativa Indica Group]	80.5	80.5	71%	1e-13	35%	
XP_002280718.1	PREDICTED: hypothetical protein [Vitis vinifera]	80.1	80.1	51%	1e-13	43%	<a href="#">UG</a>
XP_002519381.1	Serine/threonine-protein kinase PBS1, putative [Ricinus communis] >gb EEF42	80.1	80.1	53%	1e-13	42%	<a href="#">G</a>
XP_002883530.1	predicted protein [Arabidopsis lyrata subsp. lyrata] >gb EFH59789.1  predicted	79.7	79.7	52%	2e-13	46%	<a href="#">G</a>

▼ Alignments

Select All Get selected sequences

```

>[ref|XP_002516048.1|] serine-threonine protein kinase, plant-type, putative [Ricinus
communis]
gb|EF02448.1| serine-threonine protein kinase, plant-type, putative [Ricinus
communis]
Length=485
GENE ID: 8274500 NCBI:1494420 | serine-threonine protein kinase, plant-type,
putative [Ricinus communis]
Score = 88.6 bits (218), Expect = 4e-16
Identities = 54/115 (46%), Positives = 72/115 (62%), Gaps = 12/115 (10%)
Frame = +3
Query 258 AAANKKGVKT--FSHEELAKAARYFSDHNRLGGDLTGVDFEGELPNGEVVAIKRFXH 431
AA N K +E F+++EL EA FS N LG+G G VFGK LP+G VA+K+ N
Sbjct 73 AAIALKSKFQTSIFATDELEKATNGFS----NILGEGGPGVFGKGLPDRQVAVKRLKA 128
Query 432 QANPEHSEKLANNQYEMAEIILSRIEFQNIKVKIYCNDAASHRLLVYEFVPMNSL 596
+ K ++++E E + I H+N+V +IGVC +HRLLVYEFVPMNSL
Sbjct 129 GS-----KQGDREPFQVEIETIGHIS-HRNLVNLGYCIDLANRLLVYEFVPMNSL 177

>[ref|XP_002516047.1|] serine-threonine protein kinase, plant-type, putative [Ricinus
communis]
gb|EF02447.1| serine-threonine protein kinase, plant-type, putative [Ricinus
communis]
Length=432
GENE ID: 8274499 NCBI:1494410 | serine-threonine protein kinase, plant-type,
putative [Ricinus communis]
Score = 88.2 bits (217), Expect = 5e-16
Identities = 58/155 (37%), Positives = 87/155 (56%), Gaps = 17/155 (10%)
Frame = +3
Query 132 SESSSIFEAATVKKQGEIKETAIFIEANISTERGESAIPFAAANKKGVKTFPSHEELA 311
+E S +E + G+ + AI + N ST E + + R F+++ELA
Sbjct 44 AERSQLPHRQTRKQQLKFLAISL--NTSTTINEKVGQCGQ-----KVPYQRLA 93
Query 312 KAARYFSDHNRLGGDLTGVDFEGELPNGEVVAIKRFXHQAANPEHSEKLANNQYEMAEI 491
A FS + N LG G G+V+K L N +V+A+K+ K+Q + EK +++E E
Sbjct 94 AATGNS--NANKLGGGGPGEVYKGLNSQVIAYKELKYGQDNRKES----EPETILT 147

```

Go back to the BLASTX results window and click "[Taxonomy reports]"

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI Sign In Register

NCBI BLAST/ blasts/ Formatting Results - BEC5CCJH01P

Edit and Resubmit Save Search Strategies Formatting options Download

**gb|GR421296| (596 letters)**

Query ID g|242128777|gb|GR421296.1  
 Description MALC05\_B10 MALC Manihot esculenta cDNA clone MALC05\_B10.  
 Molecule type rna  
 Query Length 596

Database Name nr  
 Description All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects  
 Program BLASTX 2.2.24+ Citation

Other reports: Search Summary **Taxonomy reports**

Graphic Summary

Distribution of 102 Blast Hits on the Query Sequence

Color key for alignment scores

<40	40-60	60-80	80-200	>=200
-----	-------	-------	--------	-------

Query 0 100 200 300 400 500

See the matches ordered by species or lineage

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI Sign In Register

NCBI BLAST/ blasts/ Formatting Results - BEC5CCJH01P Reformat these Results Edit and Resubmit (Sign in above to save your search strategy)

Job Title: gb|GR421296| (596 letters)

**Tax BLAST Report**

Index

- Lineage Report
- Organism Report
- Taxonomy Report
- Help

**Lineage Report**

```

Embryophyta [land plants]
├── Tracheophyta [vascular plants]
│   ├── Spermatophyta [seed plants]
│   └── Magnoliophyta [flowering plants]
│       ├── core eudicotyledons [eudicots]
│       ├── rosids [eudicots]
│       ├── fabids [eudicots]
│       └── Malpighiales [eudicots]
│           ├── Ricinus communis 88 18 hits [eudicots] serine-threonine protein kinase, plant-type, putative [Ricinus]
│           ├── Populus trichocarpa (black cottonwood) 81 10 hits [eudicots] predicted protein [Populus trichocarpa] g|121844911|gb|
│           ├── Glycine max (soybeans) 85 3 hits [eudicots] putative receptor protein kinase PRR3 [Glycine max] g|1122
│           ├── Arabidopsis thaliana (thale-cress) 87 3 hits [eudicots] leucine-rich repeat transmembrane protein kinase, putative
│           ├── Arabidopsis lyrata subsp. lyrata 85 18 hits [eudicots] hypothetical protein ARALY08A07 495037 [Arabidopsis lyrata]
│           ├── Vitis vinifera (wine grape) 83 17 hits [eudicots] unnamed protein product [Vitis vinifera]
│           ├── Brassica napus (rapeseed) 82 1 hit [eudicots] receptor protein kinase PRR1 [Brassica napus]
│           ├── Nicotiana tabacum (tobacco) 82 1 hit [eudicots] PERK1-like protein kinase [Nicotiana tabacum]
│           ├── Sorghum bicolor (milo) 83 10 hits [monocots] hypothetical protein SORL12A07 28202710 [Sorghum bicolor]
│           ├── Zea mays (maize) 81 8 hits [monocots] hypothetical protein ZM0100275274 [Zea mays] g|1219894351|
│           ├── Oryza sativa Japonica Group (Japanese rice) 81 23 hits [monocots] hypothetical protein OsJ 33312 [Oryza sativa Japonica Group]
│           ├── Oryza sativa Indica Group (Indian rice) 81 10 hits [monocots] hypothetical protein OsJ 35473 [Oryza sativa Indica Group]
│           ├── Picea sitchensis 77 1 hit [conifers] unknown [Picea sitchensis]
│           ├── Selaginella moellendorffii 78 17 hits [club-mosses] hypothetical protein SEL000A07 230241 [Selaginella moellendorffii]
│           ├── Marchantia polymorpha 79 1 hit [liverworts] receptor-like kinase [Marchantia polymorpha]
│           └── Physcomitrella patens subsp. patens 78 8 hits [mosses] predicted protein [Physcomitrella patens subsp. patens] g|
  
```

**Organism Report**

```

Ricinus communis [eudicots] taxid 3988
ref|XP_002516248.1| serine-threonine protein kinase, plant-type... 88 4e-16
gb|EF44468.1| serine-threonine protein kinase, plant-type... 88 4e-16
ref|XP_002516247.1| serine-threonine protein kinase, plant-type... 88 5e-16
gb|EF44467.1| serine-threonine protein kinase, plant-type... 88 5e-16
ref|XP_002514297.1| ATP binding protein, putative [Ricinus... 81 5e-14
gb|EF428751.1| ATP binding protein, putative [Ricinus comm... 81 5e-14
ref|XP_002513381.1| Serine/threonine-protein kinase PRR1, ... 80 1e-13
gb|EF42998.1| Serine/threonine-protein kinase PRR1, putat... 80 1e-13
ref|XP_002516045.1| serine-threonine protein kinase, plant... 79 2e-13
gb|EF44465.1| serine-threonine protein kinase, plant-type... 79 2e-13
  
```

Go back to the BLASTX page

Click in "Algorithm parameters"

Repeat the search of GR421296 after changing the "Expect threshold to  $10^{-14}$  (1e-14) and run BLAST

Algorithm parameters

Note: Parameter values that differ from the default are highlighted in yellow and marked with a sign

General Parameters

Max target sequences: 100

Expect threshold: 1e-15

Word size: 3

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Filters and Masking

Filter:  Low complexity regions

Mask:  Mask for lookup table only  Mask lower case letters

BLAST Search database Non-redundant protein sequences (nr) using Blastx (search protein databases using a translated nucleotide query)

Show results in a new window

Observe the reduction in the number of hits

Query ID: gi|242128777|gb|GR421296.1

Description: MALC05\_B10 MALC Manihot esculenta cDNA clone MALC05\_B10.

Molecule type: rna

Query Length: 596

Database Name: nr

Description: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects

Program: BLASTX 2.2.24+ >Citation

Other reports: [Search Summary](#) [Taxonomy reports](#)

Graphic Summary

Distribution of 6 Blast Hits on the Query Sequence

Color key for alignment scores

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Pink
>=200	Red

Descriptions

Legend for links to other resources: [U](#) UniGene [G](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [X](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
XP_002516048.1	serine-threonine protein kinase, plant-type, putative [Ricin...	88.6	88.6	56%	4e-16	46%	<a href="#">G</a>
XP_002516047.1	serine-threonine protein kinase, plant-type, putative [Ricin...	88.2	88.2	78%	5e-16	37%	<a href="#">G</a>
NP_199788.1	leucine-rich repeat transmembrane protein kinase, putative [Arabidop...	87.0	87.0	57%	1e-15	41%	<a href="#">U</a> <a href="#">G</a>
XP_002864015.1	hypothetical protein ARALYDRAFT_495037 [Arabidopsis lyrata subsp. l...	85.9	85.9	74%	3e-15	36%	<a href="#">G</a>
BAE46451.1	putative receptor protein kinase PERK1 [Glycine max] >gb ACMB9502...	85.5	85.5	51%	3e-15	46%	<a href="#">G</a>
XP_002864016.1	predicted protein [Arabidopsis lyrata subsp. lyrata] >gb EPH40275.1]	84.7	84.7	53%	6e-15	43%	<a href="#">G</a>

Click on the light blue link "U" (for Unigene) in the far right column of the descriptions

▼ Descriptions

Legend for links to other resources: [U](#) UniGene [E](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer [P](#) PubChem BioAssay

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
<a href="#">XP_002516048.1</a>	serine-threonine protein kinase, plant-type, putative [Ricinus communis]	88.6	88.6	56%	4e-16	46%	<a href="#">G</a>
<a href="#">XP_002516047.1</a>	serine-threonine protein kinase, plant-type, putative [Ricinus communis]	88.2	88.2	78%	5e-16	37%	<a href="#">G</a>
<a href="#">NP_199788.1</a>	leucine-rich repeat transmembrane protein kinase, putative [Arabidopsis]	87.0	87.0	57%	1e-15	41%	<a href="#">U</a> <a href="#">G</a>
<a href="#">XP_002864015.1</a>	hypothetical protein ARALYDRAFT_495037 [Arabidopsis lyrata subsp. lyr]	85.9	85.9	74%	3e-15	36%	<a href="#">G</a>
<a href="#">BAE46451.1</a>	putative receptor protein kinase PERK1 [Glycine max] >gb ACM89502.1	85.5	85.5	51%	3e-15	46%	<a href="#">G</a>
<a href="#">XP_002864016.1</a>	predicted protein [Arabidopsis lyrata subsp. lyrata] >gb EFH40275.1  pr	84.7	84.7	53%	6e-15	46%	<a href="#">G</a>

You will be taken to NCBI's Unigene database. Click on the Arabidopsis Unigene record

NCBI UniGene ORGANIZED VIEW OF THE TRANSCRIPTOME

All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search UniGene for [ ] Go Clear

Limits Preview/Index History Clipboard Details

Display Summary Show 20 Sort By Send to

All: 1 Fungi: 0 Insects: 0 Mammals: 0 Plants: 1

1: [Leucine-rich repeat transmembrane protein kinase, putative](#)  
AT5G49770, Arabidopsis thaliana  
At.43262: 5 sequences.

Recent activity  
Turn Off Clear  
UniGene Links for Protein... (1) UniGene  
See more...

See the corresponding set of ESTs

NCBI UGID:27527 UniGene At.43262 Arabidopsis thaliana (thale cress) AT5G49770 Links

**Leucine-rich repeat transmembrane protein kinase, putative (AT5G49770)**

**SELECTED PROTEIN SIMILARITIES**  
Comparison of sequences in UniGene with selected protein reference sequences. The alignments can suggest function of a gene.

Reference Protein	Species	Id(%)	Len(aa)
<a href="#">NP_199788.1</a> leucine-rich repeat transmembrane protein kinase, putative	A. thaliana	100.0	945

**GENE EXPRESSION**  
Tissues and development stages from this gene's sequences survey gene expression. Links to other NCBI expression resources.

cDNA Sources: mixed; unspecified tissue

**MAPPING POSITION**  
Genomic location specified by transcript mapping, radiation hybrid mapping, genetic mapping or cytogenetic mapping.

Chromosome: 5

**SEQUENCES**  
Sequences representing this gene's mRNAs, ESTs, and gene predictions supported by transcribed sequences.

**mRNA sequences (3)**

<a href="#">NM_124355.2</a>	Arabidopsis thaliana leucine-rich repeat transmembrane protein kinase, putative (AT5G49770) mRNA, complete cds	PA
<a href="#">BX832089.1</a>	Arabidopsis thaliana Full-length cDNA Complete sequence from clone GSLTPGH50ZG05 of Hormone Treated Callus of strain col-0 of Arabidopsis thaliana (thale cress)	PA
<a href="#">FJ708798.1</a>	Arabidopsis thaliana leucine-rich repeat receptor-like protein kinase (LRR-RLK) mRNA, partial cds	P

**EST sequences (2)**

<a href="#">N38191.1</a>	Clone 220L1T7	mixed	P
<a href="#">AJ994789.1</a>	Clone 701499629	unspecified tissue	P

Go back to the BLASTX page and click on "blastn" tab at the top left. The same FASTA sequence or accession number GR421296 should be already in the input box. Select "Nucleotide collection" (nr) under "Database"

Under "Program selection" check "Optimize for highly similar sequences (megablast)". Click "Algorithm parameters" and see the Expect threshold

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 **Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign**

General Parameters

Max target sequences: 100

Short queries:  Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 28

Max matches in a query range: 0

Check out Discontiguous megablast by clicking its button. What parameters does this program change?

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 Discontiguous megablast (Optimize for more dissimilar sequences)

Show results in a new window

Algorithm parameters **Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign**

General Parameters

Max target sequences: 100

Short queries:  Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 11

Max matches in a query range: 0

Scoring Parameters

Match/Mismatch Scores: 2,-3

Gap Costs: Existence: 5 Extension: 2

Filters and Masking

Filter:  Low complexity regions

Mask:  Mask for lookup table only

Discontiguous Word Options

Template length: 18

Template type: Coding

Check "Somewhat similar sequences (blastn)". What parameters does this program change?

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 Blastn (Optimize for somewhat similar sequences)

Show results in a new window

Algorithm parameters Note: Parameter values that differ from the default are highlighted in yellow and marked with + sign

General Parameters

Max target sequences  Select the maximum number of aligned sequences to display

Short queries  Automatically adjust parameters for short input sequences

Expect threshold

Word size

Max matches in a query range

Scoring Parameters

Match/Mismatch Scores

Gap Costs Existence: 5 Extension: 2

Select Megablast again and click the BLAST button. Do you get any hits?

BLAST Basic Local Alignment Search Tool

My NCBI

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

NCBI/ BLAST/ blastn suite/ Formatting Results - 7ZX4V9WR01N

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#)

gb|GR421296| (596 letters)

<b>Query ID</b>	gi 242128777 gb GR421296.1	<b>Database Name</b>	nr
<b>Description</b>	MALC05_B10 MALC Manihot esculenta cDNA clone MALC05_B10.	<b>Description</b>	All GenBank+EMBL+DDJ+PDB sequences (but no EST, STS, GSS, environmental samples or phase 0, 1 or 2 HTGS sequences)
<b>Molecule type</b>	ma	<b>Program</b>	BLASTN 2.2.24+ <a href="#">Citation</a>
<b>Query Length</b>	596		

No significant similarity found. For reasons why, [click here](#)

Other reports: [Search Summary](#)

Copyright | [Disclaimer](#) | [Privacy](#) | [Accessibility](#) | [Contact](#) | [Send feedback](#)

NCBI | NLM | NIH | DGEIS

Go back to the NCBI BLAST front page and select "Primer-BLAST" under "Specialized BLAST".

**BLAST** Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help My NCBI

NCBI BLAST Home

BLAST finds regions of similarity between biological sequences. [more...](#)

**Aligning Multiple Protein Sequences? Try the COBALT Multiple Alignment Tool. (Go)**

### BLAST Assembled RefSeq Genomes

Choose a species genome to search, or [list all genomic BLAST databases.](#)

- Human
- Mouse
- Rat
- Arabidopsis thaliana*
- Oryza sativa*
- Bos taurus*
- Danio rerio*
- Drosophila melanogaster*
- Gallus gallus*
- Pan troglodytes*
- Microbes
- Apis mellifera*

### Basic BLAST

Choose a BLAST program to run.

- nucleotide\_blast** Search a nucleotide database using a nucleotide query  
Algorithms: blastn, megablast, discontinuous megablast
- protein\_blast** Search protein database using a protein query  
Algorithms: blastp, psi-blast, phi-blast
- tblastn** Search protein database using a translated nucleotide query
- tblastx** Search translated nucleotide database using a protein query
- tblastx** Search translated nucleotide database using a translated nucleotide query

### Specialized BLAST

Choose a type of specialized search (or database name in parenthesis)

- Make specific primers with **Primer-BLAST**
- Search [trace archives](#)
- Find [conserved domains](#) in your sequence (cds)
- Find sequences with similar [conserved domain architecture](#) (cdart)
- Search sequences that have [gene expression profiles](#) (GEO)
- Search [immunoglobulins](#) (igBLAST)
- Search for [SNPs](#) (snp)
- Screen sequence for [vector contamination](#) (vecscreen)
- [Align two \(or more\) sequences](#) using BLAST (b2seq)
- Search [protein or nucleotide targets](#) in PubChem BioAssay
- Search [SRA transcript and genomic libraries](#)
- Constraint Based Protein [Multiple Alignment Tool](#)
- Needleman-Wunsch [Global Sequence Alignment Tool](#)

Powered by: 1/26/2010 12:00:00 PM 1/26/2010 12:00:00 PM

**Primer-BLAST** *A tool for finding specific primers*

NCBI Primer-BLAST: Finding primers specific to your PCR template (using Primer3 and BLAST). [more...](#) [Tips for finding specific primers](#)

[Reset page](#) [Save search parameters](#) [Retrieve previous results](#)

#### PCR Template

Enter accession, gi, or FASTA sequence (A refseq record is preferred)  [Clear](#)

Range: Forward primer  From  To  [Clear](#)

Reverse primer

Or, upload FASTA file  [Browse...](#)

#### Primer Parameters

Use my own forward primer (5'→3' on plus strand)  [Clear](#)

Use my own reverse primer (5'→3' on minus strand)  [Clear](#)

PCR product size: Min 70 Max 1000

# of primers to return: 10

Primer melting temperatures (T<sub>m</sub>): Min 57.0 Opt 60.0 Max 63.0 Max T<sub>m</sub> difference 3 [Clear](#)

#### Exon/intron selection

A refseq mRNA sequence as PCR template input is required for options in the section [Clear](#)

Exon junction span: No preference [Clear](#)

Exon junction match: Exon at 5' side 7 Exon at 3' side 4

Minimal number of bases that must anneal to exons at the 5' or 3' side of the junction [Clear](#)

Primer must be separated by at least one intron on the corresponding genomic DNA [Clear](#)

Intron inclusion: Intron length range: Min 1000 Max 1000000 [Clear](#)

Click "Advanced parameters" and see that this program BLASTN with different parameters

**Advanced parameters**

**Primer Pair Specificity Checking Parameters**

Blast max number sequences: 50000

Blast expect (E) value: 30000

Blast word size: 7

Max primer pairs to screen: 1000

**Primer Parameters**

	Min	Opt	Max
PCR Product Tm			
Primer Size	15	20	25
Primer GC content (%)	20.0		80.0
GC clamp	0		
Max Poly-X	5		
Max self complementarity:	8.00		
Max 3' end complementarity:	3.00		
SNP handling	<input type="checkbox"/> Primer binding site may not contain known SNP		
Repeat filter	Automatic		
Low complexity filter	<input checked="" type="checkbox"/> Avoid low complexity region for primer selection		
Concentration of monovalent cations	50.0		
Concentration of divalent cations	0.0		
Concentration of dNTPs	0.0		
Salt correction formula	SantaLucia 1998		
Table of thermodynamic parameters	SantaLucia 1998		

Go back to the NCBI BLAST page and click on "Protein blast". Check PSI-BLAST under "Algorithm" in the next page

NCBI BLAST/ blastp suite

blastn blastp blastx tblastn tblastx

BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

**Enter Query Sequence**

Enter accession number, gi, or FASTA sequence

From  To

Or, upload file

Job Title

Align two or more sequences

**Choose Search Set**

Database:

Organism:   Exclude

Exclude:  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query:

**Program Selection**

Algorithm:  blastp (protein-protein BLAST)  PSI-BLAST (Position-Specific Iterated BLAST)  PHI-BLAST (Pattern Hit Initiated BLAST)

**BLAST** Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

Show results in a new window

Paste the hypothetical protein XP\_002536085 from *Ricinus communis* in the query box. See the PSI-BLAST threshold (0.005). Check "Show results in a new window and run BLAST"

**Algorithm parameters**

**General Parameters**

Max target sequences: 500  
 Short queries:  Automatically adjust parameters for short input sequences  
 Expect threshold: 10  
 Word size: 3  
 Max matches in a query range: 0

**Scoring Parameters**

Matrix: BLOSUM62  
 Gap Costs: Existence: 11 Extension: 1  
 Compositional adjustments: Conditional compositional score matrix adjustment

**Filters and Masking**

Filter:  Low complexity regions  
 Mask:  Mask for lookup table only  
 Mask lower case letters

**PSI/PHI BLAST**

Upload PSSM:    
 Optional  
 PSI-BLAST Threshold: 0.005  
 Pseudocount: 0

See the results above and below the threshold. See the alignments (the first is the match of the query to itself)

**Show Conserved Domains**

No putative conserved domains have been detected

Distribution of 9 Blast Hits on the Query Sequence

Color key for alignment scores

**Descriptions**

Legend for links to other resources: UniGene, GEO, Gene, Structure, Map Viewer, PubChem BioAssay

NEW - alignment score below the threshold on the previous iteration  
 - alignment was checked on the previous iteration

Run PSI-Blast iteration 2 with max 500

**Sequences producing significant alignments with E-value BETTER than threshold**

Accession	Description	Max score	Total score	Query coverage	E-value	Links
NEW XP_002336085.1	hypothetical protein R10M_1994310 [Ricinus communis] >gi223520948 gb EF2829	101.0	101.0	100%	0.0	<a href="#">C</a>
NEW XP_002316884.1	predicted protein [Populus trichocarpa] >gb EEE97496.1  predicted protein [Populus t	251	244	69%	3e-62	<a href="#">U</a> <a href="#">C</a>
NEW XP_002285182.1	PREDICTED: hypothetical protein [Vitis vinifera] >emb CR20265.3  unnamed protein	100	100	65%	7e-19	<a href="#">U</a> <a href="#">C</a>
NEW CAH80069.1	hypothetical protein [Vitis vinifera]	85.9	85.9	53%	1e-14	<a href="#">C</a>
NEW XP_002329117.1	predicted protein [Populus trichocarpa] >gb EEF06917.1  predicted protein [Populus t	64.7	64.7	29%	3e-08	<a href="#">C</a>

Run PSI-Blast iteration 2 with max 500

**Sequences with E-value WORSE than threshold**

Accession	Description	Max score	Total score	Query coverage	E-value	Links
XP_002439218.1	hypothetical protein SORBIDRAFT_02g000740 [Sorghum bicolor] >gb EE995740.1  hy	55.8	46.6	18%	0.008	<a href="#">U</a> <a href="#">C</a>
XP_002392485.1	conserved hypothetical protein [Physiphora infestans T30-4] >gb EY168772.1  cons	50.8	40.8	49%	0.44	<a href="#">C</a>
NP_001141153.1	hypothetical protein LOC100273239 [Zea mays] >gb ACF85556.1  unknown [Zea may	50.0	40.0	4%	0.76	<a href="#">U</a> <a href="#">C</a>
ACG52344.1	hypothetical protein [Zea mays]	40.0	40.0	4%	0.80	<a href="#">C</a>

```

>ref|XP_002316884.1|UG predicted protein [Populus trichocarpa]
gb|EEE97496.1|G predicted protein [Populus trichocarpa]
Length=526

GENE ID: 7460806 POPTRDRAFT_660905 | hypothetical protein [Populus trichocarpa]
(10 or fewer PubMed links)

Score = 244 bits (622), Expect = 3e-62, Method: Compositional matrix adjust.
Identities = 168/364 (46%), Positives = 227/364 (62%), Gaps = 25/364 (6%)

Query 115 TCWSSSSGSEESDNESRGRPERREKNTRKVNKVGATRRRYSRSCSSCSGHDSWSEK 174
          TC S S +ES+ E +GRPERR+ RK +N++ GA RRRYS SCSSCS HD S +
Sbjct 32 TCQSQS---DESEYERCKGRPERRDDEKRKSENIRSGAKRRRYSRSCSSCSRHDSSDF 88

Query 175 V-----TDGNVSKRLKSIITLPNEDEEGRELNRECKEEMICDHDYPPSSRNSDNDG 229
          + T N SKRL+SII LP ED E REL++D+ KEE+ DHDYPPSSRNSDNDG N
Sbjct 89 LMSNIMTGENTS KRLRSIIILPGEDSEVRELDKDKHKEEITYDHDYPPSSRNSDNDGLN 148

Query 230 KVSAYEPCVESEKRSIEIEKKEDASSFNIKTKLPSSYKNGDDHHLGSRFASDS--VG 287
          + ++R IE EK+EDA++ N K +L S K G+ H ++ D VG
Sbjct 149 -----NMEERP IEDEKREDAASNSKAIELTESNKVGEQHTRNKPGYDVGRVG 197

Query 288 KNDALEEKTSNTSEVVG SANANDLESILREKALENLRFR---GEIQTNMKSTVSKDEN 344
          ND +E+ + +V +AN +DLE++LR+KALENL+ FR G QTN KS V QKD+
Sbjct 198 TNDTKKEQNDVSGVIVNTANVDDLETVLRQKALENLKTFRSLGGFQTNAKSAVIQKDKR 257

Query 345 DVTLSLSTTKGEVNEIASLEDDGTRGSKGDSAHVQKEENIFDTENGNDVSVAMD NAY 404
          D T +S + E+ + + GTR + DSAH E+ I D G++S SA +N +
Sbjct 258 DGTAQSPFVMPQLGQTKIPKVVGTRMAGEDSAHSSLNEK--IPDGGICSESCSAKNNVH 316

Query 405 IPNQMVAGREKANMNLSPVINKPTLVTSALKQALS NATSTAMELPASKESDTAKLVIGS 464
          P+Q+A+ GREK + S NKP L+TSA ++ALSN T+T E PAS+E++ KL G+
Sbjct 317 PPDQVAIPGREKVTFASSSKNKPRLITSASRKALSNVTTTLKETPASRETNPQLASGT 376

Query 465 SKGK 468
          S G+
Sbjct 377 SIGR 380

```

Click the "GO" button by "Run PSI-Blast iteration 2 with max 500"

See the new hits suggesting potential function

Distribution of 30 Blast Hits on the Query Sequence

Mouse-over to show details and scores, click to show alignments

**Color key for alignment scores**

- <40
- 40-50
- 50-100
- 100-200
- >=200

Query 0 100 200 300 400 500

▼ Descriptions

Legend for links to other resources: [U](#) UniGene [G](#) GEO [C](#) Gene [S](#) Structure [M](#) Map Viewer [B](#) PubChem BioAssay

NEW - alignment score below the threshold on the previous iteration  
 ● - alignment was checked on the previous iteration

Run PSI-Blast iteration 3 with max

▼ Sequences producing significant alignments with E-value BETTER than threshold

Accession	Description	Max score	Total score	Query coverage	E-value	Links
<a href="#">XP_002326085.1</a>	hypothetical protein RCOM_1994310 [Ricinus communis] >gi 223520948 gb EEF26296	528	528	100%	9e-148	<a href="#">G</a>
<a href="#">XP_002316884.1</a>	predicted protein [Populus trichocarpa] >gb EEE97496.1  predicted protein [Populus tr	379	379	69%	4e-103	<a href="#">U</a> <a href="#">G</a>
<a href="#">XP_002285142.1</a>	PREDICTED: hypothetical protein [Vitis vinifera] >emb CB20265.3  unnamed protein	268	268	87%	2e-69	<a href="#">U</a> <a href="#">G</a>
<a href="#">CA580069.1</a>	hypothetical protein [Vitis vinifera]	243	243	87%	4e-62	
<a href="#">XP_002329117.1</a>	predicted protein [Populus trichocarpa] >gb EEF06917.1  predicted protein [Populus tr	180	180	29%	4e-43	<a href="#">G</a>
NEW <a href="#">XP_001191416.1</a>	PREDICTED: similar to WD repeat domain 60 [Strongylocentrotus purpuratus]	54.5	54.5	55%	3e-05	<a href="#">U</a> <a href="#">G</a>
NEW <a href="#">XP_785240.2</a>	PREDICTED: similar to WD repeat domain 60, partial [Strongylocentrotus purpuratus]	54.5	54.5	55%	3e-05	<a href="#">U</a> <a href="#">G</a>
NEW <a href="#">BAK28322.1</a>	tight junction protein ZO-1 [Hydra vulgaris]	50.3	50.3	61%	6e-04	

Look at the alignments to estimate if they are reliable

```
>ref|XP_001191416.1|UG PREDICTED: similar to WD repeat domain 60 [Strongylocentrotus
purpuratus]
Length=1077

GENE ID: 580071 LOC580071 | similar to WD repeat domain 60
[Strongylocentrotus purpuratus]

Sort alignments for this subject sequence by:
E value Score Percent identity
Query start position Subject start position

Score = 54.5 bits (129), Expect = 3e-05, Method: Composition-based stats.
Identities = 53/293 (18%), Positives = 109/293 (37%), Gaps = 12/293 (4%)

Query 25  SYSSDDSRSSVSVFSSSEDSYKHR-KSRSRTHNVKGSRRRVRSRSSSSSEESGYLKKHQRP 83
          S D R +D +HR K R + + R R ++ + +R
Sbjct 28  SRHKDKDRDKEREKPRDKDRERHRDKDRDKSHRDKHEHRKRDKDRDQKEKDKDRERRK 87

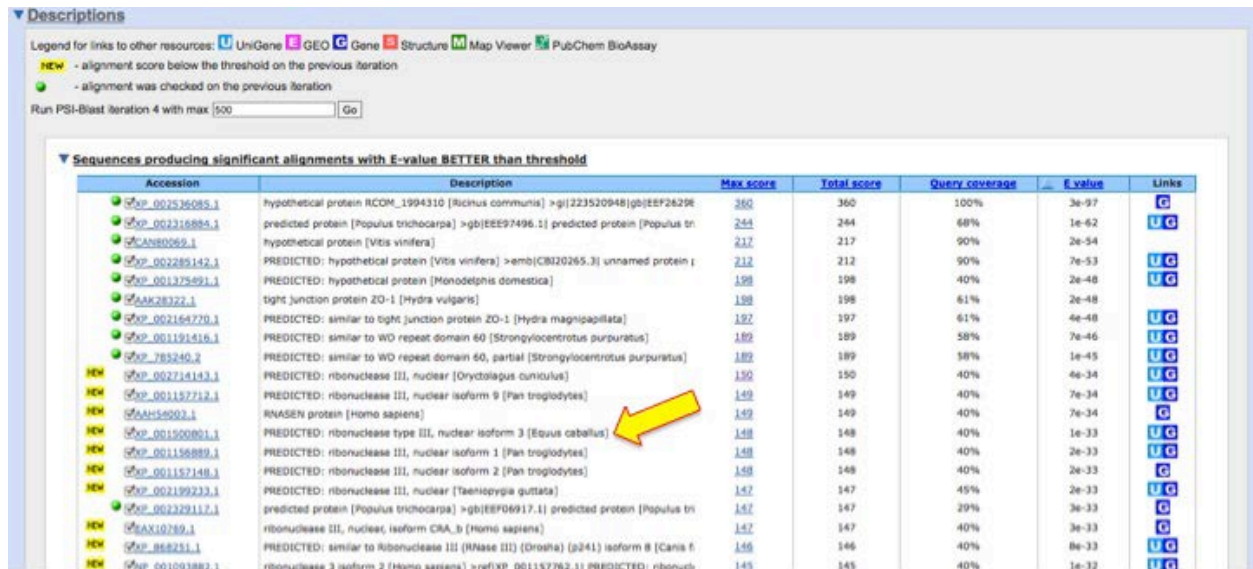
Query 84  RRKNDSEVRKETYRSKKKKAREASSSSRSCSTCWSSSSGSEESDNESSRG---RPERREK 140
          R+ +S K+ R K ++ + R +ES E RG R + REK
Sbjct 88  DREGSSKEKDRERRKDRKDKDRDKDRDKDRDKDRDKDRDKDRDKDRDKDRDKDRDKDR 147

Query 141 NTRKVNKVGKATRRYRSRSCSSCGHDWSKEKVTGDNVSKRLKSIITLPNEDEEGREL 200
          + + K + R S + + +++ D +R K + + ++
Sbjct 148 EKDRDREKERSHRDNKESAERTDSGKKERHRDR--DEKKREKESGSRDGDGKSSSTKDG 205

Query 201 NRDECKEEMICDHDYPSRSRNSNDGGNKSVSAYEPCVESEKRSIEIEKKEDASSFNI 260
          +RD+ +EE D + +SND + + +++++ + ++
Sbjct 206 DRDKRREKRRDKDRERKKEKESNDSRERRKHRDKGSDARRKREKSGERKRRKEHKE 265

Query 261 KTTKLPSSYKNGDDH-HLGSRFASDSVGK-----NDALEKTSNTSEVVGSA 307
          + + S +KNG D H R D VG+ +D E++ +++ G +
Sbjct 266 EKGRSKSEHKNGQDERHEKREDEDDVGEELKVSQKEQQMEESNQDDGYED 318
```

Run a new PSI-BLAST iteration to see more hits. Until when can you continue performing PSI-BLAST iterations?



Legend for links to other resources: UniGene GEO Gene Structure Map Viewer PubChem BioAssay

NEW - alignment score below the threshold on the previous iteration  
 ● - alignment was checked on the previous iteration

Run PSI-Blast iteration 4 with max [500] Go

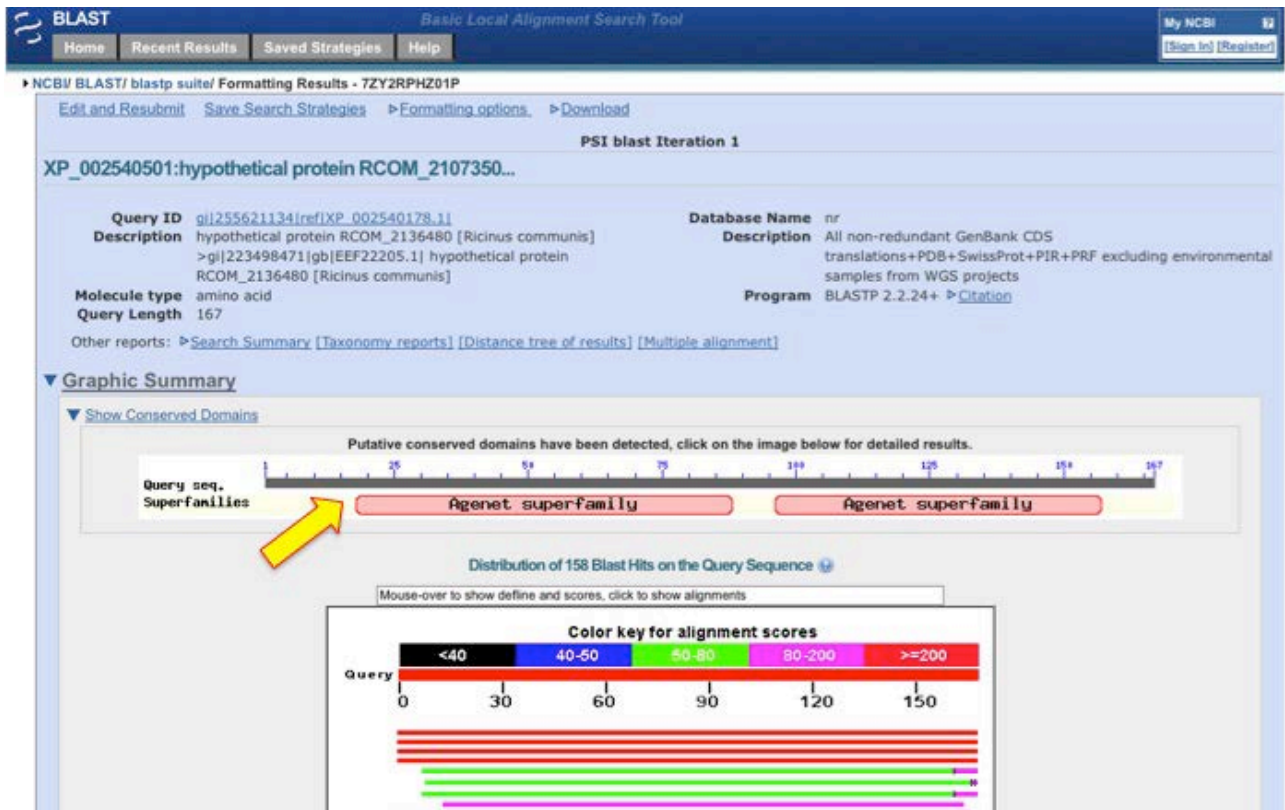
Sequences producing significant alignments with E-value BETTER than threshold

Accession	Description	Max score	Total score	Query coverage	E-value	Links
XP_002540178.1	hypothetical protein RCOM_1994310 [Ricinus communis] >gi 223520948 gb EF26296	260	360	100%	3e-97	UG
XP_002316885.1	predicted protein [Populus trichocarpa] >gi EEE97496.1  predicted protein [Populus tr	244	244	68%	1e-62	UG
CAAN00069.1	hypothetical protein [Vitis vinifera]	217	217	90%	2e-54	UG
XP_002289142.1	PREDICTED: hypothetical protein [Vitis vinifera] >emb CB120265.3  unnamed protein ;	212	212	90%	7e-53	UG
XP_001275491.1	PREDICTED: hypothetical protein [Monodelphis domestica]	198	198	40%	2e-48	UG
AAK28327.1	tight junction protein ZO-1 [Hydra vulgaris]	198	198	61%	2e-48	UG
XP_002164720.1	PREDICTED: similar to tight junction protein ZO-1 [Hydra magnipapillata]	197	197	61%	4e-48	UG
XP_001191416.1	PREDICTED: similar to WD repeat domain 60 [Strongylocentrotus purpuratus]	189	189	58%	7e-46	UG
XP_785240.2	PREDICTED: similar to WD repeat domain 60, partial [Strongylocentrotus purpuratus]	189	189	58%	1e-45	UG
XP_002714143.1	PREDICTED: ribonuclease III, nuclear [Dryobolus curvulus]	150	150	40%	4e-34	UG
XP_001157712.1	PREDICTED: ribonuclease III, nuclear isoform 9 [Pan troglodytes]	149	149	40%	7e-34	UG
AAH58003.1	RNA5EN protein [Homo sapiens]	149	149	40%	7e-34	UG
XP_001500801.1	PREDICTED: ribonuclease type III, nuclear isoform 3 [Equus caballus]	148	148	40%	1e-33	UG
XP_001158889.1	PREDICTED: ribonuclease III, nuclear isoform 1 [Pan troglodytes]	148	148	40%	2e-33	UG
XP_001157148.1	PREDICTED: ribonuclease III, nuclear isoform 2 [Pan troglodytes]	148	148	40%	2e-33	UG
XP_002199233.1	PREDICTED: ribonuclease III, nuclear [Taeniopygia guttata]	147	147	45%	2e-33	UG
XP_002329117.1	predicted protein [Populus trichocarpa] >gi EF906917.1  predicted protein [Populus tri	147	147	29%	3e-33	UG
CAK10789.1	ribonuclease III, nuclear, isoform CIA_b [Homo sapiens]	147	147	40%	3e-33	UG
XP_888251.1	PREDICTED: similar to Ribonuclease III (RNase III) [Drosophila] (g241) isoform B [Carni f	146	146	40%	8e-33	UG
XP_001093882.1	ribonuclease 3 isoform 2 [Homo sapiens] >ref XP_001157762.1  PREDICTED: rbonuc	145	145	40%	1e-32	UG

Go Back to the protein blast page

Check BLASTP under "Algorithm" and paste the hypothetical protein XP\_002540178 in the query box

See the Conserved Domain Database matches



Go back to the NCBI BLAST front page and select "nucleotide blast". Check the box "Align two or more sequences"

Paste the mRNA sequence accession number GU111719 in the top box and the genomic sequence GU111714 in the bottom box.

NCBI BLAST/ blastn suite

blastn blastx blastb blasti blasto

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

Basal page Bookmark

Enter Query Sequence

Enter accession number, gi, or FASTA sequence

Query subrange

From

To

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Enter Subject Sequence

Enter accession number, gi, or FASTA sequence

Subject subrange

From

To

Or, upload file

Program Selection

Optimize for

Highly similar sequences (megablast)

More dissimilar sequences (discontiguous megablast)

Somewhat similar sequences (blastn)

Choose a BLAST algorithm

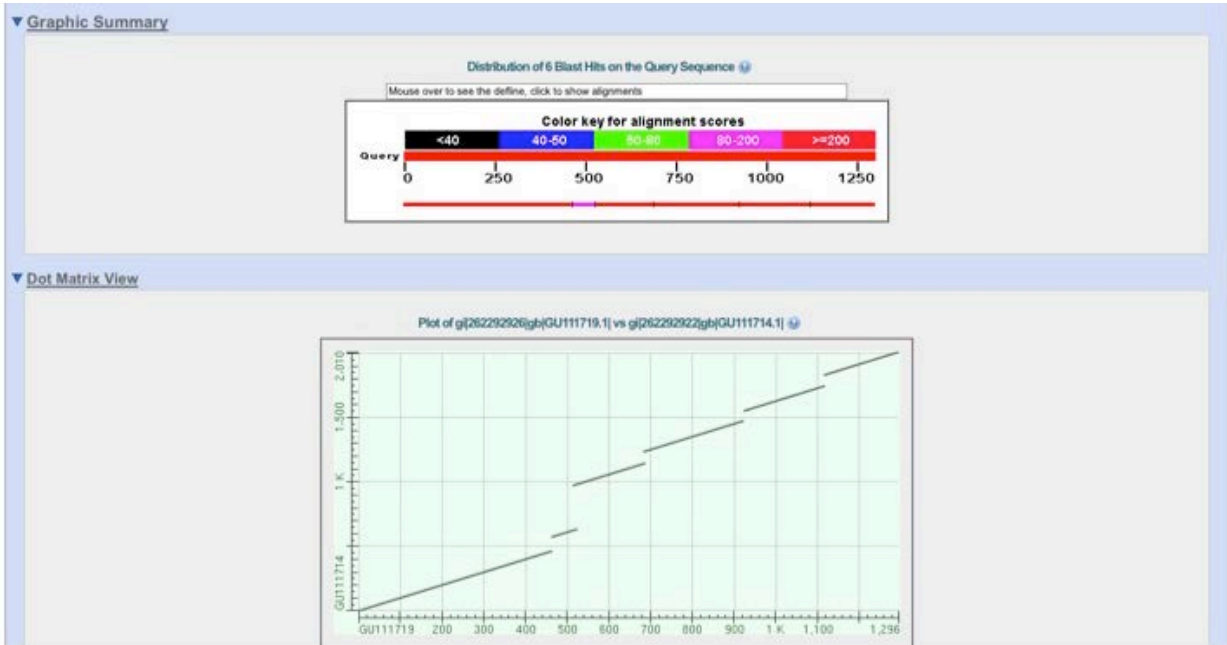
BLAST

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

Show results in a new window

Algorithm parameters

Run the BLAST alignment  
Expand the dot matrix view



## Exercises:

1. Run a nucleotide BLAST search of GR421296 (that yielded no hit using Megablast in the worked example) against the 'nr' database using Discontiguous megablast, and Blastn. Do you get any reliable hits?
2. You have this pine (*Pinus pinaster*) EST (FN868597) and you want to know if there is a putative homologue that is expressed in *Brassica rapa*. What BLAST program will you use? What database will you use? How many *B. rapa* ESTs match the pine EST? Could all these ESTs correspond to mRNAs coded by the same gene (hint: use your muscle).
3. How does the BLAST alignment of 2 sequences compares to Spidey when you align GU111719 and GU111714? (look particularly at exon 2).
4. Use the nucleotide sequence GU191335.1 as query to perform two searches: BLASTX and BLASTN versus the 'nr' database. As the sequence comes from rubber tree (*H. brasiliensis*), in both searches the first hits will be from the same species. Which is the second species hit by BLASTN? Which is the second species hit by BLASTX? Which one is more likely to be a homologue of GU191335.1?
5. BLAST the cassava EST DB927724 against the "est\_others" database. Repeat the search unchecking the filter "low complexity regions" under "Algorithm parameters". What is the difference in the results? What does it mean? (hint: click the blue question mark button next to the "low complexity regions" checkbox; click "more" for additional info).