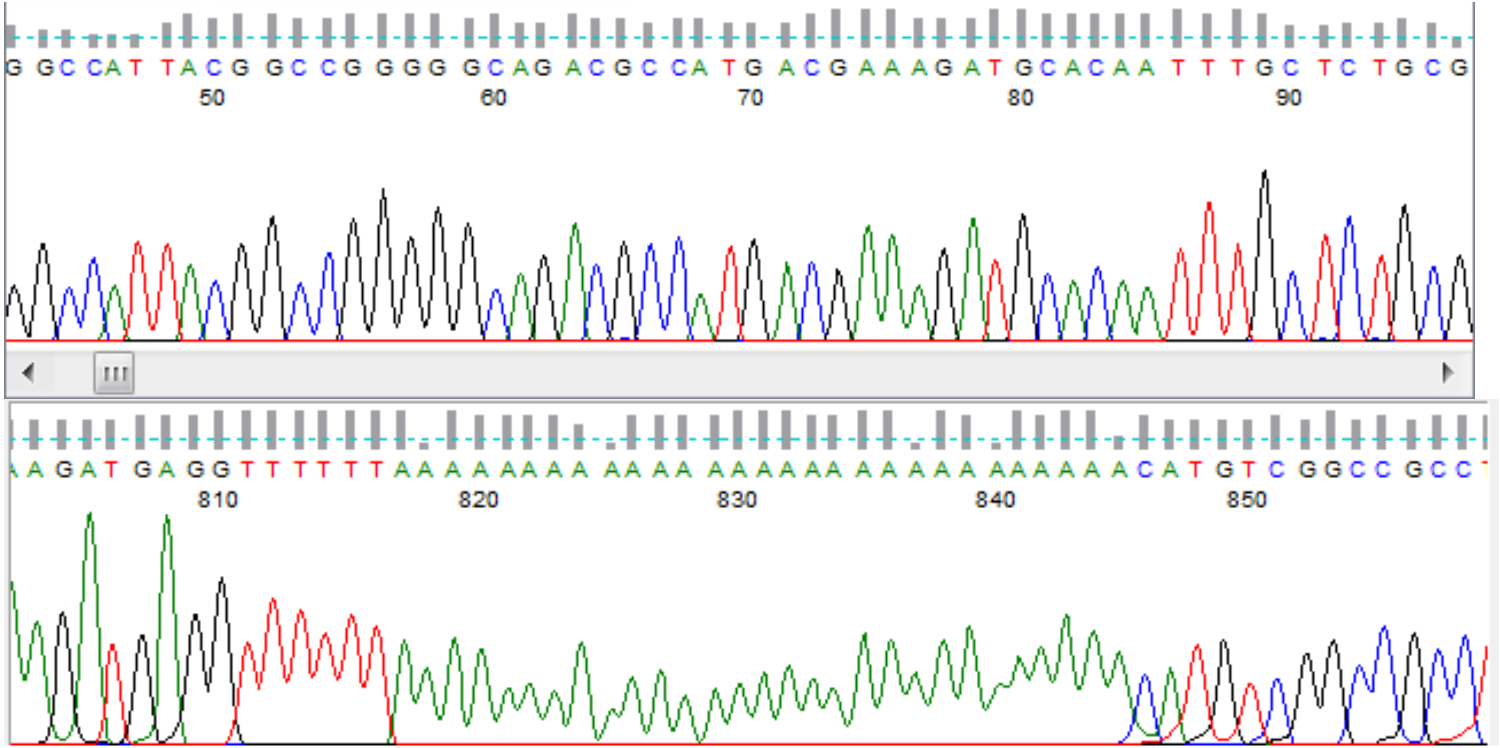
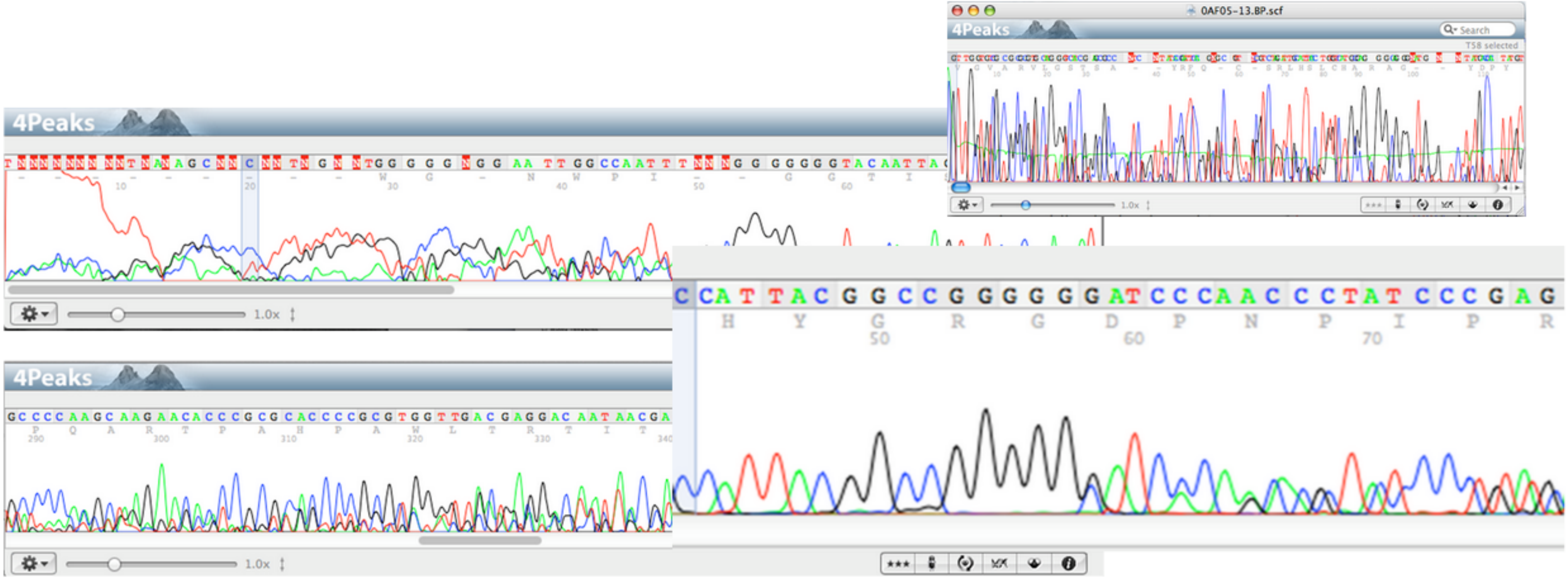
DSAP PART 2: BLAST

Review

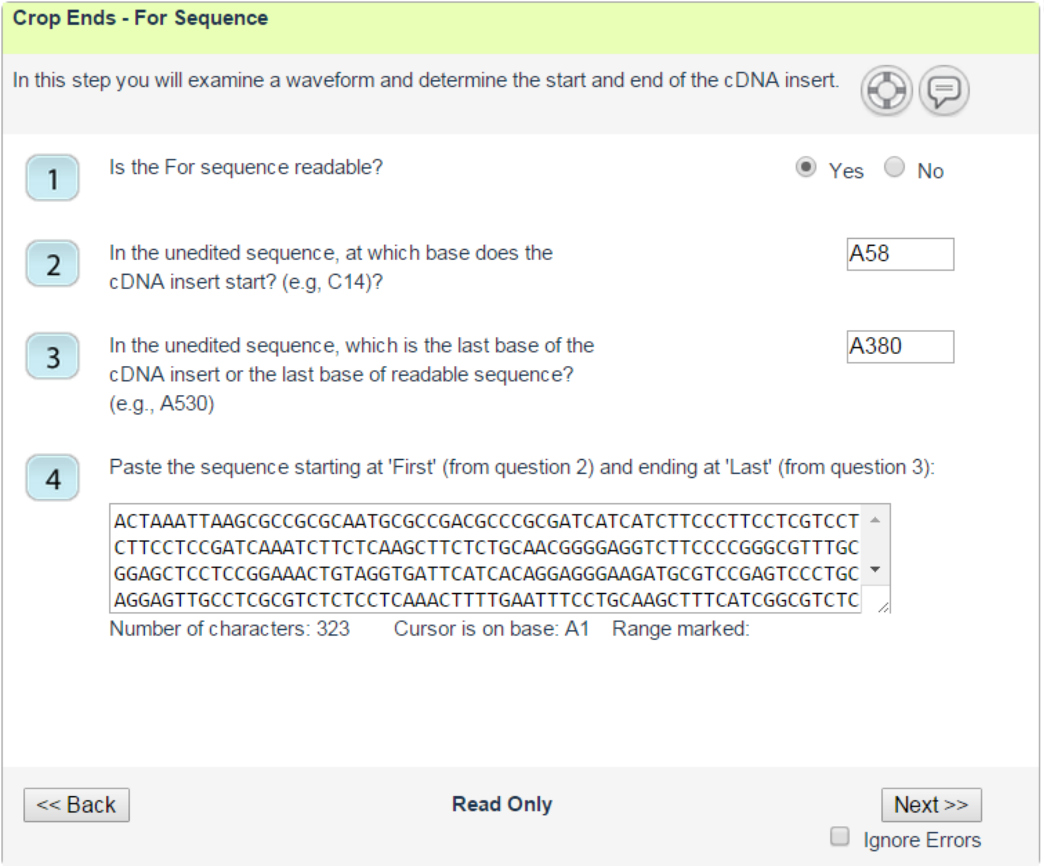

G GCCAT TACG GCCG GGG GCAG ACGC CA TG ACGAAAGAT GCACAA T T T GC TC TGCG
50
60
70
80
90
AAGAT GAGGTT TTTT TAA AAAAAA AAAA AAAAAA AAAAA AAAAAAACAT GTC GGCC GCC
810
820
830
840
850

What You Know So Far...

!?!?!?!?!?!!??!!??!!?



Review



Crop Ends - For Sequence

In this step you will examine a waveform and determine the start and end of the cDNA insert.

1 Is the For sequence readable? ● Yes ○ No

2 In the unedited sequence, at which base does the cDNA insert start? (e.g, C14)? A58

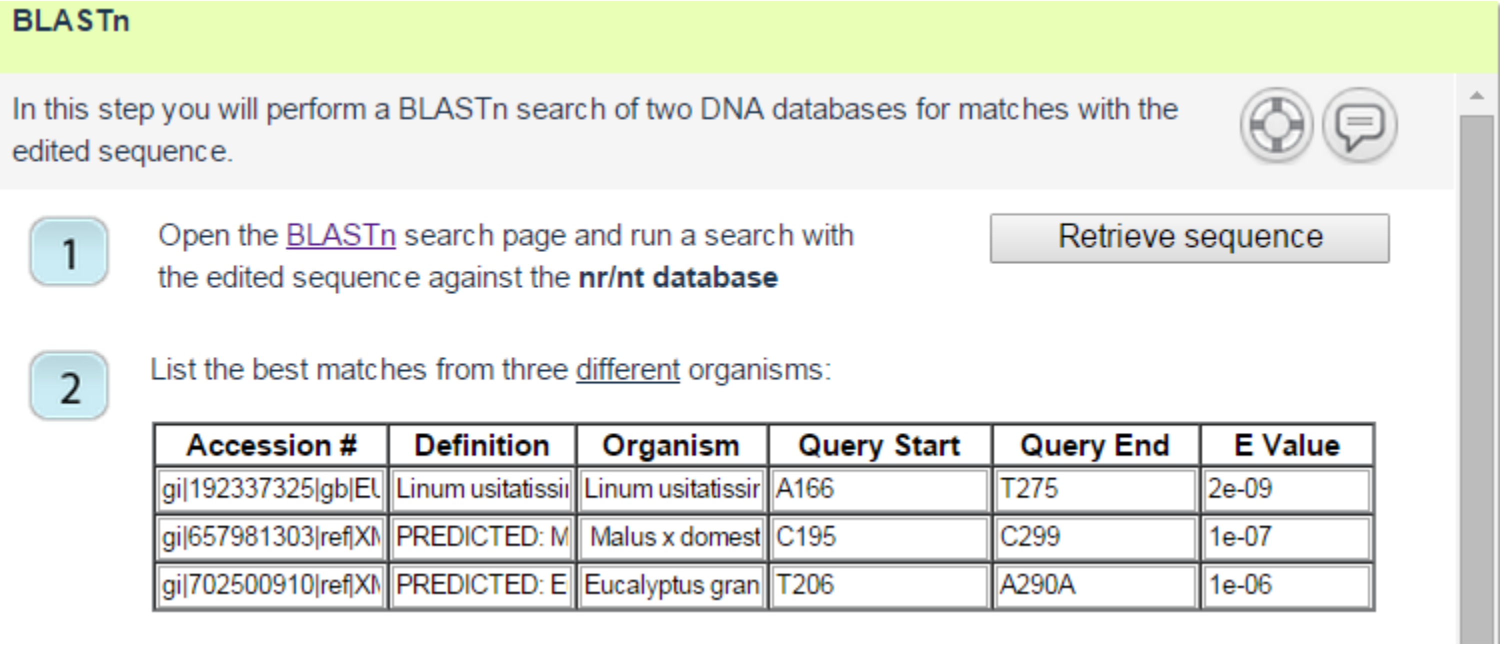
3 In the unedited sequence, which is the last base of the cDNA insert or the last base of readable sequence? (e.g., A530) A380

4 Paste the sequence starting at 'First' (from question 2) and ending at 'Last' (from question 3):

```
ACTAAATTAAGCGCCGCGCAATGCGCCGACGCCCGCGATCATCATCTTCCCTTCCTCGTCCT
CTTCCTCCGATCAAATCTTCTCAAGCTTCTCTGCAACGGGGAGGTCTTCCCCGGGCGTTTGC
GGAGCTCCTCCGGAAACTGTAGGTGATTCATCACAGGAGGGAAGATGCGTCCGAGTCCCTGC
AGGAGTTGCCTCGCGTCTCTCCTCAAACTTTTGAATTTCCTGCAAGCTTTCATCGGCGTCTC
```

Number of characters: 323 Cursor is on base: A1 Range marked:

BLASTn



BLASTn

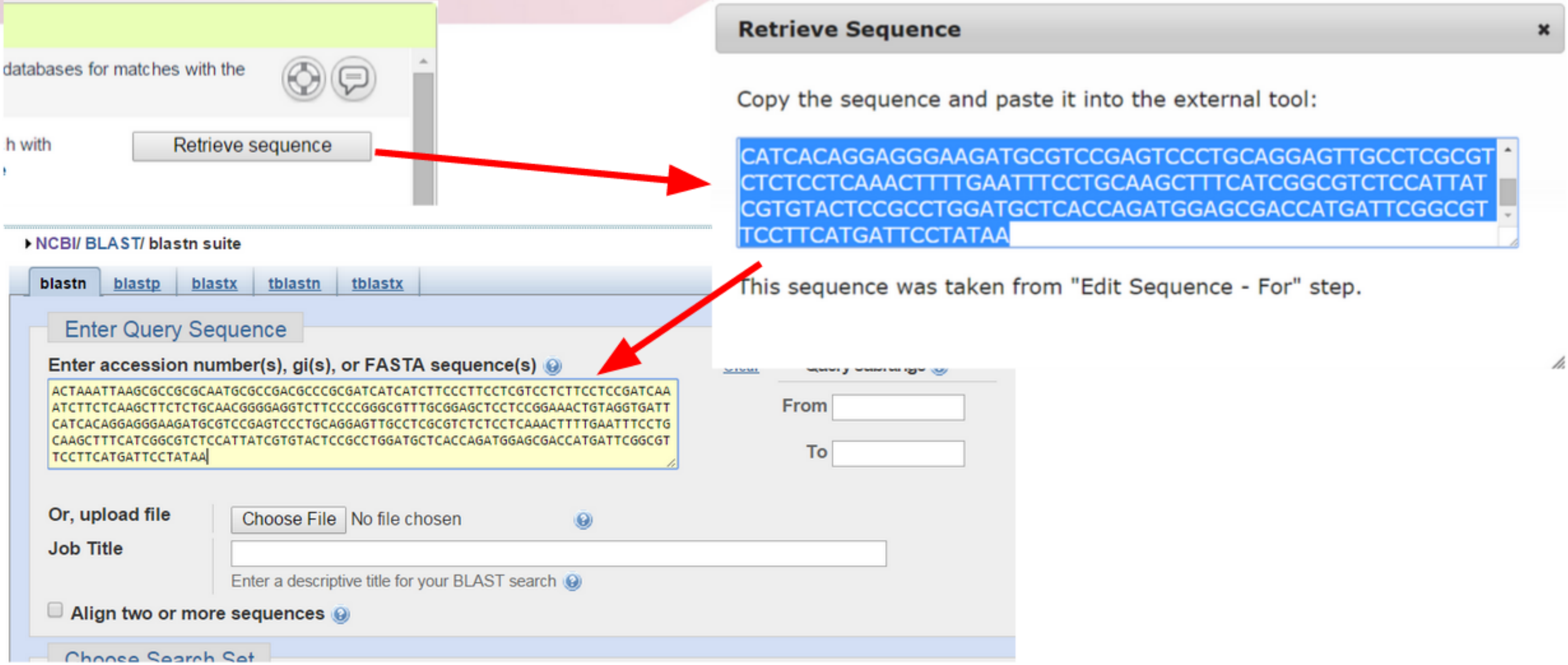
In this step you will perform a BLASTn search of two DNA databases for matches with the edited sequence.

1 Open the BLASTn search page and run a search with the edited sequence against the **nr/nt database**

2 List the best matches from three different organisms:

Accession #	Definition	Organism	Query Start	Query End	E Value
gi|192337325|gb|EU	Linum usitatissi	Linum usitatissir	A166	T275	2e-09
gi|657981303|ref|XI	PREDICTED: M	Malus x domest	C195	C299	1e-07
gi|702500910|ref|XI	PREDICTED: E	Eucalyptus gran	T206	A290A	1e-06

Running a Blastn Search



Retrieve Sequence

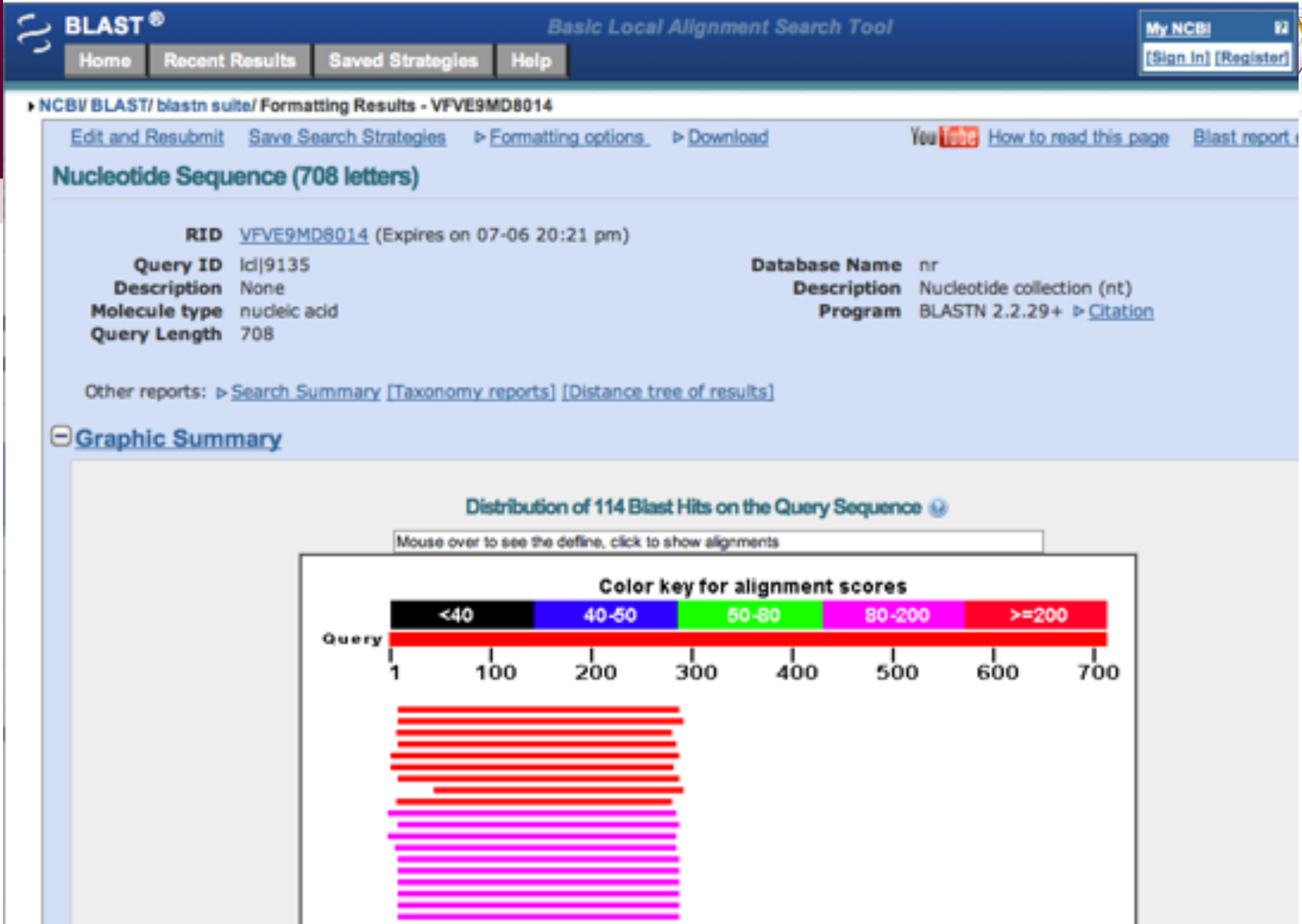
Copy the sequence and paste it into the external tool:

```
CATCACAGGAGGGAAGATGCGTCCGAGTCCCTGCAGGAGTTGCCTCGCGT
CTCTCCTCAAACTTTTGAATTTCCTGCAAGCTTTCATCGGCGTCTCCATTAT
CGTGTACTCCGCCTGGATGCTCACCAGATGGAGCGACCATGATTCGGCGT
TCCTTCATGATTCCTATAA
```

This sequence was taken from "Edit Sequence - For" step.

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

```
ACTAAATTAAGCGCCGCGCAATGCGCCGACGCCCGCGATCATCATCTTCCCTTCCTCGTCCTCTTCCTCCGATCAA
ATCTTCTCAAGCTTCTCTGCAACGGGGAGGTCTTCCCCGGGCGTTTGCGGAGCTCCTCCGGAAACTGTAGGTGATT
CATCACAGGAGGGAAGATGCGTCCGAGTCCCTGCAGGAGTTGCCTCGCGTCTCTCCTCAAACTTTTGAATTTCCTG
CAAGCTTTCATCGGCGTCTCCATTATCGTGTACTCCGCCTGGATGCTCACCAGATGGAGCGACCATGATTCGGCGT
TCCTTCATGATTCCTATAA
```



NCBI/ BLAST/ blastn suite/ Formatting Results - VFVE9MD8014

Nucleotide Sequence (708 letters)

RID VFVE9MD8014 (Expires on 07-06 20:21 pm)

Query ID lcl|9135
Description None
Molecule type nucleic acid
Query Length 708

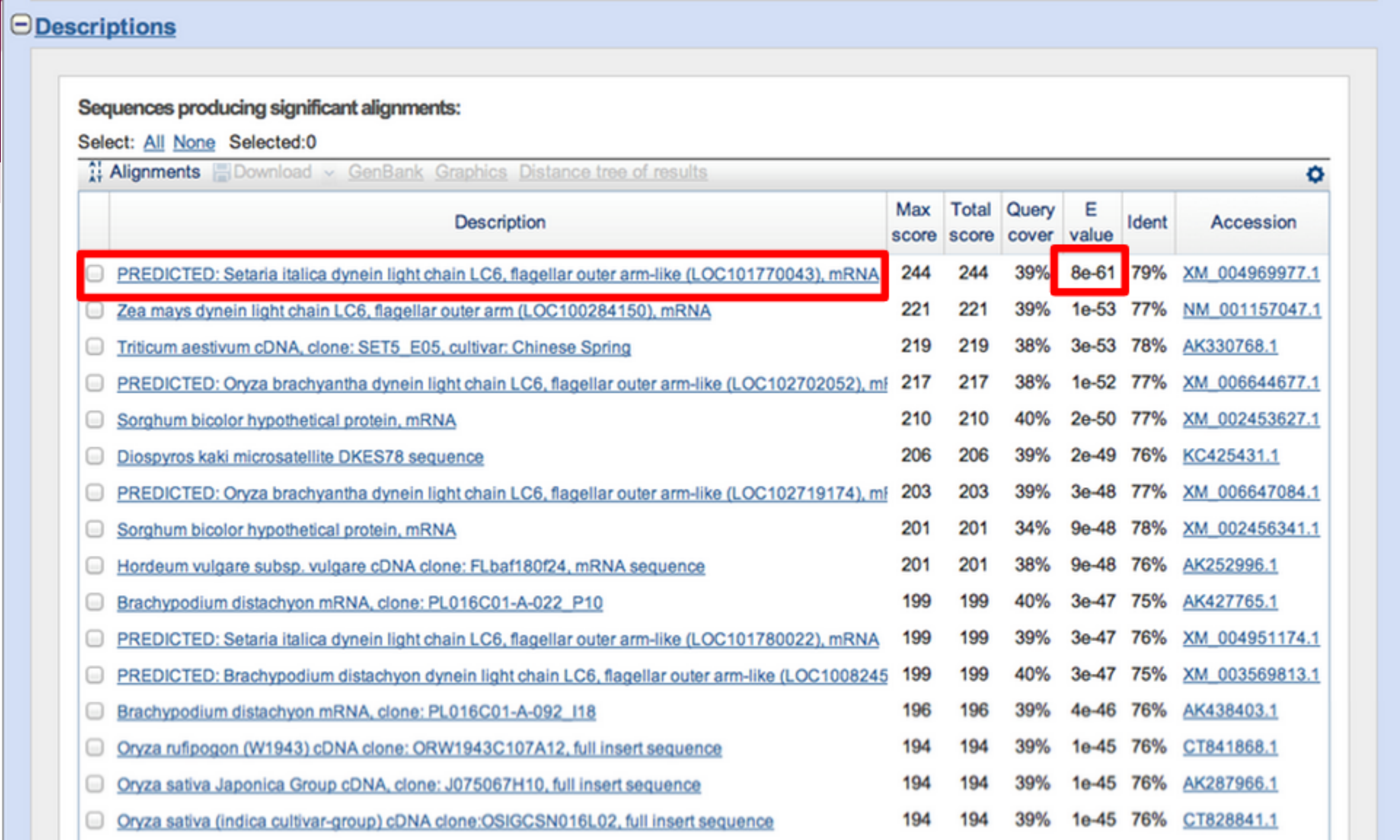
Database Name nr
Description Nucleotide collection (nt)
Program BLASTN 2.2.29+ Citation

Other reports: Search Summary [Taxonomy reports] [Distance tree of results]

Graphic Summary

Distribution of 114 Blast Hits on the Query Sequence

Mouse over to see the define, click to show alignments



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	PREDICTED: Setaria italica dynein light chain LC6, flagellar outer arm-like (LOC101770043), mRNA	244	244	39%	8e-61	79%	XM_004969977.1
☐	Zea mays dynein light chain LC6, flagellar outer arm (LOC100284150), mRNA	221	221	39%	1e-53	77%	NM_001157047.1
☐	Triticum aestivum cDNA, clone: SET5_E05, cultivar: Chinese Spring	219	219	38%	3e-53	78%	AK330768.1
☐	PREDICTED: Oryza brachyantha dynein light chain LC6, flagellar outer arm-like (LOC102702052), m	217	217	38%	1e-52	77%	XM_006644677.1
☐	Sorghum bicolor hypothetical protein, mRNA	210	210	40%	2e-50	77%	XM_002453627.1
☐	Diospyros kaki microsatellite DKES78 sequence	206	206	39%	2e-49	76%	KC425431.1
☐	PREDICTED: Oryza brachyantha dynein light chain LC6, flagellar outer arm-like (LOC102719174), m	203	203	39%	3e-48	77%	XM_006647084.1
☐	Sorghum bicolor hypothetical protein, mRNA	201	201	34%	9e-48	78%	XM_002456341.1
☐	Hordeum vulgare subsp. vulgare cDNA clone: FLbaf180f24, mRNA sequence	201	201	38%	9e-48	76%	AK252996.1
☐	Brachypodium distachyon mRNA, clone: PL016C01-A-022_P10	199	199	40%	3e-47	75%	AK427765.1
☐	PREDICTED: Setaria italica dynein light chain LC6, flagellar outer arm-like (LOC101780022), mRNA	199	199	39%	3e-47	76%	XM_004951174.1
☐	PREDICTED: Brachypodium distachyon dynein light chain LC6, flagellar outer arm-like (LOC1008245	199	199	40%	3e-47	75%	XM_003569813.1
☐	Brachypodium distachyon mRNA, clone: PL016C01-A-092_I18	196	196	39%	4e-46	76%	AK438403.1
☐	Oryza rufipogon (W1943) cDNA clone: ORW1943C107A12, full insert sequence	194	194	39%	1e-45	76%	CT841868.1
☐	Oryza sativa Japonica Group cDNA, clone: J075067H10, full insert sequence	194	194	39%	1e-45	76%	AK287966.1
☐	Oryza sativa (indica cultivar-group) cDNA clone:OSIGCSN016L02, full insert sequence	194	194	39%	1e-45	76%	CT828841.1

PREDICTED: Setaria italica dynein light chain LC6, flagellar outer arm-like (LOC101770043), mRNA

Sequence ID: ref|XM_004969977.1| Length: 791 Number of Matches: 1

Range 1: 222 to 499 GenBank Graphics

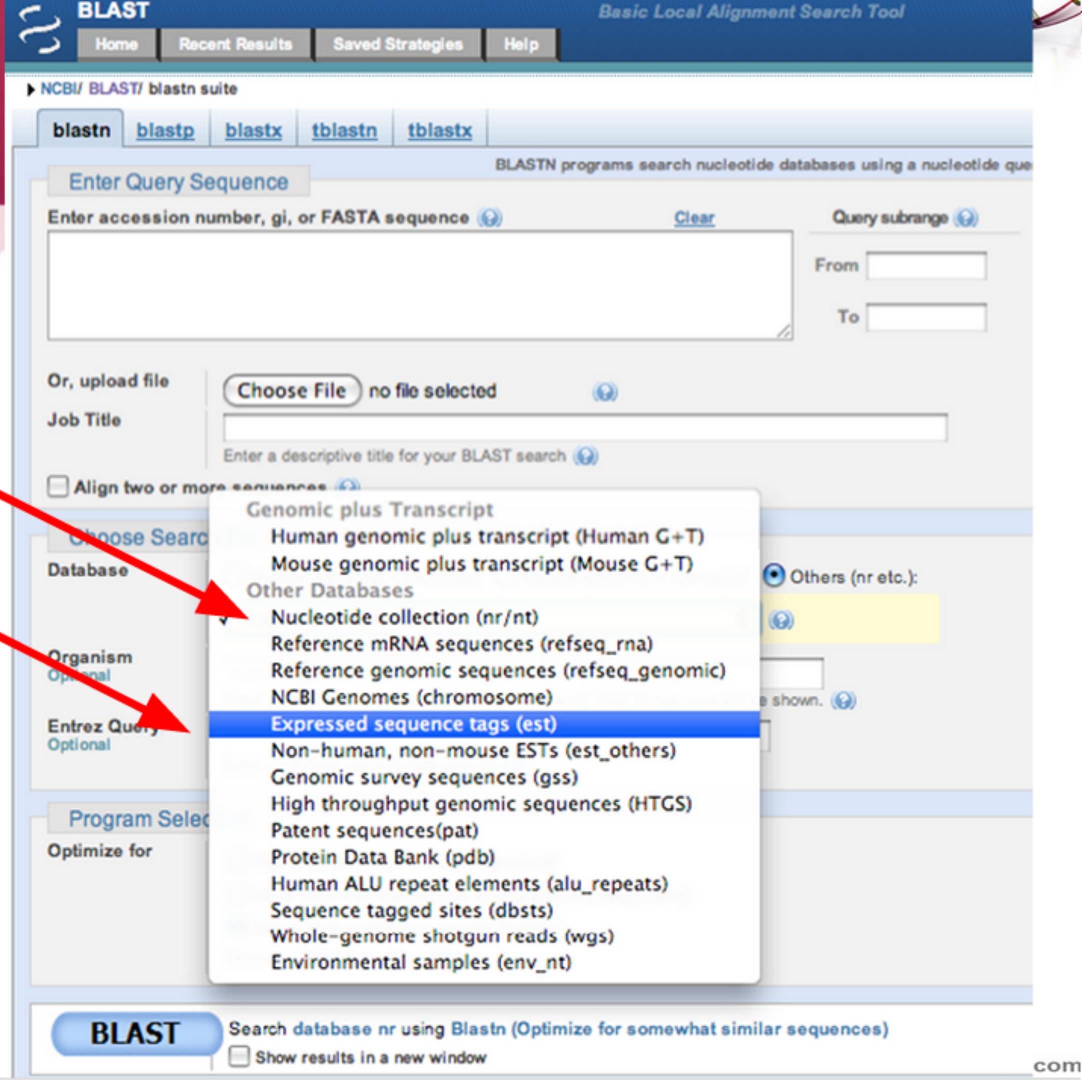
Score	Expect	Identities	Gaps	Strand
244 bits(270)	8e-61	221/278(79%)	0/278(0%)	Plus/Plus

```
Query  11   ATGTTGGAAGGGAGGGCGAGAGTAGAAGACACCGACATGCCGAGGAAGATGCAGGCGGAG  70
            ||||||||||||| |||||  || || ||||| |||||||||  ||||||||||||| ||
Sbjct  222  ATGTTGGAAGGGAAGGCGATGGTGGAGGACACGGACATGCCGGCGAAGATGCAGGCGCAG  281

Query  71   GCCATGAACGCCGCCTCTCACGCGCTCGATCTGTTCGACGTCGCGGACTGCAAGAGCCTC  130
            || |||   || || ||    || ||||| |  |||||||||   ||||||  |||| ||
Sbjct  282  GCGATGGCGGCGGCGTCCAGGGCCCTCGACCGCTTCGACGTCCTCGACTGCCGGAGCATC  341

Query  131  GCCGCGCATATCAAGAAGGAATTTGATAAGATCTACGGTCCGGGATGGCAGTGCGTCGTC  190
            || || || ||||||||||| ||||| | |||| | || ||||||||||| ||||| ||
Sbjct  342  GCGGCCCACATCAAGAAGGAGTTTGACACGATCCATGGACCGGGATGGCAATGCGTGGTG  401

Query  191  GGCTCCAGCTTCGGCTGTTTCTTCACTCACAAGAAAGGCAGCTTCATCTACTTCCGCCTG  250
            ||||||||||||||||| | |||||| ||||  || || ||||||||||||||||| ||
Sbjct  402  GGCTCCAGCTTCGGCTGCTACTTCACGCACAGCAAGGGGAGCTTCATCTACTTCCGGCTC  461

Query  251  GAGACGCTCCACTTCCTCATCTTCAAAGGCGCGGCCGC  288
            ||| |||||   |||||| |||||||||| ||||| ||
Sbjct  462  GAGTCGCTCAGGTTCCTCGTCTTCAAAGGGGCGGCAGC  499
```

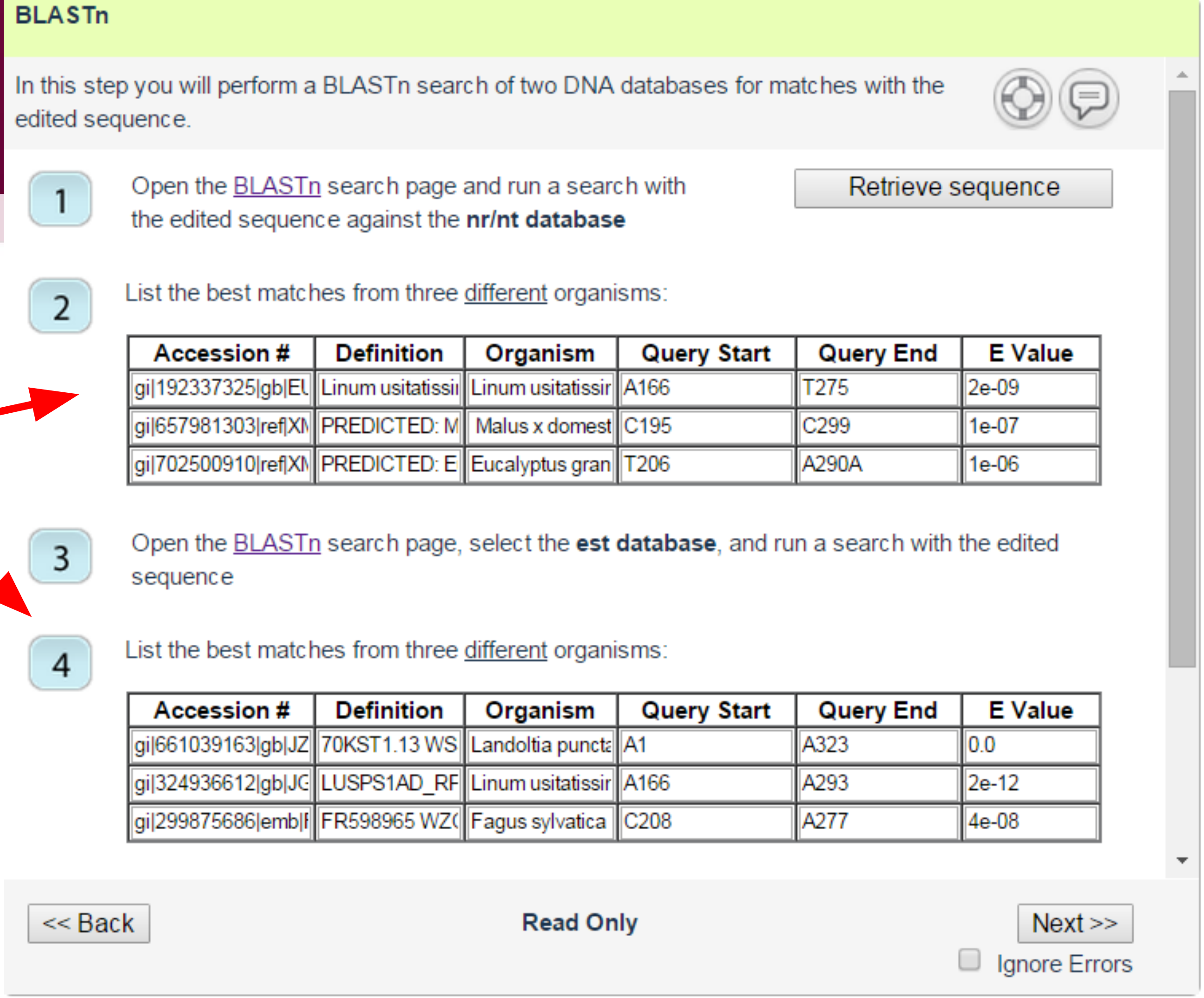
nr/nt v.s. EST

nr/nt: Nucleotide
EST: Expressed Sequence Tags

...WHY???



Search database nr using Blastn (Optimize for somewhat similar sequences)



BLASTn

In this step you will perform a BLASTn search of two DNA databases for matches with the edited sequence.

1 Open the BLASTn search page and run a search with the edited sequence against the **nr/nt database**

2 List the best matches from three different organisms:

Accession #	Definition	Organism	Query Start	Query End	E Value
gi|192337325|gb|E	Linum usitatissi	Linum usitatissi	A166	T275	2e-09
gi|657981303|ref|X	PREDICTED: M	Malus x domest	C195	C299	1e-07
gi|702500910|ref|X	PREDICTED: E	Eucalyptus gran	T206	A290A	1e-06

3 Open the BLASTn search page, select the **est database**, and run a search with the edited sequence

4 List the best matches from three different organisms:

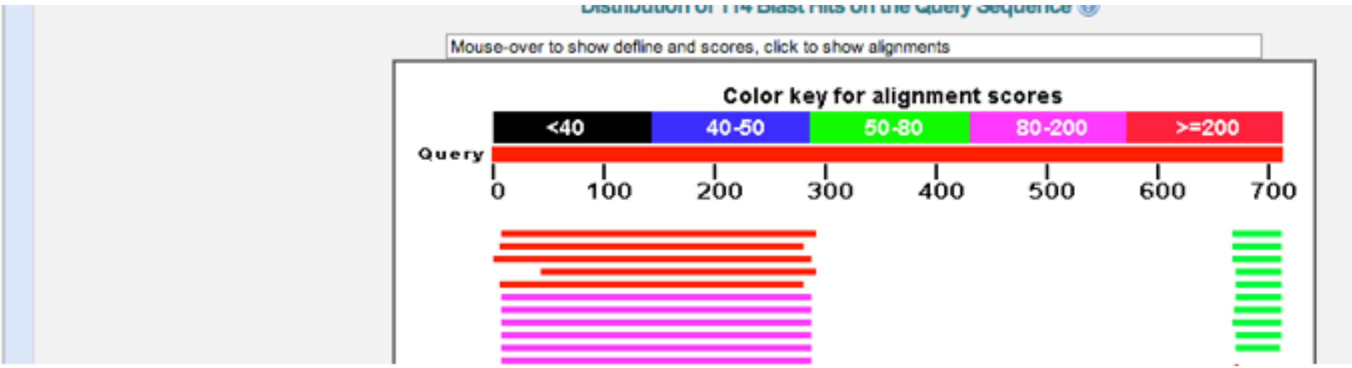
Accession #	Definition	Organism	Query Start	Query End	E Value
gi|661039163|gb|JZ	70KST1.13 WS	Landoltia punct	A1	A323	0.0
gi|324936612|gb|J	LUSPS1AD_RF	Linum usitatissi	A166	A293	2e-12
gi|299875686|emb|	FR598965 WZ	Fagus sylvatica	C208	A277	4e-08

Let’s Go More In-Depth...

Is it significant?

1. Eval

Other Things:
-Gaps
-Mismatches
-Length
-Graphic Summaries



Distribution of 114 Blast Hits on the Query Sequence

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

Gaps/Mismatches

“Degenerate Code”

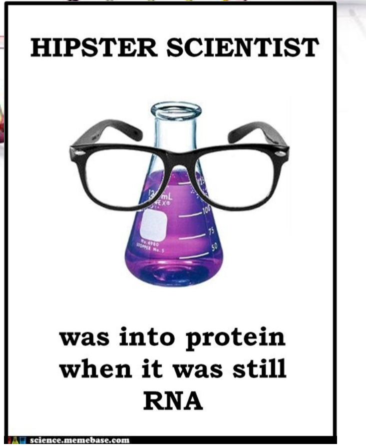
```
         C   R   E   L   L   I   L   D   A
Query    TGT CGT GAA CTC CTA ATT CTC GAC GCC
         ||| ||| ||| ||  ||  ||  ||  ||  ||
Sbjct    TGT CGT GAA CTT CTG ATC CTT GAT GCA
         C   R   E   L   L   I   L   D   A
```

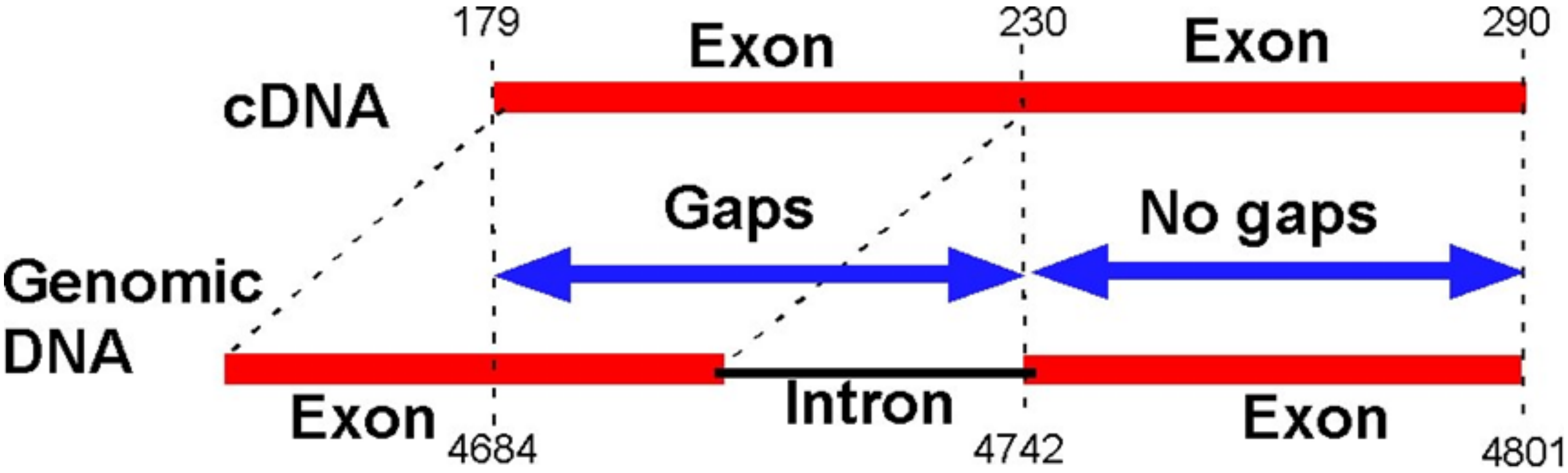
Amino Acid

```
         C  R    R  T  P  D  P  *
Query    TGTCGT-CGAACTCCTGATCCTTGA
         |||||| ||||||||||||||||||
Sbjct    TGTCGTCCGAACTCCTGATCCTTGA
         C  R  E  L  L  I  L  D
```

Why?

1. Bad Sequences
2. Different Organisms



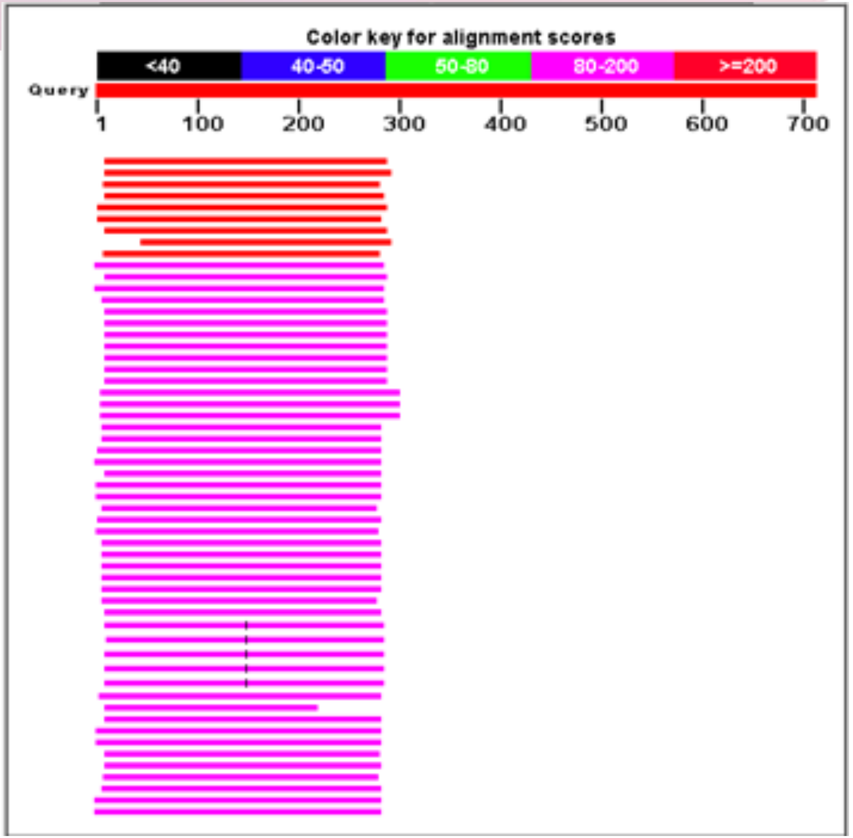


Length

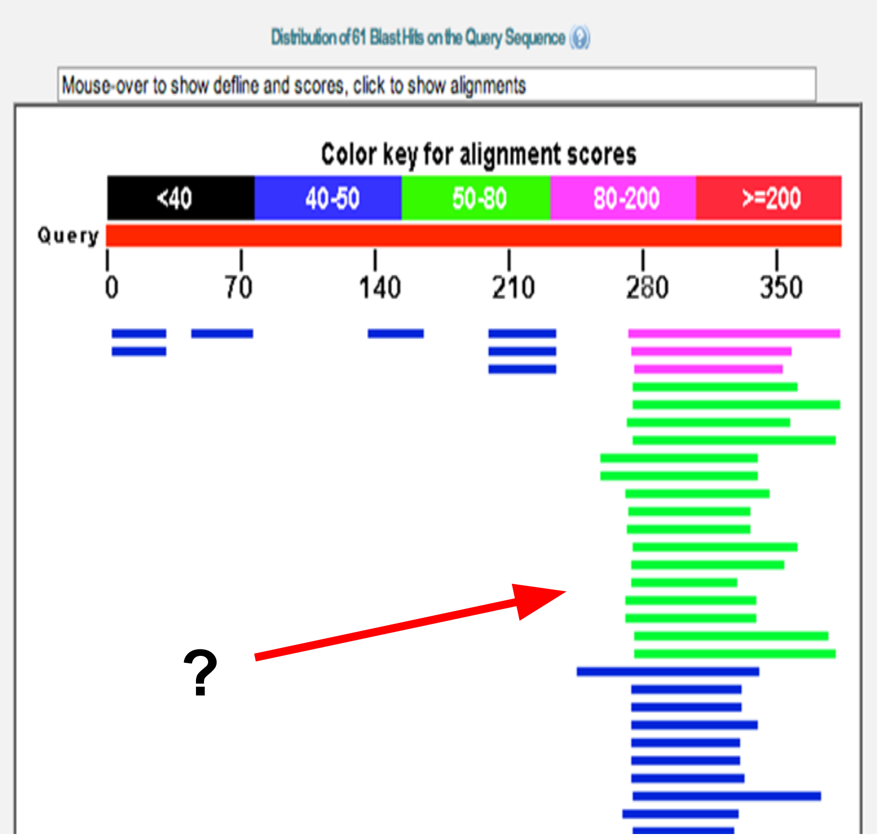
USELESS!!

```
>gi|14250883|emb|AL583809.3|CNS07EFY Human chromosome 14 DNA
sequence BAC R-736L22 of library RPCI-11 from chromosome 14 of
Homo sapiens (Human), complete sequence
 Score = 40.1 bits (20), Expect = 4.6 Identities = 20/20 (100%)

Query: 189   ttttctgaatattcataata 208
              ||||||||||||||||||||
Sbjct: 60645 ttttctgaatattcataata 60626
```

Use the Graphic Report



Color key for alignment scores

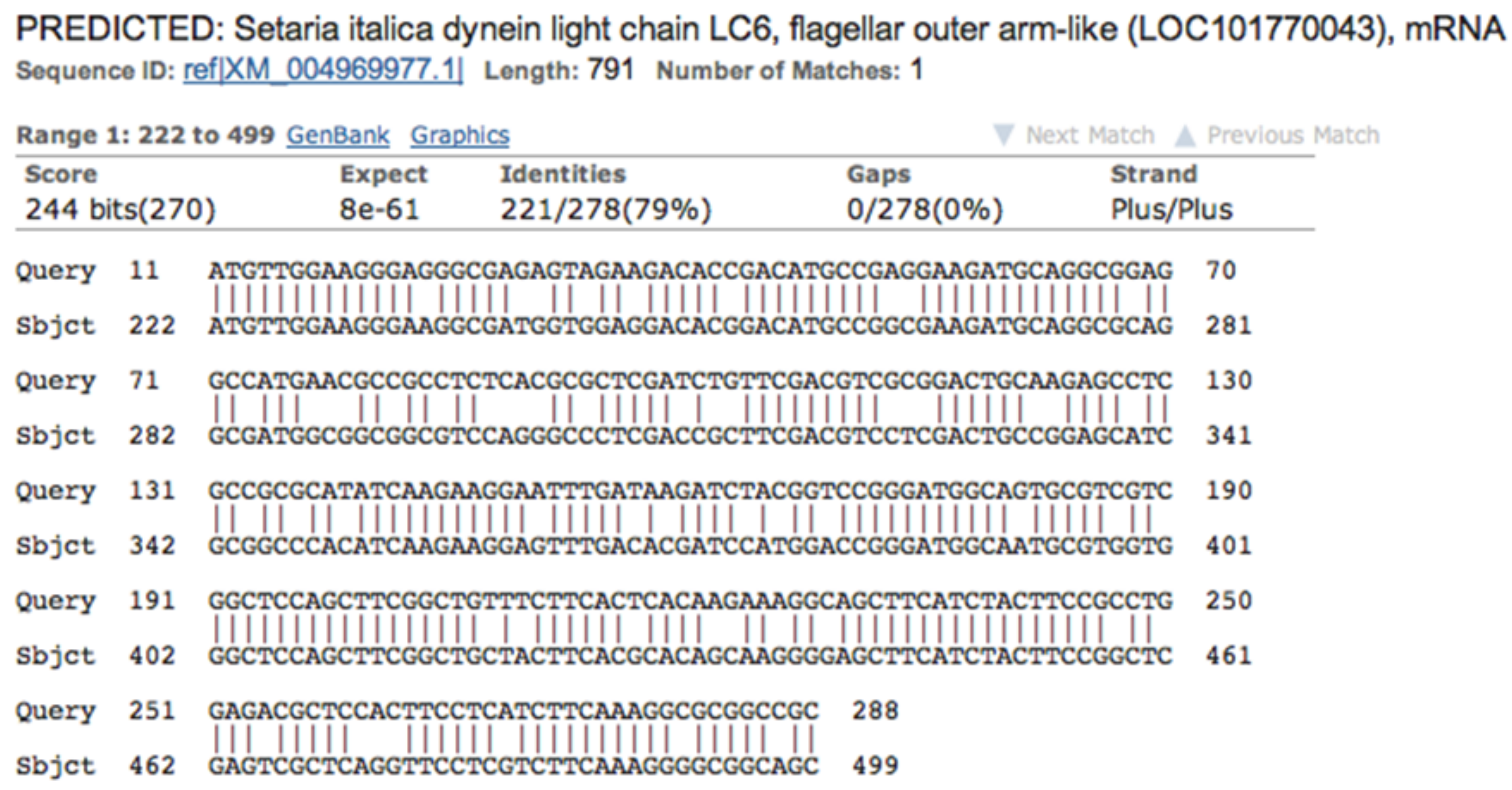


Distribution of 61 Blast Hits on the Query Sequence

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

Color key for alignment scores

Is this a good result?
Is this significant?



PREDICTED: Setaria italica dynein light chain LC6, flagellar outer arm-like (LOC101770043), mRNA

Sequence ID: ref|XM_004969977.1| **Length:** 791 **Number of Matches:** 1

Range 1: 222 to 499

Score	Expect	Identities	Gaps	Strand
244 bits(270)	8e-61	221/278(79%)	0/278(0%)	Plus/Plus

```
Query  11   ATGTTGGAAGGGAGGGCGAGAGTAGAAGACACCGACATGCCGAGGAAGATGCAGGCGGAG  70
            ||||||||||||| |||||  || || ||||| |||||||||  ||||||||||||| ||
Sbjct  222  ATGTTGGAAGGGAAGGCGATGGTGGAGGACACGGACATGCCGGCGAAGATGCAGGCGCAG  281

Query  71   GCCATGAACGCCGCCTCTCACGCGCTCGATCTGTTCGACGTCGCGGACTGCAAGAGCCTC  130
            || |||   || || ||    || ||||| |  |||||||||   ||||||  |||| ||
Sbjct  282  GCGATGGCGGCGGCGTCCAGGGCCCTCGACCGCTTCGACGTCCTCGACTGCCGGAGCATC  341

Query  131  GCCGCGCATATCAAGAAGGAATTTGATAAGATCTACGGTCCGGGATGGCAGTGCGTCGTC  190
            || || || ||||||||||| ||||| | |||| | || ||||||||||| ||||| || 
Sbjct  342  GCGGCCCACATCAAGAAGGAGTTTGACACGATCCATGGACCGGGATGGCAATGCGTGGTG  401

Query  191  GGCTCCAGCTTCGGCTGTTTCTTCACTCACAAGAAAGGCAGCTTCATCTACTTCCGCCTG  250
            ||||||||||||||||| | |||||| ||||  || || ||||||||||||||||| || 
Sbjct  402  GGCTCCAGCTTCGGCTGCTACTTCACGCACAGCAAGGGGAGCTTCATCTACTTCCGGCTC  461

Query  251  GAGACGCTCCACTTCCTCATCTTCAAAGGCGCGGCCGC  288
            ||| |||||    |||||| |||||||||| ||||| ||
Sbjct  462  GAGTCGCTCAGGTTCCTCGTCTTCAAAGGGGCGGCAGC  499
```

So, why Blastx??

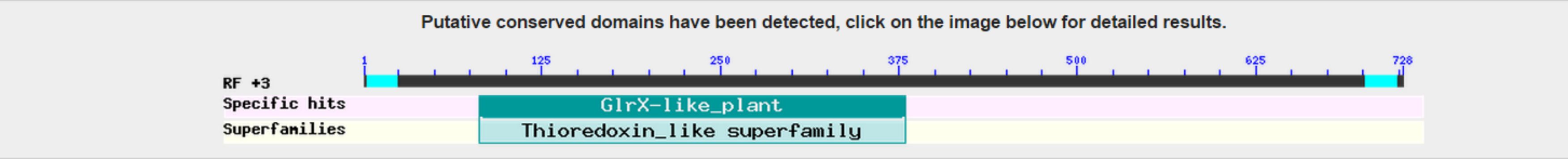
Blastn:

```
Query AGG TCG TTA CTA TCG AGG AGT AGA
            |       |           |       |
Sbjct CGT AGC CTT TTG AGT CGA TCG CGG
             16% Identity
```

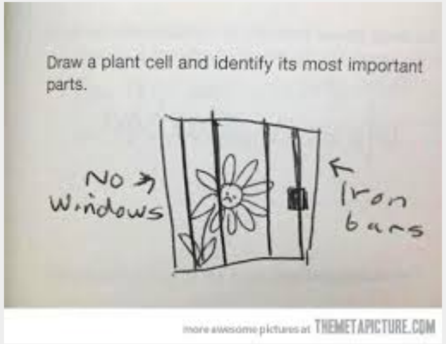
Blastx:

```
R S L L S R S R
| | | | | | | |   100% Identity
R S L L S R S R
```

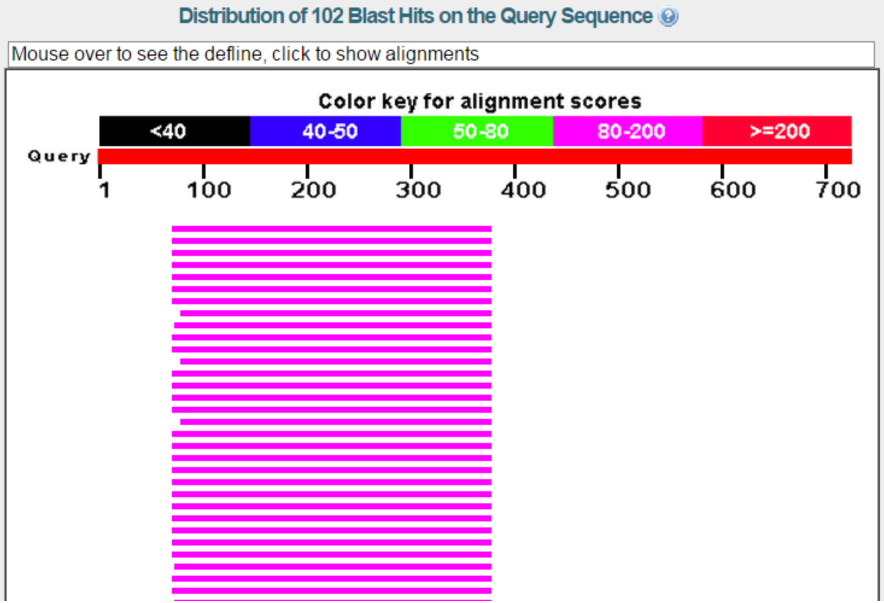
Blastx



Putative conserved domains have been detected, click on the image below for detailed results.



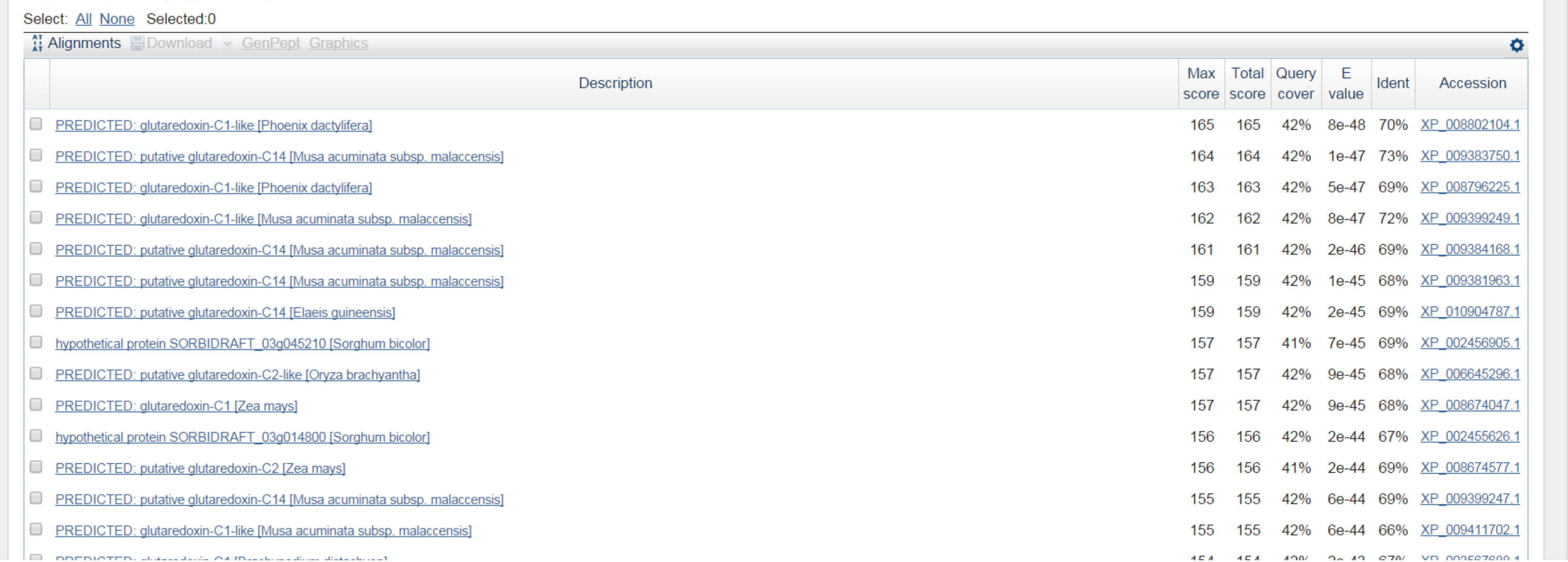
Draw a plant cell and identify its most important parts.



Distribution of 102 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments

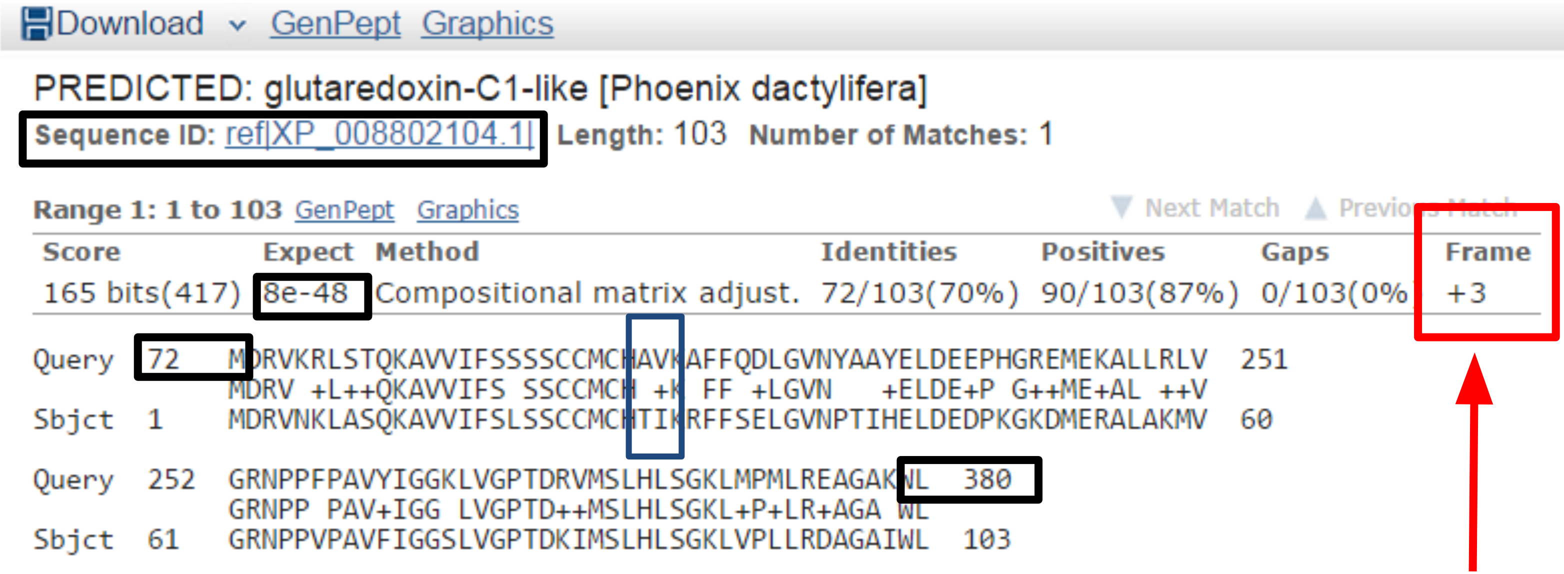
Blastx


Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	PREDICTED: glutaredoxin-C1-like [Phoenix dactylifera]	165	165	42%	8e-48	70%	XP_008802104.1
☐	PREDICTED: putative glutaredoxin-C14 [Musa acuminata subsp. malaccensis]	164	164	42%	1e-47	73%	XP_009383750.1
☐	PREDICTED: glutaredoxin-C1-like [Phoenix dactylifera]	163	163	42%	5e-47	69%	XP_008796225.1
☐	PREDICTED: glutaredoxin-C1-like [Musa acuminata subsp. malaccensis]	162	162	42%	8e-47	72%	XP_009399249.1
☐	PREDICTED: putative glutaredoxin-C14 [Musa acuminata subsp. malaccensis]	161	161	42%	2e-46	69%	XP_009384168.1
☐	PREDICTED: putative glutaredoxin-C14 [Musa acuminata subsp. malaccensis]	159	159	42%	1e-45	68%	XP_009381963.1
☐	PREDICTED: putative glutaredoxin-C14 [Elaeis guineensis]	159	159	42%	2e-45	69%	XP_010904787.1
☐	hypothetical protein SORBIDRAFT_03g045210 [Sorghum bicolor]	157	157	41%	7e-45	69%	XP_002456905.1
☐	PREDICTED: putative glutaredoxin-C2-like [Oryza brachyantha]	157	157	42%	9e-45	68%	XP_006645296.1
☐	PREDICTED: glutaredoxin-C1 [Zea mays]	157	157	42%	9e-45	68%	XP_008674047.1
☐	hypothetical protein SORBIDRAFT_03g014800 [Sorghum bicolor]	156	156	42%	2e-44	67%	XP_002455626.1
☐	PREDICTED: putative glutaredoxin-C2 [Zea mays]	156	156	41%	2e-44	69%	XP_008674577.1
☐	PREDICTED: putative glutaredoxin-C14 [Musa acuminata subsp. malaccensis]	155	155	42%	6e-44	69%	XP_009399247.1
☐	PREDICTED: glutaredoxin-C1-like [Musa acuminata subsp. malaccensis]	155	155	42%	6e-44	66%	XP_009411702.1

Frames



PREDICTED: glutaredoxin-C1-like [Phoenix dactylifera]

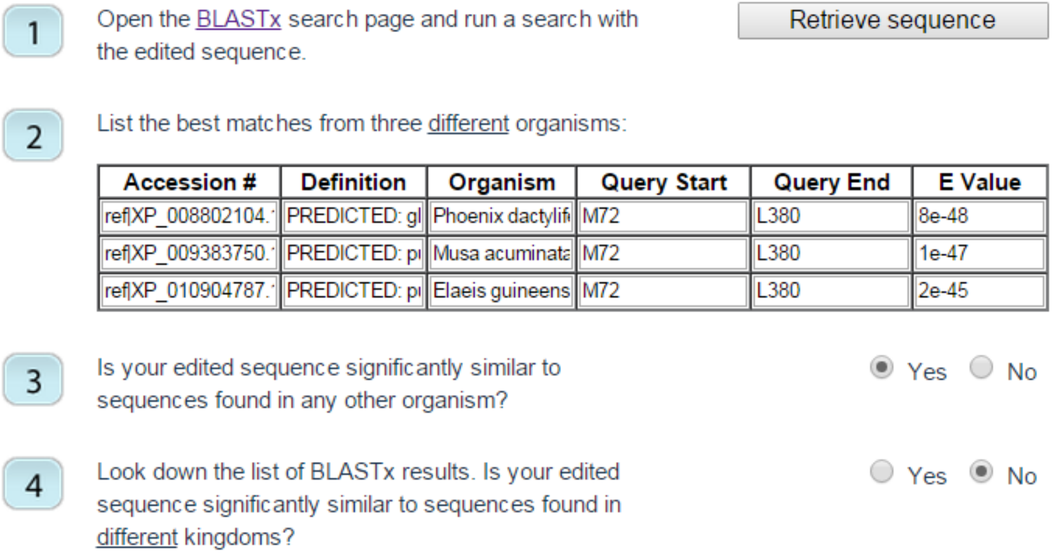
Sequence ID: ref|XP_008802104.1| **Length:** 103 **Number of Matches:** 1

Range 1: 1 to 103 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps	Frame
165 bits(417)	8e-48	Compositional matrix adjust.	72/103(70%)	90/103(87%)	0/103(0%	+3

```
Query  72   MDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQDLGVNYAAYELDEEPHGREMEKALLRLV  251
            MDRV +L++QKAVVIFS SSCCMCH +K FF +LGVN    +ELDE+P G++ME+AL ++V
Sbjct  1    MDRVNKLASQKAVVIFSLSSCCMCHTIKRFFSELGVNPTIHELDEDPKGKDMERALAKMV  60

Query  252  GRNPPFPAVYIGGKLVGPTDRVMSLHLSGKLMPMLREAGAKWL  380
            GRNPP PAV+IGG LVGPTD++MSLHLSGKL+P+LR+AGA WL
Sbjct  61   GRNPPVPAVFIGGSLVGPTDKIMSLHLSGKLVPLLRDAGAIWL  103
```

Answer the Q’s



1 Open the BLASTx search page and run a search with the edited sequence.

2 List the best matches from three different organisms:

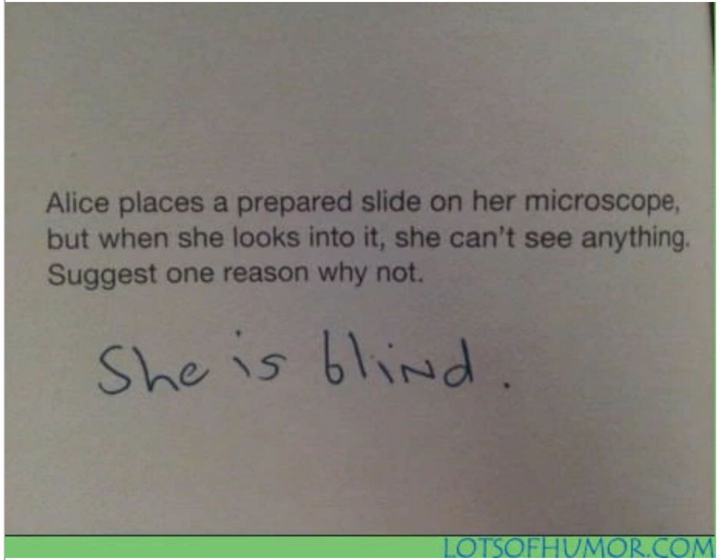
Accession #	Definition	Organism	Query Start	Query End	E Value
ref|XP_008802104.	PREDICTED: gl	Phoenix dactylif	M72	L380	8e-48
ref|XP_009383750.	PREDICTED: p	Musa acuminat	M72	L380	1e-47
ref|XP_010904787.	PREDICTED: p	Elaeis guineens	M72	L380	2e-45

3 Is your edited sequence significantly similar to sequences found in any other organism?

● Yes ○ No

4 Look down the list of BLASTx results. Is your edited sequence significantly similar to sequences found in different kingdoms?

○ Yes ● No

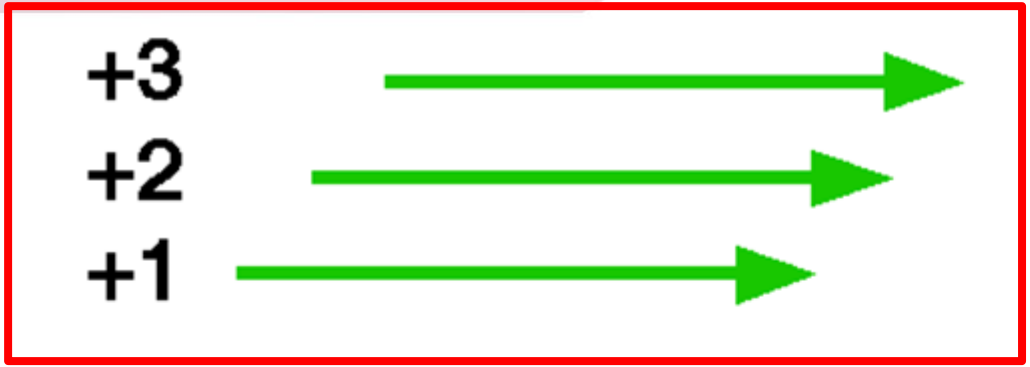


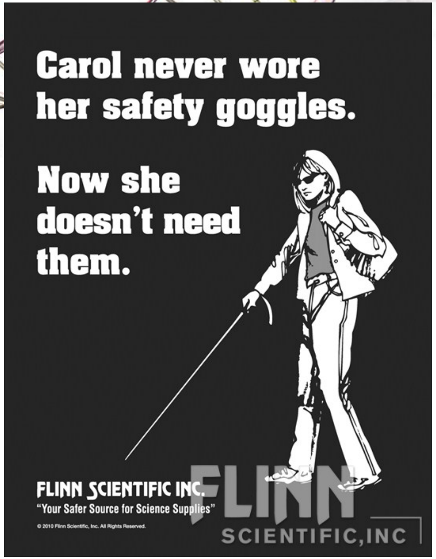
Alice places a prepared slide on her microscope, but when she looks into it, she can’t see anything. Suggest one reason why not.

She is blind.

LOTSOFHUMOR.COM

What’s a Reading Frame?





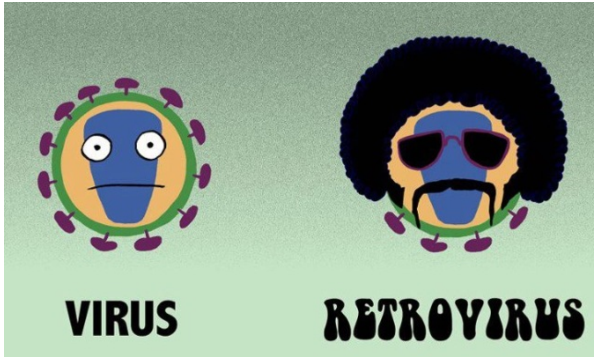
Carol never wore her safety goggles.

Now she doesn’t need them.

FLINN SCIENTIFIC INC.
“Your Safer Source for Science Supplies”
© 2010 Flinn Scientific, Inc. All Rights Reserved.

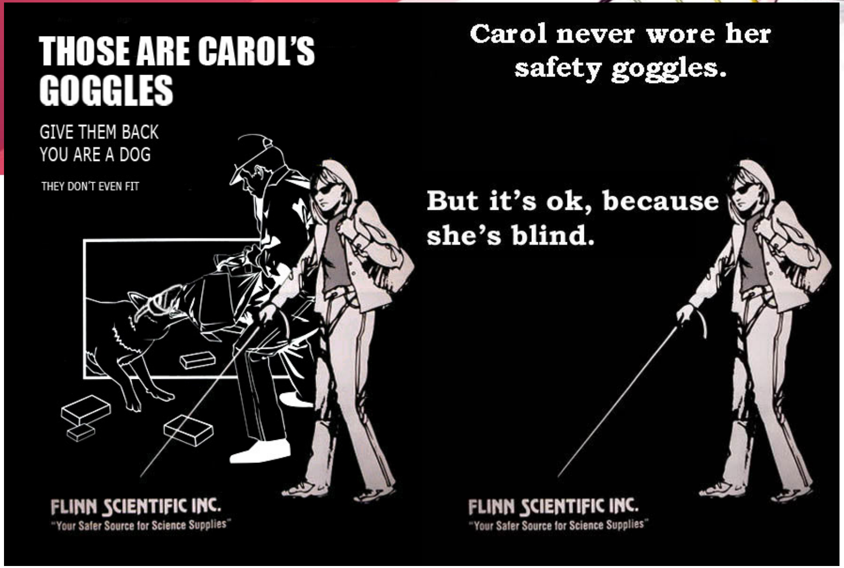
What’s a Reading Frame?

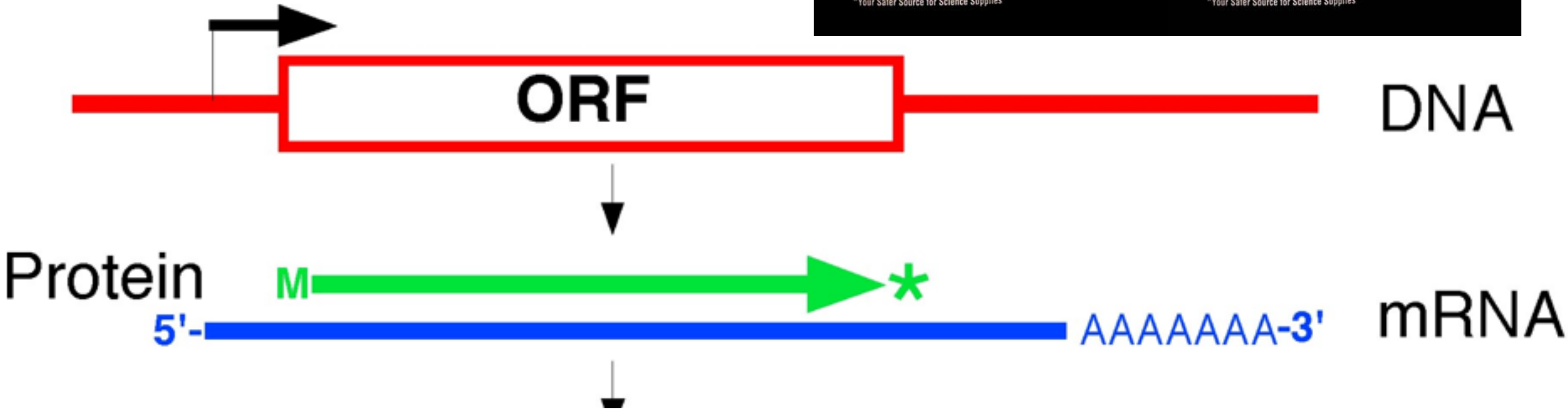
```
Query AGG TCG TTA CTA TCG AGG AGT AGA
            |       |           |       |
Sbjct CGT AGC CTT TTG AGT CGA TCG CGG
             16% Identity
```



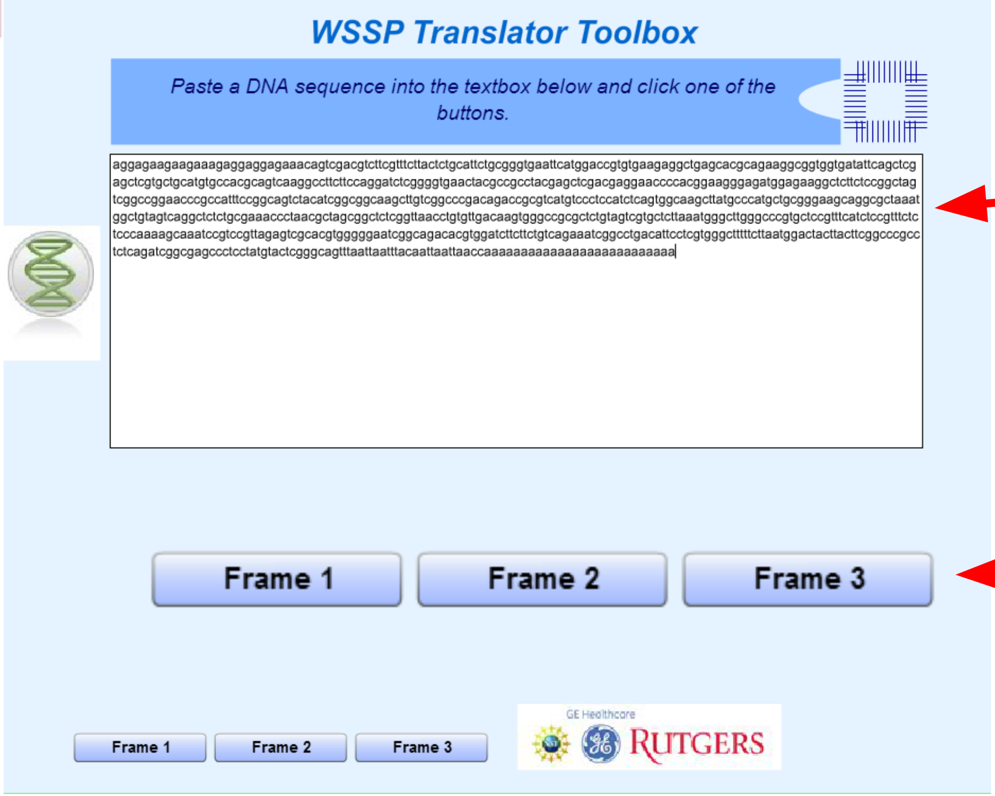
The “ORF”

“OPEN READING FRAME”





Toolbox

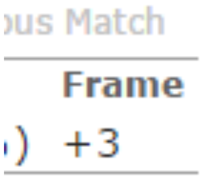


WSSP Translator Toolbox

Paste a DNA sequence into the textbox below and click one of the buttons.

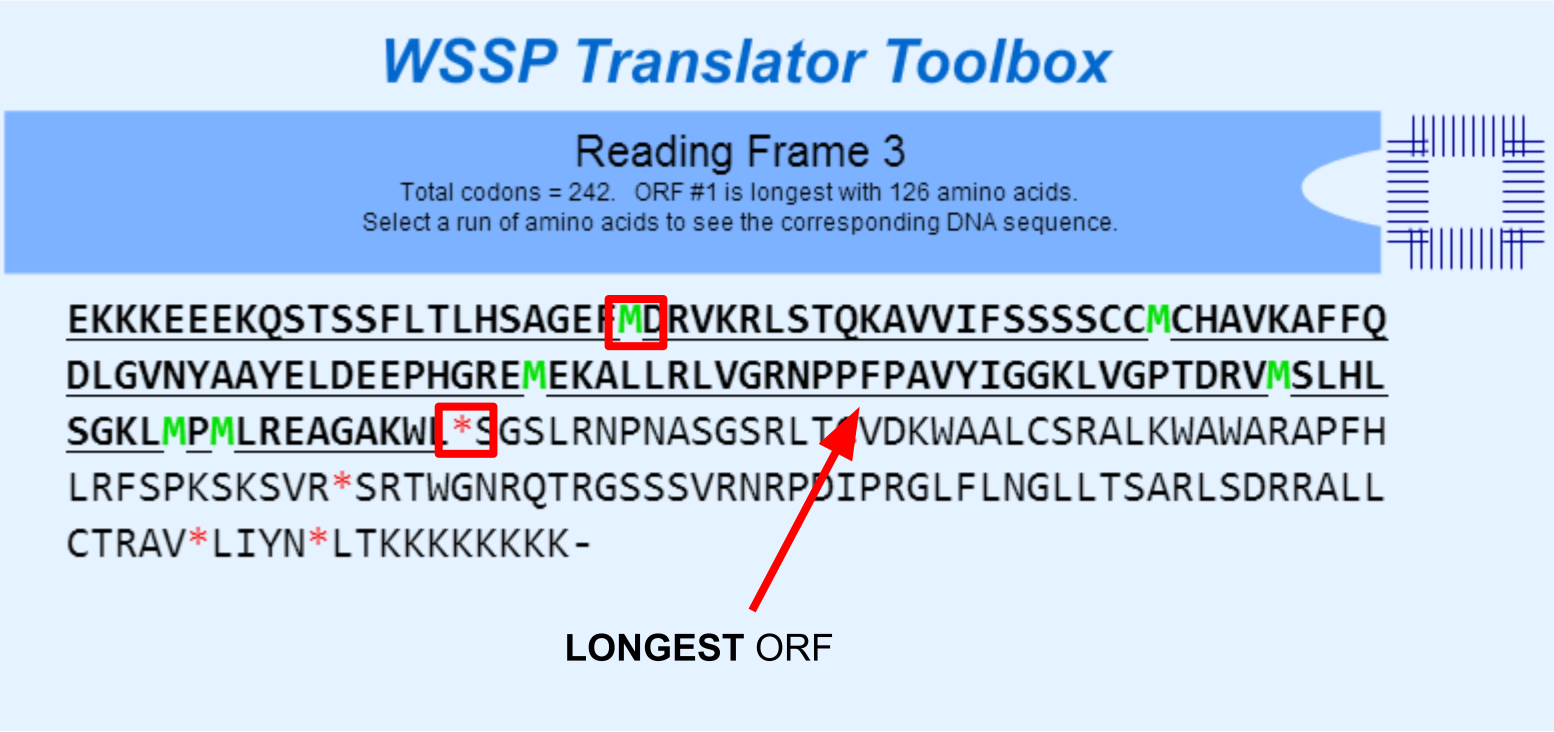
```
aggagaagaagaaagaggaggagaaacagtcgacgtcttcgtttcttactctgcattctgcgggtgaattcatggaccgtgtgaagaggctgagcacgcagaaggcggtggtgatattcagctcg
agctcgtgctgcatgtgccacgcagtcaaggccttcttccaggatctcggggtgaactacgccgcctacgagctcgacgaggaaccccacggaagggagatggagaaggctcttctccggctag
tcggccggaacccgccatttccggcagtctacatcggcggcaagcttgtcggcccgacagaccgcgtcatgtccctccatctcagtggcaagcttatgcccatgctgcgggaagcaggcgctaaat
ggctgtagtcaggctctctgcgaaaccctaacgctagcggctctcggttaacctgtgttgacaagtgggccgcgctctgtagtcgtgctcttaaatgggcttgggcccgtgctccgtttcatctccgtttctc
tcccaaaagcaaatccgtccgttagagtcgcacgtgggggaatcggcagacacgtggatcttcttctgtcagaaatcggcctgacattcctcgtgggctttttcttaatggactacttacttcggcccgcc
tctcagatcggcgagccctcctatgtactcgggcagtttaattaatttacaattaattaaccaaaaaaaaaaaaaaaaaaaaaaaaaa
```

Copy/Paste DNA



Select Your Reading Frame

Protein Sequence



WSSP Translator Toolbox

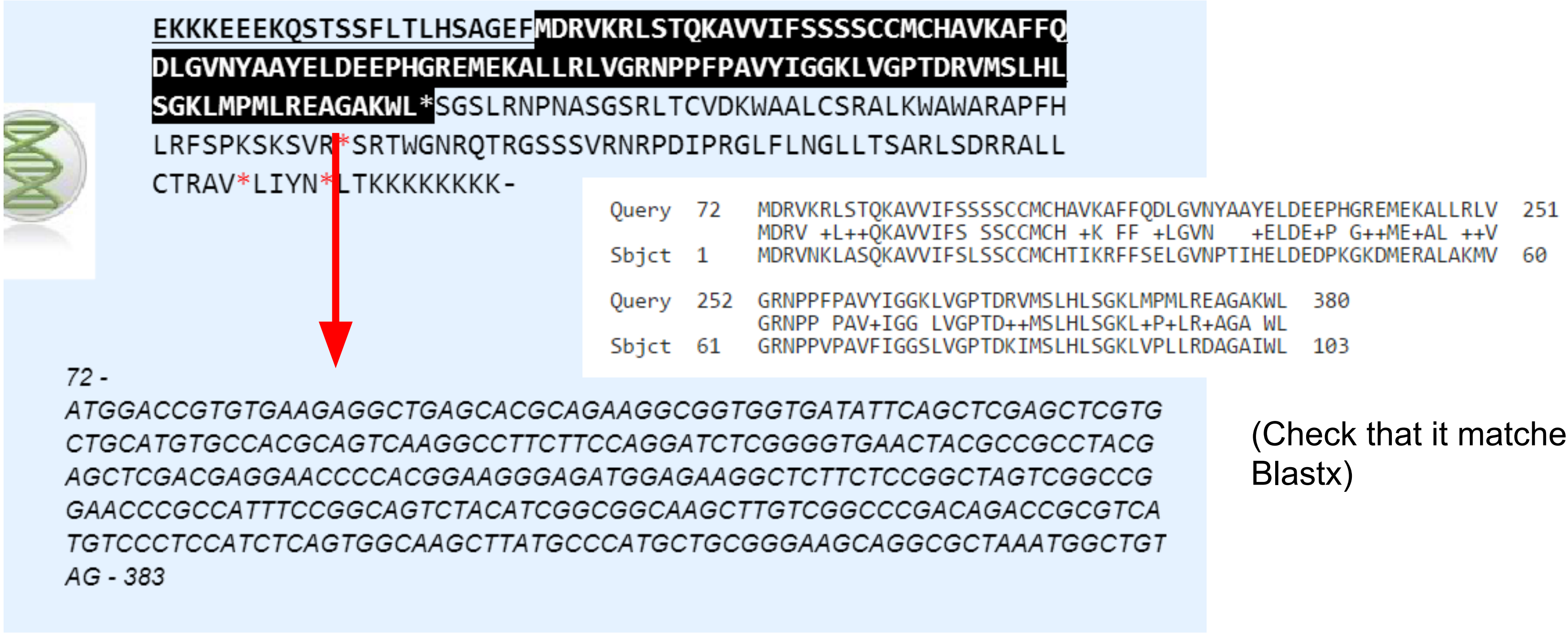
Reading Frame 3

Total codons = 242. ORF #1 is longest with 126 amino acids.
Select a run of amino acids to see the corresponding DNA sequence.

```
EKKKEEEKQSTSSFLTLHSAGEFMDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQ
DLGVNYAAYELDEEPHGREMEKALLRLVGRNPPFPAVYIGGKLVGPTDRVMSLHL
SGKLMPMLREAGAKWL*SGSLRNPNASGSRLTQVDKWAALCSRALKWAWARAPFH
LRFSPKSKSVR*SRTWGNRQTRGSSSVRNRPDIPRGLFLNGLLTSARLSDRRALL
CTRAV*LIYN*LTKKKKKKK-
```

LONGEST ORF

Finally-Your Sequence



```
EKKKEEEKQSTSSFLTLHSAGEFMDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQ
DLGVNYAAYELDEEPHGREMEKALLRLVGRNPPFPAVYIGGKLVGPTDRVMSLHL
SGKLMPMLREAGAKWL*SGSLRNPNASGSRLTCVDKWAALCSRALKWAWARAPFH
LRFSPKSKSVR*SRTWGNRQTRGSSSVRNRPDIPRGLFLNGLLTSARLSDRRALL
CTRAV*LIYN*LTKKKKKKK-
```

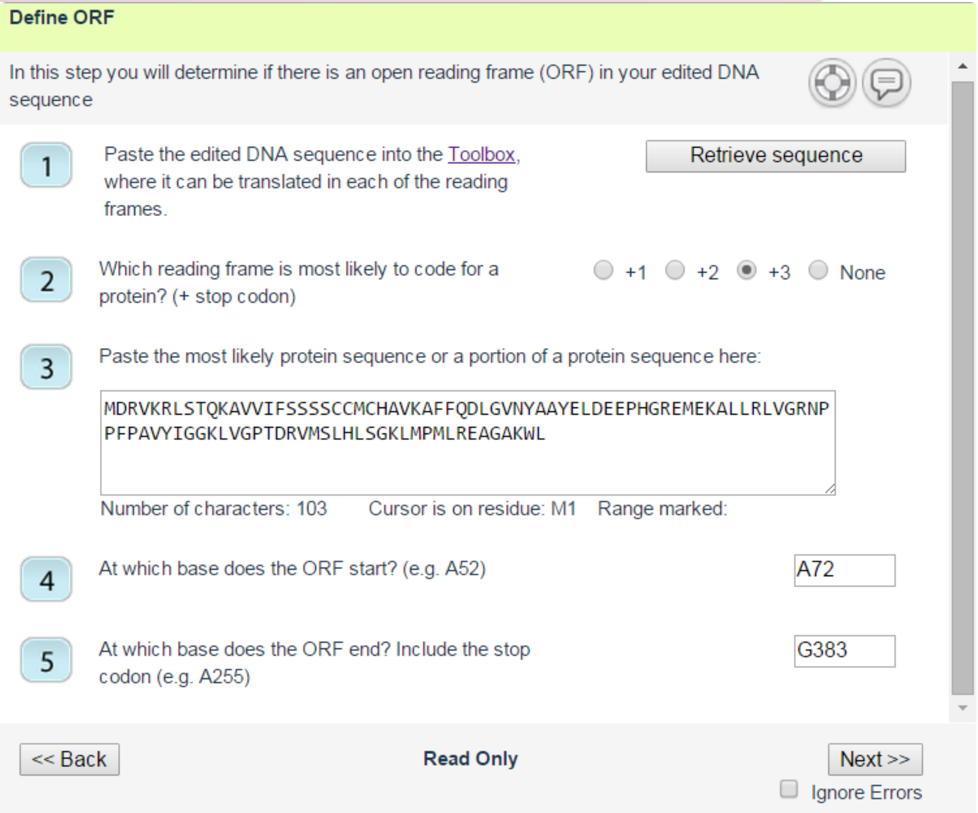
```
Query  72   MDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQDLGVNYAAYELDEEPHGREMEKALLRLV  251
            MDRV +L++QKAVVIFS SSCCMCH +K FF +LGVN   +ELDE+P G++ME+AL ++V
Sbjct  1    MDRVNKLASQKAVVIFSLSSCCMCHTIKRFFSELGVNPTIHELDEDPKGKDMERALAKMV  60

Query  252  GRNPPFPAVYIGGKLVGPTDRVMSLHLSGKLMPMLREAGAKWL  380
            GRNPP PAV+IGG LVGPTD++MSLHLSGKL+P+LR+AGA WL
Sbjct  61   GRNPPVPAVFIGGSLVGPTDKIMSLHLSGKLVPLLRDAGAIWL  103
```

72 -

ATGGACCGTGTGAAGAGGCTGAGCACGCAGAAGGCGGTGGTGATATTCAGCTCGAGCTCGTG
CTGCATGTGCCACGCAGTCAAGGCCTTCTTCCAGGATCTCGGGGTGAACTACGCCGCCTACG
AGCTCGACGAGGAACCCCACGGAAGGGAGATGGAGAAGGCTCTTCTCCGGCTAGTCGGCCG
GAACCCGCCATTTCCGGCAGTCTACATCGGCGGCAAGCTTGTCGGCCCGACAGACCGCGTCA
TGTCCCTCCATCTCAGTGGCAAGCTTATGCCCATGCTGCGGGAAGCAGGCGCTAAATGGCTGT
AG - 383

(Check that it matches Blastx)

Answer the Q's



Define ORF

In this step you will determine if there is an open reading frame (ORF) in your edited DNA sequence

1 Paste the edited DNA sequence into the Toolbox, where it can be translated in each of the reading frames.

2 Which reading frame is most likely to code for a protein? (+ stop codon)

○ +1 ○ +2 ● +3 ○ None

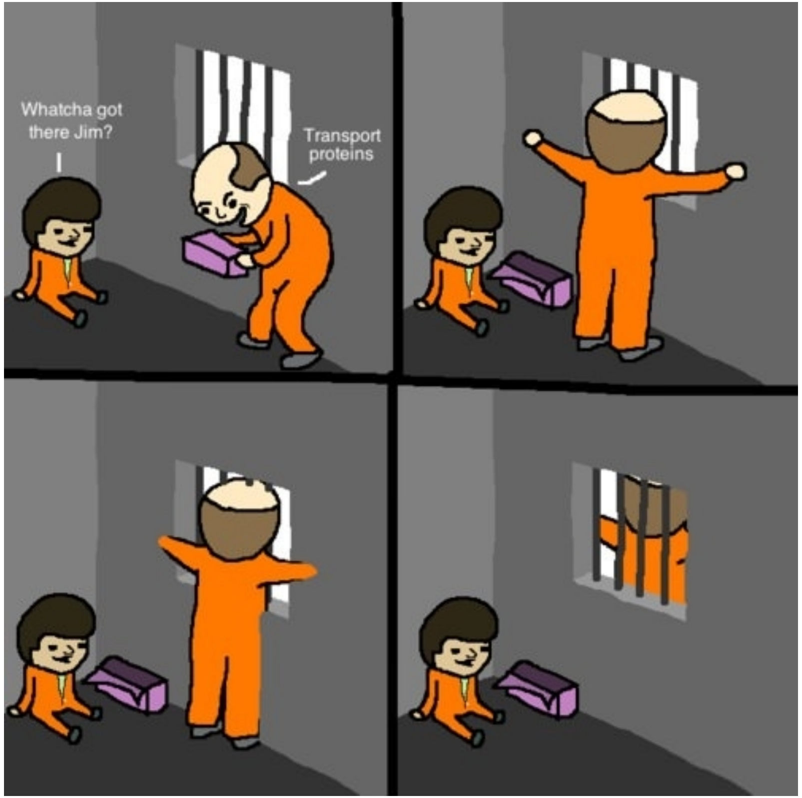
3 Paste the most likely protein sequence or a portion of a protein sequence here:

```
MDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQDLGVNYAAYELDEEPHGREMEKALLRLVGRNP
PFPAVYIGGKLVGPTDRVMSLHLSGKLMPMLREAGAKWL
```

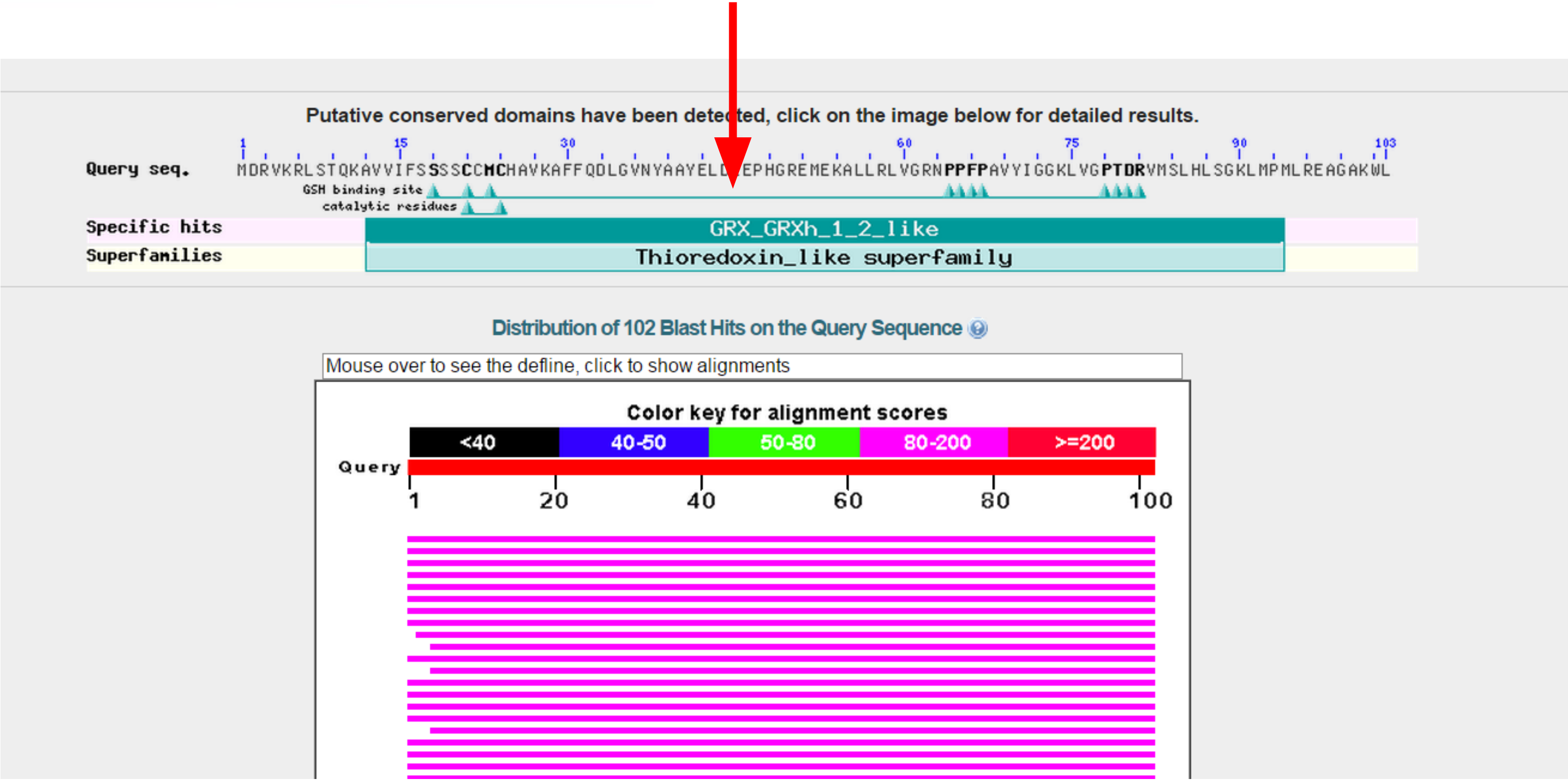
Number of characters: 103 Cursor is on residue: M1 Range marked:

4 At which base does the ORF start? (e.g. A52) A72

5 At which base does the ORF end? Include the stop codon (e.g. A255) G383



Blastp!!!



Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. MDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQDLGVNYAAYELD[illegible]EPHGREMEKALLRLVGRNPPFPAVYIGGKLVGPTDRVMSLHLSGKLMPMLREAGAKWL

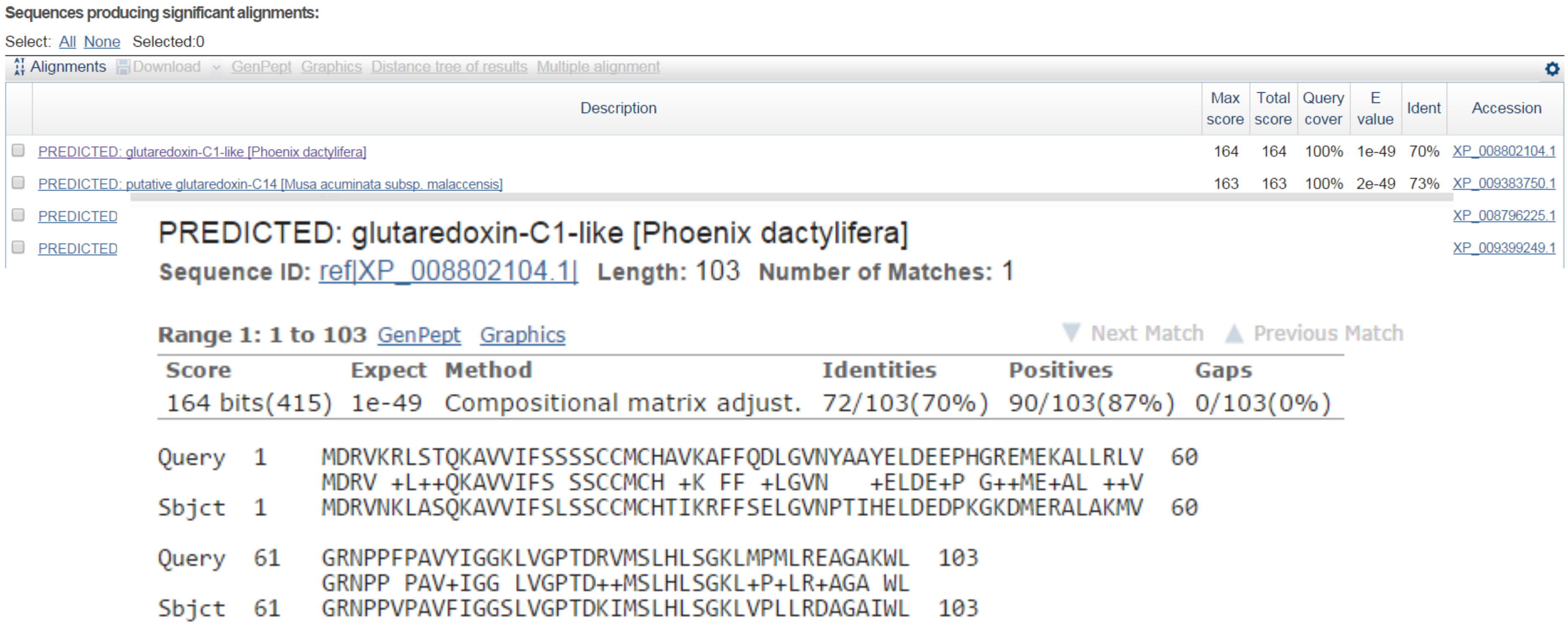
Distribution of 102 Blast Hits on the Query Sequence

Mouse over to see the defline, click to show alignments

Color key for alignment scores

<40	40-50	50-80	80-200	>=200

Blastp!!!



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	PREDICTED: glutaredoxin-C1-like [Phoenix dactylifera]	164	164	100%	1e-49	70%	XP_008802104.1
☐	PREDICTED: putative glutaredoxin-C14 [Musa acuminata subsp. malaccensis]	163	163	100%	2e-49	73%	XP_009383750.1
☐	PREDICTED						XP_008796225.1
☐	PREDICTED						XP_009399249.1

PREDICTED: glutaredoxin-C1-like [Phoenix dactylifera]

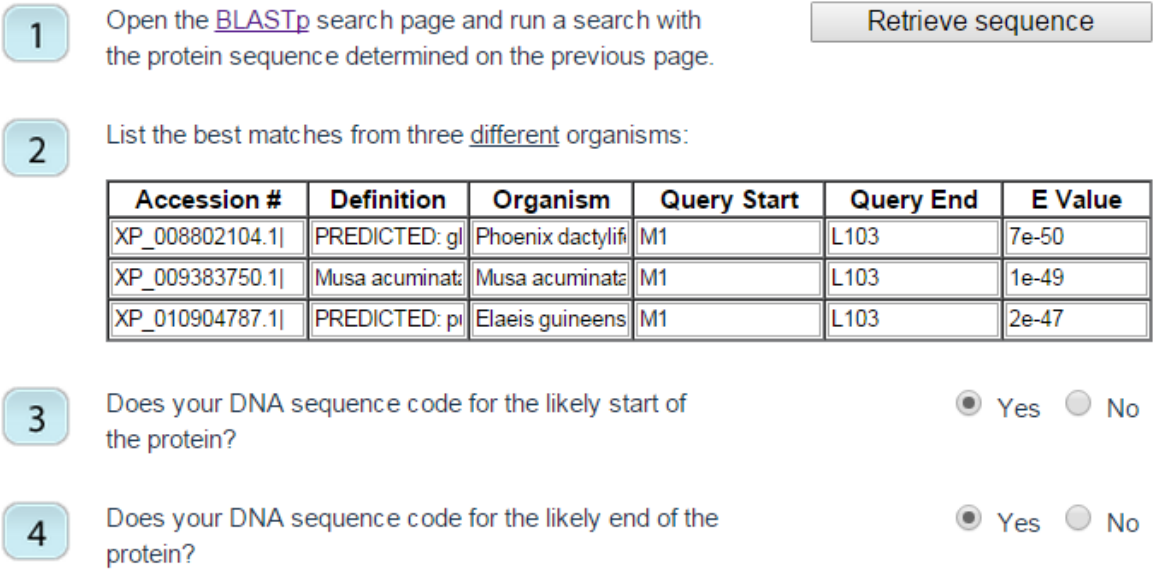
Sequence ID: ref|XP_008802104.1| **Length:** 103 **Number of Matches:** 1

Range 1: 1 to 103 GenPept Graphics

Score	Expect	Method	Identities	Positives	Gaps
164 bits(415)	1e-49	Compositional matrix adjust.	72/103(70%)	90/103(87%)	0/103(0%)

```
Query  1    MDRVKRLSTQKAVVIFSSSSCCMCHAVKAFFQDLGVNYAAYELDEEPHGREMEKALLRLV  60
            MDRV +L++QKAVVIFS SSCCMCH +K FF +LGVN   +ELDE+P G++ME+AL ++V
Sbjct  1    MDRVNKLASQKAVVIFSLSSCCMCHTIKRFFSELGVNPTIHELDEDPKGKDMERALAKMV  60

Query  61   GRNPPFPAVYIGGKLVGPTDRVMSLHLSGKLMPMLREAGAKWL  103
            GRNPP PAV+IGG LVGPTD++MSLHLSGKL+P+LR+AGA WL
Sbjct  61   GRNPPVPAVFIGGSLVGPTDKIMSLHLSGKLVPLLRDAGAIWL  103
```

Answer the Q’s



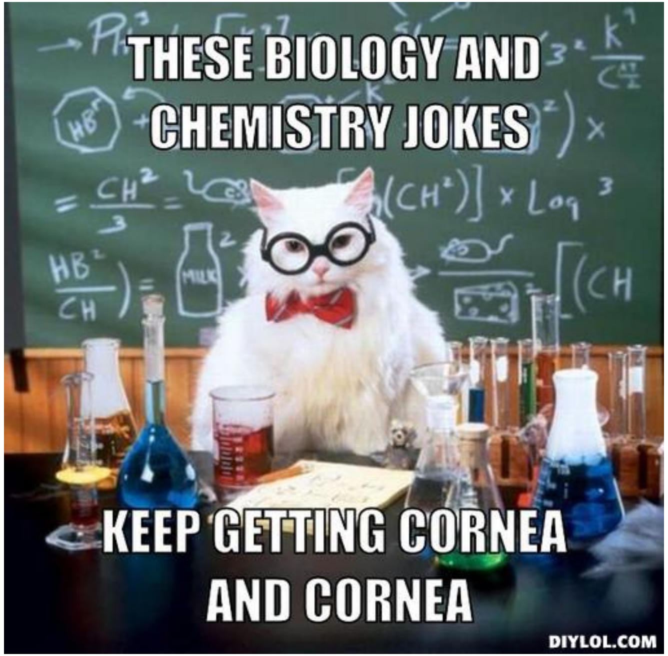
1 Open the BLASTp search page and run a search with the protein sequence determined on the previous page.

2 List the best matches from three different organisms:

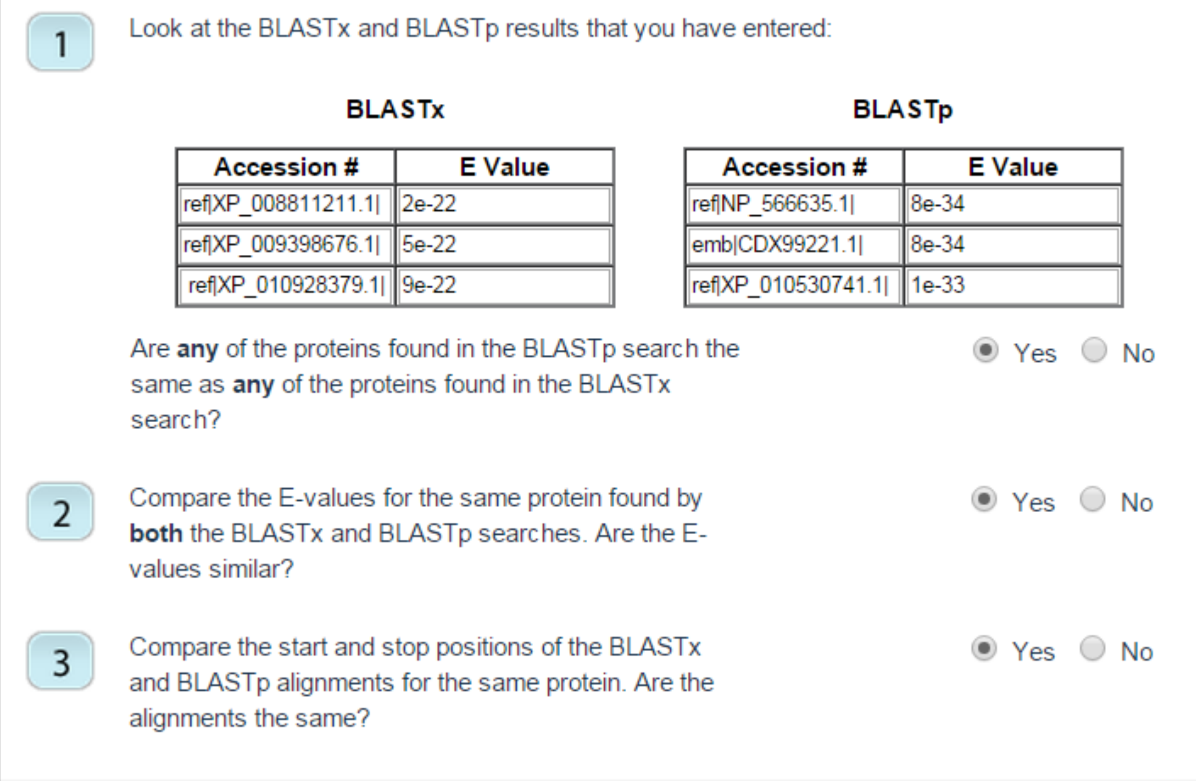
Accession #	Definition	Organism	Query Start	Query End	E Value
XP_008802104.1|	PREDICTED: gl	Phoenix dactylif	M1	L103	7e-50
XP_009383750.1|	Musa acuminat	Musa acuminata	M1	L103	1e-49
XP_010904787.1|	PREDICTED: pı	Elaeis guineens	M1	L103	2e-47

3 Does your DNA sequence code for the likely start of the protein? ● Yes ○ No

4 Does your DNA sequence code for the likely end of the protein? ● Yes ○ No



Blastp v.s. Blastx



1 Look at the BLASTx and BLASTp results that you have entered:

BLASTx

Accession #	E Value
ref|XP_008811211.1|	2e-22
ref|XP_009398676.1|	5e-22
ref|XP_010928379.1|	9e-22

BLASTp

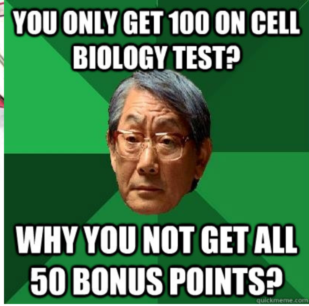
Accession #	E Value
ref|NP_566635.1|	8e-34
emb|CDX99221.1|	8e-34
ref|XP_010530741.1|	1e-33

Are **any** of the proteins found in the BLASTp search the same as **any** of the proteins found in the BLASTx search? ● Yes ○ No

2 Compare the E-values for the same protein found by **both** the BLASTx and BLASTp searches. Are the E-values similar? ● Yes ○ No

3 Compare the start and stop positions of the BLASTx and BLASTp alignments for the same protein. Are the alignments the same? ● Yes ○ No

UR DONE WITH BLAST



NOW U CAN GO ANALYZE THE STUFF!!!!!!!!!!!!!!!!!!!! <33333333333