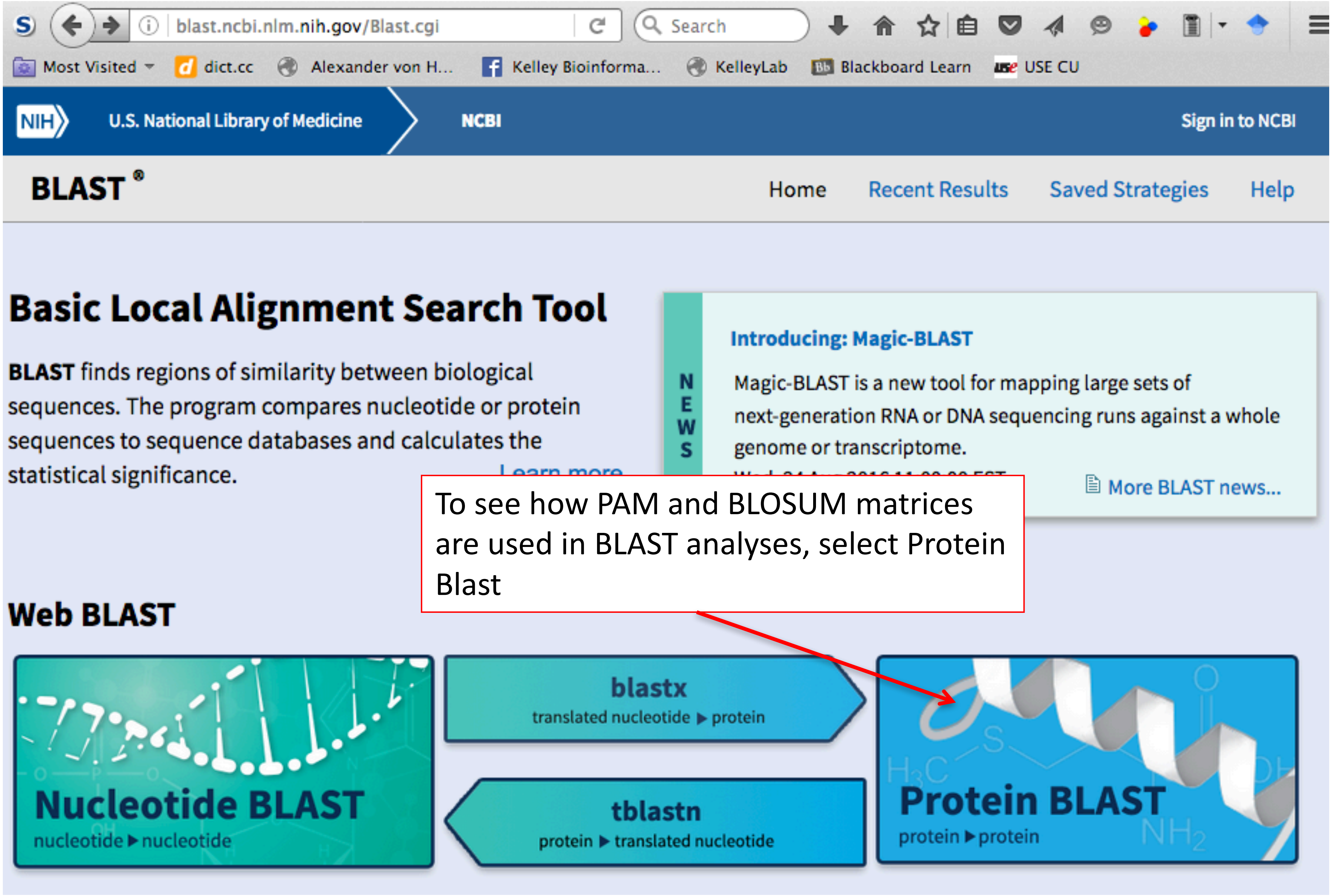


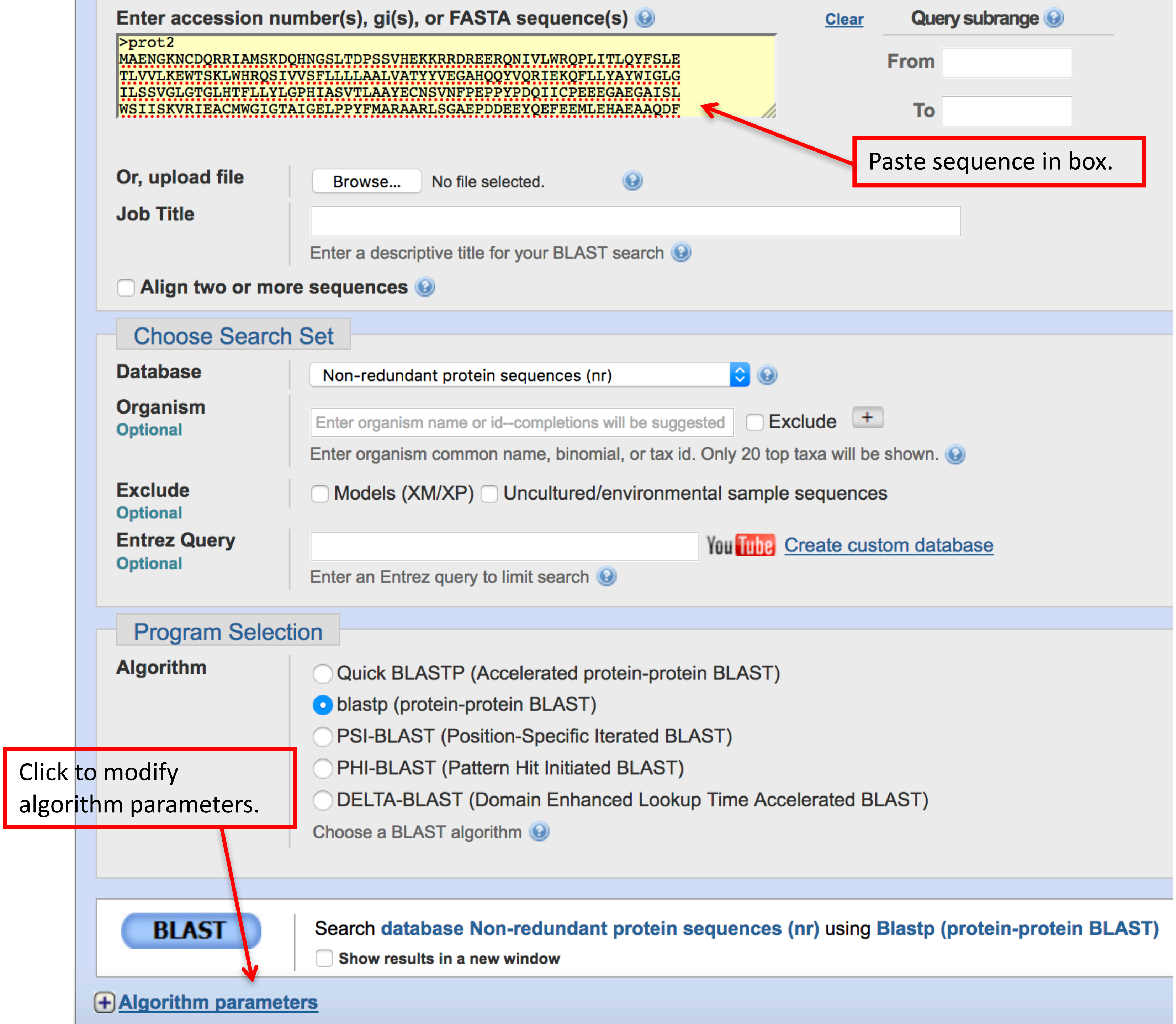
Basic Local Alignment Search Tool

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

Introducing: Magic-BLAST

Magic-BLAST is a new tool for mapping large sets of next-generation RNA or DNA sequencing runs against a whole genome or transcriptome.

To see how PAM and BLOSUM matrices are used in BLAST analyses, select Protein Blast



Paste sequence in box.

Click to modify algorithm parameters.

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

```
>prot2
MAENGKNCDQRRIAMSKDQHNGSLTDPSSVHEKKRRDREERQNIVLWRQPLITLQYFSLE
TLVVLKEWTSKLWHRQSIVVSFLLLLAALVATYYVEGAHQQYVQRIEKQFLLYAYWIGLG
ILSSVGLGTGLHTFLLYLGPHIASVTLAAYECNSVNFPEPPYPDQIICPEEEGAEGAISL
WSIISKVRIEACMWGIGTAIGELPPYFMARAARLSGAEPDDEEYQEFEEMLEHAEAAQDF
```

Query subrange

From:

To:

Or, upload file: No file selected.

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): Enter organism name or id--completions will be suggested ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

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

Entrez Query (Optional): Enter an Entrez query to limit search

Program Selection

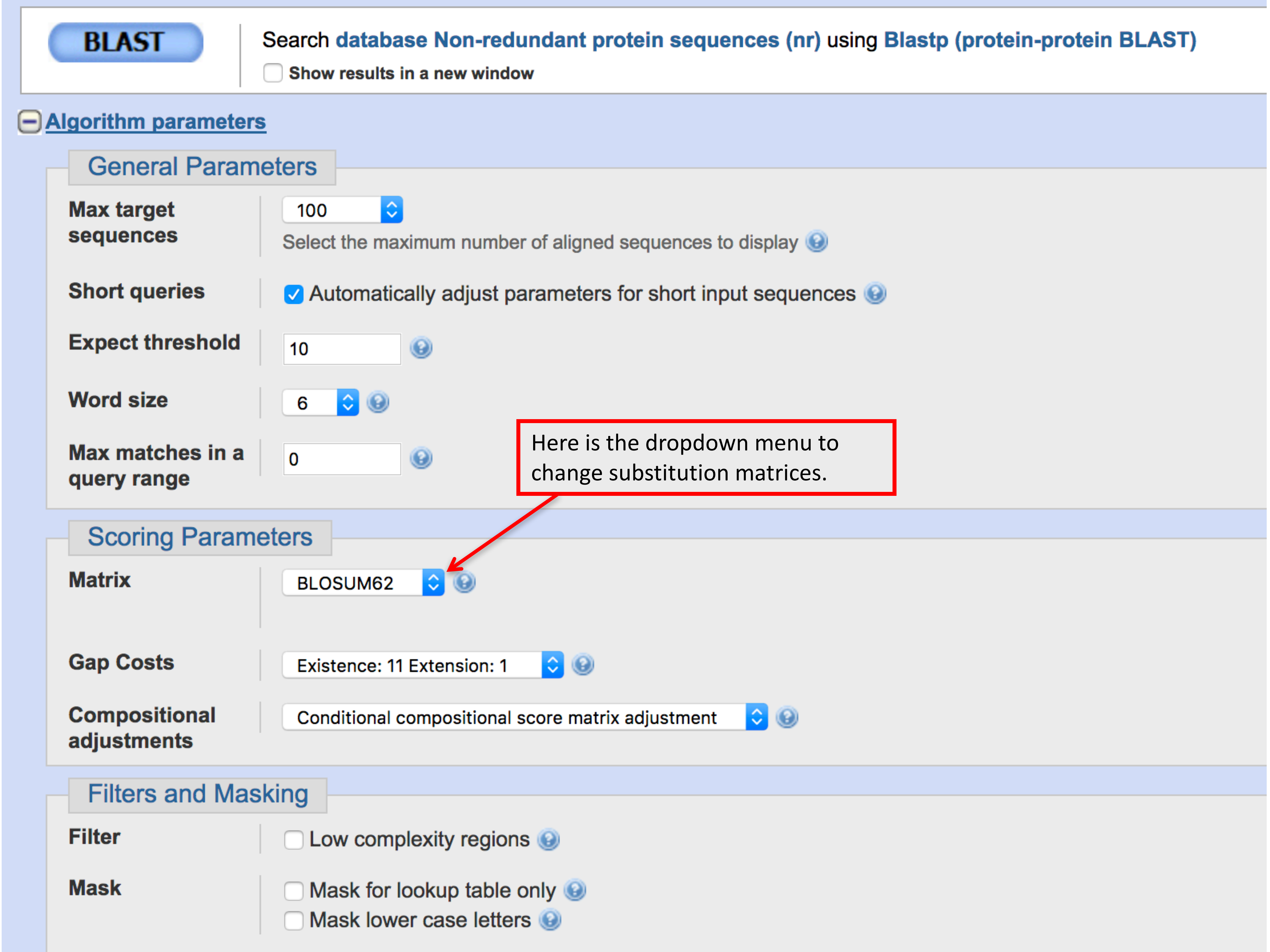
Algorithm:

○ Quick BLASTP (Accelerated protein-protein BLAST)
● blastp (protein-protein BLAST)
○ PSI-BLAST (Position-Specific Iterated BLAST)
○ PHI-BLAST (Pattern Hit Initiated BLAST)
○ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm

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

☐ Show results in a new window



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

☐ Show results in a new window

Algorithm parameters

General Parameters

Max target sequences: 100

Select the maximum number of aligned sequences to display

Short queries: ☑ Automatically adjust parameters for short input sequences

Expect threshold: 10

Word size: 6

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

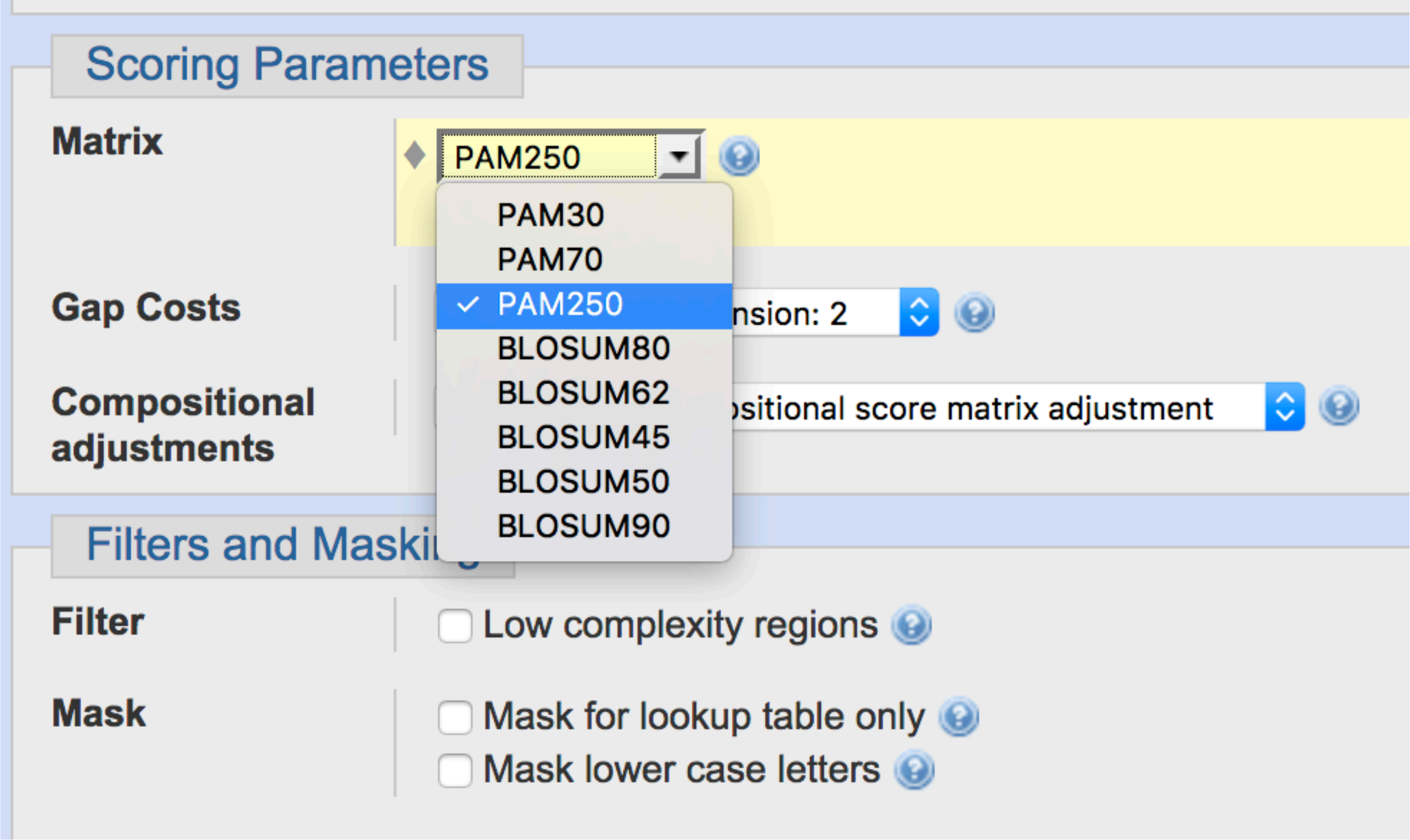
Filters and Masking

Filter: ☐ Low complexity regions

Mask: ☐ Mask for lookup table only

☐ Mask lower case letters

Here is the dropdown menu to change substitution matrices.



Scoring Parameters

Matrix PAM250

Gap Costs nsion: 2

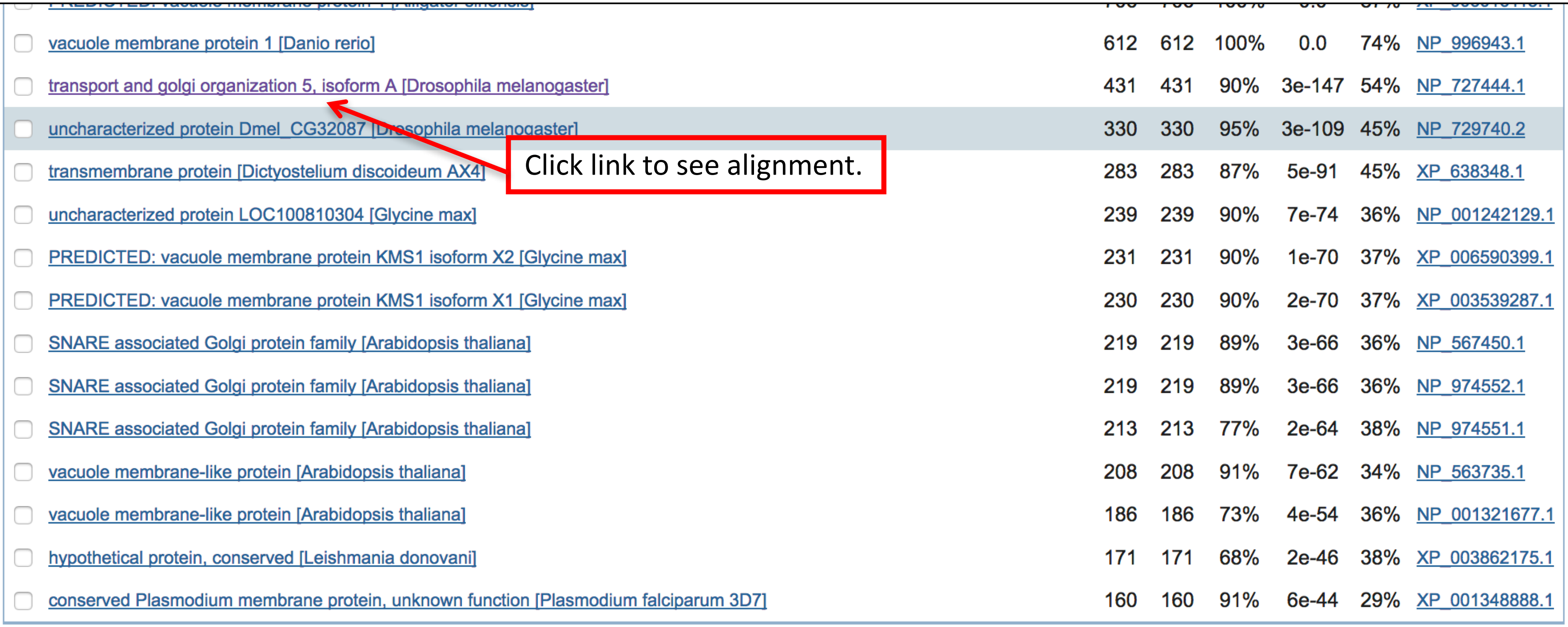
Compositional adjustments sitional score matrix adjustment

Filters and Maski

Filter ☐ Low complexity regions

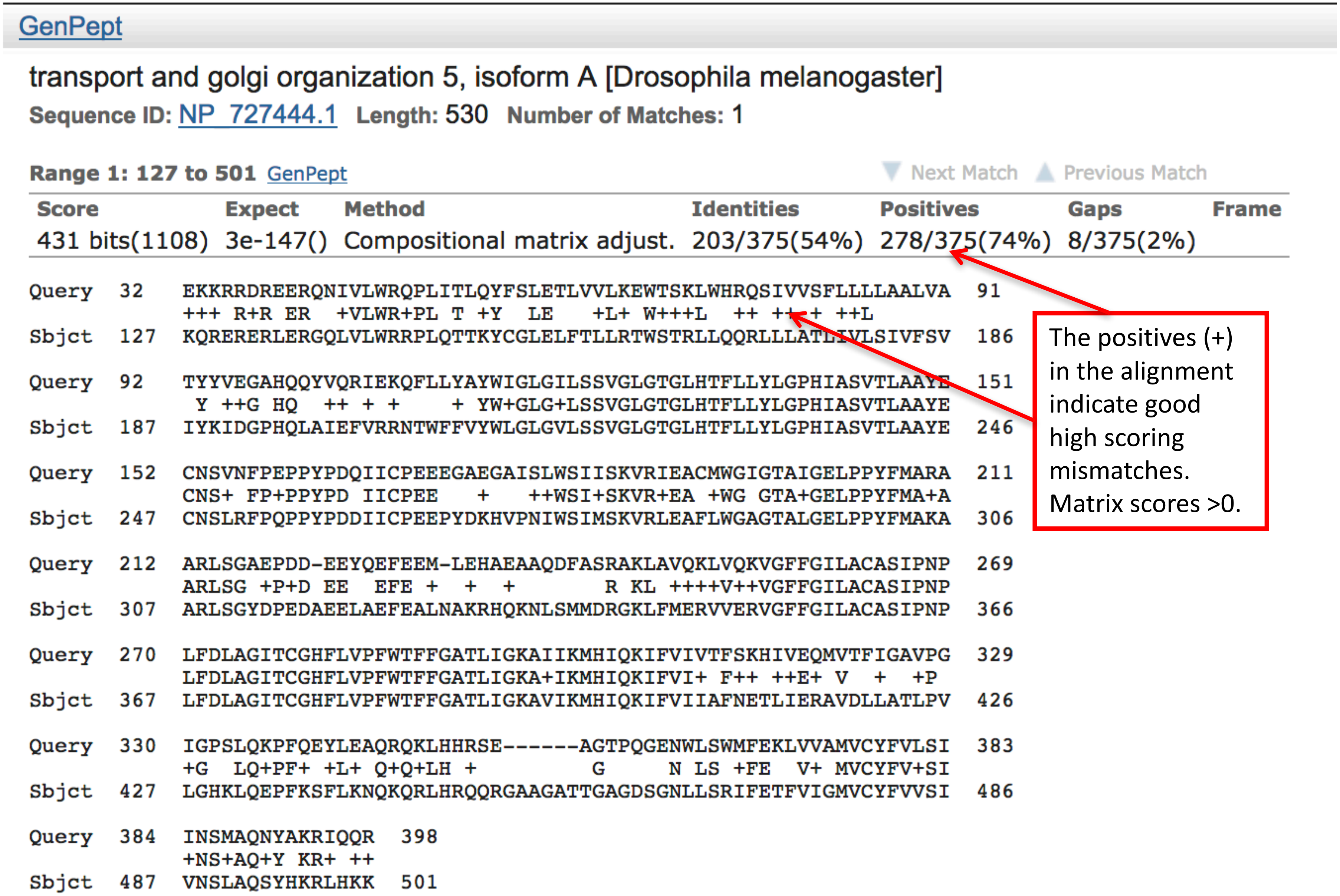
Mask ☐ Mask for lookup table only

☐ Mask lower case letters



Click link to see alignment.

☐	vacuole membrane protein 1 [Danio rerio]	612	612	100%	0.0	74%	NP_996943.1
☐	transport and golgi organization 5, isoform A [Drosophila melanogaster]	431	431	90%	3e-147	54%	NP_727444.1
☐	uncharacterized protein Dmel_CG32087 [Drosophila melanogaster]	330	330	95%	3e-109	45%	NP_729740.2
☐	transmembrane protein [Dictyostelium discoideum AX4]	283	283	87%	5e-91	45%	XP_638348.1
☐	uncharacterized protein LOC100810304 [Glycine max]	239	239	90%	7e-74	36%	NP_001242129.1
☐	PREDICTED: vacuole membrane protein KMS1 isoform X2 [Glycine max]	231	231	90%	1e-70	37%	XP_006590399.1
☐	PREDICTED: vacuole membrane protein KMS1 isoform X1 [Glycine max]	230	230	90%	2e-70	37%	XP_003539287.1
☐	SNARE associated Golgi protein family [Arabidopsis thaliana]	219	219	89%	3e-66	36%	NP_567450.1
☐	SNARE associated Golgi protein family [Arabidopsis thaliana]	219	219	89%	3e-66	36%	NP_974552.1
☐	SNARE associated Golgi protein family [Arabidopsis thaliana]	213	213	77%	2e-64	38%	NP_974551.1
☐	vacuole membrane-like protein [Arabidopsis thaliana]	208	208	91%	7e-62	34%	NP_563735.1
☐	vacuole membrane-like protein [Arabidopsis thaliana]	186	186	73%	4e-54	36%	NP_001321677.1
☐	hypothetical protein, conserved [Leishmania donovani]	171	171	68%	2e-46	38%	XP_003862175.1
☐	conserved Plasmodium membrane protein, unknown function [Plasmodium falciparum 3D7]	160	160	91%	6e-44	29%	XP_001348888.1



GenPept

transport and golgi organization 5, isoform A [Drosophila melanogaster]

Sequence ID: NP_727444.1 **Length:** 530 **Number of Matches:** 1

Range 1: 127 to 501 GenPept

Score	Expect	Method	Identities	Positives	Gaps	Frame
431 bits(1108)	3e-147()	Compositional matrix adjust.	203/375(54%)	278/375(74%)	8/375(2%)	

```
Query  32   EKKRRDREERQNIVLWRQPLITLQYFSLETLVVLKEWTSKLWHRQSIVVSFLLLLAALVA  91
            +++ R+R ER  +VLWR+PL T +Y  LE   +L+ W+++L  ++ ++ + ++L
Sbjct  127  KQRERERLERGQLVLWRRPLQTTKYCGLELFTLLRTWSTRLLQQRLLLATLIVLSIVFSV  186

Query  92   TYYVEGAHQQYVQRIEKQFLLYAYWIGLGILSSVGLGTGLHTFLLYLGPHIASVTLAAYE  151
             Y ++G HQ  ++ + +    + YW+GLG+LSSVGLGTGLHTFLLYLGPHIASVTLAAYE
Sbjct  187  IYKIDGPHQLAIEFVRRNTWFFVYWLGLGVLSSVGLGTGLHTFLLYLGPHIASVTLAAYE  246

Query  152  CNSVNFPEPPYPDQIICPEEEGAEGAISLWSIISKVRIEACMWGIGTAIGELPPYFMARA  211
            CNS+ FP+PPYPD IICPEE   +   ++WSI+SKVR+EA +WG GTA+GELPPYFMA+A
Sbjct  247  CNSLRFPQPPYPDDIICPEEPYDKHVPNIWSIMSKVRLEAFLWGAGTALGELPPYFMAKA  306

Query  212  ARLSGAEPDD-EEYQEFEEM-LEHAEAAQDFASRAKLAVQKLVQKVGFFGILACASIPNP  269
            ARLSG +P+D EE  EFE +  +  +       R KL ++++V++VGFFGILACASIPNP
Sbjct  307  ARLSGYDPEDAEELAEFEALNAKRHQKNLSMMDRGKLFMERVVERVGFFGILACASIPNP  366

Query  270  LFDLAGITCGHFLVPFWTFFGATLIGKAIIKMHIQKIFVIVTFSKHIVEQMVTFIGAVPG  329
            LFDLAGITCGHFLVPFWTFFGATLIGKA+IKMHIQKIFVI+ F++ ++E+ V  +  +P
Sbjct  367  LFDLAGITCGHFLVPFWTFFGATLIGKAVIKMHIQKIFVIIAFNETLIERAVDLLATLPV  426

Query  330  IGPSLQKPFQEYLEAQRQKLHHRSE------AGTPQGENWLSWMFEKLVVAMVCYFVLSI  383
            +G  LQ+PF+ +L+ Q+Q+LH +         G     N LS +FE  V+ MVCYFV+SI
Sbjct  427  LGHKLQEPFKSFLKNQKQRLHRQQRGAAGATTGAGDSGNLLSRIFETFVIGMVCYFVVSI  486

Query  384  INSMAQNYAKRIQQR  398
            +NS+AQ+Y KR+ ++
Sbjct  487  VNSLAQSYHKRLHKK  501
```

The positives (+) in the alignment indicate good high scoring mismatches. Matrix scores >0.