DATE: 1-11-21

WEBLEM 11

Introduction To Sequence Alignment

BLAST:

A sequence similarity search often provides the first information about a new DNA or protein sequence. A search allows scientists to infer the function of a sequence from similar sequences. There are many ways of performing a sequence similarity search, but probably the most popular method is the "Basic Local Alignment Search Tool" (BLAST) (1, 2). BLAST uses heuristics to produce results quickly. It also calculates an "expect value" that estimates how many matches would have occurred at a given score by chance, which can aid a user in judging how much confidence to have in an alignment.

As the name implies, BLAST performs "local" alignments. Most proteins are modular in nature, with one or more functional domains occurring within a protein. The same domains may also occur in proteins from different species. The BLAST algorithm is tuned to find these domains or shorter stretches of sequence similarity. The local alignment approach also means that an mRNA can be aligned with a piece of genomic DNA, as is frequently required in genome assembly and analysis. If instead BLAST started out by attempting to align two sequences over their entire lengths (known as a global alignment), fewer similarities would be detected, especially with respect to domains and motifs.

1. BLASTN:

The BLASTN nucleotide-nucleotide search looks for more distant sequences.

2. BLASTP:

BLASTP performs protein-protein sequence comparison, and its algorithm is the basis of many other types of BLAST searches such a BLASTX and TBLASTN.

3. PHI-BLAST:

PHI-BLAST (Pattern-Hit Initiated BLAST) is a search program that combines matching of regular expressions with local alignments surrounding surrounding the match. The most important features of the program have been incorporated into the BLAST software framework partly for user convenience and partly so that PHI-BLAST may be combined seamlessly with PSI-Blast. Other features that do not fit into the BLAST framework will be released later as a separate and/or separate web page query option.

4. PSI-BLAST:

PSI-BLAST first performs a BLASTP search to collect information that it then uses to produce a Position-Specific-Scoring-Matrix (PSSM). A PSSM for a query of length N is an N x 20 matrix. Each of the N columns corresponds to a letter in the query, and each column contains 20 rows. Each row corresponds to a specific residue and describes the probability of related sequences having that residue at that position. PSI-BLAST can then search a database of protein sequences with this PSSM.

FASTA:

At EMBL-EBI, we make the world's public biological data freely available to the scientific community via a range of services and tools, perform basic research and provide professional training in bioinformatics. We are part of the European Molecular Biology Laboratory (EMBL), an international, innovative and interdisciplinary research organization funded by over 20 member states, prospect and associate member states.

ASTA (pronounced FAST-AYE) is a suite of programs for searching nucleotide or protein databases with a query sequence. FASTA itself performs a local heuristic search of a protein or nucleotide database for a query of the same type. FASTX and FASTY translate a nucleotide query for searching a protein database. TFASTX and TFASTY translate a nucleotide database to be searched with a protein query. Optimal searches are available with the programs SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query against local database).

5. FASTAp;

This tool provides sequence similarity searching against protein databases using the FASTA suite of programs. FASTA provides a heuristic search with a protein query. FASTX and FASTY translate a DNA query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).

6. FASTAn:

This tool provides sequence similarity searching against nucleotide databases using the FASTA suite of programs. FASTA provides a heuristic search with a nucleotide query. TFASTX and TFASTY translate the DNA database for searching with a protein query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).

DATE: 1-11-21

WEBLEM 11/a

(URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi)

Aim:

To study the Query 'Keratin' in BLASTn.

Introduction:

As the name implies, BLAST performs "local" alignments. Most proteins are modular in nature, with one or more functional domains occurring within a protein. The same domains may also occur in proteins from different species. The BLAST algorithm is tuned to find these domains or shorter stretches of sequence similarity. The local alignment approach also means that an mRNA can be aligned with a piece of genomic DNA, as is frequently required in genome assembly and analysis. If instead BLAST started out by attempting to align two sequences over their entire lengths (known as a global alignment), fewer similarities would be detected, especially with respect to domains and motifs.

- The BLASTN nucleotide-nucleotide search looks for more distant sequences.

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.
Rhodopsin was discovered in 1876 by German physiologist Franz Christian Boll, who observed that the normally reddish purple frog retina turned pale in bright light. The fading of colour was later attributed to the destruction of rhodopsin, via a process known as bleaching. Bleaching and the subsequent regeneration of rhodopsin are major steps in the visual cycle—the series of biochemical reactions that is critical for vision in low light.

Methodology:

1. Open the homepage of blast.
2. After that click the Blastn.
3. Enter the sequence for keratin taken by Genebank.
4. Open the result page.
5. Interpret the result.

Observation:

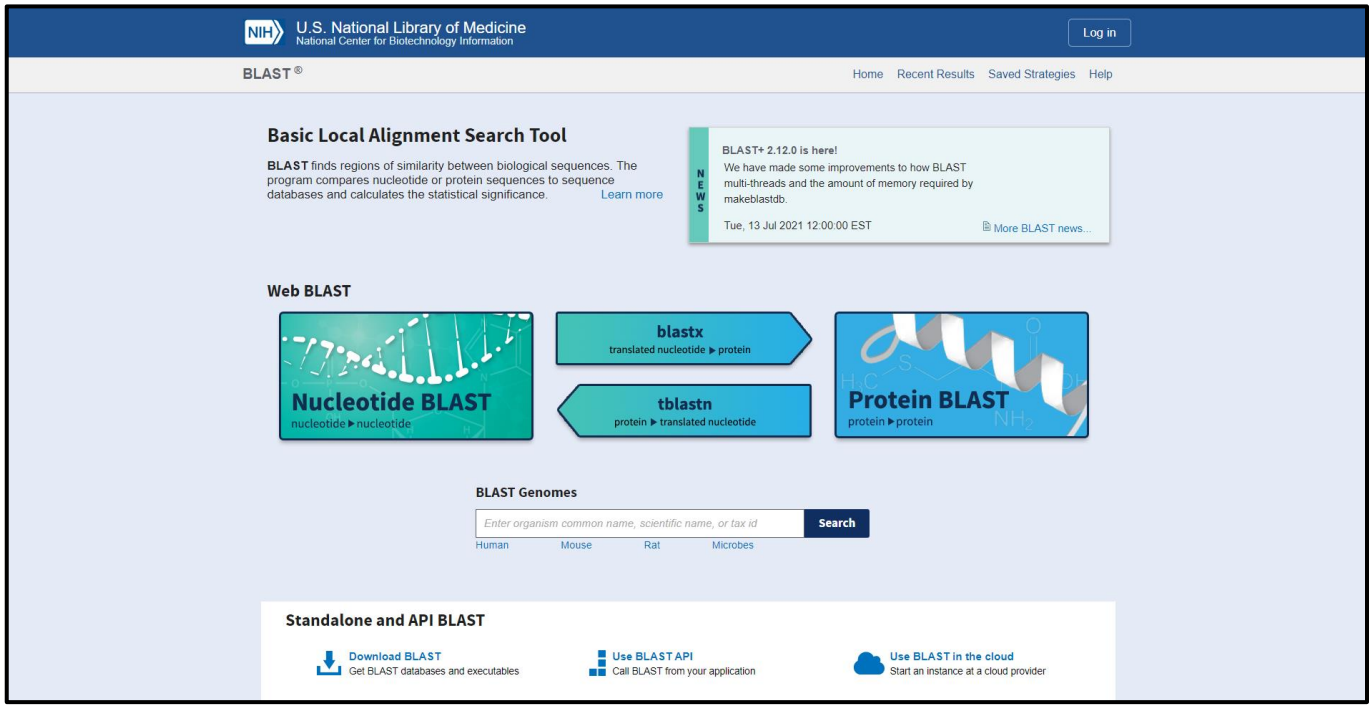


Fig1. Homepage of BLAST

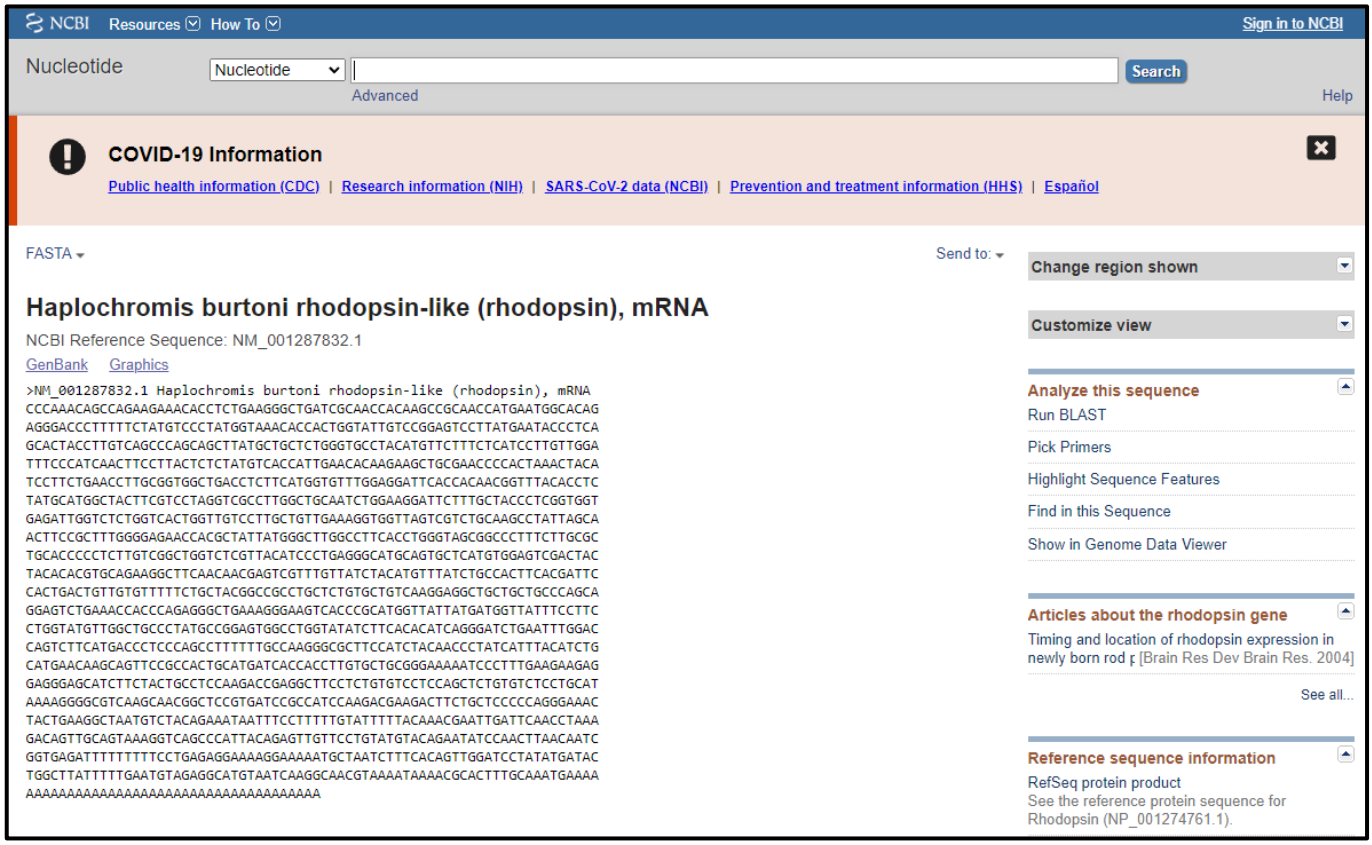


Fig2. Fasta sequence taken from genebank

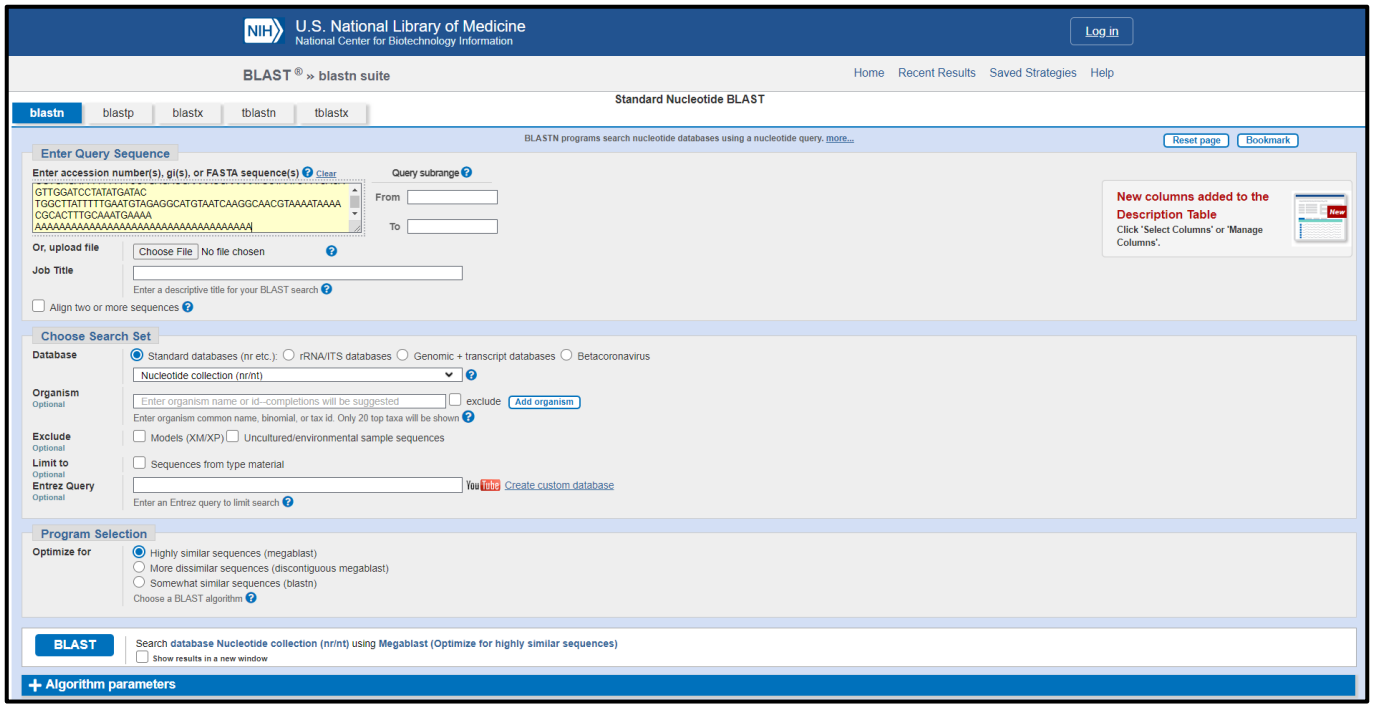


Fig3. Search bar for blastn with sequence.

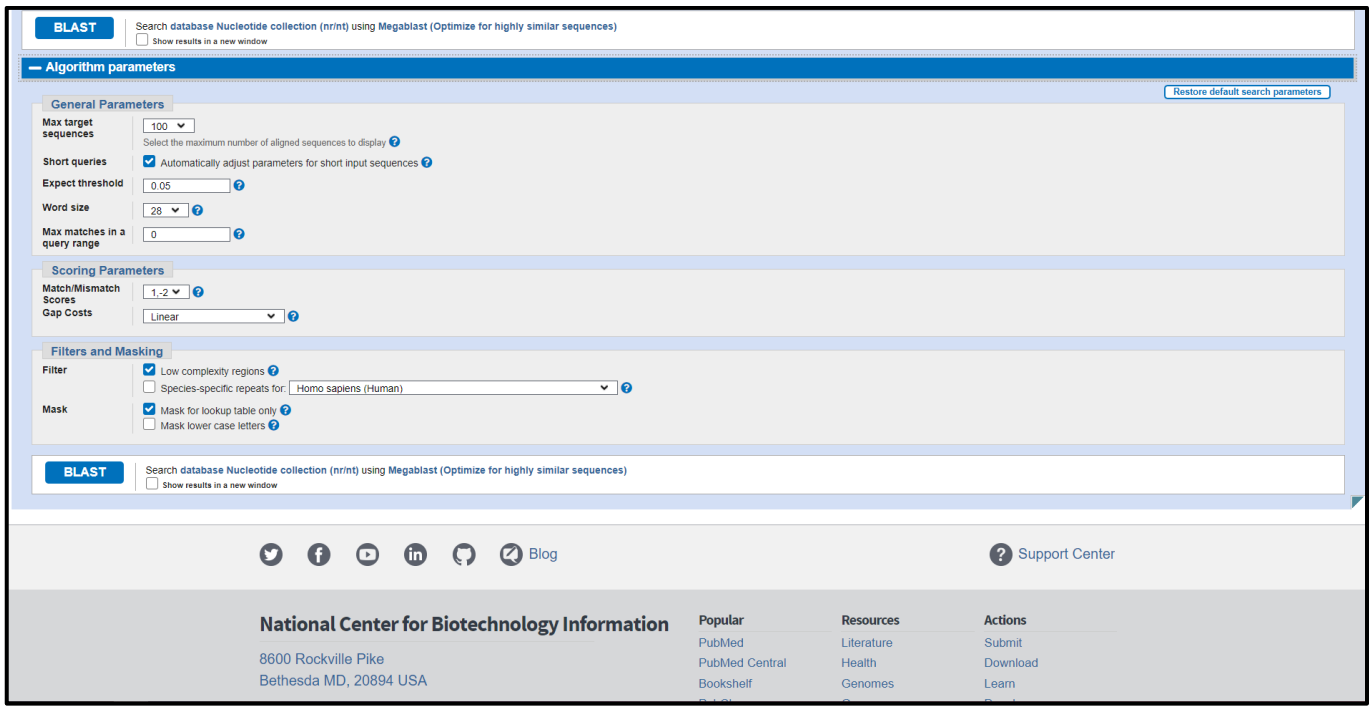


Fig4. Algorithm parameters in blastn

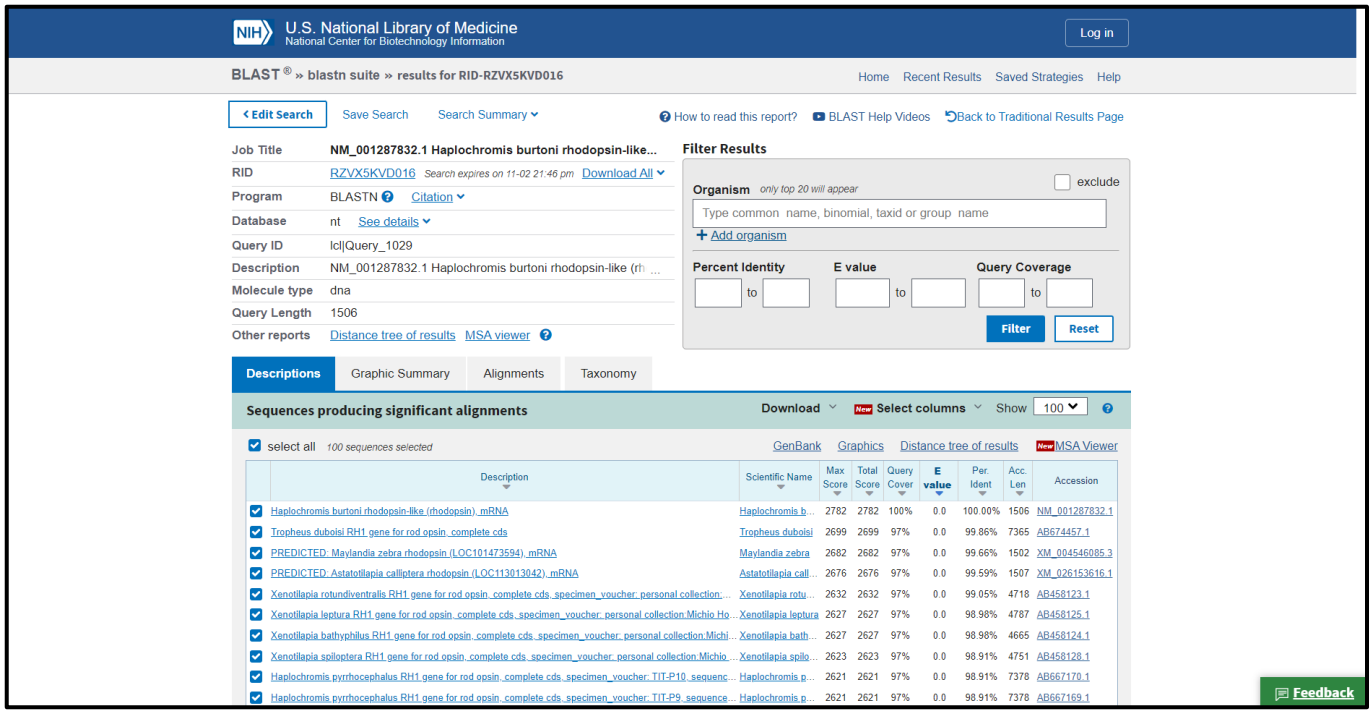


Fig5. Result page (header file) for Rhodopsin.

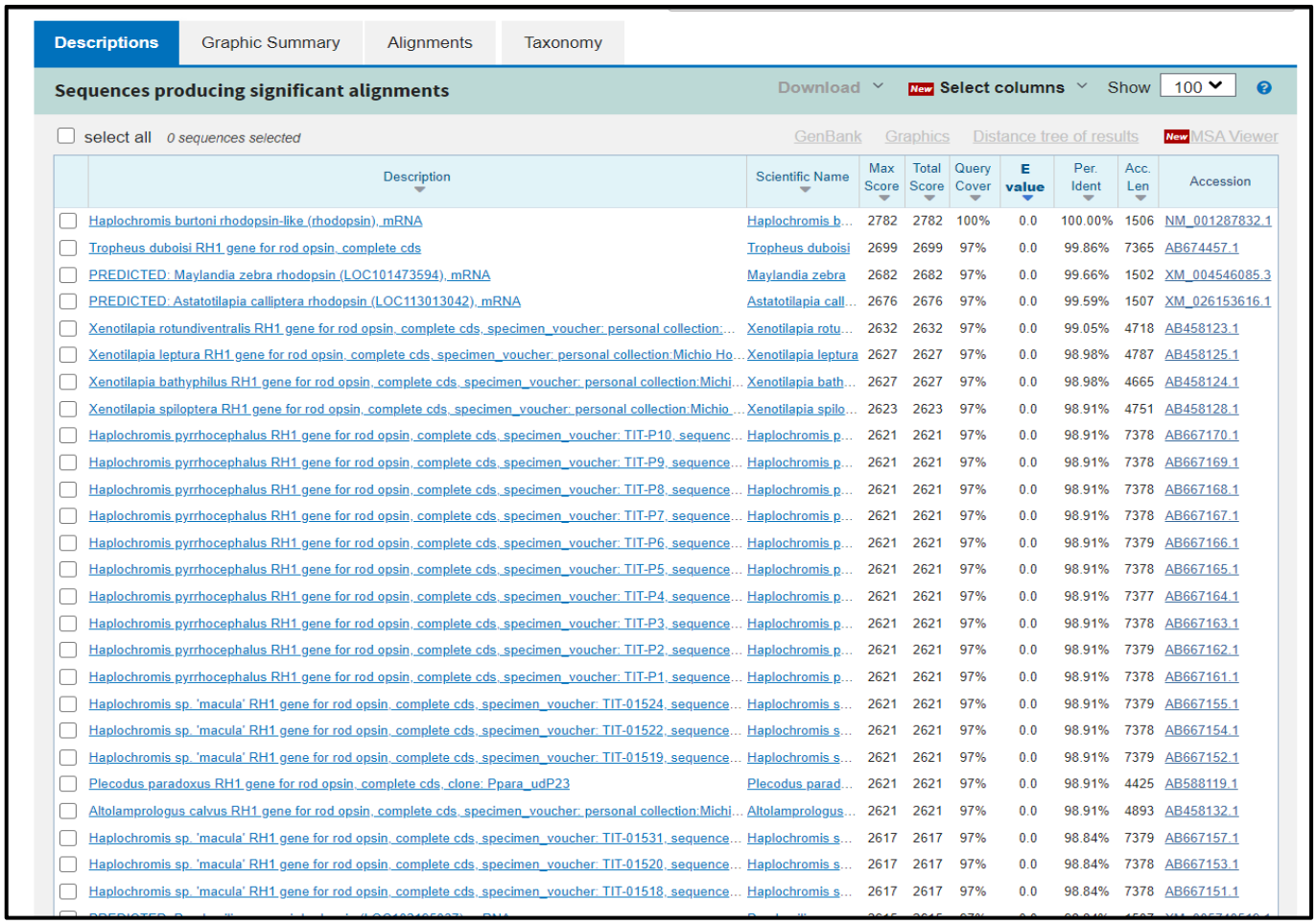


Fig6. Description for query Rhodopsin

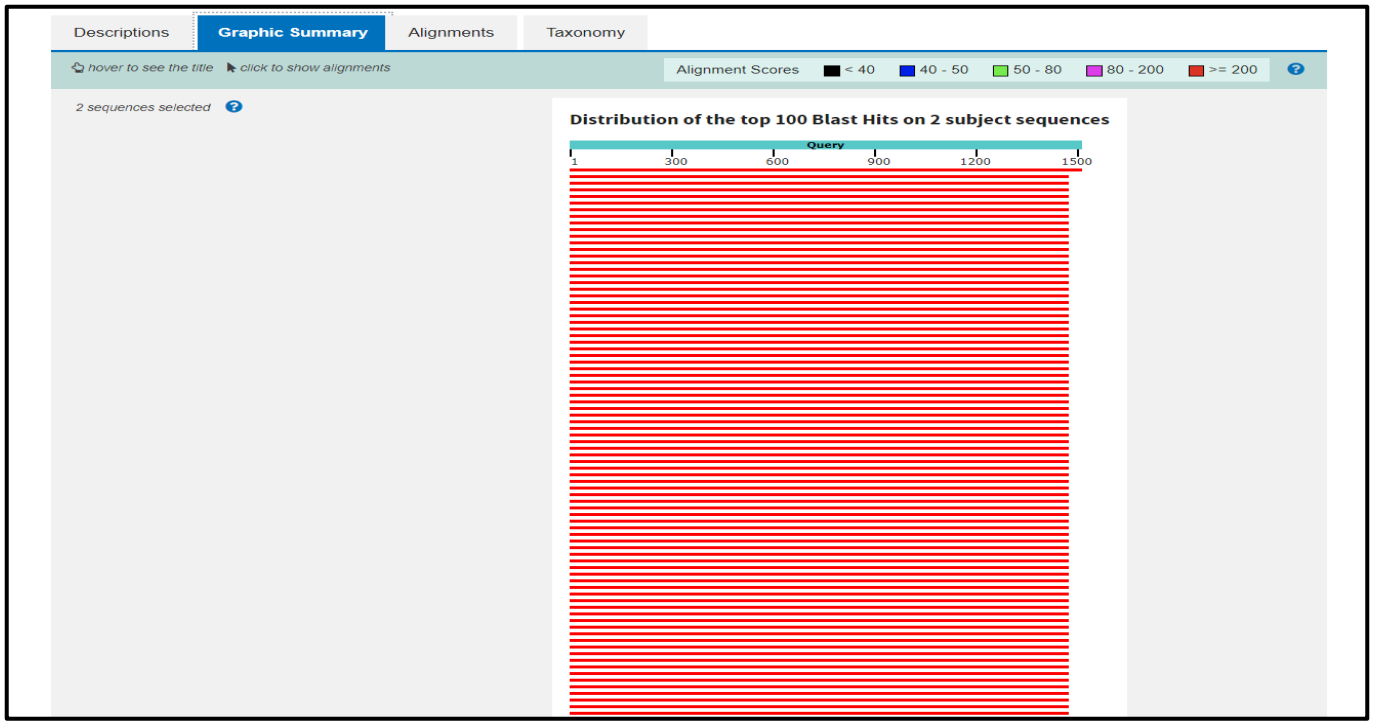


Fig7. Graphic Summary for query Rhodopsin

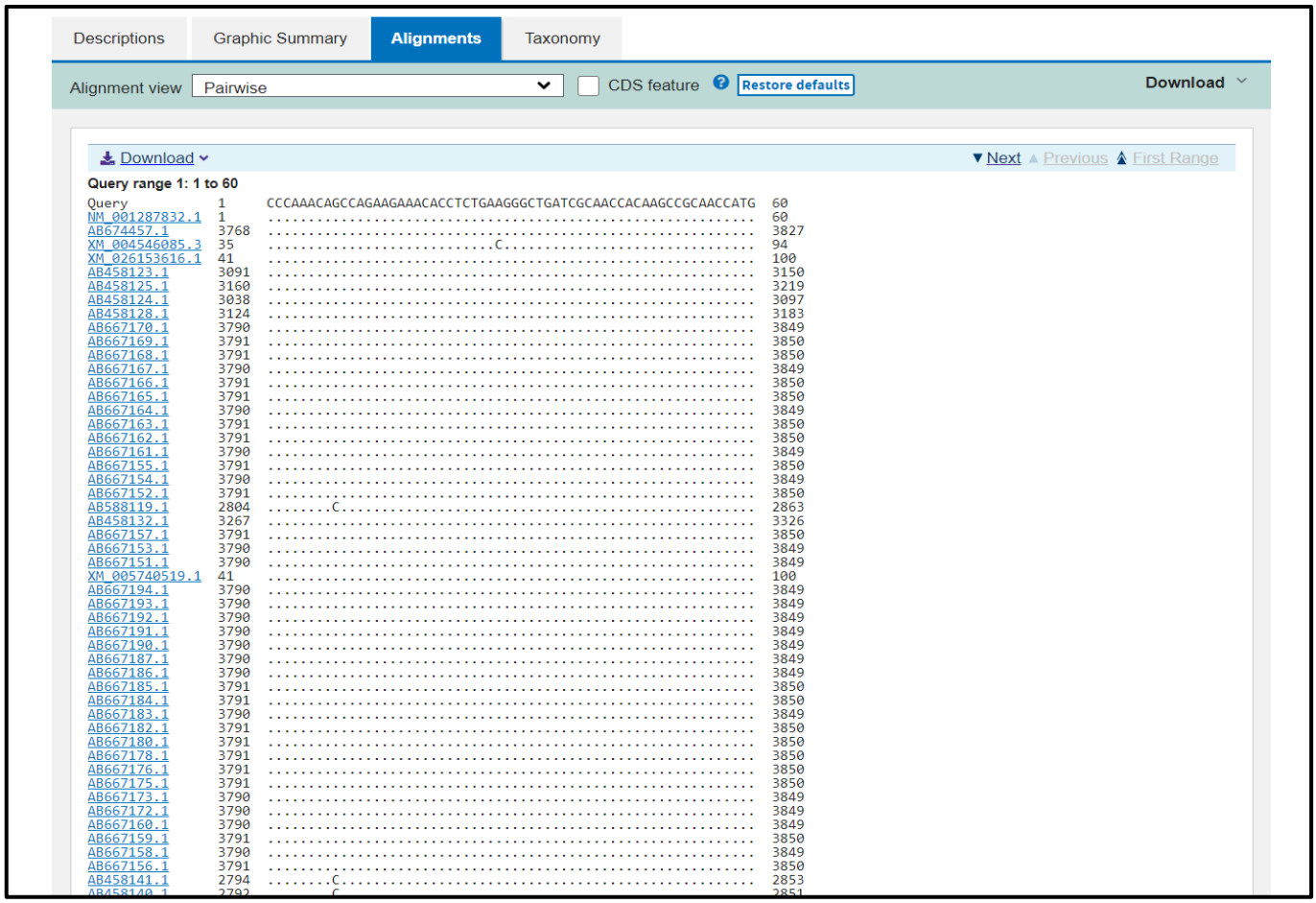
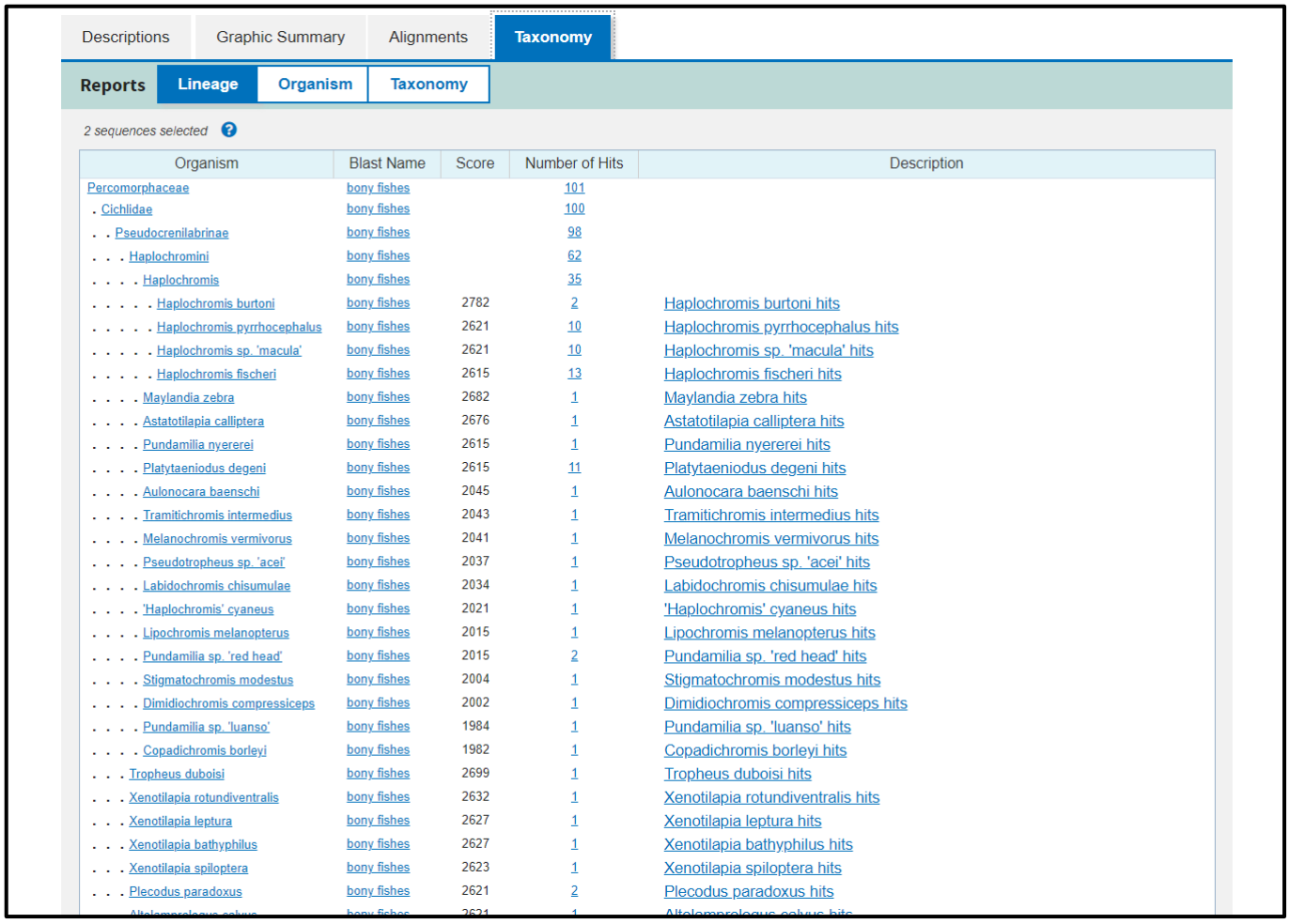


Fig8. Alignment for query Rhodopsin



Descriptions | Graphic Summary | Alignments | **Taxonomy**

Reports | **Lineage** | **Organism** | **Taxonomy**

2 sequences selected

Organism	Blast Name	Score	Number of Hits	Description
Percomorphaceae	bony fishes		101	
. Cichlidae	bony fishes		100	
. . Pseudocrenilabrinae	bony fishes		98	
. . . Haplochromini	bony fishes		62	
. . . . Haplochromis	bony fishes		35	
. . . . . Haplochromis burtoni	bony fishes	2782	2	Haplochromis burtoni hits
. . . . . Haplochromis pyrrhocephalus	bony fishes	2621	10	Haplochromis pyrrhocephalus hits
. . . . . Haplochromis sp. 'macula'	bony fishes	2621	10	Haplochromis sp. 'macula' hits
. . . . . Haplochromis fischeri	bony fishes	2615	13	Haplochromis fischeri hits
. . . . Maylandia zebra	bony fishes	2682	1	Maylandia zebra hits
. . . . Astatotilapia calliptera	bony fishes	2676	1	Astatotilapia calliptera hits
. . . . Pundamilia nyererei	bony fishes	2615	1	Pundamilia nyererei hits
. . . . Platytaeniodus degeni	bony fishes	2615	11	Platytaeniodus degeni hits
. . . . Aulonocara baenschi	bony fishes	2045	1	Aulonocara baenschi hits
. . . . Tramitichromis intermedius	bony fishes	2043	1	Tramitichromis intermedius hits
. . . . Melanochromis vermivorus	bony fishes	2041	1	Melanochromis vermivorus hits
. . . . Pseudotropheus sp. 'acei'	bony fishes	2037	1	Pseudotropheus sp. 'acei' hits
. . . . Labidochromis chisumulae	bony fishes	2034	1	Labidochromis chisumulae hits
. . . . 'Haplochromis' cyaneus	bony fishes	2021	1	'Haplochromis' cyaneus hits
. . . . Lipochromis melanopterus	bony fishes	2015	1	Lipochromis melanopterus hits
. . . . Pundamilia sp. 'red head'	bony fishes	2015	2	Pundamilia sp. 'red head' hits
. . . . Stigmatochromis modestus	bony fishes	2004	1	Stigmatochromis modestus hits
. . . . Dimidiochromis compressiceps	bony fishes	2002	1	Dimidiochromis compressiceps hits
. . . . Pundamilia sp. 'luanso'	bony fishes	1984	1	Pundamilia sp. 'luanso' hits
. . . . Copadichromis borleyi	bony fishes	1982	1	Copadichromis borleyi hits
. . . Tropheus duboisi	bony fishes	2699	1	Tropheus duboisi hits
. . . Xenotilapia rotundiventralis	bony fishes	2632	1	Xenotilapia rotundiventralis hits
. . . Xenotilapia leptura	bony fishes	2627	1	Xenotilapia leptura hits
. . . Xenotilapia bathyphilus	bony fishes	2627	1	Xenotilapia bathyphilus hits
. . . Xenotilapia spiloptera	bony fishes	2623	1	Xenotilapia spiloptera hits
. . . Plecodus paradoxus	bony fishes	2621	2	Plecodus paradoxus hits

Fig9. Taxonomy for query Rhodopsin

Results:

Here we have acquired info about the descriptions, Graphical summary, Alignment, and taxonomy related information on our query rhodopsin.

The results can be interpreted as:

Maximum score is the highest alignment score between the query sequence and the database segments.
Total score is the sum of the alignment scores of all sequences from the same db.
Percent Query coverage is the percent of the query length that is included in the aligned segments.
E-value is the measure of likeliness that sequence similarity is not random chance percent identity describes how similar the query is to the aligned sequences.

Conclusion:

BLASTN is a great tool to get the required information on a particular FASTA sequence in detail.

References:

1. BLAST: Basic Local Alignment Search Tool. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi
2. BLAST: Basic Local Alignment Search Tool. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi#alnHdr_2
3. Nucleotide BLAST: Search nucleotide databases using a nucleotide query. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=Blast-Search&LINK_LOC=blasthome
4. Haplochromis burtoni rhodopsin-like (rhodopsin), mRNA - Nucleotide - NCBI. (n.d.). Retrieved from https://www.ncbi.nlm.nih.gov/nuccore/NM_001287832.1?report=fasta

Date: 1-11-21

WEBLEM 11/b

(URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi)

Aim:

To study the Query 'Rhodopsin' in BLASTp

Introduction:

A sequences similarity search often provides the first information about a new DNA or protein sequence. A search allows scientists to infer the function of a sequence from similar sequence. There are many ways of performing a sequences similarity search, but probably the most popular method is the "Basic Local Alignment Search Tools" (BLAST). BLAST uses heuristics to produces result quickly. It also calculates an "expect value" that estimates how many matches would have occurred at a given score by chance, which can aid a user in judging how much confidence to have in an alignment.

- BLASTP performs protein-protein sequence comparison, and its algorithm is the basis of many other types of BLAST searches such as BLASTX and TBLASTN.

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.

Rhodopsin was discovered in 1876 by German physiologist Franz Christian Boll, who observed that the normally reddish purple frog retina turned pale in bright light. The fading of colour was later attributed to the destruction of rhodopsin, via a process known as bleaching. Bleaching and the subsequent regeneration of rhodopsin are major steps in the visual cycle—the series of biochemical reactions that is critical for vision in low light.

Methodology:

1. Open the homepage of blast.
2. After that click the Blastp.
3. Enter the sequence for rhodopsin taken from uniport
4. Open the result page.
5. Interpret the result.

Observation:

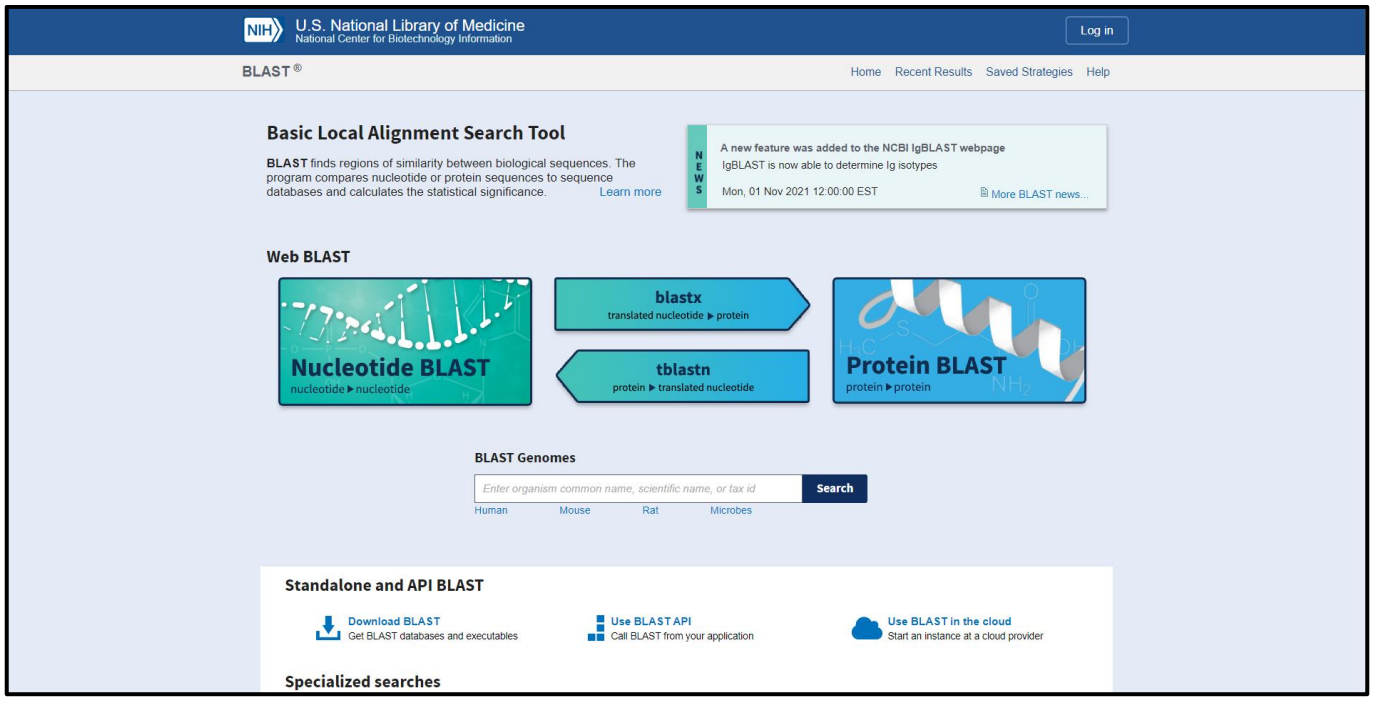


Fig1. Home page of BLAST

```
>sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX=9913 GN=RHO PE=1 SV=1
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSRYIP
EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA
```

Fig2. Fasta Sequence for query rhodopsin taken from uniport

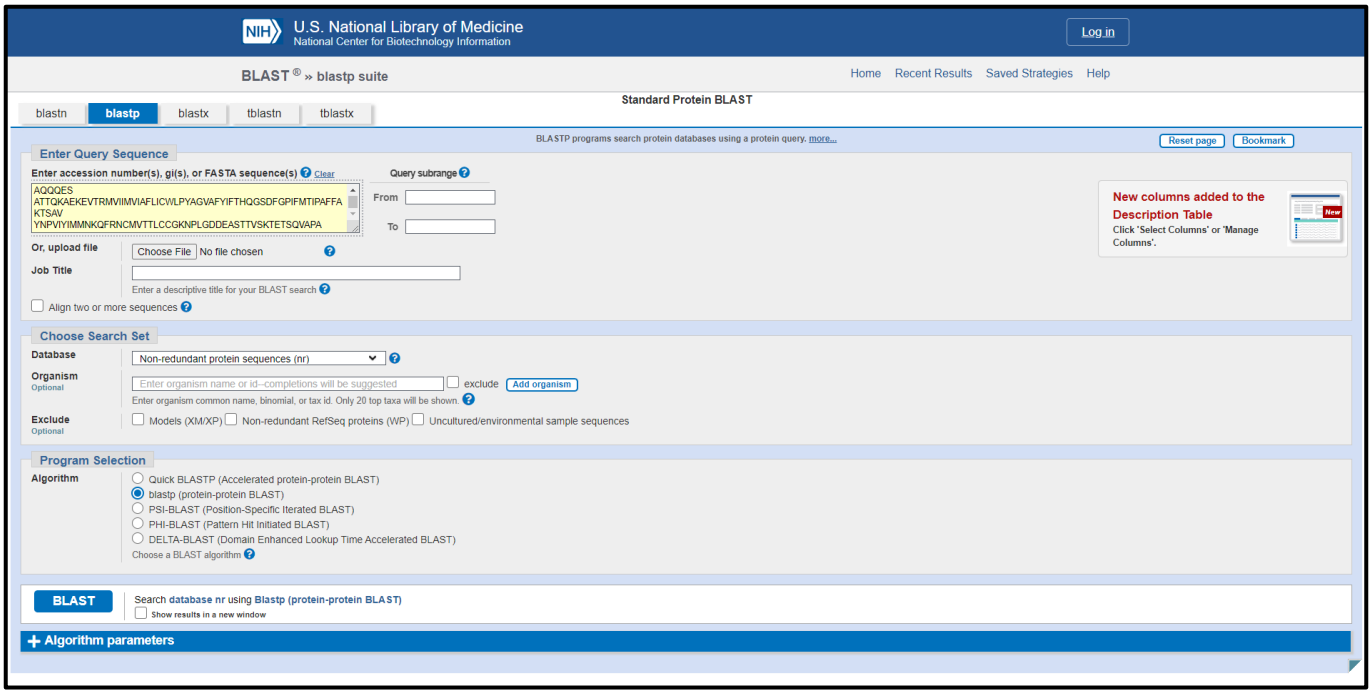


Fig3. Search page for blastp with sequence.

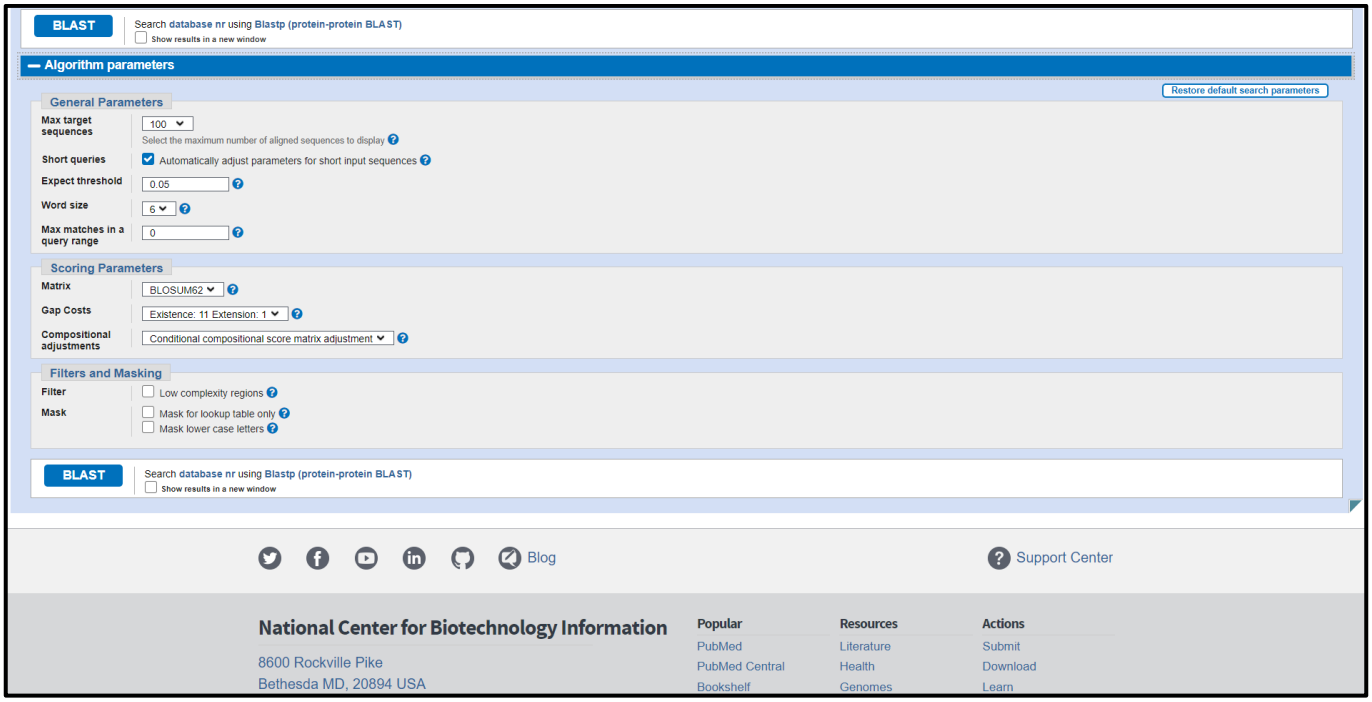


Fig4. Algorithm parameters in BLASTP

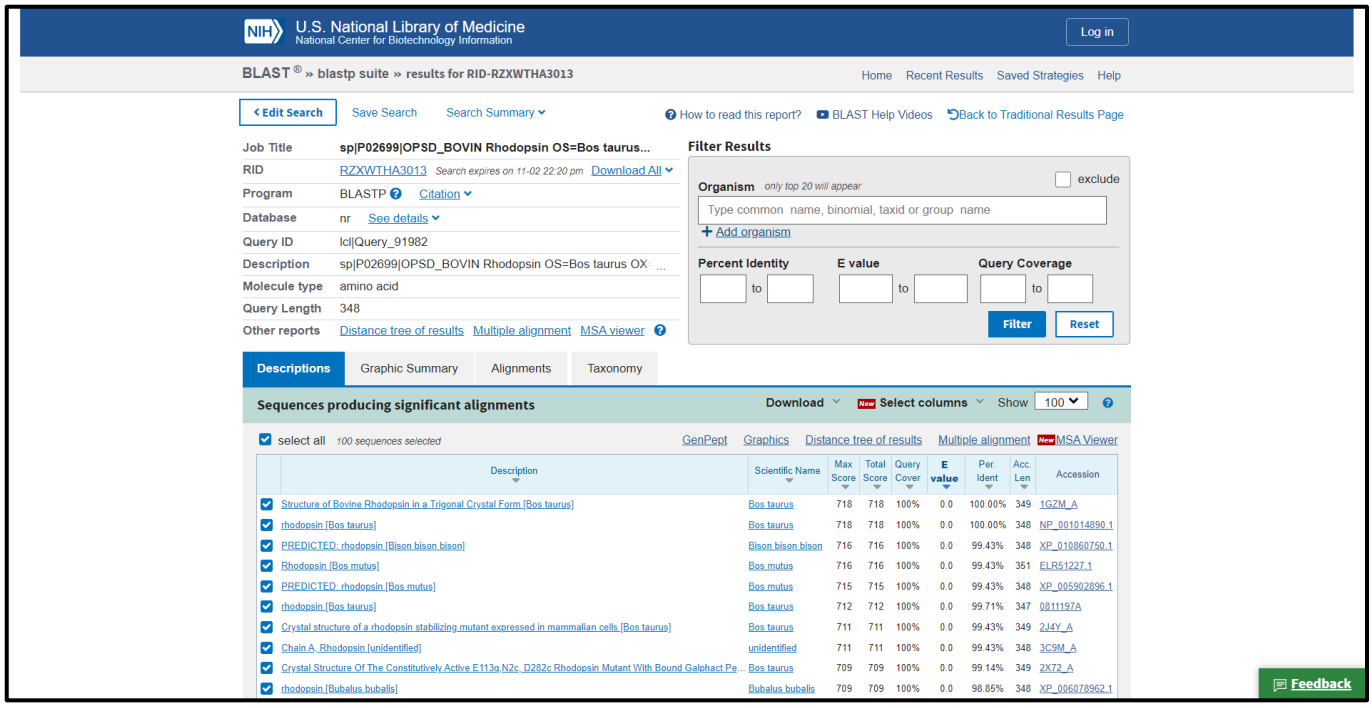


Fig5. Result page for rhodopsin

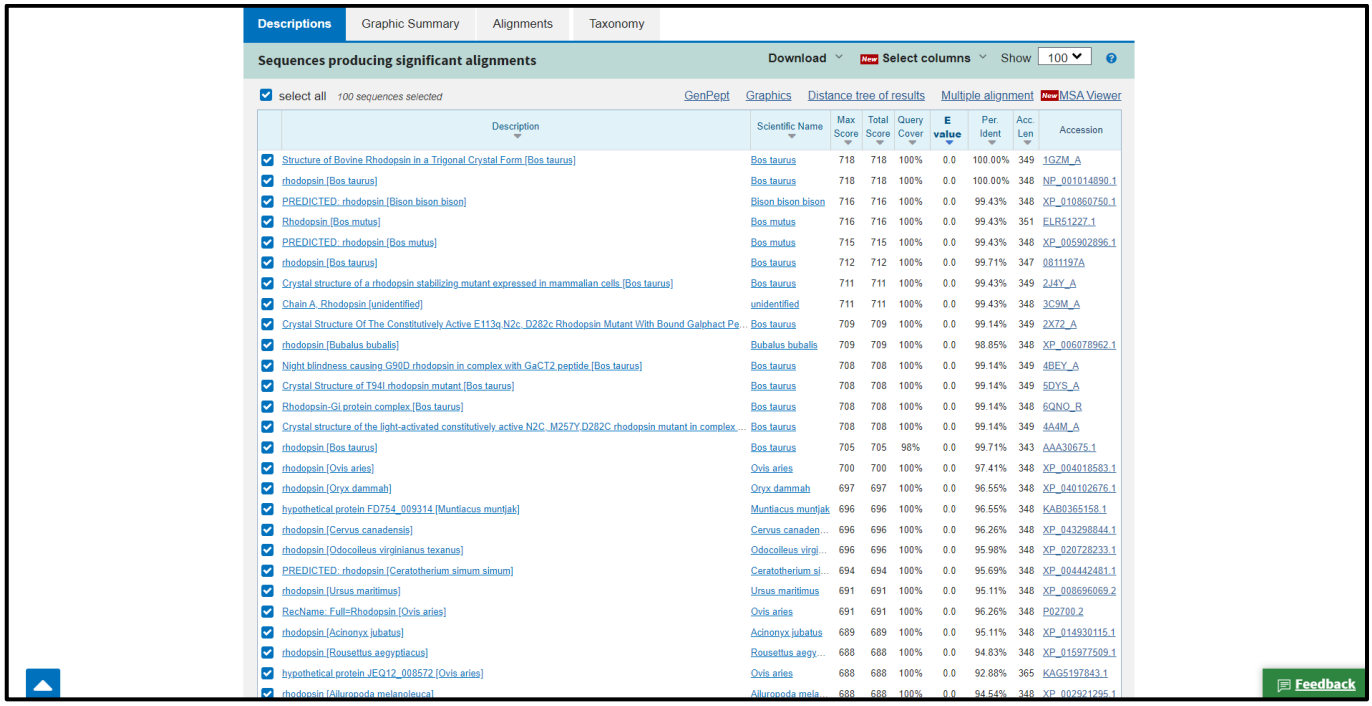


Fig6. Description for query Rhodopsin

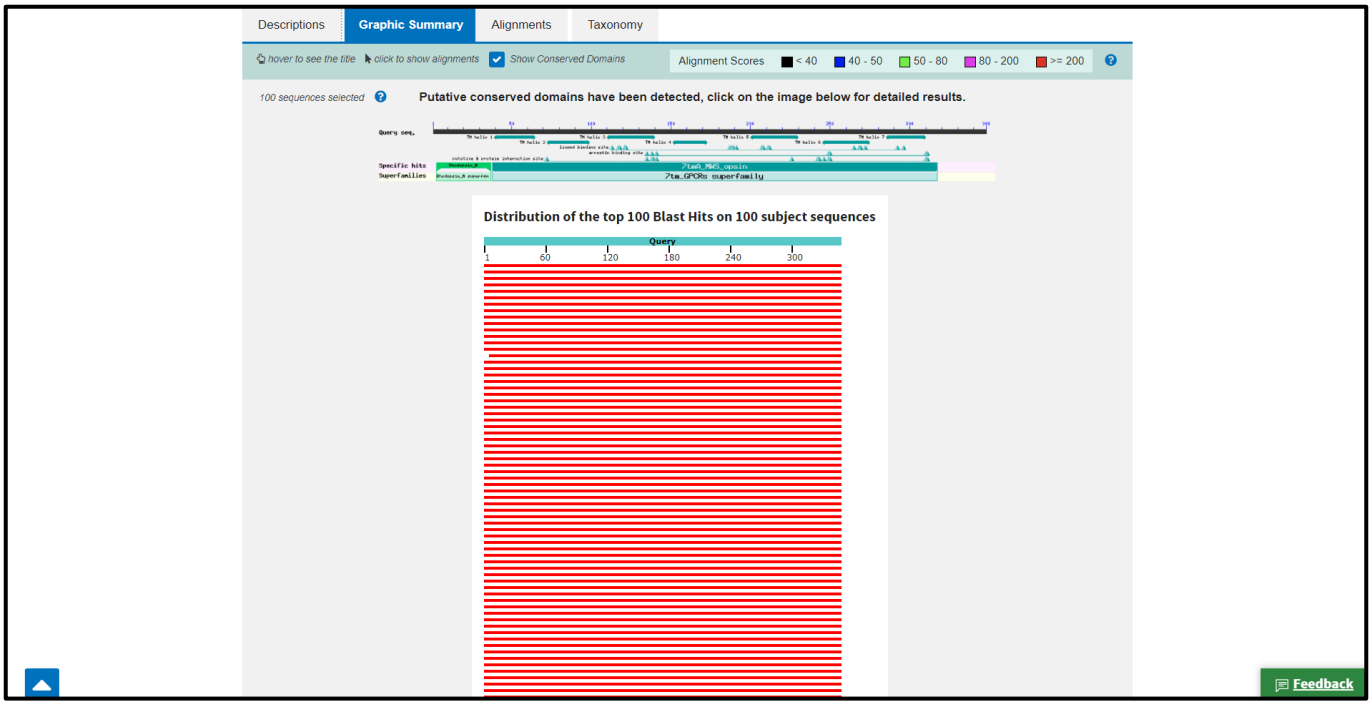


Fig7. Graphic Summary for query rhodopsin

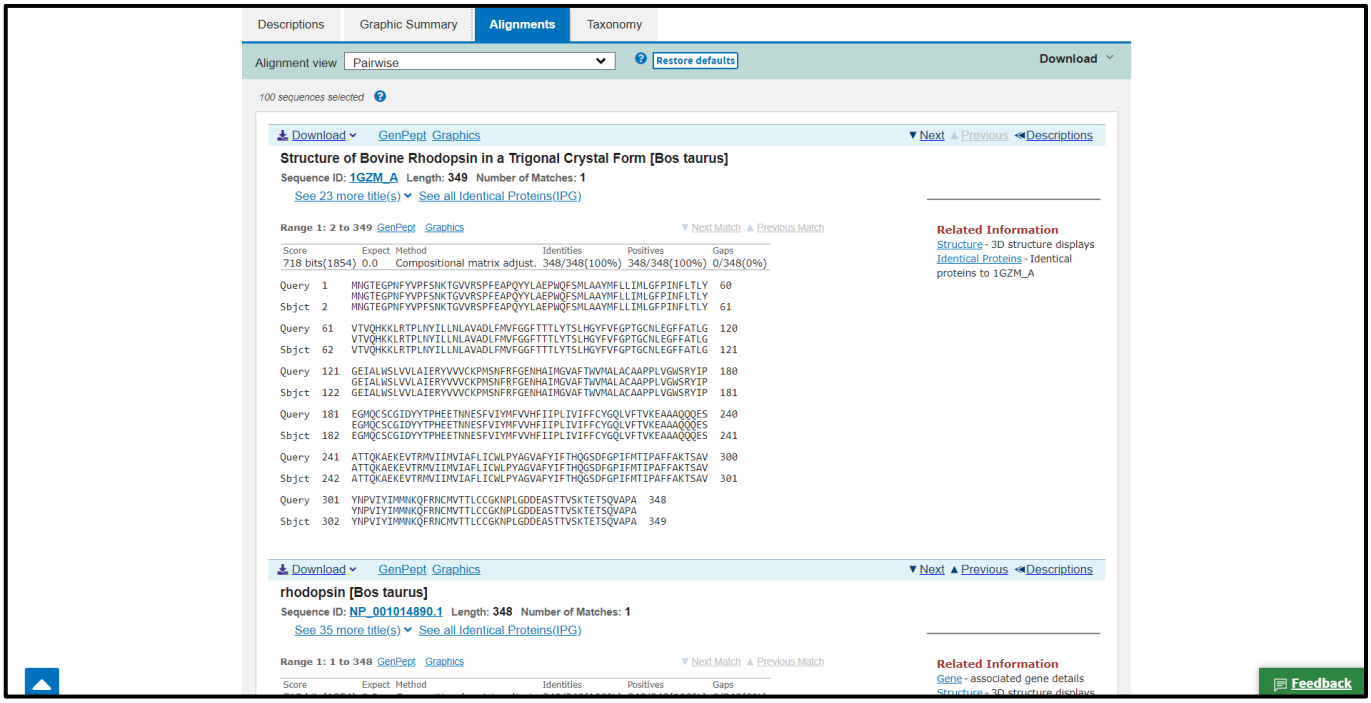


Fig8. Alignment for query rhodopsin

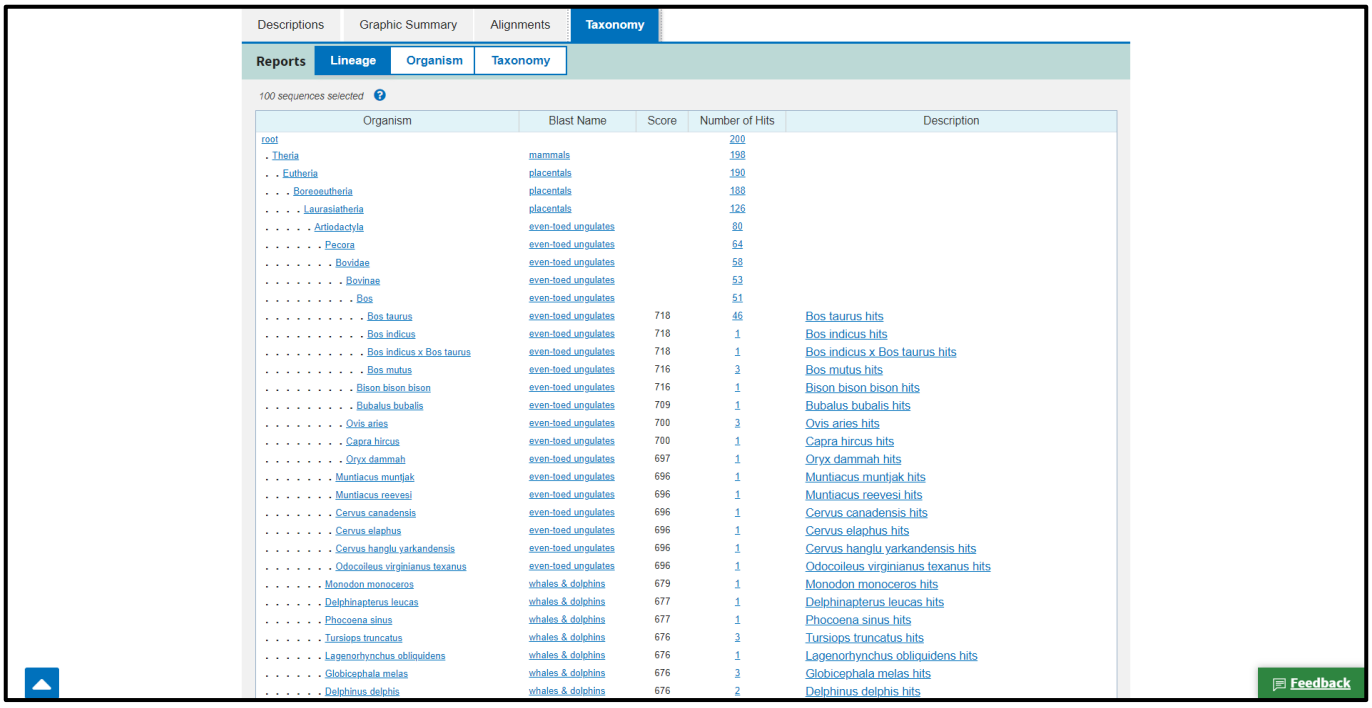


Fig9. Taxonomy for query rhodopsin

Results:

Here we have acquired info about the descriptions, Graphical summary, Alignment, and taxonomy related information on our query rhodopsin.

The results can be interpreted as:

Maximum score is the highest alignment score between the query sequence and the database segments.
Total score is the sum of the alignment scores of all sequences from the same db.
Percent Query coverage is the percent of the query length that is included in the aligned segments.
E-value is the measure of likeliness that sequence similarity is not random chance percent identity describes how similar the query is to the aligned sequences.

Conclusion:

BLASTN is a great tool to get the required information on a particular FASTA sequence in detail.

References:

1. (n.d.). Retrieved from https://www.uniprot.org/uniprot/P02699.fasta
2. BLAST: Basic Local Alignment Search Tool. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi
3. Protein BLAST: Search protein databases using a protein query. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=Blast-Search&LINK_LOC=blasthome
4. Rhodopsin. (n.d.). Retrieved from https://www.britannica.com/science/rhodopsin

DATE: 1-11-21

WEBLEM 11/c

(URL: https://www.ebi.ac.uk/Tools/sss/fasta/)

Aim:

To study the Query 'Rhodopsin' in FASTAn

Introduction:

FASTA (pronounced FAST-AYE) is a suite of programs for searching nucleotide or protein databases with a query sequence. FASTA itself performs a local heuristic search of a protein or nucleotide database for a query of the same type. FASTX and FASTY translate a nucleotide query for searching a protein database. TFASTX and TFASTY translate a nucleotide database to be searched with a protein query. Optimal searches are available with the programs SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query against local database).

FASTAn:

This tool provides sequence similarity searching against nucleotide databases using the FASTA suite of programs. FASTA provides a heuristic search with a nucleotide query. TFASTX and TFASTY translate the DNA database for searching with a protein query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.

Rhodopsin was discovered in 1876 by German physiologist Franz Christian Boll, who observed that the normally reddish purple frog retina turned pale in bright light. The fading of colour was later attributed to the destruction of rhodopsin, via a process known as bleaching. Bleaching and the subsequent regeneration of rhodopsin are major steps in the visual cycle—the series of biochemical reactions that is critical for vision in low light.

Methodology:

1. Open the homepage of FASTA.
2. After that click the FASTAn hyperlink
3. Enter the sequence for keratin taken by Genebank.
4. Open the result page.
5. Interpret the results.

Observation:

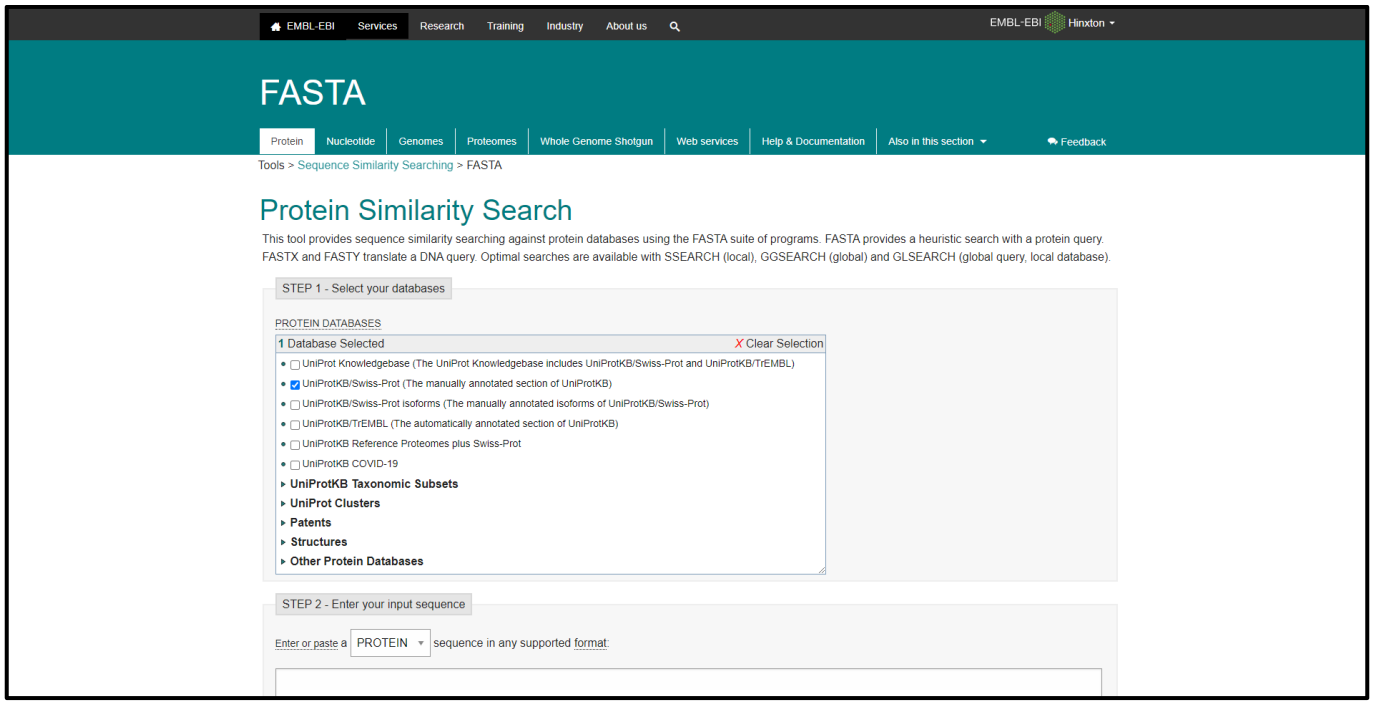


Fig1. Homepage of FASTAn.

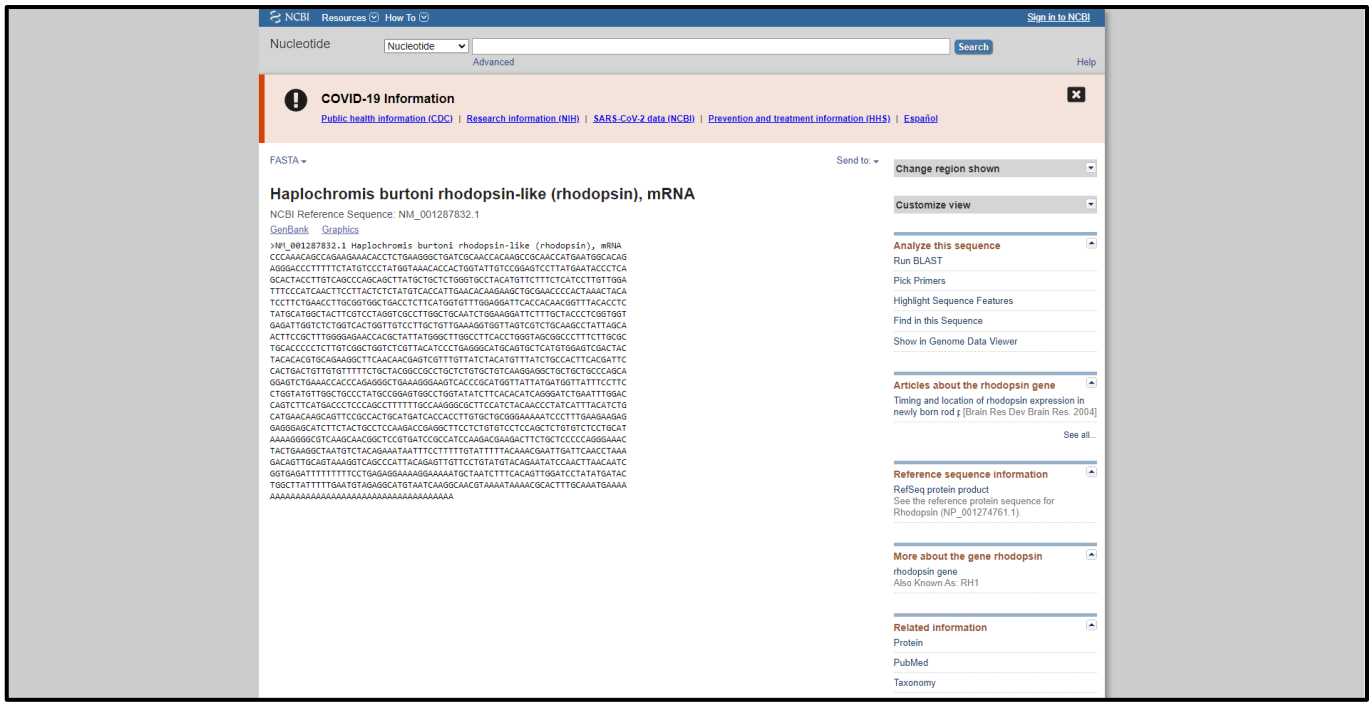


Fig2. Fasta sequence for my quesry rhodopsin on genbank

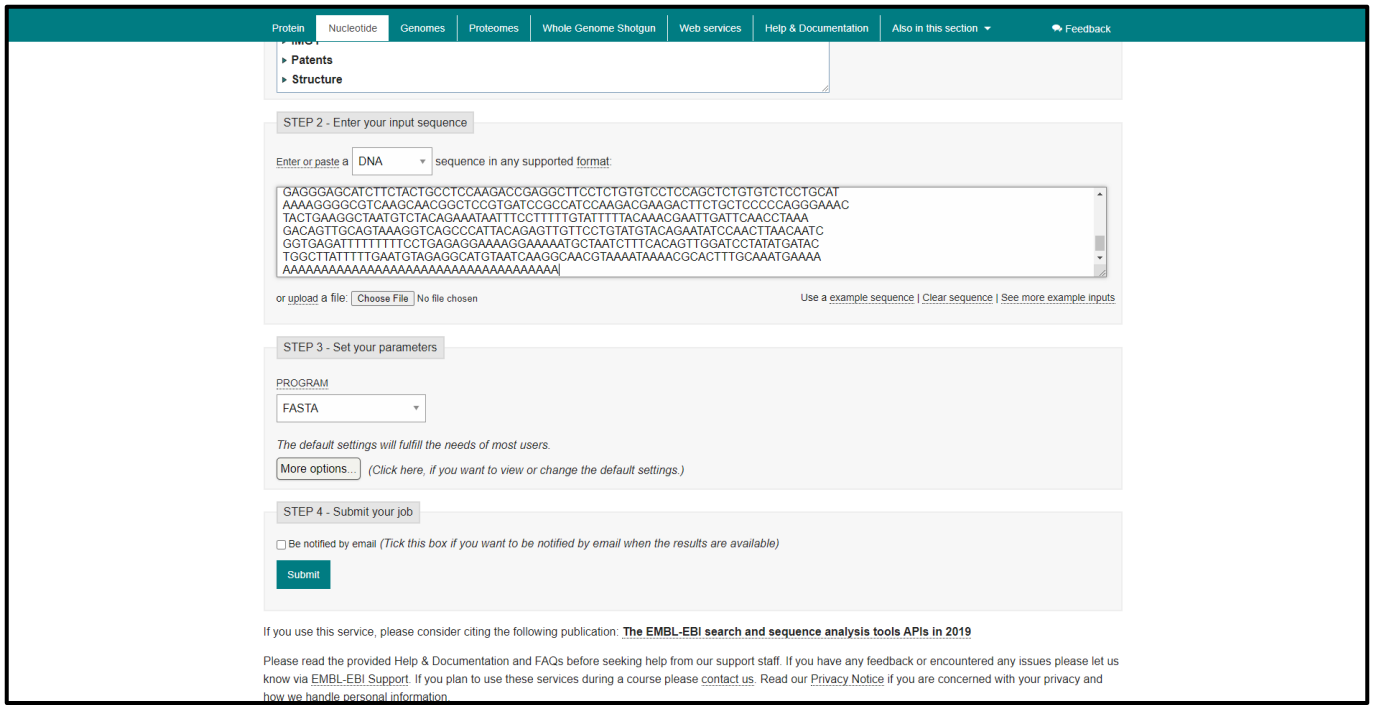


Fig3. Search bar of FASTAn with sequence.

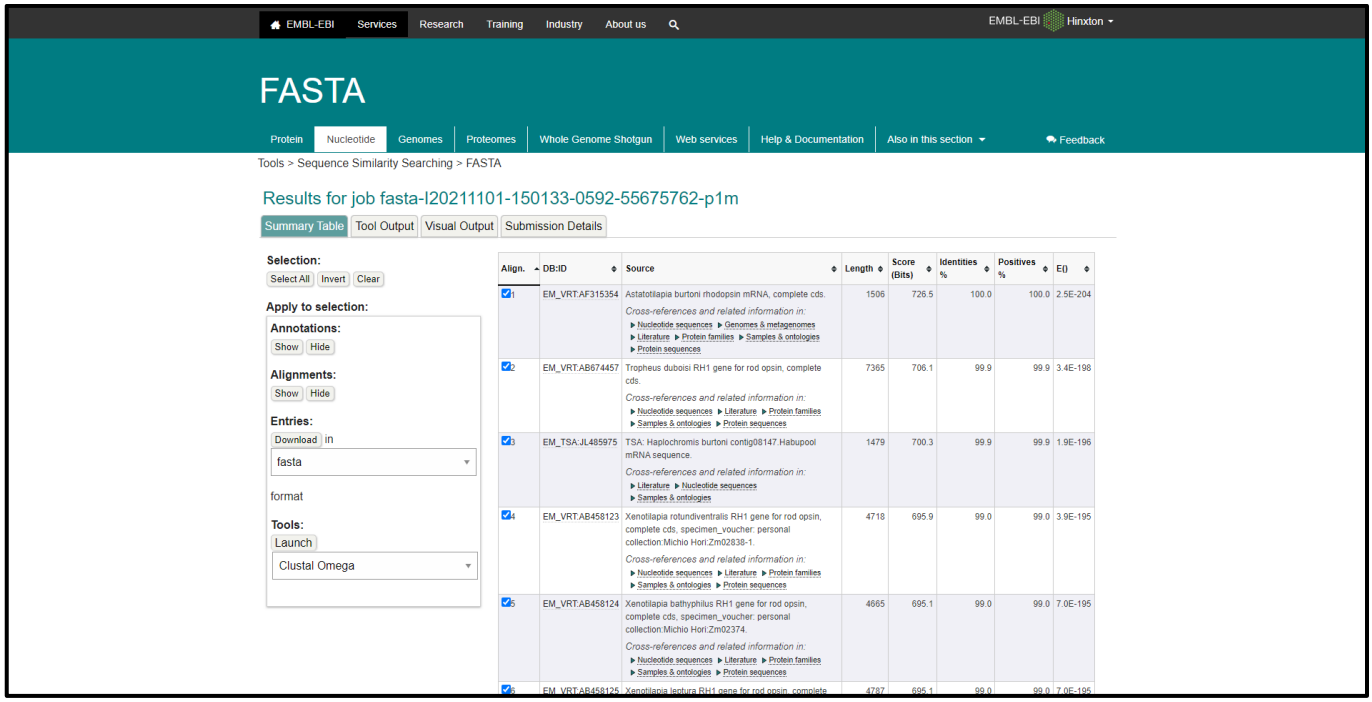


Fig4. Summary table for FASTAn with sequence

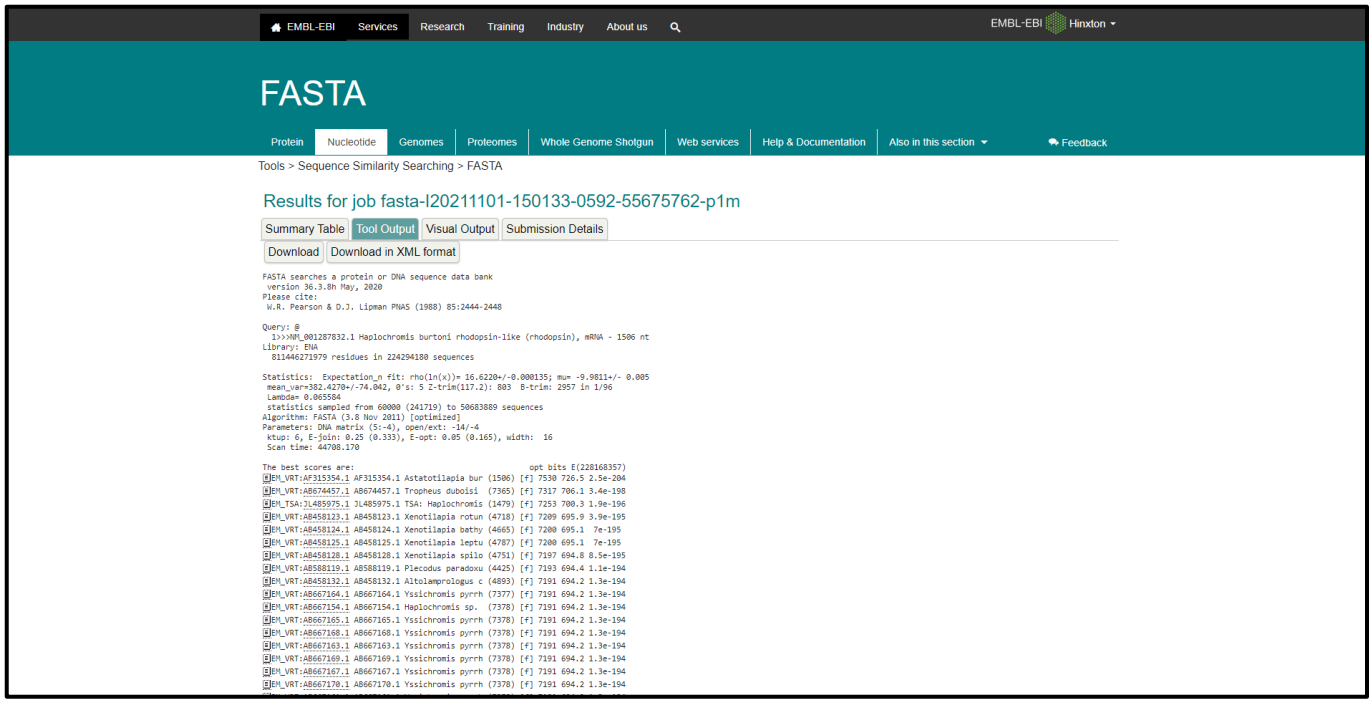


Fig5. Tool output for Rhodopsin in FASTAn

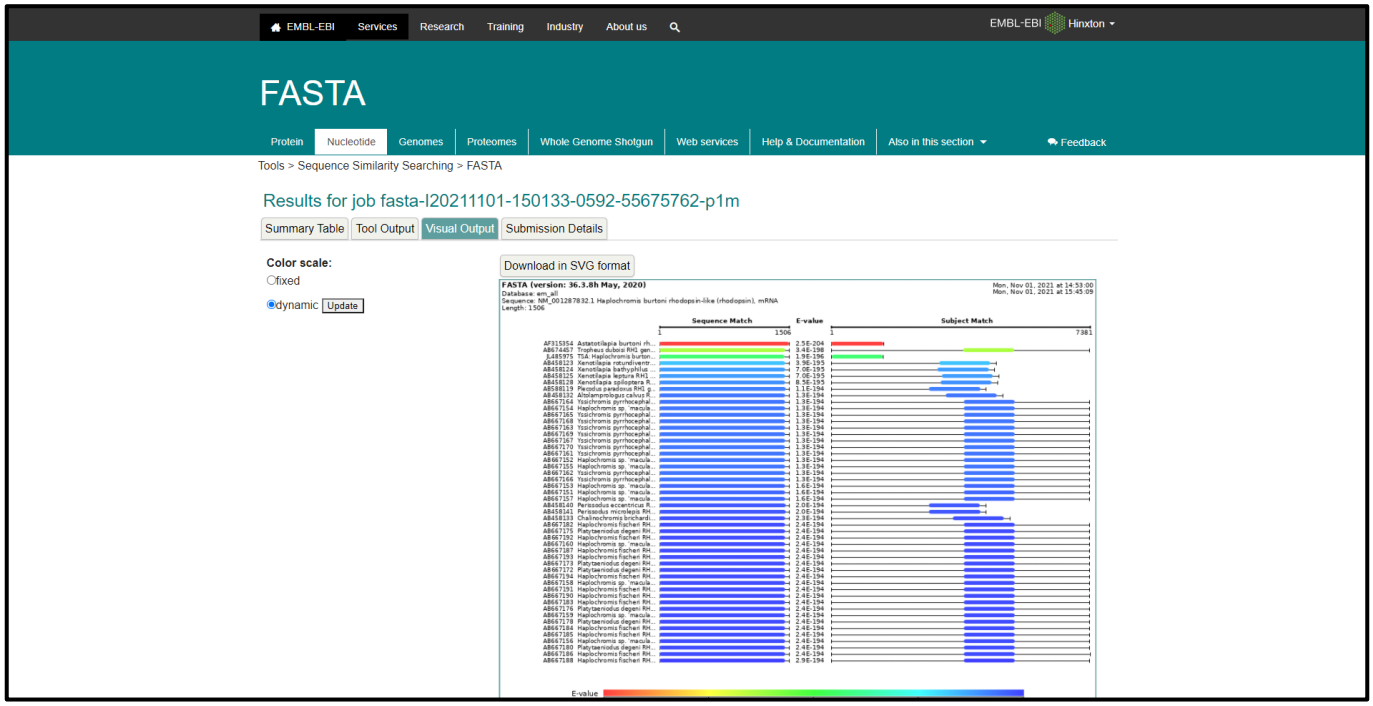


Fig6. Visual output for Rhodopsin in FASTAn

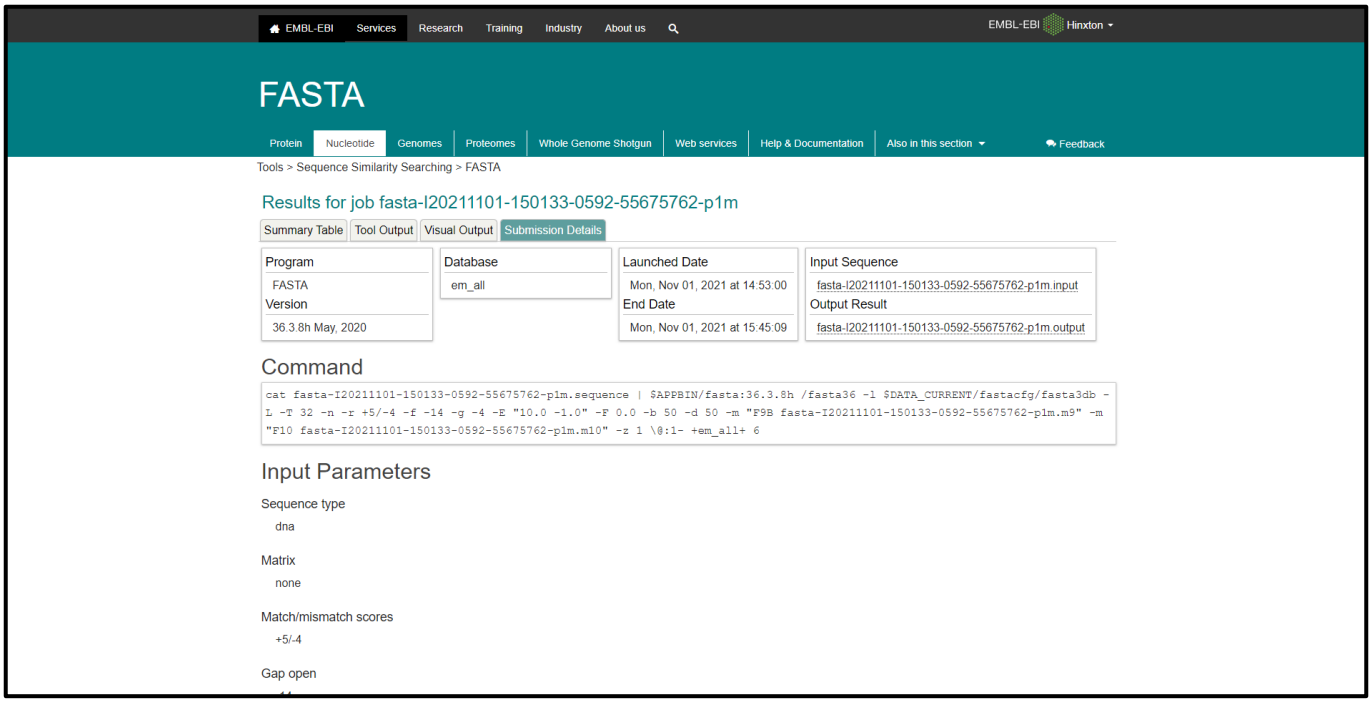


Fig7. Submission details for Rhodopsin in FASTAn

Results:

FASTAp Result divided into 5 sections summary tables, Tool output, visual output, functional predictions, Submission details.

In summary tables it shows alignment of sequence that is length, score bits, identity percentage and positive percentage.

Tools output for keratin in FASTAp shows multiple outputs for results

In visual output the color scale can be changed between fixed and dynamic, showing E-Values

Functional Prediction's color scale can also be switched between dynamic and fixed, showing E-Values

The submission details shows use the commands and input parameters that were used before we submitted the query for running.

Conclusion:

FASTA is another sequence alignment tool which is used to search similarities between sequences of DNA and proteins. The query sequence is broken down into sequence patterns or words known as k-tuples and the target sequences are searched for these k-tuples in order to find the similarities between the two.

References:

1. Rhodopsin. (n.d.). Retrieved from https://www.britannica.com/science/rhodopsin
2. (n.d.). Retrieved from https://www.uniprot.org/uniprot/P02699.fasta
3. Embl-Ebi. (n.d.). FASTA. Retrieved from https://www.ebi.ac.uk/Tools/sss/fasta/
4. Embl-Ebi. (n.d.). FASTA. Retrieved from https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=fasta-I20211101-150915-0959-51807210-p2m

DATE: 1-11-21

WEBLEM 11/d

(URL: https://www.ebi.ac.uk/Tools/sss/fasta/)

Aim:

To study the Query 'Rhodopsin' in FASTAp.

Introduction:

FASTA (pronounced FAST-AYE) is a suite of programs for searching nucleotide or protein databases with a query sequence. FASTA itself performs a local heuristic search of a protein or nucleotide database for a query of the same type. FASTX and FASTY translate a nucleotide query for searching a protein database. TFASTX and TFASTY translate a nucleotide database to be searched with a protein query. Optimal searches are available with the programs SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query against local database).

FASTAp:

This tool provides sequence similarity searching against protein databases using the FASTA suite of programs. FASTA provides a heuristic search with a protein query. FASTX and FASTY translate a DNA query. Optimal searches are available with SSEARCH (local), GGSEARCH (global) and GLSEARCH (global query, local database).

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions.

Rhodopsin was discovered in 1876 by German physiologist Franz Christian Boll, who observed that the normally reddish purple frog retina turned pale in bright light. The fading of colour was later attributed to the destruction of rhodopsin, via a process known as bleaching. Bleaching and the subsequent regeneration of rhodopsin are major steps in the visual cycle—the series of biochemical reactions that is critical for vision in low light.

Methodology:

1. Open the homepage of FASTA
2. Click on the FASTAp hyperlink
3. Enter the sequence for keratin taken by uniport
4. Open the result page.
5. Interpret the result.

Observation:

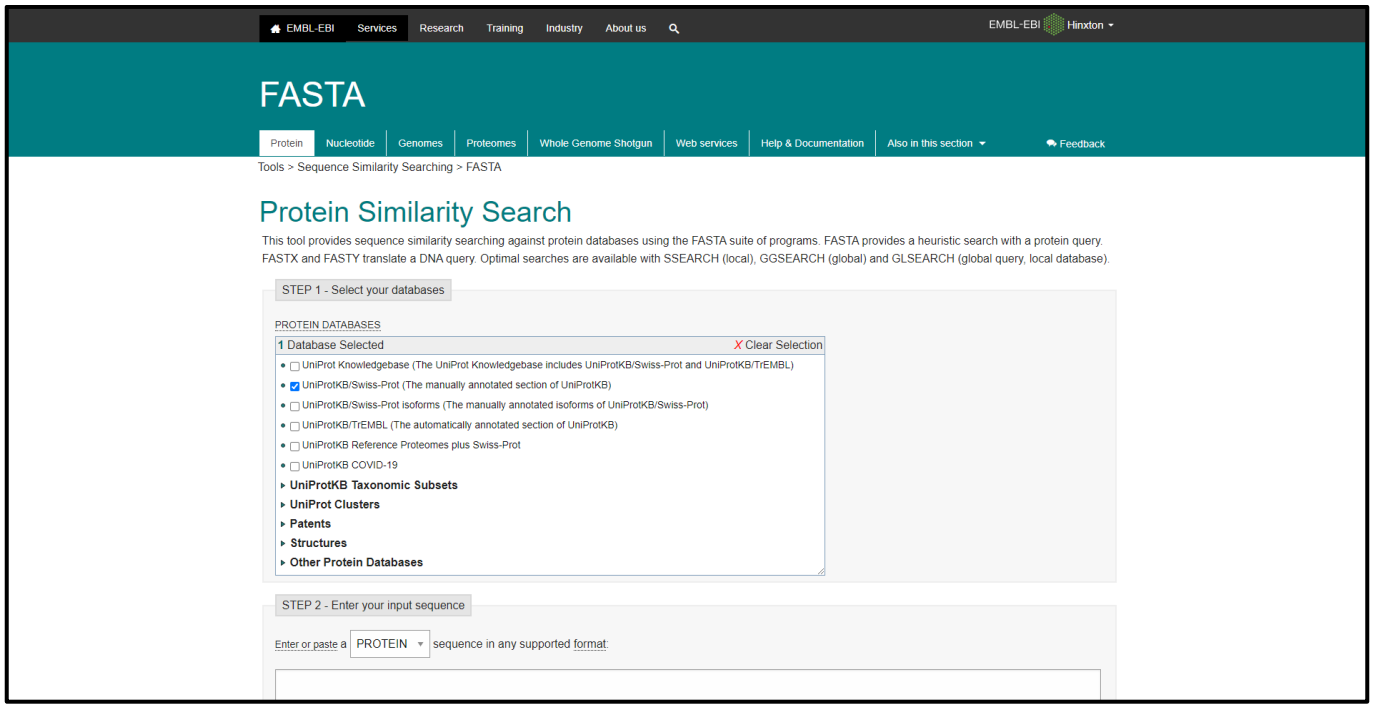


Fig1. Homepage of FASTAp

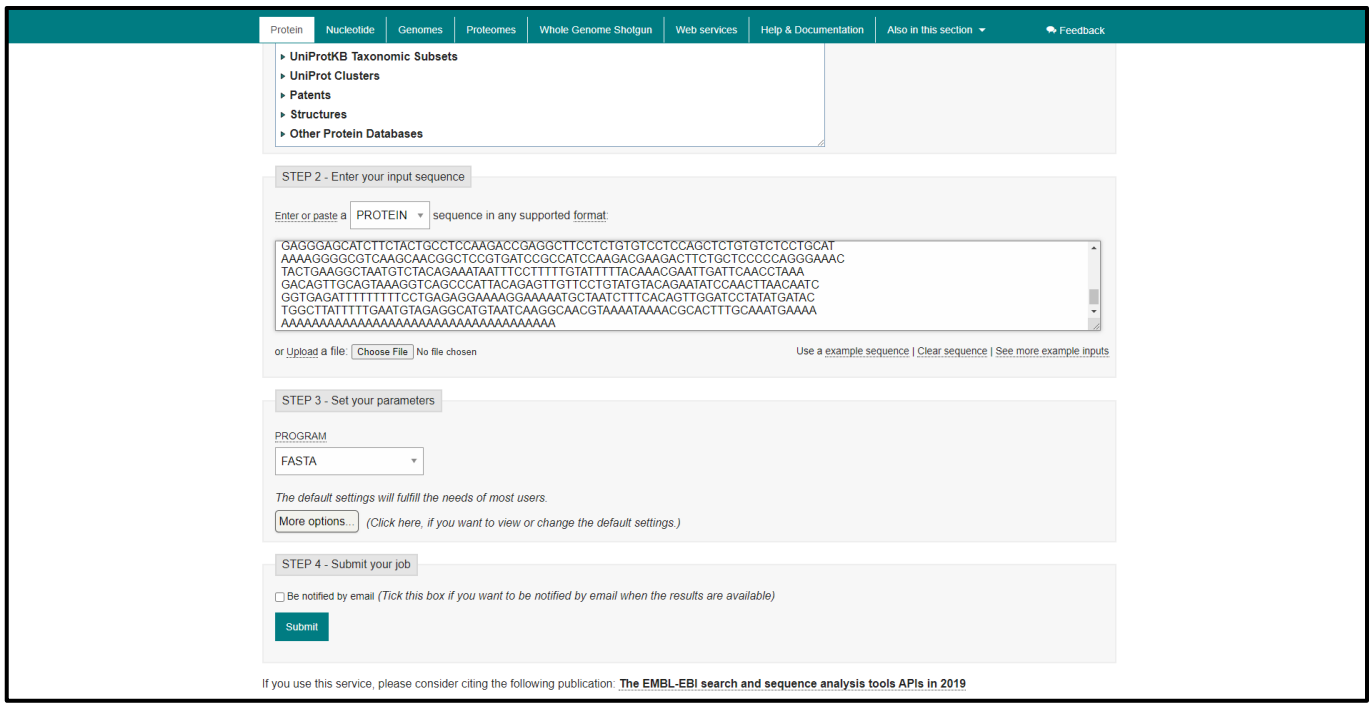


Fig2. Search bar for FASTAp with sequence for rhodopsin taken from uniprot

```
>sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX=9913 GN=RHO PE=1 SV=1
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSRYIP
EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA
```

Fig3. Fasta sequence for my query Rhodopsin

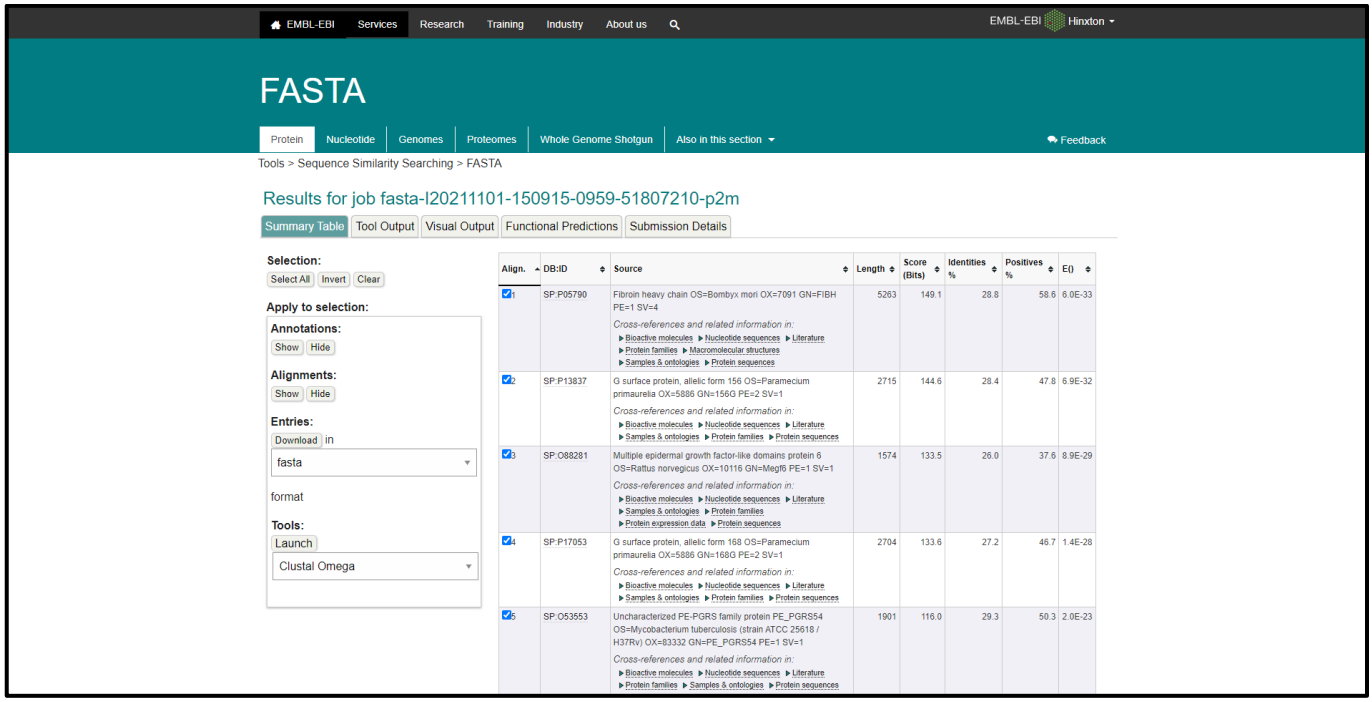


Fig4. Summary table for Rhodopsin in FASTAp

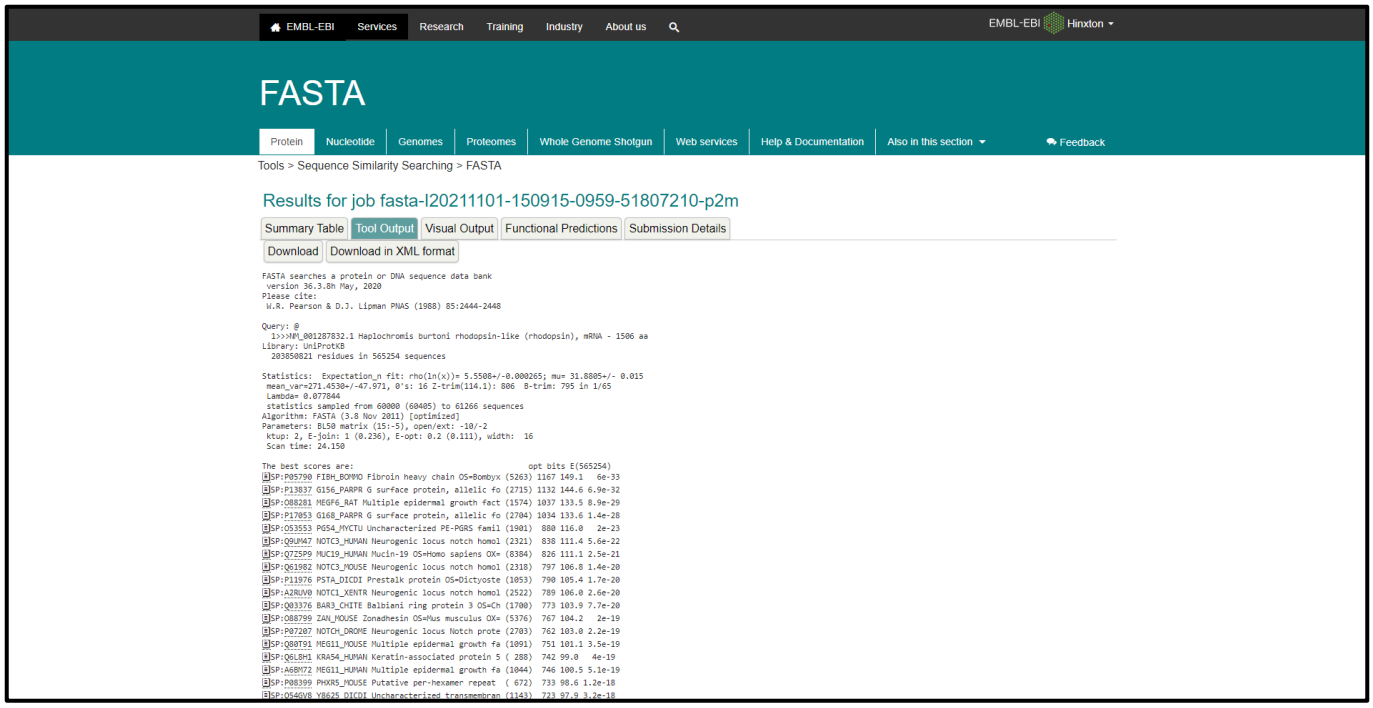


Fig5. Tool Output for Rhodopsin in FASTAp.

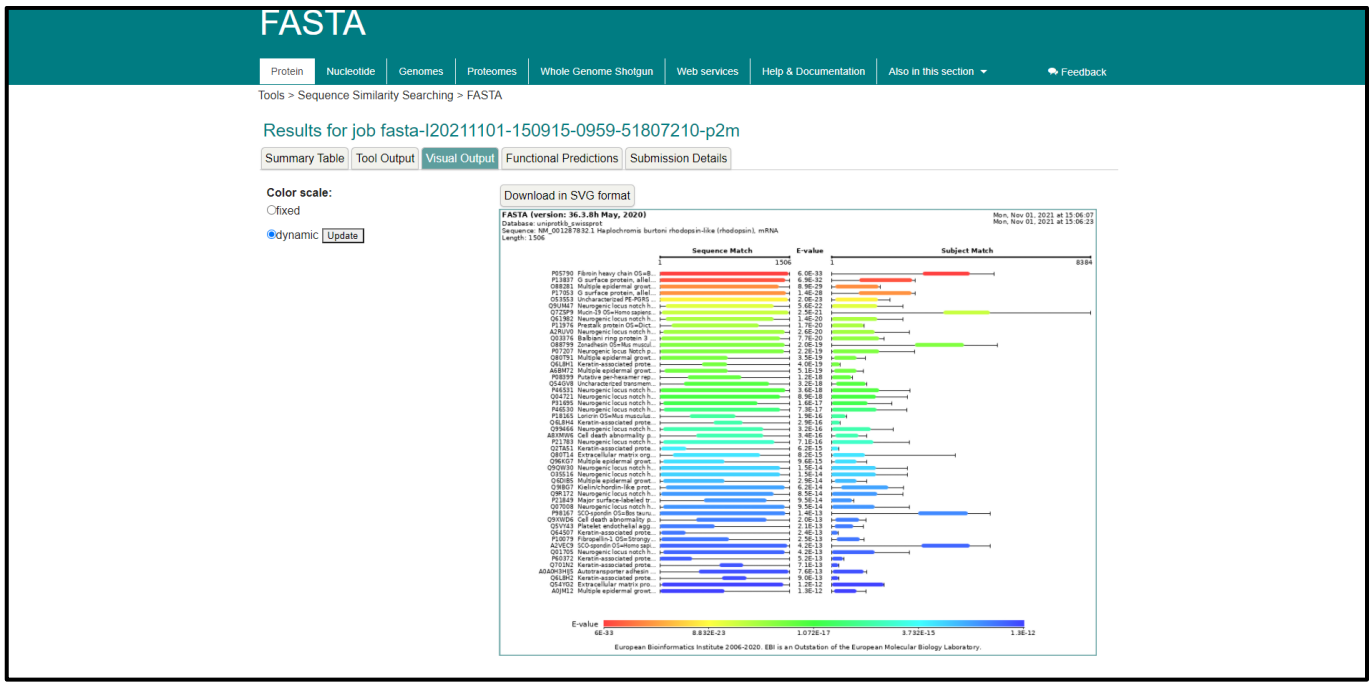


Fig6. Visual output for Rhodopsin in FASTAp

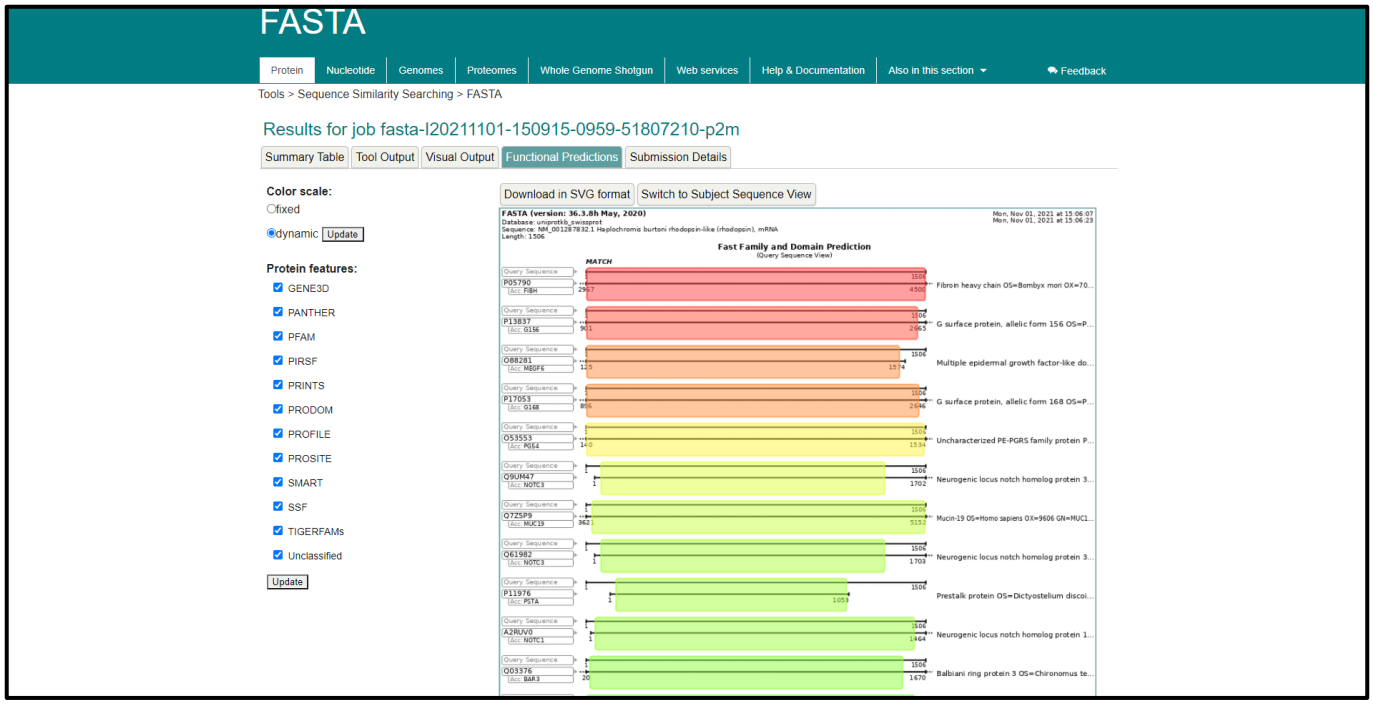


Fig7. Functional Predictions for Rhodopsin in FASTAp

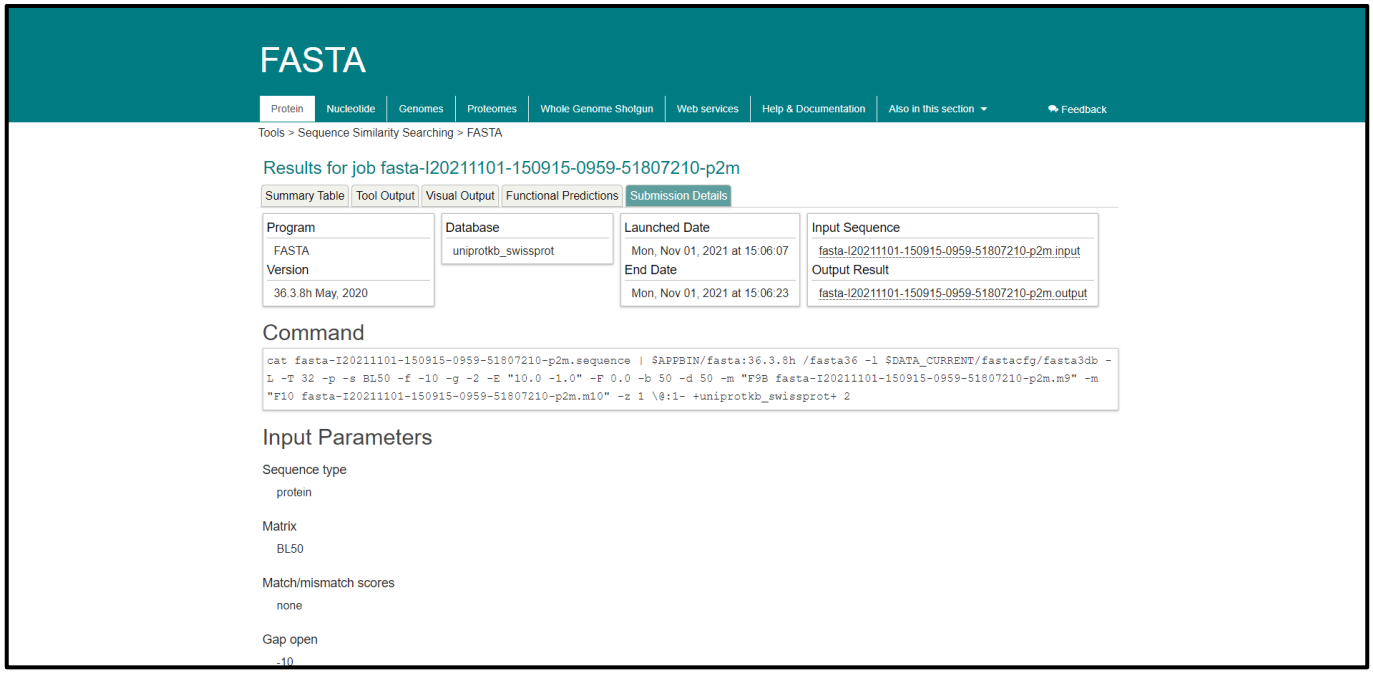


Fig8. Submission Details for Rhodopsin in FASTAp

Results:

FASTAp Result divided into 5 sections summary tables, Tool output, visual output, functional predictions, Submission details.

In summary tables it shows alignment of sequence that is length, score bits, identity percentage and positive percentage.

Tools output for keratin in FASTAp shows multiple outputs for results

In visual output the color scale can be changed between fixed and dynamic, showing E-Values

Functional Prediction's color scale can also be switched between dynamic and fixed, showing E-Values

The submission details shows use the commands and input parameters that were used before we submitted the query for running.

Conclusion:

FASTA is another sequence alignment tool which is used to search similarities between sequences of DNA and proteins. The query sequence is broken down into sequence patterns or words known as k-tuples and the target sequences are searched for these k-tuples in order to find the similarities between the two.

References:

5. Rhodopsin. (n.d.). Retrieved from https://www.britannica.com/science/rhodopsin
6. (n.d.). Retrieved from https://www.uniprot.org/uniprot/P02699.fasta
7. Embl-Ebi. (n.d.). FASTA. Retrieved from https://www.ebi.ac.uk/Tools/sss/fasta/
8. Embl-Ebi. (n.d.). FASTA. Retrieved from https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=fasta-I20211101-150915-0959-51807210-p2m

DATE: 1-11-21

WEBLEM 11/e

(URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi)

Aim:

To study the Query 'Rhodopsin' in PHI BLAST

Introduction:

A sequence similarity search often provides the first information about a new DNA or protein sequence. A search allows scientists to infer the function of a sequence from similar sequences. There are many ways of performing a sequence similarity search, but probably the most popular method is the "Basic Local Alignment Search Tool" (BLAST) (1, 2). BLAST uses heuristics to produce results quickly. It also calculates an "expect value" that estimates how many matches would have occurred at a given score by chance, which can aid a user in judging how much confidence to have in an alignment.

PHI-BLAST (Pattern-Hit Initiated BLAST) is a search program that combines matching of regular expressions with local alignments surrounding the match. The most important features of the program have been incorporated into the BLAST software framework partly for user convenience and partly so that PHI-BLAST may be combined seamlessly with PSI-BLAST. Other features that do not fit into the BLAST framework will be released later as a separate program and/or separate Web page query options. One very restrictive way to identify protein motifs is by regular expressions that must contain each instance of the motif. The PROSITE database is a compilation of restricted regular expressions that describe protein motifs. Given a protein sequence S and a regular expression pattern P occurring in S, PHI-BLAST helps answer the question: What other protein sequences both contain an occurrence of P and are homologous to S in the vicinity of the pattern occurrences? PHI-BLAST may be preferable to just searching for pattern occurrences because it filters out those cases where the pattern occurrence is probably random and not indicative of homology. PHI-BLAST may be preferable to other flavors of BLAST because it is faster and because it allows the user to express a rigid pattern occurrence requirement.

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions. Rhodopsin was discovered in 1876 by German physiologist Franz Christian Boll, who observed that the normally reddish purple frog retina turned pale in bright light. The fading of colour was later attributed to the destruction of rhodopsin, via a process known as bleaching. Bleaching and the subsequent regeneration of rhodopsin are major steps in the visual cycle—the series of biochemical reactions that is critical for vision in low light.

Methodology:

1. Open the homepage of blast.
2. After that click the BLASTp
3. Enter the sequence for rhodopsin taken by uniport.

4. Change the program selection Algorithm to PHI-BLAST
5. Enter the pattern taken by prosite
6. Open the result page
7. Interpret the result.

Observation:

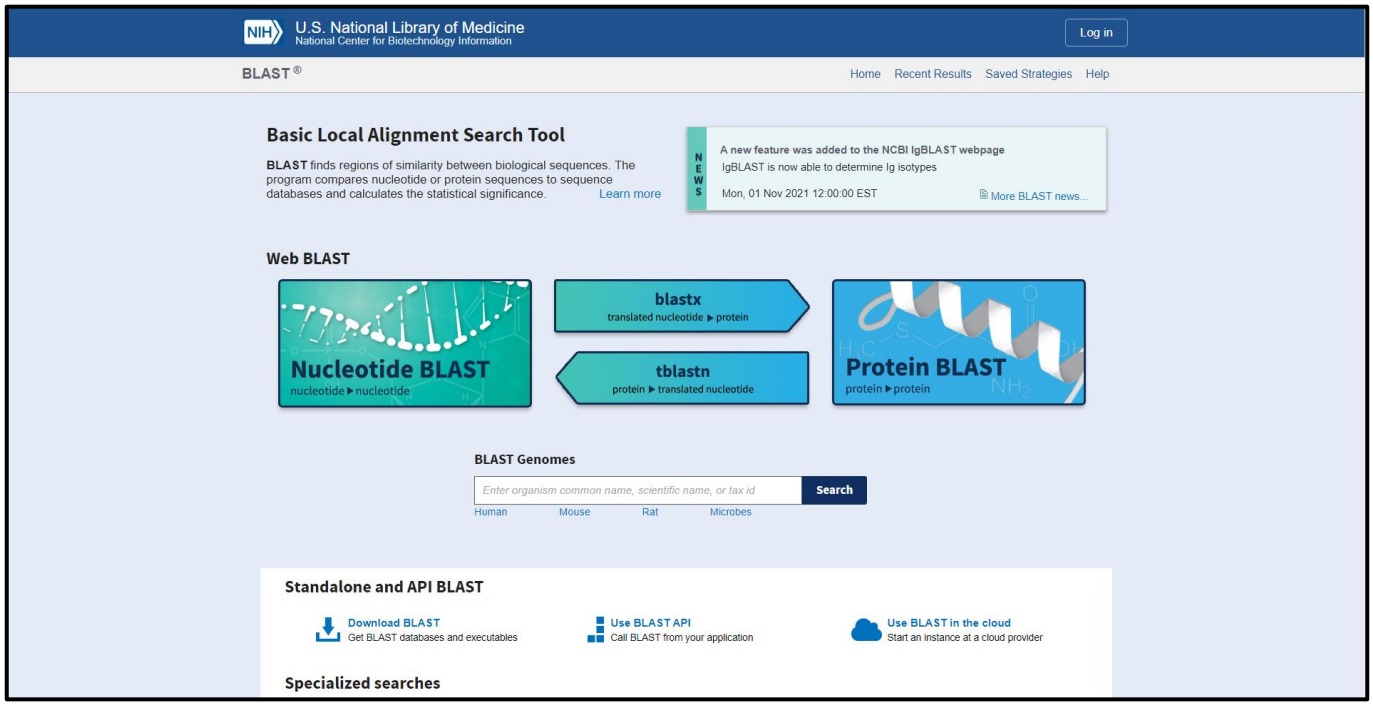


Fig1. Homepage of BLAST

```
>sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX=9913 GN=RHO PE=1 SV=1
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSRYIP
EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA
```

Fig2. FASTA sequences for query Rhodopsin

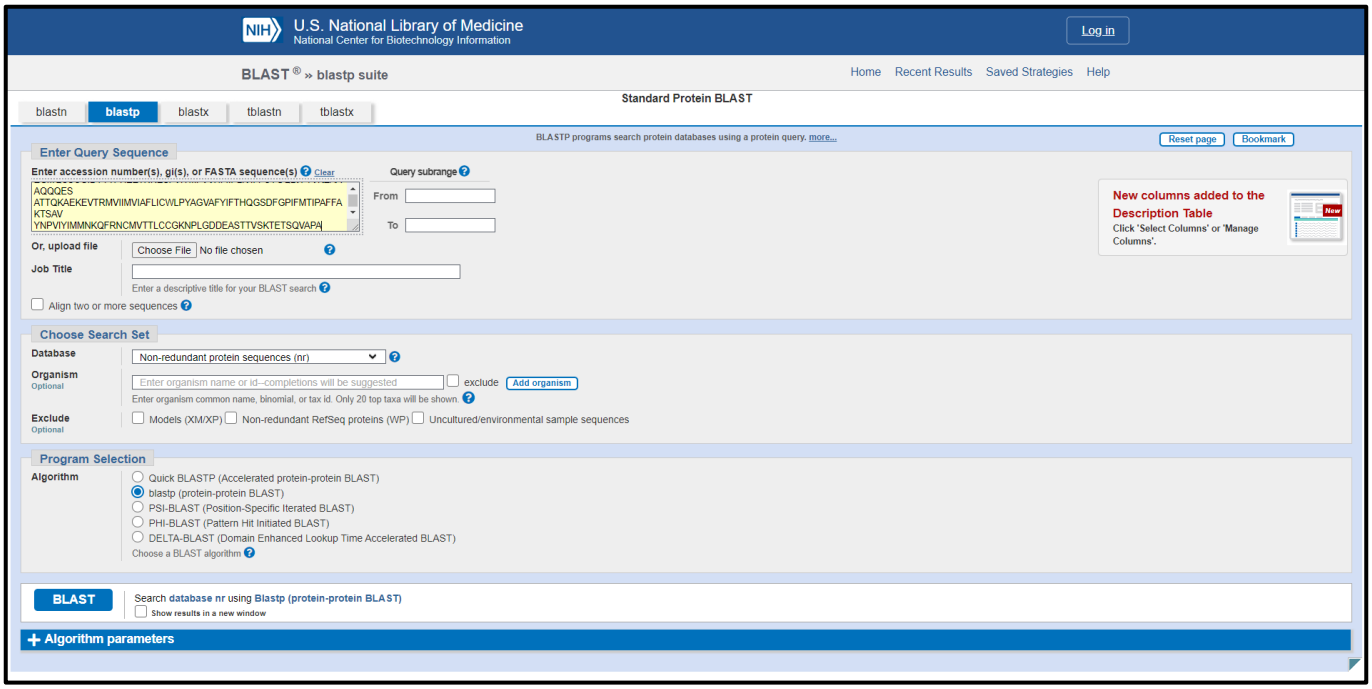


Fig3. Search bar for PHI blast with sequence pasted

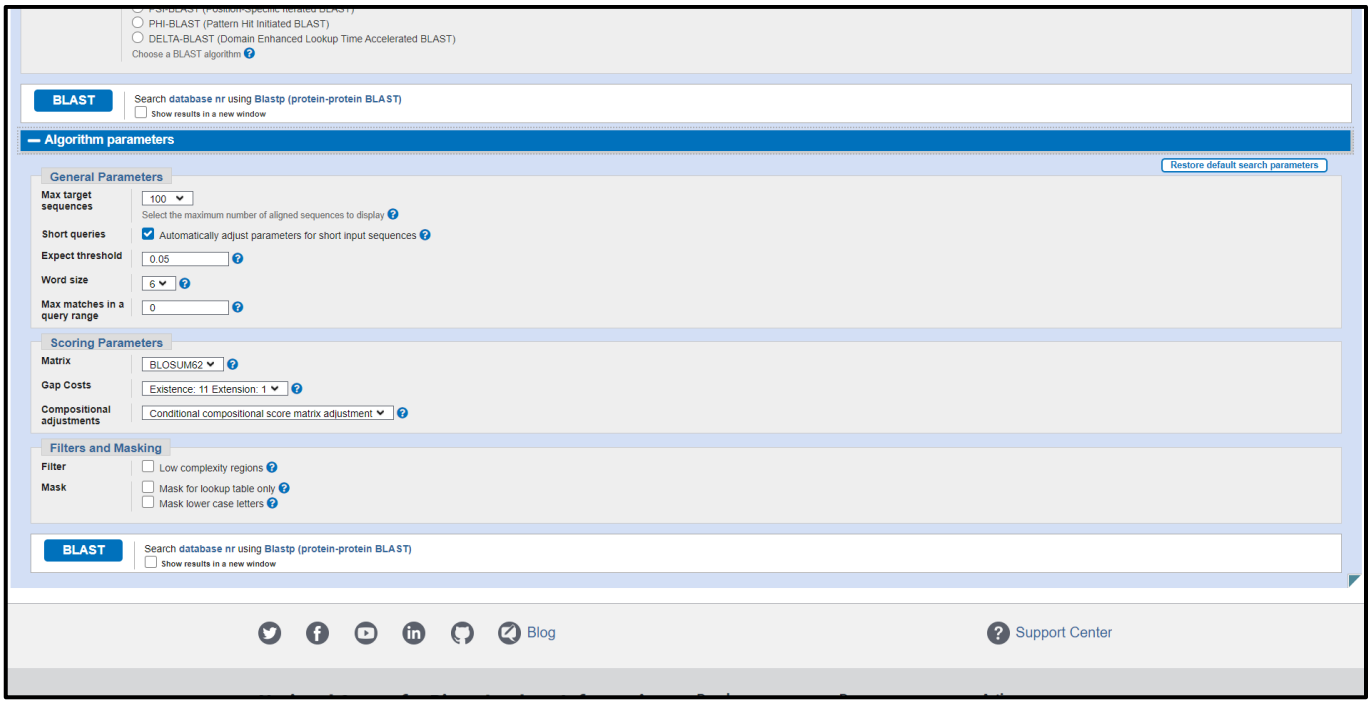


Fig4. Algorithm parameters for PHI blast for Rhodopsin

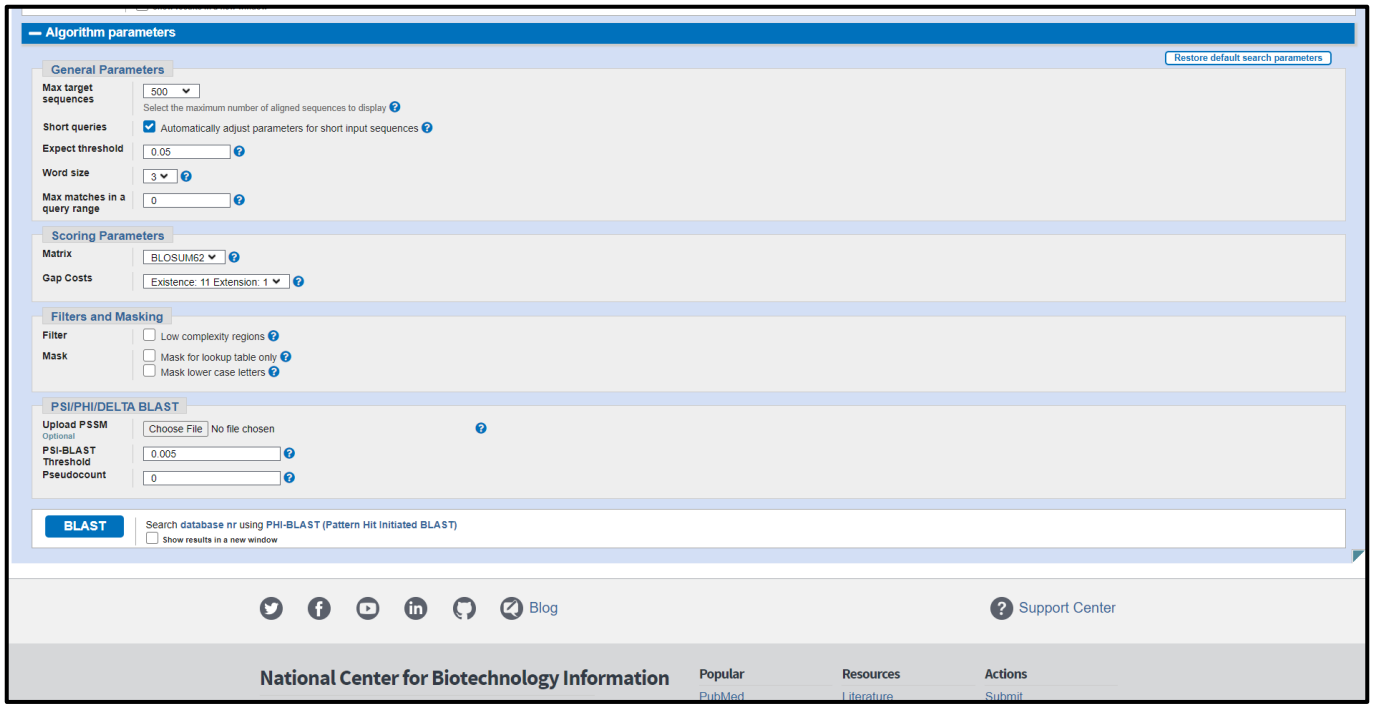


Fig5. PSI/PHI/DELTA parameters for Rhodopsin in BLAST

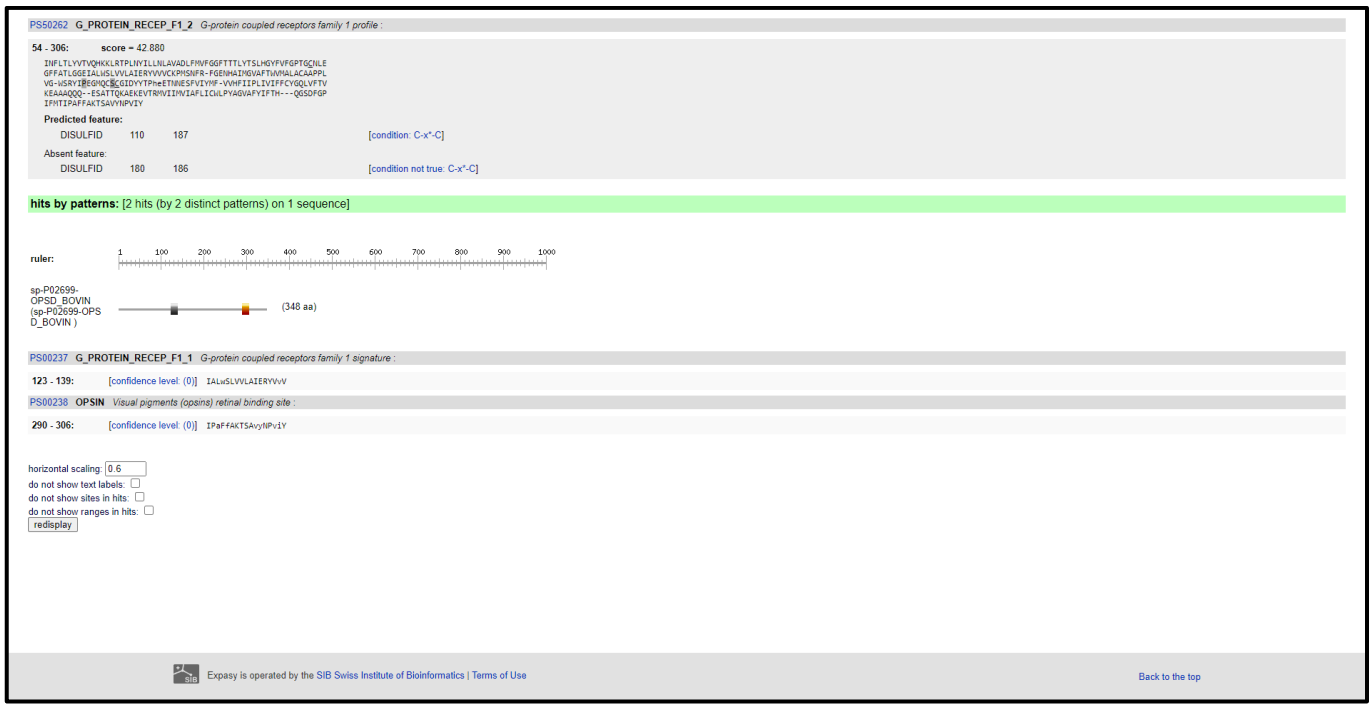


Fig6. Pattern for query in Rhodopsin in PROSITE

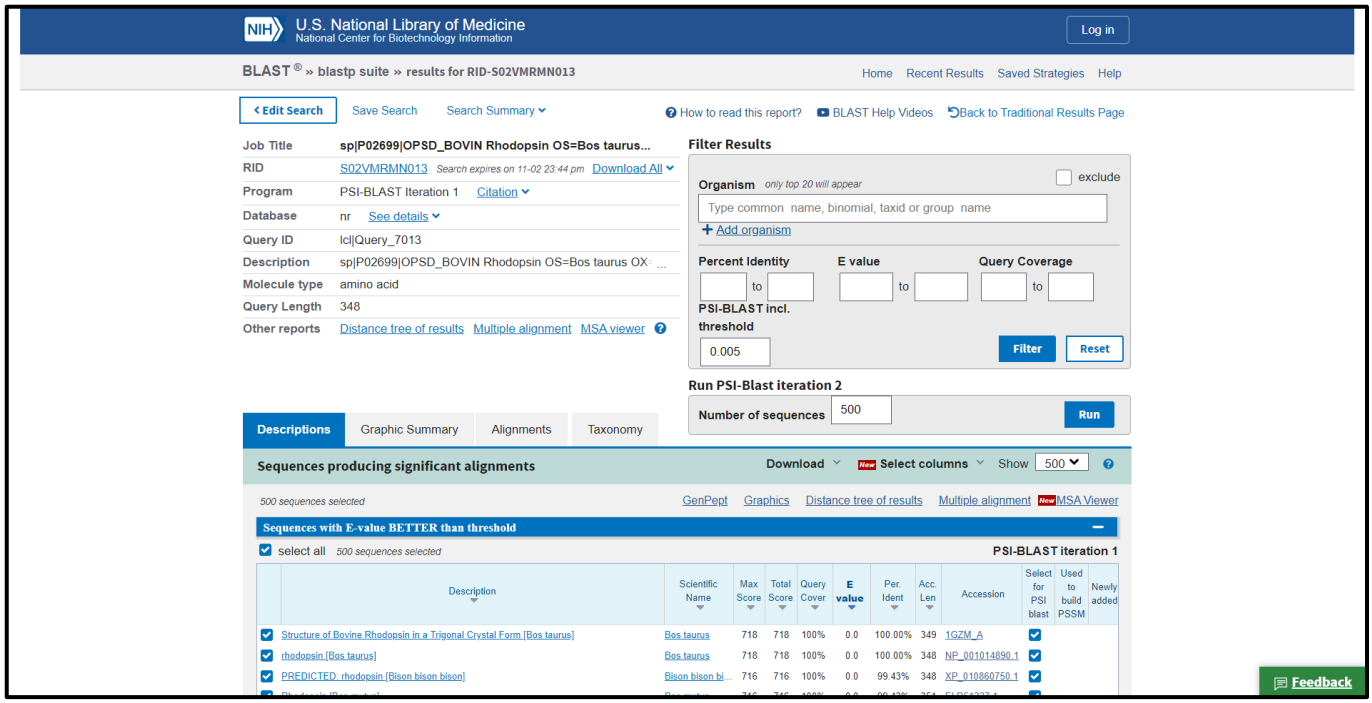


Fig7. Result page (header) for Rhodopsin

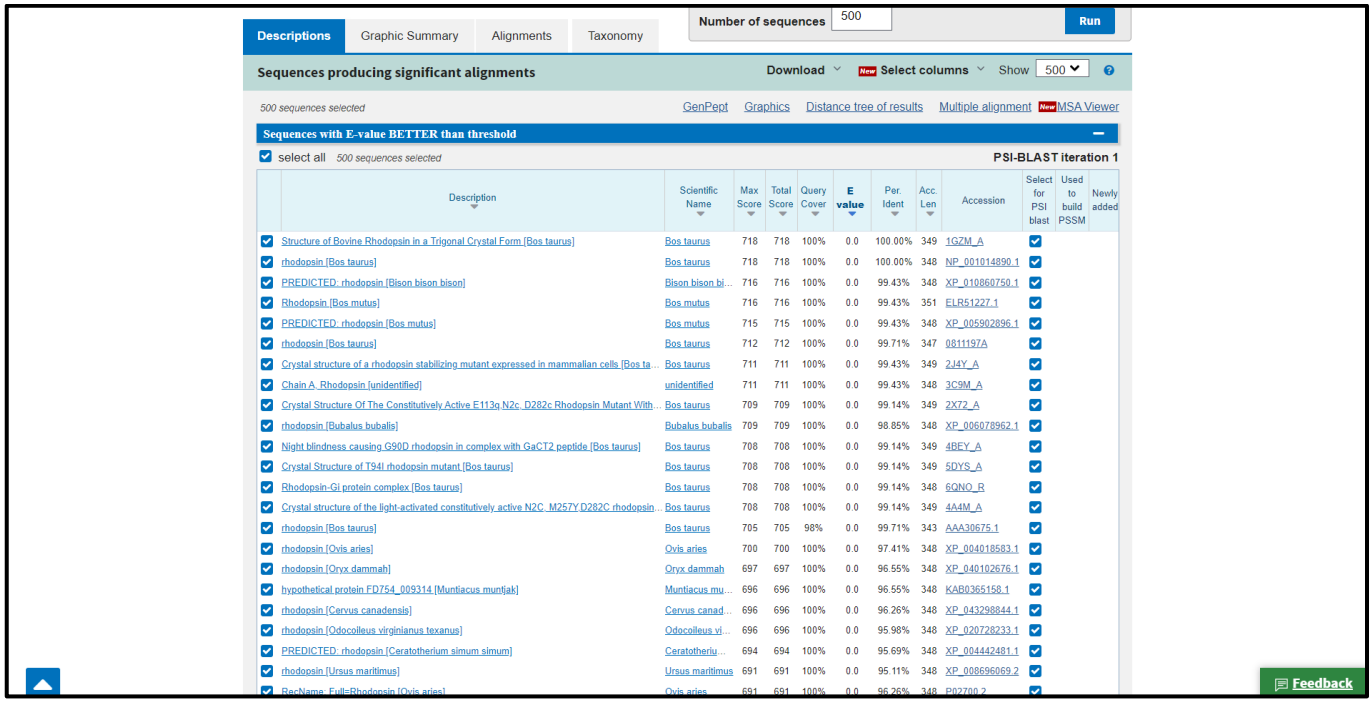


Fig8. Description for query Rhodopsin

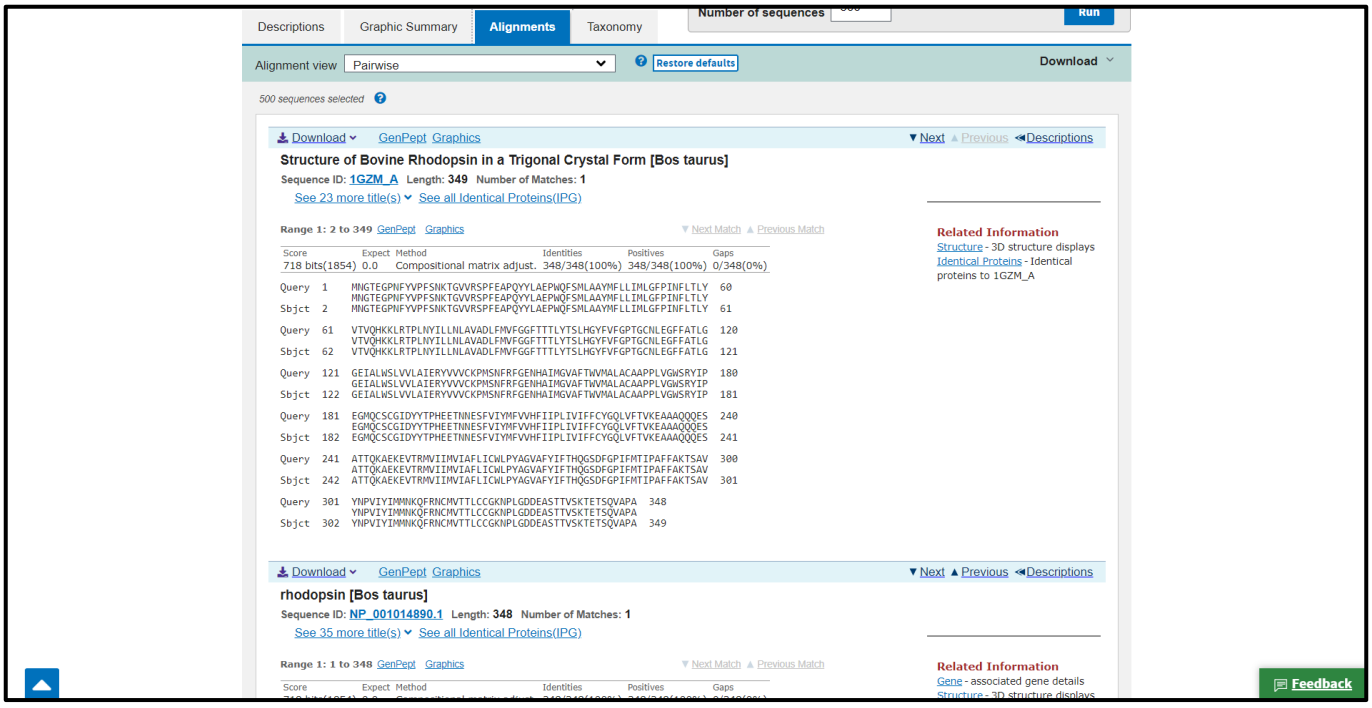


Fig9. Alignments view for query Rhodopsin

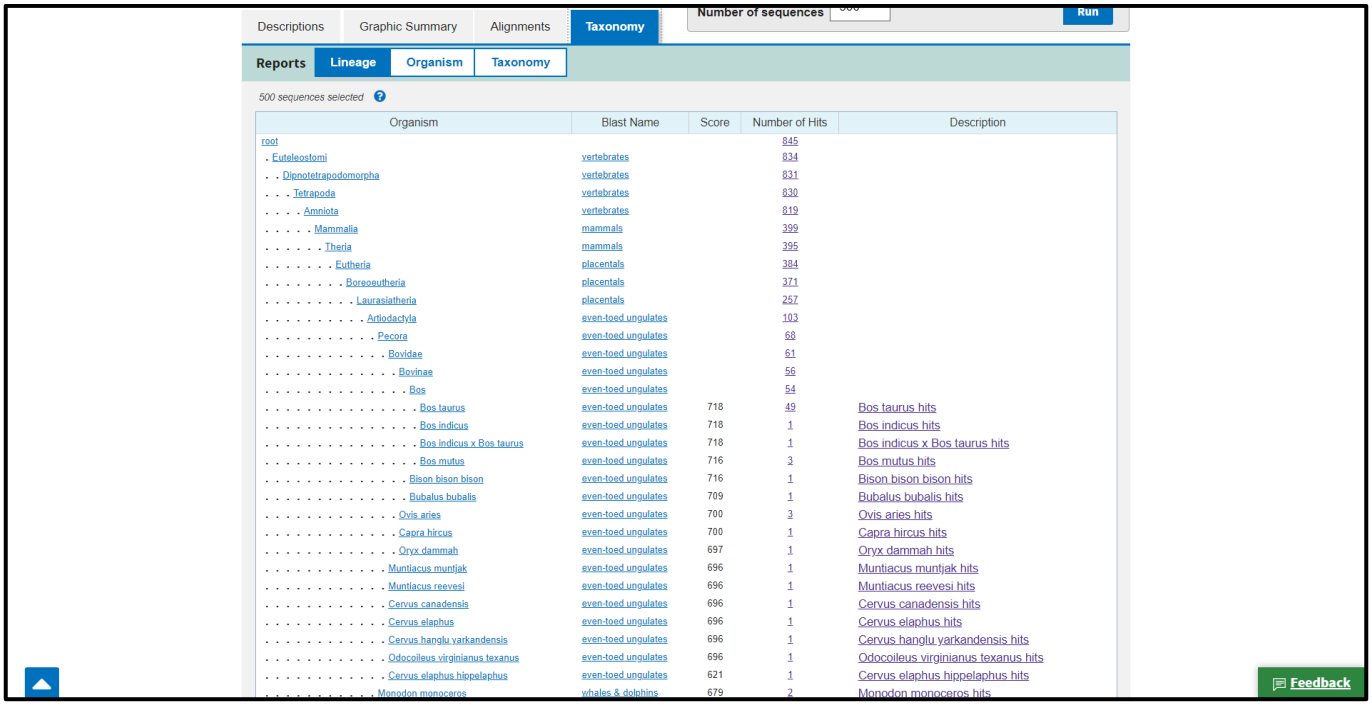


Fig10. Taxonomy for Query Rhodopsin

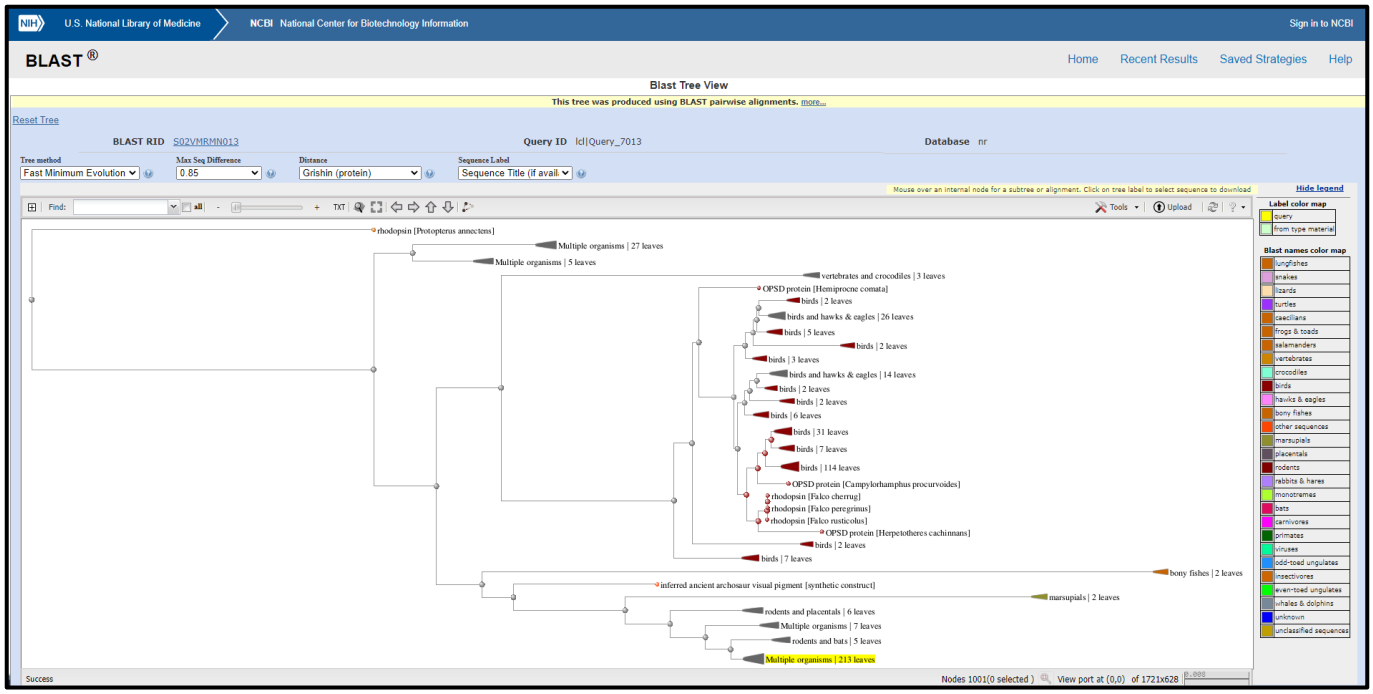


Fig11. Blast tree view for Rhodopsin

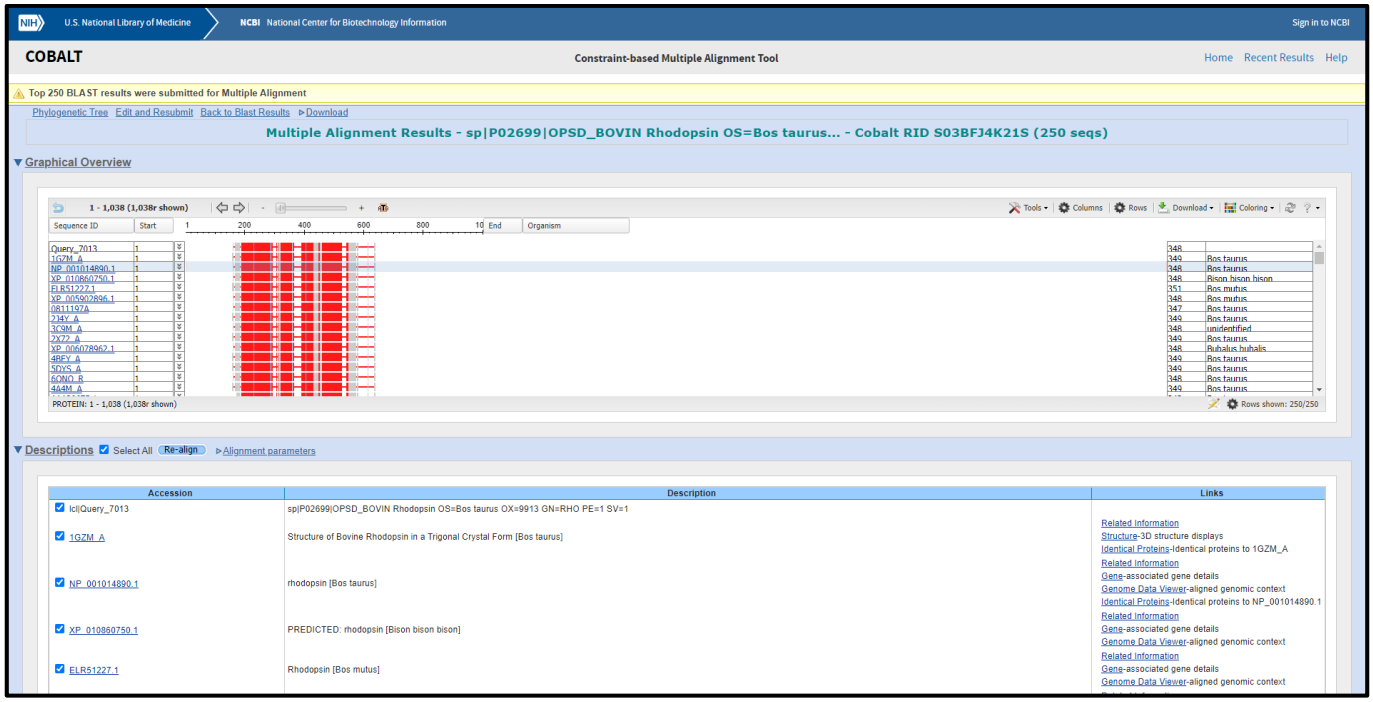


Fig12: Multiple alignment view via cobalt for rhodopsin

Results:

In PHI-BLAST the result is divided into 4 parts. Description, Graphic summary alignment, taxonomy. We can also access other tools from within the result page like BLAST tree view and COBALT.

Conclusion:

PHI-BLAST (Pattern-Hit Initiated BLAST) is a search program that combines matching of regular expressions with local alignments surrounding the match. The most important features of the program have been incorporated into the BLAST software framework.

References:

1. (n.d.). Retrieved from https://www.uniprot.org/uniprot/P02699.fasta
2. Blast Tree View Widget. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/blast/treeview/treeView.cgi?request=page&blastRID=S02VMRMN013&queryID=lcl|Query_7013&entrezLim=&ex=&exl=&exh=&ns=500
3. COBALT:Multiple Alignment Tool. (n.d.). Retrieved from https://www.st-va.ncbi.nlm.nih.gov/tools/cobalt/cobalt.cgi
4. Expasy is operated by the SIB Swiss Institute of Bioinformatics | Terms of Use. (n.d.). PROSITE. Retrieved from https://prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=97965544372.scan.gz
5. Protein BLAST: Search protein databases using a protein query. (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&BLAST_SPEC=&LINK_LOC=blasttab&LAST_PAGE=blastn
6. Rhodopsin. (n.d.). Retrieved from https://www.britannica.com/science/rhodopsin

DATE: 1-11-21

WEBLEM 11/f

(URL: https://blast.ncbi.nlm.nih.gov/Blast.cgi/)

Aim:

To study the query Rhodopsin in PSI BLAST.

Introduction:

A sequence similarity search often provides the first information about a new DNA or protein sequence. A search allows scientists to infer the function of a sequence from similar sequences. There are many ways of performing a sequence similarity search, but probably the most popular method is the "Basic Local Alignment Search Tool" (BLAST) (1, 2). BLAST uses heuristics to produce results quickly. It also calculates an "expect value" that estimates how many matches would have occurred at a given score by chance, which can aid a user in judging how much confidence to have in an alignment. PSI-BLAST first performs a BLASTP search to collect information that it then uses to produce a Position-Specific-Scoring-Matrix (PSSM). A PSSM for a query of length N is an N x 20 matrix. Each of the N columns corresponds to a letter in the query, and each column contains 20 rows. Each row corresponds to a specific residue and describes the probability of related sequences having that residue at that position. PSI-BLAST can then search a database of protein sequences with this PSSM.

Rhodopsin, also called visual purple, pigment-containing sensory protein that converts light into an electrical signal. Rhodopsin is found in a wide range of organisms, from vertebrates to bacteria. In many seeing animals, including humans, it is required for vision in dim light and is located in the retina of the eye specifically, within the tightly packed disks that make up the outer segment of the retina's photoreceptive rod cells, which are specially adapted for vision under low-light conditions. Rhodopsin was discovered in 1876 by German physiologist Franz Christian Boll, who observed that the normally reddish-purple frog retina turned pale in bright light. The fading of colour was later attributed to the destruction of rhodopsin, via a process known as bleaching. Bleaching and the subsequent regeneration of rhodopsin are major steps in the visual cycle—the series of biochemical reactions that is critical for vision in low light.

Methodology:

1. Open the homepage of BLAST.
2. After that click the BLASTp hyperlink
3. Enter the sequence for Rhodopsin taken from uniport
4. Change the program selection algorithm to PSI-BLAST
5. Change the PSI-BLAST threshold value (0.0001)
6. Open the result page.
7. Interpret the results.

Observations:

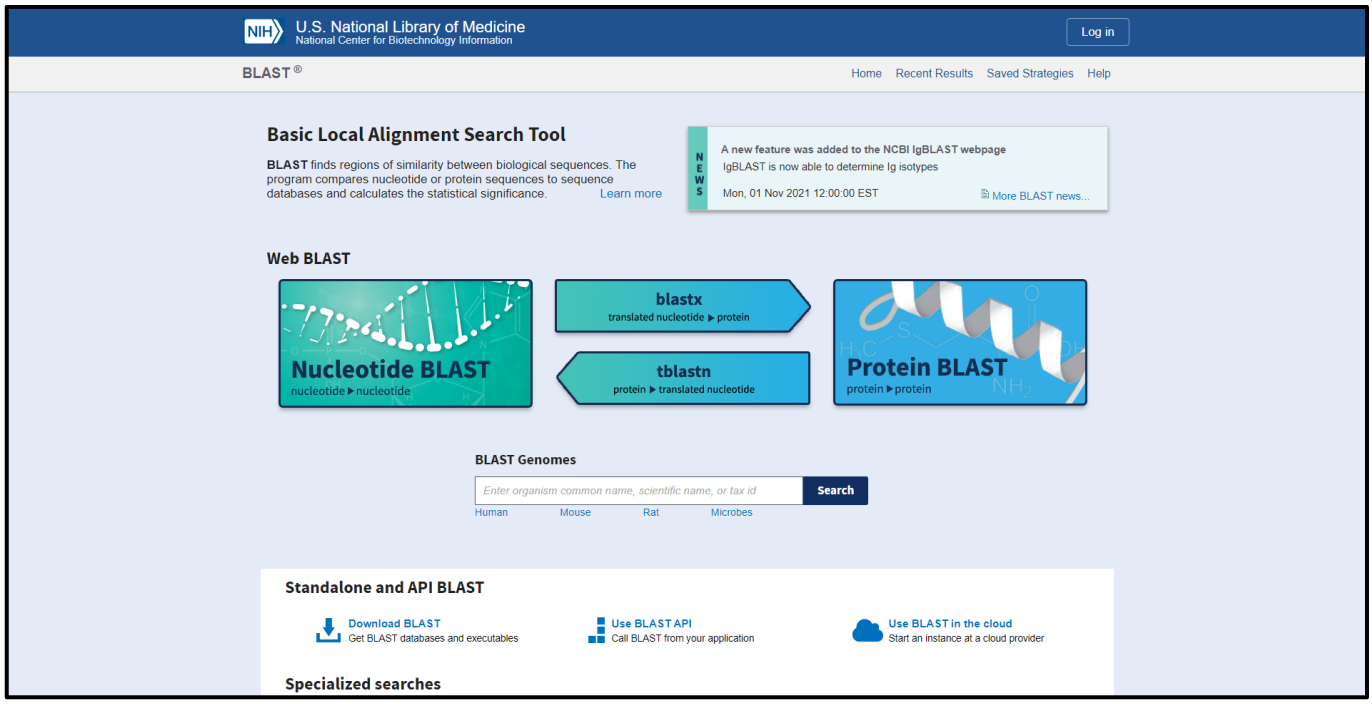


Fig1. Homepage of BLAST

```
>sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX=9913 GN=RHO PE=1 SV=1
MNGTEGPNFYVPFSNKTGVVRSPFEAPQYYLAEPWQFSMLAAYMFLLIMLGFPINFLTLY
VTVQHKKLRTPLNYILLNLAVADLFMVFGGFTTTLYTSLHGYFVFGPTGCNLEGFFATLG
GEIALWSLVVLAIERYVVVCKPMSNFRFGENHAIMGVAFTWVMALACAAPPLVGWSRYIP
EGMQCSCGIDYYTPHEETNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQQQES
ATTQKAEKEVTRMVIIMVIAFLICWLPYAGVAFYIFTHQGSDFGPIFMTIPAFFAKTSAV
YNPVIYIMMNKQFRNCMVTTLCCGKNPLGDDEASTTVSKTETSQVAPA
```

Fig2. FASTA sequence for Rhodopsin

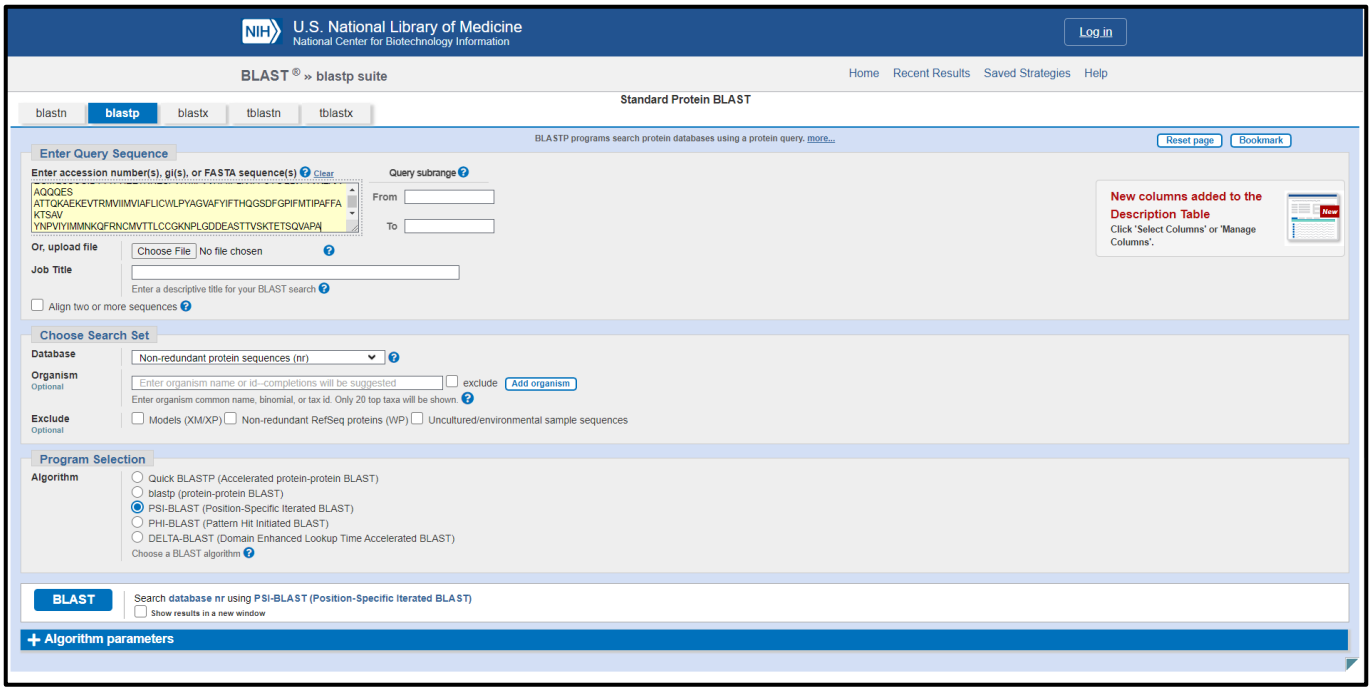


Fig3. Search bar for PSI blast with sequence for rhodopsin taken from uniport

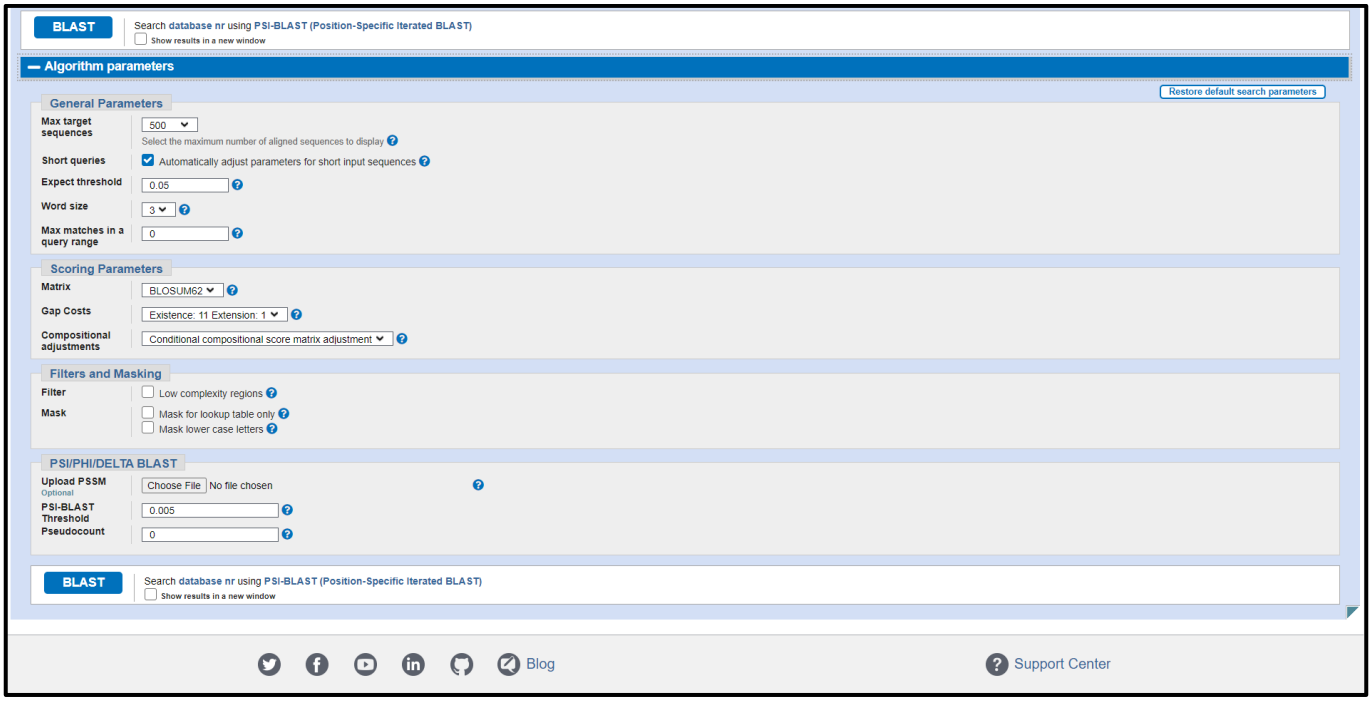


Fig4. Algorithm parameters (General, Scoring, Filters, Masking)

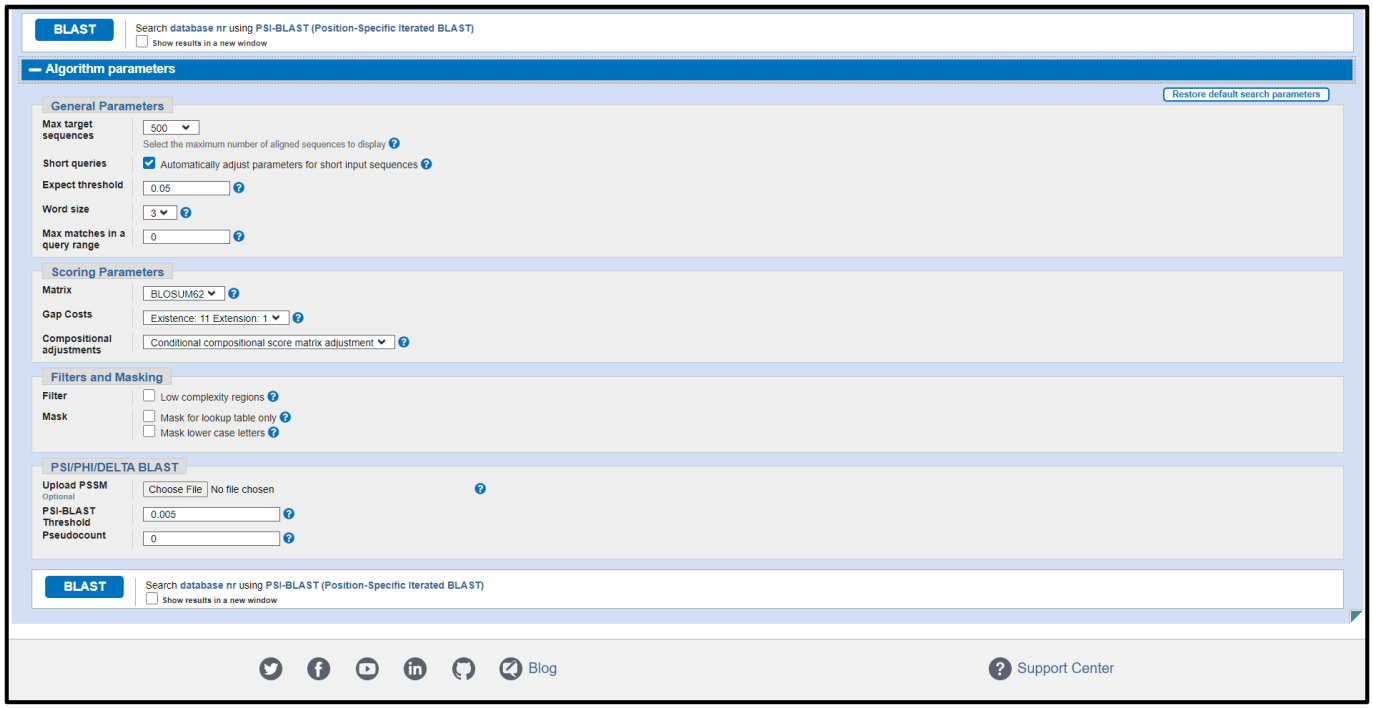


Fig5. Algorithm parameters for PSI/PHI/DELTA BLAST

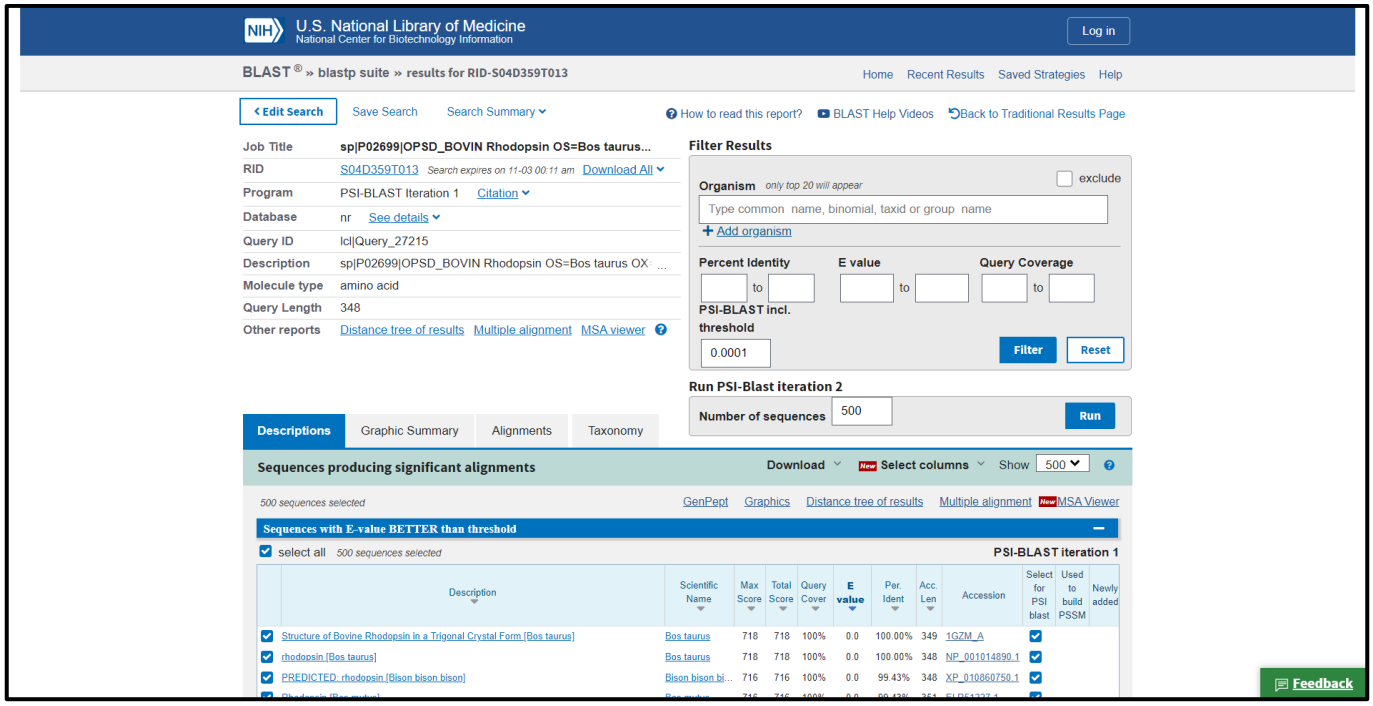


Fig6. Result page (header) for rhodopsin in PSI-BLAST

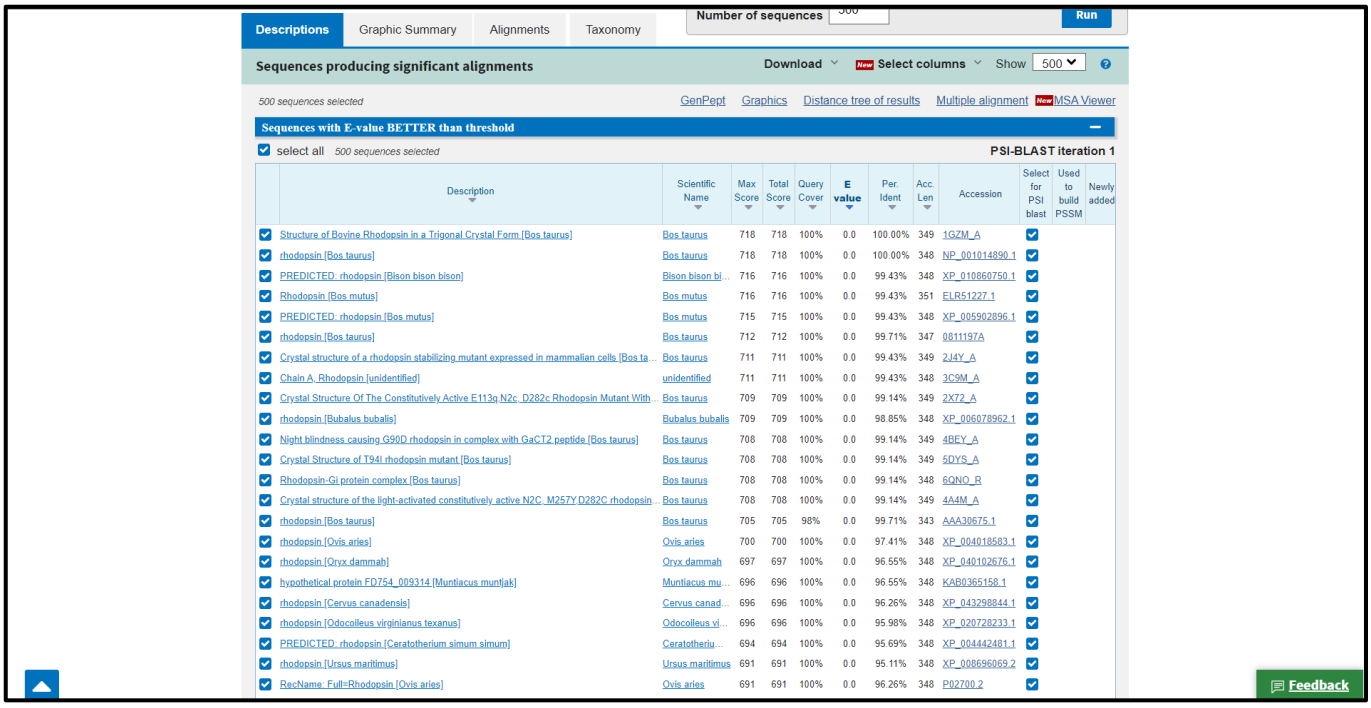


Fig7. Description for query Rhodopsin in PSI-BLAST

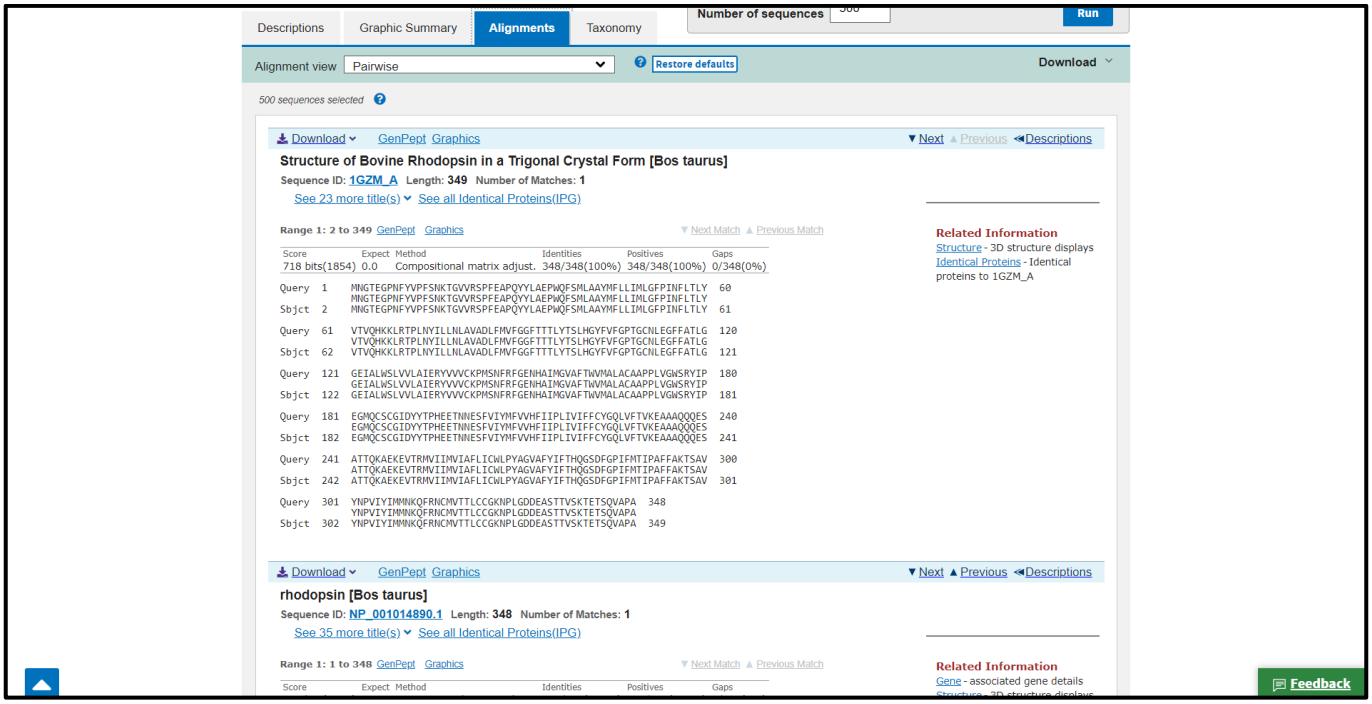


Fig8. Alignments for query Rhodopsin in PSI-BLAST

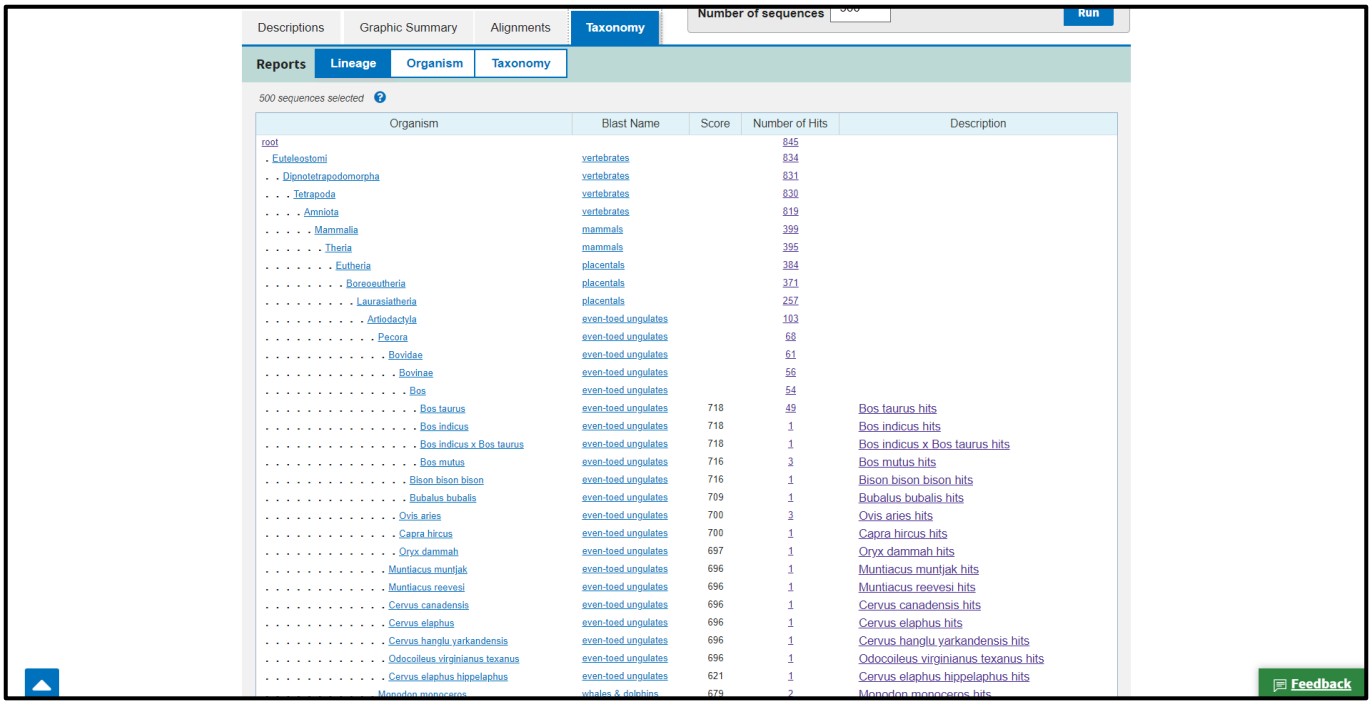


Fig9. Taxonomy for Rhodopsin in PSI-BLAST

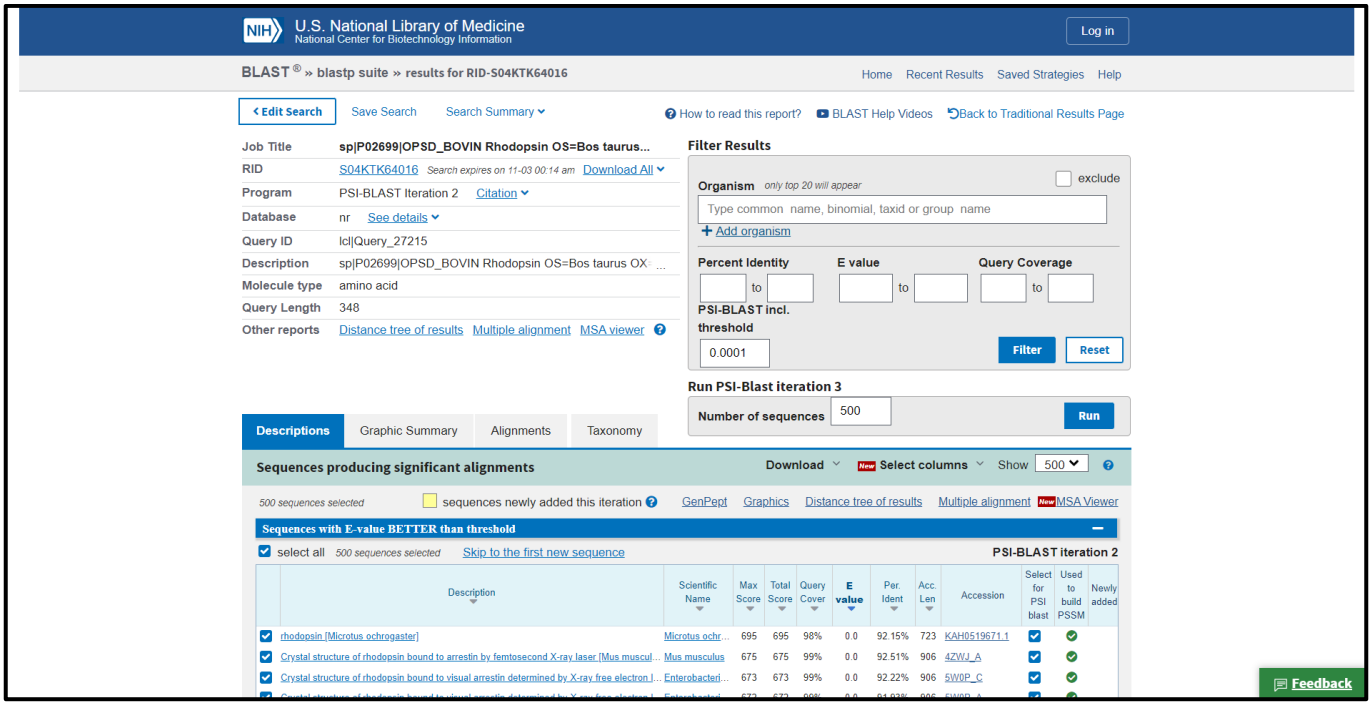


Fig10. Result page for Iteration 1

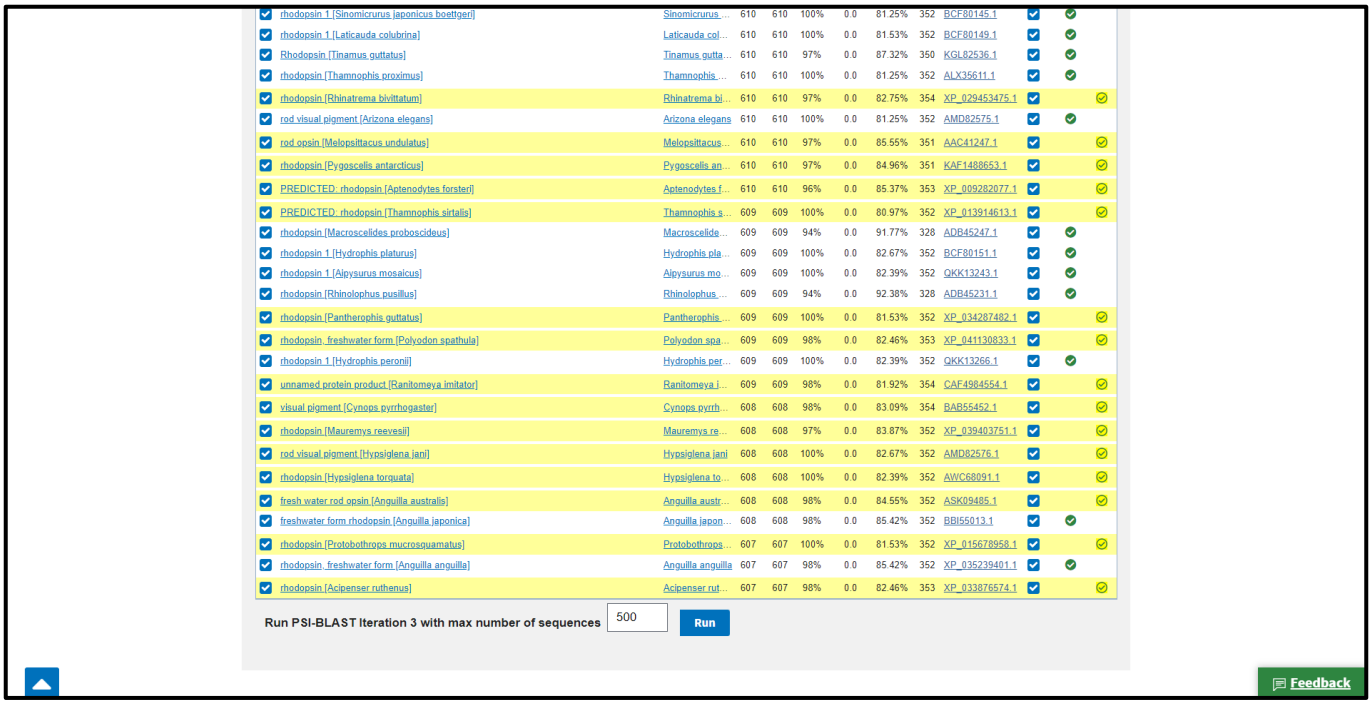


Fig11. Newly added sequences in iteration 1

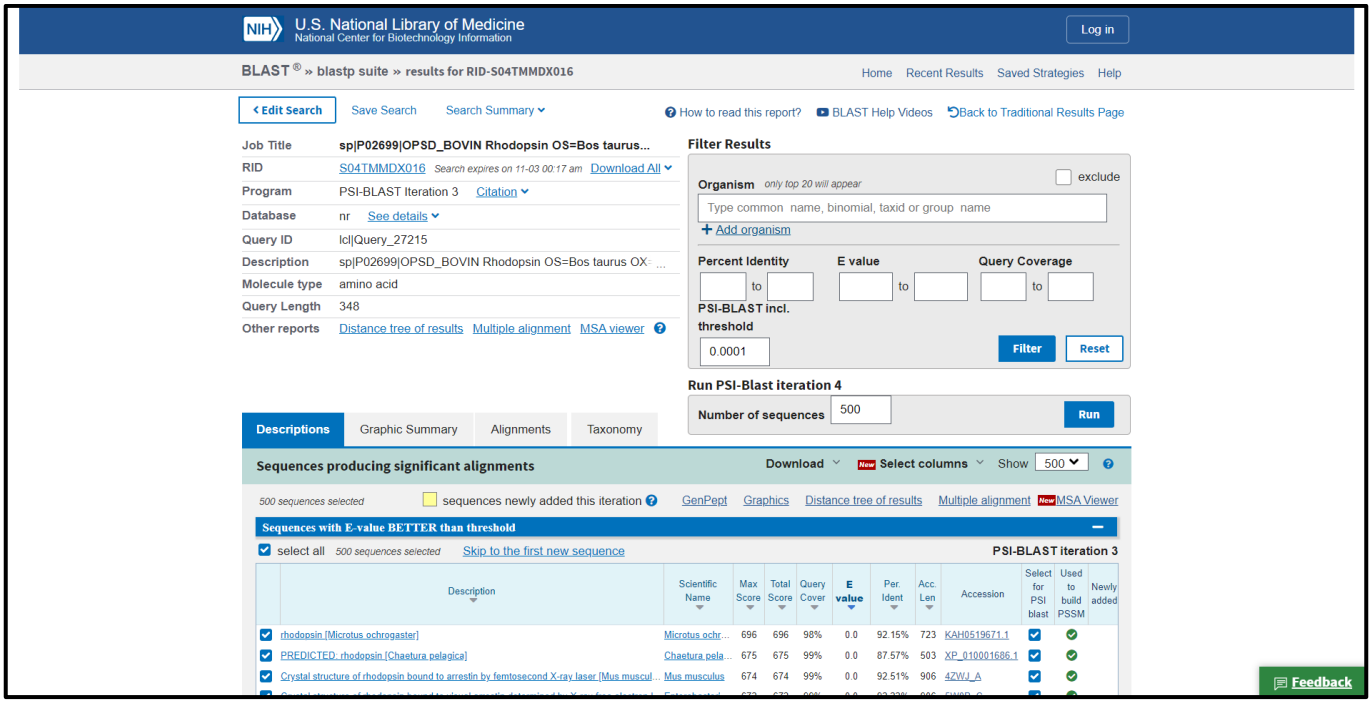
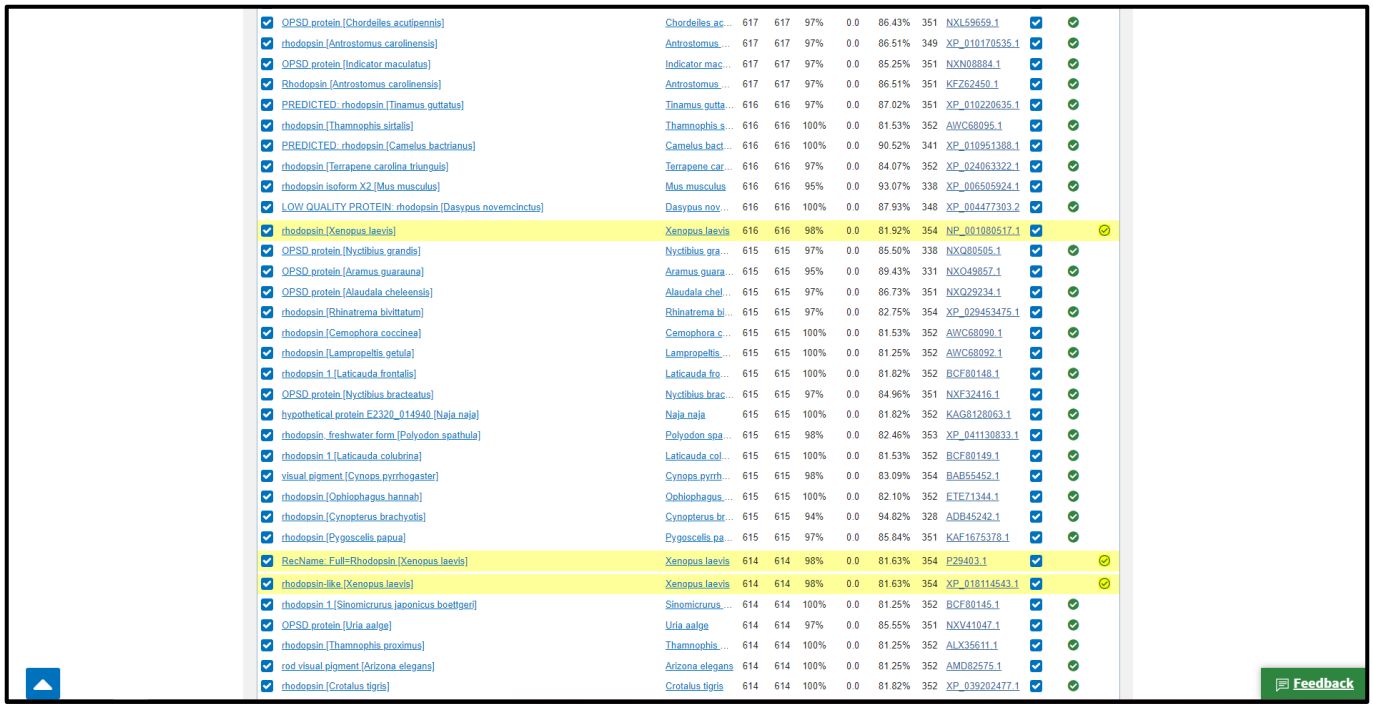


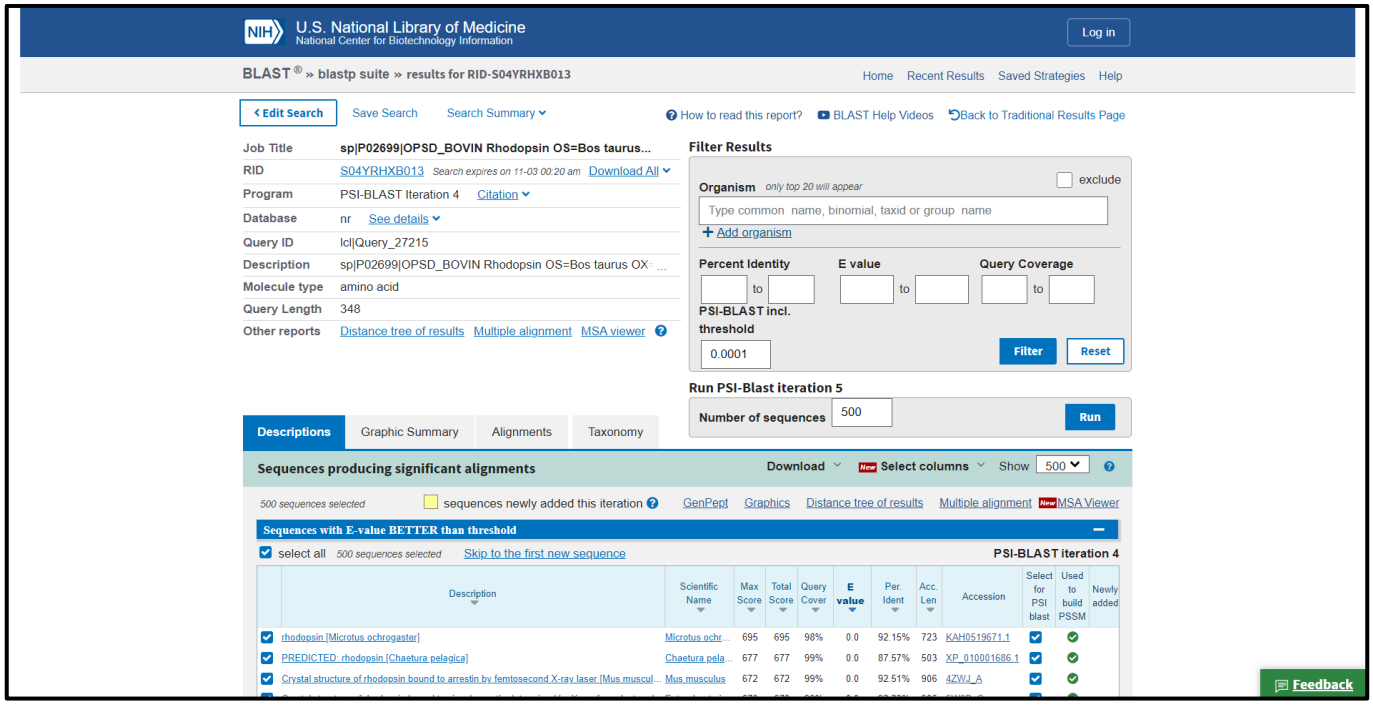
Fig12. Result page for Iteration 2



OPSD protein [Chordeiles acutipennis]	Chordeiles ac…	617	617	97%	0.0	86.43%	351	NXL59659.1	✓	✓
rhodopsin [Antrostomus carolinensis]	Antrostomus …	617	617	97%	0.0	86.51%	349	XP_010170535.1	✓	✓
OPSD protein [Indicator maculatus]	Indicator mac…	617	617	97%	0.0	85.25%	351	NXN08884.1	✓	✓
Rhodopsin [Antrostomus carolinensis]	Antrostomus …	617	617	97%	0.0	86.51%	351	KFZ62450.1	✓	✓
PREDICTED: rhodopsin [Tinamus guttatus]	Tinamus gutta…	616	616	97%	0.0	87.02%	351	XP_010220635.1	✓	✓
rhodopsin [Thamnophis sirtalis]	Thamnophis s…	616	616	100%	0.0	81.53%	352	AWC68095.1	✓	✓
PREDICTED: rhodopsin [Camelus bactrianus]	Camelus bact…	616	616	100%	0.0	90.52%	341	XP_010951388.1	✓	✓
rhodopsin [Terrapene carolina triunguis]	Terrapene car…	616	616	97%	0.0	84.07%	352	XP_024063322.1	✓	✓
rhodopsin isoform X2 [Mus musculus]	Mus musculus	616	616	95%	0.0	93.07%	338	XP_006505924.1	✓	✓
LOW QUALITY PROTEIN: rhodopsin [Dasypus novemcinctus]	Dasypus nov…	616	616	100%	0.0	87.93%	348	XP_004477303.2	✓	✓
rhodopsin [Xenopus laevis]	Xenopus laevis	616	616	98%	0.0	81.92%	354	NP_001080517.1	✓	✓
OPSD protein [Nyctibius grandis]	Nyctibius gra…	615	615	97%	0.0	85.50%	338	NXQ80505.1	✓	✓
OPSD protein [Aramus guarauna]	Aramus guara…	615	615	95%	0.0	89.43%	331	NXO49857.1	✓	✓
OPSD protein [Alaudala cheleensis]	Alaudala chel…	615	615	97%	0.0	86.73%	351	NXQ29234.1	✓	✓
rhodopsin [Rhinatrema bivittatum]	Rhinatrema bi…	615	615	97%	0.0	82.75%	354	XP_029453475.1	✓	✓
rhodopsin [Cemophora coccinea]	Cemophora c…	615	615	100%	0.0	81.53%	352	AWC68090.1	✓	✓
rhodopsin [Lampropeltis getula]	Lampropeltis…	615	615	100%	0.0	81.25%	352	AWC68092.1	✓	✓
rhodopsin 1 [Laticauda frontalis]	Laticauda fro…	615	615	100%	0.0	81.82%	352	BCF80148.1	✓	✓
OPSD protein [Nyctibius bracteatus]	Nyctibius brac…	615	615	97%	0.0	84.96%	351	NXF32416.1	✓	✓
hypothetical protein E2320_014940 [Naja naja]	Naja naja	615	615	100%	0.0	81.82%	352	KAG8128063.1	✓	✓
rhodopsin, freshwater form [Polyodon spathula]	Polyodon spa…	615	615	98%	0.0	82.46%	353	XP_041130833.1	✓	✓
rhodopsin 1 [Laticauda colubrina]	Laticauda col…	615	615	100%	0.0	81.53%	352	BCF80149.1	✓	✓
visual pigment [Cynops pyrrhogaster]	Cynops pyrrh…	615	615	98%	0.0	83.09%	354	BAB55452.1	✓	✓
rhodopsin [Ophiophagus hannah]	Ophiophagus…	615	615	100%	0.0	82.10%	352	ETE71344.1	✓	✓
rhodopsin [Cynopterus brachyotis]	Cynopterus br…	615	615	94%	0.0	94.82%	328	ADB45242.1	✓	✓
rhodopsin [Pygoscelis papua]	Pygoscelis pa…	615	615	97%	0.0	85.84%	351	KAF1675378.1	✓	✓
RecName: Full=Rhodopsin [Xenopus laevis]	Xenopus laevis	614	614	98%	0.0	81.63%	354	P29403.1	✓	✓
rhodopsin-like [Xenopus laevis]	Xenopus laevis	614	614	98%	0.0	81.63%	354	XP_018114543.1	✓	✓
rhodopsin 1 [Sinomicrurus japonicus boettgeri]	Sinomicrurus…	614	614	100%	0.0	81.25%	352	BCF80145.1	✓	✓
OPSD protein [Uria aalge]	Uria aalge	614	614	97%	0.0	85.55%	351	NXV41047.1	✓	✓
rhodopsin [Thamnophis proximus]	Thamnophis…	614	614	100%	0.0	81.25%	352	ALX35611.1	✓	✓
rod visual pigment [Arizona elegans]	Arizona elegans	614	614	100%	0.0	81.25%	352	AMD82575.1	✓	✓
rhodopsin [Crotalus tigris]	Crotalus tigris	614	614	100%	0.0	81.82%	352	XP_039202477.1	✓	✓

Feedback

Fig13. Newly found sequences in Iteration 2



U.S. National Library of Medicine
National Center for Biotechnology Information

Log in

BLAST® » blastp suite » results for RID-S04YRHXB013

Home | Recent Results | Saved Strategies | Help

‹ Edit Search | Save Search | Search Summary

How to read this report? | BLAST Help Videos | Back to Traditional Results Page

Job Title	sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus...
RID	S04YRHXB013 *Search expires on 11-03 00:20 am* Download All
Program	PSI-BLAST Iteration 4 Citation
Database	nr See details
Query ID	lcl|Query_27215
Description	sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus OX…
Molecule type	amino acid
Query Length	348
Other reports	Distance tree of results Multiple alignment MSA viewer

Filter Results

Organism *only top 20 will appear* exclude

Type common name, binomial, taxid or group name

+ Add organism

Percent Identity to | E value to | Query Coverage to

PSI-BLAST incl. threshold 0.0001

Filter | Reset

Run PSI-Blast iteration 5

Number of sequences 500 Run

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download | New Select columns | Show 500

500 sequences selected sequences newly added this iteration | GenPept | Graphics | Distance tree of results | Multiple alignment | New MSA Viewer

Sequences with E-value BETTER than threshold

select all *500 sequences selected* Skip to the first new sequence PSI-BLAST iteration 4

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	Select for PSI blast	Used to build PSSM	Newly added
rhodopsin [Microtus ochrogaster]	Microtus ochr…	695	695	98%	0.0	92.15%	723	KAH0519671.1	✓	✓	
PREDICTED: rhodopsin [Chaetura pelagica]	Chaetura pela…	677	677	99%	0.0	87.57%	503	XP_010001686.1	✓	✓	
Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser [Mus muscul…	Mus musculus	672	672	99%	0.0	92.51%	906	4ZWJ_A	✓	✓	

Feedback

Fig14. Result page for Iteration 3

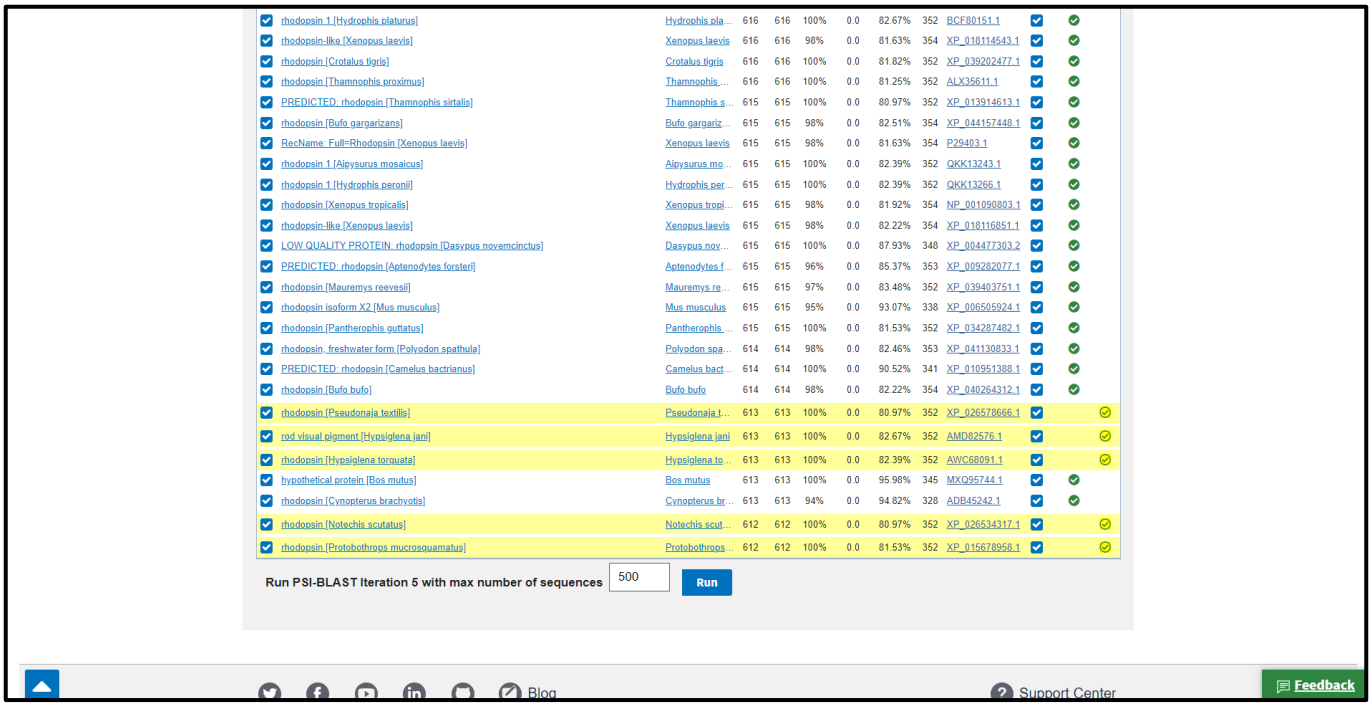


Fig15. Newly found sequences in Iteration 3

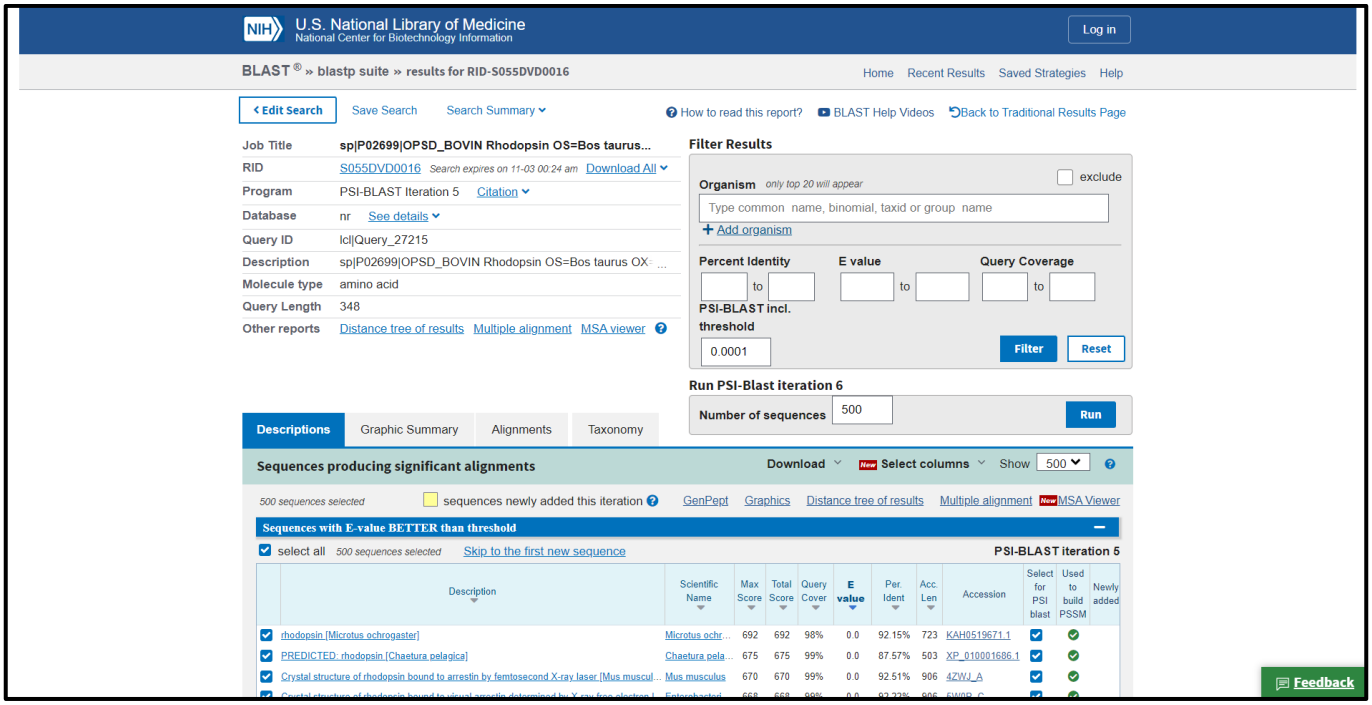


Fig16. Result page for iteration 4

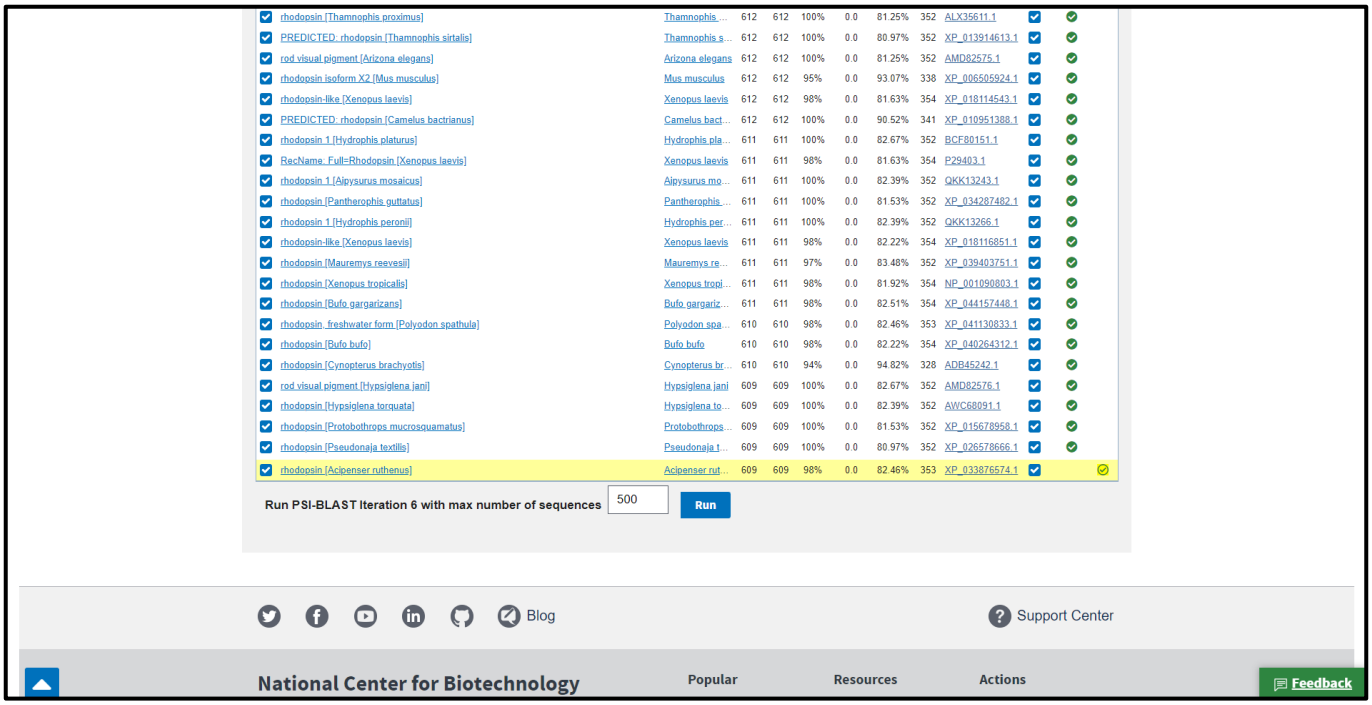


Fig17. Newly found sequences in Iteration 4

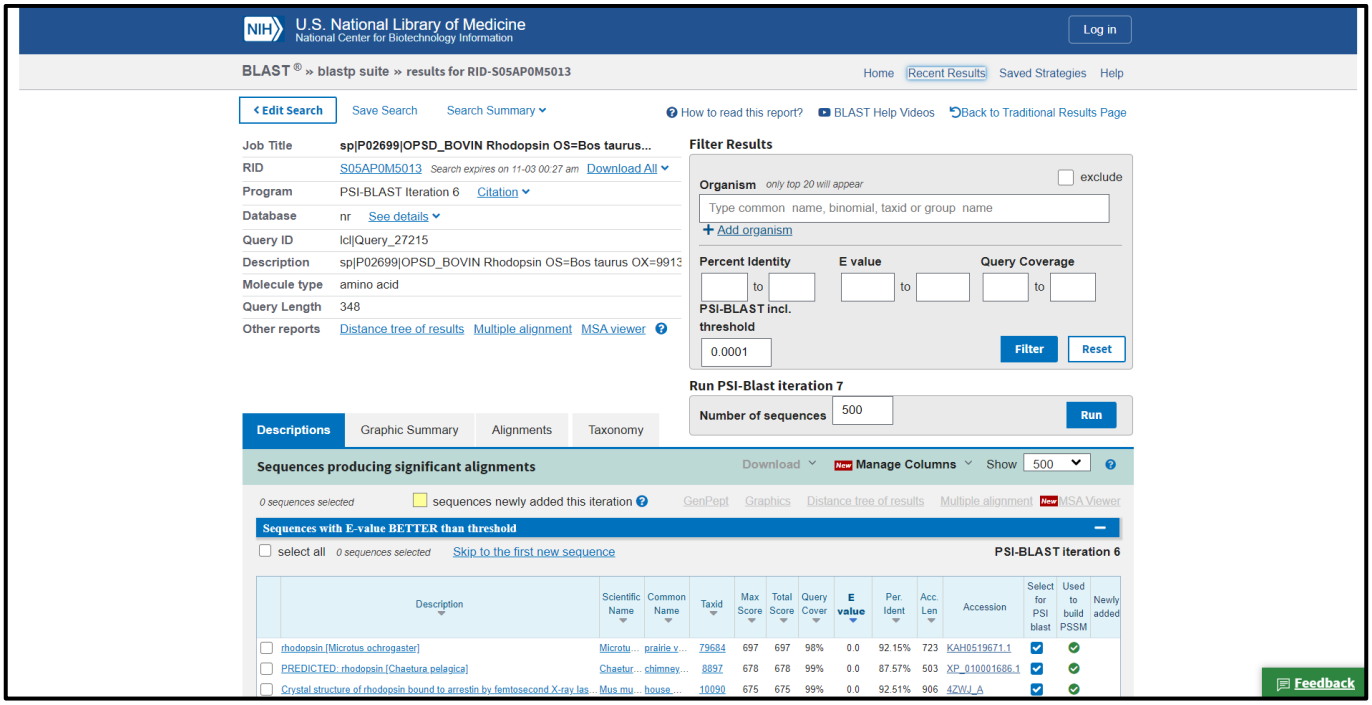


Fig18. Result page for Iteration 5

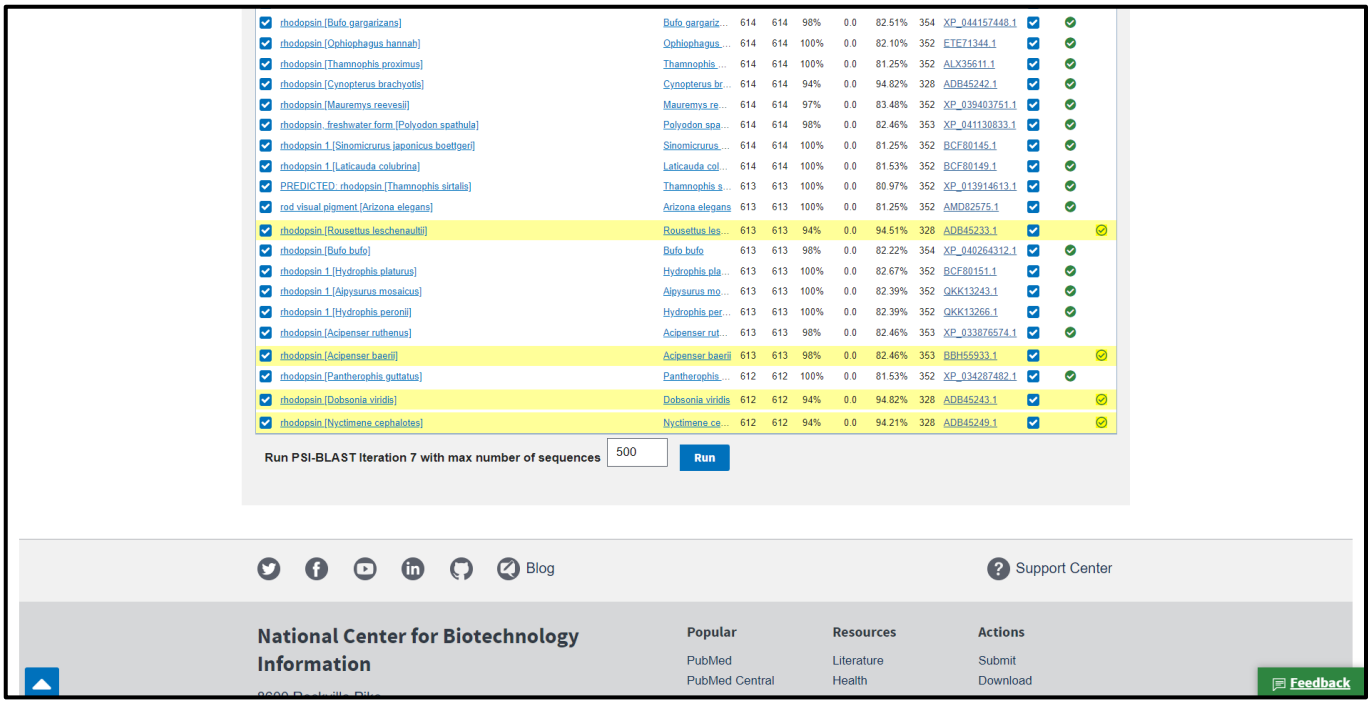


Fig19. Newly found sequences in iteration 5

Results:

In PHI-BLAST the result is divided into 4 parts. Description, Graphic summary alignment, taxonomy. We can also iterate through multiple result pages where the database searches for more and more sequence alignments

Conclusion:

The PSI-blast is used to find distantly related proteins, what you need is the results in the final iteration. The PSI-blast runs in the following five steps: Repeat steps [3] and [4] iteratively, typically 5 times. At each new search, a new PSSM is used as the query. Hope my answer could help you to understand the use of PSI-blast.

References:

1. (n.d.). Retrieved from https://www.uniprot.org/uniprot/P02699.fasta
2. NCBI Blast:sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus... (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=S04D359T013&UNIQ_OBJ_NAME=A_SearchResults_1mhZtz_215p_diTIoYKm2MJ_GTXQl_fSITZ&QUERY_INDEX=0
3. NCBI Blast:sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus... (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=S04KTK64016&UNIQ_OBJ_NAME=A_SearchResults_1mhZxS_8Gf_du382JTU5wJ_GTXQl_1aKLYe&QUERY_INDEX=0
4. NCBI Blast:sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus... (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=S04TMMDX016&UNIQ_OBJ_NAME=A_SearchResults_1mha0S_2tWP_dgi0SViy52a_GTWlT_s8Vrj&QUERY_INDEX=0
5. NCBI Blast:sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus... (n.d.). Retrieved from

https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=S04YRHXB013&UNIQ_OBJ_NAME=A_SearchResults_1mha36_1HG1_dh5iZ3ya7Ag_GTW6B_1xr8fY&QUERY_INDEX=0

6. NCBI Blast:sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus... (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=S055DVD0016&UNIQ_OBJ_NAME=A_SearchResults_1mha6X_2snk_dhWv52Lw3FN_GTWlT_TnXX2&QUERY_INDEX=0
7. NCBI Blast:sp|P02699|OPSD_BOVIN Rhodopsin OS=Bos taurus... (n.d.). Retrieved from https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&VIEW_RESULTS=FromRes&RID=S05AP0M5013&UNIQ_OBJ_NAME=A_SearchResults_1mha9H_12l1_dwXZcxLU55P_GTXQl_1s3jJ9&QUERY_INDEX=0
8. Rhodopsin. (n.d.). Retrieved from https://www.britannica.com/science/rhodopsin