



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

SECB 2103-01

BIOINFORMATICS 1

SEMESTER II 2021/2022

ALTERNATIVE ASSESSMENT

FIND-A-GENE PROJECT

LECTURER: DR HASLINA BINTI HASHIM

SUBMITTED BY:

CHONG KAH WEI

A20EC0027

INDEX

NO.	TITLE	PAGE
1.0	Introduction	1
2.0	BLAST	4
3.0	Gathering information of the “novel” gene	9
4.0	Multiple sequence alignment	12
5.0	Phylogenetic tree creation	26
6.0	Summary and discussion	28

1.0 Introduction

In this project, we are assigned to find a novel gene using BLAST searching. Before beginning this alternative assessment, we have been doing the project in a group. We started the project by choosing a protein of interest. In this alternative assessment, I had selected the human D(1B) dopamine receptor protein with the accession number NP_000789.1. After deciding the protein of interest, I used the protein accession number to carry out TBLASTN searching against Expressed Sequence Tag (est) database and *Aplysia californica* (California sea hare) organism to find the matches. The results gave the gene sequences which are possibly homologous to the D(1B) dopamine receptor protein. From the results, I chose one of the gene sequences which is the CNSN01-F-142926-501 Normalized CNS library (juvenile 1) *Aplysia californica* cDNA clone CNSN01-F-142926 5', mRNA sequence to be further analyzed to confirm whether it is a novel gene or not. In this case, I used the sequence that was retrieved from the TBLASTN results to carry out BLASTP searching against the non-redundant (nr) protein database. From the BLASTP result, I observed that there is no protein that is 100% identical to the searched sequence. The top result of the match is 98.30%. Hence, I can conclude that there is a chance that CNSN01-F-142926-501 Normalized CNS library (juvenile 1) *Aplysia californica* cDNA clone CNSN01-F-142926 5', mRNA sequence. After getting the novel protein, I performed the next step which is to gather the information of the novel gene. After that, I generate a multiple sequence alignment using the query protein, the novel protein and a few members of the family using MUSCLE and TCOFFEE. By then, I obtained the alignments with similar length. Using the alignments, I also did a phylogenetic analysis to generate a phylogenetic tree to observe the evolutionary relationship of the protein species.

Basically, we can use any known protein sequence as a starting point to get a novel protein. In this case, I am quite interested in the human dopamine receptor and I found that the D(1B) dopamine receptor protein is quite meaningful. Dopamine receptors are G-protein coupled receptors (GPCRs) that play an important role in a human's daily life. The protein is involved in many neurological processes. There are several types of dopamine receptors too, including D1, D2, D3, D4 and D5 and every receptor has different functions. Meanwhile, there is a human gene called DRD5 and it is used in making the D5 receptors,

which is also known as the D(1B) dopamine receptor protein. The protein is highly homologous to the D1 receptor and it can be considered as a D1-like receptor. Besides, D(1B) dopamine receptor protein is located in the human brain and is used in human decision making, cognition, attention and renin secretion. Thus, I found that the protein is actually quite important for humans and I decided to use the protein as the query protein of the TBLASTN searching, which is the very first step in this project. After deciding the protein of interest, I get the details such as the accession number, species and also the sequences of the protein from the NCBI protein page as shown in Figure 1.1 and Figure 1.2. By then, I am able to get the accession number of the particular protein to do further analysis.

The screenshot shows a web browser window with the URL [ncbi.nlm.nih.gov/protein/NP_000789.1](https://www.ncbi.nlm.nih.gov/protein/NP_000789.1). The page title is "D(1B) dopamine receptor [Homo sapiens]". Key details include:

- NCBI Reference Sequence:** NP_000789.1
- Identical Proteins:** FASTA, Graphics
- Sequence Information:** LOCUS NP_000789, 477 aa, linear, PRI 17-FEB-2021
- Organism:** Homo sapiens
- Keywords:** RefSeq; MANE Select.
- References:**
 - Bai R, Mei J and Hu W. Hypermethylation of DRD5 Promoter Is a Biomarker Across 12 Cancer Types. *DMO Cell Biol* 39 (11), 2052-2058 (2020)
 - Lee CG, Moon H and Park S. The effects of dopamine receptor genes on the trajectories of sport participation from adolescence through young adulthood. *Am Hum Biol* 47 (3), 256-262 (2020)
- Articles about the DRD5 gene:**
 - Hypermethylation of <i>DRD5</i> Promoter Is a Biomarker Across 12 Cance [DNA Cell Biol. 2020]
 - Role of Thioredoxin 1 in Impaired Renal Sodium Excretion of HD ₅ [J Am Heart Assoc. 2019]
 - Activation of DRD5 (dopamine receptor D5) inhibits tumor growth by autopl [Autophagy. 2017]
- See all...**
- Reference sequence information:** RefSeq genomic sequence, See the genomic reference sequence for the DRD5 gene (NG_012024.1).
- RefSeq mRNA:**

Figure 1.1: D(1B) dopamine receptor protein

The screenshot shows a web browser window displaying the sequence of the D(1B) dopamine receptor protein (NP_000789.1). The sequence is presented in a multi-line text format, with various regions and sites highlighted in blue. The sequence starts with '1 mlppgsngta ypgqfalyqq laqnavggg agappgpgsql vvtaclltl iitwllgnvl' and continues through several hundred positions. Regions and sites are annotated with identifiers like 'Region', 'Site', and 'CDS'. The browser's address bar shows the URL: ncbi.nlm.nih.gov/protein/NP_000789.1. The page includes standard navigation buttons, a search bar, and a 'Support Center' link. At the bottom, there is a dark bar with icons for weather (25°C Rain showers), system status (ENG), and date/time (10:59 AM 6/23/2022).

Figure 1.2: Sequence of D(1B) dopamine receptor protein

2.0 BLAST

Basic Local Alignment Search Tool (BLAST) is the main NCBI tool that allows rapid sequence comparison of a query sequence against a database. We can say that BLAST is actually one of the basic ways to learn about a protein or gene since the searching result will reveal all related sequences and calculate the statistical significance of all matches. In this case, NCBI provides a series of BLAST programs, including BLASTP, BLASTN, BLASTX, TBLASTX and also TBLASTN for different use. In this instance, BLASTP is used to compare a protein query to a database of proteins and the searching result will display a list of related protein sequences while BLASTN is utilized in comparing DNA query against a DNA database. On the other hand, BLASTX is carried out to compare a DNA query against a protein database. The program performs translation of a DNA sequence into six protein sequences using all six possible reading frames and then compares each of the proteins to a protein database. TBLASTN is the reverse of BLASTX since it is used when comparing a protein query to a nucleotide database. The program translates the DNA sequence in a database into six potential proteins and then compares the protein query against each of the translated proteins. TBLASTX is the most computationally intensive program in BLAST. This particular program is able to translate a DNA query from both a query and a database into six potential proteins and perform a 36 protein-protein database search. The result of TBLASTX will reveal genes that encode proteins homologous to the query.

The BLAST program that I used in this project is TBLASTN and BLASTP. TBLASTN program is used as a starting point to get a novel protein. After confirming the protein of interest, which is the NP_000789.1 D(1B) dopamine receptor protein, I use the accession number to carry out TBLASTN searching against *Aplysia californica* (California sea hare) est database as shown in Figure 2.1 and the searching result generates a list of related genes. Part of the searching results are shown in Figure 2.2 and Figure 2.3.

Enter Query Sequence

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

NP_000789.1

Query subrange ?

From [] To []

Or, upload file Choose File No file chosen ?

Job Title NP_000789-D(1B) dopamine receptor [Homo sapiens]

Enter a descriptive title for your BLAST search ?

Align two or more sequences ?

Choose Search Set

Database Expressed sequence tags (est)

Organism Aplysia californica (taxid:6500) exclude Add organism

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

Limit to Sequences from type material

Entrez Query Enter an Entrez query to limit search ?

BLAST

Search database est using Tblastn (search translated nucleotide databases using a protein query)
Show results in a new window

Figure 2.1: Input of TBLASTN

	Description	Scientific Name	Score	Cover	value	Ident	Len	Accession
<input checked="" type="checkbox"/>	CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5'... Aplysia californica	Aplysia californica	119	119	42%	3e-30	35.47%	667 EB335661.1
<input checked="" type="checkbox"/>	CNSN01-F-102105-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-102105 5'... Aplysia californica	Aplysia californica	110	110	43%	4e-27	33.01%	742 EB348863.1
<input checked="" type="checkbox"/>	CNSN01-C-000413-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-C-000413 5'... Aplysia californica	Aplysia californica	108	108	26%	5e-26	43.75%	792 EB260814.1
<input checked="" type="checkbox"/>	CNSN01-F-062370-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-062370 5'... Aplysia californica	Aplysia californica	85.5	85.5	24%	2e-18	39.17%	624 EB346158.1
<input checked="" type="checkbox"/>	CNSN01-F-063010-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-063010 5'... Aplysia californica	Aplysia californica	73.6	73.6	31%	3e-14	28.76%	638 EB346226.1
<input checked="" type="checkbox"/>	PEG002-C-011987-301 Normalized Pedal-Pleural Ganglia Aplysia californica cDNA clone PEG002-C-011987 3'... Aplysia californica	Aplysia californica	70.5	70.5	17%	1e-13	38.82%	514 EB214272.1
<input checked="" type="checkbox"/>	CNSN01-F-066686-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-066686 5'... Aplysia californica	Aplysia californica	71.2	71.2	43%	2e-13	26.64%	688 EB346577.1
<input checked="" type="checkbox"/>	CNSN01-F-048618-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-048618 5'... Aplysia californica	Aplysia californica	67.4	67.4	27%	2e-12	30.30%	542 EB297304.1
<input checked="" type="checkbox"/>	PEG003-C-228120-501 Normalized Pedal-Pleural Ganglia Aplysia californica cDNA clone PEG003-C-228120 5'... Aplysia californica	Aplysia californica	68.9	68.9	21%	2e-12	42.45%	771 EB245546.1
<input checked="" type="checkbox"/>	CNSN01-F-002002-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-002002 5'... Aplysia californica	Aplysia californica	65.9	65.9	18%	2e-12	35.56%	380 EB282651.1
<input checked="" type="checkbox"/>	PEG003-C-217630-501 Normalized Pedal-Pleural Ganglia Aplysia californica cDNA clone PEG003-C-217630 5'... Aplysia californica	Aplysia californica	68.9	68.9	21%	2e-12	42.45%	800 EB241139.1
<input checked="" type="checkbox"/>	PEG003-C-205407-501 Normalized Pedal-Pleural Ganglia Aplysia californica cDNA clone PEG003-C-205407 5'... Aplysia californica	Aplysia californica	68.9	68.9	21%	2e-12	42.45%	822 EB230890.1
<input checked="" type="checkbox"/>	CNSN01-C-004354-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-C-004354 5'... Aplysia californica	Aplysia californica	68.2	68.2	18%	4e-12	34.07%	799 EB260652.1
<input checked="" type="checkbox"/>	CNSN01-F-089997-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-089997 5'... Aplysia californica	Aplysia californica	67.0	67.0	16%	4e-12	37.35%	617 EB322814.1
<input checked="" type="checkbox"/>	CNSN01-C-008362-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-C-008362 5'... Aplysia californica	Aplysia californica	66.2	66.2	18%	6e-12	34.07%	582 EB258151.1
<input checked="" type="checkbox"/>	CNSN01-F-030457-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-030457 5'... Aplysia californica	Aplysia californica	65.9	65.9	18%	7e-12	34.44%	559 EB342970.1
<input checked="" type="checkbox"/>	CNSN01-C-001435-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-C-001435 5'... Aplysia californica	Aplysia californica	65.5	65.5	18%	1e-11	34.44%	561 EB255570.1
<input checked="" type="checkbox"/>	G1045P35RP14.T1 Aplysia californica Pooled Normalized Library Aplysia californica cDNA mRNA sequence	Aplysia californica	67.0	67.0	18%	1e-11	34.07%	847 GD208644.1
<input checked="" type="checkbox"/>	CNSN01-F-012764-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-012764 5'... Aplysia californica	Aplysia californica	65.9	65.9	27%	2e-11	31.06%	665 EB290699.1
<input checked="" type="checkbox"/>	CNSN01-F-010192-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-010192 5'... Aplysia californica	Aplysia californica	65.5	65.5	18%	2e-11	33.33%	618 EB288905.1
<input checked="" type="checkbox"/>	CNSN01-F-105086-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-105086 5'... Aplysia californica	Aplysia californica	62.8	62.8	18%	3e-11	33.33%	394 EB326306.1

Figure 2.2: Results of TBLASTN

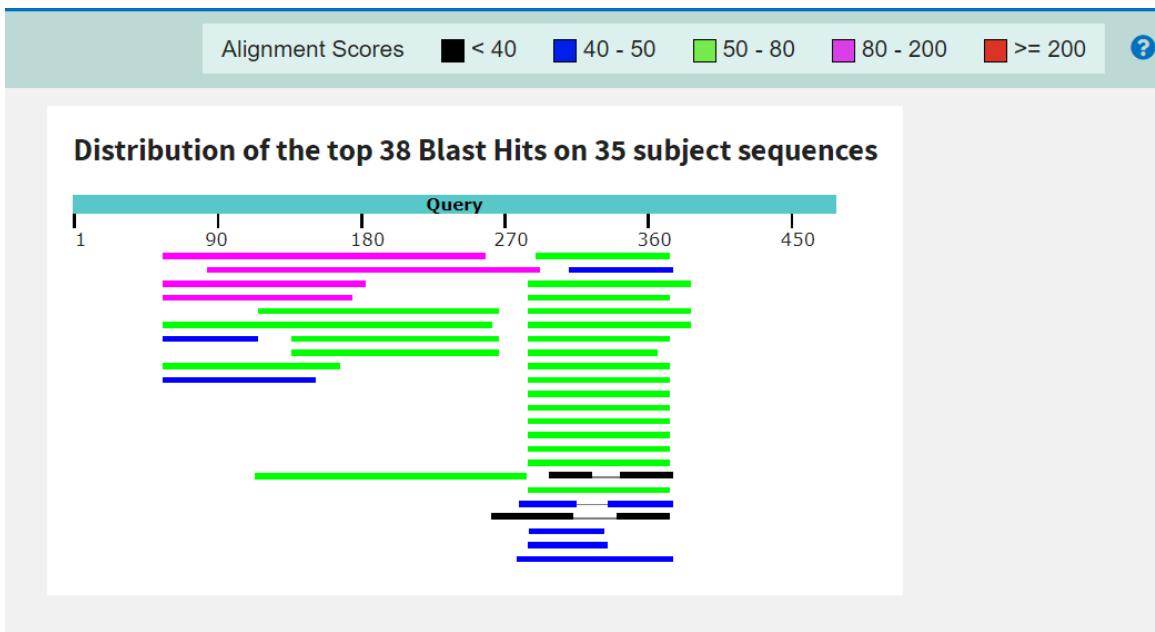


Figure 2.3: Graphic summary of TBLASTN

In this step, I have to distinguish the different results to obtain a sequence which might be a novel gene. In this case, a perfect match to my query sequence would be a sequence that is not novel, a near match would be a gene that might be novel depending on the results of further analysis and the non-homologous results would be not related to the query sequence. To differentiate the results and get a possible novel protein sequence, we need to check the E-value, maximum score, total score, percent query coverage and percent identity. E-value is a number that describes the expected hits of similar score that are found by chance in a database. In this case, the significant alignments should have an E-value that is very close to 0. Aside from that, the maximum score is the highest alignment score between the query sequence and the database sequence while the total score is the sum of alignments scores from the same database sequence that match the query sequence. Nevertheless, the percent query coverage is the percentage of how much the query sequence is covered by the match sequence and it is able to reveal how long the sequences relate to each other. On the other hand, percent identity is the percentage that describes how similar the query sequence is to the target sequence. Basically, the smaller the E-value, the larger the score, the more significant the matches. Besides, the larger percent query coverage and the percent identity, also, the higher the chance to get a better match. From the TBLASTN result, there are actually 35 matches have been generated. In this case, if the match is a perfect match, it should have an E-value which is very close to 0, a high bit

score and also 100% identical to the query sequence. However, the matches of my result can be considered as near matches since their E-value are not too large and the total scores also not too low as well as the percent identity are not 100% and in an acceptable range. On the other hand, a non-homologous sequence is an insignificant match that is not related to the query sequence. These results might get a very high E-value, very low bit score as well as a very low percent identity. Among the matches, I have chosen a match among the near matches which have a chance of becoming a novel protein. The match that I have chosen for further analysis is the top match, which is CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5', mRNA sequence since it is the most significant match. The E-value of this match is 3e-30, the max score is 119 and the percent identity is 35.47%. The alignment of this match can be seen in Figure 2.5.

The screenshot shows a BLAST search result for the CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5', mRNA sequence. The sequence ID is EB335661.1, length is 667, and there is 1 match. The alignment details are as follows:

Range 1: 69 to 596	GenBank	Graphics	Next Match	Previous Match		
Score: 119 bits(297)	Expect: 3e-30	Method: Compositional matrix adjust	Identities: 72/203(35%)	Positives: 112/203(55%)	Gaps: 28/203(13%)	Frame: +3
Query 57	GNVLVCAAIIVRSRHLRANMTNVFIVSLAVSDLFALLVMPWKA GN+ V AI+-+ L+ ++H I+ LAV+DL VA+LVMP + E++ +W G A CD+	Sbjct 69	GNIFVITAILVEKSLQ-GVSNYLILPLAVTDLLVAVLVMPLSLIYEISIHWFGLNAVCDM	245		
Query 116	WVAFDIMCSTASILNLCVISVDRYHAISRPFRYRKHMORMALMVNGLA WV+ D+C TASIL+I I+ DRYHA+S Y R L-MV + NT+S+ IS P	Sbjct 246	WVSMDVLCCSTASILHLVAIAFDRYWAVSN-IDYVRSRNARQILLMVAIVWTVSVFISMPP	422		
Query 176	VQLNNHRDQAASWGLDPNNLNWTPEEDFWEPDVNAENDSSLNR W D +P++ + C S + Y + S++ +FY	Sbjct 423	L-----FGWRHDSDDPELTGQ-CLISQDHGYTVFSTVGAFY	527		
Query 236	IPVAIMIVTYTRIYRIAQVQIRR	Sbjct 528	528 CPLLLMFVINFKIYRAARYRIRK	596		

Figure 2.5: Alignment of the match

After interpreting the data, I have considered this match as a near match which might be a novel gene and use it to proceed my project. Also, I retrieved the sequence of this protein from the TBLASTN result to perform BLASTP search against non-redundant (nr) database to check whether it is a novel protein. The particular sequence is as shown below:

```
> Aplysia californica protein (sequence taken from BLAST result)
GNIFVITAILVEKSLQ-GVSNYLILPLAVTDLLVAVLVMPLSLIYEISIHWFGLNAVCDM
WVSMDVLCCSTASILHLVAIAFDRYWAVSN-IDYVRSRNARQILLMVAIVWTVSVFISMPP
L-----FGWRHDSDDPELTGQ-CLISQDHGYTVFSTVGAFY
CPLLLMFVINFKIYRAARYRIRK
```

Then, what I have done is checking the percent identity of the results. If there is a match with 100% percent identity to a protein in the database from the same species, then the protein is not novel. In contrast, if the match is 100% but it is to a different species than the one that I started with, it is a novel protein. On the other hand, if there is no match that is 100% identical to the searched sequence, then it is likely the protein is novel. However, if there are no database matches to the original query, it means that the sequence is not actually homologous to the original query. For my result as shown in Figure 2.6, I can see that the top result is 98.30% identical to the *Aplysia californica* protein that I have taken from the previous BLAST result. I can also get the match to the original query, which is the human D(1B) dopamine receptor from the result if I restrict the database to human nr as shown in Figure 2.7. Thus, I can conclude that the *Aplysia californica* protein that I have obtained is a novel protein which is homologous to my query protein.

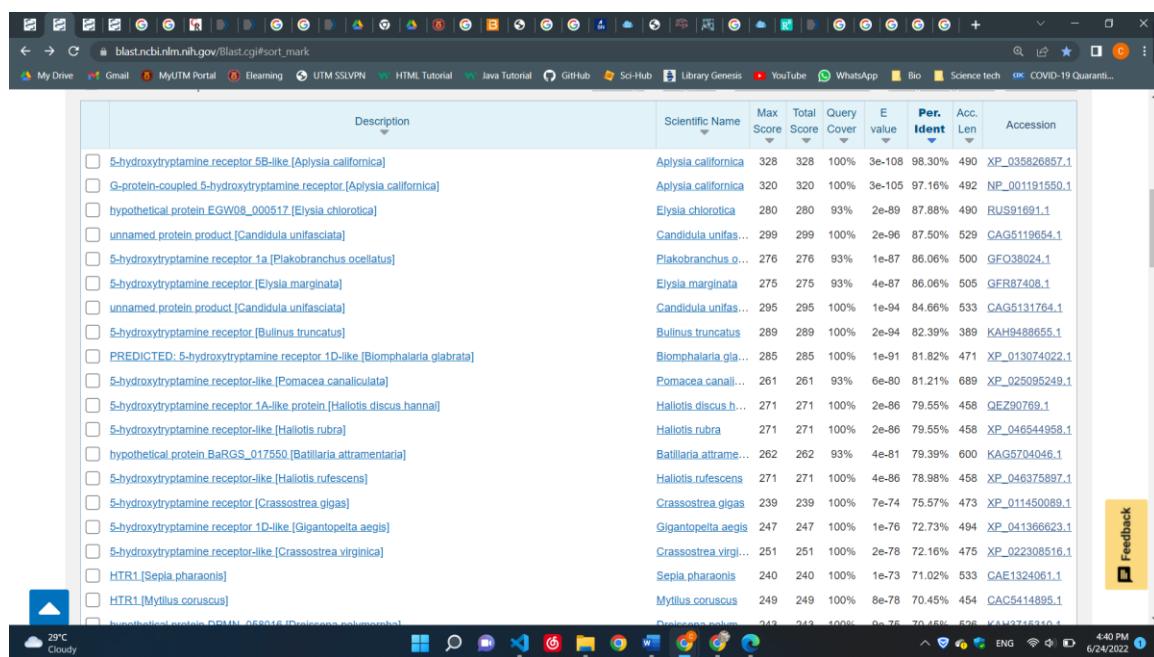


Figure 2.6: BLASTP result of *Aplysia californica* protein

<input checked="" type="checkbox"/> Crystal structure of 5-HT2AR in complex with risperidone [<i>Homo sapiens</i>]	<i>Homo sapiens</i>	97.8	97.8	93%	3e-23	39.52%	376	6A93_A
<input checked="" type="checkbox"/> D(1B) dopamine receptor [<i>Homo sapiens</i>]	<i>Homo sapiens</i>	98.6	98.6	100%	3e-23	35.47%	477	NP_000789.1
<input checked="" type="checkbox"/> Chain A_ 5-hydroxytryptamine receptor 2A_5-hydroxytryptamine receptor 2A_5-hydroxytryptamine receptor 2A_Solu...	<i>Homo sapiens</i>	97.8	97.8	93%	3e-23	39.52%	376	7WC4_A

Figure 2.7: Match to original query

3.0 Gathering the information of the “novel” gene

The gathering of information about the “novel” gene is a significant process since all of us should know the details of the gene that we found and used in the project. Through gathering the information of the “novel” gene, we can find out more important information about the sequence that might be undiscovered before. *Aplysia Californica* is actually a type of invertebrate organism and we can say that it is totally different with human. However, scientist has found that there are genes which involved in similar neural activities in the brain of both organisms, and this finding has boosted the further investigation of neurological disease such as Parkinson’s and Alzheimer’s disease. In this case, the novel protein is actually a match of human D(1B) dopamine receptor, which is also one of the gene that exist in human brain. By gathering the information of the novel gene, we might discover some useful pieces that might be able to contribute to the biological field.

Throughout the process to gather the information of this novel gene, I was interested in investigating whether the gene has same characteristics with human D(1B) dopamine receptor. However, when I was doing research about this gene, there was only a limited resource and information about this novel gene. To know more about the novel protein, I got the details from the NCBI GenBank as shown in Figure 3.1 and Figure 3.2. The basic information, including the accession number, source, sequence and so on can actually be found in this page. Besides, I also referred to the PubMed journal as shown in Figure 3.3 through the link that is provided in the NCBI page to know more about this novel gene.

Figure 3.1: Information of Aplysia Californica

Figure 3.2: Information of Aplysia Californica

The screenshot shows a web browser window with the following details:

- Title Bar:** CNSN01-F-142926-501 Normal; Neuronal Transcriptome of Aplysia
- Address Bar:** sciencedirect.com/science/article/pii/S0092867406015959/via/3Dihub
- Page Content:**
 - ScienceDirect Logo:** ScienceDirect
 - Navigation:** Journals & Books, Search, Register, Sign in
 - Journal Information:** Cell, Volume 127, Issue 7, 29 December 2006, Pages 1453-1467, CelPress logo
 - Article Title:** Neuronal Transcriptome of *Aplysia*: Neuronal Compartments and Circuitry
 - Authors:** Leonid L. Moroz^{1,2,9}, John R. Edwards^{3,9}, Sathyanarayanan V. Puthanveettil^{4,9}, Andrea B. Kohn^{1,8}, Thomas Ha^{1,2}, Andreas Heyland¹, Bjarne Knudsen¹, Anuj Sahn¹, Fahong Yu^{1,5}, Li Liu^{1,5}, Sami Jezzini^{1,2}, Peter Lovell¹, William Iannuccilli³, Minchen Chen³, Tuan Nguyen³, Huitao Sheng³, Regina Shaw^{1,3}, Sergey Kalachikov³ ... Eric R. Kandel^{4,6,8,9}
 - Section:** Resource
 - Abstract:** Identification of an estrogen receptor gene in t... Gene, Volume 540, Issue 1, 2014, pp. 26-31
 - Downloads:** Download PDF, View details
 - Related Articles:** Isolation and functional characterization of the ... International Journal of Biological Macromolecules, Vo...
 - Downloads:** Download PDF, View details
 - Comparative brain structure and visual process...** Current Biology, Volume 32, Issue 1, 2022, pp. 97-110.e4
 - Article Metrics:** Citations (252), Citation Indexes (252), Captures (252)
 - Feedback:** FEEDBACK
- Bottom Bar:** Cloudy, 38°C, 6:01 PM, 6/24/2022

Figure 3.3: PubMed journal of the novel gene

4.0 Multiple sequence alignment

Multiple sequence alignment is the alignment of three or more sequences to study the sequences' homology and evolutionary relationship. In the group project, we used MAFFT to carry out the multiple sequence alignment. However, there are still other approaches to carry out the multiple sequence alignment, such as T-COFFEE and MUSCLE. MUSCLE is an approach that are fastest, and it is most useful in aligning large number of sequences. On the other hand, T-COFFEE might be slower, but it is able to generate a more accurate result in many applications.

MUSCLE stands for Multiple Sequence Comparison by Log-Expectation. Its algorithm consists of three stages including the draft progressive, improved progressive and also the refinement. In the first stage, MUSCLE algorithm will calculate the triangular distance matrix from the pair-wise similarities which is measured using either k-mer counting or the fractional identity through global alignment. From the distance matrix, MUSCLE will construct a tree using UPGMA or neighbor-joining. Based on the tree, a progressive alignment which follows the branching order is built. In the next stage, an enhanced tree is obtained and a new set of progressive alignment is built based on this tree by computing the Kimura distance matrix. The tree will then be compared to the tree in the first stage to identify the subtrees that have changed the branching order and produced a more accurate alignment. The third stage will be the refinement. MUSCLE will choose an edge to be deleted from the tree and divide the tree into two subtrees. The profile of the two subtrees is then extracted and realign to each other to produce a new alignment.

On the other hand, T-COFFEE which stands for Tree based Consistency Objective Function for Alignment Evaluation is a consistency-based methods which utilizes the Feng-Doolittle approach to produce progressive alignments and generate a tree using the pairwise alignment. T-COFFEE will first generate a library which consists of all possible pairwise global alignments of the input sequence using the Needleman-Wunsch algorithm and 10 top-scoring local alignments. An extended library is then generated by recalculating the weights of the aligned residue. By progressive alignment, generating a distance matrix, computing the neighbor-joining tree and using dynamic programming and substitution matrix which derived form the extended library, the multiple sequences ae aligned.

By utilizing T-COFFEE and MUSCLE, I had generated a multiple sequence alignment by using the original query protein, the novel protein and 8 other members in this family. Thus, there are a total of 10 sequences that I have selected to perform the multiple sequence alignment. The sequences are as shown as below:

```
>D(1B) dopamine receptor [Homo sapiens]
MLPPGSNGTAYPGQFALYQQLAQGNAVGGSAGAPPLGPSQVVTACLLTLIIWTLGNVLVCAAIVRSRHL
RANMTN VFIVSLAVSDLFALLVMPWKAVAEVAGYWPFGAFCDVVVAFDIMCSTASILNLCVISVDRYWAI
SRPFYKRKMTQRMALVMVGLAWTLSLISFIPVQLNWHRDQAASWGGLDLPNNLANWTPWEEDFWEPDVN
AENCDSLLNRTYAISSSLISFYIPVAIMIVTYTRIYRIAQVQIRRISSLEAAHQSCRSSAACAPDTSL
RASIKKETKVLKTL SVIMGVFVCCWLFFFILNCMVPFCSGHPEGPPAGFPCVSETTFDVFVWFGWANSSLN
PVIYAFNADFQKVFAQLLGCSHFCRTPVETVNISNELISYNQDIVFHKEIAAAYIHMMNAVTGNREVD
NDEEEGPFD RMFQIYQTSPDGDPVAESVWELDCEGEISLDKITPFTPNGFH

> Aplysia californica protein (sequence taken from BLAST result)
GNIFVITAILVEKSLQ-
GVSNYLILPLAVTDLLAVLVMP SLIYEISIHWFLGNAVCDMWVSMDVL CCTASILHLVIAFDRYWA
N-IDYVRSRNARQILLMVAIVWTVSVFISM PPL-----
FGWRHD SDDPELTGQ-CLISQDHGYTVFSTVGAFYCPLLMVINF KIYRAARYRIRK
>XP_035826857.1 5-hydroxytryptamine receptor 5B-like [Aplysia
californica]
MCWHNISVQRMGRNASLLEANMYAMFPGYNITTGWPPRDNNNSNSGYFDQVTVNGSDRPELLRQYNYHEPK
YEEPVMIVICIFLGCMILAIILGNIFVITAILVEKSLQGVSNYLILSLAVTDLLAVLVMP SLIYEISIHW
FLGNAVCDMWVSMDVL CCTASILHLVIAFDRYWA SNIDYVRSRNARQILLMVAIVWTVSVFISIPPLF
GWRHD SDDPELTGQCLISQDHGYTVFSTVGAFYCPLLMVINF KIYRAARYRIRK RFGGRGGKHQALHV
PLPAVTVETSHRTNTRNSSGSDVSQDG ISMYMP SCTNANDVTRVDMESPLDVQENGQALLQQDPNLTRMLS
NTLTV PACTVS VPLPRPGTSSASNNNVTKRLRRTDKERFRREKMEMRRERKAARV LGIITGA FVVCWL PFF
VVAVVKPMCGTPCDMPSYV YSLFLW LGYVNS LINPIIYTIFNPSFRCAF NKFILRRIKS VNRIT
>GFO38024.1 5-hydroxytryptamine receptor 1a [Plakobranchus ocellatus]
MDILDATVEPGSVSSSYDV VSASP VPLSV TLKSSLMSNTPTLIGNWTTEILGNL SGNHNN TDDEKDELVIYS
YHHPRYGE MTMILICILLGGMILAIILGNIFVITAILVEKTLQGVSNYLILSLAMTDLLAVLVMP SLIYEISIHW
EISTNWFLGT VLCDMWISMDVL CCTASILHLVIAFDRYWA SNIDYI RSRNARQILFMVA VVWMV SVCIS
IPPLFGWRHDTDGPEVTGQCLISQDHGYT IFSTVGAFYCPLFLMLI INF KIYRAARYRIRK RFGRNCHH
NGLQAIHPISV VNADASQRTNTRHSSG DVSQDG VSMY MP SCTNANDVTRMDTESPLD LAETND ALLSQD
PGIQRMLSNTLTV PACTV NVTLARPGPGMSSNNN VPKRQRRGV DKERYR REKMEM RRERKAARV LGIITG
AFVVCWL PFFIIAVLMPMC GDHCYI PGYVLSLFLW LGYF NS LINPIIYTIFNPSFRV AFRKIFFRRIKTSK
RLR
>GFR87408.1 5-hydroxytryptamine receptor [Elysia marginata]
```

MQTTDGAIQTHIMSTPAHSLAGVSPPLSVSLTTSKEARYDQATFALGNVTLNFLGNLSDSNTTDKDKDEM
VIYPYHHPKYGEVTMIFICILLGGMILAIILGNIFVITAIIVEKSLQGVSNYLVLSLAMTDLLAVLVMP
SLINEISTNWFLGNVLCDMWISMDVLCCASILHLVAIAFDRYWAVSNIDYIERSRNARQILFMVAIVWMVS
VSISIPPLFGWRDESDGPEFHGQCLISQDHGYTIFSTVGAFYCPLLLMLIINFKIYRAARYRIRRKRFGNR
NGHGQHNGMQAVHVPISVVNADASQRTNMRHSSGSDVSQDGVS MYMPSCTNANDVTRMDTESPLEAETTD
ALLSQDPSLQRMLNNLTVPACTVNAMARPGSGGVLSNNNPKRAKRGADKERYRREKMEMRRERKAARV
LGIITGAFVVCWLPPFILA VLLPMCGERCVIPDYVLSLFLWLGYFNSLINPIIYTIFNPSFRVAFRKIFFR
RIRSKRLR

>KAH9488655.1 5-hydroxytryptamine receptor [Bulinus truncatus]
MFSNSTTEDSLQPMKTSASLLYQVNISDSLHVTKSLMDNVTSVLSTIQFNNTGYTTISSHTHHIPKYDE
PVMVLICILLGGMILAIILGNIFVITAILVERS LHGVSNYLILSLATT DLLAVLVMPVSLINEVSVKWFL
GPALCDMWVSMDVLCCASILHLVAIAFDRYWAVSNIDYIERSRSAKQILFMVAVVWMVSICISIPPLFGWR
HSTDGPDISGQCLISQDYGYTIFSTLGAFYCPLLLMLIINFKIYRAARYRIRRKFLGRTTFHVPLPIVT
METSQRLNNRNSSGSDVSQDGLSTNANDITRLDMESPLDEADNGQSGSEYDANLTRVFHNTLTV PANTVNL
VVTRSTTPCENNMMIQYNITIYKLRMVNVNTDINL

>XP_025095249.1 5-hydroxytryptamine receptor-like [Pomacea
canaliculata]
MRNTTDSTVGWMVGDLTTWPTPASALYISIPNVSHNYYETDSTEVGSLAARYDSHIPRFPLTTLVSICV
VLGAMIVATVLGNVFVITAIVVERSLQGVSNYLILSLAVTDLLAVLVMPPLSLLNEVSVNWYLGRLCDMW
VSMDVLCCASILHLVAIAFDRFWAVSNIDYVRRRCARQILLMIALVWFVSIAISIPPLFGWRGEMDNPEL
SGQCMISQDHGYTIFSTVGAFYCPLVLMVLNFKIYRAARYRIRRKGFVAVGGGGPGGGGGGDGGGGGGGG
AGRNSRRHHRGKVQVPMVYVEEITAQRTQRNSSGSDVSQDGYSLFNASCAHNEISRMDTGMDSGQDWNQSG
YPGDSTDGSNSNPENSYYPPTEEEVACPDPPLPNSFSSNHLRVDTIIVPSRSILLECRPPTVNNHLSLEC
RPFLGDGRPCSDGGFGGDHREVSAQDVSLVGFERDCNNSQDVGSVLLDVTNGSRRECSGSFANNGSAN
TMAGHDQDQGCVYNHEIPMTALDLPCIAIRKNSLDEELVGLKPRCVVLSKGGGGGSGNQLTVPRGSGMP
PTLGTQPNKS RANNNTKRAHNREKERHREKLEMRRERKAARVLGIITGAFVVCWLPPFILALLSPFCTSA
CSIPLEVYSLFLWLGYVNSLLNPIIYTIFNPSFRCAF RKIFFRRLRSIGR

>QEZ90769.1 5-hydroxytryptamine receptor 1A-like protein [Haliotis
discus hannai]
MENSTEELPYNVAMYIFDNVTAGLDDNL TANGTDNTVERRYTNHVPRYNEGIMIFICLILGSMILATILG
NVFVLTAILIEKSLQGVSNYLILSLGVTDLMVAVLVMPPLS LVNEISINWYLGGEVCDMWISMDVLCCASI
LHLVAISFDRYWAVSNIDYIIRRRCARQIMLMVVIVWFVSVSISIPPLFGWKDNVDPDVTGQCMISQDQGYT
IFSTVGAFYCPLTLMVLNFKIYKAARS RIRKKFGGRPRPAPPVVQVENTGNPSRNSSGSDVSQDGFSM
FNGSCVNMNEISRIESTIEGPEEAQNGNGLQTNNGLGTNGNLPPNFGPSLTVPGNVYVLSKNRSNNNSKR
ARNREKDKLRKEK IEMRRERKAARVLGIITGAFVVCWLPPF MIALIAFPCKESCTFPPIMMSFFLWLGYFN
SLLNPIIYTIFNPSFRTAFRKIFFRKLRVLVQR

>XP_041366623.1 5-hydroxytryptamine receptor 1D-like [Gigantopelta
aegis]

MPTKKTAMENFTISTITVVSLKTLDTNSTSINVSTYVLANDSADVNVGTPYDDVVTRYMMHVPYDVGI
 MTVICALLGAMILATILGNVFVITAIVIEKSLQGVSNYLVLSLAVTDLLAVLVMPISLINDSVHWFYLQ
 EVCDFWVSMVDVLCCASILHLVAISLDRFWAVSNIDYIRRCAKQIIIMVVVWFVAISISIIPPLFGWKDE
 GNPDYTGMCNCLISQDGTYTIFSTVGAFYFPLIFMLVNYKIIYIAARSIRKKNFGKPRPVPMVTVENGT
 AGHASRNSSGSDVSQDGFSMYNGSCVNNECSKFESTYEPSQMDDGPNDGAATTTTSFNGMMHVTSCGS
 NLGSTLSPNFRPSLTVPNNGNVYVVSKNRSNNNSKGARNREKDKLRKEIEMRERKAARVLGIITGAFLV
 CWLPFFLNALILPFCVSCTSHREVVSSFLWLGYFNSLLNPIIYTIFNPSFRIAFRKLFFRRLRHLKR
 >XP_022308516.1 5-hydroxytryptamine receptor-like [Crassostrea
 virginica]

MSTEQQNAGNNATGKMDSVISINLGGRMLGEALDHSNKSQIYSSLFGTRVNGTIipeIMTFDPNNESITG
 FQYEPRYDLELMVFICVLIGCVIVSIVIGNVFVISAIILEKSLQGVSNYLILSLAVTDLMAVLVMPISLI
 DEISLHWYLGPACDMWTSFDILCCTASILHLVAISLDRYWAVSNIDYIRRCKQILMMIAVVVVVAIAI
 SIAPLFGWKDTNDHPEQTGVCKISQEKAYTIFATVGSFYCPLILMVLNFKIYRAARSIRKKHINGMHTP
 IRAYKRPSITAADATTRHNSSGSDISQDGTYVYNGVNQHEEPEPVFNDEQDDATEVNTRFLVVPSNVYVL
 NKQLNPPKNSNNNSKPIAYCDKERLRKAKIEMKRERKAARVLGIITGAFIFCWLPFFLLALISPFCPSCD
 FHPAMFSVFLWLGYVNSLLNPIIYTIFNPSFRNAFRKLIFKRHRRRGR

The result of multiple sequence alignment using T-COFFEE is as shown as below.

CLUSTAL W (1.83) multiple sequence alignment

Homo_sapiens	MLPPGSNGT-AY-PGQFA--L-----
	-----YQQL-----A
Novel_aplysia_californica_protein	-----

Aplysia_californica	MCWHNISVQ-RM---
GRNASLLEANMYAMFPGYNITTGWPPRDNNNSNSGYFDQV	-----TVNGSD-R
Plakobranchus	MDILDATVE-PG-SVSSSY-DVV-
SASPVPVLSVTLKS--SLMSNT--PTLIGNWTT--EILGNL--S--GHNNTDDE	
Elysia_marginata	MQTTDGAIQ-TH-IMSTPAHSLA-
GVSPLPLSVSLTTSKEARYDQA--TFALGNVTL--NFLGNL--S--DSNTTDKD	
Bulinus_truncatus	MFSNSTED-SIQPMKTSASLLY-Q---VNISDSL-
VHVT--KSLMDNVT---SVLSTI----QFNNTG-Y	
Pomacea_canaliculata	MRNTTDSTV-GW-MVGD-----
LTT---TWPTP--AS-ALYISI---P---NVSHNSSYETDSTE-V	
Haliotis_discus_hannai	MENSTEELPYNV-AMYI-----
FDNNTV---A---GLDD--NLTANGTD--	
Gigantopelta_aegis	MPTKKTAME-NF-TISTI--TVV-S-----
LKT---LDTNS--TS-LINVST---YVLANDSA--DVFVNGTP-Y	
Crassostrea_virginica	MSTEQQNAG-NN-ATGKMD-SVI-S---
INLGGRMLGE--ALDHSN--KSQIIYSSLFGTRVNGTIIP---EIMTFD-P	

Homo_sapiens	QGNAVGGSAGAPPLGPSQVVTAC-
LLTLLIWTLLGNVLVCAAIVRSRHLRANMTNVFIVSLAVSDLFVALLVMPWKAVA	
Novel_aplysia_californica_protein	-----
GNIFVITAILVEKSLQG-VSNYLILPLAVTDLLAVLVMPSLIY	
Aplysia_californica	
PELLRQYNYHEPKYEEPVMIVICIFLGCMILAIILGNIFVITAILVEKSLQG-	
VSNYLILSLAVTDLLAVLVMPSLIY	

Plakobranchus
 KDELVYIYHHPRYGEVTMILICILLGGMILAIILGNIFVITAILVEKTLQG-
 VSNYLILSLAMTDLLAVLVMPLSLIN
 Elysia_marginata
 KDEMVIYPYHHPKYGEVTMIFICILLGGMILAIILGNIFVITAIIVEKSLOG-
 VSNYLVLSSLAMTDLLAVLVMPLSLIN
 Bulinus_truncatus
 TTISSHFTFHIPKYDEPVMVLICILLGGMILAIILGNIFVITAILVERSLOG-
 VSNYLILSLATTDLLAVLVMPLSVLIN
 Pomacea_canaliculata
 GSLAARYDSHIPRFLTTLVSICVVLGAMIVATVLGNVFVITAIIVVERSLOG-
 VSNYLILSLAVTDLLAVLVMPLSLLN
 Haliotis_discus_hannai
 NTVERRYTNHVPRYNEGIMIFICLILGSMILATILGNVFVLTAILIEKSLOG-
 VSNYLILSLGVTDLMVAVLVMPLSLVN
 Gigantopelta_aegis
 DDVVTRYMMHVPRYDVGIMTVICALLGAMILATILGNVFVITAIIVIEKSLOG-
 VSNYLVLSSLAVTDLLAVLVMPISLIN
 Crassostrea_virginica
 NNESITGFOYEPRYDLELMFICVLLGIVSIVIGNVFVISAIILEKSLOG-
 VSNYLILSLAVTDLMVAVLVMPISLID

***: * : * : .. * : . : * : : * . : * : * : * : * . : :

Homo_sapiens EVAGYWPFG-
 AFCDVWVAFDIMCSTASILNLCVISVDYWAISRPFRYKRKMTQRMALVMVGLAWTLSILISFIPVQLNW
 Novel_aplysia_californica_protein
 EISIHWFGLNAVCDMWVSMDVLCCASILHLVAIAFDRYWAVSN-
 IDYVRSRNARQILMVAIVWTVSVFISMPL-FG-
 Aplysia_californica
 EISIHWFGLNAVCDMWVSMDVLCCASILHLVAIAFDRYWAVSN-
 IDYVRSRNARQILMVAIVWTVSVFISIPPL-FG-
 Plakobranchus
 EISTNWFLGTVLCDMWISMDVLCCASILHLVAIAFDRYWAVSN-
 IDYIRSRNARQILFMVAVVMVSVSISIPPL-FG-
 Elysia_marginata
 EISTNWFLGNVLCDMWISMDVLCCASILHLVAIAFDRYWAVSN-
 IDYIRSRNARQILFMVAIVWMVSVSISIPPL-FG-
 Bulinus_truncatus
 EVSVKWFGLGPALCDMWVSMDVLCCASILHLVAIAFDRYWAVSN-
 IDYIRRSRAKQILFMVAVVMVSVSISIPPL-FG-
 Pomacea_canaliculata
 EVSVNWYLGRALCDMWVSMDVLCCASILHLVAIAFDRFWAVSN-
 IDYVRRRCARQILMIALVWFVSIAISIPPL-FG-
 Haliotis_discus_hannai
 EISINWYLGGEVCDMWISMDVLCCASILHLVAISFDRYWAVSN-
 IDYIRRRCARQIMLMVVIVWFVSVSISIPPL-FG-
 Gigantopelta_aegis
 DVSVHWYLGQEVCDFWVSMDVLCCASILHLVAISLDRFWAVSN-
 IDYIRRRCAKQIIIMVVVVWFVAISISIPPL-FG-
 Crassostrea_virginica
 EISLHWYLGPAVCDMWTSFDILCCTASILHLVAISLDRYWAVSN-
 IDYIRRRCRKQILMMIAVVVVVAIAISIAPL-FG-
 :::
 * : * . **.* : : * : * . * * : * : * . : * * : : . * : : * : : * :
 * : ..

Homo_sapiens -----
 HRDQAASWGGLDLPNNLANWTPWEEDFWEPDVNAENCDSLNRTYAISSSLISFYIPVAIMIVTYTRIYRI
AQVQIRRIS
Novel_aplysia_californica_protein -----WRHDSDDPELTG-
 QCLISQDHGYTVFSTVGAFYCPLLLMFVINFKIYRAARYRIRK--
Aplysia_californica -----WRHDSDDPELTG-
 QCLISQDHGYTVFSTVGAFYCPLLLMLVINFKIYRAARYRIRK
Plakobranchus -----WRHDTDGPEVTG-
 QCLISQDHGYTIFSTVGAFYCPLFLMLIINFKIYRAARYRIRRK
Elysia_marginata -----WRDESDGPEFHG-
 QCLISQDHGYTIFSTVGAFYCPLLLMLIINFKIYRAARYRIRRK
Bulinus_truncatus -----WRHSTDGPDISG-
 QCLISQDYGYTIFSTLGAFYCPLLLMLIINFKIYRAARYRIRRK
Pomacea_canaliculata -----WRGEMDNPELSG-
 QCMISQDHGYTIFSTVGAFYCPLVLMVLVNFKIYRAARYRIRKG
Haliotis_discus_hannai -----WKDNV-DPDVTG-
 QCMISQDQGYTIFSTVGAFYCPLTMLVLFNKIYKAARSRIRKK
Gigantopelta_aegis -----WKDEG-NPDYTG-
 MCLISQDQGYTIFSTVGAFYFPLIFMLVLNYKIYIAARSRIRKK
Crassostrea_virginica -----WKDTNDHPEQTG-
 VCKISQEKA YTIFATVGSFYCPLILMVLLNFKIYRAARSRIRKKH

*. *: .

Homo_sapiens SLERAA-----
 -----EHAQSCRSSAAC-----
Novel_aplysia_californica_protein -----

Aplysia_californica FGGRGG---K-----H---
 QALHVPLPAVTVET-SHRTNTRNSSGSDVSQDGISMYPSCTNAND
Plakobranchus FGNRNC---H-----
 HNGLQAIHVPISVNVADA-SQRTNTRHSSGSDVSQDGVS MYMPSCTNAND
Elysia_marginata FGNRNG---H-----G---
 QHNGMQAVHVPISVNVADA-SQRTNMRHSSGSDVSQDGVS MYMPSCTNAND
Bulinus_truncatus FLGRTT-----
 KFHVPLPIVTMET-SQRLNNRNSSGSDVSQDGLST-----NAND
Pomacea_canaliculata FVAVGGPGGGGGGDGGGGGGAGRNSRRHRG-
 -KVQVPMVYVEEITAQRTQRNNSGSDVSQDGYSLFNASCA-HNE
Haliotis_discus_hannai FGGR---P-----R-----
 -PAPVPPVQVEN--TGNPSRNNSGSDVSQDGFSMFNGSCVNMMNE
Gigantopelta_aegis FGGK---P-----R-----
 -PVPVPMVTENGTAGHASRNSSGSDVSQDGFSMYNGSCVNMMNE
Crassostrea_virginica INGMHT---P-----IR-
 -AYKRPSITA---ADATTRHNSSGSDISQDGYT VYNGVNQH-EE

Homo_sapiens -----

Novel_aplysia_californica_protein -----

Aplysia_californica VTRVDMES--PLDVQE-----
 ---NGQALLQ-----QDPNLTRMLS-----NTLT
Plakobranchus VTRMDTES--PLDLAE-----
 ---TNDALLS-----QDPGIQRMLS-----NTLT
Elysia_marginata VTRMDTES--PLELAE-----
 ---TTDALLS-----QDPSSLQRMLN-----NTLT

Bulinus_truncatus	ITRLDMES--PLDEAD-----
	-NGQSGSE-----YDANLTRVFH-----NTLT
Pomacea_canaliculata	ISRMDTGMDSGQDWNQSGYPGDSTTDGGSNPENSYYPVPTEEVAC-----
	PDPPLPNSFSSNHLRVDTII
Haliotis_discus_hannai	ISRIESTIEGPEE-AQNGN-----
	-GLQTNNGLLG-----TNGNLPPNFG-----PSLT
Gigantopelta_aegis	CSKFESTYEPS-Q-MDDGP-----
	NDGAATTTTSFNGMMHVTSCGSNLGSTLSPNFR-----PSLT
Crassostrea_virginica	-----PEP-VFNDTE-----
	---QDDA-----TEVNT-----RFLV

Homo_sapiens	-----

Novel_aplysia_californica_protein	-----

Aplysia_californica	VPA-----CT-----VSVPLPRPGTS--
	SASNN-----
Plakobranchus	VPA-----CT-----VNVTILARPGPGG-
	MSSNN-----
Elysia_marginata	VPA-----CT-----VNVAMARPGSGG-
	VLSNN-----
Bulinus_truncatus	VPA-----NT-----VNLVVTRST---
	TPCNN-----
Pomacea_canaliculata	VPSRSLLECRPPTVNNHLSLECRPFLGDGRPCSDGGFGGDHREVSAQDVSLSVGFERDCNNSSQDVGSVL
	LDVTNGSRR
Haliotis_discus_hannai	VPGN-----

Gigantopelta_aegis	VPNNGN-----

Crassostrea_virginica	VPSN-----V-----YVLN-KQLNPP--
	KNSNN-----

Homo_sapiens	-----

Novel_aplysia_californica_protein	-----

Aplysia_californica	-----

Plakobranchus	-----

Elysia_marginata	-----

Bulinus_truncatus	-----

Pomacea_canaliculata	ESCGSFANNGSANTMAGHDQDQGCVYNHEIPMTALDLPCIAIRKNSLDEELVGLKPRCVVLNSKGGGGGGS
	GNQLTVPRG
Haliotis_discus_hannai	-----VY-----
	-----VL-----
Gigantopelta_aegis	-----VY-----
	-----VV-----
Crassostrea_virginica	-----

Homo_sapiens	-----APDT--
SLRASIKKETKVLKTL SVMIMGVFVCCWLPFFILNCMVPFCSGHPEGP	
Novel_aplysia_californica_protein	-----
<hr/>	
Aplysia_californica	-----NVTKRL-
RRTDKERFRREKMEMRRERKAARVLGIITGAFVVVCWLPPFFVAVVKPMCGTPCDM-	
Plakobranchus	-----
NVPKRQRRGVDKERYRREKMEMRRERKAARVLGIITGAFVVVCWLPPFIIAVLMPMCGDHCYI-	
Elysia_marginata	-----
NVPKRAKGADKERYRREKMEMRRERKAARVLGIITGAFVVVCWLPPFILAVLLPMCGERCVI-	
Bulinus_truncatus	-----NM-MIQ-YN-----
<hr/>	
Pomacea_canaliculata	SGMPPTLGTQPNKSRANNNTKRA-
HNREKERHRREKLEMRRERKAARVLGIITGAFVVVCWLPPFILALLSPFCTSACSI-	
Haliotis_discus_hannai	-----SKNRSSNNNSKRA-
RNREKDKLRKEKIEMRRERKAARVLGIITGAFVVVCWLPPFMIALIAPFCKESCTF-	
Gigantopelta_aegis	-----SKNRSSNNNSKGKA-
RNREKDKLRKEKIEMRRERKAARVLGIITGAFLVCWLPPFLNALILPFCVSCTS-	
Crassostrea_virginica	-----NNSKPI-
AYCDKERLRKAKIEMKRERKAARVLGIITGAFIFCWLPFFLLALISPFCP-SCDF-	
<hr/>	
Homo_sapiens	PAGFPCVSETTFDVFVWFGWANSSLNPVIYA-
FNADFQKVFAQLLGCSHFCSRTPVETVNISNELISYNQDIVFHKEIAA	
Novel_aplysia_californica_protein	-----
<hr/>	
Aplysia_californica	-----
PSYVYSLFLWLGYVNLSLINPIIYTIFNPSFRCAFNKIFL-----	
--	
Plakobranchus	-----
PGYVLSLFLWLGYFNLSLINPIIYTIFNPSFRVAFRKIFF-----	
--	
Elysia_marginata	-----
PDYVLSLFLWLGYFNLSLINPIIYTIFNPSFRVAFRKIFF-----	
--	
Bulinus_truncatus	-----ITIYKLRMNVNTDINL-----
<hr/>	
Pomacea_canaliculata	-----
PLEVYSLFLWLGYVNSSLNPIIYTIFNPSFRCAFHKIFF-----	
--	
Haliotis_discus_hannai	-----
PPIMMSFFLWLGYFNSSLNPIIYTIFNPSFRTAFRKIFF-----	
--	
Gigantopelta_aegis	-----
HREVVSFFLWLGYFNSSLNPIIYTIFNPSFRIAFRKLFF-----	
--	
Crassostrea_virginica	-----
HPAMFSVFLWLGYVNSSLNPIIYTIFNPSFRNAFRKLIF-----	
--	
<hr/>	
Homo_sapiens	
AYIHMMMPNAVTPGNREVDNDEEGPFDRMFQIYQTSPDGDPVAESVWELDCEGEISLDKITPFTPNGF-H	
Novel_aplysia_californica_protein	-----
<hr/>	

Aplysia_californica	-----RRIKSVNRIT-----
Plakobranchus	-----RRIKTSKRLR-----
Elysia_marginata	-----RRIRSK-RLR-----
Bulinus_truncatus	-----
Pomacea_canaliculata	-----RRLRSIG--R-----
Haliotis_discus_hannai	-----RKLRLVQ--R-----
Gigantopelta_aegis	-----RRLRHLK--R-----
Crassostrea_virginica	-----KRHRRRRG-R-----

On the other hand, the result of multiple sequence alignment using MUSCLE is as shown as below.

CLUSTAL multiple sequence alignment by MUSCLE (3.8)

Homosapiens	-----MLPPGSNGTAY-----PGQFALYQ--
	-----QLAQGN-----
Crassostrea_virginica	----MSTEQQNAGNNATGKMDSVI-----
	SINLGGGRMLGEALDHNSNKSQIIYSSLFGTR
Gigantopelta_aegis	-----MPTKKTAME-----
	NFTISTITVVSLKTLDTNSTSLINVSTYVL
Haliotis_discus_hannai	-----MENSTEEL-----PYNVAMYI--
	-----FDNV-----
Pomacea_canaliculata	----MRNTTDSTVGWMVGDLTTW-----PTPASALY--
	-----ISIPN-----
Bulinus_truncatus	---MFSNSTTEDSLQPMKTSASLLY-----
	QVNISDSLVLVHTKSLMDNVTSVLSTIQFNN
Plakobranchus_ocellatus	-----MDILDATVEPGSVSSSYDVVSASPVPPLSVTL--
	-KSSLMSNTPTLIGNWTTEI
Elysia_marginata	-----MQTTDGAIQTHIMSTPAHSLAGVS-----
	PLPLSVSLTTSKEARYDQATFALGNVTLN
novel_Aplysia_californica	-----

Aplysia_californica	-----MCWHNISVQRMGRNASLL-----EANMYAMF--
	--PGYNITTGWPPRDNNNSNN

Homosapiens	VGGSGAGA-----PPLGPSQVVTAC-
	LLTLLIIWTLGNVLVCAAIVR
Crassostrea_virginica	
	VNGTIIPEIMTFDPNNESITGFQYEPRYDLELMVFICVLIGCVIVSIVIGNVFVISAIIL
Gigantopelta_aegis	
	ANDSADVNVNGTPYDDVVTRYMMHVPRYDVGIMTVICALLGAMILATILGNVFVITAIVI
Haliotis_discus_hannai	
	VAGLDDNLNTANGTDNTVERRYTNHVPRYNEGIMIFICLILGSMILATILGNVFVLTAILI
Pomacea_canaliculata	
	VSHNSSYETDSTEVGSLAARYDSHIPRFPLTTLVSICVVLGAMIVATVLGNVFVITAIVV

Bulinus_truncatus	TGYTTISSHT-----
FHHIPKYDEPVMVLICILLGGMILAIILGNIFVITAILV	
Plakobranchus_ocellatus	LGNLSGHNNNTDDEKDELV-
IYSYHHPRYGEMTMILICILLGGMILAIILGNIFVITAILV	
Elysia_marginata	LGNLSDSNTTDKDKEDEV-
IYPYHHPKYGEVTMIFICILLGGMILAIILGNIFVITAIIV	
novel_Aplysia_californica	-----
-----GNIFVITAILV	
Aplysia_californica	SGYFDQVTVNNGSDRPELLRQNYHEPKYEEPVMIVICIFLGCMILAIILGNIFVITAILV

***: :* :**:

Homosapiens	
SRHLRANMTNVFIVSLAVSDLFVALLVMPWKAVAEVAGYWPF-G-AFCDVWVAFDIMCSTA	
Crassostrea_virginica	EKSLQG-
VSNYLILSLAVTDLMVAVLVMPISLIDEISLHWYLGPACDMWTSFDILCCTA	
Gigantopelta_aegis	EKSLQG-
VSNYLVLVSLAVTDLVAVLVMPISLINDVSVHWYLQEVCFWVSMVDVLCC	
Haliotis_discus_hannai	EKSLQG-
VSNYLILSLGVTDLMVAVLVMPLSLVNEISINWYLGGEVCDMWISMDVLCC	
Pomacea_canaliculata	ERSLQG-
VSNYLILSLAVTDLVAVLVMPLSLLNEVSVNWLGRALCDMWVSMDVLCC	
Bulinus_truncatus	ERSLHG-
VSNYLILSLATTDLVAVLVMPPSLINEVSVKWFLGPALCDMWVSMDVLCC	
Plakobranchus_ocellatus	EKTLQG-
VSNYLILSLAMTDLVAVLVMPLSLINEISTNWFLGTVLCMWISMDVLCC	
Elysia_marginata	EKSLQG-
VSNYLVLVSLAMTDLVAVLVMPLSLINEISTNWFLGNVLCDMWISMDVLCC	
novel_Aplysia_californica	EKSLQG-
VSNYLILPLAVTDLVAVLVMPLSLIYEISIHWFLGNAVCDMWVSMDVLCC	
Aplysia_californica	EKSLQG-
VSNYLILSLAVTDLVAVLVMPLSLIYEISIHWFLGNAVCDMWVSMDVLCC	

.. * .. ::* ::::*. :**:***:**** . . :::

* :* .**.* ;:*: :*.*

Homosapiens	
SILNLCVISVDRYWAISRPFRYKRKMTQRMALVMGLAWTLSILISFIPVQLNWHRDQAA	
Crassostrea_virginica	SILHLVAISLDRYWAVSN-
IDYIRRRCRKQILMMIAVVVVVAIAISIAPL-----	
Gigantopelta_aegis	SILHLVAISLDRFWAVSN-
IDYIRRRCAKQIIIMVVVVWFVAISISIPIPPL-----	
Haliotis_discus_hannai	SILHLVAISFDRYWAVSN-
IDYIRRRCARQIMLMVIVWFVSVSISIPPL-----	
Pomacea_canaliculata	SILHLVAIAFDRFWAVSN-
IDYVRRRCARQILMIALVWFVSVSISIPPL-----	
Bulinus_truncatus	SILHLVAIAFDRYWAVSN-
IDYIRSRSAKQILFMVAVVWMVSICISIPPL-----	
Plakobranchus_ocellatus	SILHLVAIAFDRYWAVSN-
IDYIRSRNARQILFMVAVVWMVSVCISIPPL-----	
Elysia_marginata	SILHLVAIAFDRYWAVSN-
IDYIRSRNARQILFMVAIVWMVSVSISIPPL-----	
novel_Aplysia_californica	SILHLVAIAFDRYWAVSN-
IDYVRSRNARQILMVAIVWTVSVFISMPP-----	
Aplysia_californica	SILHLVAIAFDRYWAVSN-
IDYVRSRNARQILMVAIVWTVSVFISIPPL-----	

***:*.**.**:*. : *

* . :.*: .*: :: *: *:

Homosapiens
 SWGGLDLPPNNLANWTPWEEDFWEPDVNAENCDSLLNRTYAISSSLISFYIPVAIMIVTYT

Crassostrea_virginica -FG-----WKDTNDHPEQTGV-

CKISQEKEYTIFATVGSFYCPLILMVLLNF

Gigantopelta_aegis -FG-----WKDEGN-PDYTGM-

CLISQDQGYTIFSTVGAFYFPLIFMLVLNY

Haliotis_discus_hannai -FG-----WKDNVD-PDVTGQ-

CMISQDQGYTIFSTVGAFYCPLTLMLVLNF

Pomacea_canaliculata -FG-----WRGEMDNPELSQ-

CMISQDHGYTIFSTVGAFYCPLVLMVLNF

Bulinus_truncatus -FG-----WRHSTDGPDISQ-

CLISQDQGYTIFSTLGAFYCPLLLMLIINF

Plakobranchus_ocellatus -FG-----WRHDTDGPETGQ-

CLISQDHGYTIFSTVGAFYCPLFLMLIINF

Elysia_marginata -FG-----WRDESDGPEFHQ-

CLISQDHGYTIFSTVGAFYCPLLLMLIINF

novel_Aplysia_californica -FG-----WRHDSDDPELTGQ-

CLISQDHGYTVFSTVGAFYCPLLLMFVINF

Aplysia_californica -FG-----WRHDSDDPELTGQ-

CLISQDHGYTVFSTVGAFYCPLLLMLVINF

*: * * . * * : *

*:: ::: :** *: :*..

Homosapiens RIYRIAQVQIRR-----

-----ISSLER

Crassostrea_virginica KIYRAARSRIRKKHI-----

NGMHTPI---RAYKRPSITAAD

Gigantopelta_aegis KIYIAARSRIRKKNF-----

GGKPRPV-----PVPMVTVEN

Haliotis_discus_hannai KIYKAARSRIRKKKF-----

GGRPRPA-----PVPVVQV

Pomacea_canaliculata

KIYRAARYRIRRKGFVAVGPPGGGGGGGGGGGGGGGGGGGGGGAGRNSRRHHRGKVQVPMVYVEE

Bulinus_truncatus KIYRAARYRIRKKRF-----

LGRTTKF-----HVPLPIVTM

Plakobranchus_ocellatus KIYRAARYRIRRKRF-----

GNRNCHH--NGLQAIHVPISVVNA

Elysia_marginata KIYRAARYRIRRKRF-----

GNRNGHGQHNGMQAVHVPISVVNA

novel_Aplysia_californica KIYRAARYRIRK-----

Aplysia_californica KIYRAARYRIRKKRF-----

GRGGKH----QALHVPLPAVTW

.** * . .**.

Homosapiens AAEHAQSCRSSAAC-----

Crassostrea_virginica ATTR---HNSSGSDISQDGTVYNGVNQHEE-----

Gigantopelta_aegis GTAGHAS-

RNSSGSDVSQDGFSMYNGSCVNMNECSKFESTYEPSQMD-----

Haliotis_discus_hannai ENTGNPS-

RNSSGSDVSQDGFSMFNGSCVNMNEISRIESTI-----

Pomacea_canaliculata
NEISRMDTGMDSQDWNQSGYGDSTTDG
Bulinus_truncatus

Plakobranchus_ocellatus

Elysia_marginata

novel_Aplysia_californica

Aplysia_californica

Homosapiens
-----AP
Crassostrea_virginica
-----RFLVVP
Gigantopelta_aegis
ATTTTTSFNGMMHVTSCGSNLGSTLSPNFR-----PSLTVP
Haliothis_discus_hannai
-NLPPNFV-----PSLTVP
Pomacea_canaliculata
-PLPNSFSSNHLRVDTIIVP
Bulinus_truncatus
-NLTRVFH-----NTLTVP
Plakobranchus_ocellatus
-GIQRMLS-----NTLTVP
Elysia_marginata
-SLQRMLN-----NTLTVP
novel_Aplysia_californica

Aplysia_californica
-NLTRMLS-----NTLTVP

Homosapiens

Crassostrea_virginica

Gigantopelta_aegis

Haliothis_discus_hannai

Pomacea_canaliculata
SRSLLCRPPTVNNHLSLECRPFLGDRPCSDGGFGGDHREVSAQDVSLVGFERDC
Bulinus_truncatus

Plakobranchus_ocellatus

Elysia_marginata

novel_Aplysia_californica

Aplysia_californica

ITAQRTQ-RNSSGSDVSQDGYSLFNASCAH-
ETSQRLNNRNSSGSDVSQDGGLS-----TNANDITRLDM-
DASQRTNTRHSSGSDVSQDGVS MYMPSCTNANDVTRMDT-
DASQRTNMRHSSGSDVSQDGVS MYMPSCTNANDVTRMDT-

ETSHRTNTRNSSGSDVSQDGISM YMPSCTNANDVTRVDM-

-----PEPVFND-----TEQDDATEVNT-----
-----DGPNNDGA-----
-----EGPEEAQNGNGLQTNNGLGTNG-----
SGSNPENSYYPV-----PTEEEVACPDP-----
-----ESPLDEA-----DNGQSGSEYDA-----
-----ESPLDLA-----ETNDALLSQDP-----
-----ESPLELA-----ETTDALLSQDP-----

-----ESPLDVQ-----ENGQALLQQDP-----

--DTSL-----
--SNVY-----
NNGNVY-----
--GNVY-----
--
--ANTV-----
--ACTV-----
--ACTV-----

--ACTV-----

Homosapiens	-----
Crassostrea_virginica	-----
Gigantopelta_aegis	-----
Haliotis_discus_hannai	-----
Pomacea_canaliculata	-----
NNSSQDVGSVLLDTNGSRRESCGSFANNGSANTMAGHDQDQGCVYNHEIPMTALDLPCI	
Bulinus_truncatus	-----
Plakobranchus_ocellatus	-----
Elysia_marginata	-----
novel_Aplysia_californica	-----
Aplysia_californica	-----

Homosapiens	-----
Crassostrea_virginica	-----VLNKQLNPPKNSNNNN-----
Gigantopelta_aegis	-----VVSKNR-----SNNN-----
Haliotis_discus_hannai	-----VLSKNR-----SNNN-----
Pomacea_canaliculata	-----
AIRKNSLDEELVGLKPRCVVLNSKGGGGGSGNQLTVPRGSGMPPTLGTQPNKS RANNNT	
Bulinus_truncatus	-----
Plakobranchus_ocellatus	-----NVTLARPGPGGMSSNNNVP-----
Elysia_marginata	-----NVAMARPGSGGVLSNNNVP-----
novel_Aplysia_californica	-----
Aplysia_californica	-----SVPLPRPGTSSASNNNV-----
	T

Homosapiens	-----
RASIKKETKVLKTLCSVIMGVFVCCWLFFFILNCMVPFC SGHPEGPP	
Crassostrea_virginica	-----
KPIAYCDKERLRKAKIEMKRERKAARVLGIITGA FIFCWL PFFLLALISPFC-PSCDFHP	
Gigantopelta_aegis	-----
KGARNREKD KLRKEKIEMRRERKAARVLGIITGAFLVCWL PFFLNALILPFCSVSCTSHR	
Haliotis_discus_hannai	-----
KRARNREKD KLRKEKIEMRRERKAARVLGIITGA FVVCWL PFFMIALIAFPCKESCTFPP	
Pomacea_canaliculata	-----
KRAHNREKERHRREKLEMRRERKAARVLGIITGA FVVCWL PFFILALLSPFCTSACSIPL	
Bulinus_truncatus	-----
-----NLVVTRSTTPCN-----	

```

Plakobranchus_oellatus
RQRRGVDKERYRREKMEMRRERKAARVLGIITGAFVVCWLPPFIIAVLMPMCGDHCYIPG
Elysia_marginata
RAKRGADKERYRREKMEMRRERKAARVLGIITGAFVVCWLPPFILAVLLPMCGERCVIPD
novel_Aplysia_californica      -----
-----  

Aplysia_californica
KRLRRTDKERFRREKMEMRRERKAARVLGIITGAFVVCWLFFFVVAVVKPMCGTPCDMPS

```

```

Homosapiens          AGFPCVSETTFDVFWFGWANSSLNPVIYA-
FNADFQKVFAQLLGCSHFCSRTPVETVNI
Crassostrea_virginica   A-----
MFSVFLWLGYVNSSLNPIIYTIFNPSFRNAFRKLI-----
Gigantopelta_aegis    E-----
VVSFFLWLGYFNSSLNPIIYTIFNPSFRIAFRKLF-----
Haliotis_discus_hannai I-----
MMSFFLWLGYFNSSLNPIIYTIFNPSFRATAFRKIF-----
Pomacea_canaliculata E-----
VYSLFLWLGYVNSSLNPIIYTIFNPSFRCAFRKIF-----
Bulinus_truncatus    -----
NNMMIQYNITIYKLRMNVNTDINL-----
Plakobranchus_oellatus Y-----
VLSLFLWLGYFNSSLNPIIYTIFNPSFRVAFRKIF-----
Elysia_marginata      Y-----
VLSLFLWLGYFNSSLNPIIYTIFNPSFRVAFRKIF-----
novel_Aplysia_californica  -----
-----  

Aplysia_californica      Y-----
VYSLFLWLGYVNSSLNPIIYTIFNPSFRCAFNKIF-----

```

```

Homosapiens
SNELISYNQDIVFHKEIAAYIHMMMPNAVTPGNREVDNDEEEGPFD RMFQIYQTSPDGDP
Crassostrea_virginica
----FKRHRR-RRGR-----
Gigantopelta_aegis
----FRRRLRHLKR-----
Haliotis_discus_hannai
----FRKLRLVQR-----
Pomacea_canaliculata
----FRRRLRSIGR-----
Bulinus_truncatus
-----
Plakobranchus_oellatus
----FRRIKTSKRLR-----
Elysia_marginata
----FRRIRS-KRLR-----
novel_Aplysia_californica
-----
Aplysia_californica
----LRIKSVNRIT-----

```

Homosapiens	VAESVWELDCEGEISLDKITPFTPNGFH
Crassostrea_virginica	-----
Gigantopelta_aegis	-----

Haliothis_discus_hannai	-----
Pomacea_canaliculata	-----
Bulinus_truncatus	-----
Plakobranchus_ocellatus	-----
Elysia_marginata	-----
novel_Aplysia_californica	-----
Aplysia_californica	-----

From the result, we are able to obtain the ten protein sequences with same length. The multiple sequences are also being arranged in the way that as many characters from each sequence are matched. From that, we are able to identify the regions of similarity among the sequences which may result in evolutionary relationship among the sequences. However, there are still slightly different between the multiple sequence alignment result of the two different programs. The result of T-COFFEE contains more gaps than the result of MUSCLE. This might happen when there is more deletion in the result of T-COFFEE. We can also see that there are different number of asterisks(*) in the results and they actually represent the positions which have single, fully conserved residue. Besides, the number of the colon(:) and the period(.) are also different in both results while the colon(:) actually indicates the conservation between groups of strongly similar properties and the period(.) indicates the conservation between the groups of weakly similar properties.

5.0 Phylogenetic tree creation

In this process, I utilized MEGA X to generate a phylogenetic tree by using the same 10 sequences which I used in the last step. To create the phylogenetic tree, firstly, I copy the sequences and build new protein alignment using the MEGA X alignment explorer and export the alignment to MEGA format. There are a few selections for the alignment approaches. In this project, I chose to align the sequences by MUSCLE. Returning to the main page, I open the alignment file that I export earlier and use it to create a phylogenetic tree. In this case, we can choose either we want to generate a Maximum-Likelihood Tree, Neighbor-Joining Tree or Minimum Evolution Tree. We can also set the analysis preference to obtain the desired output. In this case, I created a bootstrapping Neighbor-Joining Tree using the settings as shown in Figure 5.1. Bootstrapping is performed using 500 for the number of bootstrap replications to identify the percent of instances in which bootstrap trees support each clade in the inferred tree. The phylogenetic tree is as shown in Figure 5.2.

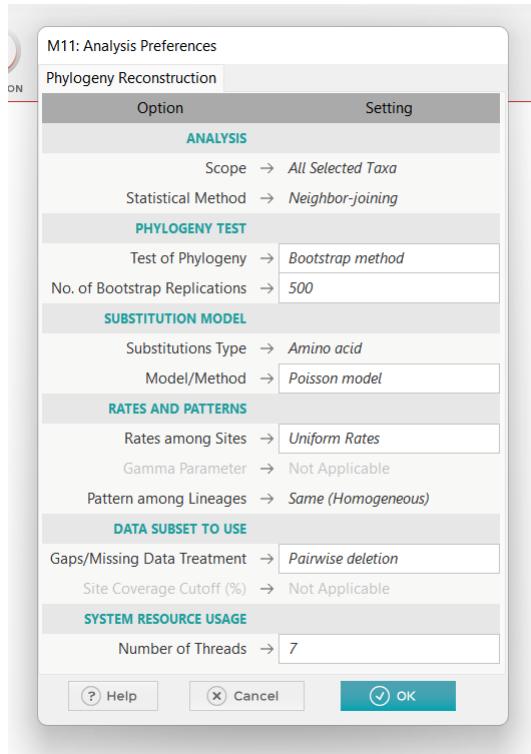


Figure 5.1: Setting of the analysis preferences

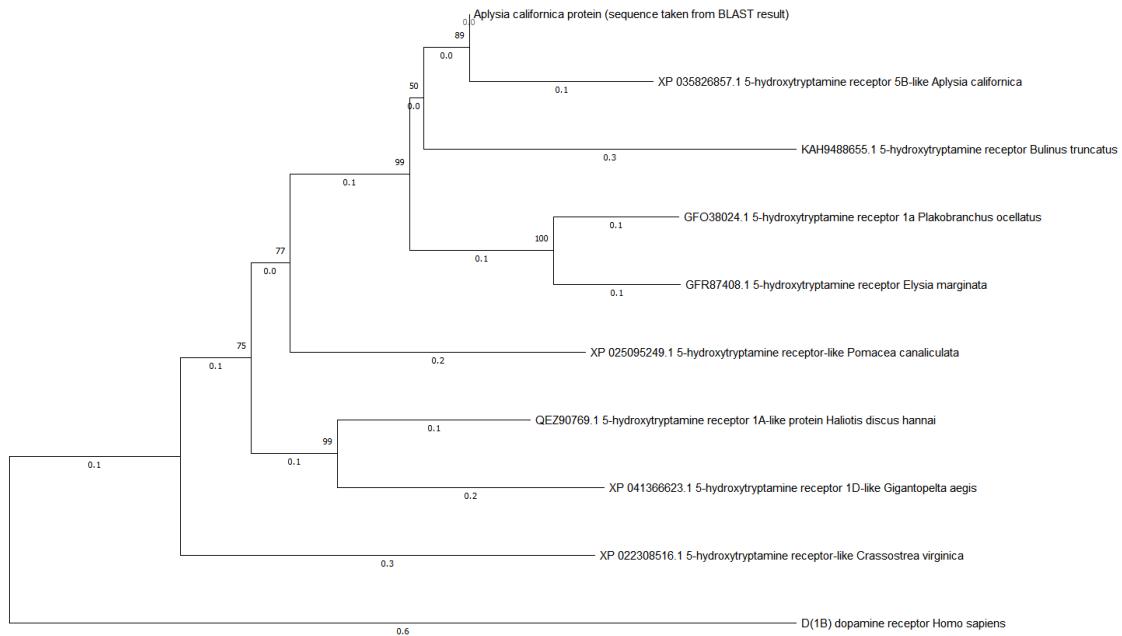


Figure 5.2: Phylogenetic tree

The phylogenetic tree reveals the hypotheses of evolutionary relationship among the group of sequences. From the result, I can actually see that the human D(1B) dopamine receptor is the most distantly related to the novel protein. This can be determined since both

sequence is actually being connected by the deep node and having a less common ancestor. This means that the ancestors of these two sequences might evolve for a very long time. Through analyzing the phylogenetic tree, we can actually understand the evolutionary pathways and the connections among the organisms. From the result also, I can know that the human D(1B) dopamine receptor and the novel protein might not share common traits.

6.0 Conclusion

In our group project, we have learned to find a novel protein by using BLAST and other related programs. To start our project, we have to decide on which known protein that we want to use. Meanwhile, in this alternative assessment, we are required to use a different protein from what we used in the group project. In this case, I have chosen the human D(1B) dopamine receptor since I am quite interested in the particular protein and I wish to know more about it. By using the TBLASTN program against *Aplysia Californica* est database, I am able to get a list of matches which are related to the human D(1B) dopamine receptor. Next, using BLASTP program against nr database for further analysis, I have discovered that the CNSN01-F-142926-501 Normalized CNS library (juvenile 1) *Aplysia californica* cDNA clone CNSN01-F-142926 5', mRNA is a novel protein. After that, I have also done the multiple sequence alignment by using the query protein, novel protein and a few proteins in the family. A phylogenetic tree is also generated to observe the origin and the evolution of the proteins from a common ancestor. Throughout the process, I got a clearer picture on what bioinformatics is in a practical way and I have built my interest towards bioinformatics. However, in this work, there is still some limitation since I only managed to get a few information about the novel proteins through NCBI and PubMed. Nevertheless, I am also limited by my biological knowledge and this has become a difficulty for me to know more clues about the novel protein that I found. Anyway, we all know that sea slug is usually used as a model organism to explore the molecular mechanisms in neuroscience. While *Aplysia Californica* is also one of the species of sea slug, this project might be able to make some contribution to the human neurological and biological field too. As a suggestion for future work, there are still a lot of novel proteins which have not been discovered yet. We as students of bioinformatics and future bioinformaticians should take our responsibility to do more analysis to discover the novel proteins which might be able to contribute in the field of biology.

CHONGKAHWEI_FINAL

by Chong Kah Wei

Submission date: 29-Jun-2022 01:43AM (UTC+0800)

Submission ID: 1864238738

File name: CHONGKAHWEI_FINAL.pdf (1.9M)

Word count: 4645

Character count: 45138



UTM
UNIVERSITI TEKNOLOGI MALAYSIA

SCHOOL OF COMPUTING
Faculty of Engineering

SECB 2103-01

BIOINFORMATICS 1

SEMESTER II 2021/2022

ALTERNATIVE ASSESSMENT

FIND-A-GENE PROJECT

LECTURER: DR HASLINA BINTI HASHIM

SUBMITTED BY:

CHONG KAH WEI

A20EC0027

INDEX

NO.	TITLE	PAGE
1.0	Introduction	1
2.0	BLAST	4
3.0	Gathering information of the “novel” gene	9
4.0	Multiple sequence alignment	12
5.0	Phylogenetic tree creation	26
6.0	Summary and discussion	28

Article Error 

1.0 Introduction

5

In this project, we are assigned to find a novel gene using BLAST searching. Before beginning this alternative assessment, we have been doing the project in a group. We started the project by choosing a protein of interest. In this alternative assessment, I had selected the human D(1B) dopamine receptor protein with the accession number NP_000789.1. After deciding the protein of interest, I used the protein accession number to carry out TBLASTN searching against Expressed Sequence Tag (est) database and *Aplysia californica* (California sea hare) organism to find the matches. The results gave the gene sequences which are possibly homologous to the D(1B) dopamine receptor protein. From the results, I chose one of the gene sequences which is the CNSN01-F-142926-501 Normalized CNS library (juvenile 1) *Aplysia californica* cDNA clone CNSN01-F-142926 5', mRNA sequence to be further analyzed to confirm whether it is a novel gene or not. In this case, I used the sequence that was retrieved from the TBLASTN results to carry out BLASTP searching against the non-redundant (nr) protein database. From the BLASTP result, I observed that there is no protein that is 100% identical to the searched sequence. The top result of the match is 98.30%. Hence, I can conclude that there is a chance that CNSN01-F-142926-501 Normalized CNS library (juvenile 1) *Aplysia californica* cDNA clone CNSN01-F-142926 5', mRNA sequence. After getting the novel protein, I performed the next step which is to gather the information of the novel gene. After that, I generate a multiple sequence alignment using the query protein, the novel protein and a few members of the family using MUSCLE and TCOFFEE. By then, I obtained the alignments with similar length. Using the alignments, I also did a phylogenetic analysis to generate a phylogenetic tree to observe the evolutionary relationship of the protein species.

Basically, we can use any known protein sequence as a starting point to get a novel protein. In this case, I am quite interested in the human dopamine receptor and I found that the D(1B) dopamine receptor protein is quite meaningful. Dopamine receptors are G-protein coupled receptors (GPCRs) that play an important role in a human's daily life. The protein is involved in many neurological processes. There are several types of dopamine receptors too, including D1, D2, D3, D4 and D5 and every receptor has different functions. Meanwhile, there is a human gene called DRD5 and it is used in making the D5 receptors,

which is also known as the D(1B) dopamine receptor protein. The protein is highly homologous to the D1 receptor and it can be considered as a D1-like receptor. Besides, D(1B) dopamine receptor protein is located in the human brain and is used in human decision making, cognition, attention and renin secretion. Thus, I found that the protein is actually quite important for humans and I decided to use the protein as the query protein of the TBLASTN searching, which is the very first step in this project. After deciding the protein of interest, I get the details such as the accession number, species and also the sequences of the protein from the NCBI protein page as shown in Figure 1.1 and Figure 1.2. By then, I am able to get the accession number of the particular protein to do further analysis.

The screenshot shows the NCBI protein page for NP_000789.1, the D(1B) dopamine receptor in Homo sapiens. Key details include:

- Protein ID:** NP_000789.1
- Length:** 477 aa
- Organism:** Homo sapiens
- References:**
 - Bai R, Mel J and Hu W. Hypermethylation of DRD5 Promoter Is a Biomarker Across 12 Cancer Types. *Am J Cell Biol* 39 (11), 2052-2058 (2020).
 - Lee CG, Moon H and Park S. The effects of dopamine receptor genes on the trajectories of sport participation from adolescence through young adulthood. *Am J Hum Biol* 47 (3), 256-262 (2020).
- Articles about the DRD5 gene:**
 - Hypermethylation of <i>DRD5</i> Promoter Is a Biomarker Across 12 Cancers [DNA Cell Biol. 2020]
 - Role of Thiodotolin 1 in Impaired Renal Sodium Excretion of h_D-5 [J Am Heart Assoc. 2019]
 - Activation of DRD5 (dopamine receptor D5) inhibits tumor growth by autophagy [Autophagy. 2017]
- Reference sequence information:** RefSeq genomic sequence (NG_012024.1), RefSeq mRNA.

Figure 1.1: D(1B) dopamine receptor protein

The screenshot shows a web browser window displaying the sequence information for the D(1B) dopamine receptor protein (NP_000789.1). The sequence is presented in a multi-line text format, starting with the header and followed by the amino acid sequence itself. The sequence begins with 'MLPOGSGNTA YPGGFLYQO LAQHNAGGS AGAPOLGPSSQ VVTACLLTLI IWTLLGNVL' and continues through several hundred positions. The browser interface includes a navigation bar at the top, a search bar, and various links and icons. At the bottom, there's a status bar showing the date and time.

```
My Drive Gmail MyUTM Portal Bearing UTM SSLVPN HTML Tutorial Java Tutorial GitHub Sci-Hub Library Genesis YouTube WhatsApp Bio Science tech COVID-19 Quarantine
```

Region
338..362 /region_name="TM helix 7"
/notes="Propagated from UniProtKB/Swiss-Prot (P21918.2)"
Site
341..360 /site_type="transmembrane region"
/notes="Propagated from UniProtKB/Swiss-Prot (P21918.2)"
CDS
3..477 /gene="DRD5"
/gene_synonym="DRD5; DRD1B; DRD1L2"
/coded_by="NM_000788..5..397..1830"
/db_xref="CCDS:CD53445.1"
/db_xref="GeneID:1816"
/db_xref="HGNC:HNGC:1026"
/db_xref="MIM:MIM:126453"
ORIGIN
1 mlpo gsgnta ypggfl yqo laqhnagg s agap olgpssq vvtac lltl iwtllgnvl
61 vcaal vlrshl l rannet nvnfl vslavsdifv allvmpukav aevag yupg fcdvv wifd
121 imcstasilm lcv isvdryw alsrpfr ykr kmtq ralavm vglavt lsl ifipq vlnw
181 hr dqpa sugg l dlpnpl anw t pweedfvev dnva end ss l mtya liss lisfy ipval
241 mityv triy iaqvg lrris sleraehaq scrssa acap dtls raiik etvl kt lsv
301 imgvfvccwl pfffl ncmvp fc shpapp agfp c vsett fdv fwf gnu nsll npviya
361 fnad fqk vfa qllgchfcs rtpvet nni nali syn gdi vfh keiaay ihmp navtp
421 gnrevndee egpfd rwfqj yqts pdgd p v aevv elde geisld kitp ftpng fh
//

You are here: NCBI > Proteins > Protein Database Support Center

FOLLOW NCBI

25°C Rain showers 10:59 AM 6/23/2022 Article Error (ETS)

Figure 1.2: Sequence of D(1B) dopamine receptor protein

2.0 BLAST

1 Basic Local Alignment Search Tool (BLAST) is the main NCBI tool that allows Article Error (ETS) 19 rapid sequence comparison of a query sequence against a database. We can say that BLAST is actually one of the basic ways to learn about a protein or gene since the searching result will reveal all related sequences and calculate the statistical significance of all matches. In this case, NCBI provides a series of BLAST programs, including BLASTP, BLASTN, BLASTX, TBLASTX and also TBLASTN for different use. In this instance, BLASTP is used to compare a protein query to a database of proteins and the searching result will display a list of related protein sequences while BLASTN is utilized in comparing DNA Article Error (ETS) 2 query against a DNA database. On the other hand, BLASTX is carried out to compare a DNA query against a protein database. The program performs translation of a DNA Article Error (ETS) 4 sequence into six protein sequences using all six possible reading frames and then compares each of the proteins to a protein database. TBLASTN is the reverse of BLASTX since it is used when comparing a protein query to a nucleotide database. The program translates the DNA sequence in a database into six potential proteins and then compares the protein query against each of the translated proteins. TBLASTX is the most computationally intensive program in BLAST. This particular program is able to translate Article Error (ETS) 11 a DNA query from both a query and a database into six potential proteins and perform a Article Error (ETS) 36 protein-protein database search. The result of TBLASTX will reveal genes that encode Article Error (ETS) proteins homologous to the query.

The BLAST program that I used in this project is TBLASTN and BLASTP. TBLASTN program is used as a starting point to get a novel protein. After confirming the protein of interest, which is the NP_000789.1 D(1B) dopamine receptor protein, I use the accession number to carry out TBLASTN searching against *Aplysia californica* (California sea hare) est Article Error (ETS) 4 database as shown in Figure 2.1 and the searching result generates a list of related genes. Part of the searching results are shown in Figure 2.2 and Figure 2.3.

The screenshot shows the TBLASTN search interface. In the 'Enter Query Sequence' field, the accession number NP_000789.1 is entered. Below it, there's a note about aligning two or more sequences. Under 'Choose Search Set', the database is set to 'Expressed sequence tags (est)'. The organism is set to 'Aplysia californica (taxid:6500)'. The search parameters include a job title 'NP_000789:D(1B) dopamine receptor [Homo sapiens]' and a descriptive title for the search.

Figure 2.1: Input of TBLASTN

The screenshot shows the TBLASTN results page. The table lists numerous hits from the Aplysia californica database. The columns include Description, Scientific Name, Score, Score, Cover, value, Ident, Len, and Accession. The first few hits are:

Description	Scientific Name	Score	Score	Cover	value	Ident	Len	Accession
CNSN01-F-142926-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	119	119	42%	3e-30	35.47%	667	EB335661.1
CNSN01-F-102105-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	110	110	43%	4e-27	33.01%	742	EB348863.1
CNSN01-C-000413-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	108	108	26%	5e-26	43.75%	792	EB260814.1
CNSN01-F-062370-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	85.5	85.5	24%	2e-18	39.17%	624	EB346158.1
CNSN01-F-063010-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	73.6	73.6	31%	3e-14	28.76%	638	EB346226.1
PE0002-C-011987-301 Normalized Pedal-Pleural Ganglia <i>Aplysia californica</i> cDNA clone PE0002-C-011987 3'-m...	<i>Aplysia californica</i>	70.5	70.5	17%	1e-13	38.82%	514	EB214272.1
CNSN01-F-066686-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	71.2	71.2	43%	2e-13	26.64%	688	EB346577.1
CNSN01-F-048818-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	67.4	67.4	27%	2e-12	30.30%	542	EB297304.1
PEG0003-C-228120-501 Normalized Pedal-Pleural Ganglia <i>Aplysia californica</i> cDNA clone PEG0003-C-228120 5'...	<i>Aplysia californica</i>	68.9	68.9	21%	2e-12	42.45%	771	EB245546.1
CNSN01-F-002002-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	65.9	65.9	18%	2e-12	35.56%	380	EB282651.1
PEG0003-C-217830-501 Normalized Pedal-Pleural Ganglia <i>Aplysia californica</i> cDNA clone PEG0003-C-217830 5'...	<i>Aplysia californica</i>	68.9	68.9	21%	2e-12	42.45%	800	EB241139.1
PE0003-C-205407-501 Normalized Pedal-Pleural Ganglia <i>Aplysia californica</i> cDNA clone PE0003-C-205407 5'...	<i>Aplysia californica</i>	68.9	68.9	21%	2e-12	42.45%	822	EB230890.1
CNSN01-C-004354-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	68.2	68.2	18%	4e-12	34.07%	799	EB280652.1
CNSN01-F-089997-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	67.0	67.0	16%	4e-12	37.35%	617	EB322814.1
CNSN01-C-008362-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	66.2	66.2	18%	6e-12	34.07%	582	EB258151.1
CNSN01-F-030457-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	65.9	65.9	18%	7e-12	34.44%	559	EB342970.1
CNSN01-C-001435-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	65.5	65.5	18%	1e-11	34.44%	569	EB255570.1
G1049P2351RP14_T1 <i>Aplysia californica</i> Pooled Normalized Library <i>Aplysia californica</i> cDNA, mRNA sequence	<i>Aplysia californica</i>	67.0	67.0	18%	1e-11	34.07%	847	EB238844.1
CNSN01-F-012764-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	65.9	65.9	27%	2e-11	31.06%	665	EB230869.1
CNSN01-F-010192-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	65.5	65.5	18%	2e-11	33.33%	618	EB288905.1
CNSN01-F-105096-501 Normalized CNS library (juvenile) 1	<i>Aplysia californica</i>	67.8	67.8	18%	3e-11	33.33%	394	EB376306.1

Figure 2.2: Results of TBLASTN

Article Error

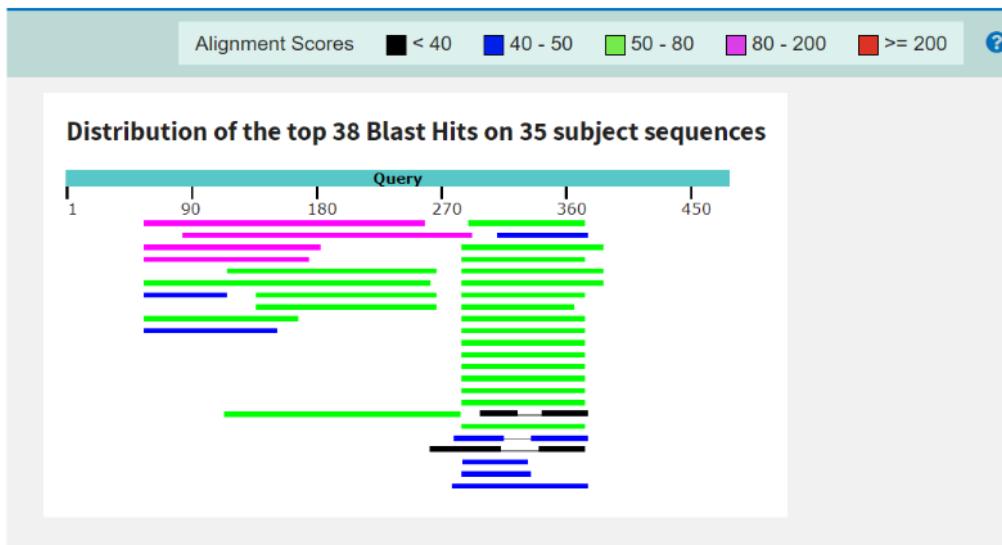


Figure 2.3: Graphic summary of TBLASTN

In this step, I have to distinguish the different results to obtain a sequence which might be a novel gene. In this case, a perfect match to my query sequence would be a sequence that is not novel, a near match would be a gene that might be novel depending on the results of further analysis and the non-homologous results would be not related to the query sequence. To differentiate the results and get a possible novel protein sequence, we need to check the E-value, maximum score, total score, percent query coverage and percent identity. E-value is a number that describes the expected hits of similar score that are found by chance in a database. In this case, the significant alignments should have an E-value that is very close to 0. Aside from that, the maximum score is the highest alignment score between the query sequence and the database sequence while the total score is the sum of alignments scores from the same database sequence that match the query sequence. Nevertheless, the percent query coverage is the percentage of how much the query sequence is covered by the match sequence and it is able to reveal how long the sequences relate to each other. On the other hand, percent identity is the percentage that describes how similar the query sequence is to the target sequence. Basically, the smaller the E-value, the larger the score, the more significant the matches. Besides, the larger percent query coverage and the percent identity, also, the higher the chance to get a better match. From the TBLASTN result, there are actually 35 matches have been generated. In this case, if the match is a perfect match, it should have an E-value which is very close to 0, a high bit

score and also 100% identical to the query sequence. However, the matches of my result can be considered as near matches since their E-value are not too large and the total scores also not too low as well as the percent identity are not 100% and in an acceptable range. On the other hand, a non-homologous sequence is an insignificant match that is not related to the query sequence. These results might get a very high E-value, very low bit score as well as a very low percent identity. Among the matches, I have chosen a match among the near matches which have a chance of becoming a novel protein. The match that I have chosen for further analysis is the top match, which is CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5', mRNA sequence since it is the most significant match. The E-value of this match is 3e-30, the max score is 119 and the percent identity is 35.47%. The alignment of this match can be seen in Figure 2.5.

The screenshot shows a BLAST search results page. The header includes 'Download' and 'GenBank Graphics' buttons, and navigation links for 'Next', 'Previous', and 'Descriptions'. The main title is 'CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5', mRNA sequence'. Below this, 'Sequence ID: EB335661.1 Length: 667 Number of Matches: 1' is displayed. A table titled 'Range 1: 69 to 596' lists the following information:

	Score	Expect	Method	Identities	Positives	Gaps	Frame
119 bits(297)	3e-30	Compositional matrix adjust.	72/203(35%)	112/203(55%)	28/203(13%)	+3	
Query 57	GNVLVCA44EVRSRHLRANHITN/FVSLAVSDLFVALVNPWIKAVAEVAGYWPFG-AFCDV	115					
	GN+ V A++ L+ +N+ Z+ LAV-DL VA+LVNP + E++ HH G A CD+						
Sbjct 69	GNIFVITAILVEKSLQ-GVSNYLILPLAVTDLLVAVLVMPISLIYEISIHWFLGNAVCMD	245					
Query 116	WVAFDIMCSTASILNLVISVDRYWAISRPFRYRKHTQRNALVMVGLAINTLSSLISFP	175					
	WV D+C TASIL+L I+ DRYNA+S Y R R L+M+ + WTS+ IS P						
Sbjct 246	WVSMDVLCCTASILHLVAIAFDRYWAVSN-IDYVRSRNARQILLMVAIVWTVSVFISMPP	422					
Query 176	VQLNIHRDQAASWGGDLPNNLAMWTPWEEDFWEPDN/NAECDSSLNRTYAIISSSLIFY	235					
	W D +P++ + C S + Y + S++ +FY						
Sbjct 423	L-----FGRHDSDDPPELTGQ-CLISQDHGYTVFSTVGAFY	527					
Query 236	IPVAIMIVTYTRIYRIAQVIRR	258					
	P+ WH V +IYR A+ +IR+						
Sbjct 528	CPLLMMFVINPKIYRAARYRINK	596					

Figure 2.5: Alignment of the match

After interpreting the data, I have considered this match as a near match which might be a novel gene and use it to proceed my project. Also, I retrieved the sequence of this protein from the TBLASTN result to perform BLASTP search against non-redundant (nr) database to check whether it is a novel protein. The particular sequence is as shown below:

```
> Aplysia californica protein (sequence taken from BLAST result)
GNIFVITAILVEKSLQ-GVSNYLILPLAVTDLLVAVLVMPISLIYEISIHWFLGNAVCMD
WVSMDVLCCTASILHLVAIAFDRYWAVSN-IDYVRSRNARQILLMVAIVWTVSVFISMPP
L-----FGRHDSDDPPELTGQ-CLISQDHGYTVFSTVGAFY
CPLLMMFVINPKIYRAARYRINK
```

Then, what I have done is checking the percent identity of the results. If there is a match with 100% percent identity to a protein in the database from the same species, then the protein is not novel. In contrast, if the match is 100% but it is to a different species than the one that I started with, it is a novel protein. On the other hand, if there is no match that is 100% identical to the searched sequence, then it is likely the protein is novel. However, if there are no database matches to the original query, it means that the sequence is not actually homologous to the original query. For my result as shown in Figure 2.6, I can see that the top result is 98.30% identical to the *Aplysia californica* protein that I have taken from the previous BLAST result. I can also get the match to the original query, which is the human D(1B) dopamine receptor from the result if I restrict the database to human nr as shown in Figure 2.7. Thus, I can conclude that the *Aplysia californica* protein that I have obtained is a novel protein which is homologous to my query protein.

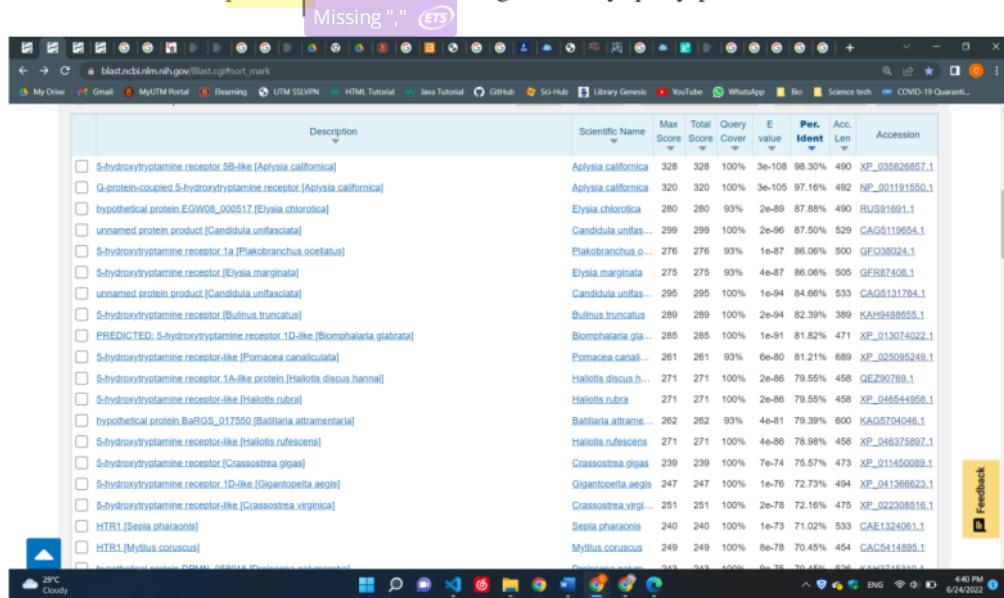


Figure 2.6: BLASTP result of *Aplysia californica* protein

<input checked="" type="checkbox"/> Crystal structure of 5-HT2AR in complex with risperidone [Homo sapiens]	<i>Homo sapiens</i>	97.8	97.8	93%	3e-23	39.52%	376	6A93_A
<input checked="" type="checkbox"/> D(1B) dopamine receptor [Homo sapiens]	<i>Homo sapiens</i>	98.6	98.6	100%	3e-23	35.47%	477	NP_000789.1
<input checked="" type="checkbox"/> Chain A: 5-hydroxytryptamine receptor 2A;5-hydroxytryptamine receptor 2A;5-hydroxytryptamine receptor 2A-Solu [Homo sapiens]	<i>Homo sapiens</i>	97.8	97.8	93%	3e-23	39.52%	376	7WIC4_A

Figure 2.7: Match to original query

3.0 Gathering the information of the “novel” gene

The gathering of information about the “novel” gene is a significant process since all of us should know the details of the gene that we found and used in the project. Through gathering the information of the “novel” gene, we can find out more important information about the sequence that might be undiscovered before. *Aplysia Californica* is actually a type of invertebrate organism and we can say that it is totally different with human. However, scientist has found that there are genes which involved in similar neural activities in the brain of both organisms, and this finding has boosted the further investigation of neurological disease such as Parkinson’s and Alzheimer’s disease. In this case, the novel protein is actually a match of human D(1B) dopamine receptor, which is also one of the gene that exist in human brain. By gathering the information of the novel gene, we might discover some useful pieces that might be able to contribute to the biological field.

Throughout the process to gather the information of this novel gene, I was interested in investigating whether the gene has same characteristics with human D(1B) dopamine receptor. However, when I was doing research about this gene, there was only a limited resource and information about this novel gene. To know more about the novel protein, I got the details from the NCBI GenBank as shown in Figure 3.1 and Figure 3.2. The basic information, including the accession number, source, sequence and so on can actually be found in this page. Besides, I also referred to the PubMed journal as shown in Figure 3.3 through the link that is provided in the NCBI page to know more about this novel gene.

CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5', mRNA sequence

GenBank EB335661.1
FASTA Graphics
Go to: ▾

LOCUS EB335661 528 bp **mRNA** linear **EST** 07-JAN-2007
DEFINITION CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica cDNA clone CNSN01-F-142926 5', mRNA sequence.
ACCESSION EB335661 REGION: 69..596
VERSION EB335661.1
DBLINK BioSample: [SAMN00150086](#)
KEYWORDS EST.
SOURCE Aplysia californica (California sea hare)
ORGANISM [Aplysia californica](#)
Eukaryota; Metazoa; Spiralia; Lophotrochozoa; Mollusca; Gastropoda; Heterobranchia; Euthyneura; Tectipleura; Aplysiidae; Aplysiidae; Aplysia.
REFERENCE 1 (bases 1 to 528)
AUTHORS Moroz,L.L., Edwards,J.R., Putmanveetil,S.V., Kohn,A.B., Ma,T., Heyland,A., Knudsen,B., Sahn,A., Yu,F., Liu,L., Jezzini,S., Lovell,P., Iannuccilli,W., Chen,M., Nguyen,T., Sheng,H., Shaw,R., Kalachikov,S., Panchin,Y., Farmerie,W., Russo,J.J., Ju,J., and Kandul,E.R.
TITLE Neuronal Transcriptome of Aplysia: Neuronal Compartments and Circuitry
JOURNAL Cell 127 (7), 1453-1467 (2006)
PUBLISHED 17398687
COMMENT Contact: Moroz LL
Department of Neuroscience & The Whitney Laboratory for Marine Bioscience

Related Information
BioSample
PubMed
Taxonomy
Full text in PMC

Recent activity
Turn Off Clear

CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica Nucleotide

Figure 3.1: Information of Aplysia California

CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica Nucleotide

COMMENT Contact: Moroz LL
Department of Neuroscience & The Whitney Laboratory for Marine Bioscience
University of Florida
9505 Ocean Shore Blvd., St. Augustine, FL 32080, USA
Tel: 904 461 4020
Fax: 904 461 4052
Email: moroz@whitney.uf.edu.

FEATURES
source Location/Qualifiers
1..528
/organism="Aplysia californica"
/mol_type="mRNA"
/db_xref="taxon:6500"
/clone="CNSN01-F-142926"
/tissue_type="Central nervous system (all ganglia)"
/clone_lib="SAMN00150086 Normalized CNS library (juvenile 1)"
/dev_stage="juvenile"
/note="Vector: pDEM1; Site_1: 5'- TAT CAA COC AGC CGA CCT CTC; Site_2: 5'- CGC AGT CGG TAC (T)13 -3' (C; mT
3' seq)CHSN01"

ORIGIN

```

1 gttataatattttgtatataac tggccatctcg tggggaaatcttcccccgg tttgtccccac
61 taaccttcatttc tgccgttgtgc aatgtactgt ctccctgtgg ctgtgtttgtt catgccttc
123 agtcttcatttc acgatgtatcg catccatgg ttttttagcca adgcgcgttg cgatgtgtgg
183 gtgtctatgtt aatgtttttcg ctgtacatgttccatattttac atctctgtggc tataggcggtt
241 gtatgtatgtt gggccgttttc aaacatgtatcg taatgttggaa gtccggaaatgc aaggccgttt
303 cttcgtgtttt ggccgttttcatgttggaa gtggccgttc taatgtttta tatcgatgtcc gctttgttcc
361 ggttgtggaaatc aatgtttttcg ccgtttttttccatgttggaa gtccggaaatgc aaggccgttt
421 caatgtttttttccatgttggaa gtggccgttc ccgtttttttccatgttggaa gtccggaaatgc aaggccgttt
481 gtgtatcaatc ttatgtatatac tggccatctcg cggtagccgg cggtagccgg taatgtttttttccatgttggaa
//
```

Related Information
CNSN01-F-142926-501 Normalized CNS library (juvenile 1) Aplysia californica Nucleotide
nbc52b01.y1 Rabbit trigeminal nerve. Unnormalized (nbc) Oryctolagus cur Nucleotide
nag32c04.y1 Rabbit eye minus lens and cornea. Unnormalized (mag) Oryctolagus cur Nucleotide
5-hydroxytryptamine receptor 5B-like [Aplysia californica] Protein
D(1B) dopamine receptor [Mus musculus] Protein

See more...

Figure 3.2: Information of Aplysia California

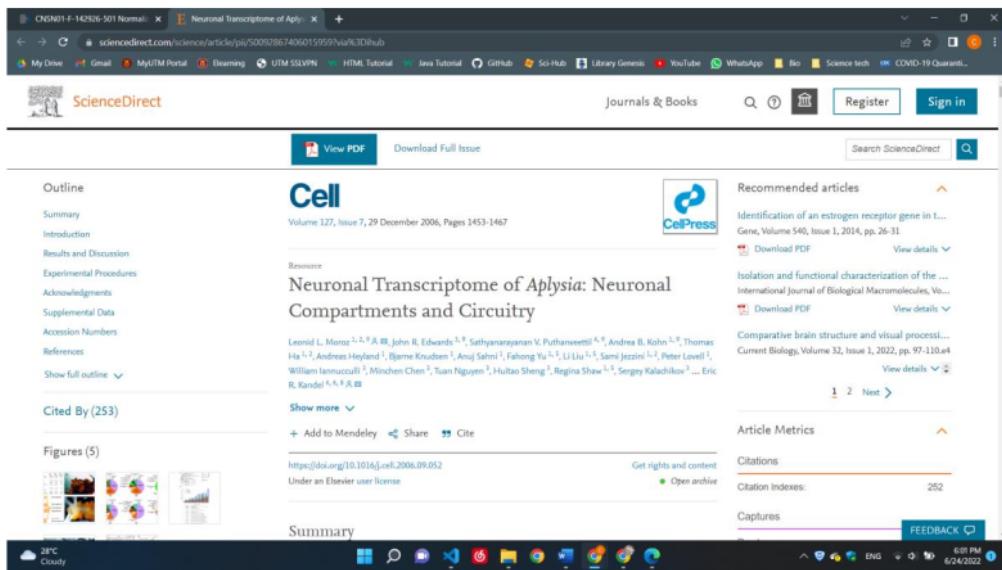


Figure 3.3: PubMed journal of the novel gene

Article Error 

20

4.0 Multiple sequence alignment

Multiple sequence alignment is the alignment of three or more sequences to study the sequences' homology and evolutionary relationship. In the group project, we used MAFFT to carry out the multiple sequence alignment. However, there are still other approaches to carry out the multiple sequence alignment, such as T-COFFEE and MUSCLE. MUSCLE is an approach that are fastest, and it is most useful in aligning large number of sequences. On the other hand, T-COFFEE might be slower, but it is able to generate a more accurate result in many applications.

MUSCLE stands for Multiple Sequence Comparison by Log-Expectation. Its algorithm consists of three stages including the draft progressive, improved progressive and also the refinement. In the first stage, MUSCLE algorithm will calculate the triangular distance matrix from the pair-wise similarities which is measured using either k-mer counting or the fractional identity through global alignment. From the distance matrix, MUSCLE will construct a tree using UPGMA or neighbor-joining. Based on the tree, a progressive alignment which follows the branching order is built. In the next stage, an enhanced tree is obtained and a new set of progressive alignment is built based on this tree by computing the Kimura distance matrix. The tree will then be compared to the tree in the first stage to identify the subtrees that have changed the branching order and produced a more accurate alignment. The third stage will be the refinement. MUSCLE will choose an edge to be deleted from the tree and divide the tree into two subtrees. The profile of the two subtrees is then extracted and realign to each other to produce a new alignment.

On the other hand, T-COFFEE which stands for Tree-based Consistency Objective Function for Alignment Evaluation is a consistency-based methods which utilizes the Feng-Doolittle approach to produce progressive alignments and generate a tree using the pairwise alignment. T-COFFEE will first generate a library which consists of all possible pairwise global alignments of the input sequence using the Needleman-Wunsch algorithm and 10 top-scoring local alignments. An extended library is then generated by recalculating the weights of the aligned residue. By progressive alignment, generating a distance matrix, computing the neighbor-joining tree and using dynamic programming and substitution matrix which derived form the extended library, the multiple sequences are aligned.

Possessive (ETS)

By utilizing T-COFFEE and MUSCLE, I had generated a multiple sequence alignment by using the original query protein, the novel protein and 8 other members in this family. Thus, there are a total of 10 sequences that I have selected to perform the multiple sequence alignment. The sequences are as shown as below:

```
>D(1B) dopamine receptor [Homo sapiens]
MLPPGSNGTAYPGQFALYQQLAQGNAVGGSAGAPPLGPSQVVTACLLLIWTLGNVLVCAAIVRSRHL
RANMTNVFIVSLAVSDLFALLVMPWKAVAEVAGYWPFGAFCDVVAFDIMCSTASILNLCVISVDRYWAI
SRPFYKRKMTQRMALVMVGLAWTLSILISFIPVQLNWHRDQAASWGGLDPNNLANWTPWEEDFWEPDVN
AENCDSSLNRTYAISSSLISFYIPVAIMIVTYTRIYRIAQVQIRRISSLERAEEHAQSCRSSAACAPDTSL
RASIKKETKVLKTLCSVIMGVFVCCWLFFFILNCMVPFCSGHPEGPPAGFPCVSETTFDVFWFGWANSSLN
PVIYAFNADFQKVFAQLLGCSHFCRTPVETVNISNELISYNQDIVFHKIEAAAYIHMPNAVTPGNREVD
NDEEEGPFDRMFQIYQTSPGDGPVAESVWELDCEGEISLDKITPFTPNGFH

>Aplysia californica protein (sequence taken from BLAST result)
Article Error (ETS)
GNIFVITAILVEKSLQ-
GVSNYLILPLAVTDLLAVLVMPPLSIIYEISIHWFLGNAVCDMWVSMVDLCCASILHLVIAIFDRYWAVS
N-IDYVRSRNARQILLMVAIVWTVSVFISMPL-----  

FGWRHDSDDPELTGQ-CLISQDHGYTVFSTVGAFYCPLLLMFVINFKIYRAARYRIRK
>XP_035826857.1 5-hydroxytryptamine receptor 5B-like [Aplysia
californica]
MCWHNISVQRMGRNASLLEANMYAMFPGYNITTGWPPRDNNNSNGYFDQVTVNGSDRPELLRQYNYHEPK
YEEPMIVICIFLGCMLAIILGNIFVITAILVEKSLQGVSNYLILSLAVTDLLAVLVMPPLSIIYEISI
WFLGNAVCDMWVSMVDLCCASILHLVIAIFDRYWAVSNIDYVRSRNARQILLMVAIVWTVSVFISIPPLF
GWRHDSDDPELTGQCLISQDHGYTVFSTVGAFYCPLLLMLVINFKIYRAARYRIRKRFGGRRGGKHQALHV
PLPAVTVETSHRTNTRNSSGSDVSQDGISMYPSCTNANDVTRDMESPLDVQENGQALLQQDPNLTRMLS
NTLTVPACTVSVPLPRPGTSSASNNNVTKRLRRTDKERFRREKMEMRERKAARVLGIITGAFVVVCWLPFF
VVAVVKPMCGTPCDMPSYVSYLFLWLGYVNSLINPIIYTIFNPSFRCAFNKIFLRRIKSVNRIT
>GFO38024.1 5-hydroxytryptamine receptor 1a [Plakobranchus ocellatus]
MDILDATVEPGSVSSSYDVVSASPVLSTVLSKSLMSNTPTLIGNWTTEILGNLSGHNNTDDEKDELVIYS
YHHPRYGEMTMILICILLGGMILAIILGNIFVITAILVEKTLQGVSNYLILSLAMTDLLAVLVMPPLSIN
EISTNWFLGTVLCDMWISMDVLCCASILHLVIAIFDRYWAVSNIDYIRSRNARQILFMVAVVWMVSCIS
IPPLFGWRHDTDGPEVTGQCLISQDHGYTIFSTVGAFYCPLFLMLIINFKIYRAARYRIRRKRFGNRCHH
NGLQAIHVPISVNADASQRTNTRHSSGSDVSQDGVSIMYPSCTNANDVTRMDTESPLDLAETNDALLSQD
PGIQRMLSNTLTVPACTVNVLARPGPGGMSSNNNVPKRQRRGVDKERYRREKMEMRERKAARVLGIITG
AFVVVCWLPFFIAVLMPMCGDHYIPGYVLSFLWLGYFNSLINPIIYTIFNPSFRVAFRKIFFRIKTSK
RLR
>GFR87408.1 5-hydroxytryptamine receptor [Elysia marginata]
```

MQTTDGAIQTHIMSTPAHSLAGVSPPLSVSLTTSKEARYDQATFALGNVTNFLGNLSDSNTTDKDKDEM
VIYPYHPKYGEVTMIFICILLGGMILAIILGNIFVITAIIVEKSLOQGVSNYLVLSLAMTDLLAVLVMPL
SLINEISTNWFLGNVLCDMWISMDVLCCCTASILHLVAIAFDRYWAVSNIDYIRSRNARQILFMVAIVWMVS
VSISIPPLFGWRDESDGPEFHGQCLISQDHGYTIFSTVGAFYCPLLLMLIINFKITYRAARYRIRRKGNGR
NGHGQHNGMQAVHVPISVVNADASQRTNMRHSSGSDVSQDGVSMMPSCTNANDVTRMDTESPLEAETTD
ALLSQDPSLQRMLNNLTVPACTVNAMARPGSGGVLSSNNVPKRAKRGADKERYRREKMEMRERKAARV
LGIITGAFVVCWLPPFILAVLLPMCGERCVIPDYVLSLFLWLGYFNSLINPIIYTIFNPSFRVAFRKIFFR
RIRSKRLR

>KAH9488655.1 5-hydroxytryptamine receptor [Bulinus truncatus]
MFNSNTTEDSLQPMKTSASLLYQVNISDSLVHVTKSLOMDNVTSVLSTIQFNNTGYTTISSHTFHHIPKYDE
PVMVLICILLGGMILAIILGNIFVITAILVERS LHGVSNYLILSLATT DLLAVLVMPLVS LINEVSVKWFL
GPALCDMWVSMDVLCCCTASILHLVAIAFDRYWAVSNIDYIRSRSAKQILFMVAVVWMVSICISIPPLFGWR
HSTDGPDISGQCLISQDYGYTIFSTLGAFYCPLLLMLIINFKITYRAARYRIRRKGRTKFHVPLPIVT
METSQRLLNRSSGSDVSQDGLSTNANDITRLDMESPLDEADNGQSGSEYDANLTRVFHNTLTVPANTVNL
VVTRSTTPCENNMMI QYNITIYKLRMNVN TDINL

>XP_025095249.1 5-hydroxytryptamine receptor-like [Pomacea
canaliculata]
MRNTTDSTVGWMVGDLTTWPTPASALYISIPNVSHNYYETDSTEVGSLAARYDISHIPRFPLTTLVSICV
VLGAMIVATVLGNVFVITAI VVERSLQGVSNYLILSLAVTDLLAVLVMPLSLLNEVSVNWLGRALCDMW
VSMDVLCCCTASILHLVAIAFDRFWAVSNIDYVRRRCARQILLMIALVWFV SIAISIPPLFGWRGEMDNPEL
SGQCMISQDHGYTIFSTVGAFYCPLVLMVLNFKIYRAARYRIRRKG FVA VGGGPGGGGGDGGGGGGGG
AGRNSRRHHRGKVQVPMVYVEEITAQRTQRNSSGSDVSQDGYSLFNASCAHNEISRMDTGMDSGQDWNQSG
Y PGDSTTDGSNSNPENSYYPVPTEEEVACPDPLPNSFSSNHLRVDTIIVPSRSILLECRRPTVNNHLSLEC
RPF LGDGRPCSDGGFGGDHREVSAQDVSLVGFGERDCNNSSQDVGSVLLDVTNGSRRECSGFANNGSAN
TMAGHDQDGCVNHEIPMTALDLPCAIRKNSLDEELVGLKPRCVV LNSKG GGGGGSGNQLTVPRGGGMP
PTLGTQPNKS RANNNTKRAHNREKERHRREKLEMRRERKAARV LGIITGAFVVCWLPPFILALLSPFCTSA
CSI PLEV YSLFLWLGYVNSLLNPIIYTIFNPSFRCAF R KIFFRRLRSIGR

>QEZ90769.1 5-hydroxytryptamine receptor 1A-like protein [Haliothis
discus hannai]
MENSTEELPYNVAMYIFDNVTAGLDDNL TANGTDNTVERRYTNHVPY NEGIMIFICILGSMILATILG
NVFVLTAILIEKSLQGVSNYLILSLGVTDLMAVLVMPLSLVNEISINWYLGGEVCDMWISMDVLCCASI
LHLVAISFDRYWAVSNIDYVRRRCARQIMLVIVWWFV SVSISIPPLFGWKDNVD PDVTGQCMISQDG YT
IFSTVGAFYCPLTLMVLNFKIYKAARS RIRKKF GGRPRPAPV PVQVENTGNPSRNSSGSDVSQDGFSM
FNGSCVN MNEISRIESTIEGPEEAQNGNGLQTNNGLGTNGNLPNFGPSLT VPGNVYVLSKNRSNNNSKR
ARNREKD KLRKEK IEMRERKAARV LGIITGAFVVCWLPPF MIALIAPFC KESCTFPPI MMSFFLWLGYFN
SLLNPIIYTIFNPSFRTAFRKIFFRKLRLVQR

>XP_041366623.1 5-hydroxytryptamine receptor 1D-like [Gigantopelta
aegis]

MPTKKTAMENFTISTITVVSLKTLDTNSTSINVSTYVLANDSADVFGNGTPYDDVVTRYMMHVPYDVGI
 MTVICALLGAMILATILGNVFVITAIVIEKSLQGVSNVLVSLAVTDLLAVLVMPISLINDVSVHWYLQG
 EVCDFWVSMVDLCCTASILHLVAISLDRFWAVSNIDYIRRCAKQIIIMVVVWFVAISISIPPLFGWKDE
 GNPDYTGMCCLISQDQGYTIFSTVGAFYFPLIFMLVLNPKIYIAARSRIRKKNFGGKPRPVPVMVTVENGT
 AGHASRNSSGSDVSQDGFSMYNGSCVNMMNECSKFESTYEPSQMDGPNDGAATTTTTSFNGMMHVTSCGS
 NLGSTLSPNFRPSLTVPNNGNVVSKNRSNNNSKGARNREKDLRKEKIEMRERKAARVLGIITGAFLV
 CWLPFFLNALILPFCSVSCTSHREVVSFLWLGYFNSLLNPITYTIFNPSFRIAFRKLFFRRLRHLKR
 >XP_022308516.1 5-hydroxytryptamine receptor-like [Crassostrea
 virginica]
 MSTEQQNAGNNATGKMDSVISINLGGRMLGEALDHNSNKSQIYYSSLFGTRVNGTIIPPEIMTFDPNNESITG
 FQYEPRYDLELMVFICVLIGCVIVSIVIGNVFVISATILEKSLQGVSNYLILSLAVTDLMVAVLVMPISLI
 DEISLHWYLGPACDMWTSFDILCCTASILHLVAISLDRYWAVSNIDYIRRCKQILMMIAVVVVVAIAI
 SIAPLFGWKDTNDHPEQTGVCKISQEAKYTIFATVGSFYCPLILMVLLNFKIYRAARSRIRKKHINGMHTP
 IRAYKRPSITAADATTRHNSSGSDISQDGYTVNGVNQHEEPEPVFNDEQDDATEVNTRFLVVPSNVYL
 NKQLNPPKNSNNNSKPIAYCDKERLRAKIEMKRERKAARVLGIITGAFIFCWLPFFLLALISPFCPSD
 FHPAMFSVFLWLGYVNSLNLNPITYTIFNPSFRNAFRKLIFKRHRRRRGR

The result of multiple sequence alignment using T-COFFEE is as shown as below.

CLUSTAL W (1.83) multiple sequence alignment

Homo_sapiens	MLPPGSNGT-AY-PGQFA--L-----
	-----YQQL-----A
Novel_aplysia_californica_protein	-----
Aplysia_californica	MCWHNISVQ-RM---
GRNASLLEANMYAMFPGYNITTGWPPRDNNNSNNSGYFDQV-----	-----TVNGSD-R
Plakobranchus	MDILDATVE-PG-SVSSSY-DVV-
SASPVPVLSVTIJKS---SLSMNT--PTLIGNWTT---EILGNL---S---GHNNTDDE	
Elysia_marginata	MQTTDGAIQ-TH-IMSTPAHSLA-
GVSPPLPLSVSLTTSKEARYDQA--TFALGNVTL---NFLGNL---S---DSNTTDKD	
Bulinus_truncatus	MFSNSTTED-SLQPMKTSASLLY-Q---VNISDSL-
-----VHVT--KSLMDNVT---SVLSTI-----QFNNTG-Y	
Pomacea_canaliculata	MRNTTDSTV-GW-MVGD-----
LTT----TWPTP--AS-ALYISI---P---NVSHNNSYETDSTE-V	
Haliothis_discus_hannai	Article Error (es) MENSTEELPYNV-AMYI-----
-----FDNVTV--MAingGLDD--NLTANGTD--	
Gigantopelta_aegis	MPTKKTAME-NF-TISTI--TVV-S-----
LKT----LDTNS--TS-LINVST---YVLANDSA--DVFGNGTP-Y	
Crassostrea_virginica	MSTEQQNAG-NN-ATGKMD-SVI-S---
INLGGRMLGE--ALDHSN--KSQIYYSSLFGTRVNGTIIP---EIMTFD-P	
 Homo_sapiens	QGNAVGGSAGAPPLGPSQVVTAC-
LLTLLIIWTLLGNVLVCAAIVRSRHLRANMTNVFIVSLAVSDLFVALLVMPWKAVA	
Novel_aplysia_californica_protein	-----
GNIFVITAILVEKSLQG-VSNYLILPLAVTDLLAVLVMPLSLIY	
Aplysia_californica	
PELLRQNYHEPKYEEPVMIVICIFLGCMILAIILGNIFVITAILVEKSLQG-	
VSNYLILSLAVTDLLAVLVMPLSLIY	

Plakobranchus
 KDELVIYSYHHPRYGEMTMILICILLGGMILAIILGNIFVITAILVEKTLQG-
 VSNYLILSLAMTDLLVAVLVMPLSLIN
 Elysia_marginata
 KDEMVIYPYHHPKYGEVTMIFICILLGGMILAIILGNIFVITAIIVEKSLQG-
 VSNYLVLISLAMTDLLVAVLVMPLSLIN
 Bulinus_truncatus
 TTISSHHTFHPIKYDEPVMLICILLGGMILAIILGNIFVITAILVERSLHG-
 VSNYLILSLATTDLLVAVLVMPPVSLIN
 Pomacea_canaliculata
 GSLAARYDSHIPRFPLTTLVSICVVLGAMIVATVLGNVFVITAIIVVERSLOG-
 VSNYLILSLAVTDLLVAVLVMPLSLLN
 Haliotis_discus_hannai
 NTVERRYTNHVPYNEGIMIFICLILGSMILATILGNVFVLTAILIEKSLQG-
 VSNYLILSLGVTDLMVAVLVMPLSLVN
 Gigantopelta_aegis
 DDVVTRYMMHVPRYDVGIMTVICALLGAMILATILGNVFVITAIIVIEKSLQG-
 VSNYLVLISLAVTDLLVAVLVMPISLIN
 Crassostrea_virginica
 NNESITGFQEPRYDLELMVFICVLIGCVIVSIVIGNVFVISAIILEKSLQG-
 VSNYLILSLAVTDLMVAVLVMPISLID

* * : * : * : . : * : . : * : * : . : * : * : * : * : * : * :

Homo_sapiens	EVAGYWPF-
AFCDVWVAFDIMCSTASILNLCVISVDRYWAISRPFRYKRKMTORMALVMVGLAWTLSILISFIPVQLNW	
Novel_aplysia_californica_protein	
EISIHWFGLGNAVCDMWVSMDVLCCCTASILHLVAIAFDRYWAVSN-	
IDYVRSRNARQILLMVAIWWTVSFISMPPL-FG-	
Aplysia_californica	
EISIHWFGLGNAVCDMWVSMDVLCCCTASILHLVAIAFDRYWAVSN-	
IDYVRSRNARQILLMVAIWWTVSFISIPPL-FG-	
Plakobranchus	
EISTNWFLGTVLCDMWISMDVLCCCTASILHLVAIAFDRYWAVSN-	
IDYIIRSRNARQILFMVAVWMVSVCISIPPL-FG-	
Elysia_marginata	
EISTNWFLGNVLCDMWISMDVLCCCTASILHLVAIAFDRYWAVSN-	
IDYIIRSRNARQILFMVAIWWMVSVSISIPPL-FG-	
Bulinus_truncatus	
EVSVKWFGLGPALCDMWVSMDVLCCCTASILHLVAIAFDRYWAVSN-	
IDYIIRSRSAKQILFMVAVWMVSVCISIPPL-FG-	
Pomacea_canaliculata	
EVSVNWLGRALCDMWVSMDVLCCCTASILHLVAIAFDRFWAVSN-	
IDYVRRRCARQILLMIALVWFVSIASISIPPL-FG-	
Haliotis_discus_hannai	
EISINWYLGGEVCDMWISMDVLCCCTASILHLVAISFDRYWAVSN-	
IDYIIRRRCARQIMLMVVVWFVSVSISIPPL-FG-	
Gigantopelta_aegis	
DVSVHWYLGQEVCDFWVMSMDVLCCCTASILHLVAISLDRFWAVSN-	
IDYIIRRRCAKQIIIMVVVWFVAISISIPPL-FG-	
Crassostrea_virginica	
EISLHWYLGPAVCDMWTSFDILCCTASILHLVAISLDRYWAVSN-	
IDYIIRRRCRKQILMMIAVWWVVAIAISIAPL-FG-	

* : * . * * . * : * : * . * * * : * . * : * : * : * . : * * : . : * : . * : : * : * :

*: .:

Homo_sapiens
 HRDQAASWGGILDLPNNLANWTPWEEDFWEPDVNAENCDSLNRTYAISSLISFYIPVAIMIVTYTRIYRI
 AQVQIRRIS
 Novel_aplysia_californica_protein -----WRHDSDDPELTG-
 QCLISQDHGYTVFSTVGAFYCPLLLMFVINFKIYRAARYRIRK--
 Aplysia_californica -----WRHDSDDPELTG-
 QCLISQDHGYTVFSTVGAFYCPLLLMLVINFKIYRAARYRIRKKR
 Plakobranchus -----WRHDTDGPEVTG-
 QCLISQDHGYTIFSTVGAFYCPLFLMLIINFKIYRAARYRIRRK
 Elysia_marginata -----WRDESDGPEFHG-
 QCLISQDHGYTIFSTVGAFYCPLLLMLIINFKIYRAARYRIRRK
 Bulinus_truncatus -----WRHSTDGPDISG-
 QCLISQDYGYTIFSTLGAFYCPLLLMLIINFKIYRAARYRIRKK
 Pomacea_canaliculata -----WRGEMDNPELSG-
 QCMISQDHGYTIFSTVGAFYCPLVLMVLVLFNKIYRAARYRIRRK
 Haliotis_discus_hannai -----WKDNV-DPDVTG-
 QCMISQDHGYTIFSTVGAFYCPLTLMLVLFNKIYKAARSRIRKK
 Gigantopelta_aegis -----WKDEG-NPDYTG-
 MCLISQDHGYTIFSTVGAFYCPLIFMLVLNYKIIYAARSRIRKK
 Crassostrea_virginica -----WKDTNDHPEQTG-
 VCKISQEKAATIFATVGSFYCPLILMVLLNFNKIYRAARSRIRKKH
 * . * : .
 * * : *:: :: : ** *: : *.: : ** *: : **:

 Homo_sapiens SLERAA-----
 -----EHAQSCRSSAAC-----
 Novel_aplysia_californica_protein -----

 Aplysia_californica FGGRRGG---K-----H---
 QALHVPLPAVTVET-SHRTNTRNSSGSDVSQDGISMYPSCTNAND
 Plakobranchus FGNRNC---H-----
 HNGLQAIHPVISVVNADA-SQRTNTRHSSGSDVSQDGVSIMYPSCTNAND
 Elysia_marginata FGNRNG---H-----G---
 QHNGMQAVHVPISSVNADA-SQRTNMRHSSGSDVSQDGVSIMYPSCTNAND
 Bulinus_truncatus FLGRTT-----
 KFHVPLPIVTMET-SQRINNRNSSGSDVSQDGGLST-----NAND
 Pomacea_canaliculata FVAVGGGGGGGGDGGGGGGGAGRNSRRHRG-
 -KVQVPMVYVVEITAQRTQRNSSGSDVSQDGYSLFNASCA-HNE
 Haliotis_discus_hannai FGGR---P-----R---
 -PAPVPVVQVEN--TGNPSRNSSGSDVSQDGFSMFNGSCVNMMNE
 Gigantopelta_aegis FGGK---P-----R---
 -PVVPVMVTVENGTAGHASRNSSGSDVSQDGFSMYNGSCVNMMNE
 Crassostrea_virginica INGMHT---P-----IR-
 -AYKRPSITA---ADATTRHNSSGSDISQDGYTVYNGVNQH-EE

 Homo_sapiens -----

 Novel_aplysia_californica_protein -----

 Aplysia_californica VTRVDMES--PLDVQE-----
 ---NGQALLQ-----QDPNLTRMLS-----NTLT
 Plakobranchus VTRMDTES--PLDLAE-----
 ---TNDALLS-----QDPGIQRMLS-----NTLT
 Elysia_marginata VTRMDTES--PLELAE-----
 ---TTDALLS-----QDPSLQRMLN-----NTLT

Bulinus_truncatus	ITRLDMES--PLDEAD-----
----NGQSGSE-----	YDANLTRVFH-----NTLT
Pomacea_canaliculata	
ISRMDTGMDSGQDWNQSGYPGDSTTDGSGSNPENSYYPVPTEEEVAC-----	
PDPPLPNSFSSNHLRVDTII	
Haliotis_discus_hannai	ISRIESTIEGPEE-AQNGN-----
-GLQTNNGLLG-----	TNGNLPPNFG-----PSLT
Gigantopelta_aegis	CSKFESTYPEPS-Q-MDDGP-----
NDGAATTTTSFNGMMHVTS CGS NLG STLSPNFR-----	PSLT
Crassostrea_virginica	-----PEP-VFNDTE-----
----QDDA-----	TEVNT-----RFLV
 Homo_sapiens	-----
 Novel_aplysia_californica_protein	-----
 Aplysia_californica	VPA-----CT-----VSVPLPRPGTS--
-----SASNN-----	
 Plakobranchus	VPA-----CT-----VNVTLARPGPGG-
-----MSSNN-----	
 Elysia_marginata	VPA-----CT-----VNVAMARP GSGG-
-----VLSNN-----	
 Bulinus_truncatus	VPA-----NT-----VNLVVTRST----
-----TPCNN-----	
 Pomacea_canaliculata	
VPSRSLLECRPPTVNNHLSLECRPFLGDGRPCSDGGFGGDHREVSAQDVSLVGFERDCNNSSQDVGSVL	
LDVTNGSRR	
Haliotis_discus_hannai	VPGN-----

Gigantopelta_aegis	VPNNNGN-----

Crassostrea_virginica	VPSN-----V-----YVLN-KQLNPP--
-----KNSNN-----	
 Homo_sapiens	-----
 Novel_aplysia_californica_protein	-----
 Aplysia_californica	-----
 Plakobranchus	-----
 Elysia_marginata	-----
 Bulinus_truncatus	-----
 Pomacea_canaliculata	
ESCGSFANNGSANTMAGHDQGCVYNHEIPMTALDLPCIAIRKNSLDEELVGLKPRCVV LNSKG GGGGGGS	
GNQLTVPRG	
Haliotis_discus_hannai	-----VY-----
-----VL-----	
Gigantopelta_aegis	-----VY-----
-----VV-----	
Crassostrea_virginica	-----

Homo_sapiens	-----APDT--
SLRA S IKKETKVLKTL S VIMGVFVCCWLPFFILNCMVPFC G HPEGP	
Novel_aplysia_californica_protein	-----
<hr/>	
Aplysia_californica	-----NVTKRL-
RRTDKERFRREKMEMRERKAARVLGIITGAFVV C WLPFFVVAVVKPMCGTPCDM-	
Plakobranchus	-----
NVPKRQR G V D KERYRREKMEMRERKAARVLGIITGAFVV C WLPFFIIAVLMPMC G DHCYI-	
Elysia_marginata	-----
NVPKRAKRGADKERYRREKMEMRERKAARVLGIITGAFVV C WLPFFILAVLLPMCGERCVI-	
Bulinus_truncatus	-----NM-MIQ-YN-----
<hr/>	
Pomacea_canaliculata	SGMPPTLGTQPNKS R ANNNTKRA-
HNREKERHRREKLEM R ERKAARVLGIITGAFVV C WLPFFILALLSPFCTSACSI-	
Haliotis_discus_hannai	-----SKNRSSNNNSKRA-
RNREKD K LRKE K IEMRERKAARVLGIITGAFVV C WLPFFMIALIAPFCKESCTF-	
Gigantopelta_aegis	-----SKNRSSNNNSKG A -
RNREKD K LRKE K IEMRERKAARVLGIITGAFLVCWLPFFLNALILPFCSVSCTS-	
Crassostrea_virginica	-----NNSKPI-
AYCDKERLRKAKIEMKRERKAARVLGIITGAFIFCWLPFFLLALISPFCP-SCDF-	
<hr/>	
Homo_sapiens	PAGFP C VSETTFDVFVWFGWANSSLNPVIYA-
FNADFQKVFAQLLGCSHFC S RTPVETVNISNELISYNQDIVFHKEIAA	
Novel_aplysia_californica_protein	-----
<hr/>	
Aplysia_californica	-----
PSYVYSLFLWLGYVN S LINPIIYTIFNPSFRCAF N KIFF-----	
--	
Plakobranchus	-----
PGYVLSLFLWLGYFN S LINPIIYTIFNPSFRVAFRKIFF-----	
--	
Elysia_marginata	-----
PDYVLSLFLWLGYFN S LINPIIYTIFNPSFRVAFRKIFF-----	
--	
Bulinus_truncatus	-----ITIYKLRMN V NTDINL-----
<hr/>	
Pomacea_canaliculata	-----
PLEVYSLFLWLGYVN S LINPIIYTIFNPSFRCAF N KIFF-----	
--	
Haliotis_discus_hannai	-----
PPIMMSFFLWLGYFN S LINPIIYTIFNPSFRTAFRKIFF-----	
--	
Gigantopelta_aegis	-----
HREVV S FFLWLGYFN S LINPIIYTIFNPSFRIAFRKLFF-----	
--	
Crassostrea_virginica	-----
HPAMFSVFLWLGYVN S LINPIIYTIFNPSFRNAFRKLIF-----	
--	
<hr/>	
Homo_sapiens	
AYIHMM P NAVTPGNREV D NEEGPFDRMFQ I YQ T SPDGDPVAESVWELDCEGEISLDKITPFTPNGF-H	
Novel_aplysia_californica_protein	-----
<hr/>	

Aplysia_californica	-----
	RRIKSVNRIT
Plakobranchus	-----
	RRIKTSKRLR
Elysia_marginata	-----
	RRIRSK-RLR
Bulinus_truncatus	-----

Pomacea_canaliculata	-----
	RRLRSIG--R
Haliotis_discus_hannai	-----
	RKLRLVQ--R
Gigantopelta_aegis	-----
	RRLRHLK--R
Crassostrea_virginica	-----
	KRHRRRRG-R

On the other hand, the result of multiple sequence alignment using MUSCLE is as shown as below.

CLUSTAL multiple sequence alignment by MUSCLE (3.8) | Frag. ETS

Homosapiens	-----MLPPGSNGTAY-----PGQFALYQ--
	--QLAQGNNA
Crassostrea_virginica	----MSTEQQNAGNNATGKMDSVI-----
SINLGGGRMLGEALDHSNKSQIIYSSLFGTR	
Gigantopelta_aegis	-----MPTKKTAME-----
NFTISTITVVSLKTLDTNSTSINVSTYVL	
Haliotis_discus_hannai	-----MENSTEEL-----PYNVAMYI--
	--FDNV
Pomacea_canaliculata	----MRNTTDSTVGWMVGDLTTW-----PTPASALY--
	--ISIPN
Bulinus_truncatus	--MFSNSTTEDSLQPMKTSASLLY-----
QVNISDSLVHVTKSLMDNVTSVLSTIQFNN	
Plakobranchus_ocellatus	-----MDILDATVEPGSVSSSYDVVSASPVPLSVTL--
-KSSLMSNTPTLIGNWTTEI	
Elysia_marginata	MQTTDGAIQTHIMSTPAHSLAGVS-----
PLPLSVSLTTSKEARYDQATFALGNVTLNF	
novel_Aplysia_californica	-----

Aplysia_californica	-----MCWHNISVQRMGRNASLL-----EANMYAMF--
	--PGYNITTGWPPRDNNNSNN
Homosapiens	VGGSGAGA-----PPLGPSQVVTAC-
LLTLLIIIWTLGNVLVCAAIVR	
Crassostrea_virginica	VNGTIPIPEIMTFDPNNESITGFQYEPRYDLELMVFICVLIGCVIVSIVIGNVFVISAIIL
Gigantopelta_aegis	ANDSADVFNNGTPYDDVVTRYMMHVPRYDVGIMTVICALLGAMILATILGNVFVITAIVI
HALIOTIS_discus_hannai	VAGLDDNLNTANGTDNTVERRYTNHVPRYNEGIMIFICLILGSMILATILGNVFVLTAILI
Pomacea_canaliculata	VSHNSSYETDSTEVGSLAARYDSHIPRFPLTTLVSICVVLGAMIVATVLGNVFVITAIVV

```

Bulinus_truncatus          TGYTTISSHT-----
FHHIPKYDEPVMVLICILLGGMILAIILGNIFVITAILV
Plakobranchus_ocellatus   LGNLSGHNNTDDEKDELV-
IYSYHHPRYGETMILICILLGGMILAIILGNIFVITAILV
Elysia_marginata          LGNLSDSNTTDKDDEMVA
IYPYHHPKYGEVTMIFICILLGGMILAIILGNIFVITAIIV
novel_Aplysia_californica -----
-----GNIFVITAILV
Aplysia_californica
SGYFDQTVNGSDRPELLRQNYHEPKYEEPVMIVICIFLGCMILAIILGNIFVITAILV

```

*** : * : *** :

Homosapiens	
SRHLRANMTNVFIVSLAVSDLFVALLVMPWKAVAEVAGYWPF	-AFCDVVWAFDIMCSTA
Crassostrea_virginica	EKSLQG-
VSNYLILSLAVTDLMAVILVMPISLIDEISLHWYLGPAVCDMWTSFDILCCTA	
Gigantopelta_aegis	EKSLQG-
VSNYLVLSLAVTDLVAVLVMPISLINDVSVHWFYLGQEVCDFWVSM	SMDVLCCTA
Haliotis_discus_hannai	EKSLQG-
VSNYLILSLGVTDLMVAVLVMPISLNEISINWYLGGEVCDMWISMDVLCCTA	
Pomacea_canaliculata	ERSLQG-
VSNYLVLSLAVTDLLVAVLVMPISLNEVS	VNWYLGALCDMWVSMDVLCCTA
Bulinus_truncatus	ERSLHG-
VSNYLILSLATTDLLVAVLVMPVSLINEVS	VKWFLGPALCDMWVSMDVLCCTA
Plakobranchus_ocellatus	EKTLQG-
VSNYLILSLAMTDLLVAVLVMPISLNEISTNWFLGT	VLCDMWISMDVLCCTA
Elysia_marginata	EKSLQG-
VSNYLVLSLAMTDLLVAVLVMPISLNEISTNWFLGNVLC	CDMWISMDVLCCTA
novel_Aplysia_californica	EKSLQG-
VSNYLILPLAVTDLVAVLVMPISLIEISIHWFLGN	AVCDMWVSMDVLCCTA
Aplysia_californica	EKSLQG-
VSNYLILSLAVTDLVAVLVMPISLIEISIHWFLGN	AVCDMWVSMDVLCCTA
	.. * .. : * : : . * : * : * : * : * . : : :
*	: * . * . * : : : : : * . *

Missing ":" 

Homosapiens	
SILNLCVISVDRYWAISRPFRYKRKMTQRMALVMVGLAWTLSILISFIPVQLNWHRDQAA	
Crassostrea_virginica	SILHLVAISLDRYWAVSN-
IDYIRRRC	TKQILMMIAVVVVVAIAISIAPL-----
Gigantopelta_aegis	SILHLVAISLDRFWAVSN-
IDYIRRRC	AKQIIIMVVVVWFVAISISIPPL-----
Haliotis_discus_hannai	SILHLVAISFDRYWAVSN-
IDYIRRRC	ARQIMLMVVIVWFVSVSISIPPL-----
Pomacea_canaliculata	SILHLVAIAFDRFWAVSN-
IDYVRRRC	ARQILLMIALVWFVSIAISIPPL-----
Bulinus_truncatus	SILHLVAIAFDRYWAVSN-
IDYIRSRS	AKQILFMVAVVWMVSICISIPPL-----
Plakobranchus_ocellatus	SILHLVAIAFDRYWAVSN-
IDYIRSRNAR	QILFMVAIVWMVSVCISIPPL-----
Elysia_marginata	SILHLVAIAFDRYWAVSN-
IDYIRSRNAR	QILFMVAIVWMVSVSISIPPL-----
novel_Aplysia_californica	SILHLVAIAFDRYWAVSN-
IDYVRSRNAR	QILLMVAIVWTVSVFISMPP-----
Aplysia_californica	SILHLVAIAFDRYWAVSN-
IDYVRSRNAR	QILLMVAIVWTVSVFISIPPL-----

Pomacea_canaliculata	ITAQRTQ-RNSSGSDVSQDGYSLFNASCAH-
NEISRMDTGMDSGQDWNQSGYPGDSTTDG	
Bulinus_truncatus	ETSQRLLNNRNSSGSDVSQDGLS-----TNANDITRLDM-

Plakobranchus_ocellatus	DASQRTNTRHSSGSDVSQDGVS MYMPSCTNANDVTRMDT-

Elysia_marginata	DASQRTNMRHSSGSDVSQDGVS MYMPSCTNANDVTRMDT-

novel_Aplysia_californica	-----

Aplysia_californica	ETSHRTNTRNNSGSDVSQDGISMYMPSCNANDVTRVDM-

 Homosapiens	-----
-----AP	
Crassostrea_virginica	-----PEPVFND-----TEQDDATEVNT-----
-----RFLVVP	
Gigantopelta_aegis	-----DGPNDGA-----
ATTTTTTSFNGMMHVTSCGSNLGSTLSPNFR	-----PSLTVP
Haliotis_discus_hannai	-----EGPEEAQNGNGLQTNNGLGTNG-----
-NLPPNFG-----PSLTVP	
Pomacea_canaliculata	SGSNPENSYYPV-----PTEEEVACPDP-----
-PLPNSFSSNHLRVDTIIVP	
Bulinus_truncatus	-----ESPLDEA-----DNGQSGSEYDA-----
-NLTRVFH-----NTLTVP	
Plakobranchus_ocellatus	-----ESPLDLA-----ETNDALLSQDP-----
-GIQRMLS-----NTLTVP	
Elysia_marginata	-----ESPLELA-----ETTDALLSQDP-----
-SLQRMLN-----NTLTVP	
novel_Aplysia_californica	-----

Aplysia_californica	-----ESPLDVQ-----ENGQALLQQDP-----
-NLTRMLS-----NTLTVP	
 Homosapiens	--DTSL-----

Crassostrea_virginica	--SNVY-----

Gigantopelta_aegis	NNGNVY-----

Haliotis_discus_hannai	--GNVY-----

Pomacea_canaliculata	--
SRSLLERCRPPTVNNHLSLECRPFLGDGRPCSDGGFFGDHREVSAQDVSLSVGFERDC	
Bulinus_truncatus	--ANTV-----

Plakobranchus_ocellatus	--ACTV-----

Elysia_marginata	--ACTV-----

novel_Aplysia_californica	-----

Aplysia_californica	--ACTV-----

Homosapiens	-----
Crassostrea_virginica	-----
Gigantopelta_aegis	-----
Haliotis_discus_hannai	-----
Pomacea_canaliculata	-----
NNSSQDVGSVLLDVTNGSRESCGSFANNGSANTMAGHDQDGCVYNHEIPMTALDLPCI	
Bulinus_truncatus	-----
Plakobranchus_ocellatus	-----
Elysia_marginata	-----
novel_Aplysia_californica	-----
Aplysia_californica	-----
 Homosapiens	-----
Crassostrea_virginica	-----VLNKQLNPPKNSNNNN-----
Gigantopelta_aegis	-----VVSKNR-----SNNN-----
Haliotis_discus_hannai	-----VLSKNR-----SNNN-----
Pomacea_canaliculata	AIRKNSLDEELVGLKPRCVVLNSKGGGGGSGNQLTVPRGSGMPPTLGTQPNKS RANNNT
Bulinus_truncatus	-----
Plakobranchus_ocellatus	-----NVTLARPGPGGMSSNNNVP-----
Elysia_marginata	-----NVAMARPGSGGVLSNNNVP-----
novel_Aplysia_californica	-----
Aplysia_californica	-----SVPLPRPGTSSASNNNV-----
 Homosapiens	-----
RASIKKETKVLTLSVIMGVFVCCWLFFFILNCMVPFCSGHPEGPP	
Crassostrea_virginica	KPIAYCDKERLRKAKIEMKRERKAARVLGIITGAFIFCWLPFFLLALISPFC-PSCDFHP
Gigantopelta_aegis	KGARNREKDRLKEKIEMRRERKAARVLGIITGAFLVCWLPFFLNALILPFCVSCTSHR
Haliotis_discus_hannai	KRARNREKDRLKEKIEMRRERKAARVLGIITGAFVVCWLPFFMIALIAPFCKESCTFPP
Pomacea_canaliculata	KRAHNREKERHRREKLEMRRERKAARVLGIITGAFLVCWLPFFILALLSPFCTSACSIPL
Bulinus_truncatus	-----
-----NLVVTRSTTPCN---	

Plakobranchus_oellatus
 QRQRGVDKERYRREKMEMRRERKAARVLGI ITGAFVVCWLPFFITAVLMPMCGDHCYIPG | Missing "," (ETS)
Elysia_marginata
 RAKRGADKERYRREKMEMRRERKAARVLGI ITGAFVVCWLPFFILAVLLPMCGERCVIPD
novel_Aplysia_californica -----

Aplysia_californica
 KRLRRTDKERFRREKMEMRRERKAARVLGI ITGAFVVCWLPFFVVAVVKPMCGTPCDMPS

Homosapiens AGFPCVSETTFDVFWFGWANSSLNPVIYA-
 FNADFQKVFAQLLGCSHFCSRTPVETVNI
Crassostrea_virginica A-----
 MFSVFLWLGYVNSSLNPIIYTIFNPSFRNAFRKLI-----
Gigantopelta_aegis E-----
 VVSFFLWLGYFNSSLNPIIYTIFNPSFRIAFRKLF-----
Haliotis_discus_hannai I-----
 MMSFFLWLGYFNSSLNPIIYTIFNPSFRTAFRKIF-----
Pomacea_canaliculata E-----
 VYSLFLWLGYVNSSLNPIIYTIFNPSFRCAFRKIF-----
Bulinus_truncatus -----
 NMMMIQYNITIYKLRMNVNTDINL-----
Plakobranchus_oellatus Y-----
 VLSLFLWLGYFNSSLNPIIYTIFNPSFRVAFRKIF-----
Elysia_marginata Y-----
 VLSLFLWLGYFNSSLNPIIYTIFNPSFRCAFNRKIF-----
novel_Aplysia_californica -----

Aplysia_californica Y-----
 VYSLFLWLGYVNSSLNPIIYTIFNPSFRCAFNRKIF-----

Homosapiens
 SNELISYNQDVFHKEIAAYIHMMMPNAVTPGNREVDNDEEEGPFDLDRMFQIYQTSPDGDP
Crassostrea_virginica -----
 ---FKRHRR-RRGR---
Gigantopelta_aegis -----
 ---FRRLRHILKR---
Haliotis_discus_hannai -----
 ---FRKLRLVQR----
Pomacea_canaliculata -----
 ---FRRLRSIGR----
Bulinus_truncatus -----

Plakobranchus_oellatus -----
 ---FRRIKTSKRLR---
Elysia_marginata -----
 ---FRRIRS-KRLR---
novel_Aplysia_californica -----

Aplysia_californica
 ---LRRIKSVNRIT-----

Homosapiens VAESVWELDCEGEISLDKITPFTPNGFH
Crassostrea_virginica -----
Gigantopelta_aegis -----

Haliothis_discus_hannai	-----
Pomacea_canaliculata	-----
Bulinus_truncatus	-----
Plakobranchus_ocellatus	-----
Elysia_marginata	-----
novel_Aplysia_californica	-----
Aplysia_californica	-----

From the result, we are able to obtain the ten protein sequences with same length. The multiple sequences are also being arranged in the way that as many characters from each sequence are matched. From that, we are able to identify the regions of similarity among the sequences which may result in evolutionary relationship among the sequences. However, there are still slightly different between the multiple sequence alignment result of the two different programs. The result of T-COFFEE contains more gaps than the result of MUSCLE. This might happen when there is more deletion in the result of T-COFFEE. We can also see that there are different number of asterisks(*) in the results and they actually represent the positions which have single, fully conserved residue. Besides, the number of the colon(:) and the period(.) are also different in both results while the colon(:) actually indicates the conservation between groups of strongly similar properties and the period(.) indicates the conservation between the groups of weakly similar properties.

5.0 Phylogenetic tree creation

In this process, I utilized MEGA X to generate a phylogenetic tree by using the same 10 sequences which I used in the last step. To create the phylogenetic tree, firstly, I copy the sequences and build new protein alignment using the MEGA X alignment explorer and export the alignment to MEGA format. There are a few selections for the alignment approaches. In this project, I chose to align the sequences by MUSCLE. Returning to the main page, I open the alignment file that I export earlier and use it to create a phylogenetic tree. In this case, we can choose either we want to generate a Maximum-Likelihood Tree, Neighbor-Joining Tree or Minimum Evolution Tree. We can also set the analysis preference to obtain the desired output. In this case, I created a bootstrapping Neighbor-Joining Tree using the settings as shown in Figure 5.1. Bootstrapping is performed using 500 for the number of bootstrap replications to identify the percent of instances in which bootstrap trees support each clade in the inferred tree. The phylogenetic tree is as shown in Figure 5.2.

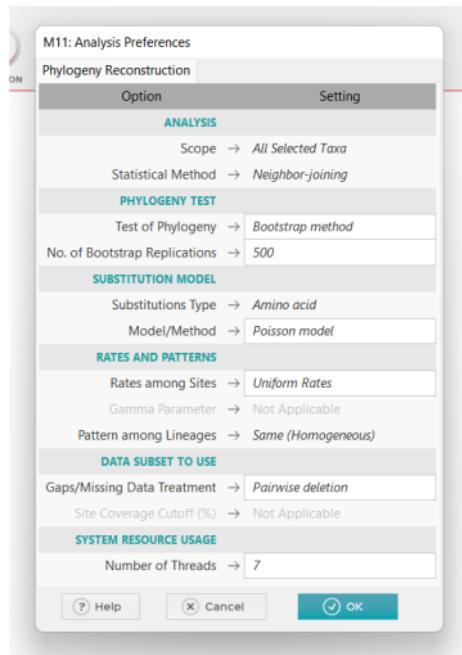


Figure 5.1: Setting of the analysis preferences

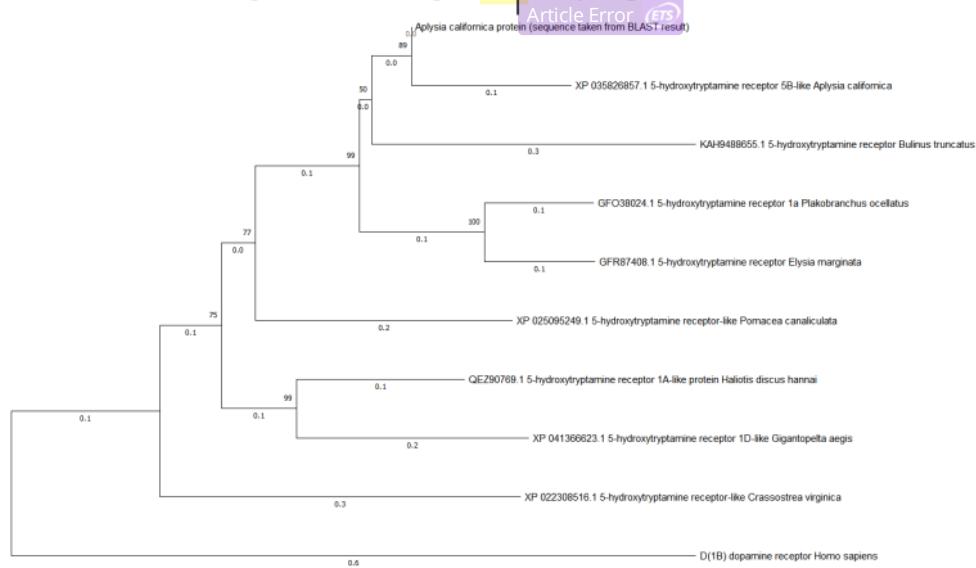


Figure 5.2: Phylogenetic tree

The phylogenetic tree reveals the hypotheses of evolutionary relationship among the group of sequences. From the result, I can actually see that the human D(1B) dopamine receptor is the most distantly related to the novel protein. This can be determined since both

sequence is actually being connected by the deep node and having a less common ancestor. This means that the ancestors of these two sequences might evolve for a very long time. Through analyzing the phylogenetic tree, we can actually understand the evolutionary pathways and the connections among the organisms. From the result also, I can know that the human D(1B) dopamine receptor and the novel protein might not share common traits.

P/V ETS

7

6.0 Conclusion

In our group project, we have learned to find a novel protein by using BLAST and other related programs. To start our project, we have to decide on which known protein that we want to use. Meanwhile, in this alternative assessment, we are required to use a different protein from what we used in the group project. In this case, I have chosen the human D(1B) dopamine receptor since I am quite interested in the particular protein and I wish to know more about it. By using the TBLASTN program against *Aplysia Californica* est database, I am able to get a list of matches which are related to the human D(1B) dopamine receptor. Next, using BLASTP program against nr database for further analysis, I have discovered that the CNSN01-F-142926-501 Normalized CNS library (juvenile 1) *Aplysia californica* cDNA clone CNSN01-F-142926 5', mRNA is a novel protein. After that, I have also done the multiple sequence alignment by using the query protein, novel protein and a few proteins in the family. A phylogenetic tree is also generated to observe the origin and the evolution of the proteins from a common ancestor. Throughout the process, I got a clearer picture on what bioinformatics is in a practical way and I have built my interest towards bioinformatics. However, in this work, there is still some limitation since I only managed to get a few information about the novel proteins through NCBI and PubMed. Nevertheless, I am also limited by my biological knowledge and this has become a difficulty for me to know more clues about the novel protein that I found. Anyway, we all know that sea slug is usually used as a model organism to explore the molecular mechanisms in neuroscience. While *Aplysia Californica* is also one of the species of sea slug, this project might be able to make some contribution to the human neurological and biological field too. As a suggestion for future work, there are still a lot of novel proteins which have not been discovered yet. We as students of bioinformatics and future bioinformaticians should take our responsibility to do more analysis to discover the novel proteins which might be able to contribute in the field of biology.

CHONGKAHWEI_FINAL

ORIGINALITY REPORT



PRIMARY SOURCES

1	library.unisel.edu.my Internet Source	3%
2	Jonathan Pevsner. "Bioinformatics and Functional Genomics", Wiley, 2009 Publication	1%
3	Submitted to Universiti Teknologi MARA Student Paper	1%
4	www.slideshare.net Internet Source	1%
5	ebin.pub Internet Source	1%
6	Submitted to National University of Ireland, Maynooth Student Paper	<1%
7	The Dopamine Receptors, 2010. Publication	<1%
8	Submitted to Sim University Student Paper	<1%
9	vdoc.pub	

Internet Source

<1 %

10 hdl.handle.net

Internet Source

<1 %

11 Submitted to University of Edinburgh

Student Paper

<1 %

12 Submitted to Drexel University

Student Paper

<1 %

13 Submitted to University of Newcastle upon
Tyne

Student Paper

<1 %

14 www.mdpi.com

Internet Source

<1 %

15 breakingmuscle.com

Internet Source

<1 %

16 Submitted to South Dakota Board of Regents

Student Paper

<1 %

17 Vinay Kumar, S. S. Maitra, Rohit Nandan
Shukla. "Environmental Metagenomics: The
Data Assembly and Data Analysis
Perspectives", Journal of The Institution of
Engineers (India): Series A, 2015

Publication

<1 %

18 repository.up.ac.za:8080

Internet Source

<1 %

19	www.coursehero.com Internet Source	<1 %
20	www.scribd.com Internet Source	<1 %
21	"Advances in Bioinformatics and Computational Biology", Springer Science and Business Media LLC, 2013 Publication	<1 %
22	repositorio.unesp.br Internet Source	<1 %
23	repositorium.sdum.uminho.pt Internet Source	<1 %

Exclude quotes On
Exclude bibliography On

Exclude matches Off

CHONGKAHWEI_FINAL

PAGE 1

PAGE 2



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 3



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word. Consider using the article **the**.



Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.



Article Error You may need to remove this article.



Missing "," Review the rules for using punctuation marks.



Article Error You may need to use an article before this word. Consider using the article **the**.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 4



Article Error You may need to remove this article.

PAGE 5



Article Error You may need to use an article before this word.

PAGE 6



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to remove this article.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 7



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 8



Article Error You may need to use an article before this word. Consider using the article **a**.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Article Error You may need to remove this article.



Proofread This part of the sentence contains an error or misspelling that makes your meaning unclear.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Article Error You may need to remove this article.



Article Error You may need to remove this article.

PAGE 9



Article Error You may need to remove this article.



S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.



Article Error You may need to use an article before this word. Consider using the article **a**.



Article Error You may need to use an article before this word.



Article Error You may need to remove this article.



Article Error You may need to use an article before this word.

PAGE 10



Missing "," Review the rules for using punctuation marks.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 11



Missing "," Review the rules for using punctuation marks.



Prep. You may be using the wrong preposition.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Possessive



Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.



S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.



Article Error You may need to use an article before this word. Consider using the article **the**.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 12

PAGE 13



Article Error You may need to use an article before this word.

PAGE 14



S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.



Article Error You may need to use an article before this word. Consider using the article **a**.



Article Error You may need to remove this article.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Article Error You may need to remove this article.



Article Error You may need to use an article before this word. Consider using the article **the**.



Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Possessive Review the rules for possessive nouns.

PAGE 15



S/V This subject and verb may not agree. Proofread the sentence to make sure the subject agrees with the verb.



Article Error You may need to use an article before this word.

PAGE 16

PAGE 17



Article Error You may need to use an article before this word.



Missing ","

PAGE 18

PAGE 19

PAGE 20

PAGE 21

PAGE 22



Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.

PAGE 23



Missing ","

PAGE 24

PAGE 25

PAGE 26

PAGE 27



Missing ","



Article Error You may need to use an article before this word.

PAGE 28



Article Error You may need to use an article before this word.



Article Error You may need to remove this article.



Article Error You may need to use an article before this word. Consider using the article **a**.



Article Error You may need to use an article before this word. Consider using the article **a**.



Frag. This sentence may be a fragment or may have incorrect punctuation. Proofread the sentence to be sure that it has correct punctuation and that it has an independent clause with a complete subject and predicate.



Sentence Cap. Review the rules for capitalization.



Missing "?" Review the rules for using punctuation marks.



Confused You have used either an imprecise word or an incorrect word.



Sentence Cap. Review the rules for capitalization.



Article Error You may need to remove this article.



Article Error You may need to remove this article.



Article Error You may need to remove this article.

PAGE 29



Article Error You may need to remove this article.



Article Error You may need to use an article before this word. Consider using the article **the**.

PAGE 30



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word.



Article Error You may need to use an article before this word. Consider using the article **the**.



Missing "," Review the rules for using punctuation marks.



Wrong Article You may have used the wrong article or pronoun. Proofread the sentence to make sure that the article or pronoun agrees with the word it describes.



P/V You have used the passive voice in this sentence. You may want to revise it using the active voice.

