

## Methodology for Phylogenetic Tree Construction

Kudipudi Srinivas<sup>2</sup>, Allam Appa Rao<sup>1</sup>, GR Sridhar<sup>3</sup>, Srinubabu Gedela<sup>1\*</sup>

<sup>1</sup>International Center for Bioinformatics & Center for Biotechnology, Andhra University, 530 003, India.

<sup>2</sup>Dept of Computer Science, Acharya Nagarjuna University, Nagajuna Nagar-522 010.

<sup>3</sup>Endocrine and Diabetes Centre, Krishna Nagar, Visakhapatnam, 530 002, India.

\*Corresponding author : Srinubabu Gedela, Email: [srinubabuau6@gmail.com](mailto:srinubabuau6@gmail.com).

Received April 20, 2008; Accepted May 15, 2008; Published May 25, 2008

**Citation:** Kudipudi S, Allam AR, Sridhar GR, Srinubabu G (2008) Methodology for Phylogenetic Tree Construction. J Proteomics Bioinform S1: S005-S011. doi:10.4172/jpb.s1000002

**Copyright:** © 2008 Kudipudi S, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

### Abstract

It has been our endower to present and explore different steps involved in the construction of a phylogenetic tree using online services. The application of different bioinformatic tools is necessary for the realization of the desirable results. A stepwise construction strategy using on line ClustalW is proposed with suitable application. The method can hope fully will aid the reader to construct the phylogenetic tree. Construction of phylogenetic tree is useful for comparison of homology of functional proteins in one species and between species.

**Keywords:** Protein; DNA; FASTA; BLAST

### Introduction

Tree representation of the family history of set of sequences that share a common ancestor is called a **Phylogenetic Tree**. A phylogeny tree shows the connection among various organisms and weight of the branches in the tree indicates time between evolutions of different organisms.

Protein Information Resource(PIR) <http://pir.georgetown.edu/>  
Swiss-Prot <http://us.expasy.org/sprot>  
Example to retrieve required Protein sequence from Swiss-Port

#### Uses of Phylogenetic Tree

- Determining the relatives of the organisms and interested.
- Identify the functionality of a gene
- Trace the origin of a gene

1.1.1.Open the link at to <http://www.expasy.org/sprot/sprot-search.html>

1.1.2.Type required protein/gene sequence name in the Gene name window , and click the Submit Query button.

1.1.3.Select the required gene from given result and click

1.1.4.To get the FASTA format, Click the FASTA format button , on the extreme right of the entry.

1.1.5.Save the Sequence into your PC

### Material and Methods

#### 1.Retrieving Required Sequence (Protein/DNA) from Major Databases.

- 1.1. Retrieving Protein sequence using major Protein databases.

Major Protein databases are

- 1.2. Retrieving DNA sequence using major Nucleotide Sequence databases

Major Nucleotide Sequence databases are

- European Molecular Biology Laboratory (EMBL) <http://www.ebi.ac.uk/>
- GenBank (National Center for Biotechnology Information, NCBI) <http://www.ncbi.nlm.nih.gov/>
- DNA databank of Japan (DDBJ) <http://www.ddbj.nig.ac.jp/>

## 2.Using BLAST to Compare Sequence of Interest (Protein/DNA) to other Sequences

2.1. Open the link at to [www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST)

2.2. Click the Standard Nucleotide-nucleotide BLAST[blastn]/Standard Protein-Protein BLAST[blastp]

Ex: Click Standard Protein-Protein BLAST[blastp]

2.3. Open the Saved FASTA-formatted sequence(Protein/DNA) from the PC

2.4. Copy and Paste into the BLAST Searching window

2.5. Deselect the Do CD-Search box.

2.6. If you use Protein Sequence don't change the Choose Database setting, because the nr (for non redundant) is the default protein database

2.7. Click the BLAST! Button

2.8. Click the Format button

2.9. When the results page appear, scroll down the page until you reach along list of sequences and save all these sequences into your PC

## 3.Preparing your Multiple Sequence Alignment

To compute your multiple sequence alignment, any of the following can be used

- ClustalW : [www.ebi.ac.uk/clustalW/Index.html](http://www.ebi.ac.uk/clustalW/Index.html)
- Dialign : [www.bibiserv.techfak.uni-bielefeld.de/dialing](http://www.bibiserv.techfak.uni-bielefeld.de/dialing)
- Tcoffee : [www.igs-server.cnrs-mrs.fr/Tcoffee](http://www.igs-server.cnrs-mrs.fr/Tcoffee)

Steps to produce multiple sequence using EBI ClustalW server

3.1. Open the link at the EBI ClustalW server at [www.ebi.ac.uk/clustalW/Index.html](http://www.ebi.ac.uk/clustalW/Index.html)

3.2. Paste the collected sequences from the sequence window

3.3. Use Output format pull-down menu to set the selection of choice

3.4 Choose Input from the Output Order pull-down menu. When output order set to input, Clustal W outputs your sequence in their original order. If the output order is set to aligned, the sequence appears in the order of they are aligned.

3.5. Click the Run Button at the bottom of page.

3.6. Save your results. Results come in three sections. First one is Pairwise section, second one is the multiple Alignment and third one is guide tree.

## 4.Computing the Tree.

Use the following methods to construct the Phylogenetic Tree

- ClustalW
- Phylip

Steps involved in a phylogenetic tree using online EBI ClustalW server

4.1. Open the link at the EBI ClustalW server at [www.ebi.ac.uk/clustalW/Index.html](http://www.ebi.ac.uk/clustalW/Index.html)

4.2. Paste your Multiple alignment into sequence window .

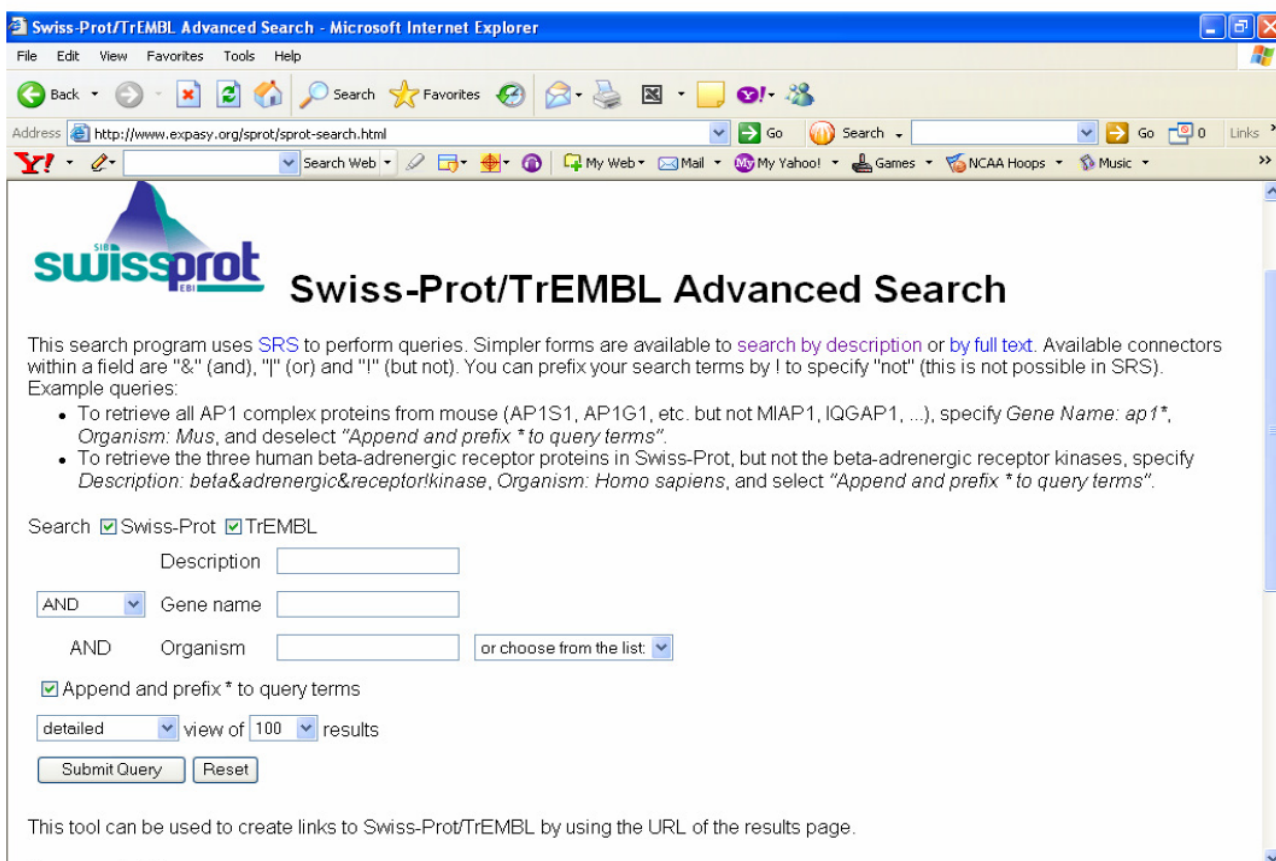
4.3. Choose NJ from Phylogenetic Tree: Tree Type drop down menu

4.4. Choose On from Correct Dist. Drop-down menu.

4.5. Choose On from the Ignore Gaps drop-down menu

4.6. Choose Phylogram from the Tree drop-down menu

4.7. Click the Run Button and view the tree.



**Swiss-Prot/TrEMBL Advanced Search**

This search program uses [SRS](#) to perform queries. Simpler forms are available to [search by description](#) or [by full text](#). Available connectors within a field are "&" (and), "|" (or) and "!" (but not). You can prefix your search terms by ! to specify "not" (this is not possible in SRS). Example queries:

- To retrieve all AP1 complex proteins from mouse (AP1S1, AP1G1, etc. but not MIAP1, IQGAP1, ...), specify *Gene Name: ap1\**, *Organism: Mus*, and deselect "Append and prefix \* to query terms".
- To retrieve the three human beta-adrenergic receptor proteins in Swiss-Prot, but not the beta-adrenergic receptor kinases, specify *Description: beta&adrenergic&receptor!kinase*, *Organism: Homo sapiens*, and select "Append and prefix \* to query terms".

Search ☒ Swiss-Prot ☒ TrEMBL

Description

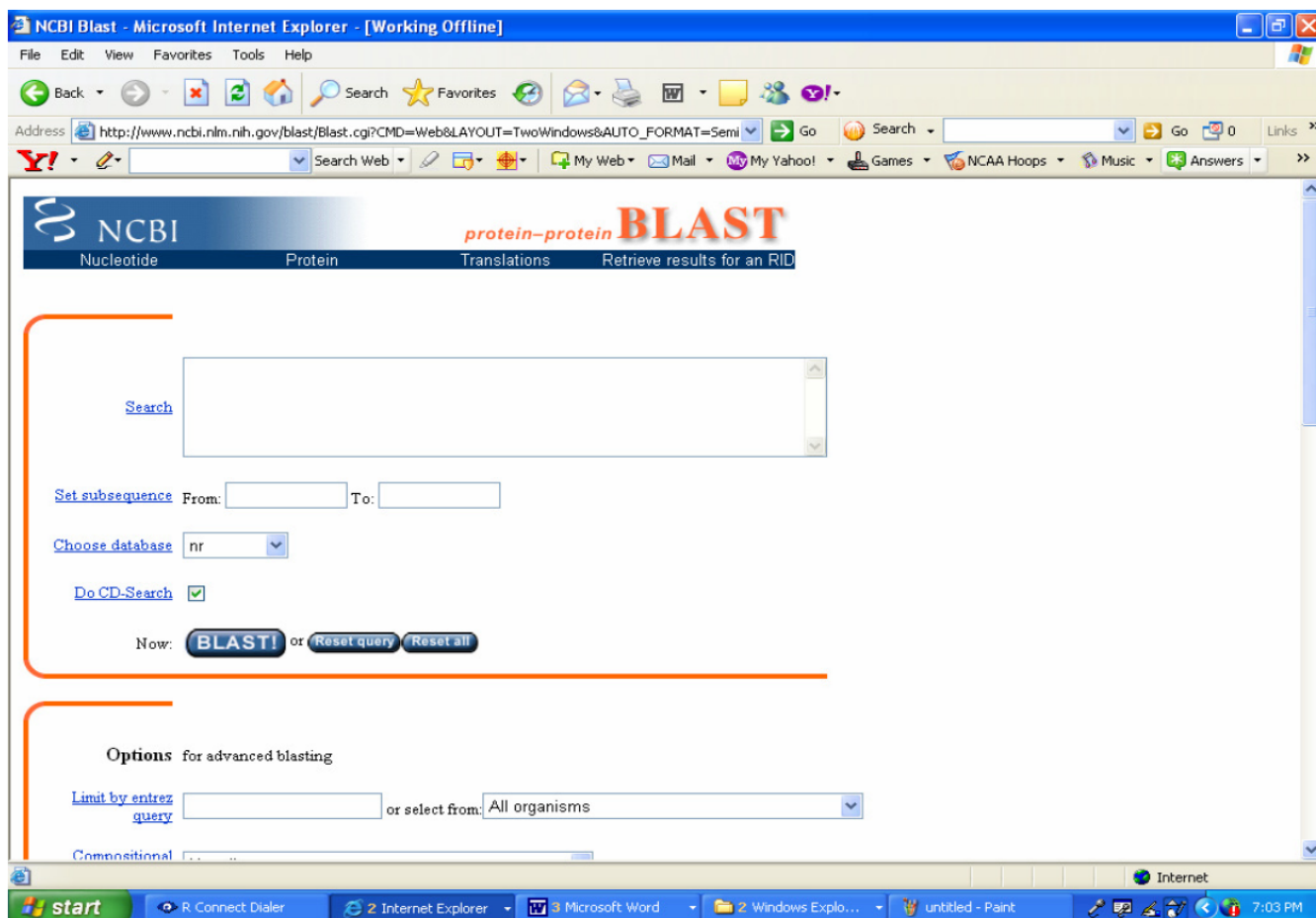
AND  Gene name

AND  Organism  or choose from the list:

☒ Append and prefix \* to query terms

detailed  view of 100  results

This tool can be used to create links to Swiss-Prot/TrEMBL by using the URL of the results page.



**NCBI protein-protein BLAST**

Nucleotide Protein Translations Retrieve results for an RID

[Search](#)

[Set subsequence](#) From:  To:

[Choose database](#) nr

[Do CD-Search](#) ☒

Now:  or

**Options** for advanced blasting

[Limit by entrez query](#)  or select from: All organisms

[Compositional](#)

NCBI Blast - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

NCBI **formatting BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Your request has been successfully submitted and put into the Blast Queue.

Query = sp|P06276|CHLE\_HUMAN Cholinesterase precursor (EC 3.1.1.8) (Acylcholine acylhydrolase) (Choline esterase II) (Butyrylcholine esterase) (Pseudocholinesterase) - Homo sapiens (Human). (602 letters)

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



The request ID is 1143553837-2915-35719900485.BLASTQ4

**Format!** or **Reset all**

The results are estimated to be ready in 12 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of a different search by entering any other valid request ID to see other recent jobs.

---

RID=1143553837-2915-35719900485.BLASTQ4, sp|P06276|CHLE\_HUMAN Cholinesterase precursor (EC 3.1.1.8) - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

Sequences producing significant alignments:

	Score (Bits)	E Value	
<a href="#">gi 4557351 ref NP_000046.1</a> butyrylcholinesterase precursor [...]	1226	0.0	<b>G</b>
<a href="#">gi 55730793 emb CAH92116.1</a> hypothetical protein [Pongo pygmaeus]	1187	0.0	
<a href="#">gi 7381418 gb AAF61480.1</a> butyrylcholinesterase [Equus caballus]	1084	0.0	
<a href="#">gi 1370277 emb CAA36308.1</a> butyrylcholinesterase [Oryctolagus...]	1079	0.0	
<a href="#">gi 74003707 ref XP_545267.2</a> PREDICTED: similar to Cholineste...	1058	0.0	<b>G</b>
<a href="#">gi 34810862 pdb 1POQ A</a> Chain A, Crystal Structure Of Soman-Ag...	1056	0.0	<b>S</b>
<a href="#">gi 57163907 ref NP_001009364.1</a> butyrylcholinesterase [Felis ...]	1055	0.0	<b>G</b>
<a href="#">gi 21362409 sp P81908 CHLE_HORSE</a> Cholinesterase (Acylcholine ...)	1051	0.0	
<a href="#">gi 2981243 gb AAC06262.1</a> butyrylcholinesterase precursor [Pa...]	1051	0.0	
<a href="#">gi 55621450 ref XP_516857.1</a> PREDICTED: butyrylcholinesterase [P...	1013	0.0	<b>G</b>
<a href="#">gi 26341068 dbj BAC34196.1</a> unnamed protein product [Mus muscu...	989	0.0	
<a href="#">gi 71896665 ref NP_033868.2</a> butyrylcholinesterase [Mus muscu...]	988	0.0	<b>G</b>
<a href="#">gi 4167981 sp Q03311 CHLE_MOUSE</a> Cholinesterase precursor (Acyl...)	985	0.0	<b>G</b>
<a href="#">gi 12621110 ref NP_075231.1</a> butyrylcholinesterase [Rattus no...]	981	0.0	<b>G</b>
<a href="#">gi 76677572 ref XP_614444.2</a> PREDICTED: similar to Cholineste...	917	0.0	<b>G</b>
<a href="#">gi 13940252 emb CAC37792.1</a> butyrylcholinesterase [Gallus gal...]	882	0.0	<b>G</b>
<a href="#">gi 7363201 emb CAA27169.1</a> acetylcholinesterase [Torpedo californ...	668	0.0	
<a href="#">gi 2248951 prf I203373A</a> acetylcholinesterase	668	0.0	
<a href="#">gi 64415 emb CAA29047.1</a> unnamed protein product [Torpedo marmor...	667	0.0	
<a href="#">gi 26130361 gb AAB86606.1</a> acetylcholinesterase catalytic subu...	658	0.0	
<a href="#">gi 58011282 gb AAW62516.1</a> acetylcholinesterase precursor [Tetra...	657	0.0	
<a href="#">gi 18858245 ref NP_571921.1</a> acetylcholinesterase [Danio rerio...]	653	0.0	<b>G</b>
<a href="#">gi 34978375 sp Q92035 ACES_BUNFA</a> Acetylcholinesterase precursor	646	0.0	
<a href="#">gi 25282401 ref NP_742006.1</a> acetylcholinesterase [Rattus nor...]	632	1e-179	<b>G</b>
<a href="#">gi 28279461 gb AAH46327.1</a> Acetylcholinesterase [Mus musculus...]	632	2e-179	<b>G</b>
<a href="#">gi 38569250 gb AAH24295.1</a> acetylcholinesterase T-form [Macaca m...]	631	2e-179	
<a href="#">gi 63101489 gb AAH94521.1</a> Acetylcholinesterase [Rattus norvegic...	631	3e-179	<b>G</b>
<a href="#">gi 45572391 ref NP_000656.1</a> acetylcholinesterase isoform E4-E...	630	4e-179	<b>G</b>
<a href="#">gi 71725857 gb AAZ39053.1</a> acetylcholinesterase synaptic isoform	630	4e-179	



ClustalW - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ebi.ac.uk/clustalw/index.html?>

Search Web Search Go

My Web Mail My Yahoo! Games NCAA Hoops Music

References

- ClustalW Help
- ClustalW FAQ
- Jalview Help
- Scores Table
- Alignment
- Guide Tree
- Colours

>> Download Software

YOUR EMAIL	ALIGNMENT TITLE	RESULTS	ALIGNMENT	CPU MODE
<input type="text"/>	Sequence	interactive	fast	single
KTUP (WORD SIZE)	WINDOW LENGTH	SCORE TYPE	TOPDIAG	PAIRGAP
def	def	percent	def	def
MATRIX	GAP OPEN	END GAPS	GAP EXTENSION	GAP DISTANCES
def	def	def	def	def

OUTPUT		PHYLOGENETIC TREE		
OUTPUT FORMAT	OUTPUT ORDER	TREE TYPE	CORRECT DIST.	IGNORE GAPS
aln w/numbers	input	none	off	off

Enter or Paste a set of Sequences in any supported format: [Help](#)

```

Sbjct 2 IILDNRV---TPLSIE--
EDIDIEYNLRPTQLEEYIGQSKVREKLRIKAAKNRGES 55

Query 357 F---LVYGA PGFSKDNNSIITRKEPQGLKI
384
          L+YG PG K + I KE + LKI
Sbjct 56 LDHVLlyGPPGLGKTTLANIIAKEMKGNLKI
86
    
```

start R. Connect Dialer 4 Microsoft Word 2 Windows Expl... untitled - Paint ClustalW - Micros... 7:48 PM

ClustalW - Opera

Opera Widgets

File Edit View Bookmarks Tools Help

Open Save Print Find Home Panels Tile Cascade Voice

New tab UniP... UniP... ExP... UniP... UniP... C... http... .NET... Tran... O Gen... UniP... ExP... file:/...

<http://www.ebi.ac.uk/cgi-bin/clustalw/result?tool=clustalw&jobid=clustalw-20060410-07562988&poll=yes>

Google

Help

- General Help
- Formats
- Gaps
- Matrix
- References
- ClustalW Help
- ClustalW FAQ
- Jalview Help
- Scores Table
- Alignment
- Guide Tree
- Colours

**ClustalW Results**

Results of search	
Number of sequences	18
Alignment score	406137
Sequence format	Pearson
Sequence type	aa
ClustalW version	1.83
Jalview	Invalid bytecode.
Output file	<a href="#">clustalw-20060410-07562988.output</a>
Alignment file	<a href="#">clustalw-20060410-07562988.aln</a>
Guide tree file	<a href="#">clustalw-20060410-07562988.dnd</a>
Your input file	<a href="#">clustalw-20060410-07562988.input</a>

[View Scores Table](#) [View Guide Tree](#) [SUBMIT ANOTHER JOB](#)

To save a result file right-click the file link in the above table and choose "Save Target As".  
If you cannot see the Jalview button, reload the page and check your browser settings to enable Java Applets.

Alignment

start EMBnet\_blastp\_8che... 2 Windows Explorer ClustalW - Opera Adobe Reader 12:19 PM

ClustalW - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.ebi.ac.uk/clustalw/index.html?>

Search Web

def def percent def def

MATRIX GAP OPEN END GAPS GAP EXTENSION GAP DISTANCES

def def def def def

OUTPUT PHYLOGENETIC TREE

OUTPUT FORMAT OUTPUT ORDER TREE TYPE CORRECT DIST. IGNORE GAPS

aln w/numbers aligned nj on on

Enter or Paste a set of Sequences in any supported format

Help

```

Sbjct 2 IILDNRV----TPLSIE--
EDIDIEYNLRPTQLEEIYIGQSKVREKLRIPIKAANKRGES 55

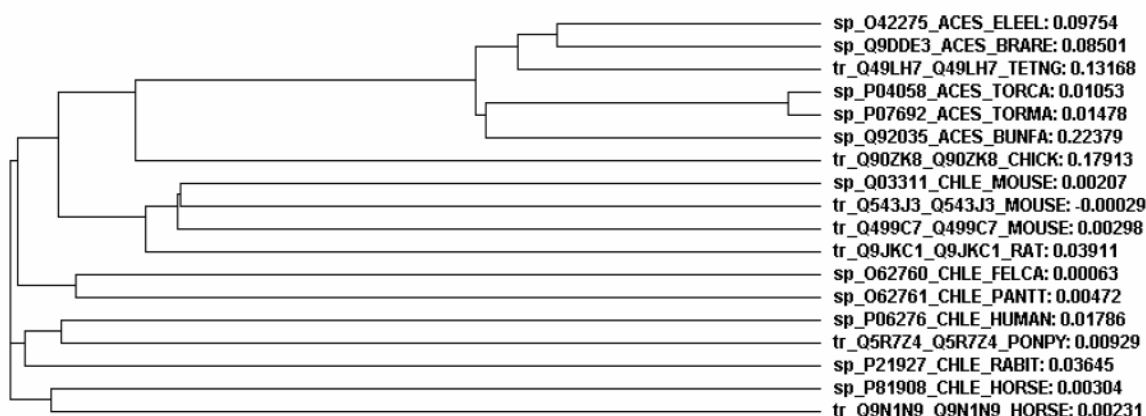
Query 357 F---LVYGAPGFSKDNNSIITRKEFQEGDKI
384
L+YG PG K + I KE + LKI
Sbjct 56 LDHVLLYGPPGLGKTTLANIIAKEMKGNLKI
86
    
```

Upload a file:  Browse... Run Reset

If you plan to use these services during a course please contact us using the email below.  
Please read the [FAQ](#) before seeking help from our support staff.

Page Maintained by [EBI Support](#) Last updated: 03/06/2006 17:19:18

### Cladogram



Show as Phylogram Tree

Hide Distances

View PH File

4.8. Results are saved

## Phylogenetic Tree Construction of Butyrylcholinesterase

### 1.Retrieving Bche Gene Sequence from Swiss-Port.

- a. Point your browser to <http://www.expasy.org/sprot/sprot-search.html>
- b. Type Gene name Bche in Gene name window ,Homo sapiens in Organism window and click the Submit Query button.
- c. <http://www.expasy.org/cgi-bin/get-entries?db=sp&db=tr&DE=&GNc=AND&GN=Bche&OC=Homo+sapiens&wild=1&view=full&num=100>
- d. Select the required gene from given result. For Ex: P06276 and get the FASTA format and save.
- e. <http://www.expasy.org/uniprot/P06276>

### 2.Using BLAST to Compare P06276 Gene Sequence to other Sequences

- a. Point your browser to [www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST) . Click the Standard Protein-Protein BLAST[blastp]
- b. [Paste P06276 FASTA formatted sequence in the sequence window](#)
- c. Press Format Option if your request has successfully executed
- d. Select required sequence and Save into your PC

### 3.Prepare Multiple Sequence Alignment with ClustalW for P06276

- a. Point your browser to the EBI ClustalW server [www.ebi.ac.uk/clustalW/Index.html](http://www.ebi.ac.uk/clustalW/Index.html)
- b. Paste collected sequences in the sequence window
- c. Click the Run Button at the bottom of page.
- d. Save your results.

### 4.Computing your Tree using ClustalW Method for P06276.

- a. Point your browser to the EBI ClustalW server [www.ebi.ac.uk/clustalW/Index.html](http://www.ebi.ac.uk/clustalW/Index.html)
- b. Paste Multiple Sequence alignment into sequence window .
- c. Choose NJ from Phylogenetic Tree: Tree Type drop down menu
- d. Choose On from Correct Dist. Drop-down menu, Choose On from the Ignore Gaps drop-down menu, Choose Phylogram from the Tree drop-down menu and Click the Run Button and view the tree.

## Conclusion

For phylogenetic tree construction, the essential performance parameters have been discussed here in detail to provide guidance to computational biologists. In particular, the stepwise discussion strategy is a very attractive goal, and this methodology, which, if appropriately used, can solve several problems and constitutes a powerful tool in the hands of researchers. The aim of this article, with application to Phylogenetic tree construction of butyrylcholinesterase has been contribute to a further research studies for homology comparison of functional proteins in species.

## Acknowledgment

This work was supported by IIT up gradation grants of AUCE (A).