

ATGC, A New Restriction Site finding Tool

Akash Kumar¹ and Mohd. Zakir Khawaja²

¹Department(s) of Bioinformatics, Uttaranchal College of Science & Technology, Dehradun, INDIA ²Department(s) of Biotechnology,Uttaranchal College of Science & Technology, Dehradun, INDIA

> **Available online at: www.isca.in, www.isca.me** Received 2nd May 2014, revised 4th June 2014, accepted 15th June 2014

Abstract

Bioinformatics include both the power of biological concept and computational method to solve biological problem. It also bridged biological field with speed and accuracy of computer. Perl language is among major computer languages which are being used to develop bioinformatics software. Finding and locating restriction site in given sequence may help to predict its function and analyse it more clearly. This tool locates the position of searched restriction enzyme in submitted nucleotide sequence which in turn may predict specific location that can be recognized by restriction enzyme. This tool also embeds few more functionalities such as to find ORF, count A,T,G,C and bad content in submitted sequence, convert Dna to protein and vice versa.

Keywords: Restriction Site, Perl.

Introduction

One of the major feature of Bioinformatics is to provide tools and software for biologist to enhance their research work¹. To develop new efficient methodologies in different field of biology and implement those algorithm in terms of computer languages is a major challenge for Bioinformatistics². Perl language is one of the most popular computer languages used by bioinformaticstics to develop their software, its extension like BIOPerl include most advanced feature to ease the work of developer³. This software is edictally designed to find Restriction site in submitted sequence while its other feature include DNA to Protein and Protein to DNA conversion, finding the percentage identity of muggy characters in given sequences with it's A,T,G,C count.

Material and Methods

The list of commonly found restriction site has been downloaded from various online available resource, then parsed and stored in text file for further processing. Input from user is used to locate the position of interested site in given sequence. This feature include the very basic algorithm for locating and printing the adjact position of interested site. The ORF finder feature extract the user input and produce six reading frame where the first three reading frame was in forward direction and another three reading frame was in reverse direction⁴. To produce six reading frames the user input was cleaved into set of the three nucleotides from each position of both of the strand and the corresponding amino acid was returned to user output⁵⁻⁶. The DNA to Protein conversion tool convert nucleotide input to protein where the user input is first converted to triplet codon and then the corresponding amino acid is returned. The Start codon is

denoted by symbol star (*) and stop codon is denoted by underscore (_). Each output come with its own seprate window where save as option is provided to store the result for future references.

Results and Discussion

The ATGC is a lightweight software which can be installed in personnel computer to support research work. The front end contain text area for user input whereas the menu bar contain four options. The first option named file menu include four important suboption that is New, Open, Save as and Exit. New option provide fresh text area for input. Open allow user to input via file whereas saveas option allow user to save input in text file, click on exit will close complete application. The other very important features of software is inside Tools option of menu bar which have further five sub option named, ATGC Count, ORF find, Restriction site found and conversion tool. The next option on menubar is Help menu which give short introduction of each section present inside Tools menu. The last option named about us contain thanks and regard and short introduction to author. The (figure-1) shows the input window example which contains a sequence used to demonstrate different feature of software. The (figure-2) shows the ATGC count badcontent output. The (figure-3) shows the six open reading frame of input sequence and (figure-5) shows the restriction site input window and output window which include the position of searched site. Dna to protein converted output is shown in (figure-6) whereas the (figure-7) shows the output window of protein to DNA conversion tool.



Figure-1 Main Input Window



Figure-2 ATGC Count Window

76 My ORF Finder	
Save	 ▼
Frame one KKKKIFFFFKKKKKKI	
Frame Two KKKKFFFFFLKKKKKK	
Frame Three KKKNFFFFF_KKKKKN	
Frame Four YFFFFFFKKKKKNFFFF 	
Frame Five IFFFFFLKKKKIFFF	
Frame Six FFFFFF_KKKKFFFF	

Figure-3 ORF Output Window



Figure-4 Restriction Site Input Window



Figure-5 Restriction Site Output Window

Res. J. Computer and IT Sci.



Figure-6 Dna To Protein Output Window

76 DNA to Protein	_ _ ×	
Save		
Your Protein to DNA Sequence -:		
CATGCTGCTGCTCATCAT1		
Figure-7		

Figure-7 Protein to Dna Output Window

Conclusion

At present there are very few standalone software is available in market for locating the restriction site position. The user interface is designed using Perl (Tk) and converted to portable executable form, with the help of Perl2Exe software which shows the uniqueness of this software and decrease dependencies to maintain the work flow.

This Software can be downloaded from https://sites.google.com/site/atgcucst

References

- 1. Frishman D., Mironov A., Mewes H.W. and Gelfand M.,Combining diverse evidence for gene recognition in completely sequenced bacterial genomes, *Nucleic Acids Research*, **26**(12), 2941-2947 (1998)
- 2. Fickett J., Recognition of protein coding regions in DNA sequences, *Nucleic Acids Research*, **17**, 5303-5318 (**1982**)
- **3.** Stajich Jason E., David Block, Kris Boulez, Steven E. Brenner, Stephen A. Chervitz, Chris Dagdigian, Georg Fuellen et al, The Bioperl toolkit: Perl modules for the life sciences, *Genome research*, **12(10)**, 1611-1618 (**2002**)
- 4. Salzberg S.L., Delcher A.L., Kasif S. and White O., Microbial gene identification using interpolated Markov models, *Nucleic Acids Research*, **26**(2), 544-548 (**1998**)
- Lukashin A.V. and Borodovsky M., Gene Mark. hmm: new solutions for gene finding, *Nucleic Acids Research*, 26(4), 1107-1115 (1998)
- 6. Dhar¹, Dwivedi Vivek and Mishra Sarad Kumar, ORF investigator: a new ORF finding tool combining Pairwise

Global Gene Alignment, *Research Journal of Recent Sciences*, 1(11), 32-35 (2012)

- 7. Skovgaard M., Jensen L.J., Brunak S., Ussery D. and Krogh A., On the total number of genes and their length distribution in completemicrobial genomes, *Trends in Genetics*, **17(8)**, 425-428 (**2001**)
- 8. Goujon, Mickael, et al., A new bioinformatics analysis tools framework at EMBL–EBI, *Nucleic acids research* 38, suppl 2, W695-W699 (2010)
- Patel Tushar S., Panchal Mayur, Prajapati Reecha, Ladumor Dhara, Kapadiya Jahnvi, Desai Piyusha and Prajapati Ashish, An Analytical Study of Various Frequent Itemset Mining Algorithm, *Res. J. Computer & IT Sci.*, 1(1), 6-9, (2013)
- 10. Bhatt T.K., Computational Studies on Calpain from Plasmodium falciparum, *Research Journal of Recent Sciences*, 1(9), 79-82, September (2012)
- 11. Kumar Akash and Dwivedi Vivek Dhar," Evolutionary Analysis and Motif Discovery in Rhodopsin from Vertebrates", *Int. Res. J. Biological Sci.*, 2(7), 6-11, (2013)