Bioinformation by Biomedical Informatics Publishing Group

open access

www.bioinformation.net

Software

ACUA: A software tool for automated codon usage analysis

Vetrivel Umashankar¹*, Vijayakumar Arunkumar¹, Sudarsanam Dorairaj²

¹Department of Bioinformatics, Vel's College of Science, Chennai, India; ²School of Genomics, Department of Zoology, Loyola College, Chennai, India; Vetrivel Umashankar* - Email: vumashankar@gmail.com; * Corresponding author

received June 28, 2007; revised August 15, 2007; accepted September 28, 2007; published online October 06, 2007

Abstract:

Currently available codon usage analysis tools lack intuitive graphical user interface and are limited to inbuilt calculations. ACUA (Automated Codon Usage Tool) has been developed to perform high throughput sequence analysis aiding statistical profiling of codon usage. The results of ACUA are presented in a spreadsheet with all perquisite codon usage data required for statistical analysis, displayed in a graphical interface. The package is also capable of on-click sequence retrieval from the results interface, and this feature is unique to ACUA.

Keywords: codon usage; ACUA; CAI; RSCU; statistical analysis

Availability: The package is available for non-commercial purposes and can be downloaded from: http://www.bioinsilico.com/acua

Background:

The present work attempts to unify all prerequisite calculations required for codon usage analysis in a common graphical platform, enabling high throughput sequence analysis and statistical profiling. Development of ACUA has been made after comprehensive studies of tools such as CodonW [1], GCUA [2], EMBOSS [3], and Jemboss [4]. ACUA has been developed by integrating calculations like codon usage table, RSCU, CAI, Nc Value, selective positional AT, GC content and its skewness. [1] The package also provides on-click access of the sequence from the result interface, and is unique to ACUA.

Methodology:

ACUA has been developed as a standalone package for codon usage analysis using Visual Basic, PERL and C++ programming languages. A snapshot of ACUA is shown in figure1.

Program Input

ACUA requires FASTA formatted nucleotide sequence(s) in a plain text file as input. Additionally, for CAI calculation, user can select the reference set either from inbuilt codon usage tables or they can provide a customized codon usage table in Emboss .cut format. Furthermore, the user shall select their preferred calculations like CAI, RSCU, Nc value, AT and GC content with their positional preferences with an optional choice for excluding non synonymous and stop codons.

Program Output

ACUA consolidates all results into an MS-Excel File (.xls) with two worksheets. The first worksheet contains CAI value, Nc Value and all the Nucleotide Positional percentages with its respective skewness, whilst the second contains codon usage table and RSCU values facilitating statistical analysis. The user can sort the results within ACUA interface by right click of mouse. Furthermore, ACUA lists complete header information of each sequence in the results, which shall aid Gene ontology classifications. Moreover, user can access their preferred sequence as output by on-click access on result of their interest and these features are present only in ACUA.

Utility:

ACUA has got the potential to serve as comprehensive platform, and also as a part of biologist's essential toolkit to extract all the prerequisite data required for statistical analysis of codon usage.

Caveats and future development:

The present version of ACUA is only compatible for single processor machine. A MPI (Message Passing Interface) version of ACUA is being developed with Rlanguage based inbuilt statistical functions, enabling cluster computing.

ISSN 0973-2063 Bioinformation 2(2): 62-63 (2007)

Bioinformation by Biomedical Informatics Publishing Group

open access

Software

www.bioinformation.net

Acua										- 7
<u>File Edit Option H</u> elp										
ڬ 🗐 🎸 📋 🐰 🚰 📉 🔕 🙆	NT Ana	lysis		•	Open E	xcel				
Input File Input File C:\acua\TestInput.txt POpen Open	A1 T			AT1 Per				CAI	ENC	
C Use My Own Codon Usage Table	100	12212A	51 11 65 4	8 12.405 6 8.868			-0.396 0.171	0.768	44.962 58.325	
○ Use Acua Generated Codon Usage Table ○ Select Codon Usage Table From List	81	2022	55 10 37 8	0 14.065			-0.29	0.553	56.024	-
Acinetobacter calcoaceticus	83	and a	92 11		1. 1. 1. 1. 1. 1.	0.137	1/2	0.547	60.123	
Options For Nucleotide Analysis	88 351	303 2	L10 11 210 28	-			-0.005 -0.148	0.541	60.238 57.653	
AT Percentage AT Skewness GC Percentage GC Skewness	90		99 12 88 11				-0.096	0.547	59.480 46.948	-
Selet AT Pos	426		L17 25	-		0.291	-0.368	0.630	53.420 53.022	-
AT1 _ GC1 _	53	47 1	L15 10	4 8.606	18.847	0.06	0.05	0.555	56.543	_
□ Options For Codon Analysis □ Codon Usage Table □ RSCU	147 51	1636-02	LO1 12 91 14	18.0000.0000000	100000000000000000000000000000000000000	2012/01/01/2020	-0.094 -0.219	0.520	48.529 52.665	-
⊽ Codon Adaptive Index ⊽ ENC	58 119	100	19 3 246 29		0		-0.269 -0.097	0.668	45.578	-
Execution Run Acua	240	1	06 16			-0.042	-0.232	0.572	54.094	- -
Windows NT 5.1 (Build 2600) CPU: Intel Pentium Total Physical Memory: 784,816K Available Physical Memory: 411,760K Total Virtual Memory: 2,097,024K Available Virtual Memory: 2,041,400K	ATGGAATT TGCCCCCT CCTGAACC AGCAGCCC ACTGGGAI AGGCGAC AACCTGAA TGACGGAI CCTGCAGO	TTCTTGG IGCTGGG GCGTGGA GACGGCC CCGCTCC TCCCTGC ACCAGAG GAGCAAC GTGGCCC	AACCACT(CAGGACC) GACCGAG GCCTACC/ AATGATC1 GAATCAT(GGTATGC) AGCATGG CGGAGTG	altus norvegicus AGACAGCAGT TGTGGTACAGI CGTCTTCTACA/ TCCTGCCTCC GCAGCTCAGAA CAGGATAAGG/ ACACCCTCAGAA CAGCCTCAGA AACATCGGCTTC ACACCCTCAAGA	TACTGTGGT GAGTGCTAC(AGATCACCAC ICTCCGCTCC GTGCAGAGG ACCAGCTGAT CTGGAAGTCC AGGCGCTTG0 GTGAGAAAAQ	CCCAAGAAA CACCCTTT CCAGCAGAG GCGCTCTTC GCCTCGCGG ACACCAGA GAGTGCGCC GAGTGCGCC GATTGGGA	AGÍCTGTGG CACCTGCCAGA CTGTCGCCAGA CTGTCGCTAC CTATGGGCC TGCAGGAGG CTATGAGCT CGCTGAGGAC ATTGACCTGG	GGATATCGC AGGAATGCTO ACCCAACGC CGGCCGACTO GAACCTCCO GGACAGGCT GATGAACAAT GTCCATGACA	TA CA GG GG TC FCC AC	

Figure 1: ACUA interface showing the sample output

Acknowledgement:

The authors are grateful to bioinsilico research team for their valuable suggestions and for hosting ACUA on web.

References:

[01] http://codonw.sourceforge.net//culong.html

- [02] J. O. McInerney, *Bioinformatics*, 14:372 (1998) [PMID: 9632833]
- [03] P. Rice, et al., Trends Genetics, 16:276 (2000) [PMID: 10827456]
- [04] T. Carver & A. Bleasby, Bioinformatics, 19:1837 (2003) [PMID: 14512356]

Edited by V. S. Mathura

Citation: Umashankar et al., Bioinformation 2(2): 62-63 (2007) License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.