Sequence analysis
LaTcOm: a web server for visualizing rare codon clusters in coding sequences
Athina Theodosiou, and Vasilis J. Promponas∗
Bioinformatics Research Laboratory, Department of Biological Sciences, University of Cyprus, PO Box 20537, CY 1678, Nicosia, Cyprus
Associate Editor: Martin Bishop

1 INTRODUCTION

transfer RNA within a cell (Ikemura, 1981); however, more recent

work by Dong et al. [1996] has shown that the correlation of codon usage and tRNA level is not perfect.

In principle, the RCC detection process is the identification of codon clusters corresponding to rare tRNA species along mRNAs, as quantified using scales of experimental cellular tRNA levels. A complete dataset of this type is only available for E. coli [Dong et al., 1996], therefore, approaches based on codon usage scales are alternatively used. Recently, two different methods have been developed for identifying RCCs in coding sequences: the %MinMax algorithm and a simplified sliding window version, three RCC detection schemes are implemented: the recently described %MinMax algorithm and a simplified sliding window approach, along with a novel modification of a linear-time algorithm for the detection of maximally scoring subsequences tailored to the RCC detection problem. Among a number of user tunable parameters, several codon-based scales relevant for RCC detection are available, including tRNA abundance values from Escherichia coli and several codon usage tables from a selection of genomes. Furthermore, useful scale transformations may be performed upon user request (e.g. linear, sigmoid). Users may choose to visualize RCC positions within the submitted sequences either with graphical representations or in textual form for further processing.

Availability: LaTcOm is freely available online at the URL http://troodos.biol.ucy.ac.cy/latcom.html.
Contact: vprobon@ucy.ac.cy
Supplementary information: Supplementary data are available at Bioinformatics online.
Received on September 17, 2011; revised on December 13, 2011; accepted on December 16, 2011

ABSTRACT

Summary: We present LaTcOm, a new web tool, which offers several alternative methods for ‘rare codon cluster’ (RCC) identification from a single and simple graphical user interface. In the current version, three RCC detection schemes are implemented: the recently described %MinMax algorithm and a simplified sliding window approach, along with a novel modification of a linear-time algorithm for the detection of maximally scoring subsequences tailored to the RCC detection problem. Among a number of user tunable parameters, several codon-based scales relevant for RCC detection are available, including tRNA abundance values from Escherichia coli and several codon usage tables from a selection of genomes. Furthermore, useful scale transformations may be performed upon user request (e.g. linear, sigmoid). Users may choose to visualize RCC positions within the submitted sequences either with graphical representations or in textual form for further processing.

Availability: LaTcOm is freely available online at the URL http://troodos.biol.ucy.ac.cy/latcom.html.
Contact: vprobon@ucy.ac.cy
Supplementary information: Supplementary data are available at Bioinformatics online.
Received on September 17, 2011; revised on December 13, 2011; accepted on December 16, 2011

1 INTRODUCTION

tRNA abundance is asymmetric (non-uniform) in cells of different organisms and can cause variation in the translation rate for each codon. It has been suggested that rare codons and their clusters are associated with translational pausing [Pierce et al., 1999; Komar and Jaenicke, 1995]. The translational rate can be maximized at mRNA regions with codons read by high cellular levels of cognate tRNA species and minimized at sites corresponding to rare tRNAs (RCCs) [Gouy and Kostrub, 2005]. A critical issue is that the concentration of isoaccepting tRNAs for a set of synonymous codons varies among organisms, tissues and stages of differentiation [Ikemura, 1985]. Moreover, different organisms show specific preferences for codons encoding the same amino acid (codon bias), reflected in the frequency of occurrence of synonymous codons in genomic DNA. In Escherichia coli, it has been shown that the non-random choice of codons is mostly attributable to the availability of transfer RNA within a cell [Ikemura, 1985]; however, more recent
Tunable parameters include the window size or the least RCC length and folding of the polypeptide chain. More specifically, RCCs according to biological evidence may be indicative of codon choices that may interfere with proper heterologous gene expression experiments, LaTcOm results could be used for the analysis of multiple sequence alignments and the study of the mechanics of translation. For example, when optimizing coding sequences for expression in E. coli, it is important to consider the spatial scan method introduced by Huang et al. (1999) to identify RCCs, and we introduce a new window-less RCC identification algorithm. It is worth mentioning that when the LaTcOm web server is being developed, another window-less RCC-detection approach, based on the spatial scan method introduced by Huang et al. (1999), was published (Ponnala, 2008). A detailed comparison of the features offered by different RCC-detection algorithms is available as Supplementary Material (Supplementary Table S1).

We anticipate that the availability of a versatile online tool for RCC identification will enable a number of analyses to be performed. For example, when optimizing coding sequences for heterologous gene expression experiments, LaTcOm results could be indicative of codon choices that may interfere with proper folding of the polypeptide chain. More specifically, RCCs according to the host organism’s tRNA abundance/codon usage may have to be preserved for expressing functional proteins. In addition, LaTcOm may be used to study patterns of translational rate within diverged protein families, or the correlation of translational rate with protein structural and functional features, such as protein disorder, aggregation and co-translational folding. Such applications may trigger extensions of the current methods, as for example for the analysis of multiple sequence alignments and the study of the mechanics of translation. For instance, Tuller et al. (2011) studied the mechanics of translation in the context of protein biogenesis.

ACKNOWLEDGEMENT

The authors wish to thank the anonymous referees for invaluable comments on the manuscript and the LaTcOm functionality. We also thank Professor Walter Ruzzo (University of Washington) and Shane Nep (University of Washington) for help with the MSS source code. We also thank Professor Zoya Ignatchova (University of Potsdam) and Dr Gong Zhang (University of Potsdam) for providing their tRNA scale, Professor Konstantinos Fokianos (University of Cyprus) for useful discussions on the RCC validation procedure and Joanna Kalvari (University of Cyprus) for helping with interfacing MSS with the LaTcOm modules.

Funding: A.G. Leventis Foundation (PhD Scholarship to A.T.); University of Cyprus.

Conflict of interest: none declared.

REFERENCES


Downloaded from https://academic.oup.com/bioinformatics/article-abstract/28/4/591/213190 by guest on 14 March 2019