JetGene – web database and toolkit for analysis of regulatory regions or nucleotide contexts at differently translated transcripts of plants

Sadovskaya N.S.¹, Mustafaev O.N.², Tyurin A.A.¹, Goldenkova-Pavlova I.V.¹

¹Tmiryazev Institute of Plant Physiology, RAS, Moscow, Russia; ²Genetic Resources Institute, ANAS, Baku, Azerbaijan

E-mail: nataliya.sadovskaya@gmail.com, orkhan@bioset.org

Key message. We create web database JetGene that allows to estimate the variation of length, nucleotide composition, codon usage frequency and to study nucleotides surrounding of the start codon. JetGene allows user to compare two samples of mRNA.

Keywords: *in silico* analysis, translation effectivity, comparison of two samples of mRNA, regulatory codes, motives

The paradox of discrepancy between the levels of mRNAs and their protein products in the eukaryotic cells, including plant cells, encountered by researchers at every step of organism development and under influence of different stress conditions at whole organism. Consequently, nowadays researchers direct their careful attention of to the study into fine mechanisms of translation.

It's known that mRNA has some regulatory codes that define the individual mRNA fate in translation. For detection of such translational determinants, researchers apply *in silico* analysis of different mRNA regions. With the aim of finding of regulatory codes in plants mRNA and their correlation with translational efficiency we create web database JetGene (<https://jetgene.bioset.org/>). This internet resource contains cDNA, CDS, 5'-UTR, 3'-UTR sequences of six kingdoms of living organisms, including plants. It is necessary to mention that information about CDS and cDNA is downloaded and updated from Ensembl regularly. JetGene has a wide toolkit that allows make a comparative analysis of sequences, in particular: (i) to estimate the variation of length, nucleotide composition, frequency of codon usage, to analyze GC-content, CpG-islands, and to study nucleotides surrounding of the start codon; (ii) to identify and define statistically significant representation of potential regulatory contexts at mRNA with different translation efficiency. It should be noted that the analysis could be performed both by full-length transcript and by coding/non-coding regions. Sequences that picked by researchers up in the analysis can be extract in fasta-format from JetGene. The graphical representation of results accompanies every step of the study. Such simple detail is greatly facilitated the work of any user.

In addition, beta-version of JetGene (<https://beta.bioset.org>) allows user to compare two samples of researcher's mRNA and to apply omics data for searching and prediction regulatory determinants of translation.

The work was supported by the Russian Science Foundation (project no. 18-14-00026; IVG-P).

JetGene – веб база данных и инструментарий для анализа регуляторных областей или нуклеотидных контекстов у дифференциально транскрируемых транскриптов растений

Садовская Н.С.¹, Мустафаев О.Н.², Тюрин А.А.¹, Голденкова-Павлова И.В.¹

¹Институт физиологии растений им. К.А. Тимирязева РАН, Москва, Россия; ²Институт генетических ресурсов

Национальной академии наук Азербайджана, Баку, Азербайджан

Аннотация. Нами создана веб база данных JetGene, позволяющая оценить вариации длины, нуклеотидного состава, частоты использования кодонов, окружение стартового кодона. Кроме того, JetGene дает возможность сравнить две выборки мРНК.

Ключевые слова: *in silico* анализ, трансляционная эффективность, сравнение двух выборок мРНК, регуляторные коды, мотивы

Парадокс несоответствия уровней мРНК и их белковых продуктов в клетках эукариот, включая растения, наблюдается на всех стадиях развития организма, а также при воздействии на организм различных стрессовых условий. В связи с этим, пристальное внимание исследователей в настоящее время направлено на изучение тонких механизмов трансляции.

Известно, что мРНК имеют регуляторные коды, которые определяют судьбу индивидуальной мРНК в трансляции. Для выявления таких детерминант трансляции применяют *in silico* анализ различных областей мРНК. С целью обнаружения регуляторных кодов в мРНК растений и их взаимосвязи с трансляционной эффективностью нами создана веб база данных JetGene (<https://jetgene.bioset.org/>), которая содержит последовательности cDNA, CDS, 5'-UTR, 3'-UTR для представителей шести основных царств живых организмов, включая растения. Необходимо упомянуть, что информация о CDS и cDNA загружена и регулярно обновляется с сервера Ensembl. JetGene имеет обширный инструментарий, позволяющий сделать детальный сравнительный анализ последовательностей, а именно (i) оценить вариации длины, нуклеотидного состава, частоты использования кодонов, проанализировать GC-состав, CpG-острова, а также изучить окружение стартового кодона трансляции; (ii) выявить и определить статистически значимую представленность потенциальных регуляторных контекстов у мРНК с разной эффективностью трансляции. Следует отметить, что анализ можно проводить как по полноразмерному транскрипту, так и по кодирующей/некодирующей областям. Последовательности, отобранные исследователем в результате анализа, могут быть экстрагированы из JetGene в fasta-формате на любом этапе работы. Графическая интерпретация результатов сопровождает каждый шаг работы и существенно облегчает работу пользователя. Кроме того, бета-версия JetGene (<https://beta.bioset.org>) позволяет сравнить две выборки мРНК пользователя и, таким образом, применить omics данные для поиска и предсказания регуляторных детерминант трансляции. Работа выполнена в рамках гранта РФФ 18-14-00026.