

Применение спектрально-аналитического метода при поиске тандемных повторов в геноме бактерий рода *Staphylococcus*

Ольшевец М. М., Назипова Н. Н., Тетюев Р. К.

om@impb.ru

Пушино, Институт математических проблем биологии РАН

Молекулы ДНК постоянно испытывают на себе действие различных мутационных процессов. Одним из наименее изученных таких явлений считается дупликация фрагментов ДНК, когда один кусок последовательности (образец) копируется несколько раз, при этом начало каждой копии образца точно пристраивается в конец предыдущей копии. Результатом такой тандемной дупликации образца является участок последовательности, который называется в геномике тандемным повтором или сателлитом. Считается, что тандемные повторы довольно часты в геномных последовательностях. Иногда с ними связывают конкретные функции, или характерные свойства, например, такие как гибкость отдельных участков ДНК.

Наиболее интересной является задача нахождения длинных неточных тандемных и разнесенных копий генетических последовательностей. Эти знания необходимы для изучения механизмов регуляции генов, принципов взаимодействия повторяющихся участков с факторами транскрипции и других вопросов.

Для решения этой проблемы авторами применяется подход, основанный на спектрально-аналитическом методе [1]. Сам метод также разрабатывается авторами с коллегами. Получается «профиль» исследуемого фрагмента генома (например GC-состав), затем с использованием спектрально-аналитического метода строится функционал оценки присутствия периодичности в профиле и на конечной стадии происходит проверка наличия повторов в тех участках генома, где значения функционала превосходят определенную пороговую величину.

В данной работе исследовались геномы двух видов болезнетворных бактерий рода *Staphylococcus* — *Staphylococcus aureus* и *Staphylococcus haemolyticus*. Известно, что эти микроорганизмы обладают значительной генетической пластичностью и устойчивы к лечению антибиотиками. Геномы *S. aureus* и *S. haemolyticus* богаты различного типа повторами (до 5% общей длины генома) и есть предположение, что насыщенность последовательности ДНК данных бактерий повторами связана с их устойчивостью к антибиотикам.

В японском Национальном институте технологии и анализа (NITE), где и были секвенированы геномы *S. aureus*, *S. haemolyticus* и некоторых других бактерий, создана база данных DOGAN [2], содержащая бактери-

альные геномы и все известные в них на данный момент тандемные повторы. Нам удалось обнаружить значительно более протяженные неточные тандемные повторы: 768 нуклеотидов для *S. aureus* и 810 нуклеотидов для *S. haemolyticus*.

В рамках единого подхода развиваются методы для поиска сателлитов любой длины и любой точности повторяемости, разнесённых несовершенных повторов, инвертированных повторов.

Сейчас авторами с коллегами проводятся исследования повторяющихся структур в геноме человека.

Работа выполнена при поддержке РФФИ, проекты № 06-01-08039, № 06-07-89303, № 07-01-00564.

Литература

- [1] Дедус Ф. Ф., Куликова Л. И., Махортых С. А., Назипова Н. Н., Панкратов А. Н., Тетуев Р. К. Аналитические методы распознавания повторяющихся структур в геномах // Докл. РАН. — 2006. — Т. 411, № 5. — С. 599–602.
- [2] National Institute of Technology and Evaluation, DOGAN (Database of the Genomes Analyzed at NITE) — www.bio.nite.go.jp/dogan/Top.