

УДК 577.3

А.В.Илатовский (4 курс, каф. ЭФ), М.Г.Петухов, к.ф.-м.н., с.н.с. (ПИЯФ РАН)

## СТАТИСТИЧЕСКИЙ АНАЛИЗ ВНЕГЕННЫХ ОБЪЕКТОВ ПОЛНОГО БАКТЕРИАЛЬНОГО ГЕНОМА

Известно, что ДНК является не только носителем генов, кодирующих аминокислотные последовательности белков, но и содержит так называемые внегенные объекты, которые необходимы для нормальной жизнедеятельности клетки. Простейшим внегенным объектом является GC-состав. Так как участок ДНК с высокой величиной GC-состава образует большее число водородных связей, его труднее расплавить, что влияет на многие функциональные свойства белков, взаимодействующих с ДНК. Так, экспериментально показано, что GC-богатые участки ДНК дестабилизируют комплексы РНК-полимеразы с ДНК, что приводит к снижению экспрессии некоторых генов [1].

Целью настоящей работы является создание методов и компьютерных программ для статистического анализа распределения различных внегенных объектов в полных бактериальных геномах и выявление областей со значимым отклонением их содержания от среднего по всему геному.

В качестве первого приложения разработанного нами метода был выбран статистический анализ распределения GC-состава в геноме бактерии *E. coli* K-12 MG1655. Статистический анализ осуществлялся с помощью созданной нами компьютерной программы, включающей следующие статистические тесты и режимы обработки данных:

- W-тест Шапиро-Вилка;
- U-тест Манна-Уитни;
- T-тест Стьюдента для двух выборок;
- точный тест Фишера;
- режимы сравнения: по среднему и по распределению;
- поиск по полному геному, белок- или РНК-кодирующей частям, некодирующей части или любому их сочетанию.

На основе статистического анализа искусственных последовательностей ДНК с наперед заданными свойствами нами были изучены возможности представленных методов по распознаванию статистически значимых отклонений величины GC-состава и определены минимальные размеры участков ДНК, поддающиеся такому анализу. В результате исследования глобального и локальных распределений GC-пар в полном геноме бактерии *E. coli* K-12 MG1655 нами показано, что:

- глобальное и локальные распределения GC-состава в этом геноме значительно отличаются от нормального распределения;
- существует значительное число локальных областей с пониженным GC-составом, в отличие от GC-богатых участков.

Мы определили как местоположение GC-аномальных областей в геноме *E. coli*, так и гены, которые могут быть затронуты их влиянием.

### ЛИТЕРАТУРА:

1. Pemberton I.K., Muskhelishvili G., Travers A.A., Buckle M. J. Mol. Biol. 2000. 299:859–64.