

**Спектрально-аналитический метод поиска повторов
в геномах и протеомах**

А. Н. Панкратов

*Институт математических проблем биологии РАН (филиал
Института прикладной математики им. М. В. Келдыша РАН)*

Многие проблемы биоинформатики связаны с задачей определения близости структур. Для первичных структур биополимеров мерой близости является схожесть последовательностей при выравнивании. Для пространственных структур понятие близости связано с классификацией структур.

Рассматривается оригинальный спектрально-аналитический метод распознавания протяженных повторов в генетических текстах. Метод основан на представлении символьной последовательности из произвольного алфавита в виде пучка кривых с последующими аппроксимацией и среднеквадратичным сравнением частей этих кривых. Поиск повторов в пространственной структуре белков возможен на основе натурального уравнения кривой, составленной из координат атомов основной цепи белка.

С алгоритмической точки зрения метод основан на спектральном индексировании, или сжатии, последовательностей на основе рядов Фурье с последующим сравнением в пространстве коэффициентов разложения. Эти операции наиболее полно используют набор инструкций процессоров, хорошо векторизуются и распараллеливаются на вычислительных системах с общей памятью.

Метод был успешно применен для нахождения тандемных мегасателлитных повторов в геномах. Базовое представление последовательности в виде пучка кривых позволяет разработать преобразования, соответствующие инвертированным и комплементарным последовательностям. Применение метода для изучения повторов в пространственной структуре белков было продемонстрировано на примере задачи распознавания

Работа выполнена при поддержке РФФИ, проект №15-29-07063.

© Панкратов А. Н., 2018

супервторичной структуры белка альфа-альфа уголок с короткой перетяжкой.

Применение аппроксимативного подхода к определению близости структур позволяет понизить вычислительную сложность задачи поиска повторов, что наиболее актуально при обработке больших данных, а также вводит понятие масштаба, что позволяет выравнивать структуры на свойственном им масштабе.

Список литературы

1. Панкратов А. Н., Пятков М. И., Тетуев Р. К., Назипова Н. Н., Дедус Ф. Ф. Поиск протяженных повторов в геномах на основе спектрально-аналитического метода. Математическая биология и биоинформатика, 2012, Т. 7, № 2, С. 476–492

2. Панкратов А. Н., Тетуев Р. К., Пятков М. И., Тойгильдин В. П., Попова Н. Н. Спектрально-аналитический метод распознавания неточных повторов в символьных последовательностях. Труды Института системного программирования РАН Т. 27. Вып. 6. 2015. С. 335–344.