Поляризационная визуализация структуры генных последовательностей: картирование предельных локальных состояний поляризации Д.А. Зимняков², А.В. Скрипаль¹, М.Г. Инкин1, О.В. Ульянова1 <u>М.С. Лаврухин¹</u>

Д.А. Зимняков², А.В. Скрипаль¹, М.Г. Инкин1, О.В. Ульянова1 <u>М.С. Лаврухин¹</u> 1Саратовский национальный исследовательский государственный университет имени Н.Г. Чернышевского

²Саратовский Государственный Технический университет имени Гагарина Ю.А.



Введение: в начале третего тысячилетия, как никогда остро встал вопрос об быстром и объёмном обнаружении и индификации болезнетворных микроорганизмов, получивших как никогда широкое распространение, вследствие благоприятных условий и высокой степени генетической изменчивости. Сейчас для получения результата секвенирования ДНК или РНК, обычно, используются программные методы на основе частичного и корреляционного анализа нуклеотидов и их групп в исследуемой последовательности.Вместе с тем, для решения подобных задач могут быть применены и новые инструментальные и инструментально-программные (гибридные) подходы, основанные на принципах когерентно-оптического и поляризационного анализа квази-случайных двумерных структур.

Цель работы: дальнейшее развитие методологии когерентнооптической визуализации и анализа генных последовательностей на основе принципов локальных стохастических преобразований состояния поляризации считывающего когерентного пучка синтезированным пространственным модулятором локальных фазовых сдвигов ортогонально поляризованных составляющих считывающего пучка и выделения в дифрагировавшем световом поле областей, соответствующих предельным значениям локальных компонентов вектора Стокса.

Поляризационная визуализация структуры генных последовательностей

Каждому триплету ставится в соответствие субматрица $\{a_{i,o}\}$, размером 2×2; между элементами матрицы и базовыми нуклеотидами установлено соответствие: $a_{0,0} \leftrightarrow A$, $a_{0,1} \leftrightarrow G$, $a_{1,0} \leftrightarrow C$, $a_{1,1} \leftrightarrow T$. Значения элементов субматрицы определяют число соответствующих нуклеотидов в триплете; таким образом, сумма всех элементов субматрицы всегда равна 3. Соответственно, минимальное число нулевых элементов в субматрице равно 1, а максимальное – 3. Матрица фазового экрана D_{i_1} размером $2N \times 2N$ синтезируется в результате последовательного объединения всех N2 субматриц. После дискретных преобразований Фурье, были получены распределения локальных значений логарифма первого (рис.1) и третьего (рис.2) компонента вектора Стокса в частотной плоскости.



Рис. 1. Модельные распределения локальных значений логарифма первого компонента вектора Стокса в частотной плоскости (scale=0.1) для: а – штамма «HuB20»; б – штамма «Ульяновск»; в – штамма «Конго»

Модельные поляризационные изображения генных последовательностей (на примере возбудителя африканской чумы свиней; штаммы «HuB20», «Конго», «Ульяновск»)

В качестве примера рассмотрим результаты компьютерного эксперимента по поляризационной визуализации генной последовательности вируса возбудителя африканской чумы свиней (штаммы «HuB20», «Конго», «Ульяновск»). На рисунке 1 в оттенках серого представлены синтезированные пространственные распределения $lg(s_{lm}^{\circ})$ в приосевой области частотной плоскости (*scale*=0.1) для трех рассматриваемых штаммов (логарифмический масштаб представления выбран вследствие весьма широких интервалов изменений интенсивности синтезированных распределений интенсивности в пределах приосевой области) На рисунке 2 приведены аналогичные распределения компонентов s_{km}^3 , характеризующих вклад левой ($s_{km}^3 = -1$) или правой ($s_{km}^3 = 1$) циркулярной составляющей в локальное состояние поляризации в точке (k,m). Обращает на себя внимание антисимметричный характер этих распределений ($s_{k-50,m-50}^3 = -s_{50-k,50-m}^3$), обусловленный применяемым алгоритмом фазовой модуляции прошедших через синтезированный экран х-и у-составляющих считывающего пучка. Также следует отметить уникальность распределений для каждого анализируемого штамма.



Рис. 2. Модельные распределения локальных значений третьего компонента вектора Стокса в частотной плоскости (scale=scale= 0.1) для: а – штамма «HuB20»; б – штамма «Ульяновск»; в-штамма «Конго»

Мутационные изменения последовательностей нуклеотидов и вариабельность отображений предельных состояний поляризации

Бинарные распределения s3_{кm} для штаммов «HuB20», «Конго» и «Ульяновск» при s3_{кm} = - 0.99 (селекция близких к циркулярной левых локальных состояний поляризации) представлены на рис. 3. При моделировании использовано максимально допустимое значение scale, равное 0.5. Для количественного сопоставления синтезируемых таким образом карт предельных состояний может быть использован коэффициент корреляции.

На рисунке 4 также приведены значения $R3_{H,U}$ (последовательности для штаммов «HuB20» и «Ульяновск») и $R3_{H,K}$ (последовательности для штаммов «HuB20» и «Конго»). Отметим отсутствие перекрытия доверительных интервалов на рисунке при малых числах замещенных триплетов; это свидетельствует о высокой чувствительности коэффициента корреляции бинарных распределений предельных локальных состояний поляризации к незначительным изменениям структуры генных последовательностей.





Рис. 3. «Панорамные» бинарные карты предельных состояний поляризации (левая эллиптическая поляризация с -0.99, 0.5) для: а – штамма «HuB20»; б – штамма «Ульяновск»; в – штамма «Конго»

Заключение

Таким образом, рассмотренный метод поляризационной визуализации структуры последовательностей нуклеотидов в ДНК и РНК биологических объектов с применением бинарного картирования предельных локальных состояний поляризации дифрагировавшего считывающего пучка демонстрирует высокую чувствительность к малым изменениям в структуре последовательностей. Синтезируемая карта предельных состояний при высоких значениях порога дискриминации по сути является уникальным идентифицирующим объектом, единственным образом соответствующим анализируемой последовательности нуклеотидов.

Необходимо отметить, что предлагаемый подход, основанный на четырехградационном матричном кодировании исходных последовательностей нуклеотидов, наложении двух фурье-образов полученной матрицы с определенным образом выбранными фазовыми сдвигами и нелинейных преобразований полученной суперпозиции с последующей бинаризацией и синтезом карт предельных состояний результатов преобразования может быть положен в основу компьютерного алгоритма идентификации объектов со сложной квази-случайной структурой.

Предложенный метод является достаточно гибким и допускает различные модификации в части задания различных алгоритмов модуляции фазы граничного поля, расширяющие функциональные возможности отображения и идентификации структуры генных последовательностей биологических объектов.

Рис. 4. Усредненные коэффициенты корреляции бинарных распределений 3_{km} (выражение (6)) для штамма «HuB20» при случайных замещениях заданного числа N_r триплетов в последовательности нуклеотидов (1, 2) в сравнении со значениями коэффициентов корреляции для пар «HuB20» - «Ульяновск» (3, 5) и «HuB20» - «Конго» (4, 6). (1, 3, 4) – коэффициент дискриминации равен -0.95; (2, 5, 6) – коэффициент дискриминации равен -0.99.

Отметим, что в достаточно широких интервалах изменений параметра отстройки s₃+1 (порядка 1.5 – 2 декад) представленные на рис. 5 зависимости допускают с приемлемой точностью степенные аппроксимации. Аналогичное поведению сложных систем вблизи критических точек (например, в области фазового перехода). В то же время, в отличие от классических критических явлений, в рассматриваемом случае установленное близкое к степенному убывание коэффициентов корреляции по мере приближения s₃_a к минимально возможному значению не является в широком смысле универсальным (показатели α для пар «HuB20»-«Ульяновск» (α≈ α≈ 0.39) и «HuB20»–«Конго» (α≈ 0.75) существенно различаются, т.е. зависят от числа замещенных триплетов. Вопрос о подобном «квази-критическом» поведении синтезированных бинарных распределений предельных состояний поляризации в зависимости от структуры синтезируемых фазовых экранов выходит за пределы данной работы и является объектом дальнейших исследований.



Рис. 5. Коэффициенты корреляции для пар «НиВ20» - «Ульяновск» (1) и «НиВ20» - «Конго» (2) в зависимости от отстройки порога дискриминации s¹ "от минимально возможного значения (-1). Сплошная и пунктирная линии демонстрируют близкий к степенному тренд в поведении зависимостей.

Благодарности. Работа выполнена при поддержке Российского научного фонда (проект № 22-21-00194).