**Введение**

Многие недавние открытия в биологии и медицине стали результатом алгоритмам и технологиям выравнивания последовательностей ДНК. Традиционные методы к сопоставлению последовательностей ДНК основаны либо на схемах глобального сопоставления, либо на эвристических схемах, которые стремятся приблизить алгоритмы глобального сопоставления, обеспечивая при этом более высокую вычислительную производительность. В этом отсчете описывается подход, использующий математическую операцию взаимной корреляции для сравнения двух последовательностей. Это может быть реализовано при помощи БПФ (Быстро преобразование Фурье) для увеличения вычислительной эффективности.

Алгоритм обобщен, примеры приведены. К ним относятся выравнивание последовательностей генов на длинных участках геномной ДНК, обнаружение сходства последовательностей у отдаленно родственных организмов, демонстрация сходства последовательностей при наличии массивных (90%) случайных точечных мутаций, сравнение последовательностей, связанных внутренними перестройками (тандемными повторами) внутри гена, и исследование гибридных белков.

Открытие генов, идентификация генетических аберраций и разработка многих новых медицинских методов лечения – все зависит от сопоставления последовательностей ДНК с ее правильным положением в геноме. Кроме того, информация о выравнивании часто имеет решающие значение для характеристики функции гена. На протяжении десятилетий методы выравнивания Dayhoff, Smith-Waterman, Needleman-Wunsch. Однако улучшения скорости или производительности, такие как алгоритмы FASTA и BLAST, часто приносят в жертву чувствительность и уверенность в качестве соответствия. На сегодняшний день мало используется взаимная корреляция и БПФ для сопоставления последовательностей. Было высказано предположение, что основания ДНК должны быть нанесены на карту комплексной плоскости до взаимной корреляции, хотя последствия этого подхода не были полностью изучены. Здесь мы дополнительно исследуем этот подход к выравниванию последовательностей ДНК, позволяющие применять полную скорость и мощность методов цифровой обработки сигналов, таких как БПФ, к проблеме выравнивания последовательностей.

**Алгоритм**

В первую очередь каждому нуклеотидному основанию присваивается комплексный номер, при этом комплементарные основания имеют противоположные знаки. Например, последовательности ДНК AACGTGT и ACG представлены двумя векторами: и . Затем с помощью взаимной корреляции может быть сформирован новый вектор.

где и представляют собой сравниваемые последовательности, а указывает комплексное сопряжение. Уравнение представляет собой некоторого рода «скользящее скалярное произведение» между векторами, представляющими две последовательности, одна из которых сдвинута на единиц относительно другой, причем находится в диапазоне от отрицательных до положительных значений. Он определяет функцию взаимной корреляции, которая вместе с представлением нуклеотидов AGCT в виде комплексных чисел, представляет нужный алгоритм.

Применение функции взаимной корреляции к выравниванию последовательностей ДНК заключается в следующем: положительный пик в действительной части функции взаимной корреляции указывает на сходство между двумя последовательностями. Отрицательный пик указывает на комплементарное сходство, когда одна последовательность состоит из комплементарных оснований в сравнении со второй последовательностью. При поиске комплементарной ДНК значение одной цепи может быть изменено на обратное перед выполнением взаимной корреляции. Для выравнивания пептидных или белковых последовательностей каждому аминокислотному остатку присваивается один или несколько комплексных номеров, и выполняется взаимная корреляция. Алгоритм был протестирован на языке Python.

Взаимная корреляция может быть выполнена с использованием БПФ, которое предоставляет быстрый процесс со скоростью . Вкратце, стратегия заключается в том, чтобы выполнить БПФ для каждой из двух последовательностей, изменить знак мнимой части одного представления области Фурье одной из последовательностей, умножить две функции области Фурье и преобразовать результат с помощью обратного БПФ.

**Результаты**

Результаты взаимного корреляционного выравнивания гена pyrG Mycobacterium tuberculosis (NCBI GeneID: 885048) длиной 10-kb.