



БАЛТИЙСКИЙ НАУЧНО-ИНЖЕНЕРНЫЙ КОНКУРС 2018

Санкт-Петербург, 5 - 8 февраля 2018

Комбинаторика циркулярных кодов

«Математика»

Сердюков Александр Николаевич, Смоленский Андрей Вадимович (научный руководитель, к.ф.-м.н., ассистент СПбГУ), место выполнения работы: в школе

Процесс перевода нуклеиновой цепочки в белок называется трансляцией и выполняется рибосомой. Удивительно, но в процессе обработки длинной цепи рибосома не сдвигается на одну-две буквы (любой сдвиг полностью нарушал бы построенный белок). На данный момент точный механизм защиты от сдвига неизвестен. Однако, существует несколько гипотез. Одна из них связана с тем, что самые частые тройки аминокислот образуют циркулярный код. Более того, когда анализ последовательностей нуклеотидов в генах стал возможен, оказалось, что у разных групп организмов эти циркулярные коды могут незначительно отличаться (максимум на 1 пару комплементарных "троек"). Циркулярный код - это код, любое слово, составленное из элементов которого и записанное по кругу может быть восстановлено лишь единственным способом. Теоретически, это свойство способно помочь рибосоме за конечное число чтений определить, не потеряла ли она порядок "букв" в тройках. Однако, определение лишь позволяет проверить, является ли данное множество циркулярным кодом, но не объясняет, как они устроены. Есть важные задачи, в которых требуется знать устройство вариативных максимальных самокомплементарных СЗ циркулярных кодов конкретной длины. Изучению и описанию их свойств посвящена данная работа.

В силу "неконструктивности" определения циркулярных кодов работа с ними осложняется сложностью их получения. Однако, интересные с прикладной точки зрения циркулярные коды уже были получены ранее. Было выбрано представление в виде графа. Но классических методы работы с графами было недостаточно. Поэтому, экспериментальная часть работы была связана с написанием собственной программы для исследования графа циркулярных кодов (и графа орбит).

Удалось доказать, что группа D4 является подгруппой внутренних автоморфизмов графа циркулярных кодов, было дано объяснение найденным регулярностям в этой графе. Была исследована структура графа, оказалось, что довольно значимыми его структурными элементами являются орбиты D4. Был найден инвариант, позволяющий определять конкретную орбиту по степеням соседних вершин, однако он оказался слабее, чем это необходимо для описания построения графа из простых структурных элементов (из 27 орбит он однозначно определяет только 20).

На данный момент найдены и проверены специальные свойства регулярности графов. Части из них дана интерпретация на уровне графов, другим - на уровне действия группы и на уровне циркулярных кодов. В дальнейшем, после более детального анализа графа орбит и поиска более сильного инварианта, будет возможно алгоритмически описать построение графа циркулярных кодов, сначала - из структурных единиц (орбит), а затем хотя бы по одному циркулярному коду.

Список литературы:

1. Arquès, Michel - "A complementary circular code in the protein coding genes"
2. Fimmel, Gonzalez - "Circular codes, symmetries and transformations"
3. Michel, Pirillo- "A relation between trinucleotide comma-free codes and trinucleotide circular codes"