

Бинарно-геномные числа и симметрии универсальных правил статистической организации геномных ДНК

Петухов С.В.

Институт машиноведения им. А.А. Благонравова РАН, Москва, Россия

spetoukhov@gmail.com

Ключевые слова: геномные ДНК; биоинформатика; правила вероятностей; многоуровневые тексты; дихотомии; симметрии; фрактальные сети

Мотивация и цель: Создатели квантовой механики П. Йордан и Э. Шредингер указывали на ключевое отличие живых тел от неодушевленных: неодушевленные объекты управляются средним случайным движением их миллионов частиц и движение отдельных частиц не существенно для целого; напротив, в живом организме избранные – генетические – молекулы обладают диктаторским влиянием на весь организм за счет квантового усиления (см. историю «квантовой биологии» [1]). Кроме того, Йордан утверждал, что упущенные наукой законы жизни являются законами вероятностей квантового мира. Целью представляемого авторского исследования является поиск этих упущенных законов жизни путем анализа статистических особенностей информационных последовательностей нуклеотидов в одноклеточных ДНК геномов высших и низших организмов из общедоступного банка генетических данных GenBank.

Методы и алгоритмы: Авторский метод анализа статистических закономерностей последовательности нуклеотидов любой из одноклеточных геномных ДНК заключается в ее бинарном представлении как последовательности нуклеотидов двух видов: пуринов (А и G), обозначаемых символом 0, и пиримидинов (С и Т), обозначаемых символом 1. В докладе этот метод поясняется примером информационной последовательности ДНК хромосомы человека № 1, содержащей около 250 миллионов нуклеотидов. Ее бинарное (пурин-пиримидиновое) представление имеет форму сверхогромного бинарного числа с около 250 миллионами битов (десятичный аналог такого числа доходит до $2^{250000000}$). Эти числа названы докладчиком «бинарно-геномными» (БГ-числами). В них предположительно отражены законы квантовой биофизики и квантовой информатики, генетическая память и пр. Математическое естествознание прежде не работало с системами таких огромных чисел (насколько докладчику известно). Для обнаружения скрытых в БГ-числах статистических закономерностей они анализировались авторским алгоритмом «статистических матрешек» (алгоритм иерархии бинарных статистик), включающим следующие стадии анализа процентного состава БГ-числа: сначала оно представляется как последовательность одиночных символов 0-1-1-0-0-... и в ней подсчитываются %0 и %1; затем оно представляется как последовательность дуплетов 01-11-00-10-... и подсчитываются проценты каждого вида бинарных дуплетов %00, %01, %10, %11; затем аналогично та же бинарная последовательность представляется как последовательность триплетов, тетраплетов, пентаплетов, ..., с подсчетом каждый раз процентов соответствующих видов n-плетов. Тем самым каждое БГ-число представляется как многослойное, каждый слой которого написан алфавитом