

КЛАСТЕРИЗАЦИЯ ОРГАНИЗМОВ ПО ХАРАКТЕРИСТИКАМ СТРОЯ ИХ ДНК

В работе рассмотрены подходы для анализа структуры знаковых последовательностей разной природы. Особо выделен анализ строя цепи, который позволяет исследовать структуру цепи безотносительно к её природе, непосредственно учитывая взаимное расположение элементов.

Даётся краткое описание строя цепи как нового математического объекта – особым образом организованного кортежа на основе данной знаковой последовательности. Определена декомпозиция строя на однородные знаковые цепи. Определены две числовые характеристики строя – средняя удалённость и регулярность элементов цепи.

Приведены вычисленные значения характеристик строя 29 нуклеотидных последовательностей живых организмов от простейших до позвоночных. Будучи упорядоченными по характеристике средней удалённости организмы разделились на три группы – между вышеупомянутыми выделилась группа беспозвоночных.

В дальнейших исследованиях производилась автоматическая кластеризация в двумерном λ -пространстве для следующих пар характеристик: <длина нуклеотидной цепи, средняя удалённость>, <регулярность, средняя удалённость>, <регулярность, нормированная удалённость>. Результаты автоматической кластеризации представлены в табличном и графическом виде и хорошо совпадают с экспертными оценками.