

Кластеризация генов митохондрий грибов и растений по триплетному составу выявляет структурированность ЭТИХ ДАННЫХ

Федотовская В.Д.^{1*}, Шпагина Т.О.¹, Колесникова А.И.¹, Садовский М.Г.^{1,2}

¹ Сибирский федеральный университет, Институт фундаментальной биологии и биотехнологии, Красноярск, Россия

² Институт вычислительного моделирования СО РАН, Красноярск, Россия

* viktorii.fedotovskaia@gmail.com

Ключевые слова: структура, функция, нуклеотидные последовательности, упругие карты, митохондриальные гены

Мотивация и цель: Выявление связи между структурой биологической молекулы (в данном случае последовательности ДНК) и ее функцией – одна из фундаментальных задач биоинформатики, далекая от решения. Исследования, посвященные этой теме, зависят от выбранного материала. Ранее указанная связь исследовалась на геномах хлоропластов [1] и на последовательности РНК генов бактерий [2].

Методы и алгоритмы: В данной работе связь между структурой, функцией и таксономией изучали на примере митохондриальных генов АТФ-синтаз грибов (*atp6*, *atp8* и *atp9*) и растений (*atp1*, *atp4*, *atp6*, *atp8* и *atp9*). Исследовали 223 генома грибов и 85 геномов растений. Из полных митохондриальных геномов были выделены последовательности указанных генов. Каждая последовательность преобразовывалась в частотный словарь триплетов с шагом рамки считывания $t = 1$. Далее на упругих картах рассматривалось распределение генов в 64-мерном пространстве, в котором осями являлись частоты триплетов. Последовательности растений и грибов рассматривались по отдельности, на разных упругих картах.

Результаты: Распределение генов на упругой карте оказалось упорядоченным: выделялись кластеры, точки в которых соответствуют одинаковым генам. Таким образом, на упругой карте с последовательностями генов грибов выделились три кластера, в каждом из которых оказались точки, соответствующие одному и тому же гену. Соответственно, на упругой карте с последовательностями генов растений обнаружилось пять геноспецифичных по составу кластеров. Связи между составом кластеров и таксономией не найдено.

Заключение и доступность: При кластеризации последовательностей генов АТФ-синтаз грибов и растений методом упругих карт доказано преобладание функции над таксономией. Данный генетический материал является специфичным, поэтому мы не можем утверждать, что обнаруженный эффект универсален.

Список литературы

1. Sadovsky M. et al. Genome structure of organelles strongly relates to taxonomy of bearers. *Bioinformatics Biomedical Engineering*. 2015:481-490.
2. Горбань А.Н. Классификация нуклеотидных последовательностей по частотным словарям обнаруживает связь между их структурой и таксономическим положением организмов. *Журн. общ. биологии*. 2003:64(1):65-77.