

Поиск редко встречающихся генов путем сравнительного анализа митохондриальных геномов зеленых растений, красных водорослей и фотосинтезирующих протистов

М.М. Иванов^{1*}, Д.В. Диброва^{1,2}

¹ Факультет биоинженерии и биоинформатики, МГУ им. М.В. Ломоносова, Москва, Россия

² НИИ ФХБ им. А.Н. Белозерского МГУ им. М.В. Ломоносова, Москва, Россия

* ivamikem@gmail.com

Цель: Ключевую роль в возникновении эукариот сыграл эндосимбиоз с предком митохондрий, после которого, согласно современным представлениям, многие гены из митохондриального генома (митогенома) были постепенно либо утрачены, либо перенесены в ядро. Обратный процесс (перенос ядерных генов в митогеном) не описан и в теории должен быть сопряжен с рядом трудностей, поэтому закодированные в митогеномах гены являются интересным объектом для изучения эволюции эукариот. Крайняя степень редукции митогенома характерна для царства Metazoa, а митогеномы растений, грибов и одноклеточных организмов намного более разнообразны. Целью настоящей работы был поиск и филогеномный анализ новых и редких белок-кодирующих генов в геномах митохондрий зеленых растений, красных водорослей и фотосинтезирующих протистов.

Материалы и методы:

1. Создание выборки полных геномов митохондрий

При поиске геномов митохондрий был использован ресурс NCBI Datasets, в котором присутствует раздел Organelles, содержащий полные геномы эукариотических органелл. Из него были загружены 1049 полных геномов, относящихся к зеленым растениям (683 вида), красным водорослям и различным фотосинтезирующим протистам (366 видов).

2. Сравнение белков с профильными HMM для MitoCOGs

Ранее были описаны MitoCOGs (митОКОГи), представляющие собой кластеры ортологичных групп (КОГи), построенные на основании митохондриальных геномов, которые описывают типичные для митохондрий гены (Kannan *et al.* (2014) MitoCOGs: clusters of orthologous genes from mitochondria and implications for the evolution of eukaryotes, *BMC Evolutionary Biology* 14:237). Мы построили профильные HMM для всего набора митОКОГов, кроме 12 наиболее крупных, которые соответствуют белкам дыхательной цепи и полностью описываются прокариотическими КОГаами (всего 131 новых профильных HMM). Для сравнения их с последовательностями белков, закодированными в митохондриальных геномах, был использован программный пакет HMMER v3.4 (<http://hmmerr.org/>).

3. Визуализация данных на таксономических деревьях

Информация о таксономической принадлежности организмов, митохондрии которых мы включили в анализ, получена из базы данных NCBI Taxonomy и представлена в виде скобочной формулы дерева. Для визуализации деревьев была использована программа iTOL v7.

Результаты: Исходя из аннотаций геномов и самостоятельного транслирования последовательностей были получены последовательности 39634 белков, относящихся к рассматриваемым группам эукариот. Из них 28322 пришлось на зеленые растения и 11312 – на протистов и красные водоросли. Для 33570 белков была обнаружена находка хотя бы одной профильной HMM из набора митОКОГов, тогда как для 6064 белков не было найдено ни одной находки (порог E-value при выравнивании $1e-5$). У зеленых растений оказалось 5397 белков, не соответствующих ни одному из митОКОГов, тогда как у остальных организмов – 667. Таким образом, больше всего редких белков оказалось у зеленых растений (около 19% от общего числа).

Выводы: В ходе проведенного анализа было показано, что во всех рассмотренных таксономических группах существует большое количество белков, которые не могут быть отнесены хотя бы к одному часто встречающемуся в митогеномах семейству белков, то есть потенциально являются редкими. Некоторые из них при этом могут быть похожи на другие известные белки прокариот: так, неожиданно в некоторых митохондриях были обнаружены гены белков, похожих на компоненты фотосистем I и II, рубиско и субъединиц цитохромного *b₆f*-комплекса. Большое количество (5850) продуктов редких митохондриальных требуют более детального анализа и сравнения с прокариотическими белками. Отдельного изучения требуют также находки тех митОКОГов, функция которых, согласно базе данных, была неизвестна.

Финансирование: Исследование выполнено в рамках государственного задания МГУ им. М.В. Ломоносова.