

Особенности эволюции С-концевого домена белков нуклеоплазминового семейства***Вяльцев Валерий Владимирович****Студент (специалист)*Московский государственный университет имени М.В.Ломоносова, Факультет
биоинженерии и биоинформатики, Москва, Россия*E-mail: vyaltsevaleriy7@gmail.com*

Белки нуклеоплазминового семейства в основном выполняют роль гистоновых шаперонов в ядре, а некоторые его представители также участвуют в таких процессах, как организация ядрышка, митоз и репарация [1]. Группы белков NPM1, NPM2 и NPM3 в данном семействе выделяются в основном для позвоночных, в то время как белки, принадлежащие беспозвоночным, причисляют к группе NPM-like белков [1]. В белках нуклеоплазминового семейства встречается два домена: N-концевой домен (он есть во всех белках семейства и ответственен за олигомеризацию) и С-концевой домен (взаимодействует с рРНК и важен для организации ядрышка), который, согласно распространенному представлению, встречается в основном в NPM1 позвоночных [1]. Однако при поиске в базе данных UniProt нам удалось обнаружить белки нуклеоплазминового семейства с С-концевым доменом у беспозвоночных, а именно у трихоплакса, некоторых представителей Cnidaria и Echinodermata. Поэтому настоящая работа направлена на исследование представленности С-концевого домена и его эволюции.

Из базы данных UniProt для анализа были собраны с учетом общих представлений о структуре белков нуклеоплазминового семейства 1070 последовательностей, принадлежащих 651 организму. Собранные последовательности принадлежали следующим таксономическим группам: Vertebrata, Echinodermata, Placozoa и Cnidaria, Lophotrochozoa, Insecta и Crustacea. Поиск белков нуклеоплазминового семейства в не представленных в UniProt группах производился с помощью tblastn в базах данных NT (Nucleotide collection), WGS (whole-genome shotgun contigs) и TSA (transcriptome shotgun assembly) преимущественно по С-концевому домену. Для получения белков из последовательности транскрипта использовалась программа ORFfinder и производилась аннотация доменов с помощью InterPro. Таким образом было найдено 15 последовательностей, принадлежащих Xenacoelomorpha, Ctenophora, Porifera, Tunicata, Hemichordata, а также Choanoflagellata. Дополнительно для некоторых найденных белков, у которых с помощью InterPro не получалось аннотировать С-концевой домен, была предсказана структура с помощью AlphaFold [2]. У всех полученных белков были взяты последовательности N-концевых доменов и выравнены с помощью программы MAFFT [3], на основе полученного выравнивания с помощью программы IQ-TREE [4] было построено филогенетическое дерево.

В результате проделанной работы было получено неразрешенное филогенетическое дерево, где единственная нетривиальная ветвь отделяет белки позвоночных от белков беспозвоночных. На нем были также отмечены группы организмов, у которых был найден С-концевой домен. Исходя из полученных данных, мы предполагаем, что С-концевой домен появился ещё у общего предка Choanoflagellata и многоклеточных животных, а общий предок групп Lophotrochozoa и Ecdysozoa, у которых С-концевой домен не был найден, утерял его. Помимо этого, Nematoda, у которых не были найдены нуклеоплазминовые белки вовсе, вероятно, потеряли их в ходе эволюции.

Источники и литература

- 1) Frehlick, L.J., Eirín-López, J.M. and Ausió, J. (2007), New insights into the nucleophosmin/nucleoplasmin family of nuclear chaperones. *Bioessays*, 29: 49-59. <http://doi.org/10.1002/bies.20512>
- 2) Jumper, J., Evans, R., Pritzel, A. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589 (2021). <https://doi.org/10.1038/s41586-021-03819-2>
- 3) Kazutaka Katoh, Daron M. Standley, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability, *Molecular Biology and Evolution*, Volume 30, Issue 4, April 2013, Pages 772–780, <https://doi.org/10.1093/molbev/mst010>
- 4) B.Q. Minh, H.A. Schmidt, O. Chernomor, D. Schrempf, M.D. Woodhams, A. von Haeseler, R. Lanfear (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.*, 37:1530-1534. <https://doi.org/10.1093/molbev/msaa015>