DOI 10.18699/SBB-2022-074

Bacteriophages as vectors of gene transfer from prokaryotes to eukaryotes

Daugavet M.[1]*, Sorokina N.[2], Dikaya V.[1], Gretsova M.[3], Podgornaya O.[1]

[1] *Institute of Cytology, RAS, St. Petersburg, Russia*

[2] *Applied Genomics Laboratory, SCAMT Institute, ITMO, St. Petersburg, Russia*

[3] *Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia*

** ka6tanka@yandex.ru*

Key words: horizontal gene transfer, eukaryotes, bacteriophages

Horizontal gene transfer (HGT) refers to the movement of genetic material between the genomes of unrelated organisms. This process is well documented for bacteria and includes conjugation, transformation and transduction. At the same time, a large amount of data on HGT cases from bacteria to eukaryote genomes has been accumulated. In our previous work, we described the ancient chimeric sequence in the genome of the ascidian *Styela rustica*, probably resulting from HGT. This sequence is coding for protein named rusticalin. According to similarity search rusticalin contain two parts: N-terminal inherited from eukaryotic ancestor and C-terminal strongly similar to bacterial sequence. Aside from this C-terminal part has reliable similarity to the enzyme of bacteriophage A500. Rusticalin homologue in the genome of another ascidian species contain specific nucleotide sequence which is very much alike bacteriophage A500 recombination site. Presence of recombination site in eukaryote genome can explain the integration of foreign genetic material and confirm bacteriophage role in this case of HGT [1].

N-terminal part of rusticalin inherited from eukaryotic ancestor is coding for special amino acid sequence called cysteine-rich repeats. This repeats are widespread among eukaryote proteins and in many cases are neighboring with conserved functional domains typical for bacteria and bacteriophages. In bacterial genomes those domains are frequently found in prophage regions showing their involvement in transduction between different bacteria. We can assume that those domains are repeatedly captured by bacteriophages during transduction process.

Eukaryotic domains with significant similarity (E value < 10^{-4}) to prokaryotic sequences belong to 13 different protein families according to Pfam database [2]. This result led us to the assumption that cysteine-rich repeats are marking the sites in the eukaryotic genomes that are susceptible to HGT. We further hypothesize that cysteine-rich repeats in eukaryotic genomes might contain multiple recombination sites for bacteriophage integration. In order to check this hypothesis we constructed the dataset of proteins with cysteine-rich repeats and dataset of bacteriophage recombination sequences described in the literature. 43 % of all genes containing cysteine-rich repeats harbor short (10–16 bp) sequences of recombination sites. While in control dataset of genes without cysteine-rich repeats, 13 % of genes contain sequences of recombination sites (10–13 bp). We can conclude that genes containing cysteine-rich repeats are enriched with short sequences similar to bacteriophage recombination sites. This may possibly facilitate integrations of foreign genetic material in eukaryote genomes involving bacteriophages as vectors.

Acknowledgements: This work was supported by Russian Foundation for Basic Research (grant No. 203490077), Russian Science Foundation (grant No. 19–74-20102) and Ministry of Science and Higher Education of the Russian Federation (Agreement No. 075-15-2021-1075, signed 28.09.2021).

References

1. Daugavet M.A., Shabelnikov S., Shumeev A., Shaposhnikova T., Adonin L.S., Podgornaya O. Features of a novel protein, rusticalin, from the ascidian Styela rustica reveal ancestral horizontal gene transfer event. *Mob DNA*. 2019;10:4. doi: 10.1186/s13100-019-0146-7.
2. Daugavet M.A., Shabelnikov S.V., Podgornaya O.I. Amino acid sequence associated with bacteriophage recombination site helps to reveal genes potentially acquired through horizontal gene transfer. *BMC Bioinformatics*. 2020;21(Suppl. 12):305. doi: 10.1186/s12859-020-03599-y.