DOI 10.18699/SBB-2022-086

Analysis of environment adaptation features of the bacteria of *Flavobacterium* genus by comparative genomics

Petrushin I.[1, 2], Gutnik D.[2], Pushmina E.[2]*, Belikov S.[1], Chernogor L.[1]

[1] *Limnological Institute, SB RAS, Irkutsk, Russia*

[2] *Irkutsk State University, Irkutsk, Russia*

* *nodreamistoobig@yandex.ru*

Key words: environment adaptation, *Flavobacterium*, polysaccharide utilization loci, comparative genomics

The bacteria in natural ecological niches evolve in a highly competitive environment. Struggling for nutrient resources and defending against pathogens bacteria adapt to the environment or host. As a rule, the more stable habitat conditions are the smaller set of genes the bacterium needs. Representatives of the *Flavobacterium* found in aquatic, bottom sediment, terrestrial, polar and glacial niches, some of them are host associated and known as animal pathogens and plant endophytes. There are some strains discovered in the engineered environment. Like many of the Bacteroidetes they are capable of decomposing a wide range of polysaccharides. For that their genome contains polysaccharide utilization loci (PULs). A number of researchers claim that genome organization and size depend on niche (aquatic or terrestrial one) (Kolton et al., 2013; Wan, 2019). Moreover, PULs composition differs for "aquatic" and "terrestrial" strains. In present work we consider to narrow the habitats for: soil (rhizosphere), host-associated, polar (glacier) and aquatic (including sediments). Besides the metabolism of polysaccharides, we are interested in the impact of other bacterial life support systems. Comparing such data of the amount of the genes involved in biochemical processes, it becomes possible to define the degree of similarity of selected strains. In a previous work a strain of *Flavobacterium* sp. SLB02 was isolated from diseased freshwater sponges of the species *Lubomirskia baicalensi*s (Petrushin et al., 2020). The genome of that bacterium has a size that is characteristic of "terrestrial" strains isolated from an organism living in fresh water. The goal of this work is analysis of environment adaptation signs of the representatives of the family *Flavobacterium* and, in particular, the strain of *Flavobacterium* sp. SLB02 by methods of comparative genomics.

Methods and Algorithms. There are more than 600 available genomes of the genus *Flavobacterium* in the NCBI GenBank database. For the analysis we took a list of strains described in (Kolton et al., 2013; Wan, 2019) and extended the list by representatives from the different niches. Genes of polysaccharide utilization (glycoside hydrolases, glycosyltransferases, polysaccharide lyases, carbohydrate esterases, carbohydrate-binding modules) were detected by dbCAN2 software (Zhang, 2018). Functional annotation of genes was made by prokaryotic genome annotation service RAST SEED. The whole genomes phylogeny was constructed using the PhyloPhlAn software based on 400 conservative protein sequences (Fig. 1). The closest strain is *Flavobacterium* sp. Leaf82, isolated from leaves of *A. thaliana*. Visualization of the strains composition based on a set of features in multidimensional space was performed by different methods: the principal component (PCA), the multidimensional scaling (MDS) and t-distributed stochastic neighbor embedding (t-SNE).

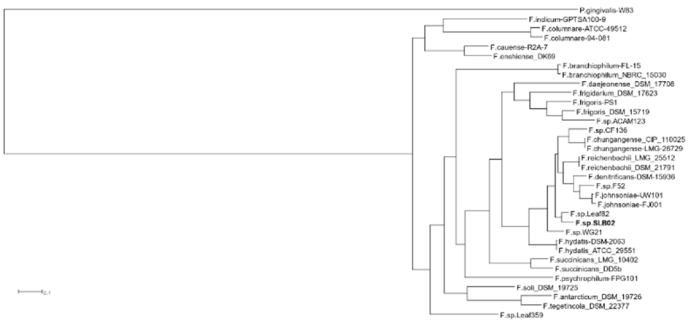

Fig. 1. Phylogenetic tree of the strain SLB02 with closely-related species of *Flavobacterium*. The multilocus tree was built based on a maximum-likelihood method of approximately 400 universal marker genes by PhyloPhlAn with the maximum-likelihood method

Results and Discussion: We performed the analysis of several groups of genes related to environment adaptation of representatives of the *Flavobacterium* genus. A diagram that allows us to estimate the similarity of strains by a set of characteristics was constructed with the help of multidimensional analysis. The most of the analyzed strains belongs to one of two clusters we determined as aquatic/sediments and soil ones. Some of host-associated strains are close to the aquatic ones, but *Flavobacterium* sp. SLB02 is close to soil, rhizosphere and endophyte samples. According to the previous studies, host-associated bacteria have several features: genome reduction, diverse CRISPR-Cas system, secretion systems and higher GC-content than for free-living species. The genome of *Flavobacterium* sp. SLB02 has reduced CRISPR-Cas system (only two spacers), has size typical to terrestrial or soil strains and has no any complete secretion systems. To study to the possible origin of *Flavobacterium* sp. SLB02 we extracted the sequences of 16S regions from the metagenomic sequencing data from water samples of Lake Baikal and performed the phylogenetic analysis, but no close related sequences were found. The applied technique can be used to choose the strains with similar genomic features, associated with habitat conditions. It can help to overcome the difficulty of defining the habitats of the bacteria with unknown origin. The aim of further work is a detailed analysis of the genomic features of adaptation to the niches or host (for the host-associated strains).

References

1. Kolton M., Sela N., Elad Y., Cytryn E. Comparative Genomic Analysis Indicates that Niche Adaptation of Terrestrial Flavobacteria Is Strongly Linked to Plant Glycan Metabolism. *PLoS One*. 2013;8(9):e76704.
2. Petrushin I., Belikov S., Chernogor L. Cooperative Interaction of Janthinobacterium sp. SLB01 and Flavobacterium sp. SLB02 in the Diseased Sponge *Lubomirskia* baicalensis. *Int J Mol Sci*. 2020;21:8128.
3. Wan X. Comparative genome analyses reveal the genomic traits and host plant adaptations of flavobacterium akiainvivens IK-1t. *Int J Mol Sci*. 2019;20(19):4910.
4. Zhang H., Yohe T., Huang L., Entwistle S., Wu P., Yang Z., Busk P., Xu W., Yin W. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res*. 2018;46(W1):W95-W101.
5. Brettin T., Davis J.J., Disz T. et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci Rep.* 2015;5:8365.