Bioniformatics reserch of CRISPR/Cas-systems in the genomes of the phytopatogenic strain *Agrobacterium fabrum* C58

I.A. Portnaia[1]*, A.Yu. Borisenko[1], Yu.P. Dzhioev[1], O.G. Karnoukhova[1], Ju.A. Markova[2], L.A. Stepanenko[1], O.N. Reva[3], N.P. Peretolchina[1], V.I. Zlobin[1]

[1] *Irkutsk State Medical University, Irkutsk, Russia*
[2] *Siberian Institute of Plant Physiology and Biochemistry, Irkutsk, Russia*
[3] *Centre for Bioinformatics and Computational Biology, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa*

DOI 10.18699/ICG-PlantGen2019-59

* e-mail: portnaya.yana.1997@yandex.ru

Abstract: The phytopathogenic bacteria of the species *Agrobacterium fabrum*, which is one of the eleven species of the *Agrobacterium tumefaciens complex*, have become important plant pathogens in the recent years. They penetrate the crown, roots and stems of plants through wounds, causing tumors in the roots. They are also optional pathogens for humans and animals. Various chemical pesticides and biological preparations are used to combat these pathogens. However, due to their environmental hazards and multiple resistance to them, it becomes necessary to look for new methods and approaches to combat them. The research into bacterial CRISPR/Cas-systems has become one of the breakthrough trends in biology, agriculture and medicine today. Due to the large number of decoded bacterial genomes, as well as computer and software technologies, it becomes possible to conduct model studies on the search for and analysis of the diversity of all forms of bacterial activity. This study presents the results of a bioinformatics search for and analysis of loci and structures of CRISPR/Cas-systems in the genome of *Agrobacterium fabrum* str. C58 presented in the GenBank database. Here are presented some software and an algorithm for searching for CRISPR/Cas-systems and screening bacteriophages identical to the spacer sequences of CRISPR-arrays. The results were obtained on the structure of their CRISPR-arrays, the diversity of CAS-proteins and phage strains detected through spacer sequences.
Key words: phytopathogenic bacteria; bioinformatics analysis; CRISPR/CAS-systems; CRISPR–arrays; cas-protein; phage strains.

1. Introduction

Bacteria that cause plant diseases are called phytopathogenic. They have different degrees of pathogenicity and belong to different genera: *Erwinia*, *Pseudomonas*, *Xanthomonas*, *Agrobacterium*, *Pectobacterium*, *Rhizobium*, and others. Plant diseases caused by bacteria are called bacterioses, which are divided into 3 groups: general (vascular), local (parenchymal) or tumors. With a general lesion, the pathogen penetrates into the vascular system of the roots, the disease is accompanied by wilting of leaves and stems, and leads to the death of the plant. A typical example of vascular bacteriosis is potato ring rot. Tumor formations on plants are cancer and tuberculosis. In the case of cancer tumors, the growth of tissue is observed; in the case of tuberculosis, cavities are filled with bacterial mucus and then are formed in the expanding tissue. The virulence factors of phytopathogenic bacteria are toxins and enzymes. Toxins interact with plant cell enzymes, inactivate them, causing the cells to die. Under the action of a number of enzymes (pectolytic, proteolytic, cellulolytic), the plant cell wall substances are split and, as a result, the pathogens freely enter cells and destroy them (Bolotin et al., 2005).

Various chemical pesticides and biological preparations are used to combat these pathogens. However, due to the environmental hazards of these drugs and the growing resistance of bacteria to them, it becomes necessary to look for new methods and approaches to overcome them. The research of the CRISPR/Cas-systems of bacteria is one of the breakthrough trends in biology, agriculture, and medicine which effectively use it nowadays. The CRISPR/Cas-system (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated proteins, or short palindromic repeats, regularly arranged by groups with CRISPR-associated proteins) is a specific adaptive protective system of prokaryotes against foreign genetic material (Gasiunas et al., 2014). CRISPR-arrays are a set of short palindromic repeats of 21–47 nucleotide pairs (b.p.) separated by unique spacer sites. Spacers complementally correspond to the gene segments of bacteriophages and plasmids to which the bacterium demonstrates resistance and carry information about the meetings of the bacterium with them during the evolutionary process (Bolotin et al., 2005). There are cas-genes nearby, whose products ensure the functioning of the CRISPR loci. There are 3 types of CRISPR/Cas-systems, differing in cas-genes and the mechanism of action of the system (Kloepper et al., 1992; Makarova et al., 2006).

In recent years, the genes and genomes of many bacterial species have been decoded. Bioinformatics methods are widely used for processing the information related to a DNA. They allow the genomes to detect and to determine the structure of nucleotide repeats of CRISPR/Cas-systems (Makarova et al., 2011). A screening of spacers using these methods makes it possible to determine the degree of bacterial resistance to specific phages and plasmids. Research in this direction is relevant both for the study of intraspecific and interspecific evolutionary processes, and for solving practical problems in the treatment of human and plant infectious dis-

eases (Abedon et al., 2011). The aim of this work is to search for and to analyze loci and structures of CRISPR/Cas-systems in the genome of *Agrobacterium fabrum* str. *C58* present in the GenBank database through the developed bioinformatics algorithm of the programs, as well as screening phage strains through CRISPR-arrays.

2. Materials and methods

The object of the study is strain 58 of *Agrobacterium fabrum* (included in the *Agrobacterium tumefaciens complex*), present in the GenBank database (No. NC_003062.2). To search for CRISPR/Cas-systems, the methods of systems modeling MacSyFinder (Macromolecular System Finder, ver. 1.0.2) (Abby et al., 2014) were used. The search for structural and functional characteristics of cas-genes was carried out using auxiliary software packages makeblastdb (ver. 2.2.28) and HMMER (ver.3.0) (Biswas et al., 2014). To search for CRISPR-arrays in the genome of the strain, five bioinformatic software search algorithms were used:

1) PILER-CR: CRISPR repeats (http://www.drive5.com/pilercr),
2) CRISPI: a CRISPR Interactive database (http://crispi.genouest),
3) CRISPRFinder (http://crispr.u-psud.fr/Server),
4) CRT: CRISPR recognition tool (http://www.room220.com/crt),
5) CRISPRDetect(http://brownlabtools.Otago.ac.nz/CRISPRDetect/predict_crispr_array.html).

Phage detection through the decoded spacers for the identified CRISPR-arrays was performed using BLASTn programs from the GenBank-Phage database, where the following programs were used: CRISPRTarget (http://bioanalysis.Otago.ac.nz/CRISPRTarget/crispr_analysis.html), Mycobacteriophage Database (http: //phagesdb.org/blast) and Phages database (http: //www.phantome. org /PhageSeed / Phage.cgi).

3. Results and discussion

To sum up, the result of our analysis is identification of cas3 (class I) and cas4 (class I-II) genes and their structural and functional characterization. Based on program matches for each site, two CRISPR-arrays were detected in the genome of *A. fabrum* str. C58. One CRISPR-array consists of three spacer sequences ranging in size from 28 to 44 nucleotide bases (b.p.) and separated by four repeats each 9 b.p. in length (CCTCCTCCC). The other CRISPR-array consists of two spacers ranging in size from 15 to 27 b.p. and separated by three repeats each 9 b.p. in length (TATCGCCAT). Using the structures of the spacers in the identified CRISPR-array, the phage strains identified, which are likely infected the *A. fabrum* str. C58 strain during its evolution. The known phage strains were identified as representatives of the bacterial genera *Mycobacterium*, *Streptomyces*, *Gordonia*, and *Arthrobacter*.

4. Conclusions

Thus, the developed bioinformatical software algorithm used in our work makes it possible to search for loci and describe the structures of the CRISPR/Cas-system of bacteria, and also makes it possible to assess the degree of their resistance to phages and plasmids. Using A. fabrum str. C58 strain as an example, it was shown that it has two CRISPR-arrays and cas-genes in its genome and that this strain is highly resistant to various types of alien phages.

Also, the number of spacers and the degree of their identity to the phage protospacers indicate the level of their impact on the strain during evolution. Therefore, the development and selection of high-quality software methods and their algorithmic constructions make it possible to identify the specific features of the studied strain that could not be noticed by other methods. The obtained information, in a long term, will allow selecting target phages for carrying out strain-specific phage therapy of plant diseases caused by phytopathogens.

References

Abby S.S., Néron B., Ménager H., Touchon M., Rocha E.P.C. Mac SyFinder: A Program to Mine Genomes for Molecular Systems with an Application to CRISPR-Cas Systems. Torres N.V. (Ed.). *PLoS ONE*. 2014;9(10): e110726. DOI 10.1371/journal.pone.0110726.

Abedon S.T., Kuhl S.J., Blasdel B.G., Kutter E.M. Phage treatment of human infections. *Bacteriophage*. 2011;1(2):66–85. DOI 10.4161/bact.1.2. 15845.

Biswas A., Gagnon J.N., Brouns S.J.J., Fineran P.C., Brown C.M. CRISPR Target: Bioinformatic prediction and analysis of crRNA targets. *RNA Biol*. 2013;10(5):817–827. DOI 10.4161/rna.24046.

Bolotin A., Quinquis B., Sorokin A., Ehrlich S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiol.* 2005;151:2551–2561. DOI 10.1099/mic.0. 28048-0.

Borisenko A.Yu., Dzhioev Yu.P., Paramonov A.I. The use bioinformatics methods for researching of CRISPR/Cas systems in the genome of Staphylococcus aureus. *Sibirskij medicinskij zhurnal* (*Irkutsk*). 2015;2:71–74.

Gasiunas G., Sinkunas T., Siksnys V. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular Mol Life Sci*. 2014;71(3):449–465. DOI 10.1007/s00018-013-1438-6.

Kloepper J.W. abd C.J. Beauchamp. A review of issues related to measuring colonization of plant roots by bacteria. *Canad J Microbiol*. 1992;38:1219–1232.

Makarova K.S., Grishin N.V., Shabalina S.A., Wolf Y.I., Koonin E.V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. 2006;1:7. DOI 10.1186/1745-6150-1-7.

Makarova K.S., Haft H.D., Barrangou R. et al. Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiol*. 2011;9(6):467–477. DOI 10.1038/nrmicro2577.

Conflict of interest. The authors declare no conflict of interest.