

## НОВЫЕ ТЕХНОЛОГИИ

УДК 576.8

Н.П. Перетолчина<sup>1</sup>, Ю.П. Джиоев<sup>1,2</sup>, А.Ю. Борисенко<sup>1</sup>, Е.А. Воскресенская<sup>3</sup>,  
А.И. Парамонов<sup>2</sup>, Л.А. Степаненко<sup>1</sup>, О.В. Колбасева<sup>1</sup>, В.И. Злобин<sup>1</sup>

### БИОИНФОРМАЦИОННЫЙ АНАЛИЗ CRISPR/CAS СИСТЕМЫ ШТАММА *YERSINIA PSEUDOTUBERCULOSIS* IP32953

<sup>1</sup> ФГБОУ ВО «Иркутский государственный медицинский университет»  
Минздрава России, Иркутск, Россия

<sup>2</sup> ФГБНУ «Научный центр проблем здоровья семьи и репродукции человека», Иркутск, Россия

<sup>3</sup> ФБУН «Санкт-Петербургский научно-исследовательский институт эпидемиологии и микробиологии  
им. Пастера», Санкт-Петербург, Россия

Цель данного исследования: поиск и расшифровка CRISPR/Cas системы с помощью пакета программ в геноме *Y. pseudotuberculosis* IP32953. С помощью объединения результатов 4 программных методов мы обнаружили в геноме штамма *Y. pseudotuberculosis* IP32953 один набор cas-генов и три локуса: YP1, YP2 и YP3. Через спейсеры кассет были обнаружены протоспейсеры в геномах: *Y. pseudotuberculosis*, *Y. intermedia*, *Y. similis*, *Salmonella phage*, *Enterobacteria phage*, и плазмиды *Y. pseudotuberculosis* IP32953 и *Y. pseudotuberculosis* IP31758.

**Ключевые слова:** CRISPR/Cas, иерсинии, биоинформатика

### BIOINFORMATIONAL ANALYSIS OF *YERSINIA PSEUDOTUBERCULOSIS* IP32953 CRISPR/CAS SYSTEM

N.P. Peretolchina<sup>1</sup>, Y.P. Dzhioev<sup>1,2</sup>, A.Y. Borisenko<sup>1</sup>, E.A. Voskresenskaya<sup>3</sup>,  
A.I. Paramonov<sup>2</sup>, L.A. Stepanenko<sup>1</sup>, O.V. Kolbaseeva<sup>1</sup>, V.I. Zlobin<sup>1</sup>

<sup>1</sup> Irkutsk State Medical University, Irkutsk, Russia

<sup>2</sup> Scientific Centre of Family Health and Human Reproduction Problems, Irkutsk, Russia

<sup>3</sup> Saint Petersburg Pasteur Institute, Saint Petersburg, Russia

The results of this study include *Yersinia pseudotuberculosis* CRISPR/Cas system structure analysis. CRISPR/Cas system is a specific adaptive protection against heterogeneous genetic elements. The object of research was the complete genome of *Y. pseudotuberculosis* IP32953 (NC\_006155). CRISPR/Cas system screening was performed by program modelling methods MacSyFinder ver. 1.0.2. CRISPR loci screening and analyzing were carried out by program package: CRISPR Recognition tool (CRT), CRISPI: a CRISPR Interactive database, CRISPRFinder, and PilerCR. Spacer sequences were used in order to find protospacers in ACLAME, GenBank-Phage and RefSeq-Plasmid databases by BLASTn search algorithm. Protospacer sequences could be found in genomes of phages, plasmids and bacteria. In last case complete genomes of bacteria were analyzed by online-tool PHAST: PHAge Search Tool. *Y. pseudotuberculosis* IP32953 has CRISPR/Cas system that consists of one sequence of cas-genes and three loci. These loci are far away from each other. Locus YP1 is situated in close proximity to cas-genes. Protospacers were found in genomes of *Y. pseudotuberculosis* PB1/+, *Y. intermedia* Y228, *Y. similis* str. 228, *Salmonella phage*, *Enterobacteria phage*, *Y. pseudotuberculosis* IP32953 plasmid pYV and plasmid of *Y. pseudotuberculosis* IP31758. Thus, the combination of four program methods allows finding CRISPR/Cas system more precisely. Spacer sequences could be used for protospacer screening.

**Key words:** CRISPR/Cas, *Yersinia*, bioinformatics

CRISPR/Cas система (Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated proteins – короткие палиндромные повторы, регулярно расположенные группами с CRISPR-ассоциированными белками) – это специфическая адаптивная система защиты бактерий от чужеродных генетических элементов. Она представляет собой набор повторов, разделённых уникальными спейсерными участками (рис. 1). Данная структура была названа CRISPR-касsetой или -локусом. В непосредственной близости от CRISPR-локуса располагаются cas-гены, продукты деятельности которых отвечают за функционирование системы. В зависимости от вида белков и соответствующих им генов CRISPR/Cas системы класси-

фицируют на 3 типа. Молекулярный механизм действия систем различного типа различается, но имеет общие черты. В частности, первые две стадии приобретения нового спейсера и транскрипции кассет не отличаются. Отличен только механизм уничтожения чужеродного генетического материала [6, 7, 11].

Молекулярный механизм действия состоит из трёх стадий: приобретение нового спейсера, транскрипция касsetы и интерференция.

Несмотря на относительно недавнее открытие и описания CRISPR/Cas системы, она может быть использована в различных практических аспектах медицины. В первую очередь, исследуя CRISPR/Cas системы различных видов патогенных бактерий,

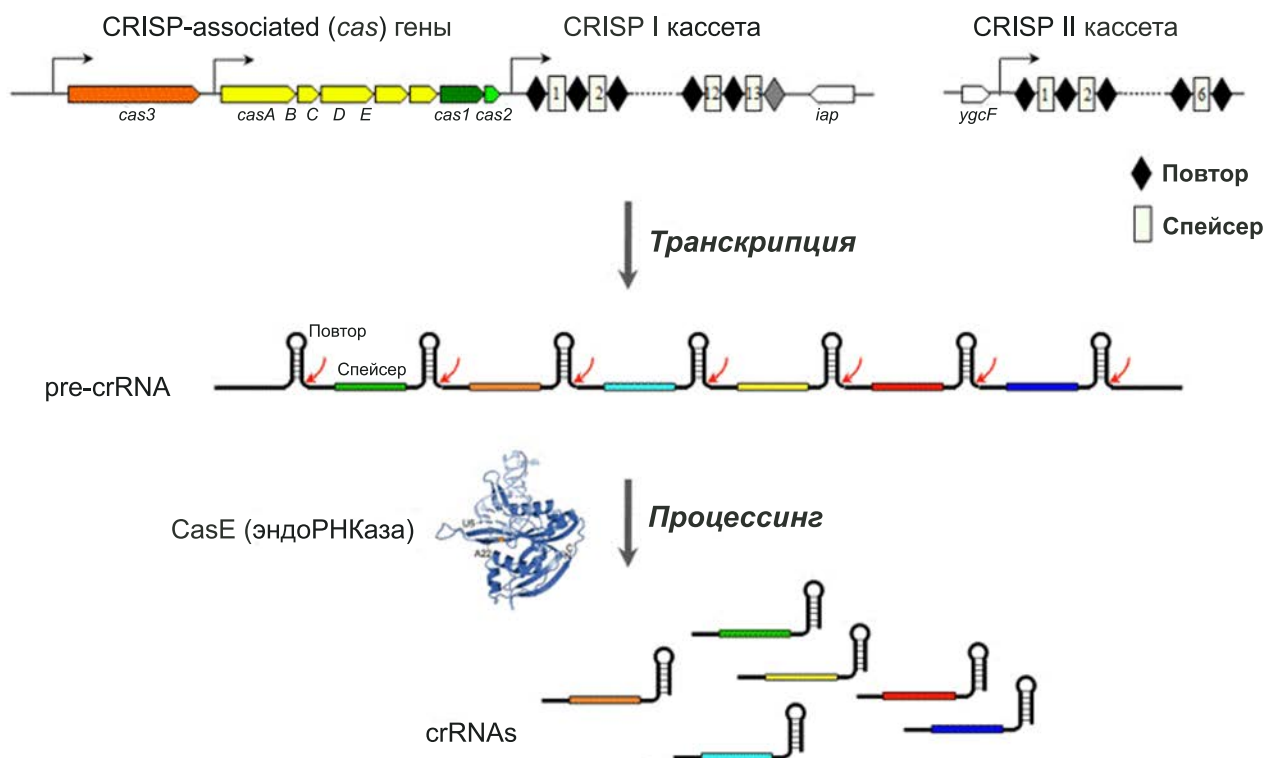


Рис. 1. Структура и механизм действия CRISPR/Cas системы.

можно предсказать устойчивость к бактериофагам, применяемым для профилактики и лечения инфекционных заболеваний. Также можно искусственно прививать бактерии от бактериофагов и плазмид. Ещё один аспект практического применения CRISPR/Cas технологий – это маркировка и типирование штаммов. Преимущество использования CRISPR-локусов для идентификации штаммов заключается прежде всего в уникальной последовательности спейсеров для каждого штамма, а также в наличии однонуклеотидных полиморфизмов SNPs (single nucleotide polymorphisms) в последовательностях спейсеров и повторов. Также некоторые штаммы могут совсем не содержать в своём геноме CRISPR-локусы или иметь только CRISPR-касеты, но не экспрессировать cas-гены [13].

В настоящее время CRISPR-локусы активно используются для типирования штаммов *Y. pestis* [13]. Но генетическое разнообразие CRISPR-систем для бактерий остальных видов *Yersinia* до сих пор остаётся малоизученным.

#### ЦЕЛЬ РАБОТЫ

Изучить структуры CRISPR/Cas системы штамма *Yersinia pseudotuberculosis* IP32953 с использованием программных методов биоинформатики.

#### МАТЕРИАЛЫ И МЕТОДЫ

В качестве объекта исследования использован геном штамма *Y. pseudotuberculosis* IP329353, представленный в базах данных GenBank (NC\_006155).

Для поиска CRISPR/Cas-систем применяли методы программного моделирования MacSyFinder (Macromolecular System Finder, ver. 1.0.2). Поиск точ-

ной гомологии последовательностей проводили при помощи установленных вспомогательных пакетов makeblastdb (ver. 2.2.28) и HMMER (ver. 3.0), а также осуществляли выявление структурных и функциональных характеристик, обнаруженных cas-генов анализируемого генома [1]. Поиск и расшифровку CRISPR-кассет производили при помощи 4 программных алгоритмов, включающих следующие программы: «PILER-CR: fast and accurate identification of CRISPR repeats», «CRISPI: a CRISPR Interactive database», «CRISPRfinder: CRISPRfinder program online» и «CRISPR Recognition Tool (CRT)» [3, 4, 8, 14]. Для скрининга фагов и плазмид по спейсерным сайтам было использовано онлайн-приложение «CRISPR Target: a tool to explore targets of CRISPRRNAs», фаговые и плазмидные базы данных ACLAME, GenBank и RefSeq и алгоритм поиска BLASTn [2, 10, 12]. Для поиска профагов использовали онлайн-приложение «PHAST: A Fast Phage Search Tool» [15].

#### РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В геноме *Y. pseudotuberculosis* IP329353 была обнаружена CRISPR/Cas система, которая включает в себя один набор cas-генов и три локуса, значительно удалённых друг от друга. При помощи MacSyFinder были описаны структурно-функциональные характеристики 3 cas- и 3 csu-генов и определён тип CRISPR/Cas системы анализируемого штамма как I F.

До настоящего времени считалось, что CRISPR/Cas система бактерий вида *Y. pseudotuberculosis* включает в себя преимущественно два локуса [5]. Однако с помощью объединения результатов 4 программных методов мы обнаружили в геноме штамма *Y. pseudotuberculosis* IP329353 три локуса: YP1, нахо-

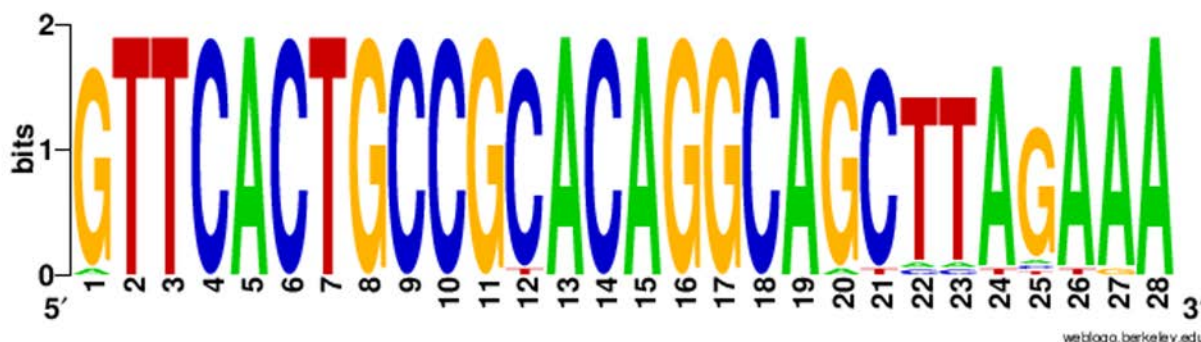


Рис. 2. Консенсусная последовательность повторов локусов *Y. pseudotuberculosis* IP32953.

дящийся в непосредственной близости от *cas*-генов, YP2 и YP3. Лидерные последовательности, с которых начинается транскрипция CRISPR-кассеты, находится перед локусами YP2 и YP3, но позади локуса YP1. Консенсусная последовательность повторов представлена на рисунке 2.

Локус YP1 состоит из 17 повторов, разделённых 16 спейсерами. Спейсеры данного локуса встречаются в локусе YP1 штаммов *Y. pseudotuberculosis* Y731, Y726, Y721, Y720, Y709, 257 и *Y. wautersii* 51. Локус YP2 состоит из 3 повторов, разделённых 2 спейсерами. Локус YP3 состоит из 5 повторов и 4 спейсеров. Спейсеры данного локуса встречаются в штаммах *Y. pseudotuberculosis* Y731, Y726, Y720, Y710.

Расшифрованные спейсерные последовательности CRISPR-кассет в дальнейшем были использованы для скрининга фагов и плазмид, к которым бактерия может проявлять устойчивость. Протоспейсеры могут встречаться и в геномах бактериофагов, и в геномах плазмид, а также в геномах бактерий. В последнем случае мы предположили наличие профага в данной части бактериального генома.

В результате поиска фагов и плазмид через спейсерные участки локуса YP1 были обнаружены протоспейсеры в геномах *Y. pseudotuberculosis* PB1+, *Y. intermedia* Y228, *Y. similis* str. 228. Протоспейсеры находятся в тех частях бактериальных хромосом, где с помощью приложения PHAST нами были обнаружены профаги. Также последовательность спейсера 3 данного локуса гомологична участкам ДНК *Salmonella* phage и *Enterobacteria* phage. Интересно, что спейсер 9 локуса YP1 идентичен последовательности гена, расположенного на плазмиде этой же бактерии: *Y. pseudotuberculosis* IP32953 плазмиды pYV. Данная плаزمиды является родоспецифической и кодирует факторы вирулентности иерсиний, а также она не является конъюгативной [9].

Спейсер 1 локуса YP2 идентичен участку плазмиды *Y. pseudotuberculosis* IP31758. Возможно, наличие гена фаговой интегразы в геноме данной плазмиды явилось триггером для приобретения спейсера *cas*-генами.

Поиск фагов и плазмид через спейсерные последовательности локуса YP3 не дал каких-либо значимых результатов на сегодняшний день. Возможно, геномы фагов и плазмид, специфичных к данным спейсерам в настоящее время ещё не представлены в базах данных.

## ВЫВОДЫ

Комбинирование известных в настоящее время программных методов поиска и расшифровки CRISPR/Cas систем позволяют более точно определить структурно-функциональные характеристики системы. Поиск фагов и плазмид через расшифрованные спейсерные последовательности с использованием доступных фаговых и плазмидных баз данных позволяет предсказать устойчивость бактерий к действию бактериофагов и к приобретению новых плазмид.

## ЛИТЕРАТУРА REFERENCES

1. Abby SS, Néron B, Ménager H, Touchon M, Rocha EP (2014). MacSyFinder: a program to mine genomes for molecular systems with an application to CRISPR-Cas systems. *PLoS one*, 9 (10), e110726.
2. Biswas A, Gagnon JN, Brouns SJ, Fineran PC, Brown CM (2013). CRISPRTarget: bioinformatic prediction and analysis of crRNA targets. *RNA biology*, 10 (5), 817-827.
3. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC, Hugenholtz P (2007). CRISPR recognition tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC bioinformatics*, 8 (1), 1.
4. Edgar RC (2007). PILER-CR: fast and accurate identification of CRISPR repeats. *BMC bioinformatics*, 8 (1), 1.
5. Eppinger M, Rosovitz MJ, Fricke WF, Rasko DA, Kokorina G, Fayolle C, Ravel J (2007). The complete genome sequence of *Yersinia pseudotuberculosis* IP31758, the causative agent of Far East scarlet-like fever. *PLoS Genet*, 3 (8), e142.
6. Gaj T, Gersbach CA, Barbas CF (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, 31 (7), 397-405.
7. Gasiunas G, Sinkunas T, Siksnys V (2014). Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and Molecular Life Sciences*, 71 (3), 449-465.
8. Grissa I, Vergnaud G, Pourcel C (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Research*, 35 (2), W52-W57.
9. Hammerl JA, Freytag B, Lanka E, Appel B, Hertwig S (2012). The pYV virulence plasmids of *Yersinia pseudotuberculosis* and *Y. pestis* contain a conserved DNA region

responsible for the mobilization by the self-transmissible plasmid pYE854. *Environmental Microbiology Reports*, 4 (4), 433-438.

10. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*, 36 (2), W5-W9.

11. Makarova KS, Haft DH, Barrangou R, Brouns SJ, Charpentier E, Horvath P, van der Oost J (2011). Evolution and classification of the CRISPR-Cas systems. *Nature Reviews Microbiology*, 9 (6), 467-477.

12. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, Fonstein M (2005). The subsystems approach to genome annotation and its use in the

project to annotate 1000 genomes. *Nucleic Acids Research*, 33 (17), 5691-5702.

13. Pourcel C, Salvignol G, Vergnaud G (2005). CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, 151 (3), 653-663.

14. Rousseau C, Gonnet M., Le Romancer M., Nicolas J (2009). CRISPI: a CRISPR interactive database. *Bioinformatics*, 25 (24), 3317-3318.

15. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS (2011). PHAST: a fast phage search tool. *Nucleic Acids Research*, gkr485.

#### Сведения об авторах Information about the authors

**Перетолчина Надежда Павловна** – аспирант ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России (664003, г. Иркутск, ул. Красного Восстания, 1; тел.: 8 (3952) 24-38-43; e-mail: nadine1lenz@gmail.com)  
**Peretolchina Nadezhda Pavlovna** – Postgraduate of Irkutsk State Medical University (664003, Irkutsk, Krasnogo Vosstaniya, 1; tel.: +7 (3952) 24-38-43; e-mail: nadine1lenz@gmail.com)

**Джиоев Юрий Павлович** – кандидат биологических наук, ведущий научный сотрудник НИИ биомедицинских технологий ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России, старший научный сотрудник лаборатории молекулярной эпидемиологии и генетической диагностики ФГБНУ «Научный центр проблем здоровья семьи и репродукции человека» (664003, г. Иркутск, ул. Тимирязева, 16; e-mail: alanir07@mail.ru)

**Dzhioev Yuri Pavlovich** – Candidate of Biological Sciences, Leading Research Officer of Scientific Research Institute of Biomedical Technologies of Irkutsk State Medical University, Senior Research Officer of the Laboratory of Molecular Epidemiology and Genetic Diagnosis of Scientific Centre of Family Health and Human Reproduction Problems (664003, Irkutsk, Timiryazev str., 16; e-mail: alanir07@mail.ru)

**Борисенко Андрей Юрьевич** – аспирант ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России

**Borisenko Andrey Yurevich** – Postgraduate of Irkutsk State Medical University

**Воскресенская Екатерина Александровна** – кандидат биологических наук ФБУН «Санкт-Петербургский научно-исследовательский институт эпидемиологии и микробиологии им. Пастера» (197101, г. Санкт-Петербург, ул. Мира, 14; e-mail: tsenevapasteur@yandex.ru)

**Voskresenskaya Ekaterina Alexandrovna** – Candidate of Biological Sciences, Saint Petersburg Pasteur Institute (197101, Saint-Petersburg, Mir str., 14; e-mail: tsenevapasteur@yandex.ru)

**Парамонов Алексей Игоревич** – лаборант-исследователь лаборатории молекулярной эпидемиологии и генетической диагностики ФГБНУ «Научный центр проблем здоровья семьи и репродукции человека» (e-mail: paramonov\_a.i@mail.ru)

**Paramonov Alexey Igorevich** – Clinical Research Assistant of the Laboratory of Molecular Epidemiology and Genetic Diagnosis of Scientific Centre of Family Health and Human Reproduction Problems (e-mail: paramonov\_a.i@mail.ru)

**Степаненко Лилия Александровна** – кандидат медицинских наук, старший научный сотрудник НИИ биомедицинских технологий ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России (e-mail: steplia@mail.ru)

**Stepanenko Lilia Alexandrovna** – Candidate of Medical Sciences, Senior Research Officer of Scientific Research Institute of Biomedical Technologies of Irkutsk State Medical University (e-mail: steplia@mail.ru)

**Колбасеева Ольга Владимировна** – кандидат биологических наук, старший научный сотрудник НИИ биомедицинских технологий ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России (e-mail: kolbaseeva@yandex.ru)

**Kolbaseeva Olga Vladimirovna** – Candidate of Biological Sciences, Senior Research Officer of Scientific Research Institute of Biomedical Technologies of Irkutsk State Medical University (e-mail: kolbaseeva@yandex.ru)

**Злобин Владимир Игоревич** – академик РАН, профессор кафедры микробиологии, вирусологии и иммунологии с курсом клинической лабораторной диагностики, директор НИИ биомедицинских технологий ФГБОУ ВО «Иркутский государственный медицинский университет» Минздрава России (e-mail: vizlobin@mail.ru)

**Zlobin Vladimir Igorevich** – Academician of RAS, Professor of the Department of Microbiology, Virology and Immunology with the Course of Clinical Laboratory Diagnostics, Director of Scientific Research Institute of Biomedical Technologies of Irkutsk State Medical University (e-mail: vizlobin@mail.ru)