

А.В.Шашкова, А.А.Горяев, Н.И.Смирнова

СТРОЕНИЕ И ФУНКЦИОНАЛЬНАЯ РОЛЬ CRISPR-СИСТЕМЫ БАКТЕРИЙ

ФГУЗ «Российский научно-исследовательский противочумный институт «Микроб», Саратов

Представлен обзор литературных данных о недавно открытой системе кластрированных, равномерно удаленных друг от друга коротких палиндромных повторов, или системе CRISPR, которая участвует в защите от проникновения чужеродной генетической информации у прокариот. Описаны особенности строения и функции CRISPR, а также предполагаемый механизм действия. Также представлены данные о наличии этой системы у возбудителей особо опасных инфекций и возможности ее использования для молекулярного типирования.

Ключевые слова: CRISPR, *cas* гены, бактериофаги, возбудители особо опасных инфекций.

A.V.Shashkova, A.A.Goryaev, N.I.Smirnova

Structure and Functional Role of Bacterial CRISPR System

Russian Research Anti-Plague Institute "Microbe", Saratov

Presented is the review of literature data on the recently discovered system of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), which takes part in the defense against penetration of alien genetic information in prokaryotes. CRISPR structural peculiarities and functions, and putative mechanism of action are described. Also presented are data on the availability of this system in particularly dangerous infections agents and on the possibility of its application for molecular typing.

Key words: CRISPR, *cas* genes, bacteriophages, agents of particularly dangerous infections.

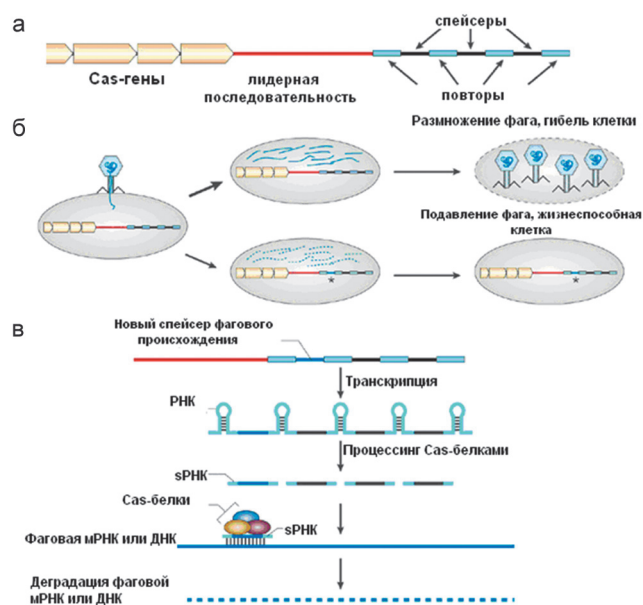
На сегодняшний день неоспоримым является тот факт, что главный вклад в эволюцию геномов прокариот вносит горизонтальный перенос генов. Между тем, приобретение нового чужеродного генетического материала может быть позитивным, негативным или нейтральным [26]. У прокариот существуют механизмы как «поощряющие» приобретение генетической информации (конъюгация, рекомбинация, трансформация и др.), так и системы защиты, ограничивающие горизонтальный перенос (системы рестрикции-модификации, сахар-неспецифические нуклеазы, abortивная инфекция и др.) [14, 30].

Кроме того, в защите от внедрения бактериофагов и конъюгативных плазмид в клетки прокариот участвуют недавно обнаруженные кластрированные, равномерно удаленные друг от друга короткие палиндромные повторы – CRISPR (от Clustered Regularly Interspaced Short Palindromic Repeats). CRISPR, впервые обнаруженные в 1987 году [15], сейчас найдены в секвенированных геномах 91,0 % архей и 45,0 % бактерий [9]. Между тем этот новый уникальный механизм «иммунитета» прокариот остается еще до конца невыясненным и требует более детального рассмотрения.

Строение CRISPR

Типичная CRISPR-система представляет собой кассету, состоящую из коротких уникальных участков – спейсеров размером 27–72 п.н. и разделяющих их строгих палиндромных повторов размером 24–47 п.н. (рисунок, а) [23, 24]. В одной CRISPR-кассете число повторов может доходить до нескольких сотен (у *Verminephrobacter eiseniae* 249 повторов), которые всегда идентичны, однако один или несколько последних могут отличаться. Повторы

CRISPR-последовательностей высококонсервативны в пределах одного вида. Наличие коротких (5–7 п.н.) палиндромов обеспечивает возможность формирования вторичной структуры «стержень-петля» зрелой crRNA (CRISPR-RNA) [18]. Спейсерные последовательности уникальны. Около 40 % обнаруженных спейсеров имеют гомологию с известными последовательностями, представленными в GenBank, из них 75 % гомологичны нуклеотидным последователь-



Строение и функционирование CRISPR:

а – строение CRISPR/cas-системы; б – клетки, не несущие спейсеров фагового происхождения, погибают при заражении фагом, а клетки, содержащие в CRISPR-локусе соответствующую фагу последовательность (на рисунке отмечена звездочкой), выживают; в – возможная модель работы системы CRISPR [по 28]

ностям бактериофагов и 20 % – плазмид [3]. Часть остальных спейсеров гомологична генам, имеющим внехромосомное происхождение, или расположенным в потенциально перемещенных участках ДНК из неродственных организмов. Однако подавляющее большинство спейсеров не имеют гомологии с известными нуклеотидными последовательностями, и это может говорить об огромном разнообразии бактериофагов в природе [3].

На 5'-конце CRISPR граничат с АТ-богатыми участками, получившими название лидерной последовательности длиной до 550 п.н. [28]. Лидерные последовательности могут служить промоторами для транскрипции множества CRISPR [10]. Также считается, что эта последовательность участвует во встраивании новых спейсеров [1].

Вблизи CRISPR были обнаружены последовательности, получившие название *cas* генов (от CRISPR-associated genes) (рисунок, а). Детальный биоинформативный анализ белков, биосинтез которых кодируют *cas* гены, показал их принадлежность к 25 различным семействам [10, 16, 21], но общим для них является содержание РНК- или ДНК-связывающих доменов, типичных для нуклеаз, хеликаз, полимераз [21]. Молекулярные функции CAS белков были предсказаны на основе анализа их аминокислотных последовательностей [21], эффектов инактивации [1, 4] или структурном сходстве [8]. CRISPR в комбинации с CAS белками формирует систему CASS (от CRISPR associated system) [10, 16]. Наиболее часто встречаются от 2 до 6 *cas* генов, при этом *cas1* является универсальным маркером CRISPR/Cas системы. Белок CAS1 относится к металлозависимым эндонуклеазам ДНК без определенной специфичности к последовательности [31]. CAS2-специфическая эндорибонуклеаза [2]. Ген *cas3* кодирует белок, состоящий из нуклеазного и хеликазного доменов [21]. D.Nan и G.Krauss показана способность нуклеазного домена белка CAS3 у *Sulfolobus solfataricus* к гидролизу двуцепочечных ДНК или РНК [12]. CAS4 является *hcsB*-подобной нуклеазой [20]. Интересным является тот факт, что у нескольких бактерий *cas4* ген объединен с *cas1*, а у других ген *cas2* с *cas3*, что, возможно, свидетельствует о совместной функции белков, кодируемых этими генами [26]. Белки CAS5 и CAS6 относятся к семейству RAMP (repeat associated mysterious protein) с предсказанной возможностью связывания и разрезания РНК [5, 8, 21]. В зависимости от состава генов и последовательностей повторов все системы CRISPR разделяются на 8 подсемейств: Ecoli, Ypest, Nmeni, Aperm, Tnear, Hmani, Mtube, Dvulg [10].

Механизм действия CRISPR/CAS

Молекулярный механизм защиты CRISPR/Cas-системы еще мало изучен. В нем, как правило, выделяют три отдельные стадии: адаптация, или иммунизация, – приобретение коротких последовательностей чужеродных нуклеиновых кислот в качестве новых спейсеров; экспрессия CRISPR – формирование зре-

лых crRNA; интерференция – узнавание чужеродной ДНК/РНК и ее разрушение (рисунок, в) [26].

В работах R.Barrangou *et al.* доказано, что в ответ на инфицирование фагами *Streptococcus thermophilus* приобрел специфическую устойчивость [1]. Резистентность к фагам была обусловлена появлением новых спейсеров в системе CRISPR, которая полностью идентична инфицирующим фагам. Экспериментально показано, что приобретение новых спейсеров осуществляется со стороны лидерной последовательности [19]. Предполагается, что лидерная последовательность содержит сайты связывания CAS белков. Поскольку новые спейсеры внедряются преимущественно в 5'-области системы, CRISPR представляет собой хронологическую запись взаимодействий бактерии с мобильными генетическими элементами [26]. Это свойство CRISPR может быть использовано для типирования, выяснения происхождения и географического распространения штаммов при эпидемиологических исследованиях, так как для штаммов, выделенных на различных территориях, характерны взаимодействия с различными бактериофагами [17, 27, 30].

Процесс распознавания и интеграции новых спейсеров в настоящее время недостаточно изучен. Считается, что важную роль в процессе адаптации играют так называемые мотивы, смежные с прото-спейсером, или PAM (от proto-spacer adjacent motif) [25]. Прото-спейсером называют последовательность в фагах или плаزمидах, гомологичную спейсеру. Предшественники спейсеров выбираются после узнавания PAM. Эти мотивы, характерные для последовательностей фагов и плазмид, отсутствуют в геноме бактерии, что позволяет отличить ДНК хозяина от ДНК чужеродного агента. Мутации в PAM вирусов или плазмид приводят к потере устойчивости бактерий к этим патогенам, несмотря на сохранение последовательности вставки в CRISPR [26]. Однако в CRISPR/Cas-системах, воздействующих на РНК, эти мотивы могут отсутствовать. До сих пор экспериментально не установлено какие *cas* гены участвуют в захвате и встраивании новых спейсеров. Предполагается участие генов *cas1* и *cas4* в этих процессах [25]. Кроме того, инактивация гена *cas7* приводит к потере бактериями способности приобретать новые спейсеры [13].

Экспрессия системы CRISPR выражается в транскрипции предшественника crRNA (pre-crRNA), которая далее подвергается процессу «созревания» и превращения в crRNAs, которые служат регуляторными последовательностями.

Транскрипция CRISPR начинается с лидерной последовательности и содержит все повторы и спейсеры этой CRISPR-структуры, так называемая pre-crRNA (от poly-spacer precursor crRNA). Далее происходит связывание pre-crRNA с комплексом CAS белков Cascade (от Cas-complex for anti-virus defense). Нуклеазы этого комплекса Cascade (предположительно CAS2) разрезают транскрипт на зрелые

сrRNA. Зрелые сrRNA длиной 39–45 нуклеотидов, содержат одну последовательность спейсера, а также 8–9 нуклеотидов повтора на 5'-конце, и более гетерогенный фрагмент части повтора на 3'-конце [11, 21]. Эти части повторов необходимы для формирования структуры стержень-петля. Кроме того, фланкирующие повторы играют важнейшую роль в способности отличать чужеродную от собственной ДНК [22].

Интерференция чужеродной ДНК или РНК обеспечивается за счет взаимодействия сrRNA и комплекса Cascade, сrRNA комплементарно узнают последовательность протоспейсера, а CAS белки обеспечивают ее разрушение. Проведенные эксперименты по получению мутантных бактерий с различным содержанием CAS белков позволили предположить ключевую роль белка CAS3, так как введение в геном этого гена заметно уменьшало чувствительность бактерий к фагам [4].

Важным остается вопрос, почему не происходит разрезание непосредственно самой системы CRISPR, несмотря на комплементарность последовательности. Последние исследования показали, что важную роль в этом играет часть повтора, фланкирующая 5'-конец сrRNA. При этом решающую роль играет наличие или отсутствие комплементарности ключевых нуклеотидов их участков. Если участок не комплементарен, что наблюдается для последовательностей чужеродных агентов, нуклеиновая кислота подвергается разрушению. В случае комплементарности с ДНК хозяина, части повторов имеют полное соответствие, поэтому разрезание не происходит [22].

CRISPR у возбудителей особо опасных инфекций

Как отмечалось выше, CRISPR обнаружены в геномах практически половины известных видов бактерий. Для нас, в первую очередь, интересна распространенность CRISPR-структур среди возбудителей особо опасных инфекций. Согласно данным, представленным в CRISPRdb [9], CRISPR-структуры присутствуют в геноме возбудителей чумы, холеры и туляремии (таблица), а у возбудителей сибирской язвы, бруцеллеза и мелиоидоза обнаружены CRISPR-подобные структуры, содержащие несколько повторов и спейсеров, но не связанные с CAS белками. Следует отметить, что в CRISPRdb представлены только штаммы, геномы которых частично или полностью секвенированы и представлены в базе данных GenBank, поскольку поиск CRISPR на данный момент осуществляется в основном с помощью биоинформативных методов.

Наиболее изучены CRISPR у *Yersinia pestis*, в геноме которой обнаружены три таких структуры, причем с одинаковыми последовательностями повторов. Данные по сравнению структуры и организации спейсеров используются при исследовании филогенетических отношений штаммов, поскольку было показано, что наличие идентичных спейсерных последовательностей в аллелях CRISPR указывает на наличие общего предка и не является резуль-

татом независимых событий [7]. Исследованные с помощью MLVA филогенетические связи штаммов *Y. pestis* были подтверждены данными анализа последовательностей CRISPR-локуса [27]. Кроме того, CRISPR-система *Y. pestis* (YPa) часто используется в качестве VNTR-маркера при типировании штаммов чумного микроба [7].

Что касается возбудителя холеры, то CRISPR/Cas-система не выявлена у штаммов *Vibrio cholerae* N16961, M66-2, MJ-1236 эльтор биоваров, но обна-

Наличие CRISPR-системы в геномах возбудителей особо опасных инфекций, представленных в CRISPRdb [9] (с небольшими модификациями)

Вид, штамм	Структуры, подобные CRISPR	CRISPR
<i>Bacillus anthracis</i> str. Ames Ancestor	2	-
<i>Bacillus anthracis</i> str. A0248	3	-
<i>Bacillus anthracis</i> str. Ames	2	-
<i>Bacillus anthracis</i> str. CDC	1	-
<i>Bacillus anthracis</i> str. Sterne	2	-
<i>Brucella abortus</i> biovar 1 str. 9-941	-	-
<i>Brucella abortus</i> S19	-	-
<i>Brucella melitensis</i> biovar <i>Abortus</i> 2308	-	-
<i>Brucella melitensis</i> 16M	1	-
<i>Brucella melitensis</i> ATCC 23457	1	-
<i>Brucella suis</i> 1330	-	-
<i>Brucella suis</i> ATCC 23445	-	-
<i>Burkholderia mallei</i> ATCC 23344	-	-
<i>Burkholderia mallei</i> NCTC 10229	-	-
<i>Burkholderia mallei</i> NCTC 10247	-	-
<i>Burkholderia mallei</i> SAVP1	-	-
<i>Burkholderia pseudomallei</i> MSHR346	-	-
<i>Burkholderia pseudomallei</i> 668	-	-
<i>Burkholderia pseudomallei</i> 1106a	1	-
<i>Burkholderia pseudomallei</i> 1710b	1	-
<i>Burkholderia pseudomallei</i> K96243	3	-
<i>Francisella tularensis</i> subsp. <i>tularensis</i> FSC 198	-	-
<i>Francisella tularensis</i> subsp. <i>tularensis</i> SCHU S4	-	-
<i>Francisella tularensis</i> subsp. <i>tularensis</i> WY96-3418	-	-
<i>Francisella tularensis</i> subsp. <i>novicida</i> U112	2	2
<i>Francisella tularensis</i> subsp. <i>holarctica</i> FTA	-	-
<i>Francisella tularensis</i> subsp. <i>holarctica</i> OSU18	-	-
<i>Francisella tularensis</i> subsp. <i>mediasiatica</i> FSC147	-	-
<i>Vibrio cholerae</i> biovar <i>eltor</i> M66-2	-	-
<i>Vibrio cholerae</i> biovar <i>eltor</i> (Mozambique) MJ-1236	-	-
<i>Vibrio cholerae</i> O1 biovar <i>eltor</i> str. N16961	-	-
<i>Vibrio cholerae</i> O1 classic O395	-	1
<i>Yersinia pestis</i> Antiqua	1	3
<i>Yersinia pestis</i> biovar <i>Microtus</i> str. 91001	-	2
<i>Yersinia pestis</i> CO92	-	3
<i>Yersinia pestis</i> KIM	1	3
<i>Yersinia pestis</i> Nepal516	-	3
<i>Yersinia pestis</i> Pestoides F	-	3

ружена у *V. cholerae* O395 классического биовара. CRISPR у этого штамма расположена на второй хромосоме и состоит из 40 повторов и 39 уникальных спейсеров. Длина последовательностей повторов равна 28 п.н. Они имеют высокую степень гомологии с последовательностями повторов некоторых других бактерий (*Klebsiella pneumoniae*, *Clostridium tepidum*). Перед областью CRISPR в геноме *V. cholerae* O395 расположены восемь *cas* генов (*cas1*, *cas2*, *cas3*) и пять подтип-специфичных генов *Escherichia coli* (*cse1*, *cse2*, *cse3*, *cse4*, *cse5e*) [6].

В заключение хотелось бы подчеркнуть, что как широкое распространение CRISPR-системы во многих бактериях и археях, так и их огромное разнообразие указывает на то, что она может быть одной из самых древних «антивирусных» оборонных систем в микробном мире. И, несмотря на огромное количество вопросов, касающихся механизма действия этой системы и роли различных CAS белков, которые все еще остаются без ответа, уже сейчас очевиден прикладной аспект использования системы CRISPR. Так, эти структуры уже широко применяются для типирования штаммов *Y. pestis*. Это говорит о том, что обнаружение и изучение распространенности CRISPR-систем у других возбудителей особо опасных инфекций на большой выборке штаммов также может открыть возможность использования CRISPR в типировании и выяснении географического происхождения штаммов. Кроме того, дальнейшее изучение структуры и функций CRISPR позволит определить возможность их использования для ограничения распространения нежелательных мобильных генетических элементов при проведении генно-инженерных работ.

СПИСОК ЛИТЕРАТУРЫ

1. Barrangou R., Fremaux C., Deveau H., Richards M., Boyaval P., Moineau S., Romero D.A., Horvath P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science*. 2007; 315(5819):1709–12.
2. Beloglazova N., Brown G., Zimmerman M. et al. A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* 2008; 283(29):20361–71.
3. Bolotin A., Quinquis B., Sorokin A., Ehrlich S.D. Clustered regularly interspaced short palindromic repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005; 151:2551–61.
4. Brouns S.J.J., Jore M.M., Lundgren M., Westa E.R., Slijkhuys R.J.H., Snijders A.P.L. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*. 2008; 321(5891):960–4.
5. Carte J., Wang R., Li H., Terns R.M., Terns M.P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* 2008; 22(24):3489–96.
6. Chakraborty S., Waise T.M.Z., Hassan F., Kabir Y., Smith M.A., Arif M. Assessment of the evolutionary origin and possibility of CRISPR-Cas (CASS) mediated RNA interference pathway in *Vibrio cholerae* O395. *In Silico Biology*. 2009; 9(4):245–54.
7. Cui Y., Li Y., Gorge O., Platonov M. E., Yan Y., Guo Z. et al. Insight into microevolution of *Yersinia pestis* by clustered regularly interspaced short palindromic repeats. *PLoS ONE*. 2008; 3(7):e2652
8. Ebihara A., Yao M., Masui R., Tanaka I., Yokoyama S., Kuramitsu S. Crystal structure of hypothetical protein TTHB192 from *Thermus thermophilus* HB8 reveals a new protein family with an RNA recognition motif-like domain. *Protein. Sci.* 2006; 15:1494–9.
9. Grissa I., Vergnaud G., Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinformatics* 2007; 8:172.
10. Haft D.H., Selengut J., Mongodin E.F., Nelson K.E. A guild of 45 CRISPR-associated (CAS) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput.*

Biol. 2005; 1(6):e60.

11. Hale C.R., Zhao P., Olson S., Duff M.O., Graveley B.R., Wells L. et al. RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell*. 2009; 139(5):945–56.
12. Han D., Krauss G. Characterisation of the endonuclease SSO2001 from *Sulfolobus solfataricus* P2. *FEBS Lett.* 2009; 583:771–6.
13. Horvath P., Coute-Monvoisin A.C., Romero D.A., Boyaval P., Fremaux C., Barrangou R. Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int. J. Food. Microbiol.* 2009; 1(131):62–70.
14. Hsia K.C., Li C.L., Yuan H.S. Structural and functional insight into sugar-nonspecific nucleases in host defense. *Curr. Opin. Struct. Biol.* 2005; 15(1):126–34.
15. Ishino Y., Shinagawa H., Makino K., Amemura M., Nakata A. Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* 1987; 169:5429–33.
16. Jansen R., Embden van J.D.A., Gastra W., Schouls L.M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* 2002; 43(6):1565–75.
17. Kamerbeek J., Schouls L., Kolk A., van Agterveld M., van Soolingen D., Kuijper S. et al. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* 1997; 35:907–14.
18. Kunin V., Sorek R., Hugenholtz P. Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol.* 2007; 8:R61.
19. Lillestøl R.K., Redder P., Garrett R.A., Brügger K. A putative viral defence mechanism in archaeal cells. *Archaea*. 2006; 2:59–72.
20. Makarova K.S., Aravind L., Grishin N.V., Rogosin I.B., Koonin E.V. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucl. Acid Res.* 2002; 30:482–96.
21. 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. *Biology direct*. 2006; 1:7.
22. Maraffini L., Sontheimer E.J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*. 2010; 463(1):568–72.
23. Mojica F.J., Diez-Villasenor C., Soria E., Juez G. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Mol. Microbiol.* 2000; 36:244–6.
24. Mojica F.J., Diez-Villasenor C., Garcia-Martinez J., Soria E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* 2005; 60:174–82.
25. Mojica F.J., Diez-Villasenor C., Garcia-Martinez J., Almedros C. Short motif sequences determine the targets of the prokaryotic CRISPR defense system. *Microbiology*. 2009; 155(3):733–40.
26. Oost van der J., Jore M.M., Westa E.R., Lundgren M., Brouns S.J.J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends in Biochemical Sciences*. 2009; 34(8): 401–7.
27. Pourcel C., Salvignol G., Vergnaud G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*. 2005; 151(3):653–63.
28. Sorek R, Kunin V., Hungenholtz P. CRISPR – a widespread system that provides acquired resistance against phages in bacteria and archae. *Nat. Rev. Microbiol.* 2008; 3(6):181–6.
29. Sturino J.M., Klaenhammer T.R. Engineered bacteriophage-defence systems in bioprocessing. 2006; 4:395–404.
30. Vergnaud G., Li Y., Gorge O., Song Y., Zhou D., Grissa I. et al. Analysis of the three *Yersinia pestis* CRISPR loci provides new tools for the investigation of ancient DNA. *Adv. Exp. Med. Biol.* 2007; 603:327–38.
31. Wiedenheft B., Zhou K., Jinek M., Coyle S., Ma W., Doudna J.A. Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure*. 2009;17(10):904–12.

Authors:

Shashkova A.V., Goryaev A.A., Smirnova N.I. Russian Research Anti-Plague Institute “Microbe”. Universitetskaya St., 46, Saratov, 410005, Russia. E-mail: microbe@san.ru

Об авторах:

Шашикова А.В., Горяев А.А., Смирнова Н.И. Российский научно-исследовательский противочумный институт «Микроб». 410005, Саратов, ул. Университетская, 46. E-mail: microbe@san.ru

Поступила 13.04.10.