(19) Евразийское патентное ведомство

(21) 201991443 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.01.14

(22) Дата подачи заявки
2016.12.14

(51) Int. Cl. ***C12N 9/22*** (2006.01)
C12N 15/113 (2010.01)

(54) ТЕРМОСТАБИЛЬНЫЕ НУКЛЕАЗЫ Cas9

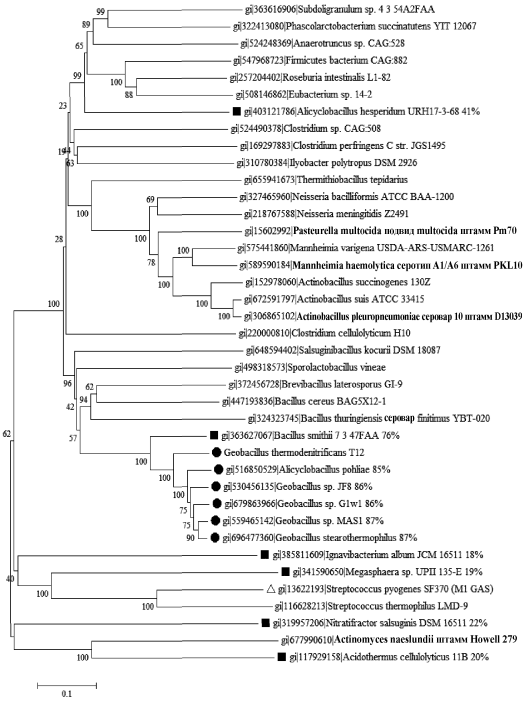
(86) **PCT/EP2016/081077**
(87) **WO 2018/108272 2018.06.21**

(71) Заявитель:
ВАГЕНИНГЕН УНИВЕРСИТЕЙТ; СТИХТИНГ ВОР ДЕ ТЕХНИШЕ ВЕТЕНСХАППЕН (NL)

(72) Изобретатель:
Ван Дер Ост Джон, Ван Краненбург Рихард (NL), Босма Эллеке Фенна (DK), Маугиакос Иоаннис (NL)

(74) Представитель:
Медведев В.Н. (RU)

(57) Термостабильные нуклеазы Cas9. Настоящее изобретение относится к области генной инженерии и, в частности, к редактированию нуклеиновых кислот и модификации генома. Настоящее изобретение относится к выделенному белку Cas или его полипептидному фрагменту, имеющему аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%. Белок или полипептид Cas способен к связыванию, расщеплению, маркированию или модификации двухцепочечного полинуклеотида-мишени при температуре в диапазоне от 30 до 100°C включительно. Изобретение также относится к выделенным молекулам нуклеиновой кислоты, кодирующим указанные нуклеазы Cas9, экспрессионным векторам и клеткам-хозяевам. По изобретению также предложены последовательности PAM, узнаваемые белком или полипептидом Cas. Нуклеазы Cas9, раскрытые в настоящем документе, являются новыми инструментами для генетической инженерии при повышенных температурах и имеют особую ценность при генетических манипуляциях с термофильными организмами, в частности микроорганизмами.



201991443 A1

201991443 A1

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-557526EA/061

ТЕРМОСТАБИЛЬНЫЕ НУКЛЕАЗЫ CAS9

ОБЛАСТЬ ТЕХНИКИ, К КОТОРОЙ ОТНОСИТСЯ ИЗОБРЕТЕНИЕ

Настоящее изобретение относится к области генной инженерии и, в частности, к редактированию нуклеиновых кислот и модификации генома. Настоящее изобретение относится к инструментам генной инженерии в виде нуклеаз, которые могут быть сконструированы для направленного сайт-специфического связывания, одноцепочечного разрыва, разрезания и модификации генетического материала; к рибонуклеопротеидам, которые проявляют активность, в частности, нуклеазную активность, в специфических сайтах генетического материала последовательности, а также к модифицированным нуклеазам и рибонуклеопротеидам для использования в качестве маркеров. Следовательно, изобретение также относится к конструктам, связанным с экспрессией, для доставки и экспрессии нуклеаз и гид-РНК в клетках. Кроме того, изобретение относится к специфичному для последовательностей редактированию нуклеиновых кислот *in vitro* или *in vivo* и к способам, используемым для достижения этой цели. Конкретной областью, к которой относится изобретение, является генетическая манипуляция с термофильными организмами, особенно микроорганизмами.

ПРЕДПОСЫЛКИ СОЗДАНИЯ ИЗОБРЕТЕНИЯ

В 2007 году впервые было продемонстрировано, что CRISPR-Cas является адаптивной иммунной системой у многих бактерий и большинства археев (Barrangou *et al.*, 2007, Science 315: 1709-1712), Brouns *et al.*, 2008, Science 321: 960-964). До настоящего времени на основе функциональных и структурных критериев были охарактеризованы три типа систем CRISPR-Cas, в большинстве из которых используются малые молекулы РНК в качестве гид-РНК для направления на комплементарные последовательности ДНК (Makarova *et al.*, 2011, Nat Rev Microbiol 9: 467-477; Van der Oost *et al.*, 2014, Nat Rev Microbiol 12: 479-492).

Недавно в лабораториях Doudna/Charpentier была проведена тщательная характеристика эффекторного фермента системы CRISPR-

Cas типа II (Cas9), включая демонстрацию того, что введение сконструированных гид-РНК для CRISPR (со специфическими спейсерными последовательностями) приводит к направленному действию на комплементарные последовательности (протоспейсеры) на плазмиде, вызывающему разрывы двойной цепи этой плазмиды (Jinek *et al.*, 2012, Science 337: 816-821). Затем в работе Jinek *et al.*, 2012, Cas9 была использована в качестве инструмента для редактирования генома.

Cas9 была использована для редактирования геномов ряда эукариотических клеток (например, рыб, растений, человека) (Charpentier and Doudna, 2013, Nature 495: 50-51).

Кроме того, Cas9 была использована для улучшения выходов гомологичной рекомбинации у бактерий путем отбора целенаправленных событий рекомбинации (Jiang *et al.*, 2013, Nature Biotechnol 31: 233-239). Для достижения этого токсический фрагмент (направляющий конструкт) совместно трансфицируют со спасающим фрагментом, несущим требуемое изменение (редактирующий конструкт, несущий точечную мутацию или делеции). Направляющий конструкт состоит из Cas9 в сочетании со сконструированным CRISPR и маркером устойчивости к антибиотикам, определяя сайт желаемой рекомбинации на хромосоме хозяина; в присутствии соответствующего антибиотика производят отбор на интеграцию направляющего конструкта в хромосому хозяина. Только тогда, когда происходит дополнительная рекомбинация редактирующего конструкта с сайтом-мишенью CRISPR на хромосоме хозяина, хозяин может избежать проблемы аутоиммунности. Следовательно, в присутствии антибиотика только желаемые (свободные от маркера) мутанты способны выживать и расти. Также представлена связанная стратегия отбора для последующего удаления интегрированного направляющего конструкта из хромосомы, с получением свободного от собственного маркера мутанта.

В последние годы было установлено, что редактирование генома, опосредуемое CRISPR-Cas, является полезным инструментом для генной инженерии. Установлено, что прокариотические системы CRISPR служат своим хозяевам в качестве адаптивных иммунных систем (Jinek *et al.*, 2012, Science 337: 816-821) и могут быть

использованы для быстрых и эффективных генетических манипуляций (Mali *et al.*, 2013, Nat Methods 10: 957-963), для которых нужны лишь модификации гид-последовательности для направления на интересующие последовательности.

Тем не менее, существует постоянная потребность в разработке средств для более эффективного специфичного для последовательности обнаружения, расщепления и манипуляций с нуклеиновыми кислотами в различных экспериментальных условиях для применения в области генетических исследований и редактирования генома. В частности, имеющиеся в настоящее время специфичные для последовательности инструменты для редактирования генома, включая Cas9, не пригодны для использования во всех условиях или для всех организмов, например, специфичные для последовательности нуклеазы являются относительно термочувствительными и поэтому не подходят для использования в случае строго термофильных микроорганизмов (которые способны расти при температуре от 41°C до 122°C и оптимально растут в диапазоне температур от >45°C до 80°C, с гипертермофилами, способными к оптимальному росту при температуре выше 80°C), например, микроорганизмов, которых используют в промышленной ферментации или в лабораторных процессах *in vitro*, проводимых при повышенных температурах.

На сегодняшний день отсутствуют экспериментальные доказательства наличия активных белков Cas9 у термофилов. На основе сравнительного скрининга генома, проведенного Chylinski *et al.* (2014; Nucleic Acids Research 42: 6091-61-05) на предмет наличия Cas9 у бактерий, было обнаружено, что система CRISPR-Cas типа II-C присутствует лишь примерно в 3,3% всех бактериальных геномов. Среди термофильных бактерий система типа II недостаточно представлена, исходя из результатов статистического анализа (P=0,0019). Кроме того, у археев не была обнаружена система типа II, однако это, предположительно, может быть связано с отсутствием белка РНКазы III (вовлеченной в систему типа II) у археев. В публикации Chylinski, *et al.* (2014; Nucleic Acids Research 42: 6091-6105) описана классификация и эволюция

систем CRISPR-Cas типа II, в частности, указаны два вида, которые имеют данные системы, однако эти виды максимально растут при 55°C и у них отсутствует строго термофильный рост с оптимальной температурой роста 60-80°C, причем гипертермофилы способны оптимально расти при температуре выше 80°C.

Несмотря на редкость системы CRISPR-Cas в бактериальных геномах и, в частности, на тот факт, что Cas9 была обнаружена только у бактерий (не археев) с оптимальными температурами роста ниже 45°C, авторы изобретения неожиданно обнаружили несколько термостабильных вариантов Cas9, которые позволяют редактировать геном при повышенных температурах. Авторы изобретения также установили оптимизированные последовательности мотива, примыкающего к протоспейсеру (PAM), которые действуют с термостабильными вариантами Cas9, позволяя редактировать геном в широком диапазоне температур, в том числе, при повышенных температурах. Эти нуклеазы Cas9 и молекулы РНК, которые были сконструированы со знанием ассоциированных последовательностей PAM, являются новыми инструментами для генной инженерии при повышенных температурах и имеют особую ценность при генетических манипуляциях с термофильными организмами; в частности, микроорганизмами.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Соответственно, настоящее изобретение относится к белку или полипептиду (Cas), связанному с изолированным кластером регулярно расположенных группами коротких палиндромных повторов (CRISPR), содержащему:

a. аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]; и/или

b. аминокислотный мотив $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и/или

c. аминокислотный мотив X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо

выбирают из гистидина или аспарагина; и/или

d. аминокислотный мотив X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и/или

e. аминокислотный мотив X_9FY$X_{10}X_{11}$REQX_{12}KEX_{13} [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

Полипептид в контексте настоящего изобретения можно рассматривать как фрагмент полноразмерного белка Cas. Такие фрагменты могут быть неактивными и быть использованы в способах и с целями, которые не связаны непосредственно со связыванием, редактированием и/или разрезанием генетического материала, например, в качестве стандартов в анализах или для вызывания выработки антител, или тому подобного.

Однако в предпочтительных вариантах осуществления белок или полипептид Cas является функциональным и способен к расщеплению, связыванию, маркированию или модификации ДНК при температуре в диапазоне от 20°C до 100°C, включительно, когда он связан с по меньшей мере одной молекулой направляющей РНК и полинуклеотидом, содержащим нуклеотидную последовательность-мишень, узнаваемую молекулой направляющей РНК. Предпочтительно, белок или полипептид Cas является функциональным и способным к указанному расщеплению, связыванию, маркированию или модификации при температуре в диапазоне от 50°C до 70°C, например, 55°C или 60°C.

В конкретных вариантах осуществления изобретения может быть предложен белок или полипептид Cas, содержащий аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]. В других вариантах осуществления белки или полипептиды Cas могут дополнительно содержать аминокислотный мотив X_1X_2CTX_3X_4 [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина.

В других вариантах осуществления белки или полипептиды Cas, описанные в настоящем документе, могут дополнительно содержать аминокислотный мотив X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина.

В других вариантах осуществления белки или полипептиды Cas, описанные в настоящем документе, могут дополнительно содержать аминокислотный мотив X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

В других вариантах осуществления белки или полипептиды Cas, описанные в настоящем документе, могут дополнительно содержать аминокислотный мотив $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

Должно быть понятно, что в соответствии с настоящим изобретением белок или полипептид Cas по изобретению может содержать любой из мотивов SEQ ID NO: 2-6, либо отдельно, либо в сочетании. Нижеследующее суммирует каждое из сочетаний мотивов, которые могут характеризовать белки или полипептиды Cas по изобретению:

EKDGKYYC [SEQ ID NO: 2].

EKDGKYYC [SEQ ID NO: 2] и $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина.

EKDGKYYC [SEQ ID NO: 2] и $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или

аспарагина.

EKDGKYYC [SEQ ID NO: 2] и $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

EKDGKYYC [SEQ ID NO: 2] и $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

EKDGKYYC [SEQ ID NO: 2] и $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой

лизин или серин.

EKDGKYYC [SEQ ID NO: 2] и $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин, и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

EKDGKYYC [SEQ ID NO: 2] и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

EKDGKYYC [SEQ ID NO: 2] и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина.

EKDGKYYC [SEQ ID NO: 2] и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

EKDGKYYC [SEQ ID NO: 2] и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

EKDGKYYC [SEQ ID NO: 2] и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5],

где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

EKDGKYYC [SEQ ID NO: 2] и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

EKDGKYYC [SEQ ID NO: 2] и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

$X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина.

$X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

$X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина,

глутамата или аргинина; и X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

$X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

$X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

$X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина.

X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин.

В другом аспекте настоящее изобретение относится к выделенному белку Cas или его полипептидному фрагменту, имеющему аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%. Предпочтительно, белок или полипептид Cas способен к связыванию, расщеплению, маркированию или модификации при температуре в диапазоне от 20°C до 100°C, включительно. Предпочтительно, белок

или полипептид Cas способен к указанному связыванию, расщеплению, маркированию или модификации при температуре в диапазоне от 50°C до 70°C, например, 55°C или 60°C. Предпочтительно, белок или полипептид Cas способен к указанному связыванию, расщеплению, маркированию или модификации при температуре в диапазоне от 30°C до 80°C, при температуре от 37°C до 78°C, предпочтительно, при температуре выше 55°C; более предпочтительно, при температуре от 55°C до 80°C; еще более предпочтительно, при температуре от 55°C до 65°C или от 60°C до 65°C.

Настоящее изобретение также относится к применению молекулы направляющей РНК и белка или полипептида Cas, предложенных в настоящем документе, для связывания, расщепления, маркирования или модификации полинуклеотида-мишени, содержащего нуклеотидную последовательность-мишень. Молекула направляющей РНК узнает нуклеотидную последовательность-мишень на нуклеотидной цепи-мишени полинуклеотида.

Полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, может быть двухцепочечным и, таким образом, содержать нуклеотидную цепь-мишень, содержащую указанную нуклеотидную последовательность-мишень, и не являющуюся мишенью нуклеотидную цепь, содержащую протоспейсерную нуклеотидную последовательность. Протоспейсерная нуклеотидная последовательность в значительной степени комплементарна нуклеотидной последовательности-мишени и образует с ней пару в двухцепочечном полинуклеотиде-мишени. Не являющаяся мишенью нуклеотидная цепь также может содержать последовательность мотива, примыкающего к протоспейсеру (PAM), в непосредственной близости к 3′-концу протоспейсерной последовательности. Последовательность PAM может иметь длину по меньшей мере 6, 7, или 8 нуклеотидов. Предпочтительно, последовательность PAM имеет цитозин в пятом положении. Предпочтительно, последовательность PAM содержит последовательность 5′-NNNNC-3′, так что с 5′-конца последовательность PAM начинается 5′-NNNNC-3′. Дополнительно или

альтернативно, последовательность PAM может иметь аденин в восьмом положении, так что последовательность PAM содержит последовательность 5′-NNNNNNNA-3′, и с 5′-конца последовательность PAM начинается 5′-NNNNNNNA-3′. Дополнительно или альтернативно, последовательность PAM может иметь цитозин в одном или более из первого, второго, третьего, четвертого и шестого положений, так что с 5′-конца последовательность PAM начинается 5′-CNNNN-3′, 5′-NCNNN-3′, 5′-NNCNN-3′, 5′-NNNCN-3′ и/или 5′-NNNNNC-3′. Предпочтительно, последовательность PAM содержит, так что с 5′-конца последовательность PAM начинается, 5′-CCCCCCNA-3′ [SEQ ID NO: 10] и, кроме того, предпочтительно, последовательность PAM содержит, так что с 5′-конца последовательность PAM начинается, 5′-CCCCCCAA-3′ [SEQ ID NO: 11]. Другие предпочтительные последовательности PAM включают 5′-ATCCCCAA-3′ [SEQ ID NO: 21] и 5′-ACGGCCAA-3′ [SEQ ID NO: 22].

Предпочтительно, белок или полипептид Cas способен к связыванию, расщеплению, маркированию или модификации при температуре в диапазоне от 40°C до 80°C, включительно, предпочтительно, в диапазоне от 45°C до 80°C, включительно, и более предпочтительно, в диапазоне от 50°C до 80°C, включительно. Например, связывание, расщепление, маркирование или модификация происходит при температуре 45°C, 46°C, 47°C, 48°C, 49°C, 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, 60°C, 61°C, 62°C, 63°C, 64°C, 65°C, 66°C, 67°C, 68°C, 69°C, 70°C, 71°C, 72°C, 73°C, 74°C, 75°C, 76°C, 77°C, 78°C, 79°C или 80°C. Более предпочтительно, белок или полипептид Cas способен к связыванию, расщеплению, маркированию или модификации при температуре в диапазоне от 55 до 65°C. В предпочтительных аспектах белок или полипептидный фрагмент Cas по изобретению может содержать аминокислотную последовательность, идентичную по меньшей мере на 75%; предпочтительно, по меньшей мере на 85%; более предпочтительно, по меньшей мере на 90%; еще более предпочтительно, по меньшей мере на 95% последовательности SEQ

ID NO: 1.

Белок или полипептид Cas может быть использован в сочетании с молекулой направляющей РНК, которая узнает нуклеотидную последовательность-мишень на нуклеотидной цепи-мишени, при этом не являющаяся мишенью нуклеотидная последовательность имеет последовательность PAM, непосредственно примыкающую к 3′-концу протоспейсерной последовательности на не являющейся мишенью цепи, как описано в настоящем документе. Таким образом, последовательность PAM может содержать последовательность 5′-NNNNC-3′, и белок Cas может связывать, расщеплять, маркировать или модифицировать цепь-мишень при температуре в диапазоне от 20°C до 100°C, включительно, предпочтительно, в диапазоне от 30°C до 90°C, включительно, в диапазоне от 37°C до 78°C, включительно, в диапазоне от 40°C до 80°C, включительно, в диапазоне от 50°C до 70°C, включительно, или в диапазоне от 55°C до 65°C, включительно. Предпочтительно, с 5′-конца последовательность PAM начинается 5′-NNNNC-3′, и белок Cas может связывать, расщеплять, маркировать или модифицировать цепь-мишень при температуре в диапазоне от 20°C до 100°C, включительно, предпочтительно, в диапазоне от 30°C до 90°C, включительно, в диапазоне от 37°C до 78°C, включительно, в диапазоне от 40°C до 80°C, включительно, в диапазоне от 50°C до 70°C, включительно, или в диапазоне от 55°C до 65°C, включительно. Предпочтительно, с 5′-конца последовательность PAM начинается 5′-NNNNNNNA-3′, и белок Cas может связывать, расщеплять, маркировать или модифицировать цепь-мишень при температуре в диапазоне от 20°C до 100°C, включительно, предпочтительно, в диапазоне от 30°C до 90°C, включительно, в диапазоне от 37°C до 78°C, включительно, в диапазоне от 40°C до 80°C, включительно, в диапазоне от 50°C до 70°C, включительно, или в диапазоне от 55°C до 65°C, включительно. Кроме того, предпочтительно, с 5′-конца последовательность PAM начинается 5′-NNNNCNNA-3′, и белок Cas может связывать, расщеплять, маркировать или модифицировать

цепь-мишень при температуре в диапазоне от 20°C до 100°C, включительно, предпочтительно, в диапазоне от 30°C до 90°C, включительно, в диапазоне от 37°C до 78°C, включительно, в диапазоне от 40°C до 80°C, включительно, в диапазоне от 50°C до 70°C, включительно, или в диапазоне от 55°C до 65°C, включительно.

Более конкретно, белок или полипептид Cas по изобретению может содержать аминокислотную последовательность, имеющую следующий процент идентичности с SEQ ID NO: 1: по меньшей мере 60%, по меньшей мере 61%, по меньшей мере 62%, по меньшей мере 63%, по меньшей мере 64%, по меньшей мере 65%, по меньшей мере 66%, по меньшей мере 67%, по меньшей мере 68%, по меньшей мере 69%, по меньшей мере 70%, по меньшей мере 71%, по меньшей мере 72%, по меньшей мере 73%, по меньшей мере 74%, по меньшей мере 75%, по меньшей мере 76%, по меньшей мере 77%, по меньшей мере 78%, по меньшей мере 79%, по меньшей мере 80%, по меньшей мере 81%, по меньшей мере 82%, по меньшей мере 83%, по меньшей мере 84%, по меньшей мере 85%, по меньшей мере 86%, по меньшей мере 87%, по меньшей мере 88%, по меньшей мере 89%, по меньшей мере 90%, по меньшей мере 91%, по меньшей мере 92%, по меньшей мере 93%, по меньшей мере 94%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98%, по меньшей мере 99%, по меньшей мере 99,5% или по меньшей мере 99,8%. Процент идентичности может составлять по меньшей мере 89%. Процент идентичности может составлять по меньшей мере 90%. Предпочтительно, процент идентичности будет составлять по меньшей мере 95%, например, 98%.

Процент идентичности аминокислотной последовательности с SEQ ID NO: 1 можно определять на основании числа идентичных положений, общих для последовательностей в выбранном окне сравнения, с учетом количества пробелов и длины каждого пробела, который необходимо ввести для оптимального выравнивания двух последовательностей.

Белок или полипептидный фрагмент Cas по изобретению может быть охарактеризован в контексте как эталонной

последовательности SEQ ID NO: 1, так и любого ее вышеупомянутого варианта с указанным процентом идентичности последовательности, отдельно или в сочетании с любым из вышеупомянутых аминокислотных мотивов (то есть, SEQ ID NOS 2 и/или 3, и/или 4, и/или 5, и/или 6) в качестве основных признаков.

Изобретение относится к применению молекулы направляющей РНК, предложенной в настоящем документе, а также белка или полипептида Cas по изобретению для связывания, расщепления, маркирования или модификации нуклеотидной цепи-мишени, содержащей нуклеотидную последовательность-мишень. Предпочтительно, указанное связывание, расщепление, маркирование или модификация происходит при температуре, указанной в настоящем документе, например, при температуре от 20 до 100°C. Изобретение также относится к способу связывания, расщепления, маркирования или модификации нуклеотидной последовательности-мишени в нуклеотидной цепи-мишени, включающему конструирование молекулы направляющей РНК, предложенной в настоящем документе, и формирование рибонуклеопротеидного комплекса, содержащего молекулу направляющей РНК и белок или полипептид Cas по изобретению. Предпочтительно, рибонуклеопротеидный комплекс осуществляет связывание, расщепление, маркирование или модификацию нуклеотидной последовательности-мишени при температуре, указанной в настоящем документе, например, при температуре от 37 до 100°C.

Варианты применения и способы по изобретению можно осуществлять, и нуклеопротеиды по изобретению можно создавать и использовать, *in vivo*, например, в бактериальных клетках. Альтернативно, варианты применения и способы по изобретению можно осуществлять, и нуклеопротеиды по изобретению можно создавать и использовать, *in vitro*. Белок Cas по изобретению может быть предоставлен в выделенной форме, например, при использовании *in vitro* или при введении в клетки путем трансфекции, белок Cas может быть гетерологично экспрессирован, например, после временной или стабильной трансформации клетки нуклеиновой кислотой, кодирующей белок Cas, молекула

направляющей РНК может быть транскрибирована с экспрессионного вектора после временной или стабильной трансформации клетки нуклеиновой кислотой, кодирующей молекулу РНК, и/или молекула РНК может быть предоставлена в выделенной форме, например, при использовании *in vitro* или при введении в клетки путем трансфекции. В предпочтительных вариантах осуществления белок или полипептид Cas экспрессируется с генома клетки-хозяина после стабильной интеграции нуклеиновой кислоты, кодирующей белок или полипептид Cas, в геноме клетки-хозяина. Таким образом, белок Cas и/или молекулу РНК можно добавлять в *in vivo* или *in vitro* среду с использованием любого искусственного или изобретенного способа введения молекулы белка или нуклеиновой кислоты в клетку, в которой она иначе не присутствует.

Полинуклеотид, содержащий нуклеотидную последовательность-мишень, может быть расщеплен белком Cas и, необязательно, расщепление может представлять собой расщепление ДНК. Нуклеотидная цепь-мишень, содержащая последовательность-мишень, может представлять собой двухцепочечную ДНК, и способ, или вариант применения, может приводить к образованию двухцепочечного разрыва в полинуклеотиде, содержащем нуклеотидную последовательность-мишень. Полинуклеотид, содержащий нуклеотидную последовательность-мишень, может представлять собой двухцепочечную ДНК, белок Cas может быть лишен способности разрезать двухцепочечную ДНК, и вариант применения или способ может приводить к выключению гена полинуклеотида.

Белок или полипептид Cas может быть предоставлен для способов, вариантов применения и нуклеопротеидов по изобретению в концентрации 250 нМ или менее, например, в концентрации 200 нМ или менее, 150 нМ или менее, 100 нМ или менее, 50 нМ или менее, 25 нМ или менее, 10 нМ или менее, 5 нМ или менее, 1 нМ или менее, или 0,5 нМ или менее. Альтернативно, белок или полипептид Cas может быть предоставлен в концентрации по меньшей мере 0,5 нМ, по меньшей мере 1 нМ, по меньшей мере 5 нМ, по меньшей мере 10 нМ, по меньшей мере 25 нМ, по меньшей мере 50 нМ, по меньшей мере 100 нМ, по меньшей мере 150 нМ, по меньшей мере 200 нМ или

по меньшей мере 250 нМ. Последовательность PAM по изобретению может иметь аденин в восьмом положении, так что последовательность PAM содержит последовательность 5′-NNNNNNNA-3′, и концентрация белка или полипептида Cas может составлять 100 нМ или менее, 50 нМ или менее, 25 нМ или менее, 10 нМ или менее, 5 нМ или менее, 1 нМ или менее или 0,5 нМ или менее. Последовательность PAM может содержать последовательность 5′-NNNNCNNA-3′, и концентрация белка или полипептида Cas может составлять 100 нМ или менее, 50 нМ или менее, 25 нМ или менее, 10 нМ или менее, 5 нМ или менее, 1 нМ или менее, или 0,5 нМ или менее. Последовательность PAM может содержать последовательность 5′-CCCCCCNA-3′ [SEQ ID NO: 10], и концентрация белка или полипептида Cas может составлять 100 нМ или менее, 50 нМ или менее, 25 нМ или менее, 10 нМ или менее, 5 нМ или менее, 1 нМ или менее, или 0,5 нМ или менее.

Кроме того, изобретение относится к нуклеиновым кислотам, кодирующим любой из вышеуказанных белков или полипептидов по изобретению. Нуклеиновые кислоты могут быть выделенными или находиться в форме экспрессионных конструктов.

Во всех вышеупомянутых аспектах настоящего изобретения аминокислотные остатки могут быть заменены консервативно или не консервативно. Консервативные аминокислотные замены представляют собой такие замены, когда аминокислотные остатки заменены другими аминокислотными остатками со сходными химическими свойствами (например, зарядом или гидрофобностью) и, следовательно, функциональные свойства полученного полипептида не изменяются.

Аналогично, специалисту в данной области должно быть понятно, что в последовательностях нуклеиновых кислот могут иметь место консервативные или не консервативные замены, не затрагивающие функцию полипептида. Консервативно модифицированные нуклеиновые кислоты представляют собой такие нуклеиновые кислоты с заменами, которые кодируют идентичные или функционально идентичные варианты аминокислотных последовательностей. Специалисту должно быть понятно, что каждый

кодон в нуклеиновой кислоте (за исключением AUG и UGG, как правило, единственных кодонов для метионина или триптофана, соответственно) может быть модифицирован с получением функционально идентичной молекулы. Соответственно, каждый молчащий вариант (то есть, с синонимичными кодонами) полинуклеотида, кодирующего полипептид по настоящему изобретению, подразумевается для каждой описанной полипептидной последовательности.

Изобретение относится к трансформированной клетке, имеющей нуклеотидную последовательность-мишень в двухцепочечном полинуклеотиде-мишени, при этом указанная клетка содержит белок или полипептид Cas, описанный в настоящем документе, и по меньшей мере одну молекулу направляющей РНК, описанную в настоящем документе, а также экспрессионный вектор, содержащий нуклеиновую кислоту, кодирующую по меньшей мере одно из указанного белка Cas и указанной молекулы направляющей РНК. Белок Cas и молекула направляющей РНК могут создавать, или допускать, возможность того, что связывание, расщепление, маркирование или модификация последовательности-мишени происходит в трансформированной клетке при повышенной температуре, или в диапазоне температур, например, от 37 до 100°C, как описано в настоящем документе. Изобретение также относится к способу связывания, расщепления, маркирования или модификации нуклеиновой кислоты-мишени в клетке, включающему либо 1) трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором, содержащим нуклеотидную последовательность, кодирующую белок или полипептид Cas по изобретению, и нуклеотидную последовательность, кодирующую молекулу направляющей РНК по изобретению; либо 2) трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором, содержащим нуклеотидную последовательность, кодирующую белок или полипептид Cas по изобретению, и дополнительным экспрессионным вектором, содержащим нуклеотидную последовательность, кодирующую молекулу направляющей РНК по изобретению; либо 3) трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором,

содержащим нуклеотидную последовательность, кодирующую белок или полипептид Cas по изобретению, и доставку молекулы направляющей РНК, описанной в настоящем документе, к клетке или в клетку. Белок или полипептид Cas может экспрессироваться с генома трансформированной клетки, например, после стабильной интеграции в геном нуклеотидной последовательности, кодирующей белок или полипептид Cas.

Изобретение также относится к наборам, включающим один или более реагентов для осуществления вариантов применения и способов по изобретению, или для получения трансформированных клеток или нуклеопротеидного комплекса по изобретению, при этом указанные наборы включают: белок или полипептид Cas по изобретению, или экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую белок или полипептид Cas по изобретению; и/или молекулу направляющей РНК по изобретению, или экспрессионный вектор, содержащий нуклеотидную последовательность, кодирующую молекулу направляющей РНК по изобретению. Наборы также могут включать инструкции по применению на практике изобретения, например, инструкции по конструированию молекулы направляющей РНК по изобретению.

Гид-РНК и последовательности-мишени

Белки Cas по изобретению позволяют проводить специфичное для последовательности связывание, расщепление, мечение, маркирование или модификацию нуклеиновых кислот-мишеней при повышенных температурах. Нуклеиновые кислоты-мишени могут представлять собой ДНК (одноцепочечные или двухцепочечные), РНК или синтетические нуклеиновые кислоты. Особенно полезным применением настоящего изобретения является специфичное для последовательности направленное воздействие и модификация геномной ДНК одним или более белками Cas по изобретению в комплексе с одной или более гид-РНК (гРНК), которые комплементарно связываются с последовательностью-мишенью геномной ДНК. Следовательно, нуклеиновая кислота-мишень, предпочтительно, представляет собой двухцепочечную ДНК. Такое направленное действие можно осуществлять *in vitro* или *in vivo*. Предпочтительно, такое направленное действие осуществляют *in*

vivo. Таким образом, белки Cas по изобретению могут быть использованы для направленного воздействия и модификации специфических последовательностей ДНК, расположенных в геномной ДНК клетки. Предполагается, что систему Cas можно использовать для модификации геномов в клетках различных типов и/или у разных организмов.

Молекулы гРНК, также называемые молекулами направляющей РНК, узнают нуклеотидную последовательность-мишень на цепи-мишени полинуклеотида. Молекулы РНК могут быть сконструированы для узнавания последовательности-мишени в двухцепочечном полинуклеотиде-мишени, при этом не являющаяся мишенью цепь содержит последовательность мотива, примыкающего к протоспейсеру (PAM), в непосредственной близости к 3'-концу протоспейсерной последовательности. В настоящем документе раскрыты последовательности PAM, которые действуют оптимальным образом с белками и полипептидами Cas по изобретению. Со знанием этих последовательностей PAM можно конструировать гРНК для использования с белками и полипептидами Cas по изобретению в диапазонах температур и при повышенных температурах по изобретению.

Соответственно, настоящее изобретение относится к рибонуклеопротеидному комплексу, содержащему белок или полипептид Cas по изобретению, описанный в настоящем документе выше, и дополнительно содержащему по меньшей мере одну молекулу РНК, которая имеет направляющую функцию в том смысле, что она узнает конкретную нуклеотидную последовательность в полинуклеотиде-мишени. Настоящее изобретение также относится к применению по меньшей мере одной молекулы направляющей РНК и белка или полипептида Cas для связывания, расщепления, маркирования или модификации нуклеотидной цепи-мишени, к способу связывания, расщепления, маркирования или модификации нуклеотидной последовательности-мишени в нуклеотидной цепи-мишени с использованием рибонуклеопротеида или нуклеопротеида по изобретению, а также к трансформированным клеткам, содержащим белок или полипептид Cas и молекулу направляющей РНК.

Полинуклеотид-мишень также может содержать определенную последовательность PAM в непосредственной близости к 3'-концу протоспейсерной последовательности, такую как последовательность PAM, приведенная в настоящем документе. Последовательность PAM может иметь длину 6, 7 или 8, или более, нуклеотидов, предпочтительно, длину 8 нуклеотидов. Предпочтительно, молекула РНК представляет собой одноцепочечную молекулу РНК, например, РНК CRISPR (crРНК), и связана, например, путем гибридизации, с tracrРНК. Направляющая РНК может представлять собой химеру crРНК и tracrРНК. Вышеупомянутые молекулы РНК могут иметь рибонуклеотидную последовательность, по меньшей мере на 90% идентичную, или комплементарную, нуклеотидной последовательности-мишени. Необязательно, молекула РНК имеет рибонуклеотидную последовательность, идентичную, или комплементарную, по меньшей мере на 90%, по меньшей мере на 91%, по меньшей мере на 92%, по меньшей мере на 93%, по меньшей мере на 94%, по меньшей мере на 95%, по меньшей мере на 96%, по меньшей мере на 97%, по меньшей мере на 98%, по меньшей мере на 99% или на 100% нуклеотидной последовательности-мишени. Предпочтительной нуклеотидной последовательностью-мишенью является ДНК.

В предпочтительном аспекте настоящее изобретение относится к рибонуклеопротеидному комплексу, описанному в настоящем документе выше, в котором по меньшей мере одна молекула направляющей РНК в значительной степени комплементарна по всей своей длине последовательности ДНК-мишени.

Молекула направляющей РНК может быть связана, или ассоциирована, с последовательностью-мишенью в составе нуклеопротеидного комплекса так, что полинуклеотид-мишень, содержащий последовательность-мишень и последовательность PAM на не являющейся мишенью цепи, может быть связан и, таким образом, являться частью нуклеопротеидного комплекса по изобретению.

Следовательно, изменение последовательности гид-РНК, которая связывается с белком Cas по изобретению, позволяет программировать белок Cas для маркирования или разрезания

двухцепочечной ДНК в сайтах, комплементарных гид-РНК.

Предпочтительно, длина по меньшей мере одной молекулы направляющей РНК в рибонуклеопротеидном комплексе по изобретению находится в диапазоне от 35 до 135 остатков, необязательно, в диапазоне от 35 до 134 остатков, от 35 до 133 остатков, от 35 до 132 остатков, от 35 до 131 остатков, от 35 до 130 остатков, от 35 до 129 остатков, от 35 до 128 остатков, от 35 до 127 остатков, от 35 до 126 остатков, от 35 до 125 остатков, от 35 до 124 остатков, от 35 до 123 остатков, от 35 до 122 остатков, от 35 до 121 остатка от 35 до 120 остатков, от 35 до 119 остатков, от 35 до 118 остатков, от 35 до 117 остатков, от 35 до 116 остатков, от 35 до 115 остатков, от 35 до 114 остатков, от 35 до 113 остатков, от 35 до 112 остатков, от 35 до 111 остатков, от 35 до 100 остатков, от 35 до 109 остатков, от 35 до 108 остатков, от 35 до 107 остатков, от 35 до 106 остатков, от 35 до 105 остатков, от 35 до 104 остатков, от 35 до 103 остатков, от 35 до 102 остатков, от 35 до 101 остатка, от 35 до 100 остатков, от 35 до 99 остатков, от 35 до 98 остатков, от 35 до 97 остатков, от 35 до 96 остатков, от 35 до 95 остатков, от 35 до 94 остатков, от 35 до 93 остатков, от 35 до 92 остатков, от 35 до 91 остатка, от 35 до 90 остатков, от 35 до 89 остатков, от 35 до 88 остатков, от 35 до 87 остатков, от 35 до 86 остатков, от 35 до 85 остатков, от 35 до 84 остатков, от 35 до 83 остатков, от 35 до 82 остатков, от 35 до 81 остатка, от 35 до 80 остатков, от 35 до 79 остатков, от 35 до 78 остатков, от 35 до 77 остатков, от 35 до 76 остатков, от 35 до 75 остатков, от 35 до 74 остатков, от 35 до 73 остатков, от 35 до 72 остатков, от 35 до 71 остатка, от 35 до 70 остатков, от 35 до 69 остатков, от 35 до 68 остатков, от 35 до 67 остатков, от 35 до 66 остатков, от 35 до 65 остатков, от 35 до 64 остатков, от 35 до 63 остатков, от 35 до 62 остатков, от 35 до 61 остатка, от 35 до 60 остатков, от 35 до 59 остатков, от 35 до 58 остатков, от 35 до 57 остатков, от 35 до 56 остатков, от 35 до 55 остатков, от 35 до 54 остатков, от 35 до 53 остатки, от 35 до 52 остатков, от 35 до 51 остатка, от 35 до 50 остатков, от 35 до 49 остатков, от 35 до 48 остатков, от 35 до 47 остатков, от 35 до 46 остатков, от 35

до 45 остатков, от 35 до 44 остатков, от 35 до 43 остатки, от 35 до 42 остатков, от 35 до 41 остатка, от 35 до 40 остатков, от 35 до 39 остатков, от 35 до 38 остатков, от 35 до 37 остатков, от 35 до 36 остатков или 35 остатков. Предпочтительно, длина по меньшей мере одной молекулы РНК находится в диапазоне от 36 до 174 остатков, от 37 до 173 остатков, от 38 до 172 остатков, от 39 до 171 остатка, от 40 до 170 остатков, от 41 до 169 остатков, от 42 до 168 остатков, от 43 до 167 остатков, от 44 до 166 остатков, от 45 до 165 остатков, от 46 до 164 остатков, от 47 до 163 остатков, от 48 до 162 остатков, от 49 до 161 остатка, от 50 до 160 остатков, от 51 до 159 остатков, от 52 до 158 остатков, от 53 до 157 остатков, от 54 до 156 остатков, от 36 до 74 остатков, от 37 до 73 остатков, от 38 до 72 остатков, от 39 до 71 остатка, от 40 до 70 остатков, от 41 до 69 остатков, от 42 до 68 остатков, от 43 до 67 остатков, от 44 до 66 остатков, от 45 до 65 остатков, от 46 до 64 остатков, от 47 до 63 остатков, от 48 до 62 остатков, от 49 до 61 остатка, от 50 до 60 остатков, от 51 до 59 остатков, от 52 до 58 остатков, от 53 до 57 остатков, от 54 до 56 остатков.

В предпочтительных аспектах настоящее изобретение относится к рибонуклеопротеидному комплексу, в котором комплементарная часть по меньшей мере одной молекулы РНК имеет длину по меньшей мере 30 остатков. Альтернативно, комплементарная часть по меньшей мере одной молекулы РНК может иметь длину 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74 или 75 остатков.

Предпочтительно, молекула направляющей РНК будет удовлетворять требованиям высокой специфичности и аффинности к последовательности нуклеиновой кислоты-мишени. Желательна константа диссоциации (K_d) в диапазоне от 1 мкМ до 1 пМ, предпочтительно, от 1 нМ до 1 пМ, более предпочтительно, 1-100 пМ, при определении методом нативного гель-электрофореза или, альтернативно, методами изотермической калориметрии титрования, поверхностного плазмонного резонанса или методами титрования на основе флуоресценции. Аффинность можно определять с

использованием анализа сдвига электрофоретической подвижности (EMSA), также называемого анализом задержки в геле (смотри Semenova *et al*. (2011) PNAS 108: 10098-10103).

Молекулу направляющей РНК, предпочтительно, моделируют на основании того, что известно в природе у прокариотов как молекулы РНК CRISPR (crРНК). Структура молекул crРНК уже установлена и описана более подробно в публикации Jore *et al*., 2011, Nature Structural & Molecular Biology 18: 529-537. Вкратце, зрелая crРНК типа I-E часто имеет длину 61 нуклеотид и состоит из 5′-области «рукоятки» из 8 нуклеотидов, «спейсерной» последовательности из 32 нуклеотидов и 3'-последовательности из 21 нуклеотида, которые образуют шпильку с тетрануклеотидной петлей (фиг. 5). Системы типа I отличаются от систем типа II (Cas9), и подробности различных систем описаны в публикации Van der Oost 2014 Nat Rev Micr 12: 479-492. В системах типа II (Cas9) существует иной механизм процессинга, в котором задействованы вторая РНК (tracrРНК) и две рибонуклеазы. Зрелая crРНК в системе типа II, вместо шпильки, остается присоединенной к фрагменту tracrРНК (фиг. 5). Однако РНК, используемая по изобретению, не обязательно должна быть создана строго по дизайну природной crРНК, будь то в отношении длины, областей или конкретных последовательностей РНК. Однако понятно, что молекулы РНК для использования по изобретению могут быть спроектированы на основе информации о последовательностях генов, имеющихся в общедоступных базах данных, или недавно открытых, а затем искусственно созданы, например, полностью или частично путем химического синтеза. Молекулы РНК по изобретению также могут быть спроектированы и получены путем экспрессии в генетически модифицированных клетках или в бесклеточных экспрессионных системах, и этот вариант может включать синтез части, или всей, последовательности РНК.

Структура и требования к crРНК в системе типа II (Cas9) также описаны в публикации Jinek *et al*., 2012, там же. В системе типа I имеется так называемая «затравочная» часть, которая образует 5′-конец спейсерной последовательности, и которая

фланкирована на ее 5′-конце 5′-рукояткой из 8 нуклеотидов. Semenova *et al*. (2011, PNAS 108: 10098-10103) обнаружили, что все остатки последовательности затравки должны быть комплементарны последовательности-мишени, хотя для остатка в положении 6 может быть допущено несоответствие (фиг. 5). В системе типа II имеется последовательность затравки из 10-12 нуклеотидов, которая расположена на 3'-конце спейсера (фиг. 5) (обзор Van der Oost 2014, там же). Аналогично, при проектировании и создании компонента РНК рибонуклеопротеидного комплекса по изобретению, направленного на локус-мишень (то есть, на последовательность), могут применяться необходимые правила соответствия и несоответствия для последовательности затравки системы типа II.

Таким образом, изобретение включает способ обнаружения и/или локализации изменения единственного основания в молекуле нуклеиновой кислоты-мишени, включающий создание контакта образца нуклеиновой кислоты с рибонуклеопротеидным комплексом по изобретению, как описано в настоящем документе выше, или с белком или полипептидом Cas и отдельным компонентом направляющей РНК по изобретению, как описано в настоящем документе выше, и при этом последовательность направляющей РНК (в том числе, когда она находится в рибонуклеопротеидном комплексе) является такой, что она различает нормальный аллель и мутантный аллель в силу изменения единственного основания, например, в положении 6 непрерывной последовательности из 8 нуклеотидных остатков.

Без связи с конкретной теорией, правило конструирования, которое может быть использовано при получении компонента направляющей РНК рибонуклеопротеидных комплексов по изобретению, включает так называемую последовательность «PAM» (мотив, примыкающий к протоспейсеру) в двухцепочечном полинуклеотиде-мишени. В системе типа I-E *E.coli* последовательность PAM может представлять собой консервативный триплет нуклеотидных остатков, например, 5′-CTT-3′, 5′-CAT-3′, 5′-CCT-3′, 5′-CAC-3′, 5′-TTT-3′, 5′-ATT-3′ и 5′-AWG-3′, где W представляет собой A, T или U. В системе типа I последовательность PAM, расположенная в цепи-мишени, как

правило, находится в положении, соответствующем 5'-концу затравки. Однако в системе типа II PAM расположена на другом конце, на замещенной, или не являющейся мишенью, цепи вблизи 3′-конца спейсера crРНК, в положении, соответствующем 3'-концу затравки (фиг. 5) (Jinek et al., 2012, в цитируемой работе). В случае Cas9 *Streptococcus pyogenes* последовательность PAM имеет консервативную пару нуклеотидных остатков 5'-NGG-3'. Недавно были охарактеризованы различные варианты Cas9 (тип IIA и тип IIC) (Ran *et al.*, 2015 Nature 520:186-191) -фиг. 1A), и были обнаружены PAM (смотри Ran *et al.*, 2015, там же - фиг. 1C). Известные в настоящее время PAM Cas9 включают: тип IIA 5′-NGGNNNN-3′ (*Streptococcus pyogenes*), 5′-NNGTNNN-3′ (*Streptococcus pasteurianus*), 5′-NNGGAAN-3′ (*Streptococcus thermophilus*), 5′-NNGGGNN-3′ (*Staphylococcus aureus*), и тип IIC 5′-NGGNNNN-3′ (*Corynebacterium difteriae*), 5′-NNGGGTN-3′ (*Campylobacter lari*), 5′-NNNCATN-3′ (*Parvobaculum lavamentivorans*), 5′-NNNNGTA-3′ (*Neiseria cinerea*). Cas9 *Geobacillus thermodenitrificans* T12 (настоящее изобретение) принадлежит к типу Type IIC (Ran *et al.*, 2015, там же). Авторы изобретения неожиданно обнаружили, что выбор последовательностей PAM для использования по изобретению может влиять на температуру, при которой белки и полипептиды Cas по изобретению будут взаимодействовать с последовательностью-мишенью. В частности, авторы изобретения обнаружили, что преимущественно 8-мерная последовательность PAM обеспечивает активность в широком диапазоне температур в случае наличия цитозина в 5-м положении после 3′-конца последовательности-мишени, и/или аденина в 8-м положении. Также предпочтительно нахождение цитозина в 1-м, 2-м, 3-м, 4-м и/или 6-м положении последовательности PAM после 3′-конца протоспейсерной последовательности.

В вариантах осуществления изобретения молекула направляющей РНК может иметь длину в диапазоне 35-75 остатков. В предпочтительных вариантах осуществления часть РНК, которая комплементарна и используется для направления на желаемую

нуклеотидную последовательность, имеет длину от 15 до 32 остатков. В контексте природной crРНК это должно соответствовать спейсерной части, как показано, например, на фиг. 1 публикации Semenova *et al.* (2011, там же).

Рибонуклеопротеидный комплекс по изобретению может иметь направляющий компонент, содержащий 8 остатков, происходящих из повтора CRISPR, в 5′-направлении для последовательности РНК, который имеет существенную комплементарность с последовательностью-мишенью ДНК. Последовательность РНК, имеющая комплементарность с последовательностью-мишенью ДНК, как должно быть понятно, в контексте crРНК соответствует спейсерной последовательности. 5'-фланкирующая последовательность РНК должна рассматриваться, как соответствующая 5'-рукоятке crРНК; как показано, например, на фигуре 1 в публикации Semenova *et al.* (2011, там же).

Рибонуклеопротеидный комплекс по изобретению может иметь последовательность, образующую шпильку и тетрануклеотидную петлю, в 3'-положении относительно последовательности направляющей РНК, которая комплементарна последовательности-мишени ДНК, то есть, в 3'-положении относительно того, что должно соответствовать 3'-рукоятке, фланкирующей спейсерную последовательность в crРНК; например, как показано на фигуре 1 в публикации Semenova *et al.* (2011, там же).

Без связи с конкретной теорией, в предпочтительном рибонуклеопротеидном комплексе и двухцепочечном полинуклеотиде-мишени не являющаяся мишенью нуклеотидная цепь, которая не спаривается с направляющей РНК рибонуклеопротеидного комплекса, может содержать непосредственно примыкающую к 3′-концу последовательность PAM, выбранную из одной или более из 5′-NNNNCNNA-3′, 5′-CNNNCNN-3′, 5′-NNNCCNN-3′, 5′-NNCNCNN-3′, 5′-NNNNCCN-3′ и 5′-NCNNCNN-3′. Предпочтительно, последовательность PAM можно выбирать из 5′-NNNNC-3′, 5′-NNNNCNNA-3′, 5′-CNNNC-3′, 5′-CNNNCNNA-3′, 5′-NCNNC-3′, 5′-NCNNCNNA-3′, 5′-NNCNC-3′, 5′-NNCNCNNA-3′, 5′-NNNCC-3′, 5′-NNNCCNNA-3′, 5′-NNNNCC-3′, 5′-NNNNCCNA-3′, 5′-CCNNC-3′, 5′-

CCNNCNNA-3′, 5′-CNCNC-3′, 5′-CNCNCNNA-3′, 5′-CNNCCN-3′, 5′-CNNCCNNA-3′, 5′-CNNNCC-3′, 5′-CNNNCCNA-3′, 5′-CCCNCN-3′, 5′-CCCNCNNA-3′, 5′-CCNCCN-3′, 5′-CCNCCNNA-3′, 5′-CCNNCC-3′, 5′-CCNNCCNA-3′, 5′-CCCCC-3′ [SEQ ID NO: 12], 5′-CCCCCNNA-3′ [SEQ ID NO: 13], 5′-CCCCCC-3′ [SEQ ID NO: 14], 5′-CCCCCCNA-3′ [SEQ ID NO: 10], 5′-NCCNC-3′, 5′-NCCNCNNA-3′, 5′-NCCCC-3′, 5′-NCCCCNNA-3′, 5′-NCCCCC-3′ [SEQ ID NO: 15], 5′-NCCCCCNA-3′ [SEQ ID NO: 16], 5′-NNCCC-3′, 5′-NNCCCNNA-3′, 5′-NNCCCC-3′, 5′-NNCCCCNA-3′, 5′-NNNCCC-3′ и 5′-NNNCCCNA-3′. Последовательность PAM может представлять собой 5′-CNCCCCAC-3′ [SEQ ID NO: 17], 5′-CCCCCCAG-3′ [SEQ ID NO: 18], 5′-CCCCCCAA-3′ [SEQ ID NO: 11], 5′-CCCCCCAT-3′ [SEQ ID NO: 19], 5′-CCCCCCAC-3′ [SEQ ID NO: 20], 5′-ATCCCCAA-3′ [SEQ ID NO: 21] или 5′-ACGGCCAA-3′ [SEQ ID NO: 22]. Предпочтительно, последовательность PAM будет иметь последовательность 5′-NNNNCNNA-3′. Однако следует понимать, что в зависимости от желательного применения и/или концентрации белка или полипептида Cas могут быть использованы другие сочетания нуклеотидов. Эти последовательности соответствуют тому, что называют «примыкающий к протоспейсеру мотив», или «PAM», в контексте природных crРНК. В системах CRISPR/Cas типа IIC эти последовательности PAM способствуют стабильному взаимодействию комплекса Cascade/crРНК с его мишенью дцДНК, для обеспечения высокой степени специфичности crРНК - как в природных системах-мишенях, так и, следовательно, предпочтительно, в случае РНК по настоящему изобретению - для последовательности-мишени. Предпочтительно, последовательность в непосредственной близости к протоспейсеру не будет представлять собой 5′-NNNCATN-3′.

Последовательности PAM по изобретению, предложенные в настоящем документе, включают последовательности, раскрытые в настоящем документе, например, в виде 6-мерных, 7-мерных или 8-мерных последовательностей. 6-мерные, 7-мерные или 8-мерные последовательности могут начинаться непосредственно в 3′-положении от протоспейсерной последовательности на не являющейся мишенью цепи, без дополнительных нуклеотидов, находящихся в

промежутке между протоспейсерной последовательностью, комплементарной последовательности, которую связывает направляющая РНК, и 5′-концом последовательности PAM. Однако следует понимать, что могут присутствовать дополнительные нуклеотиды, образующие часть последовательности PAM на 3′-конце 6-мерных, 7-мерных или 8-мерных последовательностей. Дополнительно или альтернативно, не являющаяся мишенью цепь может содержать дополнительные нуклеотиды в 3′-положении от последовательности PAM.

Нуклеопротеидный комплекс по изобретению может содержать рибонуклеопротеидный комплекс по изобретению и нуклеотидную цепь-мишень нуклеиновой кислоты, с которой связан рибонуклеопротеид.

Температуры связывания, расщепления, маркирования и модификации

Диапазон температур, включая оптимальный диапазон температур для активности, например, нуклеазной активности, белков Cas по настоящему изобретению, значительно выше диапазона у известных белков Cas. Кроме того, верхняя граница диапазона, в которой сохраняется активность, намного превышает таковую у известных белков Cas. Более высокая оптимальная температура и функциональный диапазон обеспечивают значительное преимущество для генетической инженерии при высоких температурах и, следовательно, например, при редактировании геномов термофильных микроорганизмов, многие из которых находят применение в ряде промышленных, сельскохозяйственных и фармацевтических процессов, проводимых при повышенных температурах. Таким образом, способы, варианты применения, нуклеопротеиды и трансформированные клетки по изобретению могут быть полезны в промышленных процессах, например, позволяя редактировать геном для целей метаболической инженерии. Присутствие последовательностей PAM по изобретению в непосредственной близости к протоспейсерной последовательности в не являющейся мишенью цепи приводит к повышению специфичности белков и полипептидов Cas в отношении последовательностей-мишеней, и позволяет использовать белки и полипептиды Cas при

более высоких температурах и в более широких функциональных диапазонах температур.

Предпочтительно, белки или полипептиды Cas по изобретению способны к связыванию, расщеплению, маркированию или модификации нуклеиновых кислот при температуре от 20°C до 100°C, но являются особенно полезными при повышенных температурах, например, при температуре от 41°C до 122°C, предпочтительно, при температуре от 50°C до 100°C. Белки и полипептиды Cas по изобретению способны к связыванию, расщеплению, маркированию или модификации ДНК, РНК и синтетических нуклеиновых кислот. Белки или полипептиды Cas по изобретению также являются функциональными для целей, связанных с нуклеазной активностью, редактированием генов и маркированием нуклеиновых кислот, при температурах в диапазоне от 20 до 50°C, например.

При указании в настоящем документе диапазона температур подразумевают, что конечные точки включены в раскрытый диапазон температур, то есть, диапазон является «включающим». Например, если указано, что активность имеет место при температуре в диапазоне от 20°C до 100°C, то температуры 20°C до 100°C включены в указанный диапазон.

Предпочтительно, белки или полипептиды Cas по изобретению в ассоциации с соответствующей гРНК (гид-РНК, также называемой молекулой направляющей РНК), которая узнает последовательность-мишень в молекула(ах) полинуклеотида, который должен быть связан, расщеплен, маркирован или модифицирован, осуществляют это при температурах в диапазоне от 50°C до 100°C, необязательно, в диапазоне от 55°C до 100°C, от 60°C до 100°C, от 65°C до 100°C, от 70°C до 100°C, от 75°C до 100°C, от 80°C до 100°C, от 85°C до 100°C, от 90°C до 100°C, от 95°C до 100°C. Более предпочтительно, белки Cas по изобретению расщепляют, маркируют или модифицируют нуклеиновые кислоты при температурах в диапазоне от 51°C до 99°C, от 52°C до 98°C, от 53°C до 97°C, от 54°C до 96°C, от 55°C до 95°C, от 56°C до 94°C, от 57°C до 93°C, от 58°C до 92°C, от 59°C до 91°C, от 60°C до 90°C, от 61°C до 89°C, от 62°C до 88°C, от 63°C до 87°C,

от 64°C до 86°C, от 65°C до 85°C, от 66°C до 84°C, от 67°C до 83°C, от 68°C до 82°C, от 69°C до 81°C, от 70°C до 80°C, от 71°C до 79°C, от 72°C до 78°C, от 73°C до 77°C, от 74°C до 76°C, или при температуре 75°C. Предпочтительно, белки Cas по изобретению связывают, расщепляют, маркируют или модифицируют нуклеиновые кислоты при температурах в диапазоне от 60°C до 80°C, от 61°C до 79°C, от 62°C до 78°C, от 63°C до 77°C, от 64°C до 76°C, от 60°C до 75°C, от 60°C до 70°C. Оптимально, если белки Cas по изобретению связывают, расщепляют, маркируют или модифицируют нуклеиновые кислоты при температурах в диапазоне от 60°C до 65°C, предпочтительно, при 65°C.

Молекулы направляющих РНК могут быть спроектированы для использования с белками и полипептидами Cas по изобретению, при этом молекулы направляющих РНК связываются с последовательностью-мишенью в цепи-мишени, и не являющаяся мишенью цепь дополнительно содержит последовательность PAM, приведенную в настоящем документе, непосредственно в 3′-положении от протоспейсерной последовательности. Последовательность PAM может содержать 5′-NNNNNNNA-3′, предпочтительно, 5′-NNNNCNNA-3′, например, 5′-CCCCCCNA-3′ [SEQ ID NO: 10] или 5′-CCCCCCAA-3′ [SEQ ID NO: 11], и варианты применения, способы, трансформированные клетки и нуклеопротеиды по изобретению могут обеспечивать связывание, расщепление, маркирование и/или модификацию цепи-мишени в диапазоне температур от 55°C до 65°C, предпочтительно, в диапазоне температур от 50°C до 70°C, от 40°C до 65°C, от 45°C до 75°C, от 37°C до 78°C и/или от 20°C до 80°C.

Во всех аспектах изобретения белки или полипептиды Cas могут быть получены, или могут происходить, из бактерий, археев или вирусов; или, альтернативно, могут быть синтезированы *de novo*. В предпочтительных вариантах осуществления белок или полипептид Cas по изобретению происходит из термофильного прокариотического организма, который может быть классифицирован как архей или бактерия, но, предпочтительно, является бактерией. Более предпочтительно, белок или полипептид Cas по изобретению

должен происходить из термофильной бактерии. В настоящем документе термин «термофильные» должен указывать на способность к выживанию и росту при относительно высоких температурах, например, в контексте изобретения, на способность к расщеплению, связыванию или модификации нуклеиновых кислот при температуре от 41 до 122°C (от 106 до 252°F). Предпочтительно, белок или полипептид Cas по изобретению может быть выделен из одной или более термофильных бактерий, и будет функционировать при температуре выше 60°C. Предпочтительно, белок или полипептид Cas по изобретению может быть выделен из одной или более термофильных бактерий, и будет функционировать в диапазоне температур от 60°C до 80°C и, оптимально, от 60°C до 65°C. В предпочтительных вариантах осуществления белок или полипептид Cas по изобретению происходит из *Geobacillus* sp. Более предпочтительно, белок Cas по изобретению происходит из *Geobacillus thermodenitrificans*. Еще более предпочтительно, белок Cas по изобретению происходит из *Geobacillus thermodenitrificans* T12. Белок или полипептид Cas по изобретению может происходить из вируса.

Функциональные фрагменты

Способность белков, полипептидов и рибонуклеопротеидных комплексов Cas по изобретению быть направленными на любую полинуклеотидную последовательность специфичным для последовательности образом может быть с успехом использована для того, чтобы каким-либо образом модифицировать нуклеиновую кислоту-мишень, например, путем ее расщепления и/или ее маркирования, и/или ее модификации. Поэтому следует понимать, что для достижения этого могут быть предоставлены дополнительные белки совместно с белком или полипептидом Cas. Соответственно, белки или полипептиды Cas по изобретению могут дополнительно включать по меньшей мере один функциональный фрагмент и/или белки, полипептиды или рибонуклеопротеидные комплексы Cas по настоящему изобретению могут быть предоставлены в виде части белкового комплекса, содержащего по меньшей мере один дополнительный белок. В предпочтительном аспекте настоящее

изобретение относится к белку, полипептиду или рибонуклеопротеидному комплексу Cas, в которых белок Cas или по меньшей мере один дополнительный белок дополнительно включает по меньшей мере один функциональный фрагмент. По меньшей мере один функциональный фрагмент может быть слит или связан с белком Cas. Предпочтительно, по меньшей мере один функциональный фрагмент может быть трансляционно слит с белком Cas путем экспрессии в природных или искусственных системах экспрессии белка. Альтернативно, по меньшей мере один функциональный фрагмент может быть ковалентно связан с белком Cas на этапе химического синтеза. Предпочтительно, по меньшей мере один функциональный фрагмент слит или связан с N-концом и/или C-концом белка Cas; предпочтительно, C-концом.

Желательно, чтобы по меньшей мере один функциональный фрагмент представлял собой белок. Это может быть гетерологичный белок или, альтернативно, это может быть естественный белок для видов бактерий, из которых происходит белок Cas. По меньшей мере один функциональный фрагмент может представлять собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки.

В особенно предпочтительном аспекте настоящее изобретение относится к белку, полипептиду или рибонуклеопротеидному комплексу Cas, в которых по меньшей мере один функциональный фрагмент представляет собой маркерный белок, например GFP.

Нуклеазная активность

Рибонуклеопротеид Cas по изобретению обладает активностью связывания, расщепления, маркирования или модификации нуклеиновой кислоты при определенной температуре, предпочтительно, повышенной температуре, раскрытой в настоящем

документе, например, при температуре от 50°C до 100°C. Рибонуклеопротеиды по изобретению могут быть способны к связыванию, расщеплению, маркированию или модификации ДНК, РНК или синтетических нуклеиновых кислот. В предпочтительных аспектах рибонуклеопротеиды Cas по изобретению способны расщеплять ДНК специфичным для последовательности образом, в частности, двухцепочечную ДНК.

Белки, полипептиды или рибонуклеопротеиды Cas по изобретению могут иметь более одного нуклеазного домена. Сайт-специфические нуклеазы позволяют создавать двухцепочечные разрывы (ДЦР) в выбранных положениях вдоль цепи ДНК. В целевой клетке-хозяине это позволяет создавать ДЦР в определенных заранее выбранных положениях в геноме. Создание таких разрывов сайт-специфическими нуклеазами стимулирует эндогенный механизм клеточной репарации к переориентации для осуществления вставки, делеции или модификации ДНК в нужных положениях в интересующем геноме.

Один или более сайтов с нуклеазной активностью молекулы белка или полипептида могут быть инактивированы, например, для реализации активности другого функционального фрагмента, связанного или слитого с белком или полипептидом, например, домена нуклеазы, такой как нуклеаза Fok1.

Следовательно, несмотря на тот факт, что белки, полипептиды и рибонуклеопротеиды Cas по изобретению могут обладать эндогенной нуклеазной активностью, для определенных вариантов применения может быть желательной инактивация природной нуклеазной активности белка Cas и предоставление белка или рибонуклеопротеидного комплекса Cas, в которых природная нуклеазная активность Cas9 инактивирована и белок Cas связан с по меньшей мере одним функциональным фрагментом. Снижение частоты случаев событий неправильной направленности за счет комплементации естественной активности нуклеазы Cas9 является одним из таких вариантов применения. Это может быть достигнуто, желательно, путем инактивации природной нуклеазной активности Cas9 белка или рибонуклеопротеидного комплекса Cas и

предоставления гетерологичной нуклеазы, предпочтительно, слитой с белком Cas. Соответственно, настоящее изобретение относится к белку или рибонуклеопротеидному комплексу Cas, в которых по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI. В особенно предпочтительном аспекте белок или рибонуклеопротеидный комплекс Cas по изобретению, слитый с доменом нуклеазы FokI, предоставлен в виде части белкового комплекса, предпочтительно, содержащего другой белок или рибонуклеопротеидный комплекс Cas по изобретению, слитый с доменом нуклеазы FokI, и при этом два комплекса направлены на противоположные цепи геномной ДНК-мишени.

Для некоторых вариантов применения может быть желательно полное ослабление нуклеазной активности белка, полипептида или рибонуклеопротеида Cas, например, в тех вариантах применения, когда белок или рибонуклеопротеидный комплекс Cas используют для узнавания и модификации конкретной последовательности-мишени в нуклеиновой кислоте, например, для ее маркирования в качестве части диагностического теста. В таких вариантах применения нуклеазная активность белка Cas может быть инактивирована, и функциональный фрагмент, слитый с белком Cas, может представлять собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или метки аффинной очистки.

В предпочтительном аспекте каталитически неактивный, или «мертвый», белок или полипептид Cas (мCas), лишенный нуклеазной активности, может быть связан с нуклеотидной последовательностью-мишенью и, за счет этого, стерически подавлять активность этой последовательности. Например, может быть сконструирована направляющая РНК, которая комплементарна последовательности промотора или экзона гена, так что связывание

мCas и направляющей РНК с геном приводит к стерическому подавлению инициации транскрипции или элонгации последовательности гена, тем самым подавляя экспрессию гена. Альтернативно, в способах и вариантах применения, описанных в настоящем документе, могут быть использованы модифицированные нуклеазные варианты gtCas9, которые представляют собой никазы. Никаза может быть создана путем мутации в любом из каталитических доменов HNH или RuvC нуклеазы gtCas9. Это было показано для Cas9 *S. pyogenes* (spCas) с spCas9-мутантами D10A и H840A, которые имеют неактивный нуклеазный домен RuvC или HNH, соответственно. Сочетание этих двух мутаций приводит к созданию каталитически «мертвого» варианта Cas9 (Standage-Beier, K. *et al.*, 2015, ACS Synth. Biol. 4, 1217-1225; Jinek, M. *et al.*, 2012, Science 337, 816-821; Xu, T. *et al.*, 2015, Appl. Environ. Microbiol. 81, 4423-4431). Исходя из гомологии последовательностей (фигура 3), эти остатки могут представлять собой D8 (D17 на фигуре 3) и D581 или H582 (фигура 3) в gtCas9.

В особенно предпочтительном аспекте настоящее изобретение относится к белку или к рибонуклеопротеидному комплексу Cas, в которых нуклеазная активность белка Cas инактивирована, и по меньшей мере один функциональный фрагмент представляет собой маркерный белок, например GFP. В этом случае становится возможной специфическая направленность на интересующую последовательность нуклеиновой кислоты и ее визуализация с использованием маркера, который генерирует оптический сигнал. Подходящие маркеры могут включать, например, флуоресцентный репортерный белок, например, зеленый флуоресцентный белок (GFP), желтый флуоресцентный белок (YFP), красный флуоресцентный белок (RFP), голубой флуоресцентный белок (CFP) или mCherry. Такой флуоресцентный репортерный ген предоставляет подходящий маркер для визуализации экспрессии белка, поскольку его экспрессия может быть просто и непосредственно определена путем измерения флуоресценции. Альтернативно, репортерная нуклеиновая кислота может кодировать люминесцентный белок, такой как люцифераза (например, люцифераза светляков). Альтернативно, репортерный ген может кодировать хромогенный фермент, который может быть

использован для генерирования оптического сигнала, например, такой хромогенный фермент, как бета-галактозидаза (LacZ) или бета-глюкуронидаза (Gus). Репортеры, используемые для измерения экспрессии, также могут представлять собой антигенные пептидные метки. Другие репортеры или маркеры также известны в данной области, и их можно использовать по мере необходимости.

Поскольку маркер может быть визуализирован, в некоторых вариантах осуществления, когда нуклеиновая кислота-мишень представляет собой РНК, в частности мРНК, возможно количественное определение транскрипционной активности гена путем обнаружения и количественного определения оптического сигнала, обеспечиваемого маркером, в частности, когда оптический сигнал, генерируемый маркером, прямо пропорционален количеству продукта экспрессии. Вследствие этого, в предпочтительных вариантах осуществления изобретения белки или рибонуклеопротеиды Cas по изобретению могут быть использованы для анализа продуктов экспрессии гена, представляющего интерес.

В одном аспекте gtCas9, описанные в настоящем документе, могут быть использованы в способе опосредованной гомологичной рекомбинацией (ГР) модификации генома в микробных клетках. Такие способы включают ГР и сайт-специфическую активность gtCas9, в результате имеет место контрселекция за счет активности gtCas9, приводящей к удалению микроорганизмов, которые не имеют нужную модификацию, введенную в результате ГР.

Таким образом, способы и варианты применения, предложенные в настоящем документе, способствуют осуществлению процесса гомологичной рекомбинации на первом этапе, в результате чего микробный геном может быть модифицирован за счет нужной мутации, и на втором этапе не модифицированные клетки могут становиться мишенью для рибонуклеазного комплекса gtCas9, вносящего ДЦР в геномы не модифицированных клеток. Вследствие отсутствия эффективного механизма репарации за счет негомологичного соединения концов (NHEJ) у большинства микроорганизмов ДЦР, как правило, приводит к гибели клеток. Таким образом, данные способы и варианты применения приводят к увеличению общей популяции микробных клеток с нужной мутацией и при этом элиминации любых

не модифицированных микробных клеток. Предпочтительно, такие способы и варианты применения используют в случае микробов, которые практически не имеют эндогенный репарационный механизм NHEJ. Альтернативно, способы и варианты применения можно использовать в случае микробов, которые имеют эндогенный репарационный механизм NHEJ. Способы и варианты применения, описанные в настоящем документе, могут быть применены к микробам, которые имеют эндогенный репарационный механизм NHEJ, но у которых либо репарационный механизм NHEJ условно ослаблен, либо активность NHEJ подвергнута нокауту.

В способах и вариантах применения, предложенных в настоящем документе, можно использовать полинуклеотидную последовательность гомологичной рекомбинации, которая имеет по меньшей мере одно несовпадение с гид-РНК, так что гид-РНК теряет способность узнавать модифицированный геном. Это означает, что рибонуклеазный комплекс gtCas9 не будет узнавать модифицированный геном. Вследствие этого, ДЦР не могут быть внесены рибонуклеазным комплексом gtCas9 и модифицированные таким образом клетки будут выживать. Однако для клеток с не модифицированными геномами все-еще будет характерна существенная комплементарность с гид-РНК и, как следствие, может происходить сайт-специфическое расщепление рибонуклеазным комплексом gtCas9.

В другом аспекте способов и вариантов применения по изобретению путь предотвращения расщепления рибонуклеазным комплексом gtCas9 микробного генома заключается не столько в модификации или элиминации последовательности, являющейся мишенью для гид-РНК, а скорее связан с последовательностью PAM, необходимой для рибонуклеазного комплекса gtCas9. PAM либо модифицируют, либо элиминируют, чтобы рибонуклеазный комплекс gtCas9 не мог «увидеть» специфический сайт разрезания. Таким образом, способы и варианты применения по изобретению могут включать те, в которых используют полинуклеотидную последовательность гомологичной рекомбинации, которая не содержит последовательность PAM, узнаваемую рибонуклеазным комплексом gtCas9. Таким образом, ДЦР не могут быть внесены рибонуклеазным комплексом gtCas9, и модифицированные за счет ГР

клетки будут выживать. Однако не модифицированные клетки все еще будут узнаваться рибонуклеазным комплексом gtCas9 и его гид-РНК, и, как следствие, будет иметь место сайт-специфическое расщепление.

Таким образом, в настоящем документе предложены способы и варианты применения, в которых используют ГР для модификации генома микроорганизмов. Предпочтительно, каждая из расположенных выше и ниже фланкирующих последовательностей имеют длину от 0,5 тысяч пар нуклеотидов (т.п.н.) до 1,0 т.п.н. Однако также возможна рекомбинация с использованием более длинных или более коротких фрагментов. Полинуклеотид для гомологичной рекомбинации может дополнительно содержать полинуклеотидную последовательность между расположенными выше и ниже фланкирующими областями. Эта полинуклеотидная последовательность может, например, содержать модификацию, которая должна быть введена в микробный геном.

Хотя для гомологичной рекомбинации необходимо, чтобы расположенные выше и ниже фланкирующие последовательности имели существенную комплементарность с областями-мишенями, несовпадения также допустимы. Таким образом, в некоторых вариантах осуществления известно, что гомологичная рекомбинация происходит между сегментами ДНК со значительной степенью гомологии с расположенными выше и ниже фланкирующими последовательностями. В альтернативных вариантах осуществления расположенные выше и ниже фланкирующие последовательности полностью комплементарны областям-мишеням. Расположенные выше и ниже фланкирующие последовательности не обязательно должны быть идентичны по размеру. Однако в некоторых случаях расположенные выше и ниже фланкирующие последовательности идентичны по размеру. Эффективность гомологичной рекомбинации будет варьироваться в зависимости от вероятности гомологичной рекомбинации фрагмента фланкирующей последовательности с наименьшей длиной. Однако, даже если процесс гомологичной рекомбинации является неэффективным, способ, описанный в настоящем документе, позволит успешно выбирать любую микробную клетку, имеющую желательную модификацию, из не модифицированных

микробных клеток. Гомологичная рекомбинация также позволяет осуществлять крупные делеции (например, 50 т.п.н. или более), охватывающие целые кластеры генов. Гомологичную рекомбинацию также используют для рекомбинационной инженерии, которая является хорошо известным способом, позволяющим производить рекомбинацию более коротких фрагментов (45-100 нт). Способы и варианты применения, описанные в настоящем документе, необязательно, могут дополнительно включать использование по меньшей мере другого полинуклеотида гомологичной рекомбинации или полинуклеотида, содержащего последовательность, кодирующую полинуклеотид гомологичной рекомбинации, имеющий последовательность, в значительной степени комплементарную второй области-мишени, содержащей мишень в микробном геноме.

В предпочтительных вариантах осуществления в способах и вариантах применения, описанных в настоящем документе, используют полинуклеотид гомологичной рекомбинации, который представляет собой ДНК. В некоторых вариантах осуществления ДНК является одноцепочечной. В других вариантах осуществления ДНК является двухцепочечной. В следующих вариантах осуществления ДНК является двухцепочечный и переносимой в плазмиде.

ГР в способах и вариантах применения, предложенных в настоящем документе, можно использовать для удаления полинуклеотидной последовательности из микробного генома. Альтернативно, ГР в способах и вариантах применения, предложенных в настоящем документе, можно использовать для вставки одного или более генов, или их фрагментов, в микробный геном. В качестве дополнительной альтернативы, ГР в способах и вариантах применения, предложенных в настоящем документе, можно использовать для модификации или замены по меньшей мере одного нуклеотида в микробном геноме. Следовательно, способы и варианты применения, предложенные в настоящем документе, могут быть использованы для любой желательной модификации генома.

Альтернативно, gtCas9, описанные в настоящем документе, можно использовать в опосредованном ГР способе модификации генома в микробных клетках, в котором активность gtCas9 приводит к внесению ДЦР и может индуцировать клеточную ГР в микробных

клетках, как было показано для spCas9 (Jiang *et al.* (2013) Nature Biotech, 31, 233-239; Xu *et al.* (2015) Appl Environ Microbiol, 81, 4423-4431; Huang *et al.* (2015) Acta Biochimica et Biophysica Sinica, 47, 231-243).

Альтернативно, гомологичная рекомбинация может быть облегчена за счет рекомбинационной инженерии, например, путем введения олигонуклеотида в микробную клетку, экспрессирующую ген, кодирующий RecT или бета-белок, как описано в обзоре Mougiakos *et al.* ((2016), Trends Biotechnol. 34: 575-587). В следующем варианте осуществления Cas9 можно комбинировать с мультиплексной автоматизированной генной инженерией (MAGE), как описано в публикации Ronda *et al.* ((2016), Sci. Rep. 6: 19452.)

В тексте заявки эталонные последовательности белков Cas по изобретению могут быть определены, как нуклеотидная последовательность, кодирующая аминокислотную последовательность. Например, аминокислотная последовательность мотивов, приведенных в SEQ ID NO: 2-6, также включает все нуклеотидные последовательности, которые кодируют данную аминокислотную последовательность.

Соответственно, настоящее изобретение также относится к выделенной молекуле нуклеиновой кислоты, кодирующей белок Cas, содержащий:

a. аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]; и/или

b. аминокислотный мотив $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и/или

c. аминокислотный мотив X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и/или

d. аминокислотный мотив X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и/или

e. аминокислотный мотив $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6],

где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин;

при этом белок Cas способен к связыванию, расщеплению, маркированию или модификации ДНК при температуре от 50°C до 100°C при связывании с по меньшей мере одной молекулой направляющей РНК и полинуклеотидом, содержащим нуклеотидную последовательность-мишень, узнаваемую молекулой направляющей РНК.

В другом аспекте настоящее изобретение также относится к выделенной молекуле нуклеиновой кислоты, кодирующей белок (Cas), связанный с кластером регулярно расположенных группами коротких палиндромных повторов (CRISPR), имеющий аминокислотную последовательность SEQ ID NO: 1 или последовательность, по меньшей мере на 77% идентичную ей.

В другом аспекте настоящее изобретение также относится к выделенной молекуле нуклеиновой кислоты, дополнительно содержащей по меньшей мере одну нуклеотидную последовательность, кодирующую пептид, который в процессе трансляции слит с белком Cas.

В другом аспекте настоящее изобретение также относится к выделенной молекуле нуклеиновой кислоты, в которой по меньшей мере одна нуклеотидная последовательность, слитая с молекулой нуклеиновой кислоты, кодирующей белок Cas, кодирует белок, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко)-активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, структурирующего ДНК белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или метки аффинной очистки.

Экспрессионные векторы

Нуклеиновые кислоты по настоящему изобретению могут быть

выделены. Однако для того, чтобы экспрессия чувствительного конструкта нуклеиновой кислоты могла быть осуществлена в выбранной клетке, полинуклеотидная последовательность, кодирующая белок или рибонуклеопротеид Cas, предпочтительно, должна быть предоставлена в экспрессионном конструкте. В некоторых вариантах осуществления полинуклеотид, кодирующий белок или рибонуклеопротеид Cas, может быть предоставлен в виде части подходящего экспрессионного вектора. В некоторых вариантах осуществления экспрессионный вектор по настоящему изобретению (с нуклеотидной последовательностью, кодирующей аминокислотные остатки, которые при экспрессии будут слиты с белком Cas, или без нее) может дополнительно содержать нуклеотидную последовательность, кодирующую молекулу направляющей РНК, как описано выше в настоящем документе. Следовательно, такие экспрессионные векторы могут быть использованы в подходящем хозяине для получения рибонуклеопротеидного комплекса по изобретению, который может быть направлен на нужную нуклеотидную последовательность. Альтернативно, нуклеотидные последовательности, кодирующие молекулу направляющей РНК, описанную выше в настоящем документе, могут быть предоставлены в отдельном экспрессионном векторе или, альтернативно, могут быть доставлены в клетку-мишень другими способами.

Подходящие экспрессионные векторы можно варьировать в зависимости от клетки-реципиента, соответственно, они могут включать регуляторные элементы, которые обеспечивают экспрессию в клетке-мишени, и, предпочтительно, которые способствуют высоким уровням экспрессии. Такие регуляторные последовательности могут обладать способностью влиять на транскрипцию или трансляцию гена, или продукта гена, например, в отношении инициации, точности, скорости, стабильности, последующего процессинга и мобильности.

Такие элементы могут включать, например, сильные и/или конститутивные промоторы, 5-' и 3'-UTR, энхансеры транскрипции и/или трансляции, фактор транскрипции или связывающие белок последовательности, последовательности сайтов инициации и терминации транскрипции, сайты связывания рибосом, сайты

рекомбинации, последовательности полиаденилирования, смысловые или антисмысловые последовательности, последовательности, обеспечивающие правильную инициацию транскрипции, и, необязательно, поли-А-сигналы, обеспечивающие терминацию транскрипции и стабилизацию транскрипта в клетке-хозяине. Регуляторные последовательности могут иметь растительное, животное, бактериальное, грибковое или вирусное происхождение и, предпочтительно, могут происходить из того же организма, что и клетка-хозяин. Очевидно, что подходящие регуляторные элементы будут варьироваться в зависимости от интересующей клетки-хозяина. Например, регуляторные элементы, которые способствуют высокому уровню экспрессии в клетках-хозяевах прокариот, таких как *E. coli*, могут включать промоторы pLac, T7, P(Bla), P(Cat), P(Kat), trp или tac. Регуляторные элементы, которые способствуют высокому уровню экспрессии в клетках-хозяевах эукариот, могут включать промотор AOX1 или GAL1 в дрожжах или промоторы CMV или SV40, энхансер CMV, энхансер SV40, активатор транскрипции вируса простого герпеса VIP16 или включение интрона глобина в клетках животных. В растениях высокий уровень конститутивной экспрессии может быть получен, например, с помощью промотора убиквитина 1 *Zea mays* или промоторов 35S и 19S вируса мозаики цветной капусты.

Подходящие регуляторные элементы могут быть конститутивными, вследствие этого они управляют экспрессией в большинстве условий окружающей среды или стадий развития, специфичными или индуцируемыми стадией развития. Предпочтительно, промотор является индуцируемым для управления экспрессией в ответ на сигналы окружающей среды, химические сигналы или сигналы развития, такие как температура, свет, химикаты, засуха и другие стимулы. Соответственно, можно выбирать промоторы, допускающие экспрессию интересующего белка на конкретных стадиях развития или в ответ на вне- или внутриклеточные условия, сигналы или внешние стимулы. Например, существует ряд промоторов для использования в *E. coli*, которые обеспечивают высокий уровень экспрессии на определенных стадиях роста (например, промотор osmY стационарной фазы) или в ответ на

конкретные стимулы (например, промотор HtpG теплового шока).

Подходящие экспрессионные векторы могут содержать дополнительные последовательности, кодирующие селективные маркеры, которые позволяют отбирать указанный вектор в подходящей клетке-хозяине и/или в конкретных условиях.

Изобретение также относится к способу модификации нуклеиновой кислоты-мишени в клетке, включающему трансфекцию, трансформацию или трансдукцию клетки любым из экспрессионных векторов, описанных выше в настоящем документе. Способы трансфекции, трансформации или трансдукции хорошо известны специалистам в данной области. Если используют один экспрессионный вектор для обеспечения экспрессии рибонуклеопротеидного комплекса по изобретению, и если направляющую РНК добавляют непосредственно в клетку, то можно использовать один и тот же, или разные методы трансфекции, трансформации или трансдукции. Аналогично, если используют один экспрессионный вектор для обеспечения экспрессии рибонуклеопротеидного комплекса по изобретению, и если используют другой экспрессионный вектор для получения направляющей РНК *in situ* в результате экспрессии, то можно использовать один и тот же, или разные методы трансфекции, трансформации или трансдукции.

В других вариантах осуществления мРНК, кодирующую белок или полипептид Cas, вводят в клетку, так что комплекс Cascade экспрессируется в клетке. Направляющую РНК, которая направляет комплекс белка Cas к нужной последовательности-мишени, также вводят в клетку, одновременно, раздельно или последовательно с мРНК, так что в клетке образуется необходимый рибонуклеопротеидный комплекс.

Соответственно, изобретение также относится к способу модификации, то есть расщепления, мечения, маркирования или связывания, нуклеиновой кислоты-мишени, включающему создание контакта нуклеиновой кислоты с рибонуклеопротеидным комплексом, описанным выше в настоящем документе.

Кроме того, изобретение также относится к способу модификации нуклеиновой кислоты-мишени, включающему создание

контакта нуклеиновой кислоты с белком или полипептидом Cas, описанным выше в настоящем документе, в дополнение к молекуле направляющей РНК, описанной выше в настоящем документе.

Таким образом, в соответствии с вышеуказанными способами модификацию нуклеиновой кислоты-мишени можно выполнять *in vitro* и в бесклеточной среде. В бесклеточной среде добавление каждого из: нуклеиновой кислоты-мишени, белка Cas и молекулы направляющей РНК может происходить одновременно, последовательно (в любом желаемом порядке) или раздельно. Таким образом, можно добавлять нуклеиновую кислоту-мишень и направляющую РНК в реакционную смесь одновременно, а затем отдельно на более поздней стадии добавлять белок или полипептид Cas по изобретению.

В равной мере модификацию нуклеиновой кислоты-мишени можно осуществлять *in vivo*, то есть, *in situ* в клетке, будь то выделенная клетка или часть многоклеточной ткани, органа или организма. В контексте цельной ткани и органов и в контексте организма способ желательно осуществлять *in vivo* или, альтернативно, можно осуществлять путем выделения клетки из целой ткани, органа или организма, обработки клетки рибонуклеопротеидным комплексом в соответствии со способом, с последующим возвратом клетки, обработанной рибонуклеопротеидным комплексом, в ее прежнее местоположение или в другой участок того же самого, или другого, организма.

В данных вариантах осуществления для рибонуклеопротеидного комплекса, или белка или полипептида Cas, необходима соответствующая форма доставки в клетку. Такие подходящие системы и способы доставки хорошо известны специалистам в данной области техники и включают, но не ограничиваются ими, микроинъекцию в цитоплазму или ядро. В предпочтительных способах доставки используют аденоассоциированный вирус (AAV); такая система доставки не вызывает заболевание у человека и была одобрена для клинического применения в Европе.

Соответственно, настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени, включающему создание контакта нуклеиновой кислоты:

a. с рибонуклеопротеидным комплексом, описанным выше в настоящем документе; или

b. с белком или белковым комплексом, описанным выше в настоящем документе, и молекулой РНК, описанной выше в настоящем документе.

В следующем аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в клетке, включающему трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором, содержащим нуклеотидные последовательности, кодирующие рибонуклеопротеидный комплекс, описанный выше в настоящем документе; или, альтернативно, трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором, содержащим нуклеотидные последовательности, кодирующие белок или белковый комплекс, описанный выше в настоящем документе, и дополнительным экспрессионным вектором, содержащим нуклеотидную последовательность, кодирующую молекулу направляющей РНК, описанную выше в настоящем документе.

В следующем аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в клетке, включающему трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором, содержащим нуклеотидные последовательности, кодирующие белок или белковый комплекс, описанный выше в настоящем документе, и последующую доставку в клетку молекулы направляющей РНК, описанной выше в настоящем документе.

В вариантах осуществления, в которых молекулу гид-РНК (гРНК) (то есть, направляющей РНК) и белок или полипептид Cas предоставляют отдельно, а не в виде части рибонуклеопротеидного комплекса, для молекулы гРНК необходима соответствующая форма доставки в клетку, одновременно, раздельно или последовательно с белком или белковым комплексом Cas. Такие формы введения РНК в клетки хорошо известны специалистам в данной области и могут включать доставку *in vitro* или *ex vivo* с использованием обычных методов трансфекции. Можно использовать любой из физических методов, таких как микроинъекция и электропорация, а также соосаждение кальцием, коммерчески доступные катионные полимеры и

липиды, а также проникающие в клетку пептиды и проникающие в клетку (биолистические) частицы. Например, в качестве носителей для доставки в цитоплазму и/или в ядро могут быть использованы вирусы, особенно предпочтительными являются AAV, например, посредством (обратимого) слияния белкового комплекса Cas по изобретению или рибонуклеопротеидного комплекса по изобретению с вирусной частицей.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок или репортерный белок, и маркерный белок или репортерный белок связывается с нуклеиновой кислотой-мишенью; предпочтительно, при этом маркер представляет собой флуоресцентный белок, например, зеленый флуоресцентный белок (GFP).

В вышеупомянутых способах модификации нуклеиновой кислоты-мишени функциональный фрагмент может представлять собой маркер, и маркер связывается с нуклеиновой кислотой-мишенью; предпочтительно, при этом маркер представляет собой белок; необязательно, флуоресцентный белок, например, зеленый флуоресцентный белок (GFP), желтый флуоресцентный белок (YFP), красный флуоресцентный белок (RFP) или mCherry. Независимо от применения *in vitro*, *ex vivo* или *in vivo*, способы по изобретению могут быть использованы для прямой визуализации локуса-мишени в молекуле нуклеиновой кислоты, предпочтительно, в форме структуры более высокого порядка, такой как суперспирализованная плазмида или хромосома, или одноцепочечной нуклеиновой кислоты-мишени, такой как мРНК. Для прямой визуализации локуса-мишени можно использовать электронную микрографию или флуоресцентную микроскопию. Однако следует понимать, что в контексте способов по изобретению в качестве маркера могут быть использованы и другие виды меток, включая молекулы органических красителей, радиоактивные метки и спиновые метки, которые могут представлять собой малые молекулы.

В способах модификации нуклеиновой кислоты-мишени по изобретению, в которых нуклеиновая кислота-мишень представляет собой дцДНК, функциональный фрагмент может представлять собой

нуклеазу или геликазу-нуклеазу, и модификация, предпочтительно, представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе. Таким путем может быть осуществлено уникальное специфичное для последовательности разрезание ДНК за счет использования подходящего функционального фрагмента, слитого с рибонуклеопротеидным комплексом. Выбранная последовательность компонента РНК конечного рибонуклеопротеидного комплекса обеспечивает желаемую специфичность последовательности для действия функционального фрагмента.

Таким образом, изобретение также относится к способу негомологичного соединения концов молекулы дцДНК в желаемом локусе в клетке для удаления по меньшей мере части нуклеотидной последовательности из молекулы дцДНК; необязательно, для нокаута функции гена или генов, при этом способ включает создание двухцепочечных разрывов с использованием любого из способов модификации нуклеиновой кислоты-мишени, описанных выше в настоящем документе.

Изобретение также относится к способу гомологичной рекомбинации нуклеиновой кислоты с молекулой дцДНК в желаемом локусе в клетке для модификации существующей нуклеотидной последовательности или вставки желаемой нуклеотидной последовательности, при этом способ включает создание двухцепочечного разрыва в желаемом локусе с использованием любого из способов модификации нуклеиновой кислоты-мишени, описанных выше в настоящем документе.

Таким образом, изобретение также относится к способу модификации экспрессии генов в организме, включающему модификацию последовательности нуклеиновой кислоты-мишени в соответствии с любым из способов, описанных выше в настоящем документе, при этом нуклеиновая кислота представляет собой дцДНК, и функциональный фрагмент выбирают из фермента, модифицирующего ДНК (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции.

Изобретение также относится к способу модификации экспрессии генов в организме, включающему модификацию последовательности нуклеиновой кислоты-мишени в соответствии с

любым из способов, описанных выше в настоящем документе, при этом нуклеиновая кислота представляет собой мРНК, и функциональный фрагмент представляет собой рибонуклеазу; необязательно, выбранную из эндонуклеазы, 3'-экзонуклеазы или 5'-экзонуклеазы.

Нуклеиновая кислота-мишень может представлять собой ДНК, РНК или синтетическую нуклеиновую кислоту. Предпочтительно, нуклеиновая кислота-мишень представляет собой ДНК; предпочтительно, дцДНК.

Однако нуклеиновая кислота-мишень может представлять собой РНК; предпочтительно, мРНК. Следовательно, альтернативно, настоящее изобретение также относится к способам модификации нуклеиновой кислоты-мишени, в которых нуклеиновая кислота-мишень представляет собой РНК.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени, при этом нуклеиновая кислота представляет собой дцДНК, по меньшей мере один функциональный фрагмент представляет собой нуклеазу или геликазу-нуклеазу, и модификация представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в клетке, при этом модификация приводит к выключению экспрессии генов в желаемом локусе; и при этом способ включает этапы:

a. создания двухцепочечных разрывов в молекуле дцДНК; и

b. репарации молекулы дцДНК в клетке путем негомологичного соединения концов (NHEJ).

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в клетке, при этом существующая нуклеотидная последовательность модифицируется или удаляется и/или желаемая нуклеотидная последовательность вводится в желаемый локус, при этом способ включает этапы:

a. создания двухцепочечного разрыва в желаемом локусе; и

b. репарации молекулы дцДНК в клетке путем гомологичной рекомбинации.

В другом аспекте настоящее изобретение относится к способу

модификации экспрессии генов в клетке, включающему модификацию последовательности нуклеиновой кислоты-мишени, описанную выше в настоящем документе; при этом нуклеиновая кислота представляет собой дцДНК, и функциональный фрагмент выбирают из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции.

В другом аспекте настоящее изобретение относится к способу модификации экспрессии генов в клетке, включающему модификацию последовательности нуклеиновой кислоты-мишени, описанную выше в настоящем документе; при этом нуклеиновая кислота представляет собой мРНК, и функциональный фрагмент представляет собой рибонуклеазу; необязательно, выбранную из эндонуклеазы, 3'-экзонуклеазы или 5'-экзонуклеазы.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени, описанному выше в настоящем документе, при этом способ осуществляют при температуре от 45°C до 100°C. Предпочтительно, способ осуществляют при температуре 50°C или выше. Более предпочтительно, способ осуществляют при температуре от 55°C до 80°C. Оптимально, способ осуществляют при температуре от 60°C до 65°C. Альтернативно, способ можно осуществлять при температуре от 20°C до 45°C. Более предпочтительно, при температуре от 30°C до 45°C. Еще более предпочтительно, при температуре от 37°C до 45°C.

В любом из способов модификации нуклеиновой кислоты-мишени, описанных выше в настоящем документе, клетка может представлять собой прокариотическую клетку или, альтернативно, может представлять собой эукариотическую клетку.

Клетки-хозяева

Преимуществом настоящего изобретения является широкая область применения, и клетки-хозяева по настоящему изобретению могут быть получены из любого поддающегося генетическим манипуляциям организма, который можно культивировать. Соответственно, настоящее изобретение относится к клетке-хозяину, трансформированной способом, описанным выше в настоящем документе. Изобретение относится к трансформированной клетке,

имеющей нуклеотидную последовательность-мишень в двухцепочечном полинуклеотиде-мишени, при этом указанная клетка содержит белок или полипептид Cas, описанный в настоящем документе, и по меньшей мере одну молекулу направляющей РНК, описанную в настоящем документе, и экспрессионный вектор, содержащий нуклеиновую кислоту, кодирующую по меньшей мере одно из указанного белка Cas и указанной молекулы направляющей РНК.

Подходящие клетки-хозяева могут быть прокариотическими или эукариотическими. В частности, для использования по настоящему изобретению могут быть выбраны обычно используемые клетки-хозяева, включая прокариотические или эукариотические клетки, которые являются генетически доступными и которые можно культивировать, например, прокариотические клетки, клетки грибов, растительные клетки и клетки животных, включая клетки человека (но не эмбриональные стволовые клетки). Предпочтительно, клетки-хозяева будут выбраны из прокариотической клетки, клетки грибов, растительной клетки, клетки простейших или клетки животных. Предпочтительные клетки-хозяева для использования по настоящему изобретению, как правило, получают от видов, которые обычно демонстрируют высокие темпы роста, легко поддаются культивированию и/или трансформации, характеризуются коротким временем генерации, от видов, которые имеют установленные, связанные с ними генетические ресурсы, или от видов, подвергнутых селекции, модификации или синтезу для оптимальной экспрессии гетерологичного белка в определенных условиях. В предпочтительных вариантах осуществления изобретения, в которых интересующий белок в итоге будет использован в конкретных промышленных, сельскохозяйственных, химических или терапевтических целях, подходящая клетка-хозяин может быть выбрана на основе желаемых конкретных условий или клеточного контекста, в которых предполагается использовать интересующий белок. Предпочтительно, клетка-хозяин будет представлять собой прокариотическую клетку. В предпочтительных вариантах осуществления клетка-хозяин представляет собой бактериальную клетку. Клеткой-хозяином может быть, например, клетка

Escherichia coli (*E. coli*). Предпочтительно, клетка-хозяин будет представлять собой клетку термофильной бактерии.

Способы и варианты применения по изобретению, описанные в настоящем документе, могут быть использованы для модификации геномов бактериальных клеток. В конкретных вариантах осуществления бактерии представляют собой термофильные бактерии, предпочтительно, бактерии выбирают из: *Acidithiobacillus* sp., включая ***Acidithiobacillus caldus***; *Aeribacillus* sp., включая *Aeribacillus pallidus*; *Alicyclobacillus* sp., включая *Alicyclobacillus acidocaldarius*, *Alicyclobacillus acidoterrestris*, *Alicyclobacillus cycloheptanicus I*, *Alicyclobacillus hesperidum*; *Anoxybacillus* sp., включая *Anoxybacillus caldiproteolyticus*, *Anoxybacillus flavithermus*, *Anoxybacillus rupiensis*, *Anoxybacillus tepidamans*; *Bacillus* sp., включая ***Bacillus caldolyticus***, ***Bacillus caldotenax***, ***Bacillus caldovelox***, ***Bacillus coagulans***, ***Bacillus clausii***, ***Bacillus licheniformis***, ***Bacillus methanolicus***, ***Bacillus smithii including Bacillus smithii ET138***, ***Bacillus subtilis***, *Bacillus thermocopriae*, *Bacillus thermolactis*, ***Bacillus thermoamylovorans***, *Bacillus thermoleovorans*; *Caldibacillus* sp., включая *Caldibacillus debilis*; *Caldicellulosiruptor* sp., включая ***Caldicellulosiruptor bescii***, ***Caldicellulosiruptor hydrothermalis***, *Caldicellulosiruptor kristjanssonii*, *Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor lactoaceticus*, *Caldicellulosiruptor obsidiansis*, *Caldicellulosiruptor owensensis*, *Caldicellulosiruptor saccharolyticus*; *Clostridium* sp., включая *Clostridium clariflavum*, *Clostridium straminisolvens*, *Clostridium tepidiprofundi*, *Clostridium thermobutyricum*, ***Clostridium thermocellum***, ***Clostridium thermosuccinogenes***, *Clostridium thermopalmarium*; *Deinococcus* sp., включая ***Deinococcus cellulosilyticus***, ***Deinococcus deserti***, ***Deinococcus geothermalis***, ***Deinococcus murrayi***, ***Deinococcus radiodurans***; *Defluviitalea* sp., включая *Defluviitalea phaphyphila*, *Desulfotomaculum* sp., включая *Desulfotomaculum carboxydivorans*, *Desulfotomaculum nigrificans*, *Desulfotomaculum salinum*, *Desulfotomaculum solfataricum*;

Desulfurella sp., включая *Desulfurella acetivorans*; *Desulfurobacterium* sp., включая *Desulfurobacterium thermolithotrophum*; *Geobacillus* sp., включая *Geobacillus icigianus*, *Geobacillus caldoxylosilyticus*, *Geobacillus jurassicus*, *Geobacillus galactosidasius*, ***Geobacillus kaustophilus***, *Geobacillus lituanicus*, ***Geobacillus stearothermophilus***, *Geobacillus subterraneus*, *Geobacillus thermantarcticus*, *Geobacillus thermocatenulatus*, ***Geobacillus thermodenitrificans***, ***Geobacillus thermoglucosidans***, ***Geobacillus thermoleovorans***, *Geobacillus toebii*, ***Geobacillus uzenensis***, *Geobacillus vulcanii*, *Geobacillus zalihae*; *Hydrogenobacter* sp., включая *Hydrogenobacter thermophiles*; *Hydrogenobaculum* sp., включая *Hydrogenobaculum acidophilum*; *Ignavibacterium* sp., включая *Ignavibacterium album*; *Lactobacillus* sp., включая ***Lactobacillus bulgaricus***, ***Lactobacillus delbrueckii***, *Lactobacillus ingluviei*, ***Lactobacillus thermotolerans***; *Marinithermus* sp., включая *Marinithermus hydrothermalis*; *Moorella* sp., включая ***Moorella thermoacetica***; *Oceanithermus* sp., включая *Oceanithermus desulfurans*, *Oceanithermus profundus*; *Paenibacillus* sp., включая *Paenibacillus* sp. J2, *Paenibacillus marinum*, *Paenibacillus thermoaerophilus*; *Persephonella* sp., включая *Persephonella guaymasensis*, *Persephonella hydrogeniphila*, *Persephonella marina*; *Rhodothermus* sp., включая ***Rhodothermus marinus***, *Rhodothermus obamensis*, *Rhodothermus profundi*; *Sulfobacillus* sp., включая *Sulfobacillus acidophilus*; *Sulfurihydrogenibium* sp., включая *Sulfurihydrogenibium azorense*, *Sulfurihydrogenibium kristjanssonii*, *Sulfurihydrogenibium rodmanii*, *Sulfurihydrogenibium yellowstonense*; *Symbiobacterium* sp., включая *Symbiobacterium thermophilum*, *Symbiobacterium toebii*; *Thermoanaerobacter* sp., включая ***Thermoanaerobacter brockii***, ***Thermoanaerobacter ethanolicus***, *Thermoanaerobacter italicus*, *Thermoanaerobacter kivui*, *Thermoanaerobacter marianensis*, *Thermoanaerobacter mathranii*, *Thermoanaerobacter pseudoethanolicus*, *Thermoanaerobacter wiegelii*; *Thermoanaerobacterium* sp., включая *Thermoanaerobacterium aciditolerans*, *Thermoanaerobacterium aotearoense*,

Thermoanaerobacterium ethanolicus, ***Thermoanaerobacterium pseudoethanolicus***, ***Thermoanaerobacterium saccharolyticum***, ***Thermoanaerobacterium thermosaccharolyticum***, ***Thermoanaerobacterium xylanolyticum***; *Thermobacillus* sp., включая *Thermobacillus composti*, *Thermobacillus xylanilyticus*; *Thermocrinis* sp., включая *Thermocrinis albus*, *Thermocrinis ruber*; *Thermodulfatator* sp., включая *Thermodesulfatator atlanticus*, *Thermodesulfatator autotrophicus*, *Thermodesulfatator indicus*; *Thermodesulfobacterium* sp., включая *Thermodesulfobacterium commune*, *Thermodesulfobacterium hydrogeniphilum*; *Thermodesulfobium* sp., включая *Thermodesulfobium narugense*; *Thermodesulfovibrio* sp., включая *Thermodesulfovibrio aggregans*, *Thermodesulfovibrio thiophilus*, *Thermodesulfovibrio yellowstonii*; *Thermosipho* sp., включая *Thermosipho africanus*, *Thermosipho atlanticus*, *Thermosipho melanesiensis*; *Thermotoga* sp., включая ***Thermotoga maritima***, *Thermotoga neopolitana*, ***Thermotoga* sp. RQ7**; *Thermovibrio* sp., включая *Thermovibrio ammonificans*, *Thermovibrio ruber*; *Thermovirga* sp., включая *Thermovirga lienii*, и *Thermus* sp., включая ***Thermus aquaticus***, ***Thermus caldophilus***, ***Thermus flavus***, ***Thermus scotoductus***, ***Thermus thermophilus***; *Thiobacillus neapolitanus*.

В другом аспекте способ или вариант применения, описанный в настоящем документе, может быть использован для модификации бактерий, которые являются мезофильными. В предпочтительных вариантах осуществления бактерий выбирают из: *Acidithiobacillus* sp., включая ***Acidithiobacillus caldus***; *Actinobacillus* sp., включая ***Actinobacillus succinogenes***; *Anaerobiospirillum* sp., включая *Anaerobiospirillum succiniciproducens*; *Bacillus* sp., включая ***Bacillus alcaliphilus***, ***Bacillus amyloliquefaciens***, *Bacillus circulans*, ***Bacillus cereus***, ***Bacillus clausii***, *Bacillus firmus*, *Bacillus halodurans*, *Bacillus lautus*, *Bacillus lentus*, ***Bacillus licheniformis***, ***Bacillus megaterium***, *Bacillus pumilus*, ***Bacillus subtilis***, ***Bacillus thuringiensis***; *Basfia* sp., включая ***Basfia succiniciproducens***; *Brevibacillus* sp., включая ***Brevibacillus brevis***; ***Brevibacillus laterosporus***; *Clostridium*

sp., включая ***Clostridium acetobutylicum***, ***Clostridium autoethanogenum***, ***Clostridium beijerinkii***, *Clostridium carboxidivorans*, ***Clostridium cellulolyticum***, ***Clostridium ljungdahlii***, ***Clostridium pasteurianum***, ***Clostridum perfringens***, *Clostridium ragsdalei*, *Clostridium saccharobutylicum*, *Clostridium saccharoperbutylacetonium*; *Corynebacterium* sp., включая ***Corynebacterium glutamicum***; *Desulfitobacterium* sp., включая ***Desulfitobacterium dehalogenans***, *Desulfitobacterium hafniense*; *Desulfotomaculum* sp., включая *Desulfotomaculum acetoxidans*, *Desulfotomaculum gibsoniae*, *Desulfotomaculum reducens*, *Desulfotomaculum ruminis*; *Enterobacter* sp., включая *Enterobacter asburiae*; *Enterococcus* sp., включая ***Enterococcus faecalis***; *Escherichia* sp., включая ***Escherichia coli***; *Lactobacillus* sp., включая ***Lactobacillus acidophilus***, ***Lactobacillus amylophilus***, ***Lactobacillus amylovorus***, *Lactobacillus animalis*, *Lactobacillus arizonensis*, ***Lactobacillus bavaricus***, ***Lactobacillus brevis***, ***Lactobacillus buchneri***, ***Lactobacillus bulgaricus***, ***Lactobacillus casei***, ***Lactobacillus corynoformis***, ***Lactobacillus crispatus***, ***Lactobacillus curvatus***, ***Lactobacillus delbrueckii***, ***Lactobacillus fermentum***, ***Lactobacillus gasseri***, ***Lactobacillus helveticus***, ***Lactobacillus johnsonii***, ***Lactobacillus pentosus***, ***Lactobacillus plantarum***, ***Lactobacillus reuteri***, ***Lactobacillus rhamnosus***, ***Lactobacillus sakei***, ***Lactobacillus salivarius***, ***Lactobacillus sanfriscensis***; *Mannheimia* sp., включая ***Mannheimia succiniciproducens***; *Paenibacillus* sp., включая ***Paenibacillus alvei***, *Paenibacillus beijingensis*, *Paenibacillus borealis*, *Paenibacillus dauci*, *Paenibacillus durus*, *Paenibacillus graminis*, ***Paenibacillus larvae***, *Paenibacillus lentimorbus*, *Paenibacillus macerans*, *Paenibacillus mucilaginosus*, *Paenibacillus odorifer*, ***Paenibacillus polymyxa***, *Paenibacillus stellifer*, *Paenibacillus terrae*, *Paenibacillus wulumuqiensis*; *Pediococcus* sp., включая ***Pediococcus acidilactici***, *Pediococcus claussenii*, *Pediococcus ethanolidurans*, ***Pediococcus pentosaceus***; *Salmonella typhimurium*; *Sporolactobacillus* sp., включая *Sporolactobacillus inulinus*, *Sporolactobacillus laevolacticus*; *Staphylococcus aureus*;

Streptococcus sp., включая ***Streptococcus agalactiae***, ***Streptococcus bovis***, *Streptococcus equisimilis*, *Streptococcus feacalis*, *Streptococcus mutans*, *Streptococcus oralis*, *Streptococcus pneumonia*, *Streptococcus pyogenes*, *Streptococcus salivarius*, ***Streptococcus thermophilus***, *Streptococcus sobrinus*, *Streptococcus uberis*; *Streptomyces* sp., включая *Streptomyces achromogenes*, *Streptomyces avermitilis*, ***Streptomyces coelicolor***, *Streptomyces griseus*, ***Streptomyces lividans***, ***Streptomyces parvulus***, ***Streptomyces venezuelae***, ***Streptomyces vinaceus***; *Tetragenococcus* sp., включая ***Tetragenococcus halophilus***, и *Zymomonas* sp., включая ***Zymomonas mobilis***.

В следующем аспекте способ или вариант применения, описанный в настоящем документе, может быть использован для модификации генома дрожжей или грибов. В конкретных вариантах осуществления грибы являются мезофильными, предпочтительно, грибы выбирают из: *Aspergillus* sp., включая, но без ограничения, ***Aspergillus nidulans***, ***Aspergillus niger***, *Aspergillus terreus*, *Aspergillus oryzae* и *Aspergillus terreus*, более предпочтительно, *Aspergillus* sp. представляет собой *Aspergillus nidulans* или *Aspergillus niger*. Альтернативно, мезофильные виды грибов могут представлять собой вид *Candida*.

Изобретение также относится к применению способа, описанного в настоящем документе, для модификации видов дрожжей или грибов, которые являются термофильными, предпочтительно, грибы или дрожжи выбирают из: *Aspergillus* sp., включая ***Aspergillus fumigatus***, ***Aspergillus nidulans***, *Aspergillus terreus*, *Aspergillus versicolor*; *Canariomyces* sp., включая *Canariomyces thermophile*; *Chaetomium* sp., включая *Chaetomium mesopotamicum*, ***Chaetomium thermophilum***; *Candida* sp., включая *Candida bovina*, *Candida sloofii*, *Candida thermophila*, ***Candida tropicalis***, ***Candida krusei*** (=*Issatchenkia orientalis*); *Cercophora* sp., включая *Cercophora coronate*, *Cercophora septentrionalis*; *Coonemeria* sp., включая *Coonemeria aegyptiaca*; *Corynascus* sp., включая *Corynascus thermophiles*; *Geotrichum* sp., включая *Geotrichum candidum*; *Kluyveromyces* sp., включая ***Kluyveromyces fragilis***, ***Kluyveromyces marxianus***; *Malbranchea*

sp., включая *Malbranchea cinnamomea*, *Malbranchea sulfurea*; *Melanocarpus* sp., включая *Melanocarpus albomyces*;*Myceliophtora* sp., включая *Myceliophthora fergusii*, *Myceliophthora thermophila*; *Mycothermus* sp., включая *Mycothermus thermophiles* (=*Scytalidium thermophilum*/*Torula thermophila*); *Myriococcum* sp., включая *Myriococcum thermophilum*; *Paecilomyces* sp., включая *Paecilomyces thermophila*; *Remersonia* sp., включая *Remersonia thermophila*; *Rhizomucor* sp., включая ***Rhizomucor pusillus***, *Rhizomucor tauricus*; ***Saccharomyces*** sp., включая ***Saccharomyces cerevisiae***, ***Schizosaccharomyces*** sp., включая ***Schizosaccharomyces pombe***, Scytalidium sp., включая *Scytalidium thermophilum*; *Sordaris* sp., включая *Sordaria thermophila*; *Thermoascus* sp., включая *Thermoascus aurantiacus*, *Thermoascus thermophiles*; *Thermomucor* sp., включая *Thermomucor indicae-seudaticae*, и *Thermomyces* sp., включая *Thermomyces ibadanensis*, ***Thermomyces lanuginosus***.

В вышеприведенных списках микроорганизмы, выделенные **жирным шрифтом**, как установлено, являются особенно подходящими/пригодными для использования по настоящему изобретению.

Некоторые предпочтительные варианты осуществления настоящего изобретения включают один или более термофильных микроорганизмов, выбранных из: термофильных бацилл, включая *Aeribacillus, Alicyclobacillus, Anoxybacillus, Bacillus, Geobacillus*; *Paenibacillus* sp.; термофильных клостридий, включая *Anaerobacter, Anaerobacterium, Caldicellulosiruptor, Clostridium, Moorella, Thermoanaerobacter, Thermoanaerobacterium, Thermobrachium, Thermohalobacter* sp., или одного или более термофильных видов *Lactobacillus*, а также мезофильные бактерии, выбранные из *Bacillus* sp., *Escherichia coli* и *Lactobacillus* sp.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

Далее изобретение будет подробно описано со ссылкой на конкретные варианты осуществления и со ссылкой на прилагаемые чертежи, в которых:

На фигуре 1 показано филогенетическое дерево

последовательностей белков Cas9, построенное методом присоединения соседей. Были включены все последовательности, имеющие сходство последовательностей со штаммом T12 выше 40% на основании результатов pBLAST или PSI-BLAST, и хорошо охарактеризованные в настоящее время последовательности (*S. pyogenes, S. thermophiles* и *A. naeslundii*), а также все идентифицированные в настоящее время последовательности, имеющие идентичность ниже 40%. Для всех термофильных последовательностей процент идентичности с T12 указан после названия штамма. Числовые идентификаторы генов (gi) указаны перед названием вида. Пояснение: заштрихованные кружки: термофильные (оптимум выше 60°C) последовательности Cas9, заштрихованные квадраты: термотолерантные (оптимум <50°C) последовательности Cas9, незаштрихованный треугольник: последовательность Cas9, наиболее часто используемая в настоящее время для редактирования генома, мезофильного происхождения; без символа: мезофильная последовательность Cas9. Величины в узловых точках представляют собой значения бутстрап-анализа 1000-репликатов; масштабная шкала представляет собой оценочные аминокислотные замены для каждого сайта.

На фигуре 2 показано филогенетическое дерево последовательностей генов Cas9, построенное методом присоединения соседей. Идентичность на уровне генов была крайне слабой; для выравнивания генов использовали последовательности из тех же организмов, из которых были использованы последовательности для выравнивания белков. Числовые идентификаторы генов (gi) указаны перед названием вида. Пояснение: заштрихованные кружки: термофильные (оптимум выше 60°C) последовательности Cas9, заштрихованные квадраты: термотолерантные (оптимум <50°C) последовательности Cas9, незаштрихованный треугольник: последовательность Cas9, наиболее часто используемая в настоящее время для редактирования генома, мезофильного происхождения; без символа: мезофильная последовательность Cas9. Величины в узловых точках представляют собой значения бутстрап-анализа 1000-репликатов.

На фигуре 3 показано выравнивание белковых последовательностей Cas9 для gtCas9 (SEQ ID NO: 1) (тип II-C) и хорошо охарактеризованных последовательностей типа II-C (*A. naeslundii*/«ana»; SEQ ID NO: 8) и типа II-A (*S. pyogenes*/«pyo»; SEQ ID NO: 9 и *S. thermophilus*). Важные остатки активного центра высоко консервативны и указаны черными стрелками. Белковые домены, описанные для Ana-Cas9 и Pyo-Cas9 (Jinek, *et al.*, 2014, Science 343: 1247997), указаны заштрихованными боксами и буквами того же цвета. Домен узнавания PAM определен для системы типа II-A *S. pyogenes*, но ни для одной из систем типа II-C, и, вследствие этого, указан только в последовательности *S. pyogenes*.

На фигуре 4 показана белковая структура Cas9 *A. naeslundii* (Cas9-Ana) (Jinek *et al.*, 2014). gtCas9 относится к той же системе CRISPR типа II-C, и остатки активного центра могут быть идентифицированы.

На фигуре 5 показано сравнение crРНК-направляемого нацеливания на комплементарную дцДНК. Спаривание оснований показано пунктирными линиями. РНК изображена черным цветом, ДНК - серым. Спаривание оснований между спейсером crРНК и протоспейсером-мишенью показано толстой черной пунктирной линией, спаривание оснований между цепями ДНК и между цепями РНК показано толстыми серыми пунктирными линями. Указан 5'-конец crРНК. Обратите внимание, что PAM (маленький белый бокс) в типе I находится ниже цепи-мишени (протоспейсера), тогда как в типе II он находится на другом конце сдвинутой цепи. Аналогично, затравка (предсказанная последовательность гида, с которой начинается спаривание оснований с цепью-мишенью ДНК, и где недопустимы ошибочные спаривания) находится близко к PAM, и, в силу этого, отличается в типах I и II (Van der Oost, 2014, там же). На панели A показана схема системы Cascade типа I *E. coli*. crРНК имеет внутренний спейсер (серый бокс, 31-32 нт, который позволяет узнавать мишень), фланкированный с двух сторон 8-нт 5′-рукояткой и 29-нт 3′-рукояткой, имеющей структуру типа «стебель-петля» (шпилька) (Jore 2011, там же). На панели B показана схема

системы Cas9 типа II *S. pyogenes*. Происходит спаривание оснований crРНК с tracrРНК, что делает возможным процессинг РНКазой III (противоположные черные треугольники). Кроме того, 5'-конец crРНК обрезается РНКазой (черный треугольник), что, как правило, приводит к образованию 20-нт спейсера. Обратите внимание, что можно вводить синтетическую петлю для связывания crРНК и tracrРНК, в результате чего образуется единая гид-РНК (егРНК) (Jinek *et al.*, 2012, там же).

На фигуре 6 показано выравнивание последовательностей системы CRISPR типа IIc *G.thermodenitrificans* T12.

На фигуре 7 показано шесть одиночных хитов, полученных с целью *in silico* предсказания PAM для gtCas9.

На фигуре 8 показано графическое представление паттернов WebLogo, объединяющее результаты выравнивания, приведенного на Фигуре 7. WebLogo создавали с использованием weblogo.berkeley.edu.

На фигуре 9 представлены результаты *in vitro* анализа расщепления при 60°C плазмид-мишеней очищенной gtCas9. Плазмиды содержали конкретные 8-нуклеотидные варианты последовательности PAM.

На фигуре 10 представлены результаты *in vitro* анализов для изучения эффекта концентрации gtCas9 с использованием плазмиды-мишени с последовательностью PAM CCCCCCAA [SEQ ID NO: 11].

На фигуре 11 представлены результаты *in vitro* анализов в диапазоне температур с использованием плазмиды-мишени с последовательностью PAM CCCCCCAA [SEQ ID NO: 11].

На фигуре 12 представлены результаты *in vivo* редактирования генома клеток *Bacillus smithii* ET138 с использованием gtCas9 и 8-нт последовательностей PAM, в виде роста или отсутствия колоний клеток *Bacillus smithii* ET138 на чашках с селективной средой, как описано в примере 9. Колонии указаны стрелками на фигуре 12.

На фигуре 13 представлены результаты ПЦР-скрининга колоний клеток, в которых ген *pyrF* был делетирован. Колонии были получены после трансформации клеток *Bacillus smithii* ET138

конструктом 3 (отрицательный контроль). 15 колоний были подвергнуты скринингу, но не в одной из них не был обнаружен делеционный генотип - полоса, соответствующая размеру 2,1 т.п.н., но вместо этого, у всех был обнаружен генотип дикого типа - полоса, соответствующая размеру 2,9 т.п.н., как описано в примере 9.

На фигуре 14 представлены результаты ПЦР-скрининга колоний клеток, в которых ген *pyrF* был делетирован. Колонии были получены после трансформации клеток *Bacillus smithii* ET138 конструктом 1 (последовательность PAM ATCCCCAA [SEQ ID NO: 21]). 20 колоний были подвергнуты скринингу, и у одной был обнаружен делеционный генотип - полоса, соответствующая размеру 2,1 т.п.н., в то время как у остальных наблюдали как полосу, соответствующую размеру 2,9 т.п.н., и так и делеционный генотип - полосу, соответствующую размеру 2,1 т.п.н., как описано в примере 9. Не были обнаружены генотипы исключительно дикого типа.

Ниже приведены полинуклеотидные и аминокислотные последовательности белков Cas, используемых в соответствии с изобретением.

[SEQ ID NO: 1] АК последовательность белка Cas9 *Geobacillus thermodenitrificans* T12

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRIFDRAENPKTGESLALPRRLARSARRR
LRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNNDELARILLHLAK
RRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVAR
DDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPK
ATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFK
GLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYGKGAAKSFRPIDFDTFGYALTMFKDDTDI
RSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTAC
ERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIELARELSQ
SFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLQPIEIER
LLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLGSERWQQFETFVLTNK
QFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRI
TAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPQFPQPW
PHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGID
ERSGKIQTVVKKKLSEIQLDKTGHFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKK

NGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPIYTIDMMKGILPN
KAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIDSS
NGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL*

[SEQ ID NO: 7] Последовательность ДНК Cas9 *Geobacillus thermodenitrificans* T12

ATGAAGTATAAAATCGGTCTTGATATCGGCATTACGTCTATCGGTTGGGCTGTCATTAA
TTTGGACATTCCTCGCATCGAAGATTTAGGTGTCCGCATTTTTGACAGAGCGGAAAACCCGAAA
ACCGGGGAGTCACTAGCTCTTCCACGTCGCCTCGCCCGCTCCGCCCGACGTCGTCTGCGGCGTC
GCAAACATCGACTGGAGCGCATTCGCCGCCTGTTCGTCCGCGAAGGAATTTTAACGAAGGAAGA
GCTGAACAAGCTGTTTGAAAAAAAGCACGAAATCGACGTCTGGCAGCTTCGTGTTGAAGCACTG
GATCGAAAACTAAATAACGATGAATTAGCCCGCATCCTTCTTCATCTGGCTAAACGGCGTGGAT
TTAGATCCAACCGCAAGAGTGAGCGCACCAACAAAGAAAACAGTACGATGCTCAAACATATTGA
AGAAAACCAATCCATTCTTTCAAGTTACCGAACGGTTGCAGAAATGGTTGTCAAGGATCCGAAA
TTTTCCCTGCACAAGCGTAATAAAGAGGATAATTACACCAACACTGTTGCCCGCGACGATCTTG
AACGGGAAATCAAACTGATTTTCGCCAAACAGCGCGAATATGGGAACATCGTTTGCACAGAAGC
ATTTGAACACGAGTATATTTCCATTTGGGCATCGCAACGCCCTTTTGCTTCTAAGGATGATATC
GAGAAAAAAGTCGGTTTCTGTACGTTTGAGCCTAAAGAAAAACGCGCGCCAAAAGCAACATACA
CATTCCAGTCCTTCACCGTCTGGGAACATATTAACAAACTTCGTCTTGTCTCCCCGGGAGGCAT
CCGGGCACTAACCGATGATGAACGTCGTCTTATATACAAGCAAGCATTTCATAAAAATAAAATC
ACCTTCCATGATGTTCGAACATTGCTTAACTTGCCTGACGACACCCGTTTTAAAGGTCTTTTAT
ATGACCGAAACACCACGCTGAAGGAAAATGAGAAAGTTCGCTTCCTTGAACTCGGCGCCTATCA
TAAAATACGGAAAGCGATCGACAGCGTCTATGGCAAAGGAGCAGCAAAATCATTTCGTCCGATT
GATTTTGATACATTTGGCTACGCATTAACGATGTTTAAAGACGACACCGACATTCGCAGTTACT
TGCGAAACGAATACGAACAAAATGGAAAACGAATGGAAAATCTAGCGGATAAAGTCTATGATGA
AGAATTGATTGAAGAACTTTTAAACTTATCGTTTTCTAAGTTTGGTCATCTATCCCTTAAAGCG
CTTCGCAACATCCTTCCATATATGGAACAAGGCGAAGTCTACTCAACCGCTTGTGAACGAGCAG
GATATACATTTACAGGGCCAAAGAAAAAACAGAAAACGGTATTGCTGCCGAACATTCCGCCGAT
CGCCAATCCGGTCGTCATGCGCGCACTGACACAGGCACGCAAAGTGGTCAATGCCATTATCAAA
AAGTACGGCTCACCGGTCTCCATCCATATCGAACTGGCCCGGGAACTATCACAATCCTTTGATG
AACGACGTAAAATGCAGAAAGAACAGGAAGGAAACCGAAAGAAAAACGAAACTGCCATTCGCCA
ACTTGTTGAATATGGGCTGACGCTCAATCCAACTGGGCTTGACATTGTGAAATTCAAACTATGG
AGCGAACAAAACGGAAAATGTGCCTATTCACTCCAACCGATCGAAATCGAGCGGTTGCTCGAAC
CAGGCTATACAGAAGTCGACCATGTGATTCCATACAGCCGAAGCTTGGACGATAGCTATACCAA
TAAAGTTCTTGTGTTGACAAAGGAGAACCGTGAAAAAGGAAACCGCACCCCAGCTGAATATTTA
GGATTAGGCTCAGAACGTTGGCAACAGTTCGAGACGTTTGTCTTGACAAATAAGCAGTTTTCGA
AAAAGAAGCGGGATCGACTCCTTCGGCTTCATTACGATGAAAACGAAGAAAATGAGTTTAAAAA

TCGTAATCTAAATGATACCCGTTATATCTCACGCTTCTTGGCTAACTTTATTCGCGAACATCTC
AAATTCGCCGACAGCGATGACAAACAAAAAGTATACACGGTCAACGGCCGTATTACCGCCCATT
TACGCAGCCGTTGGAATTTTAACAAAAACCGGGAAGAATCGAATTTGCATCATGCCGTCGATGC
TGCCATCGTCGCCTGCACAACGCCGAGCGATATCGCCCGAGTCACCGCCTTCTATCAACGGCGC
GAACAAAACAAAGAACTGTCCAAAAAGACGGATCCGCAGTTTCCGCAGCCTTGGCCGCACTTTG
CTGATGAACTGCAGGCGCGTTTATCAAAAAATCCAAAGGAGAGTATAAAAGCTCTCAATCTTGG
AAATTATGATAACGAGAAACTCGAATCGTTGCAGCCGGTTTTTGTCTCCCGAATGCCGAAGCGG
AGCATAACAGGAGCGGCTCATCAAGAAACATTGCGGCGTTATATCGGCATCGACGAACGGAGCG
GAAAAATACAGACGGTCGTCAAAAAGAAACTATCCGAGATCCAACTGGATAAAACAGGTCATTT
CCCAATGTACGGGAAAGAAAGCGATCCAAGGACATATGAAGCCATTCGCCAACGGTTGCTTGAA
CATAACAATGACCCAAAAAAGGCGTTTCAAGAGCCTCTGTATAAACCGAAGAAGAACGGAGAAC
TAGGTCCTATCATCCGAACAATCAAAATCATCGATACGACAAATCAAGTTATTCCGCTCAACGA
TGGCAAAACAGTCGCCTACAACAGCAACATCGTGCGGGTCGACGTCTTTGAGAAAGATGGCAAA
TATTATTGTGTCCCTATCTATACAATAGATATGATGAAAGGGATCTTGCCAAACAAGGCGATCG
AGCCGAACAAACCGTACTCTGAGTGGAAGGAAATGACGGAGGACTATACATTCCGATTCAGTCT
ATACCCAAATGATCTTATCCGTATCGAATTTCCCCGAGAAAAAACAATAAAGACTGCTGTGGGG
GAAGAAATCAAAATTAAGGATCTGTTCGCCTATTATCAAACCATCGACTCCTCCAATGGAGGGT
TAAGTTTGGTTAGCCATGATAACAACTTTTCGCTCCGCAGCATCGGTTCAAGAACCCTCAAACG
ATTCGAGAAATACCAAGTAGATGTGCTAGGCAACATCTACAAAGTGAGAGGGGAAAAGAGAGTT
GGGGTGGCGTCATCTTCTCATTCGAAAGCCGGGGAAACTATCCGTCCGTTATAA

ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

Пример 1: Выделение *Geobacillus thermodenitrificans*

Микроорганизм *G. thermodenitrificans* был неожиданно обнаружен в библиотеке из ± 500 изолятов при поиске термофила, способного расщеплять лигноцеллюлозные субстраты в анаэробных условиях. Сначала создавали библиотеку из ± 500 изолятов, которая, после нескольких раундов селекции путем выделения на целлюлозе и ксилане, была сокращена до 110 изолятов. Эта библиотека из 110 изолятов состояла только из изолятов *Geobacillus*, при этом на 79% библиотека состояла из *G. thermodenitrificans*.

Выделенный штамм *G. thermodenitrificans* был назван «T12». Белок Cas9 из *G. thermodenitrificans* T12 был назван «gtCas9».

Пример 2: Определение основных консенсусных последовательностей для Cas9 в *Geobacillus thermodenitrificans*

Были выполнены следующие поиски в базе данных и выравнивания:

Выполняли поиск с помощью pBLAST и nBLAST на сервере BLAST в собственной организации, при этом последовательность либо белка, либо гена, *G. thermodenitrificans* T12 использовали в качестве искомой последовательности. Эта база данных в последний раз была обновлена в мае 2014 г. и, таким образом, не содержит самые последние добавленные геномы *Geobacillus*, однако нормальный онлайн поиск BLAST не использовали для предотвращения обнародования последовательности T12. Последовательности с установленной степенью идентичности, составляющей более 40% при поиске BLAST, включены в фигуру 1.

Для включения более новых данных о последовательностях использовали последовательность *Geobacillus* MAS1 (с наибольшей степенью родства с gtCas9) для проведения поиска PSI-BLAST на веб-сайте NCBI (Johnson *et al.*, 2008 Nucleic Acids Res. 36(Web Server issue): W5-9). Проводили два последовательных раунда PSI-BLAST, в которых лишь последовательности, отвечающие следующим критериям, использовали в следующем раунде: минимальное покрытие последовательности 96% в первом раунде и 97% во втором и третьем раунде, минимальная идентичность 40%, только по одному штамму на вид.

Последовательности, полученные после проведения поиска PSI-BLAST, а также последовательности с более, чем 40% идентичностью с T12, полученные после поиска pBLAST на внутреннем сервере, которые не проявились при PSI-BLAST, выравнивали совместно с хорошо охарактеризованными в настоящее время мезофильными последовательностями и всеми идентифицированными в настоящее время термофильными последовательностями, также, если они были связаны более отдаленным родством, на основании чего создавали филогенетическое дерево методом присоединения соседей (смотри фигуру 1). Выравнивание проводили в Mega6 с использованием ClustalW, после чего создавали дерево методом присоединения соседей и проводили бутстрап-анализ с использованием 1000 репликатов.

При проведении поиска BLASTn с использованием *Geobacillus*

sp. MAS1 в качестве искомой последовательности, был идентифицирован лишь Cas9 *Geobacillus sp.* JF8 с идентичностью 88%, что указывало на очень небольшую гомологию на генном уровне. На фигуре 2 показано построенное методом присоединения соседей филогенетическое дерево последовательностей генов Cas9, выровненных при помощи Clustal.

Белковые последовательности *G. thermodenitrificans* T12, *A. naeslundii* и *S. pyogenes* дополнительно анализировали на гомологию белковых доменов (смотри фигуру 3) путем выравнивания их в CloneManager с использованием BLOSUM62 с параметрами по умолчанию.

Пример 3: Идентификация основных аминокислотных мотивов, необходимых для функции CAS9, и тех, которые придают термостабильность, в термофильных нуклеазах Cas9

Процентная идентичность вышеописанных выровненных белковых последовательностей показана на фигуре 1. gtCas9 относится к типу II-C. Наиболее хорошо изученной и недавно кристаллизованной структурой системы типа II-C является система из *Actinomyces naeslundii* (Jinek *et al.*, 2014, Science 343: 1247997). Эта белковая последовательность имеет лишь 20% идентичности с gtCas9, но может быть использована для оценки высоко консервативных остатков. Две хорошо охарактеризованные системы типа II-A (*S. pyogenes* и *S. thermophilus*) также были включены в анализ (Jinek *et al.*, 2014, Science 343: 1247997; Nishimasu *et al.*, 2014, Cell 156: 935-949). Выровненные последовательности этих четырех белков представлены на фигуре 3; на фигуре 4 показана белковая структура, определенная для *A. naeslundii* («Ana-Cas9») (Jinek *et al.*, 2014, Science 343: 1247997). Длины Cas9 из t12 (gtCas9) и *Actinomyces naeslundii* являются очень сходными (*A. naeslundii* 1101 ак, gtCas9 1082 ак), и ожидается, что gtCas9 будет иметь аналогичную белковую структуру, однако это еще предстоит определить, поскольку общая идентичность последовательности с cas9-Ana составляет лишь 20%. Все остатки активного центра, описанные Jinek *et al.* (Jinek *et al.*, 2014, Science 343: 1247997) в Cas9 из *A. naeslundii* и *S. pyogenes*, могут быть идентифицированы в gtCas9 (смотри фигуру 3). PAM-

связывающий домен был определен для системы типа II-A *S. pyogenes*, но не для одной из систем типа II-C, и, вследствие этого, указан лишь в последовательности *S. pyogenes*. Кроме того, сайт узнавания PAM сильно варьируется, не только между системами CRISPR, но также и среди видов, содержащих одну и ту же систему.

Пример 4: Определение последовательности PAM для gtCas9 *G. thermodenitrificans*

Установлено, что прокариотические системы CRISPR служат своим хозяевам в качестве адаптивных иммунных систем (Jinek *et al.*, 2012, Science 337: 816-821) и могут быть использованы для быстрых и эффективных генетических манипуляций ((Mali *et al.*, 2013, Nat Methods 10: 957-963).

Белки Cas действуют как специфичные для последовательности нуклеазы для систем CRISPR типа II (Makarova *et al.*, 2011, Nat Rev Micro 9: 467-477). Небольшие молекулы crРНК, которые состоят из «спейсера» (мишени), связанного с областью повторов, являются продуктами транскрипции и процессинга локусов CRISPR. «Спейсеры» обычно происходят из генома бактериофагов и мобильных генетических элементов, однако они также могут быть сконструированы для нацеливания на конкретную нуклеотидную последовательность при генной инженерии (Bikard *et al.*, 2013, Nucleic Acids Research 41: 7429-7437). Молекулы crРНК используются Cas9 в качестве направляющих элементов при идентификации их ДНК-мишеней. Область спейсера идентична области ДНК, являющейся мишенью для расщепления, «протоспейсеру» (Brouns *et al.*, 2012, Science 337: 808-809). PAM (примыкающий к протоспейсеру мотив), находящийся рядом с протоспейсером, необходим для узнавания мишени Cas9 (Jinek *et al.*, 2012, Science 337: 816-821).

Для проведения *in vitro* или *in vivo* исследований определения PAM для систем типа II, необходимо *in silico* предсказать матрицу CRISPR для системы, экспрессирующий tracrРНК модуль. Матрицу CRISPR используют для идентификации модуля crРНК. Экспрессирующая tracrРНК последовательность расположена либо в пределах 500-п.н. окна, фланкирующего Cas9, либо между генами Cas и локусом CRISPR (Chylinski, K., *et al.* (2014)

Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* 42, 6091-6105). tracrРНК должна состоять из 5′-последовательности с высоким уровнем комплементарности с прямыми повторами матрицы CRISPR, за которой следует предсказанная структура с не менее, чем двумя структурами типа стебель-петля и Rho-независимый сигнал терминации транскрипции (Ran, F.A., *et al.* (2015) In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186-191). Молекулу crРНК и tracrРНК затем можно использовать для конструирования химерной молекулы егРНК. 5′-конец егРНК состоит из укороченного спейсера длиной 20 нт, за которым следует укороченный повтор матрицы CRISPR длиной 16-20 нт. За повтором следует соответствующий укороченный анти-повтор и стебель-петля модуля tracrРНК. Части повтора и анти-повтора в егРНК, как правило, соединены линкером GAAA (Karvelis, T., *et al.* (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253).

Гены *cas* (cas9, за которым следуют гены cas1 и cas2) системы CRISPR типа IIc *G. thermodenitrificans* T12 транскрибируются с использованием антисмысловой цепи хромосомы T12. За геном cas2 следует фрагмент ДНК длиной 100 п.н., который в процессе транскрипции образует структуру РНК с несколькими петлями. Очевидно, что данная структура действует в качестве терминатора транскрипции.

Матрица CRISPR с 11 повторами и 10 спейсерными последовательностями расположена выше последовательности терминации транскрипции, и лидер матрицы расположен на 5′-конце матрицы. Ожидается, что локус ДНК, который транскрибируется в tracrРНК, расположен ниже гена *cas9*. Выравнивание 325-п.н. последовательности непосредственно ниже гена *cas9* с 36-п.н. повтором из матрицы CRISPR показало, что существует 36-п.н. последовательность в локусе tracrРНК, почти идентичная повтору (показано на фигуре 6). Этот результат привел авторов изобретения к заключению, что направление транскрипции локуса tracrРНК должно быть противоположным направлению транскрипции

матрицы CRISPR. Следовательно, 5′-конец tracrРНК будет комплементарен 3′-концу crРНК, что приводит к образованию необходимой для Cas9 молекулы двойной РНК.

Пример 5: Создание мишени со случайными PAM

Два разных спейсера из локуса CRISPR II штамма T12 *G. thermodenitrificans* амплифицировали методом ПЦР с использованием геномной ДНК *G. thermodenitrificans* T12 в качестве матрицы. Использовали две пары вырожденных праймеров для амплификации каждого спейсера:

Во-первых, использовали пару, которая приводит к введению шести случайных нуклеотидов выше фрагмента «протоспейсера», что ведет к продуцированию пула протоспейсеров со случайными последовательностями PAM.

Во-вторых, использовали пару, которая приводит к введению шести случайных нуклеотидов ниже фрагмента «протоспейсера», что ведет к продуцированию пула протоспейсеров со случайными последовательностями PAM.

Полученные фрагменты лигировали в вектор pNW33n, получая 4 пула конструктов «протоспейсера», каждый со всеми возможными 4096 разными комбинациями 6-нуклеотидых PAM. Скомпонованную ДНК использовали для трансформации клеток *G. thermodenitrificans* T12. Клетки высевали на селективную среду с хлорамфениколом, более 2×10^6 клеток из каждого пула протоспейсера объединяли. Плазмидную ДНК экстрагировали из пулов, область-мишень амплифицировали методом ПЦР, и продукты отправляли для глубокого секвенирования. PAM с наименьшими считываемыми последовательностями следует считать активными, и процесс следует повторять лишь с конструктами pNW33n, которые содержат спейсеры с этими PAM. Снижение эффективности трансформации *G. thermodenitrificans* T12 должно свидетельствовать об активности PAM.

Пример 6: *In vitro* определение последовательностей PAM для gtCas9

Конструирование вектора pRham:$cas9_{gt}$

Ген $cas9_{gt}$ амплифицировали методом ПЦР из генома *G.*

thermodenitrificans T12, используя праймеры BG6927 и BG6928, и объединяли с вектором pRham C-His Kan (Lucigen) в одной смеси. Смесь использовали для трансформации термокомпетентных клеток *E. cloni* в соответствии с предоставленным протоколом. 100 мкл из трансформационной среды высевали на чашки с LB+50 канамицином для роста в течение ночи при 37°C. Из образовавшихся одиночных колоний *E. cloni*:: pRham:$cas9_{gt}$ случайным образом выбирали 3 и инокулировали в 10 мл среды LB, содержащей 50 мкг/мл канамицина. Из культур готовили глицериновые стоки путем добавления стерильного глицерина к 1 мл из каждой культуры до конечной концентрации 20% (об/об). Глицериновые стоки хранили при -80°C. Оставшиеся 9 мл из каждой культуры использовали для выделения плазмиды в соответствии с протоколом «GeneJET Plasmid Miniprep Kit» (Thermoscientific). Плазмиды отправляли для подтверждения последовательности $cas9_{gt}$, и было подтверждено, что одна из плазмид содержит ген с правильной последовательностью. Соответствующую культуру впоследствии использовали для гетерологичной экспрессии и очистки gtCas9.

Гетерологичная экспрессия gtCas9 в векторе *E. cloni*::pRham:$cas9_{gt}$

Готовили предварительную культуру *E. cloni*:: pRham:$cas9_{gt}$ после инокуляции 10 мл LB+50 канамицина соответствующими глицериновыми стоками. После роста в течение ночи при 37°C и 180 об/мин 2 мл из предварительной культуры использовали для инокуляции 200 мл среды LB+50 канамицина. Культуру *E. cloni*::pRham: $cas9_{gt}$ инкубировали при 37°C, 180 об/мин до достижения OD_{600} 0,7. Затем индуцировали экспрессию gtCas9 путем добавления L-рамнозы до конечной концентрации 0,2% масс/об. Экспрессии позволяли продолжаться в течение 8 ч, после чего культуры центрифугировали в течение 10 минут при 4700 об/мин, 4°C, для сбора клеток. Среду отбрасывали, и осажденные клетки либо хранили при -20°C, либо использовали для получения бесклеточного экстракта (БКЭ) в соответствии со следующим протоколом:

1. Ресупендировать осадок в 20 мл буфера для обработки

ультразвуком (20 мМ натрий-фосфатный буфер (pH 7,5), 100 мМ NaCl, 5 мМ $MgCl_2$, 5%(об/об) глицерина, 1 мМ ДТТ)

2. Разрушать 1 мл клеток путем обработки ультразвуком (8 импульсов по 30 секунд, в промежутке охлаждение на льду в течение 20 секунд)

3. Центрифугировать в течение 15 минут при 35000 g, 4°C, для осаждения нерастворимых фрагментов

4. Удалить супернатант и хранить его при 4°C или на льду.

Проектирование и конструирование библиотеки PAM направляющего модуля егРНК для gtCas9

После *in silico* определения экспрессирующего tracrРНК модуля ДНК в геноме штамма T12 *G. thermodenitrificans* (смотри пример 4, выше), проектировали экспрессирующий единую гид-РНК (егРНК) модуль ДНК, объединяющий модули для crРНК и tracrРНК системы CRISPR/Cas9 в одной молекуле. Спейсер на 5′-конце егРНК проектировали комплементарным с протоспейсером плазмидной библиотеки, и модуль помещали под транскрипционный контроль промотора T7. Модуль ДНК для пT7_егРНК синтезировали с использованием Baseclear и вводили в вектор pUC57, получая вектор pUC57:пT7_егРНК. DH5α-компетентные клетки *E. coli* (NEB) трансформировали вектором, и трансформационную среду высевали на чашки с LB-агаром, содержащим 100 мкг/мл ампициллина. Чашки инкубировали в течение ночи при 37°C. Три из образовавшихся отдельных колоний инокулировали в 10 мл среды LB, содержащей 100 мкг/мл ампициллина. Из культур готовили глицериновые стоки путем добавления стерильного глицерина к 1 мл из каждой культуры до конечной концентрации 20% (об/об). Глицериновые стоки хранили при -80°C. Оставшиеся 9 мл из каждой культуры использовали для выделения плазмиды в соответствии с протоколом «GeneJET Plasmid Miniprep Kit» (Thermoscientific). Выделенную плазмиду использовали в качестве ПЦР-матрицы для амплификации модуля пT7_егРНК. Получали модуль ДНК для пT7_егРНК длиной 218-п.н. (из которых первые 18 п.н. соответствуют пT7) с использованием праймеров BG6574 и BG6575. После завершения ПЦР-смесь подвергали электрофорезу в 1,5% агарозном геле. Полосу, соответствующую

нужному размеру, вырезали и очищали в соответствии с протоколом «Zymoclean™ Gel DNA Recovery Kit».

In vitro транскрипцию (IVT) проводили с использованием набора «HiScribe™ T7 High Yield RNA Synthesis Kit» (NEB). Очищенный модуль ДНК для пТ7_егРНК использовали в качестве матрицы. Смесь для IVT смешивали с равным объемом красителя RNA loading dye (NEB) и нагревали при 70°C в течение 15 минут для разрушения вторичной структуры. После тепловой обработки смесь для IVT подвергали электрофорезу в денатурирующих условиях в ПААГ с мочевиной, и полученный полиакриламидный гель погружали на 10 минут в 100 мл 0,5х буфера TBE, содержащего 10 мкл SYBR Gold (Invitrogen), для окрашивания. Полосу, соответствующую нужному размеру (200 нт), вырезали, и егРНК очищали в соответствии со следующим протоколом очистки РНК:

1. Измельчить содержащие РНК фрагменты геля шпателем и добавить 1 мл буфера для элюирования РНК, оставить на ночь при комнатной температуре.

2. Разнести 330-мкл аликвоты в новые 1,5-мл пробирки.

3. Добавить 3 объема (990 мкл) предварительно охлажденного (-20°C) 100% EtOH.

4. Инкубировать в течение 60 минут при -20°C.

5. Центрифугировать в течение 20 минут при 13000 об/мин в микроцентрифуге при комнатной температуре.

6. Удалить EtOH, промыть осадок 1 мл 70% EtOH.

7. Центрифугировать в течение 5 минут при 13000 об/мин в микроцентрифуге при комнатной температуре.

8. Удалить 990 мкл супернатанта.

9. Выпарить оставшийся EtOH в термомиксере при 55°C в течение 15-20 минут.

10. Ресуспендировать осадок в 20 мкл MQ, хранить при -20°C.

Проектирование и конструирование библиотеки 7-нт PAM и линеаризация библиотеки

Проектировали и конструировали библиотеку PAM на основе вектора pNW33n. 20-п.н. протоспейсер вводили в вектор, фланкированный с его 3′-стороны вырожденной последовательностью

длиной 7 нуклеотидов; вырожденная последовательность служит в качестве PAM, и когда протоспейсер фланкирован правильной PAM, он может быть узнан в качестве мишени егРНК, связанной с Cas9, и расщеплен. Библиотеку PAM готовили в соответствии со следующим протоколом:

1. Приготовить вставку двухцепочечной ДНК SpPAM путем отжига одноцепочечных олигонуклеотидов 1 (BG6494) и 2 (BG6495) ДНК

I. 10 мкл 10x NE буфер 2,1

II. 1 мкл 50 мкМ олиго 1 (~1,125 мкг)

III. 1 мкл 50 мкМ олиго 2 (~1,125 мкг)

IV. 85 мкл MQ

V. Инкубировать смесь при 94°C в течение 5 мин и охлаждать до 37°C со скоростью 0,03°C/сек.

2. Добавить 1 мкл фрагмента Кленова 3'->5' экзополимеразы (NEB) в каждую смесь отожженных олигонуклеотидов, а затем добавить 2,5 мкл 10 мкМ дНТФ. Инкубировать при 37°C в течение 1 ч, а затем при 75°C в течение 20 мин.

3. Добавить ферменты рестрикции: 2 мкл HF-BamHI и 2 мкл BspHI к 46 мкл смеси отжига. Инкубировать при 37°C в течение 1 ч. Этот процесс приведет к образованию вставки SpPAMbb с липкими концами. Использовать набор Zymo DNA cleaning and concentrator kit (Zymo Research) для очистки полученной вставки.

4. Расщепить pNW33n ферментами HF-BamHI и BspHI (NEB) и очистить 3400-п.н. линейный фрагмент pNW33nbb с липкими концами, используя Zymo DNA cleaning and concentrator kit (Zymo Research).

5. Лигировать 50 нг pNW33nBB с 11 нг вставки SPPAMbb, используя лигазу NEB T4 в соответствии с предоставленным протоколом. Очистить смесь для лигирования, используя Zymo DNA cleaning and concentrator kit (Zymo Research).

6. Трансформировать электрокомпетентные клетки DH10b (200 мкл клеток с 500 нг ДНК). Восстанавливать клетки в среде SOC (200 мкл клеток в 800 мкл SOC) в течение часа, а затем инокулировать восстановленные клетки в 50 мл чреды LB+12,5

мкг/мл хлорамфеникола. Инкубировать культуру в течение ночи при 37°C и 180 об/мин.

7. Выделить плазмидную ДНК из культуры клеток, используя JetStar 2.0 maxiprep kit (GENOMED).

8. Провести рестрикцию SapI (NEB) в соответствии с предоставленным протоколом для линеаризации выделенной плазмиды.

Проектирование и проведение реакций определения PAM

Готовили следующую реакционную смесь для расщепления с целью индуцированного gtCas9 введения дцДНК разрывов в представителей библиотеки PAM, содержащих правильную PAM ниже 3'-конца протоспейсера-мишени:

1. 2,5 мкг БКЭ *E. cloni*::pRham:$cas9_{gt}$ на реакцию

2. егРНК до конечной концентрации 30 нМ

3. 200 нг линеаризованной библиотеки PAM на реакцию

4. 2 мкл буфера расщепления (100 мМ натрий-фосфатный буфер (pH 7,5), 500 мМ NaCl, 25 мМ $MgCl_2$, 25% (об/об) глицерина, 5 мМ ДТТ)

5. добавить MQ воду до конечного объема 20 мкл

Реакционную смесь инкубировали в течение 1 ч при 60°C, а затем реакцию останавливали после добавления 4 мкл 6x красителя для нанесения на гель (NEB). Затем реакционную смесь нагружали на 1% агарозный гель. Электрофорез в геле проводили в течение 1 ч и 15 мин при 100 В, а затем гель инкубировали в течение 30 мин в 100 мл 0,5x буфера TAE, содержащего 10 мкл красителя SYBR Gold dye (ThermoFisher). После визуализации полос ДНК в синем свете полосу, соответствующую успешно расщепленным и содержащим PAM фрагментам ДНК, вырезали из геля и очищали из геля с использованием «Zymoclean™ Gel DNA Recovery Kit» в соответствии с предоставленным протоколом.

Маркировка PAM-содержащих расщепленных gtCAs9 фрагментов ДНК для секвенирования

Cas9-индуцированные разрывы ДНК, как правило, бывают внесены между 3-и и 4-м нуклеотидом протоспейсера, проксимально к последовательности PAM. В результате, невозможно спроектировать пару праймеров, которые позволят ПЦР-

амплифицировать PAM-содержащую часть расщепленных фрагментов ДНК, чтобы впоследствии секвенировать и определить последовательность PAM. Для этой цели использовали 5-этапный процесс:

Этап 1: A-наращивание с помощью Taq полимеразы

A-наращивание представляет собой способ нематричного добавления аденина к 3'-концу «тупой» двухцепочечной молекулы ДНК с использованием Taq полимеразы

Компоненты реакции:

- gtCas9-расщепленные и PAM-содержащие фрагменты ДНК - 200 нг
- 10X буфер ThermoPol® (NEB) -5 мкл
- 1 мМ дАТФ - 10 мкл
- Taq ДНК-полимераза (NEB) -0,2 мкл
- H_2O - до конечного объема реакционной смеси 50 мкл
- Время инкубации - 20 мин
- Температура инкубации - 72°C

Этап 2: Конструирование адаптеров секвенирования

Два комплементарных коротких олигонуклеотида оцДНК фосфорилировали и отжигали, получая адаптер секвенирования для PAM-проксимального сайта фрагментов ДНК из этапа 1. Один из олигонуклеотидов имел дополнительный тимин на его 3'-конце для облегчения лигирования адаптера к A-наращенным фрагментам.

Фосфорилирование адаптерных олигонуклеотидов (отдельные реакции фосфорилирования для каждого олигонуклеотида)

- 100 мкМ маточный раствор олигонуклеотида - 2 мкл
- 10X буфер T4 ДНК-лигазы (NEB) -2 мкл
- Стерильная MQ вода - 15 мкл
- T4 Полинуклеотид-киназа (ПНК) (NEB) -1 мкл
- Время инкубации - 60 мин
- Температура инкубации - 37°C
- Инактивация T4 ПНК - 65°C в течение 20 мин

Отжиг фосфорилированных олигонуклеотидов

- Олигонуклеотид 1-5 мкл из соответствующей смеси для фосфорилирования

- Олигонуклеотид 1-5 мкл из соответствующей смеси для фосфорилирования

- Стерильная MQ вода - 90 мкл

- Инкубировать фосфорилированные олигонуклеотиды при 95°C в течение 3 минут. Медленно охлаждать реакционную смесь при комнатной температуре в течение ~ от 30 мин до 1 часа

Этап 3: Лигирование gtCas9-расщепленных, A-наращенных фрагментов с адаптерами секвенирования

Продукты из этапов 1 и 2 лигировали в соответствии со следующим протоколом:

- 10X буфер T4 ДНК-лигазы - 2 мкл

- Продукт из этапа 1-50 нг

- Продукт из этапа 2-4 нг

- T4 ДНК-лигаза - 1 мкл

- Стерильная MQ вода - до 20 мкл

- Время инкубации - 10 мин

- Температура инкубации - 20-25°C

- Инактивация нагреванием 65°C в течение 10 мин

Этап 4: ПЦР-амплификация PAM-содержащего фрагмента длиной 150 нуклеотидов

5 мкл из смеси для лигирования этапа 4 использовали в качестве матрицы для ПЦР-амплификации при помощи Q5 ДНК-полимеразы (NEB). Олигонуклеотид с тиминовым удлинением из этапа 2 использовали в качестве прямого праймера, и обратный праймер был спроектирован для отжига с 150 нуклеотидами, расположенными ниже последовательности PAM.

Ту же последовательность амплифицировали с использованием не обработанной gtCas9 ДНК PAM-библиотеки в качестве матрицы. Оба ПЦР-продукта очищали в геле и отправляли для секвенирования спаренных концов в системе Illumina HiSeq 2500 (Baseclear).

Анализ результатов секвенирования и определение последовательностей-кандидатов PAM

После анализа результатов секвенирования были сконструированы следующие частотные матрицы. Матрицы отражают относительную представленность каждого нуклеотида в каждом

положении PAM в расщепленных и не расщепленных gtCas9 библиотеках:

Не расщепленные	**пол1**	**пол2**	**пол3**	**пол4**	**пол5**	**пол6**	**пол7**
A	19,22	20,83	19,12	24,43	24,59	21,75	18,22
C	34,75	30	31,9	30,54	25,96	27,9	27,17
T	19,16	22,19	25,34	21,28	26,09	26	21,56
G	26,87	26,98	23,64	23,75	23,36	24,35	33,05

Расщепленные	**пол1**	**пол2**	**пол3**	**пол4**	**пол5**	**пол6**	**пол7**
A	10,63	18,65	14,6	14,49	3,36	8,66	27,54
C	**66,22**	**49,59**	**56,82**	**60,35**	**92,4**	**62,26**	34,94
T	8,09	11,21	19,12	12,15	2,35	14,66	5,58
G	15,05	20,54	9,45	13,01	1,89	14,43	31,94

Данные результаты показывают явное предпочтение в отношении мишеней с цитозином в 5-м положении PAM и предпочтение в отношении мишеней с цитозинами в первых 4 положениях PAM.

Пример 7: *In silico* предсказание PAM для gtCas9

In silico предсказания PAM являются возможными, если в геномных базах данных доступно достаточно последовательностей протоспейсера. *In silico* предсказание PAM для gtCas9 начинали с идентификации хитов спейсеров из матрицы CRISPR в геноме штамма T12 *G. thermodenitrificans* путем сравнения с последовательностями в геномных базах данных, таких как GenBank. Использовали инструмент «CRISPR finder» (http://crispr.u-psud.fr/Server/) для идентификации локусов-кандидатов CRISPR в T12. Результат идентификации локусов CRISPR затем загружали в инструмент «CRISPR target» (http://bioanalysis.otago.ac.nz/CRISPRTarget/crispr_analysis.html), который производит поиск в выбранных базах данных и выдает результат в виде подходящих протоспейсеров. Затем проводили скрининг этих последовательностей протоспейсеров с целью поиска уникальных хитов и в отношении комплементарности со спейсерами - например, несовпадения в затравочной последовательности предположительно считали ложноположительными хитами и исключали из дальнейшего анализа. Идентичность хитов с последовательностями профага и (интегрированными) плазмидами

свидетельствовала о том, что полученные хиты являлись истинно положительными. В целом, в результате этого процесса было получено 6 одиночных хитов (фигура 7). Впоследствии фланкирующие области (3′-конец для нуклеазы gtCas типа II) оставшихся, уникальных протоспейсерных хитов выравнивали и сравнивали на наличие консенсусных последовательностей с использованием WebLogo (http://weblogo.berkeley.edu/logo.cgi) (Crooks GE, Hon G, Chandonia JM, Brenner SE WebLogo: A sequence logo generator, *Genome Research*, 14:1188-1190, (2004)) (фигура 8).

Результаты *in silico* сравнивали с экспериментальными результатами *in vitro* идентификации PAM (смотри пример 6), в которых было отмечено предпочтение для цитозина в качестве 5-го остатка последовательности PAM.

Пример 8: Определение 8-нуклеотидной последовательности PAM для gtCas9

Данные *in silico* из примера 8 свидетельствовали о том, что у gtCas9 наблюдалось некоторое предпочтение в отношении аденозина в 8-м положении, вследствие этого, проводили дальнейшие эксперименты по определению PAM, в которых также тестировали 8-е положение последовательности PAM. Это согласуется с характеристикой последовательности PAM для Cas9 мезофильной бактерии *Brevibacillus laterosporus* SSP360D4 (Karvelis *et al.*, 2015), которая, как было установлено, находится между 5-м до 8-м положениями на 3′-конце протоспейсера.

Конкретные варианты последовательности PAM длиной 8 нуклеотидов тестировали с gtCas9:

1) CNCCCCAC [SEQ ID NO: 17],
2) CCCCCCAG [SEQ ID NO: 18],
3) CCCCCCAA [SEQ ID NO: 11],
4) CCCCCCAT [SEQ ID NO: 19],
5) CCCCCCAC [SEQ ID NO: 20],
6) NNNNTNNC (отрицательный контроль для PAM)

После проведения *in vitro* анализа расщепления при 60°C с направлением на эти (не линеаризованные) плазмиды очищенной gtCas9 и той же егРНК, что и ранее (смотри пример 6), наблюдали

повышенную расщепляющую активность gtCas9 при использовании последовательности CCCCCCAA [SEQ ID NO: 11] в качестве PAM (фигура 9). Однако расщепляющая активность была четко детектируемой в случае всех протестированных последовательностей PAM, даже в случае последовательности PAM - отрицательного контроля наблюдали слабую полосу, указывающую на расщепление. Без связи с конкретной теорией, возможно, что использование gtCas9 в высокой концентрации привело к расщеплению, наблюдаемому в случае отрицательного контроля. В целом, сообщалось, что использование Cas9 в высоких концентрациях в *in vitro* анализах приводит к Cas9-индуцированному расщеплению ДНК без строгой необходимости в PAM.

Как правило, известно, что концентрация Cas9 влияет на эффективность Cas9-индуцированного расщепления ДНК (более высокая концентрация Cas9 приводит к более высокой активности Cas9). Это также наблюдали при проведении *in vitro* анализов с использованием плазмиды-мишени с последовательностью PAM CCCCCCAA [SEQ ID NO: 11] и разных концентраций gtCas9 (фигура 10).

Плазмиду-мишень с последовательностью PAM CCCCCCAA [SEQ ID NO: 11] в *in vitro* анализах, описанных выше, использовали в широком диапазоне температур от 38 до 78°C (фигура 11). Удивительно, но gtCas9 была активна при всех температурах, демонстрируя наивысшую активность при температуре от 40,1 до 64,9°C.

Таким образом, оптимальный диапазон температур для Cas9 из видов *Geobacillus* является более широким, чем для белков Cas, охарактеризованных до настоящего времени. Аналогично, верхняя граница диапазона, при которой сохраняется нуклеазная активность, гораздо выше, чем граница для известных белков Cas. Более высокая оптимальная температура и функциональный диапазон обеспечивает значительное преимущество для генной инженерии при высоких температурах и, следовательно, для редактирования геномов термофильных микроорганизмов, которые находят применение в ряде промышленных, сельскохозяйственных и фармацевтических

процессов, проводимых при повышенных температурах.

Пример 9: *In vivo* редактирование генома *Bacillus smithii* ET138 при помощи gtCas9 и 8-нуклеотидных последовательностей PAM

Для подтверждения того, что 8-нуклеотидные PAM также узнаваемы gtCas9 *in vivo*, был спланирован эксперимент для удаления гена *pyrF* в геноме *Bacillus smithii* ET138 при 55°C.

Данный способ включает введение матричного конструкта для гомологичной рекомбинации, в котором имеются области, комплементарные последовательностям выше и ниже гена-мишени (*pyrF)*, в клетки *B. smithii* ET 138. Введение матрицы позволяет использовать процесс гомологичной рекомбинации для введения матрицы для гомологичной рекомбинации (без гена *pyrF)* в геном, так что она также заменяет ген ДТ *pyrF* в геноме клетки.

Включение gtCas9 и егРНК в конструкт для гомологичной рекомбинации можно использовать для внесения двухцепочечных разрывов ДНК (ДЦРД) в бактериальные геномы, которые содержат ген *pyrF* ДТ. ДЦРД в бактериальном геноме, как правило, приводит к клеточной гибели. Таким образом, егРНК, которая узнает последовательность в *pyrF* ДТ, может приводить к ДЦРД и гибели клеток, содержащих только *pyrF* ДТ. Внесение ДЦРД также зависит от нахождения соответствующей последовательности PAM ниже 3'-конца протоспейсера, узнаваемого gtCas9.

Плазмиду pNW33n использовали в качестве каркаса для клонирования:

i) гена $cas9_{gt}$ под контролем разработанного в собственной организации репрессируемого глюкозой промотора; и

ii) областей на 1 т.п.н. выше и 1 т.п.н. ниже гена *pyrF* в геноме *B. smithii* ET138 в качестве матрицы для гомологичной рекомбинации, которая приведет к делеции гена *pyrF* из генома *B. smithii* ET138; и

iii) единой гид-РНК (егРНК), экспрессирующей модуль под транскрипционным контролем конститутивного промотора.

Создавали три отдельных конструкта, в которых последовательности единой гид-РНК отличались в первых 20 нуклеотидах, которые соответствуют последовательности,

направляющей gtCas9 к ее специфической ДНК-мишени в геноме (также известной как спейсер). Были разработаны три разные спейсерные последовательности для нацеливания на три разных протоспейсера-кандидата, все в гене *pyrF B. smithii* ET138. В настоящем документе конструкты названы конструктами 1, 2 и 3, соответственно.

Три разных протоспейсера-мишени имели на их 3′-конце следующую последовательность-кандидат PAM:

1. TCCATTCC (отрицательный контроль в соответствии с результатами *in vitro* анализов; 3′-конец протоспейсера, являющегося мишенью егРНК, закодированной на конструкте номер 3)

2. ATCCCCAA (3′-конец протоспейсера, являющегося мишенью егРНК, закодированной на конструкте номер 1; [SEQ ID NO: 21])

3. ACGGCCAA (3′-конец протоспейсера, являющегося мишенью егРНК, закодированной на конструкте номер 2, [SEQ ID NO: 22]).

После трансформации клеток *B. smithii* ET 138 одним из трех конструктов и высевания на селективную среду были получены следующие результаты:

1. Когда клетки были трансформированы конструктом, направленным на протоспейсер, который имел последовательность TCCATTCC отрицательного контроля PAM на 3′-конце (конструкт номер 3), эффективность трансформации не была изменена (фигура 12 A). Число колоний было в том же диапазоне, что и число колоний после трансформации конструктом положительного контроля pNW33n (фигура 12 B). Из 15 колоний, для которых выполняли ПЦР колоний для скрининга на колонии, в которых ген *pyrF* был делетирован, ни в одной не был обнаружен делеционный генотип - полоса, соответствующая размеру 2,1 т.п.н., во всех случаях наблюдали полосу, соответствующую ожидаемому размеру - 2,9 т.п.н. (фигура 13). Это указывает на то, что тестируемая PAM действительно не была узнана gtCas9 *in vivo*.

2. Когда клетки были трансформированы конструктом номер 1, было получено лишь несколько колоний (фигура 12 C), при сравнении с положительным контролем (клетками, трансформированными pNW33n). Для 20 колоний выполняли ПЦР

колоний для скрининга на колонии, в которых ген *pyrF* был делетирован. Большинство (19) колоний имели генотип как дикого типа, так и с делецией *pyrF*, при том, что одна колония имела генотип с делецией *pyrF* (фигура 14). Этот результат указывает на то, что последовательность PAM ATCCCCAA [SEQ ID NO: 21] узнается *in vivo* gtCas9, поскольку отсутствовали генотипы только ДТ. Сниженная эффективность трансформации также является показателем того, что часть клеточной популяции была редуцирована, что может быть связано с клеточной гибелью, вызванной ДЦРД в клетках с генотипом только ДТ вследствие успешного нацеливания на них gtCas9.

3. Когда клетки были трансформированы конструктом номер 2, колонии полностью отсутствовали (фигура 12 D). Отсутствие колоний является показателем того, что все клетки популяции успешно стали мишенью gtCas9, следствием чего явилась гибель клеток, вызванная ДЦРД. Это свидетельствует о том, что последовательность PAM ACGGCCAA [SEQ ID NO: 22] узнается gtCas9.

Полученные результаты показывают, что gtCas9 является активной при 55°C *in vivo* с вышеуказанными последовательностями PAM, этот факт согласуется с результатами *in vitro* определения PAM. Кроме того, она может быть использована в качестве инструмента редактирования генома при той же температуре в сочетании с находящейся на плазмиде матрицей для гомологичной рекомбинации.

Следующий раздел описания состоит из нумерованных параграфов, в которых просто приведены признаки изобретения, уже описанные в настоящем документе. Нумерованные параграфы в данном разделе не являются пунктами формулы изобретения. Пункты формулы изобретения приведены ниже в разделе «Формула изобретения».

1. Белок или полипептид (Cas), связанный с изолированным кластером регулярно расположенных группами коротких палиндромных повторов (CRISPR), содержащий:

a. аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]; и/или

b. аминокислотный мотив $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2

независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4 представляет собой одно из аланина, глутамата или аргинина; и/или

c. аминокислотный мотив X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и/или

d. аминокислотный мотив X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и/или

e. аминокислотный мотив $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин;

при этом белок Cas способен к расщеплению нуклеиновой кислоты при температуре от 50°C до 100°C, когда он связан с по меньшей мере одной молекулой направляющей РНК и полинуклеотидом, содержащим нуклеотидную последовательность-мишень, узнаваемую молекулой направляющей РНК.

2. Выделенный белок или полипептидный фрагмент Cas, имеющий аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%, при этом белок Cas способен к связыванию, расщеплению, модификации или маркированию полинуклеотида, содержащего нуклеотидную последовательность-мишень, при температуре от 50°C до 100°C, когда он связан с по меньшей мере одной молекулой РНК, которая узнает последовательность-мишень.

3. Белок или полипептидный фрагмент Cas по параграфу 1 или 2, при этом белок или фрагмент Cas способен к связыванию, расщеплению, маркированию или модификации нуклеиновой кислоты при температуре от 50°C до 75°C, предпочтительно, при температуре выше 60°C; более предпочтительно, при температуре от 60°C до 80°C; более предпочтительно, при температуре от 60°C до 65°C.

4. Белок или полипептидный фрагмент Cas по любому из

параграфов 1-3, при этом связывание, расщепление, маркирование или модификация нуклеиновой кислоты представляет собой расщепление ДНК.

5. Белок или полипептидный фрагмент Cas по любому из предшествующих параграфов, при этом аминокислотная последовательность представляет собой аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%.

6. Белок или полипептидный фрагмент Cas по любому из предшествующих параграфов, при этом белок Cas может быть получен из бактерии, архея или вируса.

7. Белок или полипептидный фрагмент Cas по любому из предшествующих параграфов, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

8. Рибонуклеопротеидный комплекс, содержащий белок Cas по любому из предшествующих параграфов и содержащий по меньшей мере одну молекулу направляющей РНК, которая узнает последовательность в полинуклеотиде-мишени.

9. Рибонуклеопротеидный комплекс по параграфу 8, при этом молекула направляющей РНК содержит crРНК и, необязательно, tracrРНК.

10. Рибонуклеопротеидный комплекс по любому из параграфов 7-9, при этом длина по меньшей мере одной молекулы РНК находится в диапазоне 35-135 нуклеотидных остатков.

11. Рибонуклеопротеидный комплекс по параграфу 8 или 9, при этом последовательность-мишень имеет длину 31 или 32 нуклеотидных остатков.

12. Белок или полипептид Cas по любому из параграфов 1-7 или рибонуклеопротеидный комплекс по любому из параграфов 8-11, при этом белок или полипептид предоставлен в виде части белкового комплекса, содержащего по меньшей мере один дополнительный функциональный или не функциональный белок.

13. Белок, полипептид или рибонуклеопротеидный комплекс Cas по параграфу 12, при этом белок или полипептид Cas и/или по меньшей мере один дополнительный белок также содержит по меньшей

мере один функциональный фрагмент.

14. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по параграфу 13, при этом по меньшей мере один функциональный фрагмент слит или связан с N-концом и/или C-концом белка, полипептида или рибонуклеопротеидного комплекса Cas; предпочтительно, N-концом.

15. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по параграфу 13 или 14, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки.

16. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по параграфу 15, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

17. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по параграфу 15 или 16, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

18. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по любому из параграфов 15-17, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок, например, GFP.

19. Выделенная молекула нуклеиновой кислоты, кодирующая белок или полипептид Cas, содержащий:

a. аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]; и/или

b. аминокислотный мотив $X_1X_2CTX_3X_4$ [SEQ ID NO: 3], где X_1 независимо выбирают из изолейцина, метионина или пролина, X_2 независимо выбирают из валина, серина, аспарагина или изолейцина, X_3 независимо выбирают из глутамата или лизина, и X_4

представляет собой одно из аланина, глутамата или аргинина; и/или

c. аминокислотный мотив X_5LKX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина; и/или

d. аминокислотный мотив X_7VYSX_8K [SEQ ID NO: 5], где X_7 представляет собой глутамат или изолейцин и X_8 представляет собой одно из триптофана, серина или лизина; и/или

e. аминокислотный мотив $X_9FYX_{10}X_{11}REQX_{12}KEX_{13}$ [SEQ ID NO: 6], где X_9 представляет собой аланин или глутамат, X_{10} представляет собой глутамин или лизин, X_{11} представляет собой аргинин или аланин, X_{12} представляет собой аспарагин или аланин и X_{13} представляет собой лизин или серин;

при этом белок или полипептид Cas способен к связыванию, расщеплению, маркированию или модификации ДНК при температуре от 50°C до 100°C, когда он связан с по меньшей мере одной молекулой направляющей РНК и полинуклеотидом, содержащим нуклеотидную последовательность-мишень, узнаваемую молекулой направляющей РНК.

20. Выделенная молекула нуклеиновой кислоты, кодирующая белок (Cas), связанный с изолированным кластером регулярно расположенных группами коротких палиндромных повторов (CRISPR), имеющий аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%; или его полипептидный фрагмент.

21. Выделенная молекула нуклеиновой кислоты по параграфу 19 или 20, дополнительно содержащая по меньшей мере одну нуклеотидную последовательность, кодирующую аминокислотную последовательность, которая в процессе трансляции слита с белком или полипептидом Cas.

22. Выделенная молекула нуклеиновой кислоты по параграфу 21, при этом по меньшей мере одна нуклеотидная последовательность, слитая с молекулой нуклеиновой кислоты, кодирующей белок или полипептид Cas, кодирует белок, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-

метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или метки аффинной очистки.

23. Экспрессионный вектор, содержащий молекулу нуклеиновой кислоты по любому из параграфов 19-22.

24. Экспрессионный вектор по параграфу 23, дополнительно содержащий нуклеотидную последовательность, кодирующую по меньшей мере одну молекулу направляющей РНК.

25. Способ модификации нуклеиновой кислоты-мишени, включающий создание контакта нуклеиновой кислоты с:

a. рибонуклеопротеидным комплексом по любому из параграфов 6-11; или

b. белком или белковым комплексом по любому из параграфов 12-18 и по меньшей мере одной молекулой направляющей РНК по любому из параграфов 6-11.

26. Способ модификации нуклеиновой кислоты-мишени в клетке, включающий трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором по параграфу 24; или, альтернативно, трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором по параграфу 23 и дополнительным экспрессионным вектором, содержащим нуклеотидную последовательность, кодирующую молекулу направляющей РНК по любому из параграфов 6-11.

27. Способ модификации нуклеиновой кислоты-мишени в клетке, включающий трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором по параграфу 23, с последующей доставкой молекулы направляющей РНК по любому из параграфов 6-11 к клетке или в клетку.

28. Способ модификации нуклеиновой кислоты-мишени по любому из параграфов 25-28, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок или репортерный белок, и маркерный белок или репортерный белок связывается с нуклеиновой кислотой-мишенью; предпочтительно, при этом маркер представляет собой флуоресцентный белок, например,

зеленый флуоресцентный белок (GFP).

29. Способ по любому из параграфов 25-28, при этом нуклеиновая кислота-мишень представляет собой ДНК; предпочтительно, дцДНК.

30. Способ по любому из параграфов 25-28, при этом нуклеиновая кислота-мишень представляет собой РНК.

31. Способ модификации нуклеиновой кислоты-мишени по параграфу 29, при этом нуклеиновая кислота представляет собой дцДНК, по меньшей мере один функциональный фрагмент представляет собой нуклеазу или геликазу-нуклеазу, и модификация представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе.

32. Способ выключения экспрессии гена в желаемом локусе в соответствии с любым из способов по любому из параграфов 26, 27, 29 или 31.

33. Способ модификации, либо удаления и/или вставки желаемой нуклеотидной последовательности в желаемом локусе в соответствии с любым из способов по любому из параграфов 26, 27, 29 или 31.

34. Способ модификации экспрессии гена в клетке, включающий модификацию нуклеотидной последовательности-мишени в соответствии со способом по любому из параграфов 25-29; при этом нуклеиновая кислота представляет собой дцДНК, и функциональный фрагмент выбирают из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции.

35. Способ модификации экспрессии гена в клетке, включающий модификацию нуклеотидной последовательности-мишени в соответствии со способом по параграфу 30, при этом нуклеиновая кислота представляет собой мРНК, и функциональный фрагмент представляет собой рибонуклеазу; необязательно, выбранную из эндонуклеазы, 3'-экзонуклеазы или 5'-экзонуклеазы.

36. Способ модификации нуклеиновой кислоты-мишени по любому из параграфов 25-35, который применяют при температуре от 50°C до 100°C.

37. Способ модификации нуклеиновой кислоты-мишени по параграфу 36, который применяют при температуре 60°C или выше, предпочтительно, от 60°C до 80°C, более предпочтительно, от 60°C до 65°C.

38. Способ по любому из параграфов 25-37, при этом клетка представляет собой прокариотическую клетку.

39. Способ по любому из параграфов 25-38, при этом клетка представляет собой эукариотическую клетку.

40. Клетка-хозяин, трансформированная способом по любому из параграфов 22-36.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Применение по меньшей мере одной молекулы направляющей РНК и белка Cas для связывания, расщепления, маркирования или модификации двухцепочечного полинуклеотида-мишени, содержащего нуклеотидную последовательность-мишень, при этом:

двухцепочечный полинуклеотид-мишень содержит нуклеотидную цепь-мишень, содержащую указанную нуклеотидную последовательность-мишень, и не являющуюся мишенью нуклеотидную цепь, содержащую протоспейсерную нуклеотидную последовательность, комплементарную нуклеотидной последовательности-мишени;

белок Cas имеет аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%;

по меньшей мере одна молекула направляющей РНК узнает последовательность-мишень;

не являющаяся мишенью нуклеотидная цепь дополнительно содержит последовательность мотива, примыкающего к протоспейсеру (PAM), в непосредственной близости к 3′-концу протоспейсерной нуклеотидной последовательности, при этом последовательность PAM представляет собой 5′-NNNNCNN-3′.

2. Применение по п. 1, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 100°C, при температуре от 30°C до 80°C, при температуре от 37°C до 78°C, предпочтительно, при температуре выше 55°C; более предпочтительно, при температуре от 55°C до 80°C; еще более предпочтительно, при температуре от 55°C до 65°C или от 60°C до 65°C.

3. Применение по п. 1 или п. 2, при этом полинуклеотид, содержащий нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

4. Применение по любому из пунктов 1-3, при этом нуклеотидная цепь-мишень, содержащая последовательность-мишень, представляет собой двухцепочечную ДНК, и указанное применение

приводит к двухцепочечному разрыву в полинуклеотиде, содержащем нуклеотидную последовательность-мишень.

5. Применение по п. 1 или п. 2, при этом полинуклеотид, содержащий нуклеотидную последовательность-мишень, представляет собой двухцепочечную ДНК, белок Cas лишен способности разрезать двухцепочечную ДНК, и указанное применение приводит к выключению гена полинуклеотида.

6. Применение по любому из предшествующих пунктов, при этом последовательность PAM представляет собой 5′-NNNNCNNA-3′, 5′-CNNNCNN-3′, 5′-NNNCCNN-3′, 5′-NNCNCNN-3′, 5′-NNNNCCN-3′ и/или 5′-NCNNCNN-3′.

7. Применение по любому из предшествующих пунктов, при этом последовательность PAM представляет собой 5′-CCCCCCNA-3′ [SEQ ID NO: 10], предпочтительно, при этом последовательность PAM представляет собой 5′-CCCCCCAA-3′ [SEQ ID NO: 11].

8. Применение по любому из предшествующих пунктов, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

9. Применение по любому из предшествующих пунктов, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

10. Применение по любому из предшествующих пунктов, при этом молекула направляющей РНК содержит crРНК и tracrРНК.

11. Применение по любому из предшествующих пунктов, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

12. Применение по любому из предшествующих пунктов, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

13. Применение по любому из предшествующих пунктов, при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

14. Применение по любому из предшествующих пунктов, при этом белок Cas предоставлен в виде части белкового комплекса, содержащего по меньшей мере один дополнительный функциональный

или не функциональный белок, необязательно, при этом по меньшей мере один дополнительный белок также содержит по меньшей мере один функциональный фрагмент.

15. Применение по п. 13 или п. 14, при этом белок Cas или дополнительный белок содержит по меньшей мере один функциональный фрагмент, слитый или связанный с N-концом и/или С-концом белка или белкового комплекса Cas; предпочтительно, С-концом.

16. Применение по любому из пунктов 13-15, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки, например, зеленый флуоресцентный белок (GFP).

17. Применение по п. 16, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

18. Применение по п. 16 или п. 17, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

19. Применение по любому из пунктов 16-18, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

20. Способ связывания, расщепления, маркирования или модификации двухцепочечного полинуклеотида-мишени, при этом двухцепочечный полинуклеотид-мишень содержит нуклеотидную цепь-мишень, содержащую нуклеотидную последовательность-мишень, и не являющуюся мишенью нуклеотидную цепь, содержащую протоспейсерную нуклеотидную последовательность, комплементарную нуклеотидной последовательности-мишени, включающий:

а. проектирование по меньшей мере одной молекулы

направляющей РНК, при этом молекула направляющей РНК узнает последовательность-мишень в цепи-мишени, и не являющаяся мишенью цепь дополнительно содержит последовательность мотива, примыкающего к протоспейсеру (PAM), в непосредственной близости к 3′-концу протоспейсерной последовательности, причем последовательность PAM представляет собой 5′-NNNNCNN-3′;

b. формирование рибонуклеопротеидного комплекса, содержащего молекулу направляющей РНК и белок Cas, при этом выделенный белок Cas имеет аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%; и

c. связывание, расщепление, маркирование или модификацию полинуклеотида-мишени рибонуклеопротеидным комплексом.

21. Способ по п. 20, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 100°C, при температуре от 30°C до 80°C, при температуре от 37°C до 78°C, предпочтительно, при температуре выше 55°C; более предпочтительно, при температуре от 55°C до 80°C; еще более предпочтительно, при температуре от 55°C до 65°C или от 60°C до 65°C.

22. Способ по п. 20 или п. 21, при этом двухцепочечный полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

23. Способ по любому из пунктов 20-22, при этом полинуклеотид-мишень представляет собой двухцепочечную ДНК, и указанное применение приводит к двухцепочечному разрыву в полинуклеотиде.

24. Способ по п. 20 или п. 21, при этом полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, представляет собой двухцепочечную ДНК, белок Cas лишен способности разрезать двухцепочечную ДНК, и указанный способ приводит к выключению гена полинуклеотида-мишени.

25. Способ по любому из пунктов 20-24, при этом

последовательность PAM представляет собой 5′-NNNNCNNA-3′, 5′-CNNNCNN-3′, 5′-NNNCCNN-3′, 5′-NNCNCNN-3′, 5′-NNNNCCN-3′ и/или 5′-NCNNCNN-3′.

26. Способ по любому из пунктов 20-25, при этом последовательность PAM представляет собой 5′-CCCCCCNA-3′ [SEQ ID NO: 10], предпочтительно, при этом последовательность PAM представляет собой 5′-CCCCCCAA-3′ [SEQ ID NO: 11].

27. Способ по любому из пунктов 20-26, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

28. Способ по любому из пунктов 20-27, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

29. Способ по любому из пунктов 20-28, при этом молекула направляющей РНК содержит crРНК и tracrРНК.

30. Способ по любому из пунктов 20-29, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

31. Способ по любому из пунктов 20-30, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

32. Способ по любому из пунктов 20-31, при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

33. Способ по любому из пунктов 20-32, при этом белок Cas предоставлен в виде части белкового комплекса, содержащего по меньшей мере один дополнительный функциональный или не функциональный белок, необязательно, при этом по меньшей мере один дополнительный белок также содержит по меньшей мере один функциональный фрагмент.

34. Способ по п. 32 или 33, при этом белок Cas или дополнительный белок содержит по меньшей мере один функциональный фрагмент, слитый или связанный с N-концом и/или C-концом белка или белкового комплекса Cas; предпочтительно, C-концом.

35. Способ по любому из пунктов 32-34, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки, например, зеленый флуоресцентный белок (GFP).

36. Способ по п. 35, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

37. Способ по п. 35 или п. 36, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

38. Способ по любому из пунктов 35-37, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

39. Применение по п. 16 или способ по п. 35, при этом двухцепочечный полинуклеотид-мишень представляет собой дцДНК, по меньшей мере один функциональный фрагмент представляет собой нуклеазу или геликазу-нуклеазу, и модификация представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе.

40. Применение по п. 16 или способ по п. 35, при этом двухцепочечный полинуклеотид-мишень представляет собой дцДНК, функциональный фрагмент выбирают из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции, и связывание, расщепление, маркирование или модификация приводит к модификации экспрессии гена.

41. Применение по п. 16 или способ по п. 35, при этом указанное связывание, расщепление, маркирование или модификация происходит *in vivo*.

42. Применение по любому из пунктов 1-4, 6-19 или 39, или способ по любому из пунктов 20-23, 25-39, при этом связывание,

расщепление, маркирование или модификация приводит к модификации или удалению и/или вставке желаемой нуклеотидной последовательности в желаемом локусе, и/или при этом связывание, расщепление, маркирование или модификация приводит к выключению экспрессии гена в желаемом локусе.

43. Трансформированная клетка, содержащая двухцепочечный полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, при этом двухцепочечный полинуклеотид-мишень содержит нуклеотидную цепь-мишень, содержащую указанную нуклеотидную последовательность-мишень, и не являющуюся мишенью нуклеотидную цепь, содержащую протоспейсерную нуклеотидную последовательность, комплементарную нуклеотидной последовательности-мишени, при этом указанная клетка содержит:

белок (Cas), связанный с изолированным кластером регулярно расположенных группами коротких палиндромных повторов (CRISPR), имеющий аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%;

по меньшей мере одну молекулу направляющей РНК, которая узнает нуклеотидную последовательность-мишень в нуклеотидной цепи-мишени, при этом не являющаяся мишенью цепь дополнительно содержит последовательность мотива, примыкающего к протоспейсеру (PAM), в непосредственной близости к 3′-концу протоспейсерной последовательности, причем последовательность PAM представляет собой 5′-NNNNCNN-3′; и

экспрессионный вектор, содержащий нуклеиновую кислоту, кодирующую по меньшей мере одно из указанного белка Cas и указанной молекулы направляющей РНК.

44. Трансформированная клетка по п. 43, при этом белок Cas и молекула направляющей РНК осуществляют связывание, расщепление, маркирование или модификацию полинуклеотида-мишени в клетке, и связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 100°C, при температуре от 30°C до 80°C, при температуре от 37°C до 78°C, предпочтительно, при температуре выше 55°C; более

предпочтительно, при температуре от 55°C до 80°C; еще более предпочтительно, при температуре от 55°C до 65°C или от 60°C до 65°C.

45. Трансформированная клетка по п. 43 или п. 44, при этом нуклеотидная цепь-мишень, содержащая нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

46. Трансформированная клетка по любому из пунктов 43-45, при этом полинуклеотид-мишень, содержащий последовательность-мишень, представляет собой двухцепочечную ДНК, и указанное связывание, расщепление, маркирование или модификация приводит к двухцепочечному разрыву в полинуклеотиде-мишени.

47. Трансформированная клетка по п. 43 или п. 44, при этом полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, представляет собой двухцепочечную ДНК, белок Cas лишен способности разрезать двухцепочечную ДНК, и указанное связывание, расщепление, маркирование или модификация приводит к выключению гена полинуклеотида-мишени.

48. Трансформированная клетка по любому из пунктов 43-47, при этом последовательность PAM представляет собой 5′-NNNNCNNA-3′, 5′-CNNNCNN-3′, 5′-NNNCCNN-3′, 5′-NNCNCNN-3′, 5′-NNNNCCN-3′ и/или 5′-NCNNCNN-3′.

49. Трансформированная клетка по любому из пунктов 43-48, при этом последовательность PAM представляет собой 5′-CCCCCCNA-3′ [SEQ ID NO: 10], предпочтительно, при этом последовательность PAM представляет собой 5′-CCCCCCAA-3′ [SEQ ID NO: 11].

50. Трансформированная клетка по любому из пунктов 43-49, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

51. Трансформированная клетка по любому из пунктов 43-50, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

52. Трансформированная клетка по любому из пунктов 43-51, представляющая собой прокариотическую клетку.

53. Трансформированная клетка по любому из пунктов 43-51, представляющая собой эукариотическую клетку.

54. Трансформированная клетка по любому из пунктов 43-53, при этом молекула направляющей РНК содержит crРНК и tracrРНК.

55. Трансформированная клетка по любому из пунктов 43-54, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

56. Трансформированная клетка по любому из пунктов 43-55, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

57. Трансформированная клетка по любому из пунктов 43-56, при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

58. Трансформированная клетка по любому из пунктов 43-57, при этом белок Cas предоставлен в виде части белкового комплекса, содержащего по меньшей мере один дополнительный функциональный или не функциональный белок, необязательно, при этом по меньшей мере один дополнительный белок также содержит по меньшей мере один функциональный фрагмент.

59. Трансформированная клетка по п. 57 или 58, при этом белок Cas или дополнительный белок содержит по меньшей мере один функциональный фрагмент, слитый или связанный с N-концом и/или С-концом белка или белкового комплекса Cas; предпочтительно, N-концом.

60. Трансформированная клетка по любому из пунктов 57-59, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки, например, зеленый флуоресцентный белок (GFP).

61. Трансформированная клетка по п. 60, при этом

естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

62. Трансформированная клетка по любому из пунктов 57-61, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

63. Трансформированная клетка по любому из пунктов 57-61, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

64. Трансформированная клетка по любому из пунктов 57-62, при этом двухцепочечный полинуклеотид-мишень представляет собой дцДНК, по меньшей мере один функциональный фрагмент представляет собой нуклеазу или геликазу-нуклеазу, и модификация представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе.

65. Трансформированная клетка по любому из пунктов 57-61 или способ по п. 35, при этом двухцепочечный полинуклеотид-мишень представляет собой дцДНК, функциональный фрагмент выбирают из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции, и связывание, расщепление, маркирование или модификация приводит к модификации экспрессии гена.

66. Трансформированная клетка по любому из пунктов 57-62, при этом белок Cas экспрессируется с экспрессионного вектора.

67. Трансформированная клетка по любому из пунктов 43-66, при этом связывание, расщепление, маркирование или модификация приводит к модификации или удалению и/или вставке желаемой нуклеотидной последовательности в желаемом локусе, и/или при этом связывание, расщепление, маркирование или модификация приводит к выключению экспрессии гена в желаемом локусе.

68. Нуклеопротеидный комплекс, содержащий белок Cas, по меньшей мере одну молекулу направляющей РНК, которая узнает нуклеотидную последовательность-мишень в двухцепочечном полинуклеотиде-мишени, и полинуклеотид-мишень, при этом

белок Cas имеет аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на

77%;

двухцепочечный полинуклеотид-мишень содержит нуклеотидную цепь-мишень, содержащую указанную нуклеотидную последовательность-мишень, не являющуюся мишенью нуклеотидную цепь, содержащую протоспейсерную нуклеотидную последовательность, комплементарную нуклеотидной последовательности-мишени, и последовательность мотива, примыкающего к протоспейсеру (PAM), в непосредственной близости к 3′-концу протоспейсерной последовательности, причем последовательность PAM представляет собой 5′-NNNNCNN-3′.

69. Нуклеопротеидный комплекс по п. 68, существующий при температуре от 20°C до 100°C, при температуре от 30°C до 80°C, при температуре от 37°C до 78°C, предпочтительно, при температуре выше 55°C; более предпочтительно, при температуре от 55°C до 80°C; еще более предпочтительно, при температуре от 55°C до 65°C или от 60°C до 65°C.

70. Нуклеопротеидный комплекс по п. 68 или п. 69, при этом двухцепочечный полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

71. Нуклеопротеидный комплекс по любому из пунктов 68-70, при этом полинуклеотид-мишень, содержащий последовательность-мишень, представляет собой двухцепочечную ДНК, и указанное связывание, расщепление, маркирование или модификация приводит к двухцепочечному разрыву в полинуклеотиде-мишени.

72. Нуклеопротеидный комплекс по п. 68 или п. 69, при этом полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, представляет собой двухцепочечную ДНК, белок Cas лишен способности разрезать двухцепочечную ДНК, и присутствие указанного нуклеопротеидного комплекса приводит к выключению гена полинуклеотида-мишени.

73. Нуклеопротеидный комплекс по любому из пунктов 68-72, при этом последовательность PAM представляет собой 5′-NNNNCNNA-3′,

5′-CNNNCNN-3′, 5′-NNNCCNN-3′, 5′-NNCNCNN-3′, 5′-NNNNCCN-3′ и/или 5′-NCNNCNN-3′.

74. Нуклеопротеидный комплекс по любому из пунктов 68-73, при этом последовательность PAM представляет собой 5′-CCCCCCNA-3′ [SEQ ID NO: 10], предпочтительно, при этом последовательность PAM представляет собой 5′-CCCCCCAA-3′ [SEQ ID NO: 11].

75. Нуклеопротеидный комплекс по любому из пунктов 68-74, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

76. Нуклеопротеидный комплекс по любому из пунктов 68-75, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

77. Нуклеопротеидный комплекс по любому из пунктов 68-76, который находится в прокариотической клетке.

78. Нуклеопротеидный комплекс по любому из пунктов 68-76, который находится в эукариотической клетке.

79. Нуклеопротеидный комплекс по любому из пунктов 68-78, при этом молекула направляющей РНК содержит crРНК и tracrРНК.

80. Нуклеопротеидный комплекс по любому из пунктов 68-79, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

81. Нуклеопротеидный комплекс по любому из пунктов 68-80, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

82. Нуклеопротеидный комплекс по любому из пунктов 68-81, при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

83. Нуклеопротеидный комплекс по любому из пунктов 68-82, при этом белок Cas предоставлен в виде части белкового комплекса, содержащего по меньшей мере один дополнительный функциональный или не функциональный белок, необязательно, при этом по меньшей мере один дополнительный белок также содержит по меньшей мере один функциональный фрагмент.

84. Нуклеопротеидный комплекс по п. 82 или 83, при этом белок Cas или дополнительный белок содержит по меньшей мере один

функциональный фрагмент, слитый или связанный с N-концом и/или С-концом белка или белкового комплекса Cas; предпочтительно, С-концом.

85. Нуклеопротеидный комплекс по любому из пунктов 82-84, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки, например, зеленый флуоресцентный белок (GFP).

86. Нуклеопротеидный комплекс по п. 85, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

87. Нуклеопротеидный комплекс по любому из пунктов 82-86, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

88. Нуклеопротеидный комплекс по любому из пунктов 82-86, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

89. Нуклеопротеидный комплекс по любому из пунктов 82-87, при этом нуклеиновая кислота представляет собой дцДНК, по меньшей мере один функциональный фрагмент представляет собой нуклеазу или геликазу-нуклеазу, и полинуклеотид-мишень имеет одноцепочечный или двухцепочечный разрыв в желаемом локусе.

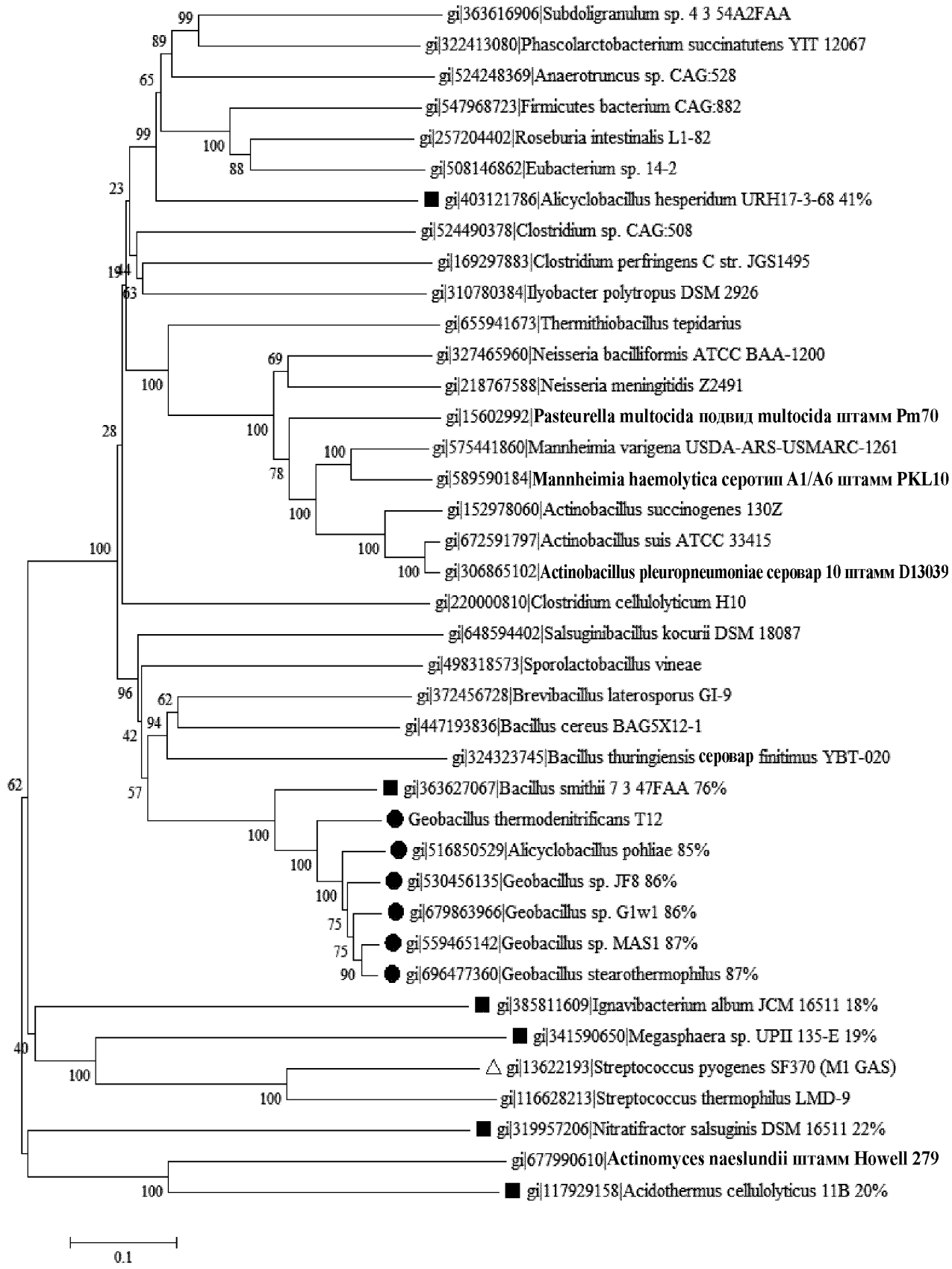
90. Нуклеопротеидный комплекс по любому из пунктов 82-86, при этом нуклеиновая кислота представляет собой дцДНК, функциональный фрагмент выбирают из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции, и образование нуклеопротеидного комплекса приводит к модификации экспрессии гена.

91. Нуклеопротеидный комплекс по любому из пунктов 68-90,

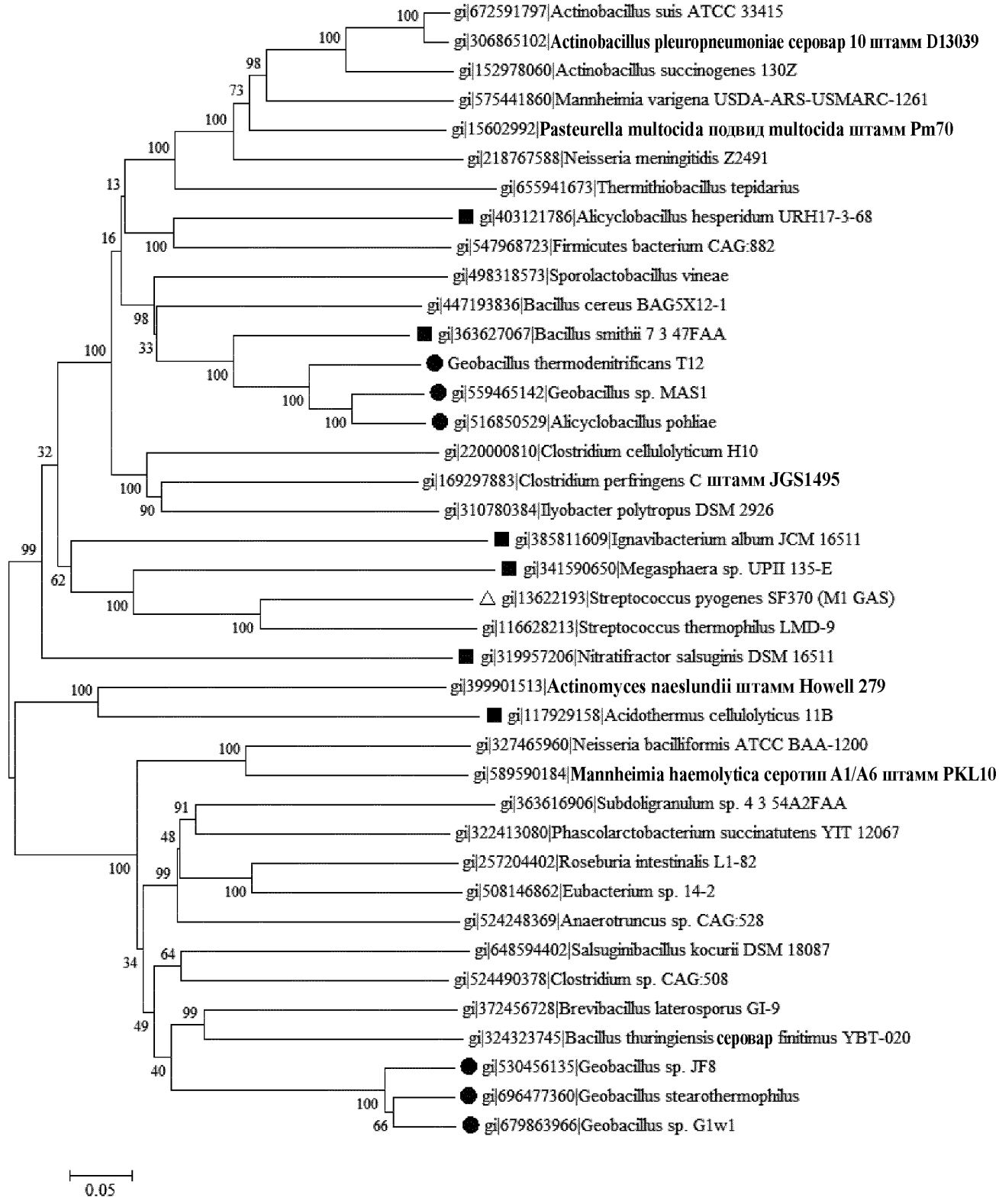
при этом образование нуклеопротеидного комплекса приводит к модификации или удалению и/или вставке желаемой нуклеотидной последовательности в желаемом локусе, и/или при этом образование нуклеопротеидного комплекса приводит к выключению экспрессии гена в желаемом локусе.

По доверенности

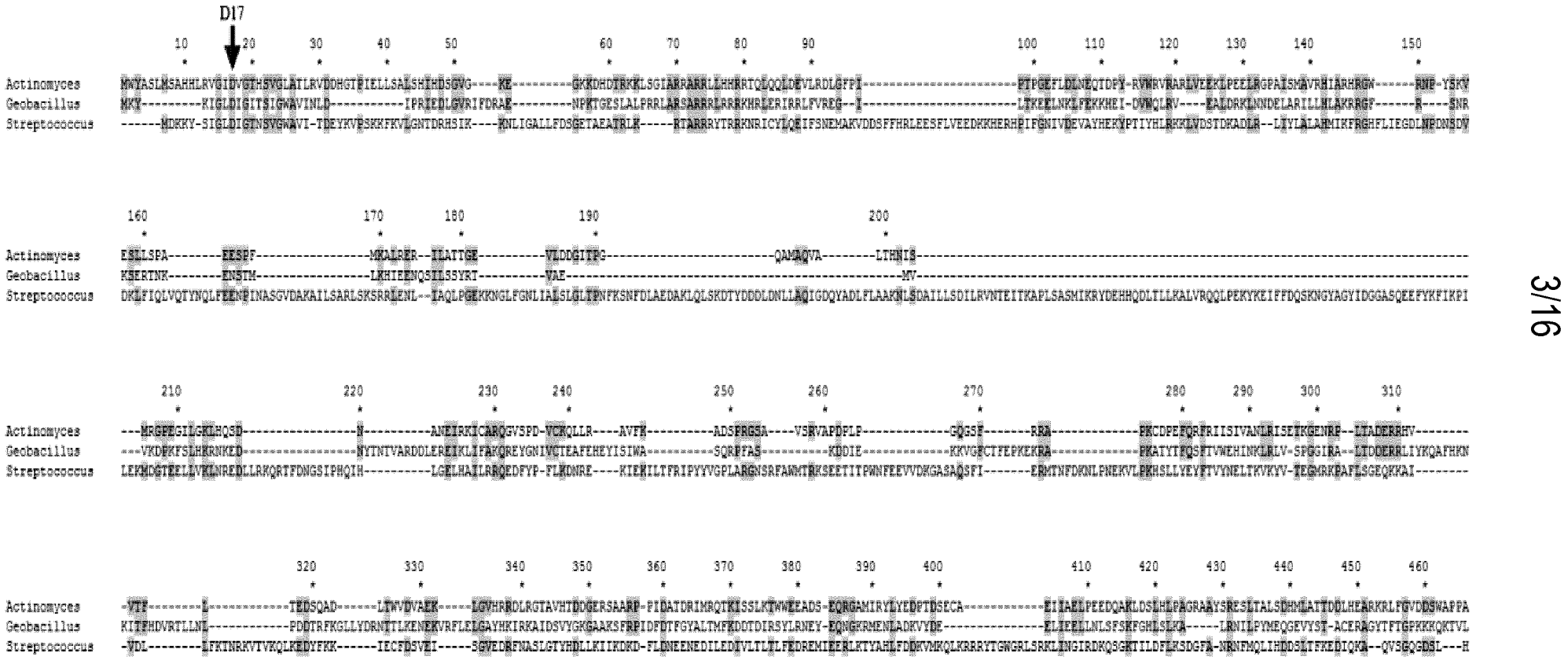
ФИГ.1


gi|363616906|Subdoligranulum sp. 4 3 54A2FAA
gi|322413080|Phascolarctobacterium succinatutens YIT 12067
gi|524248369|Anaerotruncus sp. CAG:528
gi|547968723|Firmicutes bacterium CAG:882
gi|257204402|Roseburia intestinalis L1-82
gi|508146862|Eubacterium sp. 14-2
gi|403121786|Alicyclobacillus hesperidum URH17-3-68 41%
gi|524490378|Clostridium sp. CAG:508
gi|169297883|Clostridium perfringens C str. JGS1495
gi|310780384|Ilyobacter polytropus DSM 2926
gi|655941673|Thermithiobacillus tepidarius
gi|327465960|Neisseria bacilliformis ATCC BAA-1200
gi|218767588|Neisseria meningitidis Z2491
gi|15602992|Pasteurella multocida подвид multocida штамм Pm70
gi|575441860|Mannheimia varigena USDA-ARS-USMARC-1261
gi|589590184|Mannheimia haemolytica серотип A1/A6 штамм PKL10
gi|152978060|Actinobacillus succinogenes 130Z
gi|672591797|Actinobacillus suis ATCC 33415
gi|306865102|Actinobacillus pleuropneumoniae серовар 10 штамм D13039
gi|220000810|Clostridium cellulolyticum H10
gi|648594402|Salsuginibacillus kocurii DSM 18087
gi|498318573|Sporolactobacillus vineae
gi|372456728|Brevibacillus laterosporus GI-9
gi|447193836|Bacillus cereus BAG5X12-1
gi|324323745|Bacillus thuringiensis серовар finitimus YBT-020
gi|363627067|Bacillus smithii 7 3 47FAA 76%
Geobacillus thermodenitrificans T12
gi|516850529|Alicyclobacillus pohliae 85%
gi|530456135|Geobacillus sp. JF8 86%
gi|679863966|Geobacillus sp. G1w1 86%
gi|559465142|Geobacillus sp. MAS1 87%
gi|696477360|Geobacillus stearothermophilus 87%
gi|385811609|Ignavibacterium album JCM 16511 18%
gi|341590650|Megasphaera sp. UPII 135-E 19%
gi|13622193|Streptococcus pyogenes SF370 (M1 GAS)
gi|116628213|Streptococcus thermophilus LMD-9
gi|319957206|Nitratifractor salsuginis DSM 16511 22%
gi|677990610|Actinomyces naeslundii штамм Howell 279
gi|117929158|Acidothermus cellulolyticus 11B 20%
0.1

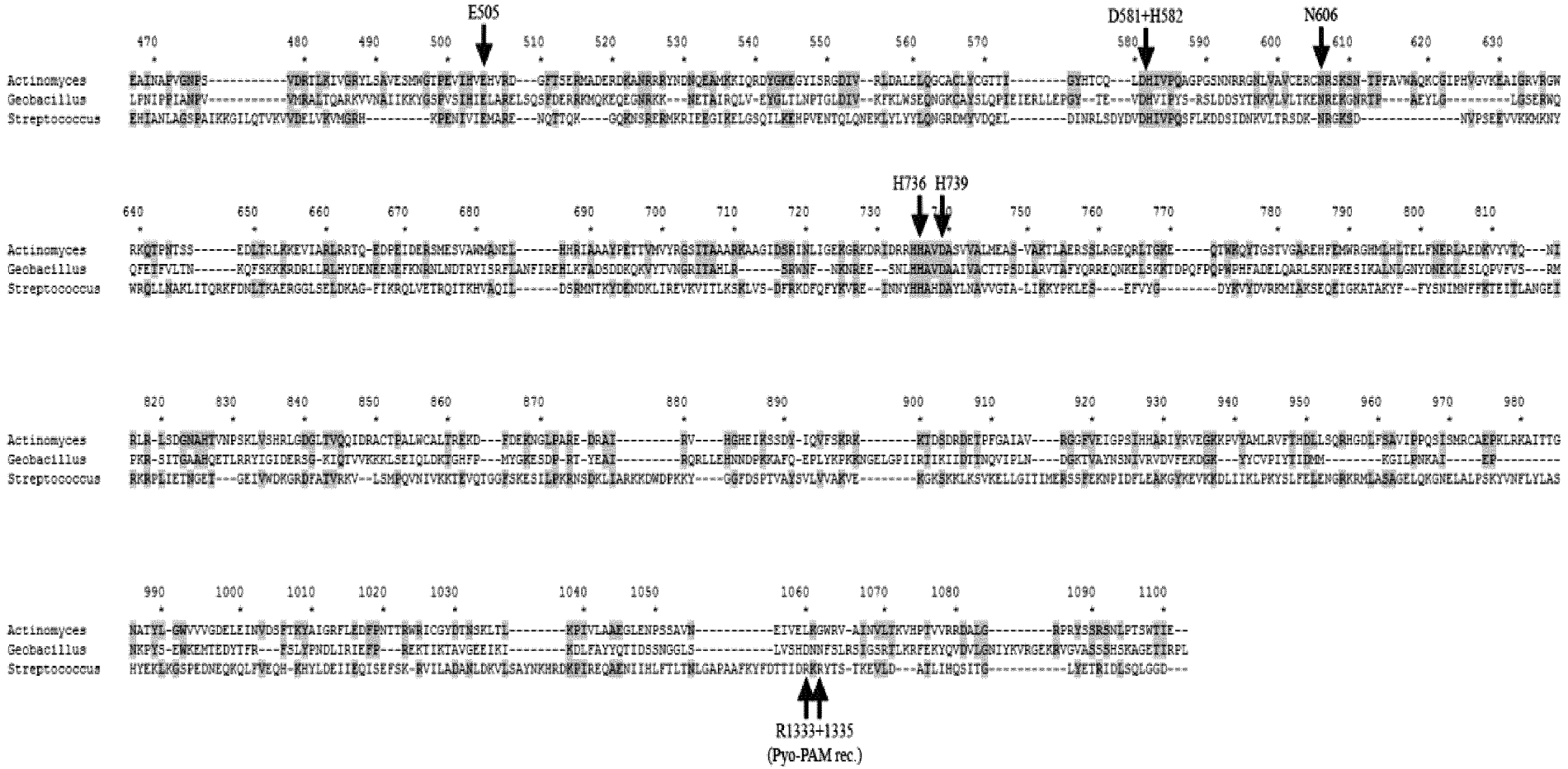
ФИГ.2


gi|672591797|Actinobacillus suis ATCC 33415
gi|306865102|Actinobacillus pleuropneumoniae серовар 10 штамм D13039
gi|152978060|Actinobacillus succinogenes 130Z
gi|575441860|Mannheimia varigena USDA-ARS-USMARC-1261
gi|15602992|Pasteurella multocida подвид multocida штамм Pm70
gi|218767588|Neisseria meningitidis Z2491
gi|655941673|Thermithiobacillus tepidarius
gi|403121786|Alicyclobacillus hesperidum URH17-3-68
gi|547968723|Firmicutes bacterium CAG:882
gi|498318573|Sporolactobacillus vineae
gi|447193836|Bacillus cereus BAG5X12-1
gi|363627067|Bacillus smithii 7 3 47FAA
Geobacillus thermodenitrificans T12
gi|559465142|Geobacillus sp. MAS1
gi|516850529|Alicyclobacillus pohliae
gi|220000810|Clostridium cellulolyticum H10
gi|169297883|Clostridium perfringens C штамм JGS1495
gi|310780384|Ilyobacter polytropus DSM 2926
gi|385811609|Ignavibacterium album JCM 16511
gi|341590650|Megasphaera sp. UPII 135-E
gi|13622193|Streptococcus pyogenes SF370 (M1 GAS)
gi|116628213|Streptococcus thermophilus LMD-9
gi|319957206|Nitratifractor salsuginis DSM 16511
gi|399901513|Actinomyces naeslundii штамм Howell 279
gi|117929158|Acidothermus cellulolyticus 11B
gi|327465960|Neisseria bacilliformis ATCC BAA-1200
gi|589590184|Mannheimia haemolytica серотип A1/A6 штамм PKL10
gi|363616906|Subdoligranulum sp. 4 3 54A2FAA
gi|322413080|Phascolarctobacterium succinatutens YIT 12067
gi|257204402|Roseburia intestinalis L1-82
gi|508146862|Eubacterium sp. 14-2
gi|524248369|Anaerotruncus sp. CAG:528
gi|648594402|Salsuginibacillus kocurii DSM 18087
gi|524490378|Clostridium sp. CAG:508
gi|372456728|Brevibacillus laterosporus GI-9
gi|324323745|Bacillus thuringiensis серовар finitimus YBT-020
gi|530456135|Geobacillus sp. JF8
gi|696477360|Geobacillus stearothermophilus
gi|679863966|Geobacillus sp. G1w1
0.05

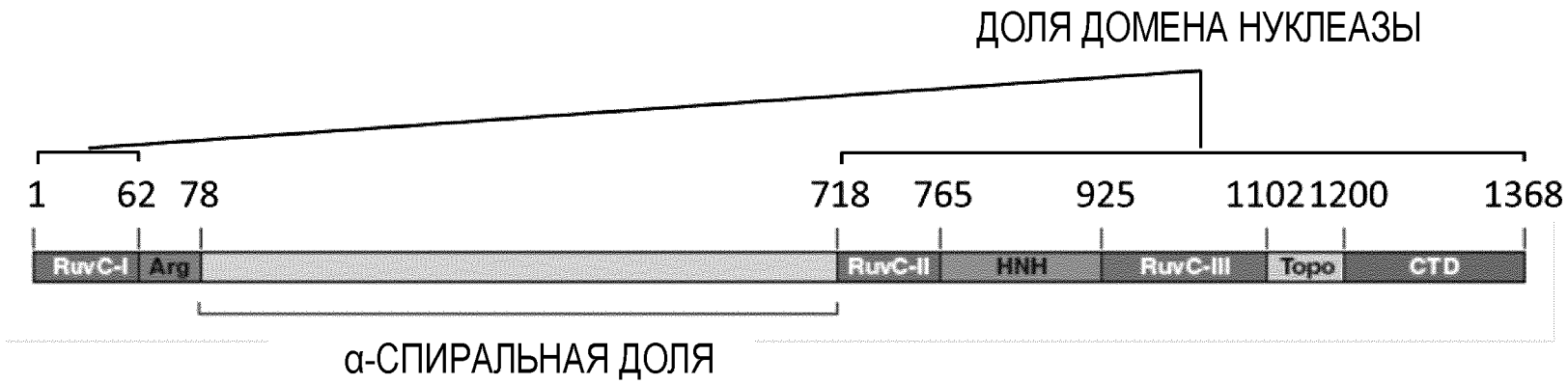
ФИГ.3



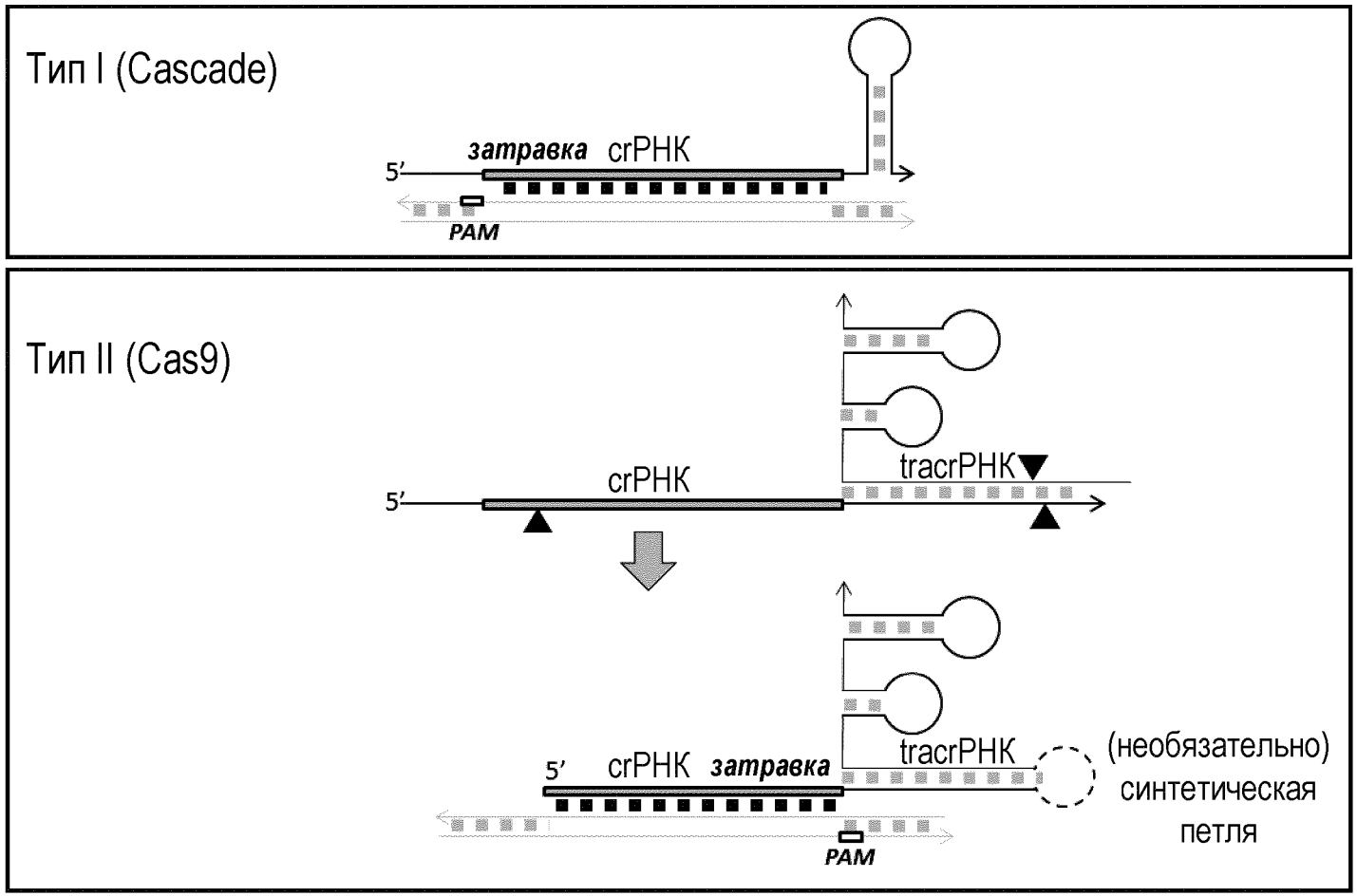
ФИГ.3 (продолжение)



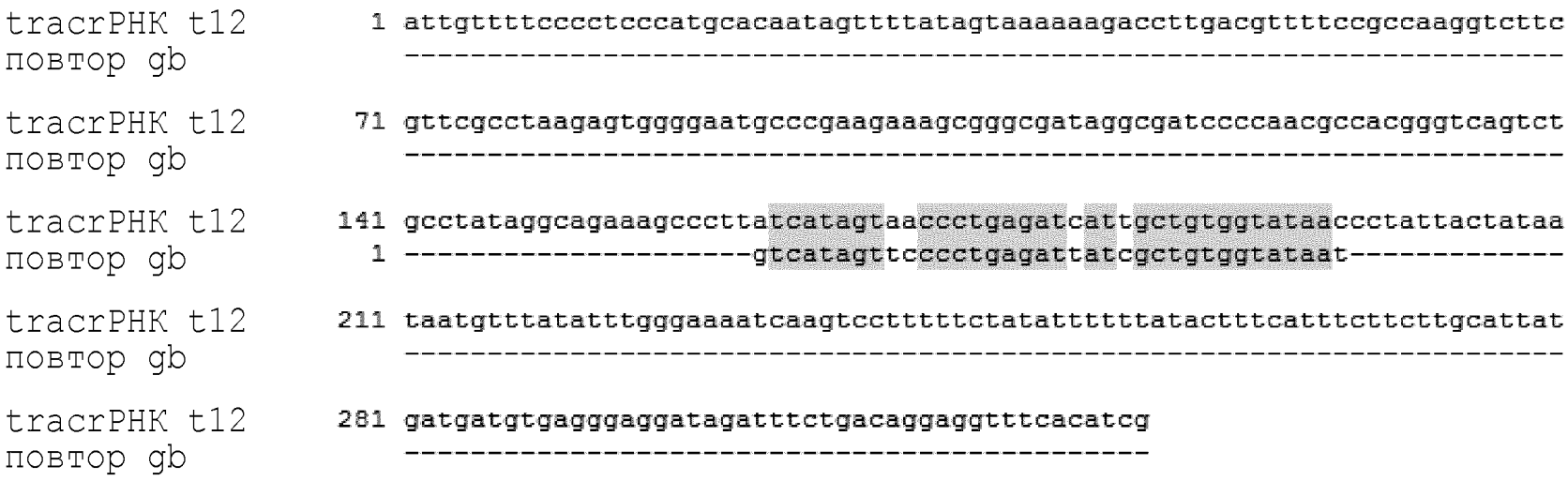
ФИГ.4


ДОЛЯ ДОМЕНА НУКЛЕАЗЫ
1
62
78
718
765
925
1102
1200
1368
RuvC-I
Arg
RuvC-II
HNH
RuvC-III
Topo
CTD
α-СПИРАЛЬНАЯ ДОЛЯ

ФИГ.5

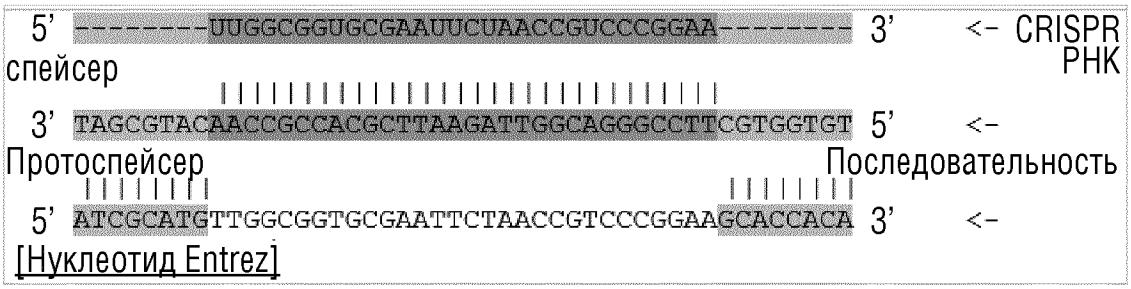

Тип I (Cascade)
затравка
crРНК
5′
PAM
Тип II (Cas9)
crРНК
tracrРНК
5′
5′
crРНК
затравка
tracrРНК
(необязательно)
синтетическая
петля
PAM

ФИГ.6


tracrРНК t12 1 attgttttccccctcccatgcacaatagttttatagtaaaaaagaccttgacgttttccgccaaggtcttc
повтор gb ----------
tracrРНК t12 71 gttcgcctaagagtggggaatgcccgaagaaagcgggcgataggcgatccccaacgccacgggtcagtct
повтор gb ----------
tracrРНК t12 141 gcctataggcagaaagcccttatcatagtaaccctgagatcattgctgtggtataaccctattactataa
повтор gb 1 ----------gtcatagttcccctgagattatcgctgtggtataat----------
tracrРНК t12 211 taatgtttatatttgggaaaatcaagtccttttttctatatttttttatactttcatttcttcttgcattat
повтор gb ----------
tracrРНК t12 281 gatgatgtgagggaggatagatttctgacaggaggtttcacatcg
повтор gb ----------

ФИГ.7

1. Сопоставление с: **DQ453159 DQ453159 Geobacillus вирус E2, полный геном** (DQ453159) положение: 39980-39951, с: **спейсер4** CRISPR №4 Спейсер №1 (спейсер4_4_1) положение: 1-30, Цепь: +



Баллов: 30

2. Сопоставление с: **Bacillus alveayuensis штамм 24KAM51 LG50_053, последовательность всего генома, секвенированная методом «дробовика»** (NZ_JYCE01000053) положение: 110918-110889, с: **спейсер9** CRISPR №9 Спейсер №1 (спейсер9_9_1) положение: 1-30, Цепь: +

Баллов: 30

3. Сопоставление с: **Anoxybacillus flavithermus WK1, полный геном** (CP000922) положение: 701422-701450, с: **спейсер5** CRISPR №5 Спейсер №1 (спейсер5_5_1) положение: 1-29, Цепь: -

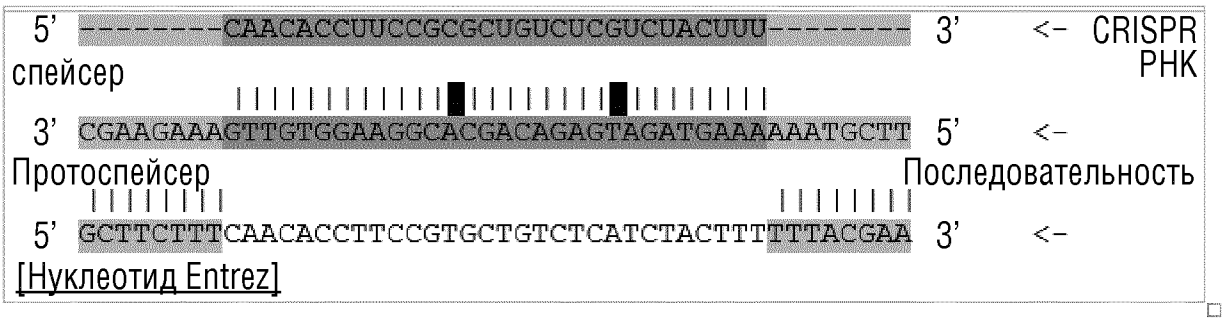
Баллов: 29

4. Сопоставление с: **Geobacillus kaustophilus штамм Et23 LG51_086, последовательность всего генома, секвенированная методом «дробовика»** (NZ_JYCF01000086) положение: 15565-15537, с: **спейсер5** CRISPR №5 Спейсер №1 (спейсер5_5_1) положение: 1-29, Цепь: +

ФИГ.7 (продолжение)

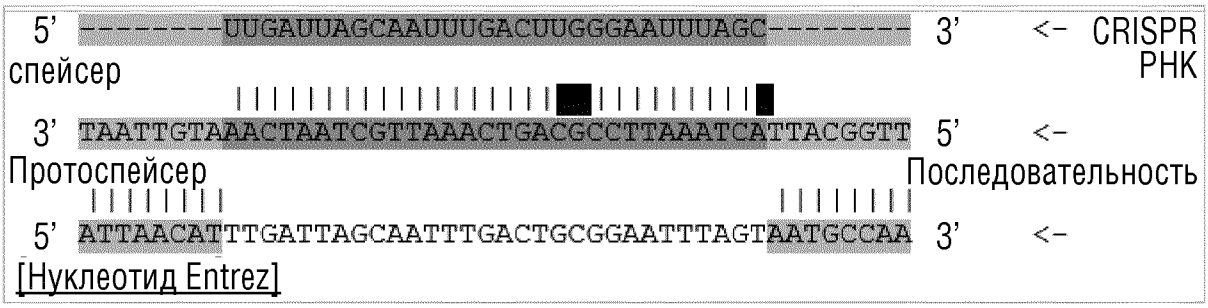
Баллов: 29

5. Сопоставление с: **DQ453159 DQ453159 Geobacillus вирус E2, полный геном** (DQ453159) положение: 6492-6463, с: **спейсер6** CRISPR №6 Спейсер №1 (спейсер6_6_1) положение: 1-30, Цепь: +

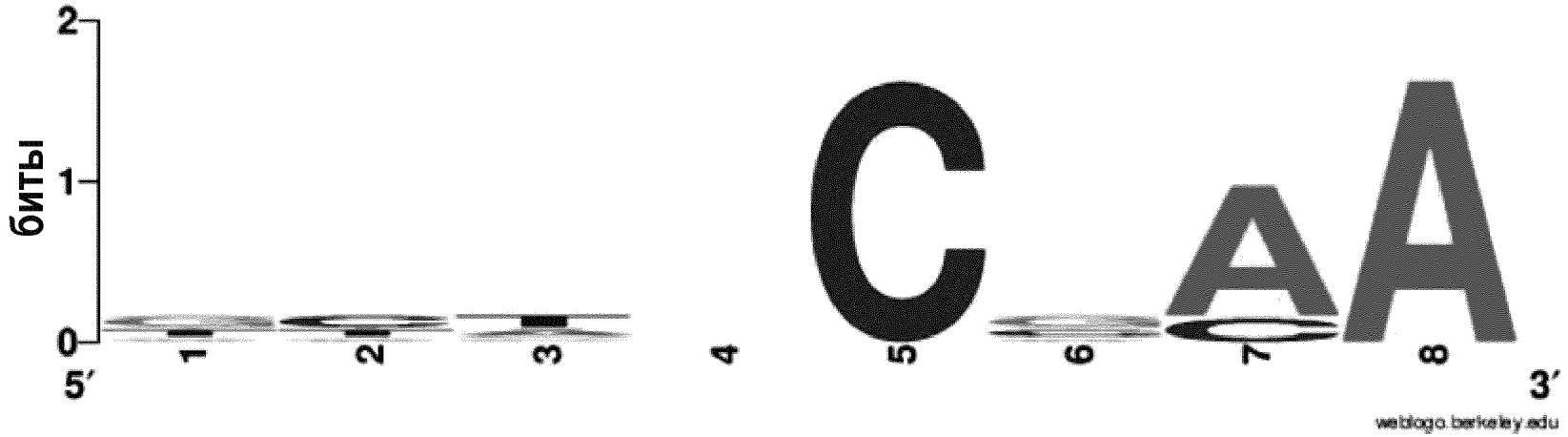


Баллов: 26

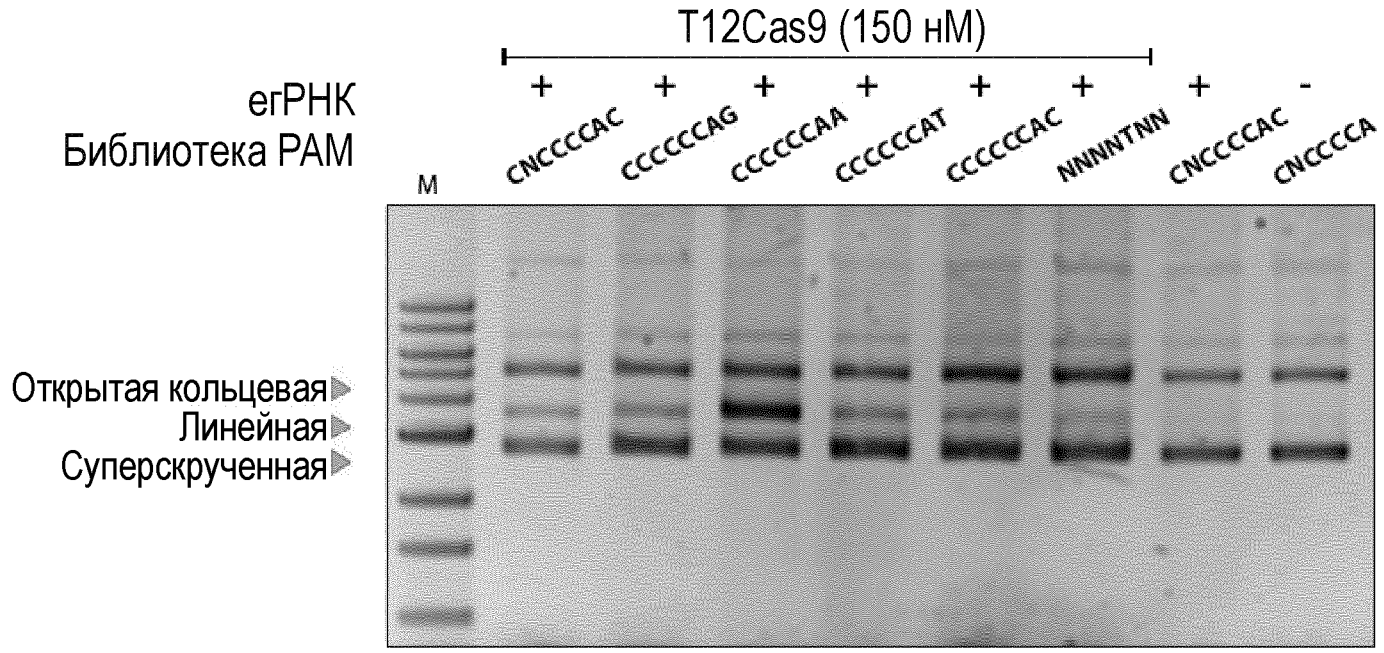
6. Сопоставление с: **Pasteurella bettyae CCUG 2042 contig00003, последовательность всего генома, секвенированная методом «дробовика»** (NZ_AJSX01000041) положение: 137780-137808, с: **спейсер7** CRISPR №7 Спейсер №1 (спейсер7_7_1) положение: 1-30, Цепь: -



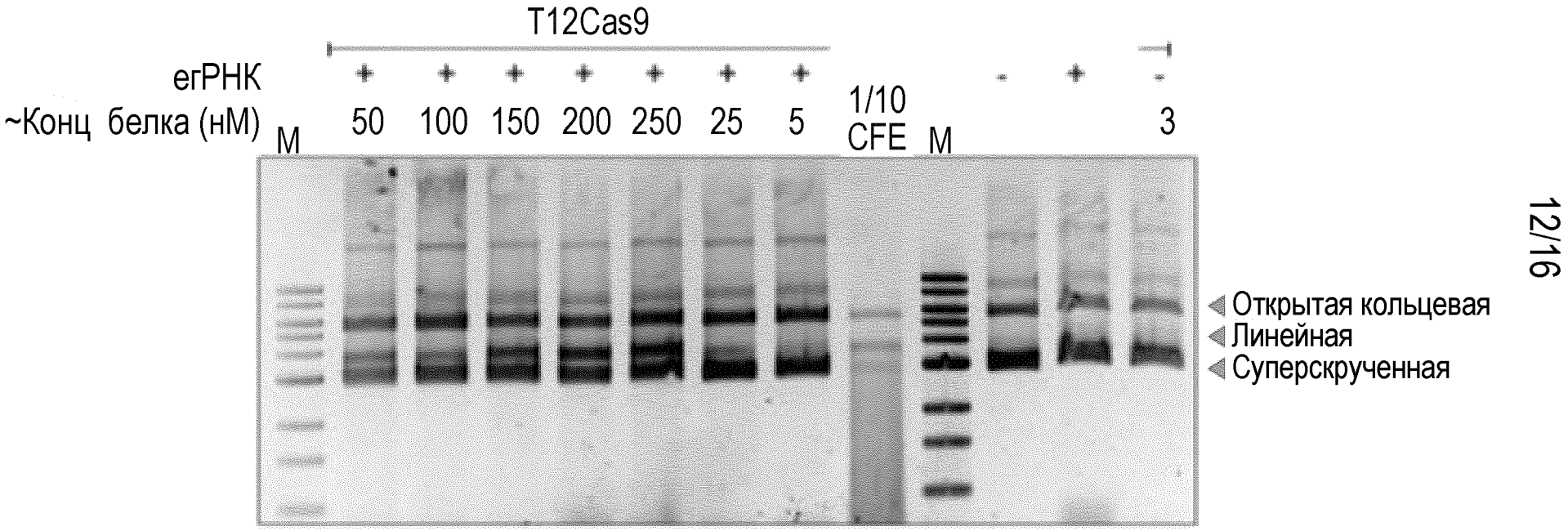
ФИГ.8



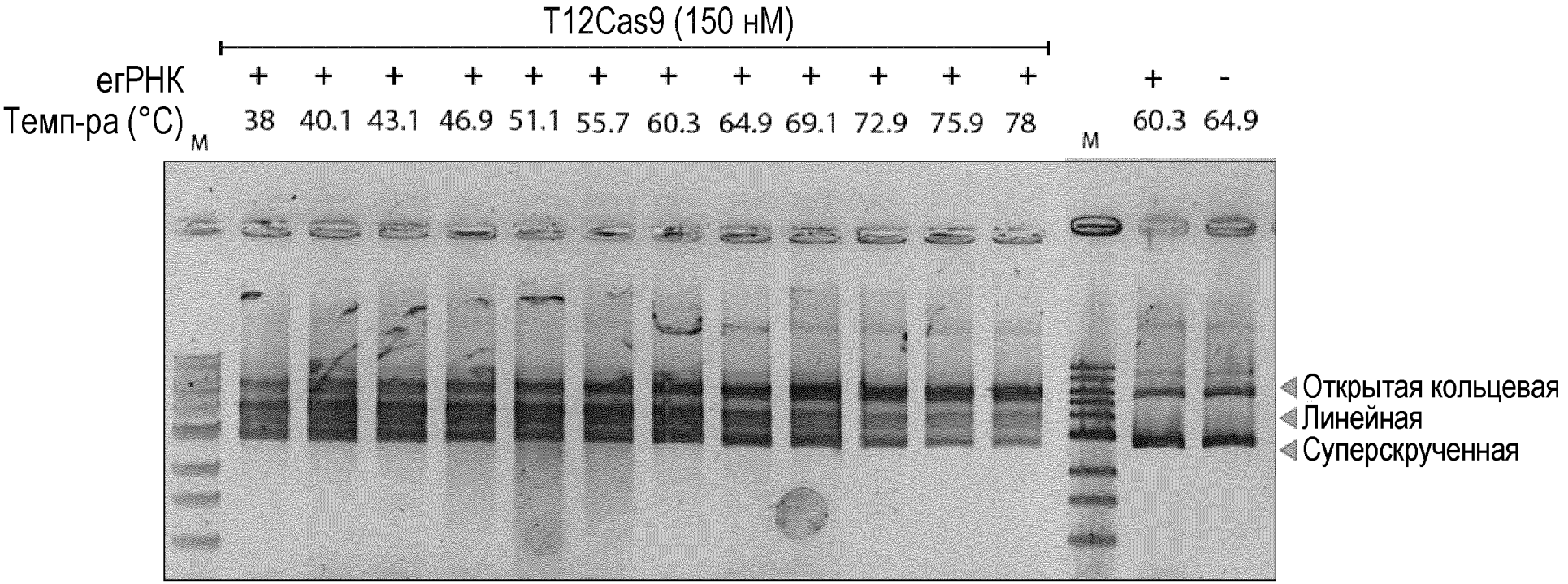
ФИГ.9


T12Cas9 (150 нМ)
+ + + + + + + -
егРНК
Библиотека PAM
CNCCCCAC
CCCCCCAG
CCCCCCAA
CCCCCCAT
CCCCCCAC
NNNNTNN
CNCCCCAC
CNCCCCA
M
Открытая кольцевая
Линейная
Суперскрученная

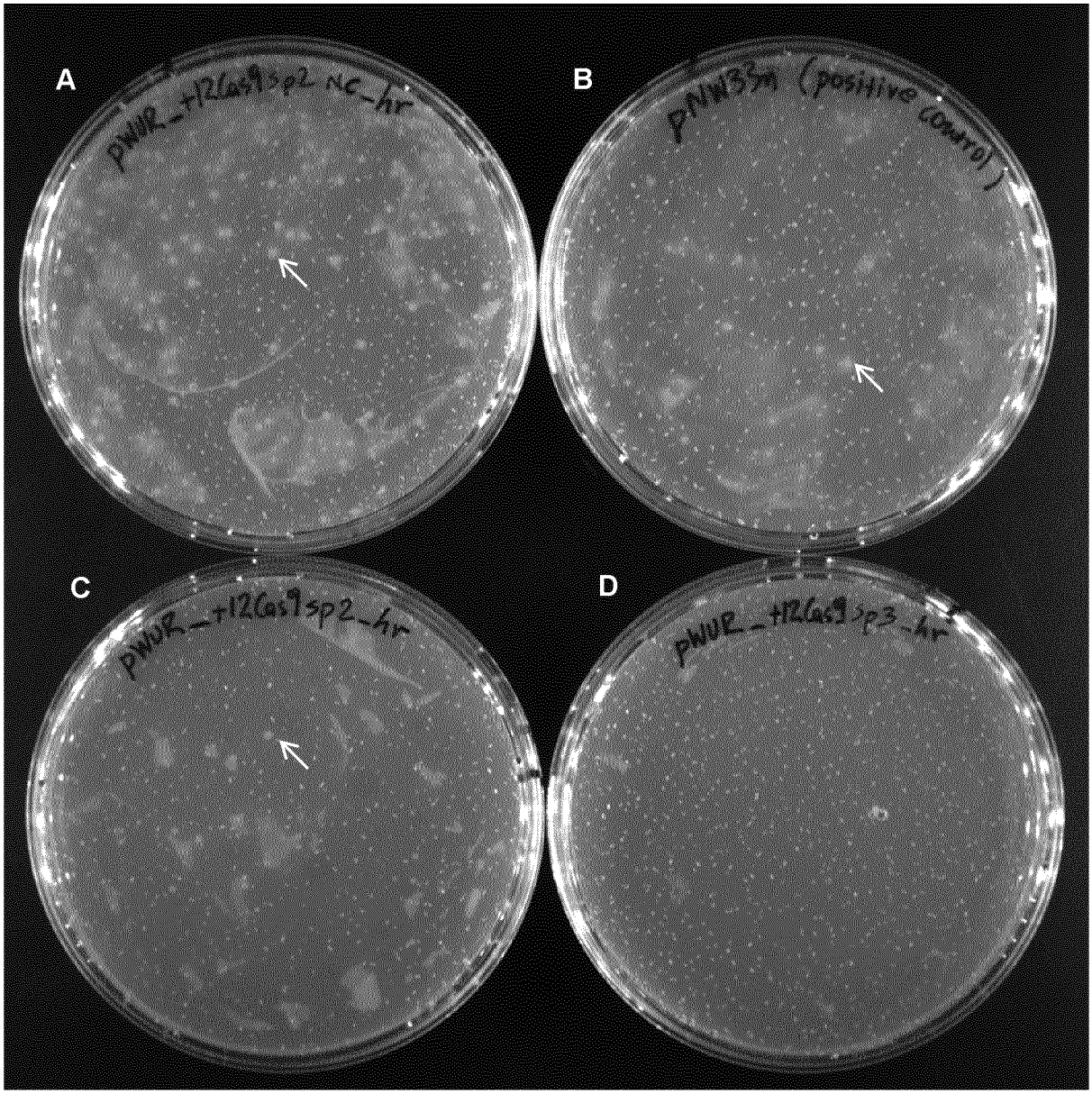
ФИГ.10


T12Cas9
егРНК
~Конц белка (нМ)
M
50
100
150
200
250
25
5
1/10
CFE
M
3
Открытая кольцевая
Линейная
Суперскрученная

ФИГ.11

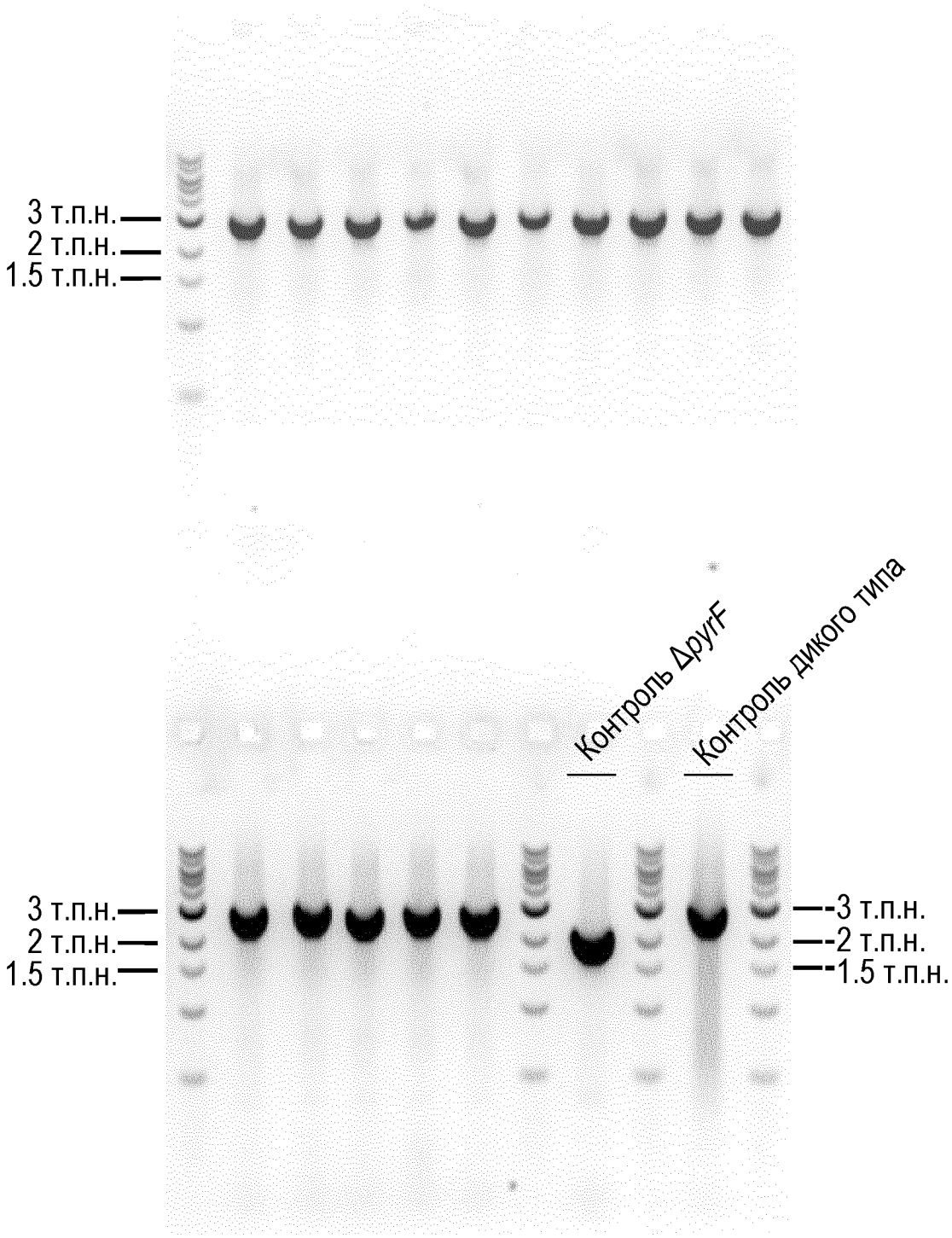

T12Cas9 (150 нМ)
егРНК + + + + + + + + + + + + + -
Темп-ра (°C) м 38 40.1 43.1 46.9 51.1 55.7 60.3 64.9 69.1 72.9 75.9 78 м 60.3 64.9
Открытая кольцевая
Линейная
Суперскрученная

ФИГ.12

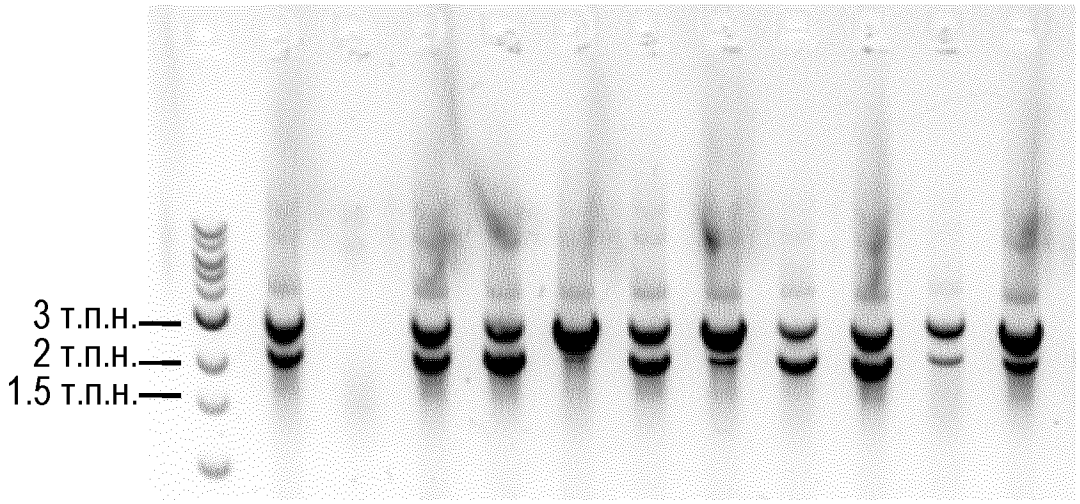


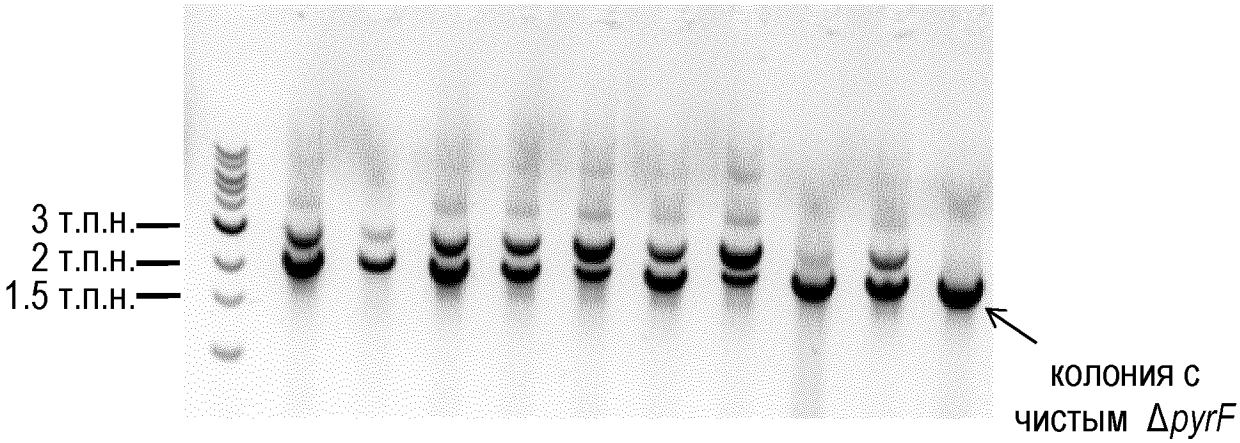
Колонии указаны стрелками

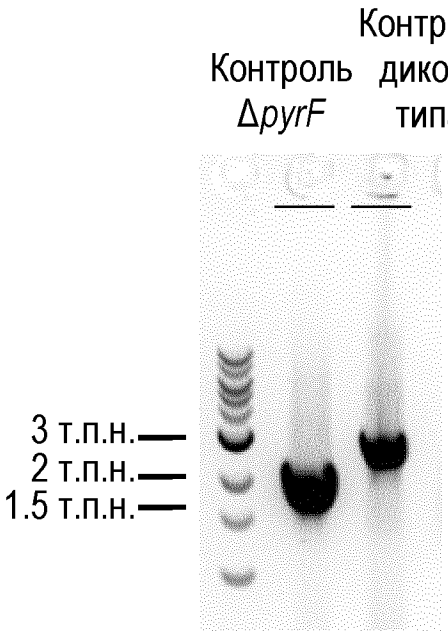
ФИГ.13


3 т.п.н.
2 т.п.н.
1.5 т.п.н.
Контроль ΔpyrF
Контроль дикого типа
3 т.п.н.
2 т.п.н.
1.5 т.п.н.
3 т.п.н.
2 т.п.н.
1.5 т.п.н.

ФИГ.14


3 т.п.н.
2 т.п.н.
1.5 т.п.н.


3 т.п.н.
2 т.п.н.
1.5 т.п.н.
колония с чистым ΔpyrF


Контроль ΔpyrF
Контроль дикого типа
3 т.п.н.
2 т.п.н.
1.5 т.п.н.