



(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.01.13(51) Int. Cl. C12N 9/22 (2006.01)
C12N 15/113 (2010.01)(22) Дата подачи заявки
2017.08.16

(54) ТЕРМОСТАБИЛЬНЫЕ НУКЛЕАЗЫ Cas9

(31) PCT/EP2016/081077

(32) 2016.12.14

(33) EP

(86) PCT/EP2017/070797

(87) WO 2018/108339 2018.06.21

(71) Заявитель:

ВАГЕНИНГЕН УНИВЕРСИТЕЙТ;
СТИХТИНГ ВОР ДЕ ТЕХНИШЕ
ВЕТЕНСХАППЕН (NL)

(72) Изобретатель:

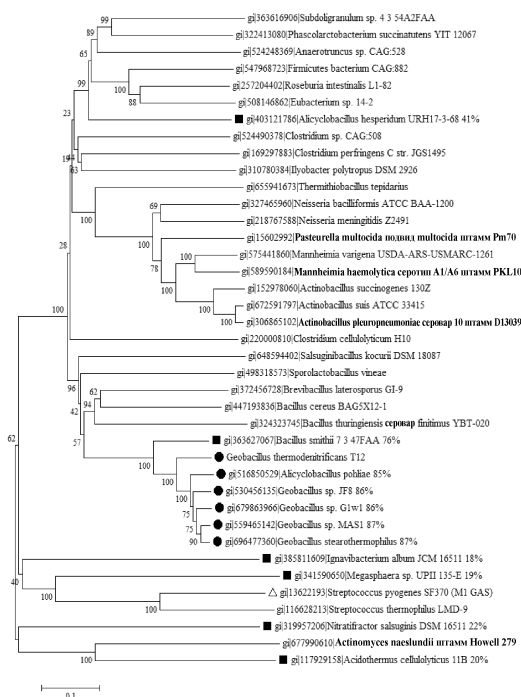
Ван Дер Ост Джон, Ван Краненбург
Рихард (NL), Босма Эллеке
Фенна (DK), Маугиакос Иоаннис,
Моханраджу Прартхана (NL)

(74) Представитель:

Медведев В.Н. (RU)

(57) 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 20 до 100°C включительно. Изобретение также относится к выделенным молекулам нуклеи-

новой кислоты, кодирующим указанные нуклеазы Cas9, экспрессионным векторам и клеткам-хозяевам. По изобретению также предложены последовательности РАМ, узнаваемые белком или полипептидом Cas. Нуклеазы Cas9, раскрытые в настоящем документе, являются новыми инструментами для генетической инженерии в целом и в частности при повышенных температурах.



ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-557549EA/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.

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 два класса систем CRISPR-Cas, каждый из которых включает три типа, в большинстве из которых используются малые молекулы РНК в качестве гид-РНК для направления на комплементарные последовательности ДНК (Makarova *et al.*, 2015, Nat Rev Microbiol 13: 722-736; Mahanraju *et al.*, 2016, Science 353: aad5147).

Недавно в лабораториях Doudna/Charpentier была проведена тщательная характеристика эффекторного фермента системы CRISPR-Cas класса 2/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 аутоиммунитета. 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 геномной 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: 6091-6105) 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 геномной инженерии при повышенных температурах и имеют особую ценность при генетических манипуляциях с термофильными организмами; в частности, микроорганизмами.

Недавно была проведена филогенетическая переоценка рода термофильных бактерий *Geobacillus*, которая привела к созданию нового рода, *Parageobacillus*. Вследствие этого, некоторые виды, ранее отнесенные к роду *Geobacillus*, в систематике были отнесены к роду *Parageobacillus* и соответствующим образом переименованы (Aliyu et al., (2016) *Systematic and Applied Microbiology* 39:527–533).

Кластер регулярно расположенных группами коротких палиндромных повторов (CRISPR) и связанные с CRISPR белки (Cas) обеспечивают адаптивный и наследуемый иммунитет у прокариот против инвазивных генетических элементов (Brouns et al. *Science* 321, (2008); Barrangou et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315, (2007); Wright et al. *Cell* 164, 29–44 (2016); Mohanraju et al. *Science* 353, aad5147 (2016)). Системы CRISPR-Cas подразделяют на два класса (1 и 2) и шесть типов (I–VI), в зависимости от их

сложности и характерного белка (Makarova *et al.* *Nat. Rev. Microbiol.* 13, 722-736 (2015)). Системы класса 2, включая CRISPR-Cas9 типа II и CRISPR-Cas12a типа V (прежнее название CRISPR-Cpf1), недавно были использованы в качестве инструментов редактирования генома как для эукариот (Komor *et al.* *Cell* 168, 20-36 (2017); Puchta, *Curr. Opin. Plant Biol.* 36, 1-8 (2017); Xu *et al.* *J. Genet. Genomics* 42, 141-149 (2015); Tang *et al.* *Nat. Plants* 3, 17018 (2017); Zetsche *et al.* *Nat. Biotechnol.* 35, 31-34 (2016)), так и для прокариот (Mougiakos, *et al.* *Trends Biotechnol.* 34, 575-587 (2016)). Эти системы входят в число самых простых известных систем CRISPR-Cas, поскольку они вносят направленные двухцепочечные разрывы (ДЦР) в ДНК за счет рибонуклеопротеидного (РНП) комплекса, образованного одной эндонуклеазой Cas и гид-РНК.

До настоящего времени Cas9 *Streptococcus pyogenes* (SpCas9) является наиболее хорошо охарактеризованной и наиболее широко используемой Cas9 для редактирования генома. Хотя были охарактеризованы несколько других систем типа II, ни одна из них не происходит из термофильного организма (Nakade, *et al.* *Bioengineered* 1-9 (2017). doi:10.1080/21655979.2017.1282018). Характеризация таких систем CRISPR-Cas была бы интересна для фундаментального понимания процессов, а также для разработки новых вариантов применения.

Хотя основные генетические инструменты доступны для ряда термофилов (Taylor *et al.* *Microb. Biotechnol.* 4, 438-448 (2011); Olson, *et al.* *Curr. Opin. Biotechnol.* 33, 130-141 (2015); Zeldes, *et al.* *Front. Microbiol.* 6, 1209 (2015)), эффективность этих инструментов все-еще остается слишком низкой для полного исследования и использования этой интересной группы микроорганизмов. С учетом полученных авторами изобретения данных, показывающих, что SpCas9 не активна *in vivo* при температуре $\geq 42^{\circ}\text{C}$, они ранее разработали на основе SpCas9 инструмент для генной инженерии факультативных термофилов, объединяющий гомологичную рекомбинацию при повышенных температурах и основанную на SpCas9 контрселекцию при умеренных

температурах (Mougiakos *et al.* *ACS Synth. Biol.* 6, 849–861 (2017)). Однако основанный на Cas9 инструмент редактирования и выключения генов для облигатных термофилов все еще отсутствует, поскольку SpCas9 не активна при температуре 42°C или выше (Mougiakos *et al.* *ACS Synth. Biol.* 6, 849–861 (2017)), и до настоящего времени ни одна термофильная Cas9 не была охарактеризована.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

Авторы изобретения обнаружили и охарактеризовали ThermoCas9: направляемую РНК ДНК-эндонуклеазу из системы CRISPR-Cas типа IIC термофильной бактерии *Geobacillus thermodenitrificans* T12. Авторы изобретения показали, что она, неожиданно, обладает активностью *in vitro* в широком диапазоне температур, продемонстрировали важность структуры еРНК для термостабильности и применили ThermoCas9 для *in vivo* редактирования генома в широком диапазоне температур.

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

- а. аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]; и/или
- б. 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
- с. аминокислотный мотив X_5LXX_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
- д. 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
- е. 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.

Во избежание сомнений, белки, полипептиды или нуклеиновые кислоты, кодирующие белки, Cas по настоящему изобретению могут быть также названы «GtCas9» или «ThermoCas9». Термины «GtCas9» и «ThermoCas9» в тексте спецификации использованы взаимозаменяемо и имеют одно и то же значение.

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_5LXX_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:

EKDGKYUC [SEQ ID NO: 2].

EKDGKYUC [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.

EKDGKYUC [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_5LXX_6IE [SEQ ID NO: 4], где X_5 независимо выбирают из метионина или фенилаланина и X_6 независимо выбирают из гистидина или аспарагина.

EKDGKYUC [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_5LXX_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.

EKDGKYUC [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_5LXX_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.

EKDGKYUC [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_5LXX_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.

EKDGKYUC [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.

EKDGKYUC [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.

EKDGKYUC [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.

EKDGKYUC [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.

EKDGKYUC [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.

EKDGKYUC [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.

EKDGKYUC [SEQ ID NO: 2] и X₅LKX₆IE [SEQ ID NO: 4], где X₅ независимо выбирают из метионина или фенилаланина и X₆ независимо выбирают из гистидина или аспарагина; и X₉FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], где X₉ представляет собой аланин или глутамат, X₁₀ представляет собой глутамин или лизин, X₁₁ представляет собой аргинин или аланин, X₁₂ представляет собой аспарагин или аланин и X₁₃ представляет собой лизин или серин.

EKDGKYUC [SEQ ID NO: 2] и X₇VYSX₈K [SEQ ID NO: 5], где X₇ представляет собой глутамат или изолейцин и X₈ представляет собой одно из триптофана, серина или лизина; и X₉FYX₁₀X₁₁REQX₁₂KEX₁₃ [SEQ ID NO: 6], где X₉ представляет собой аланин или глутамат, X₁₀ представляет собой глутамин или лизин, X₁₁ представляет собой аргинин или аланин, X₁₂ представляет собой аспарагин или аланин и X₁₃ представляет собой лизин или серин.

X₁X₂CTX₃X₄ [SEQ ID NO: 3], где X₁ независимо выбирают из изолейцина, метионина или пролина, X₂ независимо выбирают из валина, серина, аспарагина или изолейцина, X₃ независимо выбирают из глутамата или лизина, и X₄ представляет собой одно из аланина, глутамата или аргинина; и X₅LKX₆IE [SEQ ID NO: 4], где X₅ независимо выбирают из метионина или фенилаланина и X₆ независимо выбирают из гистидина или аспарагина.

X₁X₂CTX₃X₄ [SEQ ID NO: 3], где X₁ независимо выбирают из изолейцина, метионина или пролина, X₂ независимо выбирают из валина, серина, аспарагина или изолейцина, X₃ независимо выбирают из глутамата или лизина, и X₄ представляет собой одно из аланина, глутамата или аргинина; и X₅LKX₆IE [SEQ ID NO: 4], где X₅ независимо выбирают из метионина или фенилаланина и X₆ независимо выбирают из гистидина или аспарагина; и X₇VYSX₈K [SEQ ID NO: 5], где X₇ представляет собой глутамат или изолейцин и X₈ представляет собой одно из триптофана, серина или лизина.

X₁X₂CTX₃X₄ [SEQ ID NO: 3], где X₁ независимо выбирают из изолейцина, метионина или пролина, X₂ независимо выбирают из валина, серина, аспарагина или изолейцина, X₃ независимо выбирают из глутамата или лизина, и X₄ представляет собой одно из аланина, глутамата или аргинина; и X₅LKX₆IE [SEQ ID NO: 4], где X₅

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_5LXX_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%, при этом белок Cas или его полипептидный фрагмент содержит любые из следующих мотивов или аминокислот, отдельно или в сочетании:

Домен RuvC-I, содержащий IGLDIGITSIG [SEQ ID NO: 23], предпочтительно, IGLDIGITSIGWAVINLD [SEQ ID NO: 24].

Мостиковый домен, содержащий RSARR [SEQ ID NO: 25],

предпочтительно, PRRLARSARRRLRRRKHRLERIRRL [SEQ ID NO: 26];
и/или

Домен α -спиральной/узнающей доли, содержащий WQLR [SEQ ID NO: 27]; и/или

Домен α -спиральной/узнающей доли, содержащий HLAKRRG [SEQ ID NO: 28], предпочтительно, LARILLHLAKRRG [SEQ ID NO: 29];
и/или

Домен α -спиральной/узнающей доли, содержащий IFAKQ [SEQ ID NO: 30], предпочтительно, EIKLIFAKQ [SEQ ID NO: 31]; и/или

Домен α -спиральной/узнающей доли, содержащий IWASQR [SEQ ID NO: 32]; и/или KVGFCSTFEPKEKRAPK [SEQ ID NO: 33]; и/или FTVWENINKLRL [SEQ ID NO: 34]; и/или

Домен RuvC-II, содержащий мотив IANPVVMRALTQ [SEQ ID NO: 35], предпочтительно, IANPVVMRALTQARKVVNAIIKKYG [SEQ ID NO: 36];
и/или

Домен RuvC-II, содержащий мотив ELAR [SEQ ID NO: 37], предпочтительно, IHIELARE [SEQ ID NO: 38]; и/или

Домен HNH, содержащий мотив QNGKCA Y [SEQ ID NO: 39], предпочтительно, IVKFKLWSEQNGKCA Y [SEQ ID NO: 40]; и/или

Домен HNH, содержащий мотив VDHVIP [SEQ ID NO: 41], предпочтительно, VDHVIPYSRSLDDSYTNKVL [SEQ ID NO: 42]; и/или

Домен RuvC-III, содержащий мотив DTRYISRFLAN [SEQ ID NO: 43]; и/или

Домен RuvC-III, содержащий мотив VYTVNGRITAHLSRW [SEQ ID NO: 44]; и/или

Домен RuvC-III, содержащий мотив HNAVDA [SEQ ID NO: 45], предпочтительно, HNAVDAAIVA [SEQ ID NO: 46].

Предпочтительно, 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%, при этом белок Cas или его полипептидный фрагмент содержит каждый из аминокислотных мотивов [SEQ ID NO: 23] - [SEQ ID NO: 46] в сочетании.

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 расщеплению, связыванию, маркированию или модификации при температуре в диапазоне от 20°C до 70°C, например, 25°C, 55°C, 60°C или 65°C. Предпочтительно, белок или полипептид Cas способен к указанному расщеплению, связыванию, маркированию или модификации при температуре в диапазоне от 50°C и 70°C, например, 55°C или 60°C. Предпочтительно, белок или полипептид Cas способен к указанному расщеплению, связыванию, 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'-NNNNNNA-3', и с 5'-конца последовательность PAM начинается 5'-NNNNNNA-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'-NNCNC-3', 5'-NNNCN-3' и/или 5'-NNNNNC-3'. Необязательно, последовательность PAM содержит, так что с 5'-конца последовательность PAM начинается, 5'-CCCCCNA-3' [SEQ ID NO: 10] и, кроме того, предпочтительно, последовательность PAM содержит, так что с 5'-конца последовательность PAM начинается, 5'-CCCCCAA-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 РАМ, непосредственно примыкающую к 3'-концу протоспейсерной последовательности на не являющейся мишенью цепи, как описано в настоящем документе. Таким образом, последовательность РАМ 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 РАМ 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 РАМ 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 РАМ начинается 5'-NNNNCNNA-3' [SEQ ID NO: 47], 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 человеческих клетках. Необязательно, варианты применения и способы по изобретению можно осуществлять, и нуклеопротеиды по изобретению можно создавать и использовать, *in vivo*, например, в человеческих клетках, которые не являются эмбриональными стволовыми клетками. Необязательно, варианты применения и способы по изобретению можно осуществлять, и нуклеопротеиды по изобретению можно создавать и использовать, *in vivo*, например, в человеческих клетках в способах, которые не включают модификацию генетических особенностей зародышевой линии человека. 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' [SEQ ID NO: 47], 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'-CCCCCNNA-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 человеческой 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. Необязательно, трансформированная человеческая клетка представляет собой выделенную человеческую клетку.

Необязательно, человеческая клетка не является эмбриональной стволовой клеткой. 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 человеческой 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. Необязательно, человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, способ не включает модификацию генетических особенностей зародышевой линии человека.

Изобретение также относится к наборам, включающим один или более реагентов для осуществления вариантов применения и способов по изобретению, или для получения трансформированных человеческих 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 человеческой клетки. Предполагается, что систему Cas можно использовать для модификации геномов в клетках различных типов и/или у разных организмов. Необязательно, белки и системы Cas по изобретению могут быть использованы для модификации генома человеческих клеток разных типов, в частности, выделенных человеческих клеток, за исключением эмбриональных стволовых клеток.

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 (РАМ), в непосредственной близости к 3'-концу протоспейсерной последовательности. В настоящем документе раскрыты последовательности РАМ, которые действуют оптимальным образом с белками и полипептидами Cas по изобретению. Со знанием этих последовательностей РАМ 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 человеческим клеткам, содержащим белок или полипептид Cas и молекулу направляющей РНК. Полинуклеотид-мишень также может содержать определенную последовательность РАМ в непосредственной близости к 3'-концу протоспейсерной последовательности, такую как последовательность РАМ, приведенная в настоящем документе. Последовательность РАМ 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. Необязательно, применение по меньшей мере одной молекулы направляющей РНК и белка или полипептида Cas для связывания, расщепления, маркирования или модификации нуклеотидной цепи-мишени, а также способ связывания, расщепления, маркирования или модификации нуклеотидной последовательности-мишени в нуклеотидной цепи-мишени с использованием рибонуклеопротеида или нуклеопротеида по изобретению, можно осуществлять в человеческой клетке. Необязательно, человеческая клетка будет выделенной. Необязательно, человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, белок или полипептид, или рибонуклеопротеид, или нуклеопротеид Cas по изобретению не будет использован для модификации генетических особенностей зародышевой линии человека.

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 РАМ 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 остатков, 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 остатков, 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 тетра-нуклеотидной 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 Microbiol 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, имеющих 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 «РАМ» (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'-СТТ-3', 5'-САТ-3', 5'-ССТ-3', 5'-САС-3', 5'-ТТТ-3', 5'-АТТ-3' и 5'-АWГ-3', где W представляет собой А, Т или У. 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 crPНК, 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. 1А), и были обнаружены PAM (смотри Ran et al., 2015, там же - фиг. 1С). 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 diphtheriae*), 5'-NNGGGTN-3' (*Campylobacter lari*), 5'-NNNCATN-3' (*Parvobaculum lavamentivorans*), 5'-NNNNGTA-3' (*Neisseria 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 конкретных аспектах взаимодействие с последовательностью-мишенью в широком диапазоне температур, например, от 20°C до 100°C, от 20°C до 80°C, от 30°C до 80°C, от 20°C до 70°C или от 25°C до 65°C, может быть достигнуто за счет использования последовательности РАМ 5'-NNNNCVAA-3' [SEQ ID NO: 48]. Для первых 4 положений РАМ конкретные предпочтения отсутствуют. Вследствие этого, первые 4 нуклеотида могут представлять собой любые нуклеотиды (NNNN). Предпочтительно, взаимодействие с последовательностью-мишенью в таком широком диапазоне температур может быть достигнуто за счет использования последовательности РАМ 5'-NNNNCSAA-3' [SEQ ID NO: 49]. Оптимально, РАМ может иметь последовательность 5'-NNNNCGAA-3' [SEQ ID NO: 50] или 5'-NNNNCCAA-3' [SEQ ID NO: 51].

Если взаимодействие с последовательностью-мишенью необходимо при температуре $\geq 30^\circ\text{C}$, например, от 30°C до 100°C, предпочтительно, в диапазоне от 30°C до 70°C, от 30°C до 65°C или от 45°C до 65°C, последовательность РАМ оптимально может представлять собой последовательность 5'-NNNNCNAA-3' [SEQ ID NO: 52] или 5'-NNNNCMCA-3' [SEQ ID NO: 53]. Для первых 4 положений РАМ конкретные предпочтения отсутствуют. Вследствие этого, первые 4 нуклеотида могут представлять собой любые нуклеотиды (NNNN). Необязательно, например, последовательность РАМ может представлять собой 5'-CCCCNAA-3' или 5'-CCCCMCA-3'. Необязательно, например, последовательность РАМ может быть выбрана из 5'-CCCCAAA-3', 5'-CCCCATA-3', 5'-CCCCAGA-3', 5'-CCCCACA-3', 5'-CCCCATA-3', 5'-CCCCCTTA-3', 5'-CCCCCTGA-3', 5'-CCCCCTCA-3', 5'-CCCCGAA-3', 5'-CCCCGTA-3', 5'-CCCCGGA-3', 5'-CCCCGCA-3', 5'-CCCCCAA-3' [SEQ ID NO: 11], 5'-CCCCCTTA-3', 5'-CCCCCGA-3' или 5'-CCCCCCTTA-3'.

В вариантах осуществления изобретения молекула направляющей РНК может иметь длину в диапазоне 35-200 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' [SEQ ID NO: 47], 5'-CNNNCNN-3', 5'-NNNCCNN-3', 5'-NNCNCNN-3', 5'-NNNNCCN-3' и 5'-NCNNCNN-3'. Необязательно, например, последовательность PAM можно выбирать из 5'-NNNNC-3', 5'-NNNNCNNA-3' [SEQ ID NO: 47], 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'-ATCCCCCCAA-3' [SEQ ID NO: 21] или 5'-ACGGCCAA-3' [SEQ ID NO: 22]. Предпочтительно, последовательность PAM будет иметь последовательность 5'-NNNNCNNA-3' [SEQ ID NO: 47]. Однако следует понимать, что в зависимости от желательного применения и/или концентрации белка или полипептида Cas могут быть использованы другие сочетания нуклеотидов. В частности, для первых 4 положений PAM конкретные предпочтения отсутствуют. Вследствие этого, первые 4 нуклеотида могут представлять собой любые нуклеотиды (NNNN). Эти последовательности соответствуют тому, что называют «примыкающий к протоспейсеру мотив», или «PAM», в контексте природных crPНК. В системах CRISPR/Cas типа IIC эти последовательности PAM способствуют стабильному взаимодействию комплекса Cascade/crPНК с его мишенью дцДНК, для обеспечения высокой степени специфичности crPНК - как в природных системах-мишенях, так и, следовательно, предпочтительно, в случае PНК по настоящему изобретению - для последовательности-мишени. Предпочтительно, последовательность в непосредственной близости к протоспейсеру не будет представлять собой 5'-NNNCATN-3'.

Кроме того, последовательность PAM может представлять собой последовательность 5'-NNNNCNNA-3' [SEQ ID NO: 47], например, 5'-NNNNCNAA-3' [SEQ ID NO: 52] или 5'-NNNNCMCA-3' [SEQ ID NO: 53].

Одним из ограничений мезофильной SpCas9 является то, что она проявляет активность лишь при температуре 25–44°C; при более высоких температурах активность SpCas9 быстро снижается до неподдающихся обнаружению уровней (Mougiakos *et al.*, 2017, ACS Synth Biol. 6: 849–861). В отличие от диапазона температур 25–44°C ее мезофильного ортолога SpCas9, ThermoCas9 по настоящему изобретению является активной *in vitro* в более широком диапазоне температур 20–70°C. Более широкие границы активности и стабильности ThermoCas9 позволяют использовать ее в молекулярно-биологических методах, для которых требуются манипуляции с ДНК при температурах 20–70°C, а также использовать ее в жестких условиях, требующих более надежной ферментативной активности. Вследствие этого, ThermoCas9 может быть использована в качестве инструмента для редактирования генома человеческих клеток. Необязательно, указанные человеческие клетки не будут являться эмбриональными стволовыми клетками.

В дополнение к наличию широкого функционального температурного диапазона активности, то есть, способности к функционированию как при низких, так и при высоких, температурах, например, при 20°C и 70°C или при 20°C и 65°C, или при 25°C и 65°C, возможность манипулировать диапазоном температур, при которых ThermoCas9 способна к направленному расщеплению или связыванию, или при которых направленное расщепление или связывание происходит эффективно, путем изменения структурных особенностей ThermoCas9 или связанных элементов (таких как, например, еРНК или tracrRNA), позволило бы иметь более высокий уровень контроля над манипуляциями с последовательностью нуклеиновой кислоты. Однако до настоящего времени было недостаточно информации о молекулярных детерминантах активности Cas9 при конкретных температурах.

Авторы изобретения выявили несколько факторов, которые важны для придания ThermoCas9 термостабильности, одним из них является предпочтение ThermoCas9 в отношении PAM. Предпочтения ThermoCas9 в отношении PAM являются очень строгими для активности в нижней части температурного диапазона ($\leq 30^\circ\text{C}$), в то

время как больше вариабельности PAM допустимо для активности при температурах в диапазоне от умеренных до оптимальных (от 37°C до 60°C). В силу этого, последовательность PAM можно изменять для достижения наиболее эффективного связывания, расщепления, маркирования или модификации мишени при конкретной температуре. Это обеспечивает большую гибкость в использовании ThermoCas9 в зависимости от конкретного варианта применения. Например, в некоторых вариантах применения может быть желателен очень широкий диапазон температур для связывания, расщепления, маркирования или модификации мишени, например, от 20°C до 70°C, предпочтительно, от 20°C до 65°C или от 25°C до 65°C. Связывание, расщепление, маркирование или модификация последовательности-мишени в таком широком диапазоне температур может быть достигнуто за счет использования последовательности PAM 5'-NNNNCVAA-3' [SEQ ID NO: 48]. Предпочтительно, связывание, расщепление, маркирование или модификация последовательности-мишени в таком широком диапазоне температур может быть достигнуто за счет использования последовательности PAM 5'-NNNNCSAA-3' [SEQ ID NO: 49], например, 5'-NNNNCGAA-3' [SEQ ID NO: 50] или 5'-NNNNCCAA-3' [SEQ ID NO: 51]. Для первых 4 положений PAM конкретные предпочтения отсутствуют. Вследствие этого, первые 4 нуклеотида могут представлять собой любые нуклеотиды (NNNN). Необязательно, например, 5'-CCCCCGAA-3' или 5'-CCCCCCAA-3' [SEQ ID NO: 11].

Если связывание, расщепление, маркирование или модификация мишени необходимы при температуре $\geq 30^\circ\text{C}$, например, от 30°C до 100°C, предпочтительно, в диапазоне от 30°C до 70°C, от 30°C до 65°C или от 45°C до 65°C, последовательность PAM оптимально может представлять собой последовательность 5'-NNNNCNAA-3' [SEQ ID NO: 52] или 5'-NNNNCMCA-3' [SEQ ID NO: 53]. Для первых 4 положений PAM конкретные предпочтения отсутствуют. Вследствие этого, первые 4 нуклеотида могут представлять собой любые нуклеотиды (NNNN). Необязательно, например, последовательность PAM может представлять собой 5'-CCCCNAA-3' или 5'-CCCCMCA-3'.

Необязательно, например, последовательность РАМ можно выбирать из 5'-СССССААА-3', 5'-СССССАТА-3', 5'-СССССАГА-3', 5'-СССССАСА-3', 5'-СССССТАА-3', 5'-СССССТТА-3', 5'-СССССТГА-3', 5'-СССССТСА-3', 5'-СССССГАА-3', 5'-СССССГТА-3', 5'-СССССГГА-3', 5'-СССССГСА-3', 5'-ССССССАА-3' [SEQ ID NO: 11], 5'-ССССССТА-3', 5'-ССССССГА-3' или 5'-СССССССА-3'.

Последовательности РАМ 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 РАМ. Однако следует понимать, что могут присутствовать дополнительные нуклеотиды, образующие часть последовательности РАМ 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 РАМ.

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. Таким образом, способы, варианты применения, нуклеопротеиды и трансформированные человеческие клетки по изобретению могут быть полезны для редактирования генома при повышенных температурах. 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.

Вследствие значительно более высокой термостабильности белки Cas по настоящему изобретению сохраняют функцию, например, нуклеазную активность, в гораздо более широком диапазоне температур, чем известные белки Cas9. 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 человеческих клеток при повышенных температурах. Необязательно, за исключением генетической инженерии эмбриональных стволовых клеток. Более широкие границы активности и стабильности ThermoCas9 позволяют использовать ее в молекулярно-биологических методах, для которых требуются манипуляции с ДНК в широком диапазоне температур, например, при температурах 20–70°C, а также использовать ее в жестких условиях, требующих более надежной ферментативной активности.

Авторы изобретения также успешно продемонстрировали, что белки Cas по изобретению могут быть использованы для прямого транскрипционного контроля последовательностей-мишеней, например, для выключения транскрипции за счет специфичного для последовательности связывания с последовательностями-мишенями. Вследствие этого, ThermoCas9 также можно использовать в качестве

инструмента транскрипционного контроля в человеческих клетках, например, для выключения или активации транскрипции генов-мишеней. Таким образом, ThermoCas9 также можно использовать в качестве инструмента выключения генов в человеческих клетках. Необязательно, белки Cas по изобретению не используют в эмбриональных стволовых клетках. Необязательно, белки Cas по изобретению не используют для модификации генетических особенностей зародышевой линии человека.

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 молекуле(ах) полинуклеотида, который должен быть связан, расщеплен, маркирован или модифицирован, осуществляют это при температурах в диапазоне от 20°C до 100°C, необязательно, в диапазоне от 20°C до 70°C, от 20°C до 65°C, от 25°C до 70°C, от 25°C до 65°C, от 55°C до 100°C, от 50°C до 70°C, от 55°C до 70°C

или от 55°C до 65°C.

Предпочтительно, белки или полипептиды Cas по изобретению в ассоциации с соответствующей гРНК (гид-РНК, также называемой молекулой направляющей РНК), которая узнает последовательность-мишень в молекуле(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 РАМ, приведенную в настоящем документе, непосредственно в 3'-положении от протоспейсерной последовательности. Последовательность РАМ

может представлять собой 5'-NNNNNNNA-3', предпочтительно, 5'-NNNNCNNA-3' [SEQ ID NO: 47], необязательно, например, 5'-CCCCCNA-3' [SEQ ID NO: 10] или 5'-CCCCCAA-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.

Последовательность PAM можно изменять для достижения более эффективного расщепления мишени при конкретной температуре. Это обеспечивает большую гибкость в использовании белков Cas по настоящему изобретению в зависимости от конкретного варианта применения. Если активность связывания, расщепления, маркирования или модификации, например, активность расщепления, необходима в диапазоне температур от 20°C до 100°C, предпочтительно, от 20°C до 70°C или от 20°C до 65°C, или от 25°C до 65°C, то эту активность можно осуществлять, используя последовательность PAM 5'-NNNNCVAA-3' [SEQ ID NO: 48], предпочтительно, активность в таком диапазоне температур можно осуществлять, используя последовательность PAM 5'-NNNNCSAA-3' [SEQ ID NO: 49], например, 5'-NNNNCGAA-3' [SEQ ID NO: 50] или 5'-NNNNCCAA-3' [SEQ ID NO: 51]. Необязательно, например, 5'-CCCCGAA-3' [SEQ ID NO: 52] или 5'-CCCCCAA-3' [SEQ ID NO: 11].

Авторы изобретения установили, что термостабильность ThermoCas9 возрастает при связывании с гид-РНК (егРНК) и образовании рибонуклеопротеидного комплекса. Гид-РНК (егРНК), предпочтительно, может содержать tracrРНК и crРНК. В такой структуре гид-РНК, предпочтительно, может содержать crРНК, которая содержит нуклеотидный фрагмент спейсера и фрагмент повтора. crРНК, предпочтительно, может иметь длину 17-20 нт. Необязательно, crРНК может иметь длину 17 нт. Альтернативно, crРНК может иметь длину 18 нт, длину 19 нт или длину 20 нт. Гид-РНК также может содержать tracrРНК (фрагмент анти-повтора

(который спаривается с фрагментом повтора crРНК)). tracrРНК и crРНК могут быть разделены синтетическим линкером. Предпочтительной структурой гид-РНК является следующая: 5'-[crРНК (17-20 нуклеотидный фрагмент спейсера и фрагмент повтора) - (необязательно: синтетическая петля для связывания двух РНК) - tracrРНК (фрагмент анти-повтора (который спаривается с фрагментом повтора crРНК) и некоторые переменные структуры стебель-петля (описанные ниже), которые в некоторых системах могут быть в некоторой степени укорочены)]-3'.

Как правило, tracrРНК будет предоставлена в виде части химерной единой гид-РНК (egРНК), например, содержащей crРНК и tracrРНК. tracrРНК может состоять из области анти-повтора, за которой следуют одна или более шпилечных структур, предпочтительно, две или более шпилечных структур, или более предпочтительно, три или более шпилечных структур. Наличие полноразмерной шпильки из повтора/анти-повтора (образованной 3'-концом части crРНК (повтор) и 5'-концом комплементарной части tracrРНК (анти-повтор) в синтетической химерной egРНК, которые связаны 4-нуклеотидным линкером, например, 5'-GAAA-3')) на дистальном конце спейсера действует в качестве якоря для нуклеазы, однако она не важна для выбора мишени и активности расщепления. Например, делеции на дистальном конце спейсера, например, делеция длиной вплоть до 50 нт в части tracrРНК, может быть допустима, с небольшим влиянием, или без влияния, на эффективность расщепления ДНК. Соответственно, например, делеция на дистальном конце спейсера в полноразмерной шпильке из повтора/анти-повтора может составлять вплоть до 50 нт, вплоть до 45 нт, вплоть до 40 нт, вплоть до 35 нт, вплоть до 30 нт, вплоть до 25 нт, вплоть до 20 нт, вплоть до 15 нт, вплоть до 10 нт или вплоть до 5 нт, без отрицательного влияния на активность расщепления в отношении ДНК-мишени.

Неожиданно авторы изобретения также обнаружили, что структура tracrРНК влияет на термостабильность ThermoCas9 и эффективность активности, в частности, активности расщепления. В частности, число шпилечных (или типа стебель-петля) структур в

tracrRNA или еgRNA можно изменять для достижения наибольшей эффективности связывания, расщепления, маркирования или модификация мишени при конкретной температуре. Это обеспечивает большую гибкость в использовании белков Cas по настоящему изобретению в зависимости от конкретного варианта применения. Необязательно, tracrRNA или еgRNA могут быть предоставлены с нуклеотидной последовательностью, которая способна к образованию одной или более структур стебель-петля, двух или более структур стебель-петля, или трех или более структур стебель-петля. Необязательно, tracrRNA или еgRNA могут быть предоставлены с нуклеотидной последовательностью, которая спроектирована для образования одной или более структур стебель-петля, двух или более структур стебель-петля, или трех или более структур стебель-петля. Предпочтительно, еgRNA будет предоставлена с нуклеотидной последовательностью, которая способна к образованию по меньшей мере трех структур стебель-петля.

Необязательно, если активность связывания, расщепления, маркирования или модификации, например, активность расщепления, необходима в диапазоне температур от 20°C до 60°C, предпочтительно, от 37°C до 60°C или при 37°C, 40°C, 45°C, 50°C, 55°C или 60°C, то эту активность можно осуществлять, используя последовательность еgRNA, которая способна к образованию одной или более структур стебель-петля.

Необязательно, если активность связывания, расщепления, маркирования или модификации, например, активность расщепления, необходима в диапазоне температур от 20°C до 65°C, предпочтительно, от 37°C до 65°C, более предпочтительно, от 45°C до 55°C или при 37°C, 40°C, 45°C, 50°C, 55°C или 60°C, то эту активность можно осуществлять, используя последовательность еgRNA, которая способна к образованию двух или более структур стебель-петля.

Необязательно, если активность связывания, расщепления, маркирования или модификации, например, активность расщепления, необходима в диапазоне температур от 20°C до 100°C, предпочтительно, от 20°C до 70°C, более предпочтительно, от 37°C

до 65°C или при 37°C, 40°C, 45°C, 50°C, 55°C, 60°C или 65°C, то эту активность можно осуществлять, используя последовательность егРНК, которая способна к образованию трех или более структур стебель-петля.

Предпочтительно, часть егРНК, соответствующая tracrРНК, будет содержать последовательность AAGGGCUUUCUGCCUAUAGGCAGACUGCCC [SEQ ID NO: 54], которая является примером 5'-шпильки. Предпочтительно, часть егРНК, соответствующая tracrРНК, будет дополнительно содержать последовательность GUGGCGUUGGGGAUCGCCUAUCGCC [SEQ ID NO: 55], которая является примером «средней» шпильки. Предпочтительно, часть егРНК, соответствующая tracrРНК, будет дополнительно содержать последовательность CGCUUUCUUCGGGCAUUCSSCACUCUUAGGCGUUUU [SEQ ID NO: 56], которая является примером 3'-шпильки.

Предпочтительно, часть егРНК, соответствующая tracrРНК, будет содержать последовательность AAGGGCUUUCUGCCUAUAGGCAGACUGCCCGUGGCGUUGGGGAUCGCCUAUCGCC [SEQ ID NO: 57], то есть, включающую 5'-шпильку и среднюю шпильку.

Предпочтительно, часть егРНК, соответствующая tracrРНК, может содержать последовательность; AAGGGCUUUCUGCCUAUAGGCAGACUGCCCGUGGCGUUGGGGAUCGCCUAUCGCCCGCUUUCUU CGGGCAUUCSSCACUCUUAGGCGUUUU [SEQ ID NO: 58], то есть, включающую 5'-шпильку, среднюю шпильку и 3'-шпильку.

Авторы изобретения обнаружили, что число предсказанных структур стебель-петля в каркасе tracrРНК играет чрезвычайно важную роль в расщеплении ДНК, в частности, при повышенных температурах. Они определили, что, хотя наличие трех структур стебель-петля в каркасе tracrРНК не является необходимым для активности расщепления, в случае присутствия всех трех петель расщепление является наиболее эффективным при всех температурах в диапазоне, это свидетельствует о том, что полноразмерная tracrРНК необходима для оптимального расщепления ДНК за счет ThermoCas9 при повышенных температурах. Напротив, удаление 3'-шпильки приводит к снижению эффективности расщепления. Более

того, авторы изобретения обнаружили, что удаление как средней шпильки, так и 3'-шпильки, приводит к резкому снижению эффективности расщепления ThermoCas9, в частности, на верхней и нижней границах функциональных диапазонов температур. Предпочтительно, если связывание, расщепление, маркирование или модификация последовательности-мишени необходимы при повышенных температурах, например, от 45°C до 100°C, от 50°C до 100°C, от 50°C до 70°C, от 50°C до 65°C, от 55°C до 65°C или в широком диапазоне температур, например, от 20°C до 100°C, от 20°C до 70°C, от 20°C до 65°C. Предпочтительно, ThermoCas9, связанная с еРНК с тремя структурами стебель-петля, будет оставаться стабильной и способной к связыванию, расщеплению, маркированию или модификации последовательности-мишени в течение по меньшей мере 1 мин, по меньшей мере 2 мин, по меньшей мере 3 мин, по меньшей мере 4 мин или по меньшей мере 5 мин, предпочтительно, 5 мин, при выбранной температуре в диапазоне от 20°C до 100°C, от 20°C до 70°C, от 20°C до 65°C, от 45°C до 100°C, от 50°C до 100°C, от 50°C до 70°C, от 50°C до 65°C или от 55°C до 65°C.

Кроме того, авторы изобретения также установили, что длину последовательности спейсера еРНК можно изменять для изменения эффективности активности ThermoCas9, например, активности связывания, расщепления, маркирования или модификации. Как правило, последовательность спейсера будет иметь длину в диапазоне от 18 нт до 25 нт. Необязательно, последовательность спейсера будет иметь длину 18 нт, 19 нт, 20 нт, 21 нт, 22 нт, 23 нт, 24 нт или 25 нт. Предпочтительно, будет использован спейсер длиной 19 нт, 20 нт, 21 нт или 23 нт, поскольку белки Cas9 по изобретению расщепляют последовательности-мишени с наибольшей эффективностью, когда они связаны с молекулами еРНК, имеющими такие длины спейсеров. Эффективность расщепления значительно падает при использовании спейсера длиной 18 нт. Предпочтительно, длина спейсера будет составлять 23 нт.

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-метиلاзы, 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 FokI.

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-метиلاзы, 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 (dCas), 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 dCas 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.

Предпочтительно, мутации D8A и H582A в gtCas9 (ThermoCas9) можно использовать для создания каталитически неактивного, или «мертвого», белка Cas или полипептидного варианта ThermoCas9 (dCas), лишённого нуклеазной активности. Такой dCas может найти применение, например, в качестве эффективного термоактивного инструмента выключения транскрипции CRISPRi, способного стабильно и специфически связываться с элементами ДНК без внесения разрывов в дцДНК. Предпочтительно, такая система сможет, помимо прочего, значительно облегчать изучение метаболизма человеческих клеток.

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 специфической направленностью 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. Такие способы включают ГР и сайт-специфическую активность gtCas9, в результате имеет место контрелекция за счет активности gtCas9, приводящей к удалению

клеток, 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 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. Данные способы и варианты применения приводят к увеличению общей популяции клеток с нужной мутацией и при этом элиминации любых не модифицированных 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 клеток, которые имеют эндогенный репарационный механизм NHEJ. Способы и варианты применения, описанные в настоящем документе, могут быть применены к клеткам, 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 РАМ, необходимой для рибонуклеазного комплекса *gtCas9*. РАМ 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 РАМ, 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 клетки. 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 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 клетку, имеющую желательную модификацию, из не модифицированных 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 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 двухцепочечной и переносимой в плазмиде.

ГР в способах и вариантах применения, предложенных в настоящем документе, можно использовать для удаления полинуклеотидной последовательности из генома. Альтернативно, ГР в способах и вариантах применения, предложенных в настоящем документе, можно использовать для вставки одного или более генов, или их фрагментов, в 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 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 клетках, в котором активность *gtCas9* приводит к внесению ДЦР и может индуцировать клеточную ГР в 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 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 геной 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

с. 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. Необязательно, белок Cas способен к указанному расщеплению, маркированию или модификации в человеческих клетках. Необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, при этом указанное расщепление, маркирование или модификация не приводит к изменению генетических особенностей зародышевой линии человека.

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.

Нуклеазная активность ThermoCas9: двухвалентные катионы

Ранее охарактеризованные мезофильные эндонуклеазы Cas9 используют двухвалентные катионы для катализа образования ДЦР в ДНК-мишени. Авторы изобретения продемонстрировали, что ThermoCas9 могут опосредовать расщепление дцДНК в присутствии любого из следующих двухвалентных катионов: Mg^{2+} , Ca^{2+} , Mn^{2+} , Co^{2+} , Ni^{2+} и Cu^{2+} .

Нуклеазная активность ThermoCas9: субстраты

Авторы изобретения также неожиданно обнаружили, что, несмотря на сообщения о том, что некоторые системы типа IIC эффективно разрезают одноцепочечную ДНК ((Ma, et al., *Mol. Cell* 60, 398-407 (2015); Zhang, et al., *Mol. Cell* 60, 242-255 (2015)), ThermoCas9 неспособна к прямому расщеплению оцДНК. Нуклеазная активность ThermoCas9 ограничена дцДНК в качестве субстратов.

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. Очевидно, что подходящие регуляторные элементы будут варьироваться в зависимости от интересующей клетки-хозяина. Регуляторные элементы, которые способствуют высокому уровню экспрессии в человеческих клетках-хозяевах, 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 человеческих клетках.

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 человеческой 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, отдельно 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. Необязательно, указанная нуклеиновая кислота находится в человеческой клетке. Необязательно, указанная человеческая клетка не является эмбриональной стволовой клеткой.

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 отдельно. 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 человеческой клетке, будь то выделенная клетка или часть многоклеточной ткани, органа или организма. В контексте цельной ткани и органов и в контексте организма способ желателен осуществлять *in vivo* или, альтернативно, можно осуществлять путем выделения человеческой 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. Необязательно, указанная человеческая клетка не является эмбриональной стволовой клеткой.

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:

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

б. 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 человеческой 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. Необязательно, указанная нуклеиновая кислота-мишень находится в человеческой клетке, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой.

В следующем аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в человеческой 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. Необязательно, указанная нуклеиновая кислота-мишень находится в человеческой клетке, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой.

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 человеческую 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 плаزمиды 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 человеческой 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. Необязательно, указанный способ применяют к человеческой клетке, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой.

Изобретение также относится к способу гомологичной рекомбинации нуклеиновой кислоты с молекулой дцДНК в желаемом локусе в человеческой 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. Необязательно, указанный способ применяют к человеческой клетке, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой.

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; и при этом указанный организм является человеком.

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'-эксонуклеазы или 5'-эксонуклеазы; и при этом указанный организм является человеком.

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 в человеческой клетке, 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. Необязательно, указанная человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, при этом указанный способ не приводит к модификации генетических особенностей зародышевой линии человека.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в человеческой 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 негомологического соединения концов (NHEJ); необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой.

Необязательно, при этом указанный способ не приводит к модификации генетических особенностей зародышевой линии человека.

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

а. создания двухцепочечного разрыва в желаемом локусе; и

б. репарации молекулы дцДНК в клетке путем гомологичной рекомбинации; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, при этом указанный способ не приводит к модификации генетических особенностей зародышевой линии человека.

В другом аспекте настоящее изобретение относится к способу модификации экспрессии генов в человеческой 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; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, при этом указанный способ не приводит к модификации генетических особенностей зародышевой линии человека.

В другом аспекте настоящее изобретение относится к способу модификации экспрессии генов в человеческой 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; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, при

этом указанный способ не приводит к модификации генетических особенностей зародышевой линии человека.

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; необязательно, при этом нуклеиновая кислота-мишень находится в человеческой клетке. Необязательно, указанная человеческая клетка не является эмбриональной стволовой клеткой. Необязательно, при этом указанный способ не приводит к модификации генетических особенностей зародышевой линии человека.

В любом из способов модификации нуклеиновой кислоты-мишени, описанных выше в настоящем документе, клетка может представлять собой человеческую клетку. Необязательно, в любом из способов модификации нуклеиновой кислоты-мишени, описанных в настоящем документе, указанная человеческая клетка не может представлять собой эмбриональную стволовую клетку. Необязательно, указанные способы не приводят к модификации генетических особенностей зародышевой линии человека.

Клетки-хозяева

Преимуществом способов, описанных в настоящем документе, является широкая область применения, и клетки-хозяева по настоящему изобретению могут быть получены из любых человеческих клеток, которые можно культивировать. Соответственно, настоящее изобретение относится к человеческой клетке, трансформированной способом, описанным выше в настоящем документе. Изобретение относится к трансформированной человеческой 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. Необязательно, указанная человеческая клетка не может представлять собой эмбриональную стволовую клетку.

Для использования по настоящему изобретению могут быть выбраны обычно используемые клетки-хозяева, включая человеческие клетки, которые являются генетически доступными и которые можно культивировать, например, включая человеческие клетки (но не эмбриональные стволовые клетки). Предпочтительные клетки-хозяева для использования по настоящему изобретению, как правило, являются клетками, которые обычно демонстрируют высокие темпы роста, легко поддаются культивированию и/или трансформации, характеризуются коротким временем генерации, имеют установленные, связанные с ними генетические ресурсы, или клетками, подвергнутыми 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 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 человеческую клетку. Способы и варианты применения изобретения, описанные в настоящем документе, могут быть использованы для модификации геномов человеческих клеток.

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-реплик; 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-реплик.

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*/«ана»; SEQ ID NO: 8) и типа II-A (*S. pyogenes*/«пуо»; 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 линиями. 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 А 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 В 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 (crРНК) (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 CCCCCAA [SEQ ID NO: 11].

На фигуре 11 представлены результаты *in vitro* анализов в диапазоне температур с использованием плазмиды-мишени с последовательностью PAM CCCCCAA [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 конструктором 1 (последовательность PAM ATCCSSAA [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.

На фигуре 15 показан локус CRISPR-Cas типа IIC *Geobacillus thermodenitrificans* T12, кодирующий термостабильный гомолог Cas9, ThermoCas9.

(A) Схематическое изображение геномного локуса, кодирующего ThermoCas9. Доменная структура ThermoCas9, основанная на сравнении последовательностей, с предсказанными остатками активных центров, выделенными красным цветом. Показана гомологичная модель ThermoCas9, созданная с использованием Phyre 2 (Kelley *et al.* *Nat. Protoc.* 10, 845-858 (2015)), домены изображены разными цветами.

(B) Филогенетическое дерево ортологов Cas9, которые в высокой степени идентичны ThermoCas9. Эволюционный анализ выполняли в MEGA7 (Kumar *et al.* *Mol. Biol. Evol.* 33, 1870-1874 (2016)).

(C) Результаты SDS-ПААГ ThermoCas9 после очистки металл-аффинной хроматографией и гель-фильтрацией. Расположение полученной одной полосы согласуется с теоретической молекулярной массой 126 кДа апо-ThermoCas9.

На фигуре 16 показан анализ PAM для ThermoCas9.

(A) Схематическое изображение *in vitro* анализа расщепления

для выяснения положения и идентификации (5'-NNNNNNN-3') мотива, примыкающего к протоспейсеру (PAM). Черные треугольники указывают положение участка расщепления.

(B) Последовательность logo консенсусного 7-нт PAM для ThermoCas9, полученная путем сравнительного анализа расщепления при помощи ThermoCas9 библиотек-мишеней. Высота букв в каждом положении определяется содержанием информации.

(C) Продление структуры PAM до 8-го положения с использованием *in vitro* анализа расщепления. Четыре линейаризованные плазмиды-мишени, каждая из которых содержит одну конкретную PAM 5'-CCCCCAN-3', инкубировали с ThermoCas9 и егРНК при 55°C в течение 1 часа, затем анализировали методом электрофореза в агарозном геле.

(D) *In vitro* анализы расщепления ДНК-мишеней с разными PAM при 30°C и 55°C. Шестнадцать линейаризованных плазмид-мишеней, каждая из которых содержит одну конкретную PAM 5'-CCCCCNNA-3' [SEQ ID NO: 13], инкубировали с ThermoCas9 и егРНК, затем анализировали на эффективность расщепления методом электрофореза в агарозном геле. Смотри также фигуру 21.

На фигуре 17 показано, что ThermoCas9 является активной в широком диапазоне температур, и ее термостабильность возрастает при связывании с егРНК.

(A) Схематическое изображение егРНК и соответствующей ДНК-мишени. ДНК-мишень изображена в виде прямоугольника с черным контуром, и PAM изображена в виде темно-серого горизонтального эллипса с черным контуром. crРНК изображена в виде темно-серого прямоугольника с черным контуром, и сайт, в котором 3'-конец crРНК связан с 5'-концом tracrРНК, изображен в виде черного вертикального эллипса. Черный бокс с белыми буквами и светло-серый бокс с черными буквами изображают предсказанные три и две петли на 3'-стороне tracrРНК, соответственно. 41-нт укорочение области повтора/анти-повтора, образованной комплементарными 3'-концом crРНК и 5'-концом tracrРНК, изображено в виде длинной светло-серой вертикальной пунктирной линии. Предсказанное 3'-

положение первой петли tracrRNA отмечено черным треугольником и черной пунктирной линией. Предсказанное 3'-положение второй петли tracrRNA отмечено белым треугольником и черной пунктирной линией. Предсказанное 3'-положение третьей петли tracrRNA отмечено белым треугольником и белой пунктирной линией.

(B) Важность предсказанных трех структур стебель-петля в каркасе tracrRNA тестировали, осуществляя транскрипцию укороченных вариантов egRNA и оценивая их способность направлять ThermoCas9 для расщепления ДНК-мишени при разных температурах. Показаны средние значения для по меньшей мере двух биологических реплик, планки погрешностей соответствуют S.D.

(C) Для определения максимальной температуры анализировали эндонуклеазную активность РНП комплекса ThermoCas9:egRNA после инкубации при 60°C, 65°C и 70°C в течение 5 или 10 мин. Добавляли предварительно нагретую ДНК-субстрат, и реакционную смесь инкубировали в течение 1 часа при соответствующей температуре.

(D) Сравнение активного диапазона температур ThermoCas9 и SpCas9 в анализах активности, проводимых после 5-мин инкубации при указанной температуре. Добавляли предварительно нагретую ДНК-субстрат, и реакционную смесь инкубировали в течение 1 часа при той же температуре.

На фигуре 18 показана модификация генома термофилов с использованием ThermoCas9.

(A) Схематическое изображение основного конструкта pThermoCas9_Δинтересующий ген (goi). Ген *thermocas9* вводили либо в pNW33n (*B. smithii*), либо в pEMG (*P. putida*), вектор. Фланкирующие последовательности для гомологичной рекомбинации были введены выше *thermocas9* и включали 1-т.п.н. (*B. smithii*) или 0,5 т.п.н. (*P. putida*) расположенную выше область и 1-т.п.н. или 0,5 т.п.н. расположенную ниже область от интересующего гена (goi) в геноме-мишени. Экспрессирующий egRNA модуль вводили ниже гена *thermocas9*. Поскольку точка начала репликации (ori), белок репликации (rep), маркер устойчивости к антибиотику (AB) и возможные вспомогательные элементы (AE) являются специфическими для каркаса, они представлены пунктирной линией.

(B) Результаты электрофореза в агарозном геле, показывающие полученные продукты геном-специфической ПЦР для десяти колоний после процесса удаления при помощи *ThermoCas9* гена *pyrF* из генома *B. smithii* ET 138. Все десять колоний имели генотип $\Delta pyrF$, и одна колония представляла собой чистый мутант $\Delta pyrF$, лишенный продукта дикого типа.

(C) Схематическое изображение основного конструкта p*ThermoCas9i_goi*. Имея целью экспрессию каталитически неактивной *ThermoCas9* (*Thermo-dCas9:D8A*, мутант H582A), вносили соответствующие мутации для создания гена *thermo-dcas9*. Ген *thermo-dcas9* был введен в вектор p*NW33n*. Экспрессирующий eРНК модуль вводили ниже *thermo-dcas9*.

(D) Графическое представление данных, относящихся к продуцированию, росту и результатам ОТ-кПЦР из эксперимента по выключению гена *ldhL* с использованием *Thermo-dCas9*. Графики представляют продуцирование лактата, оптическую плотность при 600 нм и процентную долю транскрипции *ldhL* в культурах с примененным подавлением, в сравнении с контрольными культурами. Показаны средние значения для по меньшей мере двух биологических реплик, планки погрешностей соответствуют S.D.

На фигуре 19 показано множественное выравнивание последовательностей ортологов Cas9 типа II-A, B и C. Последовательности белков Cas9 из *Streptococcus pyogenes* (Sp), *Streptococcus thermophilus* (St), *Wolinella succinogenes* (Ws), *Neisseria meningitidis* (Nm), *Actinomyces naeslundii* (An) и *Geobacillus thermodenitrificans* (*Thermo*) были выровнены с использованием ClustalW1 в MEGA7 2 с параметрами по умолчанию; ESPrpt3 использовали для визуализации. Строго консервированные остатки показаны белым текстом на сером фоне; сходные остатки показаны черным текстом в белых вертикальных прямоугольниках с черным контуром. Пирамиды указывают на два консервативных нуклеазных домена во всех последовательностях. Горизонтальные черные стрелки и спирали указывают на β -цепи и α -спирали, соответственно, во вторичной структуре SpCas9 (белковая база данных nr 4CMP4). Структурные домены указаны для SpCas9 и

ThermoCas9 с использованием той же цветовой схемы, что и на фигуре 15А.

На фигуре 20 показаны результаты *in silico* определения PAM. На панели (А) показаны два хита, полученные для фаговых геномов с использованием CRISPRtarget6. На панели (В) показана последовательность logo консенсусного 7-нт PAM для ThermoCas9, полученная путем *in silico* анализа PAM. Высота букв в каждом положении определяется содержанием информации.

На фигуре 21 показано определение PAM для ThermoCas9. *In vitro* анализы расщепления для ДНК-мишеней с разными PAM при 20°C, 37°C, 45°C и 60°C. Семь (20°C) или шестнадцать (37°C, 45°C, 60°C) линейаризованных плазмид-мишеней, каждая из которых содержит конкретную PAM 5'-CCCCNNA-3' [SEQ ID NO: 13], инкубировали с ThermoCas9 и егРНК, затем анализировали методом электрофореза в агарозном геле.

На фигуре 22 показана активность ThermoCas9 в широком диапазоне температур с использованием егРНК, содержащей одну петлю. Важность предсказанных трех структур стебель-петля в каркасе tracrРНК тестировали, осуществляя транскрипцию укороченных вариантов егРНК и оценивая их способность направлять ThermoCas9 для расщепления ДНК-мишени при разных температурах. Выше показан эффект одной петли на активность ThermoCas9 при разных температурах. Показаны средние значения для по меньшей мере двух биологических реплик, планки погрешностей соответствуют S.D.

На фигуре 23 показано, что ThermoCas9 опосредует направленное действие на дцДНК, используя двухвалентные катионы в качестве катализаторов, и не расщепляет оцДНК. На панели (А) показано *in vitro* расщепление плазмидной ДНК ThermoCas9 с ЭДТА и разными ионами металлов. M=1 т.п.н. маркеры ММ ДНК. На панели (В) показана активность ThermoCas9 в отношении оцДНК-субстратов. M=10-п.н. маркеры ММ ДНК.

На фигуре 24 показан выбор спейсера для эксперимента по выключению *ldhL*. Схематическое представление отжига спейсера (егРНК)-протоспейсера во время процесса выключения *ldhL*;

выбранный протоспейсер находится на нематричной цепи и на 39 нт ниже иницирующего кодона гена *ldhL*.

На фигуре 25 представлена карта плазмиды pThermoCas9_pp Δ pyrF, состоящей из каркаса pEMG, фланкирующей области *pyrF Pseudomonas putida* и гена *thermocas9*, а также направляющей на *pyrF Pseudomonas putida* еРНК.

На фигуре 26 представлены результаты капиллярного электрофореза в геле, показывающие продукты, полученные при проведении геном-специфической ПЦР на полученных колониях после процесса удаления за счет ThermoCas9 гена *pyrF* из генома *Pseudomonas putida*. Полоса 1854 п.н. и полоса 1112 п.н. соответствуют генотипу *pyrF* и Δ *pyrF*, соответственно.

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

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

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRI FDRAENPKTGESLALPRRLARSARRR
 LRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNDELARILLHLAK
 RRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVAR
 DDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPK
 ATYTFQSFTVWEHINKLRLVSPGGIRALTDDERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFK
 GLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYKGGAAKSFRPIDFDTFGYALTMFKDDTDI
 RSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTAC
 ERAGYTFTGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAI IKKYGSPVSIHIELARELSQ
 SFDERRKMQKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLOPIEIER
 LLEPGYTEVDHVI PYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGLSERWQQFETFVLTNK
 QFSKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRI
 TAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAFYQRREQNKELSKKTDPOFPQPW
 PHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGID
 ERSGKIQTVVKKLSEIQLDKGTGHFFMYGKESDPRTYEAIRQLLEHNNDPKKAFQEPLYKPKK
 NGELGPIIRTIKI IDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYCVPIYTIIDMMKGILPN
 KAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQTIIDSS
 NGGLSLVSHDNNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL*

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

thermodenitrificans T12

ATGAAGTATAAAATCGGTCTTGATATCGGCATTACGTCTATCGGTTGGGCTGTCATTAA
TTTGGACATTCTCGCATCGAAGATTTAGGTGTCCGCATTTTTGACAGAGCGGAAAACCCGAAA
ACCGGGGAGTCACTAGCTCTTCCACGTGCCTCGCCCGCTCCGCCGACGTCGTCTGCGGCGTC
GCAAACATCGACTGGAGCGCATTCGCCGCCTGTTTCGTCCGCGAAGGAATTTAACGAAGGAAGA
GCTGAACAAGCTGTTTGAAAAAAGCACGAAATCGACGTCTGGCAGCTTCGTGTTGAAGCACTG
GATCGAAAATAAATAACGATGAATTAGCCCGCATCCTTCTTCATCTGGCTAAACGGCGTGGAT
TTAGATCCAACCGCAAGAGTGAGCGCACCAACAAAGAAAACAGTACGATGCTCAAACATATTGA
AGAAAACCAATCCATTCTTTCAAGTTACCGAACGGTTGCAGAAATGGTTGTCAAGGATCCGAAA
TTTTCCCTGCACAAGCGTAATAAAGAGGATAATTACACCAACACTGTTGCCCGCGACGATCTTG
AACGGGAAATCAAACCTGATTTTCGCCAAACAGCGCGAATATGGGAACATCGTTTGCACAGAAGC
ATTTGAACACGAGTATATTTCCATTTGGGCATCGCAACGCCCTTTTGCTTCTAAGGATGATATC
GAGAAAAAAGTCGGTTTTCTGTACGTTTGAGCCTAAAGAAAAACGCGCGCCAAAAGCAACATACA
CATTCCAGTCCTTCACCGTCTGGGAACATATTAACAAACTTCGTCTTGTCTCCCCGGGAGGCAT
CCGGGCACTAACCGATGATGAACGTCGTCTTATATAACAAGCAAGCATTTTCATAAAAATAAAATC
ACCTTCCATGATGTTTGAACATTTGCTTAACTTGCCCTGACGACACCCGTTTTAAAGGTCTTTTTAT
ATGACCGAAACACCACGCTGAAGGAAAATGAGAAAGTTCGCTTCCTTGAACCTCGGCGCCTATCA
TAAAATACGGAAAGCGATCGACAGCGTCTATGGCAAAGGAGCAGCAAATCATTTTCGTCCGATT
GATTTTGATACATTTGGCTACGCATTAACGATGTTTAAAGACGACACCGACATTCGCAGTTACT
TGCGAAACGAATACGAACAAAATGGAAAACGAATGGAAAATCTAGCGGATAAAGTCTATGATGA
AGAATTGATTGAAGAACTTTTAAACTTATCGTTTTCTAAGTTTGGTCATCTATCCCTTAAAGCG
CTTCGCAACATCCTTCCATATATGGAACAAGGCGAAGTCTACTCAACCGCTTGTGAACGAGCAG
GATATACATTTACAGGGCCAAAGAAAAACAGAAAACGGTATTGCTGCCGAACATTCGCGCGAT
CGCCAATCCGGTCGTCATGCGCGCACTGACACAGGCACGCAAAGTGGTCAATGCCATTATCAAA
AAGTACGGCTCACCGGTCTCCATCCATATCGAACTGGCCCGGGAACATCACAATCCTTTGATG
AACGACGTAATAATGCAGAAAGAACAGGAAGGAAACCGAAAGAAAAACGAAACTGCCATTGCGCA
ACTTGTTGAATATGGGCTGACGCTCAATCCAACCTGGGCTTGACATTGTGAAATTCAAACCTATGG
AGCGAACAAAACGGAAAATGTGCCTATTTACTCCAACCGATCGAAATCGAGCGGTTGCTCGAAC
CAGGCTATACAGAAGTCGACCATGTGATTCCATACAGCCGAAGCTTGGACGATAGCTATACCAA
TAAAGTTCTTGTGTTGACAAAGGAGAACCGTGAAAAAGGAAACCGCACCCAGCTGAATATTTA
GGATTAGGCTCAGAACGTTGGCAACAGTTCGAGACGTTTGTCTTGACAAATAAGCAGTTTTCGA
AAAAGAAGCGGGATCGACTCCTTCGGCTTCATTACGATGAAAACGAAGAAAATGAGTTTAAAAA
TCGTAATCTAAATGATACCCGTTATATCTCACGCTTCTTGGCTAACTTTATTCGCGAACATCTC
AAATTCGCCGACAGCGATGACAAACAAAAAGTATACACGGTCAACGGCCGTATTACCGCCCATT
TACGCAGCCGTTGGAATTTTAAACAAAACCGGGAAGAATCGAATTTGCATCATGCCGTGATGC
TGCCATCGTCGCCTGCACAACGCCGAGCGATATCGCCCGAGTCACCGCCTTCTATCAACGGCGC

GAACAAAACAAAGAАСТGTCCAAAАAGACGGATCCGCAGTTTCCGCAGCCTTGGCCGCACTTTG
 CTGATGAACTGCAGGCGCGTTTATCAAAAАATCCAAAGGAGAGTATAААAGCTCTCAATCTTGG
 AAATTATGATAACGAGAAACTCGAATCGTTGCAGCCGGTTTTTGTCTCCCGAATGCCGAAGCGG
 AGCATAACAGGAGCGGCTCATCAAGAAACATTGCGGCGTTATATCGGCATCGACGAACGGAGCG
 GAAAAATACAGACGGTCGTCAAAAAGAACTATCCGAGATCCAACTGGATAААACAGGTCATTT
 CCCAATGTACGGGAAAGAAAGCGATCCAAGGACATATGAAGCCATTCGCCAACGGTTGCTTGAA
 CATAACAATGACCCAAAААAGGCGTTTCAAGAGCCTCTGTATAААACCGAAGAAGAACGGAGAAC
 TAGGTCSTATCATCCGAACAATCAAAATCATCGATACGACAAATCAAGTTATTCCGCTCAACGA
 TGGCAААACAGTCGCCTACAACAGCAACATCGTGCGGGTTCGACGTCTTTGAGAAAGATGGCAАА
 TATTATTGTGTCCSTATCTATACAATAGATATGATGAAAGGGATCTTGCCAAACAAGGCGATCG
 AGCCGAACAAACCGTACTCTGAGTGGAAGGAAATGACGGAGGACTATAACATTCGATTCACTCT
 ATACCCAAATGATCTTATCCGTATCGAATTTCCCGAGAAAААACAATAАAGACTGCTGTGGGG
 GAAGAAATCAAAATTAAGGATCTGTTTCGCCTATTATCAAACCATCGACTCCTCCAATGGAGGGT
 TAAGTTTGGTTAGCCATGATAACAАСТTTTCGCTCCGCAGCATCGGTTCAAGAACCSTCAAACG
 ATTCGAGAAATACCAAGTAGATGTGCTAGGCAACATCTACAAAGTGAGAGGGGAAAAGAGAGTT
 GGGGTGGCGTCACTTTCTCATTCGAAAGCCGGGGAAACTATCCGTCCGTTATAА

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 «Т12». Белок Cas9 из *G. thermodenitrificans* Т12 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 rBLAST и nBLAST на сервере BLAST в собственной организации, при этом последовательность либо белка, либо гена, *G. thermodenitrificans* Т12 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 обнаружения 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 Мегаб с использованием ClustalW, после чего создавали дерево методом присоединения соседей и проводили бутстрап-анализ с использованием 1000 реплик.

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 геномном 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 crPНК, которые состоят из «спейсера» (мишени), связанного с областью повтора, 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 crPНК 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 (мотив, примыкающий к протоспейсеру), 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 tracrPНК модуль. Матрицу CRISPR используют для идентификации модуля crPНК. Экспрессирующая tracrPНК 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). tracrPНК должна состоять из 5'-последовательности с высоким уровнем комплементарности с прямым повтором 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 еРНК. 5'-конец еРНК 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 еРНК, 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 конструкторов «протоспейсера», каждый со всеми возможными 4096 разными комбинациями 6-нуклеотидных 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. coli* 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 OD₆₀₀ 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 MgCl₂, 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 РАМ направляющего модуля егРНК для qtCas9

После *in silico* определения экспрессирующего *tracr*РНК модуля ДНК в геноме штамма T12 *G. thermodenitrificans* (смотри пример 4, выше), проектировали экспрессирующий единую гид-РНК (егРНК) модуль ДНК, объединяющий модули для *cr*РНК и *tracr*РНК 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 пТ7_егРНК синтезировали с использованием Baseclear и вводили в вектор pUC57, получая вектор pUC57:пТ7_егРНК. DH5 α -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 пТ7_егРНК. Получали модуль ДНК для пТ7_егРНК длиной 218-п.н. (из которых первые 18 п.н. соответствуют пТ7) 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 0,5x 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 РАМ и линеаризация библиотеки

Проектировали и конструировали библиотеку РАМ 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 РАМ, и когда протоспейсер фланкирован правильной РАМ, он может быть узан в качестве мишени егРНК, связанной с Cas9, и расщеплен. Библиотеку РАМ 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 отоженных 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 линейаризации выделенной плазмиды.

Проектирование и проведение реакций определения PAM

Готовили следующую реакционную смесь для расщепления с целью индуцированного gтCas9 введения дцДНК разрывов в представители библиотеки PAM, содержащих правильную PAM ниже 3'-конца протоспейсера-мишени:

1. 2,5 мкг ВКЭ *E. coli*::pRham:cas9_{gt} на реакцию
2. егРНК до конечной концентрации 30 нМ
3. 200 нг линейаризованной библиотеки PAM на реакцию
4. 2 мкл буфера расщепления (100 мМ натрий-фосфатный буфер (рН 7,5), 500 мМ NaCl, 25 мМ MgCl₂, 25% (об/об) глицерина, 5 мМ ДТТ)
5. MQ вода до конечного объема 20 мкл.

Реакционную смесь инкубировали в течение 1 ч при 60°C, а затем реакцию останавливали после добавления 4 мкл бх красителя для нанесения на гель (NEB). Затем реакционную смесь загружали 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 gтCas9 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: А-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
- H₂O - 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 тиминным 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 линейаризованные) 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 РАМ.

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 РАМ СССССАА [SEQ ID NO: 11] и разных концентраций *gtCas9* (фигура 10).

Плазмиду-мишень с последовательностью РАМ СССССАА [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 РАМ

Для подтверждения того, что 8-нуклеотидные РАМ 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 РАМ 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 РАМ:

1. ТССАТТСС (отрицательный контроль в соответствии с результатами *in vitro* анализов; 3'-конец протоспейсера, являющегося мишенью егРНК, закодированной на конструкторе номер 3)
2. АТССССАА (3'-конец протоспейсера, являющегося мишенью егРНК, закодированной на конструкторе номер 1; [SEQ ID NO: 21])
3. АСГГССАА (3'-конец протоспейсера, являющегося мишенью егРНК, закодированной на конструкторе номер 2, [SEQ ID NO: 22]).

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

1. Когда клетки были трансформированы конструктором, направленным на протоспейсер, который имел последовательность ТССАТТСС отрицательного контроля РАМ на 3'-конце (конструктор номер 3), эффективность трансформации не была изменена (фигура 12 А). Число колоний было в том же диапазоне, что и число колоний после трансформации конструктором положительного контроля рNW33n (фигура 12 В). 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 РАМ действительно не была узнана *gtCas9 in vivo*.

2. Когда клетки были трансформированы конструктором номер 1, было получено лишь несколько колоний (фигура 12 С), при сравнении с положительным контролем (клетками, трансформированными рNW33n). 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 РАМ АТССССАА [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 конструктором 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.

Пример 10: Идентификация и очистка ThermoCas9

Авторы изобретения недавно выделили и секвенировали штамм T12 *Geobacillus thermodenitrificans*, грамположительной, умеренно термофильной бактерии с оптимальной температурой роста 65°C (Daas *et al. Biotechnol. Biofuels* 9, 210 (2016)). В противоположность прежним заявлениям, что системы CRISPR-Cas типа II отсутствуют у термофильных бактерий (Li *et al. Nucleic Acids Res.* 44, e34-e34 (2016)), результаты секвенирования показали наличие системы CRISPR-Cas типа IIC в геноме *G. thermodenitrificans* T12 (фигура 15A). Предсказано, что эндонуклеаза Cas9 данной системы (ThermoCas9) является относительно небольшой (1082 аминокислоты) в сравнении с другими ортологами Cas9, такими как SpCas9 (1368 аминокислот). Размеры отличаются, главным образом, за счет укороченной доли REC, как было продемонстрировано для других ортологов Cas9 небольшого размера (фигура 19) (Ran *et al. Nature* 520, 186-191 (2015)). Кроме того, ожидалось, что ThermoCas9

будет активна по меньшей мере в районе температурного оптимума *G. thermodenitrificans* T12 (Daas et al. *Biotechnol. Biofuels* 9, 210 (2016)). Используя последовательность ThermoCas9 в качестве искомой последовательности, авторы изобретения провели BLAST-Р поиск в неизбыточной базе данных белковых последовательностей NCBI и обнаружили ряд высокоидентичных ортологов Cas9 (идентичность 87-99% на уровне белка, Таблица 1), главным образом, в роде *Geobacillus*, это свидетельствует в пользу того, что ThermoCas9 является частью высококонсервативной защитной системы термофильных бактерий (фигура 15B). Эти характеристики указывают на то, что она может быть потенциальным кандидатом для использования в качестве инструмента редактирования генома и выключения генов для термофильных микроорганизмов, и в условиях, когда требуется повышенная надежность белка.

Авторы изобретения сначала выполнили *in silico* предсказание модулей crРНК и tracrРНК системы CRISPR-Cas *G. thermodenitrificans* T12 с использованием ранее описанного подхода (Mougiakos et al. *Trends Biotechnol.* 34, 575-587 (2016); Ran et al. *Nature* 520, 186-191 (2015)). На основании этого предсказания была сконструирована 190-нт химерная еРНК путем связывания предсказанных полноразмерных crРНК (30-нт спейсер, за которым следует 36-нт повтор) и tracrРНК (36-нт анти-повтор, за которым следует 88-нт последовательность с тремя предсказанными шпилечными структурами). ThermoCas9 была гетерологично экспрессирована в *E. coli* и очищена до гомогенности. Следуя гипотезе о том, что нагружение еРНК на ThermoCas9 может стабилизировать белок, авторы изобретения инкубировали очищенную апо-ThermoCas9 и ThermoCas9, нагруженную *in vitro* транскрибированной еРНК, при 60°C и 65°C в течение 15 и 30 мин. SDS-ПААГ анализ показал, что очищенная ThermoCas9 денатурирует при 65°C, но не при 60°C, в то время как температура денатурации комплекса ThermoCas9-еРНК превышает 65°C (фигура 15C). Продемонстрированная термостабильность ThermoCas9 подразумевает ее потенциал в качестве термотолерантного инструмента CRISPR-Cas9 для редактирования генома, что побудило авторов изобретения

проанализировать некоторые соответствующие молекулярные особенности более подробно.

Таблица 1. Результаты анализа pBLAST последовательностей белков Cas9 с фигуры 1В в сравнении с ThermoCas9.

Вид	% идентичности ^a
<i>Geobacillus</i> 47C-IIb	99
<i>Geobacillus</i> 46C-IIa	89
<i>Geobacillus</i> LC300	89
<i>Geobacillus jurassicus</i>	89
<i>Geobacillus</i> MAS1	88
<i>Geobacillus stearothermophilus</i>	88
<i>Geobacillus stearothermophilus</i> ATCC 12980	88
<i>Geobacillus</i> Sah69	88
<i>Geobacillus stearothermophilus</i>	88
<i>Geobacillus kaustophilus</i>	88
<i>Geobacillus stearothermophilus</i>	88
<i>Geobacillus</i> genomosp. 3	87
<i>Geobacillus</i> genomosp. 3	87
<i>Geobacillus subterraneus</i>	87
<i>Effusibacillus pohliae</i>	86

Пример 11: Определение PAM для ThermoCas9

Первым шагом в направлении характеристики ThermoCas9 было *in silico* предсказание ее предпочтений в отношении PAM для успешного расщепления ДНК-мишени. Авторы изобретения использовали 10 спейсеров локуса CRISPR *G. thermodenitrificans* T12 для поиска потенциальных протоспейсеров в вирусных и плазмидных последовательностях с использованием CRISPRtarget (Biswas *et al.* *PNK Biol.* 10, 817-827 (2013)). Поскольку лишь два хита были получены для фаговых геномов (фигура 20А), было решено продолжать использовать *in vitro* подход к определению PAM. Авторы изобретения *in vitro* транскрибировали предсказанную последовательность егРНК, которая содержала спейсер для основанного на ThermoCas9 нацеливания на линейные дцДНК субстраты с сочетающимся протоспейсером. Протоспейсер был

фланкирован на его 3'-конце случайными последовательностями длиной 7 пар нуклеотидов (п.н.). После проведения анализов расщепления при помощи ThermoCas9 при 55°C расщепленные представители библиотеки (вместе с не являющимся мишенью образцом библиотеки в качестве контроля) были подвергнуты глубокому секвенированию и сравнению с целью выявления предпочтений ThermoCas9 в отношении PAM (фигура 16A). Результаты секвенирования показали, что ThermoCas9 вносит двухцепочечные разрывы в ДНК, которые, по аналогии с мезофильными вариантами Cas9, расположены в основном между 3-м и 4-м нуклеотидами, проксимальными к последовательности PAM. Более того, расщепленные последовательности показали, что ThermoCas9 узнает PAM 5'-NNNNCNR-3', со слабым предпочтением в отношении цитозина в 1-м, 3-м, 4-м и 6-м положениях PAM (фигура 16B). Недавние исследования выявили важность 8-го положения PAM для узнавания мишени некоторыми ортологами Cas9 типа IIC (Karvelis *et al. Genome Biol.* 16, 253 (2015); Kim *et al. Genome Res.* 24, 1012-9 (2014)). С учетом этого, и принимая во внимание результаты *in silico* предсказания PAM для ThermoCas9, авторы изобретения провели дополнительные анализы определения PAM. Результаты показали оптимальную эффективность нацеливания в случае присутствия аденина в 8-м положении PAM (фигура 16C). Интересно, что, невзирая на ограниченное число хитов, вышеупомянутое *in silico* предсказание PAM (фигура 20B) также указывало на важность цитозина в 5-м и аденина в 8-м положениях PAM.

Для дальнейшего прояснения неопределенности в отношении 6-го и 7-го положений PAM авторы изобретения создали набор из 16 разных фрагментов ДНК-мишени, в которых сочетающийся протоспейсер был фланкирован PAM 5'-CCCCNNA-3' [SEQ ID NO: 13]. Проводили анализы расщепления этих фрагментов (каждый с уникальным сочетанием 6-го и 7-го нуклеотидов), в которых разные компоненты (ThermoCas9, направляющая егРНК, дцДНК-мишень) были предварительно нагреты отдельно при разных температурах (20, 30, 37, 45, 55 и 60°C) в течение 10 мин до их объединения и инкубации в течение 1 часа при соответствующей температуре анализа. Когда

анализы проводили при температуре от 37°C до 60°C, все из разных ДНК-субстратов были расщеплены (фигура 16D, фигура 21). Однако наиболее расщепленные фрагменты-мишени имели последовательности PAM (с 5-го по 8-е положения PAM) 5'-CNAА-3' и 5'-СМСА-3', в то время как наименее расщепленные мишени содержали PAM 5'-САКА-3'. При 30°C наблюдали расщепление лишь ДНК-субстратов с оптимальными последовательностями PAM (с 5-го по 8-е положения PAM) 5'-CNAА-3' и 5'-СМСА-3' (фигура 2D). И наконец, при 20°C мишенями служили лишь ДНК-субстраты с последовательностями PAM (с 5-го по 8-е положения PAM) 5'-CVAA-3' и 5'-СССА (фигура 21), что делает эти последовательности наиболее предпочтительными последовательностями PAM. Эти результаты демонстрируют, что на нижней границе температуры ThermoCas9 расщепляет лишь фрагменты с предпочтительной последовательностью PAM. Эту характеристику можно использовать при процессах редактирования *in vivo*, например, чтобы избежать нецелевых эффектов.

Пример 12: Термостабильность и укорочения

Предсказанная *tracr*РНК состоит из области анти-повтора, за которой следуют три шпилечные структуры (фигура 17A). Использование *tracr*РНК наряду с *cr*РНК для образования химерной *eg*РНК привело к успешному направленному расщеплению ДНК-субстрата. Было обнаружено, что делеция длиной 41 нт на дистальном конце спейсера в полноразмерной шпильке повтора-анти-повтора (фигура 17A), что, скорее всего, больше напоминает естественное состояние двойной структуры гида, оказывает небольшое, или не оказывает, влияние на эффективность расщепления ДНК. Эффект дальнейшего укорочения предсказанных шпилек (фигура 17A) на эффективность расщепления ThermoCas9 оценивали путем проведения серии расщеплений, в которых все компоненты (*eg*РНК, ThermoCas9, ДНК-субстрат) предварительно нагревали отдельно при разных температурах (37-65°C) в течение 1, 2 и 5 мин перед их объединением и инкубацией в течение 1 часа при разных температурах анализа (37-65°C). Судя по всему, число предсказанных структур стебель-петля в каркасе *tracr*РНК играет

очень важную роль в расщеплении ДНК; в случае наличия всех трех петель эффективность расщепления была наивысшей при всех протестированных температурах, при этом эффективность снижалась при удалении 3'-шпильки (фигура 17В). Кроме того, эффективность расщепления резко падала при удалении как средней шпильки, так и 3'-шпильки (фигура 22). При том, что предварительное нагревание ThermoCas9 при 65°C в течение 1 или 2 мин приводило к заметному расщеплению, активность расщепления была ликвидирована после 5-мин инкубации. Анализ термостабильности показал, что варианты eгРНК без 3'-структуры стебель-петля имеют сниженную стабильность белка ThermoCas9 при 65°C, это указывает на то, что полноразмерная tracrРНК необходима для оптимального расщепления ДНК ThermoCas9 при повышенных температурах. Кроме того, авторы изобретения также варьировали длины спейсерной последовательности (от 25 до 18 нт) и обнаружили, что при длине спейсера 23, 21, 20 и 19 нт мишени расщепляются с наибольшей эффективностью. Эффективность расщепления значительно падает при использовании спейсера длиной 18 нт.

In vivo РНП комплекс ThermoCas9:eгРНК, вероятно, образуется в течение нескольких минут. В совокупности с вышеприведенными результатами это мотивировало авторов изобретения на оценку активности и термостабильности РНП. Предварительно сформированный РНП комплекс нагревали при 60, 65 и 70°C в течение 5 и 10 мин перед добавлением предварительно нагретой ДНК и последующей инкубацией в течение 1 часа при 60, 65 и 70°C. Поразительно, но авторы изобретения наблюдали, что РНП с ThermoCas9 был активным при температуре вплоть до 70°C, несмотря на его предварительное нагревание в течение 5 мин при 70°C (фигура 17С). Этот результат подтвердил предположение авторов изобретения о том, что стабильность ThermoCas9 сильно коррелирует со связыванием соответствующей направляющей eгРНК (Ma et al., *Mol. Cell* 60, 398-407 (2015)).

В некоторых вариантах применения было бы полезно, если бы ThermoCas9 имела широкий температурный диапазон активности, то

есть, была бы функциональной как при низких, так и при высоких, температурах. Кроме того, в некоторых обстоятельствах было бы полезно, если бы активность ThermoCas9 можно было ограничить более узкими диапазонами температур, например, иметь активность только при низких или только при высоких температурах. Следовательно, возможность изменять диапазон температур, при котором ThermoCas9 способна к направленному расщеплению или связыванию, или при котором направленное расщепление или связывание происходит эффективно, за счет изменения структурных особенностей ThermoCas9 или связанных элементов (таких как eгРНК), позволило бы иметь более высокий уровень контроля над манипуляциями с последовательностью нуклеиновой кислоты. Вследствие этого, авторы изобретения провели сравнение температурного диапазона ThermoCas9 с таковым для Cas9 *Streptococcus pyogenes* (SpCas9). Для обоих гомологов Cas9 проводили *in vitro* анализы активности при температуре от 20 до 65°C. Оба белка инкубировали в течение 5 мин при соответствующей температуре анализа перед добавлением молекул eгРНК и ДНК-мишени. В соответствии с результатами предыдущего анализа²⁶, мезофильная SpCas9 была активна лишь при температуре от 25 до 44°C (фигура 17D); выше этой температуры активность SpCas9 быстро снижалась до неподдающихся обнаружению уровней. Напротив, активность расщепления ThermoCas9 можно было обнаружить при температуре от 25 до 65°C (фигура 17D). Это указывает на возможность использования ThermoCas9 в качестве инструмента редактирования генома как для термофильных, так и мезофильных организмов.

Ранее охарактеризованные мезофильные эндонуклеазы Cas9 используют двухвалентные катионы в качестве катализаторов образования ДЦР в ДНК-мишени (Jinek *et al.* *Science* 337, 816–821 (2012); Chen *et al.* *J. Biol. Chem.* 289, 13284–13294 (2014)). Для определения того, какие катионы участвуют в расщеплении ДНК за счет ThermoCas9, проводили анализы расщепления плазмид в присутствии одного из следующих двухвалентных катионов: Mg²⁺, Ca²⁺, Mn²⁺, Co²⁺, Ni²⁺ и Cu²⁺; анализ с хелатирующим катионы агентом

ЭДТА включали в качестве отрицательного контроля. Как и ожидалось, дцДНК-мишень была расщеплена в присутствии двухвалентных катионов и оставалась интактной в присутствии ЭДТА (фигура 23А). На основании информации о том, что некоторые системы типа IIC эффективно расщепляют одноцепочечную ДНК (Ma *et al. Mol. Cell* 60, 398-407 (2015); Zhang *et al. Mol. Cell* 60, 242-255 (2015)), авторы изобретения протестировали активность ThermoCas9 в отношении оцДНК-субстратов. Однако расщепление отсутствовало, это свидетельствовало о том, что ThermoCas9 представляет собой дцДНК-нуклеазу (фигура 23В).

Пример 13: Удаление гена в термофиле *B. Smithii* при помощи ThermoCas9

Авторы изобретения решили разработать инструмент на основе ThermoCas9 для редактирования генома термофильных бактерий. В настоящем документе авторы изобретения приводят подтверждение концепции с использованием *Bacillus smithii* ET 138, культивируемой при 55°C. С целью использования минимума генетических частей, авторы изобретения использовали подход с одной плазмидой. Они сконструировали набор плазмид pThermoCas9 на основе pNW33n, содержащих ген *thermocas9* под контролем естественного промотора *xytL* (P_{xytL}), матрицу гомологичной рекомбинации для репарации индуцированных Cas9 двухцепочечных разрывов ДНК в интересующем гене, и экспрессионный модуль eгРНК под контролем конститутивного промотора *pta* (P_{pta}) из *Bacillus coagulans* (фигура 4А).

Первой целью было удаление полноразмерного гена *pyrF* из генома *B. smithii* ET 138. Производные от pNW33n плазмиды pThermoCas9_bs Δ pyrF1 и pThermoCas9_bs Δ pyrF2 использовали для экспрессии разных гид-РНК для ThermoCas9 со спейсерами, направленными на разные сайты гена *pyrF*, при этом третья плазида (pThermoCas9_ctrl) содержала случайный не направляющий спейсер в экспрессионном модуле eгРНК. Трансформация компетентных клеток *B. smithii* ET 138 при 55°C контрольными плазмидами pNW33n (без гид-РНК) и pThermoCas9_ctrl привела к образованию ~200 колоний в каждом случае. Из 10 подвергнутых

скринингу колоний pThermoCas9_ctrl, ни одна не имела генотип $\Delta pyrF$, что подтверждало результаты предыдущих исследований в том, что гомологичная рекомбинация в *B. smithii* ET 138 является не достаточной для получения чистых мутантов (Mougiakos *et al.* *ACS Synth. Biol.* 6, 849–861 (2017); Vosma *et al.* *Microb. Cell Fact.* 14, 99 (2015)). Напротив, трансформация плазмидами pThermoCas9_bs $\Delta pyrF1$ и pThermoCas9_bs $\Delta pyrF2$ привела к образованию 20 и 0 колоний, соответственно, что свидетельствовало о *in vivo* активности ThermoCas9 при 55°C и подтверждало вышеописанный широкий температурный диапазон белка *in vitro*. Из десяти подвергнутых скринингу колоний pThermoCas9_ $\Delta pyrF1$, одна представляла собой чистый мутант $\Delta pyrF$, в то время как остальные имели смешанный генотип дикого типа/ $\Delta pyrF$ (фигура 4B), что подтверждало применимость системы, поскольку спроектированная управляемая гомологией репарация гена-мишени *pyrF* была успешной. Тем не менее, в строго контролируемой основанной на SpCas9 системе контрселекции, которую ранее разработали авторы изобретения, эффективность удаления гена *pyrF* была выше (Olson *et al.*, *Curr. Opin. Biotechnol.* 33, 130–141 (2015)). Низкое число полученных трансформантов и чистых мутантов при использовании инструмента на основе ThermoCas9 может быть объяснено низкой эффективностью гомологичной рекомбинации в клетках *B. smithii* (Olson *et al.*, *Curr. Opin. Biotechnol.* 33, 130–141 (2015)) в сочетании с конститутивной экспрессией высокоактивной ThermoCas9. Ожидается, что использование строго контролируемого промотора приведет к повышению эффективности.

Пример 14: Удаление гена в мезофиле *Pseudomonas putida* при помощи ThermoCas9

Для расширения сферы применения инструмента редактирования генома на основе ThermoCas9 и для оценки того, могут ли результаты, полученные *in vitro*, быть подтверждены *in vivo*, ее активность в мезофильной грамотрицательной бактерии *P. putida* KT2440 оценивали путем комбинирования гомологичной рекомбинации и основанной на ThermoCas9 контрселекции. До настоящего времени

для данного организма не сообщалось о генно-инженерном инструменте на основе Cas9. Авторы изобретения вновь использовали подход с одной плазмидой. Они сконструировали плазмиду pThermoCas9_pp Δ pyrF на основе pEMG, содержащую ген *thermocas9* под контролем индуцируемого 3-метилбензоатом промотора Pm, матрицу гомологичной рекомбинации для удаления гена *pyrF* и экспрессионный модуль eгРНК под контролем конститутивного промотора P3. После трансформации клеток *P. putida* KT2440 и подтверждения методом ПЦР интеграции плазмиды колонию инокулировали в селективную жидкую среду и культивировали в течение ночи при 37°C. Ночную культуру использовали для инокуляции селективной среды, и экспрессию ThermoCas9 индуцировали 3-метилбензоатом. Затем разведенную культуру высевали на не селективную среду, содержащую 3-метилбензоат. Для сравнения проводили параллельный эксперимент без индукции экспрессии ThermoCas9 3-метилбензоатом. Процесс привел к получению 76 колоний в случае индуцированной культуры и 52 колоний в случае не индуцированной культуры. В индуцированной культуре 38 колоний (50%) имели чистый делеционный генотип и 6 колоний имели смешанный дикого типа/делеционный генотип. Напротив, лишь 1 колония (2%) в не индуцированной культуре имела делеционный генотип, и не были получены колонии со смешанным дикого типа/делеционным генотипом (фигура 24). Эти результаты показывают, что ThermoCas9 может быть использована в качестве эффективного инструмента контрселекции для мезофила *P. putida* KT2440, выращиваемого при 37°C.

Пример 15: Выключение гена при помощи ThermoCas9

Эффективный термоактивный инструмент выключения транскрипции CRISPRi в настоящее время отсутствует. Такая система могла бы быть полезной для целого ряда вариантов применения. Например, такая система могла бы сильно облегчить исследование метаболизма термофилов. Каталитически «мертвый» вариант ThermoCas9 мог бы служить для этой цели за счет стабильного связывания с элементами ДНК без внесения разрывов в дцДНК. С этой целью авторы изобретения идентифицировали

каталитические домены RuvC и HNH в ThermoCas9 и внесли соответствующие мутации D8A и H582A для создания «мертвой» (d)ThermoCas9. После подтверждения спроектированной последовательности Thermo-dCas9 гетерологично продуцировали, очищали и использовали для *in vitro* анализа расщепления с той же ДНК-мишенью, которую использовали в вышеописанных анализах ThermoCas9; расщепление отсутствовало, что подтверждало каталитическую инактивацию нуклеазы.

На пути к разработке инструмента CRISPRi на основе Thermo-dCas9 авторы изобретения поставили цель выключения транскрипции экспрессируемого на высоком уровне гена *ldhL* из генома *B. smithii* ET138. Они сконструировали векторы pThermoCas9i_*ldhL* и pThermoCas9i_ctrl на основе pNW33n. Оба вектора содержали ген *thermo-dCas9* под контролем промотора P_{xy1L} и экспрессионный модуль eгРНК под контролем конститутивного промотора P_{pta} (фигура 4С). Плазмида pThermoCas9i_*ldhL* содержала спейсер для направления на нематричную цепь ДНК на 5'-конце гена *ldhL* в *B. smithii* ET 138 (фигура S7). Выбор положения и цепи-мишени был основан на результатах предыдущих исследований (Bikard *et al. Nucleic Acids Res.* 41, 7429-7437 (2013); Larson *et al. Nat. Protoc.* 8, 2180-2196 (2013)) с целью эффективной понижающей регуляции гена *ldhL*. Плазмида pThermoCas9i_ctrl содержала случайный не направляющий спейсер в экспрессионном модуле eгРНК. Конструкты использовали для трансформации компетентных клеток *B. smithii* ET 138 при 55°C, с последующим высеванием клеток на чашки с LB2 агаром, что привело к получению равных количеств колоний. Две из примерно 700 колоний на конструкт были выбраны для культивирования в микроаэробных условиях продуцирования лактата в течение 24 часов, как описано ранее (Bosma *et al. Appl. Environ. Microbiol.* 81, 1874-1883 (2015)). Рост культур pThermoCas9i_*ldhL* был снижен на 50% в сравнении с ростом культур pThermoCas9i_ctrl (фигура 4E). Авторы изобретения ранее показали, что делеция гена *ldhL* приводит к сильной задержке роста *B. smithii* ET 138 вследствие отсутствия основанной на Ldh НАД⁺-регенерирующей способности в микроаэробных условиях (Bosma *et al. Microb. Cell Fact.* 14, 99

(2015)). Таким образом, наблюдаемое снижение роста, по всей вероятности, было вызвано ингибированием транскрипции гена *ldhL* и последующим окислительно-восстановительным дисбалансом вследствие утраты НАД⁺-регенерирующей способности. Действительно, ВЭЖХ-анализ показал 40% снижение продуцирования лактата культурами с выключенным геном *ldhL*, и ОТ-кПЦР анализ показал, что уровни транскрипции гена *ldhL* были значительно снижены в культурах pThermoCas9i_ *ldhL* в сравнении с культурами pThermoCas9i_ctrl (фигура 4E).

Пример 16: Резюме

Большинство вариантов применения CRISPR-Cas основаны на направляемом РНК вмешательстве в ДНК белков системы CRISPR-Cas класса 2, таких как Cas9 и Cas12a (Komor *et al.*, *Cell* 168, 20-36 (2017); Puchta, *Curr. Opin. Plant Biol.* 36, 1-8 (2017); Xu *et al.* *J. Genet. Genomics* 42, 141-149 (2015); Tang *et al.* *Nat. Plants* 3, 17018 (2017); Zetsche *et al.* *Nat. Biotechnol.* 35, 31-34 (2016); Mougiakos *et al.*, *Trends Biotechnol.* 34, 575-587 (2016)). До настоящей работы иммунные системы CRISPR-Cas класса 2 не были выявлены и охарактеризованы в термофильных микроорганизмах, в противоположность представленным в большом количестве системам CRISPR-Cas класса 1, имеющимся у термофильных бактерий и архей (Makarova *et al.*, *Nat. Rev. Microbiol.* 13, 722-736 (2015); Weinberger *et al.*, *MBio* 3, e00456-12 (2012)), некоторые из которых были использованы для редактирования генома термофилов (Li *et al.* *Nucleic Acids Res.* 44, e34-e34 (2016)). В результате область применения технологий CRISPR-Cas была в основном ограничена температурами ниже 42°C вследствие мезофильной природы используемых эндонуклеаз Cas. Таким образом, это исключало применение данных технологий для облигатных термофилов и в экспериментальных подходах, требующих повышенных температур и/или повышенной стабильности белков.

Авторы изобретения охарактеризовали ThermoCas9, ортолог Cas9 из термофильной бактерии *G. thermodenitrificans* T12, штамма, ранее выделенного из компоста (Daas *et al.*, *Biotechnol. Biofuels* 9, 210 (2016)). Сбор информации позволил выявить в

геномах других термофилов дополнительные ортологи Cas9, почти идентичные ThermoCas9, и впервые было продемонстрировано, что системы CRISPR-Cas типа II действительно существуют у термофилов, по меньшей мере в некоторых ветвях родов *Bacillus* и *Geobacillus*. Авторы изобретения показали, что ThermoCas9 является активной *in vitro* в широком диапазоне температур 20-70°C, который гораздо шире, чем диапазон 25-44°C ее мезофильного ортолога SpCas9. Более широкие границы активности и стабильности ThermoCas9 позволяют использовать ее в молекулярно-биологических методах, для которых требуются манипуляции с ДНК при температурах 20-70°C, а также использовать ее в жестких условиях, требующих более надежной ферментативной активности. Более того, авторы изобретения выявили несколько факторов, которые важны для придания ThermoCas9 термостабильности. Во-первых, авторы изобретения продемонстрировали, что предпочтения ThermoCas9 в отношении PAM являются очень строгими для проявления активности в нижней части температурного диапазона ($\leq 30^\circ\text{C}$), в то время как большая вариабельность PAM допускается для активности при умеренных или оптимальных температурах (37-60°C). Во-вторых, авторы изобретения продемонстрировали, что активность и термостабильность ThermoCas9 сильно зависят от связывания с соответствующей направляющей егРНК. Без связи с конкретной теорией, авторы изобретения предполагают, что эта стабилизация мультидоменного белка Cas9, скорее всего, является результатом значительного конформационного изменения из открытого/гибкого состояния в более компактное состояние, как описано для SpCas9 при связывании гид-РНК (Jinek *et al. Science* 343, 1247997-1247997 (2014)).

На основании описанных в настоящем документе характеристик новой ThermoCas9 авторы изобретения успешно разработали инструменты для редактирования генома строго термофильных прокариот. Они показали, что ThermoCas9 является активной *in vivo* при 55°C и 37°C, и адаптировали современные генно-инженерные технологии на основе Cas9 для термофила *B. smithii* ET 138 и мезофила *P. putida* KT2440. Вследствие широкого температурного

диапазона ThermoCas9 ожидается, что простой, эффективный, основанный на одной плазмиде подход к использованию ThermoCas9 будет пригодным для широкого спектра термофильных и мезофильных микроорганизмов, которые могут расти при температурах от 37°C вплоть до 70°C. Это является дополнением к существующим технологиям для мезофильных организмов, позволяя использовать их для большой группы организмов, для которых до настоящего времени эти эффективные инструменты были недоступны.

Скрининг природных ресурсов на новые ферменты с желательными свойствами, безусловно, является ценным. Результаты предыдущих исследований указывали на то, что адаптация мезофильного ортолога Cas9 к более высоким температурам, с направленной эволюцией и белковой инженерией, будет наилучшим подходом к конструированию термофильного белка Cas9²⁹. Вместо этого, авторы изобретения выявили монофилетическую группу Cas9 в некоторых термофильных бактериях и превратили один из этих термостабильных вариантов ThermoCas9 в мощный инструмент редактирования генома как для термофильных, так и для мезофильных организмов. Благодаря данному исследованию авторы изобретения еще более расширили потенциал основанных на Cas9 технологий редактирования генома и открыли новые возможности для использования технологий на основе Cas9 в новых вариантах применения в жестких условиях или в условиях, в которых необходима активность в широком диапазоне температур.

Пример 17: Материалы и методы

а. Штаммы и условия роста бактерий

Умеренный термофил *B. smithii* ET 138 $\Delta sigF \Delta hsdR$ (Mougiakos, et al., (2017) ACS Synth. Biol. 6, 849-861) использовали в экспериментах с редактированием и выключением гена с помощью ThermoCas9. Его выращивали в среде LB2 (Bosma, et al. *Microb. Cell Fact.* 14, 99 (2015)) при 55°C. Во всех экспериментах для чашек использовали 30 г агара (Difco) на литр среды. При необходимости, добавляли хлорамфеникол в концентрации 7 мкг/мл. Для экспрессии белка клетки *E. coli* Rosetta (DE3) выращивали в среде LB в колбах при 37°C в шейкере-инкубаторе при

120 об/мин до достижения значения OD_{600nm} 0,5 после переключения температуры на 16°C. Через 30 мин индуцировали экспрессию добавлением изопропил-1-тио-β-d-галактопиранозидом (IPTG) до конечной концентрации 0,5 мМ, после чего инкубацию продолжали при 16°C. Для клонирования конструктов РАМ, для 6-го, 7-го и 8-го положений, DH5-альфа компетентные клетки *E. coli* (NEB) трансформировали в соответствии с руководством, предоставленным производителем, и выращивали в течение ночи на чашках с LB-агаром при 37°C. Для клонирования библиотеки выроденных 7-нт РАМ электрокомпетентные клетки DH10B *E. coli* трансформировали в соответствии со стандартными методами (Sambrook, Fritsch & Maniatis, T. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory, 1989) и выращивали на чашках с LB-агаром при 37°C в течение ночи. *E. coli* DH5α λpir (Invitrogen) использовали для конструирования плазмиды *P. putida* с использованием метода трансформации, описанного в публикации Ausubel et al. (*Current Protocols in Molecular Biology*. (John Wiley & Sons, Inc., 2001). doi:10.1002/0471142727). Для всех штаммов *E. coli*, при необходимости, использовали хлорамфеникол в концентрации 25 мг/л и канамицин в концентрации 50 мг/л. Штаммы *Pseudomonas putida* KT2440 (DSM 6125) культивировали при 37°C в среде LB, если не указано иначе. При необходимости, добавляли канамицин в концентрации 50 мг/л и 3-метилбензоат в концентрации 3 мМ.

в. Экспрессия и очистка ThermoCas9

ThermoCas9 ПЦР-амплифицировали из генома *G. thermodenitrificans* T12, затем клонировали, гетерологично экспрессировали в *E. coli* Rosetta (DE3) и очищали методом ЖХБР с сочетанием этапов Ni²⁺-аффинной, анионообменной и гель-фильтрационной хроматографии. Последовательность гена встраивали в плазмиду рML-1В (полученную из UC Berkeley MacroLab, Addgene #29653) безлигазным клонированием с использованием олигонуклеотидов (Таблица 2), с получением конструкта для экспрессии белка, кодирующего последовательность полипептида ThermoCas9 (остатки 1-1082), слитую с N-концевой меткой,

содержащей гексагистидиновую последовательность и сайт расщепления протеазы вируса гравировки табака (TEV). Для экспрессии каталитически неактивного белка ThermoCas9 (Thermo-dCas9), вносили точечные мутации D8A и H582A методом ПЦР и подтверждали секвенированием ДНК.

Таблица 2 | Олигонуклеотиды, использованные в данном исследовании

		Последовательность	Описание	SEQ ID
Конструирование библиотеки РАМ	BG6 494	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC NNNNNNNCTAGATCCTTTTAAATTA AAAATGAAGT TTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с 7-нт случайной последовательностью РАМ	59
	BG6 495	TATGCCGGATCCTCAGACCAAGTTTACTCATATAT ACTTTAGATTGATTTAAAACCTTCATTTTAAATTTAA AAGGATCTAG	ОБ для конструирования <i>in vitro</i> последовательностей-мишеней ДНК	60
	BG7 356	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-T-	Адаптер при отжиге с BG7357 лигируется к А-наращенным расщепленным ThermoCas9 фрагментам	61
	BG7 357	CTGTCTCTTATACACATCTGACGCTGCCGACGA	Адаптер при отжиге с BG7356 лигируется к А-наращенным расщепленным ThermoCas9 фрагментам	62
	BG7 358	TCGTCGGCAGCGTCAG	ПР адаптер секвенирования для ПЦР-амплификации расщепленных ThermoCas9 фрагментов	63
	BG7 359	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGG ACCATGATTACGCCAAGC	ОБ адаптер секвенирования для ПЦР-амплификации расщепленных ThermoCas9 фрагментов	64
	BG7 616	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGG GTCATGAGATTATCAAAAAGGATCTTC	ОБ адаптер секвенирования для ПЦР-амплификации контрольных фрагментов	65
	BG8 157	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCAGCTAGATCCTTTTAAATTA AAAATGAA GTTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с РАМ «CCCCCAG»	66
	BG8 158	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCCAAGCTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с РАМ «CCCCCCAA»	67

BG8 159	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCATCTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с РАМ «CCCCCCAT»	68
BG8 160	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCCACCTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с РАМ «CCCCCCAC»	69
BG8 161	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC NNNNTNNCTAGATCCTTTTAAATTA AAAATGAAGT TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с РАМ «NNNNTNN»	70
BG8 363	ACGGTTATCCACAGAATCAG	ПР для ПЦР-линеаризации библиотек для идентификации РАМ	71
BG8 364	CGGGATTGACTTTTAAAAAAGG	ОБ для ПЦР-линеаризации библиотек для идентификации РАМ	72
BG8 763	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCAAACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «АА»	73
BG8 764	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCATACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «АТ»	74
BG8 765	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCAGACTAGATCCTTTTAAATTA AAAATGAA GTTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «АГ»	75
BG8 766	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCACACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «АС»	76
BG8 767	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCTAACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ТА»	77

BG8 768	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCTTACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ТТ»	78
BG8 769	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCTGACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ТГ»	79
BG8 770	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCTCACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ТС»	80
BG8 771	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCGAACTAGATCCTTTTAAATTA AAAATGAA GTTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ГА»	81
BG8 772	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCGTACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ГТ»	82
BG8 773	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCGGACTAGATCCTTTTAAATTA AAAATGAA GTTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ГГ»	83
BG8 774	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCGCACTAGATCCTTTTAAATTA AAAATGAA GTTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «ГС»	84
BG8 775	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCAACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «СА»	85
BG8 776	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCTACTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «СТ»	86

	BG8 777	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCGACTAGATCCTTTTAAATTA AAAATGAA GTTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «СG»	87
	BG8 778	TATGCCTCATGAGATTATCAAAAAGGATCTTCAC CCCCCCTAGATCCTTTTAAATTA AAAATGAAG TTTTAAATCAATC	ПР для конструирования <i>in vitro</i> ДНК-мишени с положениями РАМ 6 и 7 «СC»	88
Модуль егРНК для <i>in vitro</i> транскрипции	BG6 574	AAGCTTGAAATAATACGACTCACTATAGG	ПР для ПЦР-амплификации матрицы егРНК для первого процесса идентификации РАМ (30-нт спейсер)	89
	BG6 576	AAAAAAGACCTTGACGTTTTCC	ПР для ПЦР-амплификации матрицы егРНК для первого процесса идентификации РАМ	90
	BG9 307	AAGCTTGAAATAATACGACTCACTATAGGTGAGAT TATCAAAAAGGATCTTCACGTC	ОБ для ПЦР-амплификации матрицы егРНК для всех процессов идентификации РАМ за исключением первого (25-нт спейсер)	91
	BG9 309	AAAACGCCTAAGAGTGGGGAATG	ОБ для ПЦР-амплификации матрицы егРНК с 3 шпильками для всех процессов идентификации РАМ за исключением первого	92
	BG9 310	AAAAGGCGATAGGCGATCC	ОБ для ПЦР-амплификации матрицы егРНК с 2 шпильками для всех процессов идентификации РАМ за исключением первого	93
	BG9 311	AAAACGGGTCAGTCTGCSTATAG	ОБ для ПЦР-амплификации матрицы егРНК с 1 шпилькой для всех процессов идентификации РАМ за исключением первого	94
	BG9 308	AAGCTTGAAATAATACGACTCACTATAGGTGAGAT TATCAAAAAGGATCTTCACGTC	рТ7 и 25-нт спейсер егРНК ПР	95

	BG1 0118	AAGCTTGAAATAATACGACTCACTATAGGAGATTA TCAAAAAGGATCTTCACGTCA	рТ7 и 24-нт спейсер егРНК ПР	96
	BG1 0119	AAGCTTGAAATAATACGACTCACTATAGGAAGATT ATCAAAAAGGATCTTCACGTCATAG	рТ7 и 23-нт спейсер егРНК ПР	97
	BG1 0120	AAGCTTGAAATAATACGACTCACTATAGGATTATC AAAAAGGATCTTCACGTCATAGT	рТ7 и 22-нт спейсер егРНК ПР	98
	BG1 0121	AAGCTTGAAATAATACGACTCACTATAGGAATTAT CAAAAAGGATCTTCACGTCATAGTT	рТ7 и 21-нт спейсер егРНК ПР	99
	BG1 0122	AAGCTTGAAATAATACGACTCACTATAGGTTATCA AAAAGGATCTTCACGTCATAGTT	рТ7 и 20-нт спейсер егРНК ПР	100
	BG1 0123	AAGCTTGAAATAATACGACTCACTATAGGTATCAA AAAGGATCTTCACGTCATAGTTC	рТ7 и 19-нт спейсер егРНК ПР	101
	BG1 0124	AAGCTTGAAATAATACGACTCACTATAGGATCAAA AAGGATCTTCACGTCATAGTTC	рТ7 и 18-нт спейсер егРНК ПР	102
	BG9 312	AAAACGCCTAAGAGTGGGGAATGCCCGAAGAAAG CGGGCGATAGGCGATCC	3 петли егРНК ОН ОБ	103
	BG8 191	AAGCTTGGCGTAATCATGGTC	Для конструирования плазмиды рThermoCas9_ctrl и рThermoCas9_bsΔpyrF1/2	104
	BG8 192	TCATGAGTTCCTCATGTTGTG	Для конструирования плазмиды рThermoCas9_ctrl и рThermoCas9_bsΔpyrF1/2	105
Конструкты для редактирования и выключения	BG8 194	tatggcgaatcacaacatgggaactcatgaGAACATCCTCTTTCTTA G	Для конструирования плазмиды рThermoCas9_ctrl и рThermoCas9_bsΔpyrF1/2	106
	BG8 195	gccgatatcaagaccgattttatacttcatTTAAGTTACCTCCTCGAT TG	Для конструирования плазмиды рThermoCas9_ctrl и рThermoCas9_bsΔpyrF1/2	107

BG8 196	ATGAAGTATAAAAATCGGTCTTG	Для конструирования плазмиды pThermoCas9_ctrl и pThermoCas9_bsΔpyrF1/2	108
BG8 197	TAACGGACGGATAGTTTC	Для конструирования плазмиды pThermoCas9_ctrl и pThermoCas9_bsΔpyrF1/2	109
BG8 198	gaaagccggggaaactatccgtccgttataAATCAGACAAAATGGC CTGCTTATG	Для конструирования плазмиды pThermoCas9_ctrl и pThermoCas9_bsΔpyrF1/2	110
BG8 263	gaactatgacactttatcttcagaatggacGTATAACGGTATCCATTT TAAGAATAATCC	Для конструирования плазмиды pThermoCas9_ctrl	111
BG8 268	accgttatacgtccattctgaaaaataaagtGTCATAGTCCCCTGAG AT	Для конструирования плазмиды pThermoCas9_ctrl	112
BG8 210	aacagctatgaccatgattacgccaaagcttCCCTCCCATGCACAATA G	Для конструирования плазмиды pThermoCas9_ctrl и pThermoCas9_bsΔpyrF1/2	113
BG8 261	gaactatgacatcatggagtttaaatccaGTATAACGGTATCCATT TTAAGAATAATCC	Для конструирования pThermoCas9_bsΔpyrF1	114
BG8 266	accgttatactggatttaaaactccatgatGTCATAGTCCCCTGAGA T	Для конструирования pThermoCas9_bsΔpyrF2	115
BG8 317	gaactatgaccaccagcttacatcaacaaGTATAACGGTATCCATT TTAAGAATAATCC	Для конструирования pThermoCas9_ΔbspyrF2	116
BG8 320	accgttatactgttgatgtaagctgggtgGTCATAGTCCCCTGAG AT	Для конструирования pThermoCas9_bsΔpyrF2	117
BG9 075	STATCGGCATTACGTCTATC	Для конструирования pThermoCas9i_ctrl	118
BG9 076	GCGTCGACTTCTGTATAGC	Для конструирования pThermoCas9i_ctrl	119
BG9 091	TGAAGTATAAAAATCGGTCTTGCTATCGGCATTACG TCTATC	Для конструирования pThermoCas9i_ctrl	120

BG9 092	CAAGCTTCGGCTGTATGGAATCACAGCGTCGACTT CTGTATAGC	Для конструирования pThermoCas9i_ctrl	121
BG9 077	GCTGTGATTCCATACAG	Для конструирования pThermoCas9i_ctrl	122
BG9 267	GGTGCAGTAGGTTGCAGCTATGCTTGTATAACGGT ATCCAT	Для конструирования pThermoCas9i_ctrl	123
BG9 263	AAGCATAGCTGCAACCTACTGCACCGTCATAGTTC CCCTGAGATTATCG	Для конструирования pThermoCas9i_ctrl	124
BG9 088	TCATGACCAAATCCCTTAACG	Для конструирования pThermoCas9i_ctrl	125
BG9 089	TTAAGGGATTTTGGTCATGAGAACATCCTCTTTCTT AG	Для конструирования pThermoCas9i_ctrl	126
BG9 090	GCAAGACCGATTTTATACTTCATTTAAG	Для конструирования pThermoCas9i_ctrl	127
BG9 548	GGATCCCATGACGCTAGTATCCAGCTGGGTCATAG TTCCCCTGAGATTATCG	Для конструирования pThermoCas9i_ldhL	128
BG9 601	TTCAATATTTTTTTTGAATAAAAAATACGATACAA TAAAAATGTCTAGAAAAAGATAAAAAATG	Для конструирования pThermoCas9i_ldhL	129
BG9 600	TTTTTTATTCAAAAAAATATTGAATTTTAAAAAT GATGGTGCTAGTATGAAG	Для конструирования pThermoCas9i_ldhL	130
BG9 549	CCAGCTGGATACTAGCGTCATGGGATCCGTATAAC GGTATCCATTTAAGAATAATCC	Для конструирования pThermoCas9i_ldhL	131
BG8 552	TCGGGGTTCGTTCCCTTG	ПР для проверки геномной делеции <i>purF</i> проверки нокаута	132
BG8 553	CTTACACAGCCAGTGACGGAAC	ОБ для проверки геномной делеции <i>purF</i> проверки нокаута	133

BG2 365	GCCGGCGTCCCGGAAAACGA	Для конструирования pThermoCas9_ppΔpyrF	134
BG2 366	GCAGGTCGGGTTCCTCGCATCCATGCCCCGAACT	Для конструирования pThermoCas9_ppΔpyrF	135
BG2 367	ggcttcggaatcgtttccgggacgccggcACGGCATTGGCAAGGC CAAG	Для конструирования pThermoCas9_ppΔpyrF	136
BG2 368	gacacaggcatcgggGCAGGGTCTCTTGGCAAGTC	Для конструирования pThermoCas9_ppΔpyrF	137
BG2 369	gccaagagaccctgCACCGATGCCTGTGTGCGAACC	Для конструирования pThermoCas9_ppΔpyrF	138
BG2 370	cttggcggaaaacgtcaaggctctttttacACGCGCATCAACTTCAA GGC	Для конструирования pThermoCas9_ppΔpyrF	139
BG2 371	atgacgagctgttcaccagcagcgcTATTATTGAAGCATTTATC AGGG	Для конструирования pThermoCas9_ppΔpyrF	140
BG2 372	GTAAAAAAGACCTTGACGTTTTTC	Для конструирования pThermoCas9_ppΔpyrF	141
BG2 373	tatgaagcgggcatTTGAAGACGAAAGGGCCTC	Для конструирования pThermoCas9_ppΔpyrF	142
BG2 374	taatagcgtgctggtgaacagctcGTCATAGTTCCCCTGAGAT TATCG	Для конструирования pThermoCas9_ppΔpyrF	143
BG2 375	tggagtcataaacatATGAAGTATAAAATCGGTCTTG	Для конструирования pThermoCas9_ppΔpyrF	144
BG2 376	cccttcgtctcAAATGGCCCGCTTCATAAGCAG	Для конструирования pThermoCas9_ppΔpyrF	145
BG2 377	gattttatacTTCATATGTTTCATGACTCCATTATTATTG	Для конструирования pThermoCas9_ppΔpyrF	146

	BG2 378	gggggcatggatgCGAGGAACCCGACCTGCATTGG	Для конструирования рThermoCas9_ppΔpyrF	147
	BG2 381	ACACGGCGGATGCACTTACC	ПР для подтверждения интеграции плазмиды и делеции <i>pyrF</i> в <i>P. putida</i>	148
	BG2 382	TGGACGTGTA CTTCGACAAC	ОБ для подтверждения делеции <i>pyrF</i> в <i>P. putida</i>	149
	BG2 135	ACACGGCGGATGCACTTACC	ОБ для подтверждения интеграции плазмиды в <i>P. putida</i>	150
Праймеры для секвенирования	BG8 196	TGGACGTGTA CTTCGACAAC	<i>thermocas9</i> послед. 1	151
	BG8 197	TAACGGACGGATAGTTTC	<i>thermocas9</i> послед. 2	152
	BG6 850	GCCTCATGAATGCAGCGATGGTCCGGTGTTC	<i>pyrF</i> , верхняя	153
	BG6 849	GCCTCATGAGTTC CATGTTGTGATTC	<i>pyrF</i> , нижняя	154
	BG6 769	CAATCCA ACTGGGCTTGAC	<i>thermocas9</i> послед. 3	155
	BG6 841	CAAGAACTTTATTGGTATAG	<i>thermocas9</i> послед. 4	156
	BG6 840	TTGCAGAAATGGTTGTCAAG	<i>thermocas9</i> послед. 5	157
	BG9 215	GAGATAATGCCGACTGTAC	Каркас рNW33n послед. 1	158

	BG9 216	AGGGCTCGCCTTTGGGAAG	Каркас pNW33n послед. 2	159
	BG9 505	GTTGCCAACGTTCTGAG	<i>thermocas9</i> послед. 6	160
	BG9 506	AATCCACGCCGTTTAG	<i>thermocas9</i> послед. 7	161
Анализы расщепления	BG8 363	ACGGTTATCCACAGAATCAG	ПР для ПЦР-линеаризации ДНК-мишени	162
	BG8 364	CGGGATTGACTTTTAAAAAAGG	ОБ для ПЦР-линеаризации ДНК-мишени	163
	BG9 302	AAACTTCATTTTTAATTTAAAAGGATCTAGAACCC CCCGTGAAGATCCTTTTGGATAATCTCATGACCAA AATCCCTTAACGTGAGTTTTTCGTTCCACTGAGCGTC AGACCCCGTAGAAA	Олигонуклеотид нематричной цепи для анализов расщепления оцДНК	164
	BG9 303	TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAAC TCACGTTAAGGGATTTTGGTCATGAGATTATCAAA AAGGATCTTCACCCCCCACTAGATCCTTTTAAA TTAAAAATGAAGTTT	Олигонуклеотид матричной цепи для анализов расщепления оцДНК	165
	BG9 304	TTTCTACGGGGTCTGACGCTCAGTGGAACGAAAAC TCACGTTAAGGGATTTTGGTCATGAGATTATCAAA AAGGATCTTCACGGGGGGTTCTAGATCCTTTTAAA TTAAAAATGAAGTTT	Олигонуклеотид матричной цепи для анализов расщепления оцДНК	166

Экспрессия ThermoCas9 и OT-кПЦР	BG7 886	TACTTCCAATCCAATGCAAAGTATAAAATCGGTCT TGATATCG	ПР БЛК_thermocas9	167
	BG7 887	TTATCCACTTCCAATGTTATTATAACGGACGGATA GTTTCCCCGGCTTTC	ОБ БЛК_thermocas9	168
Экспрессия ThermoCas9	BG9 665	ATGACGAAAGGAGTTTCTTATTATG	ОБ для кПЦР проверки <i>ldhl</i>	169
	BG9 666	AACGGTATCCGTGATTAAG	ПР для кПЦР проверки <i>ldhl</i>	170

Сайты рестрикции выделены курсивом. Последовательности РАМ выделены красным цветом. Области спейсера выделены жирным шрифтом. Нуклеотиды, выделенные строчным шрифтом, соответствуют «липким» концам праймера для HiFi сборки ДНК. БЛК: Безлигазное клонирование; ПР: прямой праймер; ОБ: обратный праймер.

Белки экспрессировали в клетках *E. coli* штамма Rosetta 2 (DE3). Культуры выращивали до OD_{600nm} 0,5-0,6. Экспрессию индуцировали добавлением IPTG до конечной концентрации 0,5 мМ, и инкубацию продолжали при 16°C в течение ночи. Клетки собирали центрифугированием, и клеточный осадок ресуспендировали в 20 мл буфера для лизиса (50 мМ фосфат натрия, pH 8, 500 мМ NaCl, 1 мМ ДТТ, 10 мМ имидазол) с добавлением ингибиторов протеаз (Roche cOmplete, без ЭДТА) и лизоцима. После гомогенизации клетки лизировали путем обработки ультразвуком (Sonoplus, Bandelin), используя ультразвуковой зонд MS72 Microtip Probe (Bandelin), в течение 5-8 минут, цикл состоял из 2-с импульса и 2,5-с паузы, с амплитудой 30%, а затем центрифугировали при 16000 x *g* в течение 1 часа при 4°C для удаления нерастворимого материала. Осветленный лизат фильтровали через 0,22-микронные фильтры (Mdi membrane technologies) и наносили на никелевую колонку (HisTrap HP, GE Lifesciences), промывали и затем элюировали 250 мМ имидазолом. Фракции, содержащие ThermoCas9, объединяли и диализовали в течение ночи в диализном буфере (250 мМ KCl, 20 мМ HEPES/КОН и 1 мМ ДТТ, pH 7,5). После диализа образец разбавляли 1:1 в 10 мМ HEPES/КОН, pH 8, и наносили на колонку с гепарин-сефарозой FF, предварительно уравновешенную в буфере IEX-A (150 мМ KCl, 20 мМ HEPES/КОН, pH 8). Колонку промывали IEX-A, а затем элюировали градиентом IEX-C (2 мМ KCl, 20 мМ HEPES/КОН, pH 8). Образец концентрировали до 700 мкл перед нанесением на колонку для гель-фильтрации (HiLoad 16/600 Superdex 200) методом ЖХБР (AKTA Pure). Фракции после гель-фильтрации анализировали методом SDS-ПААГ; фракции, содержащие ThermoCas9, объединяли, концентрировали до 200 мкл (50 мМ фосфат натрия, pH 8, 2 мМ ДТТ, 5% глицерина, 500 мМ NaCl) и либо использовали непосредственно для биохимических анализов, либо замораживали при -80°C для хранения.

c. *In vitro* синтез егРНК

Модуль егРНК конструировали, производя слияние предсказанных последовательностей crРНК и tracrРНК с линкером 5'-GAAA-3'. Экспрессирующую егРНК последовательность ДНК

помещали под транскрипционный контроль промотора T7. Он был синтезирован (Baseclear, Leiden, The Netherlands) и предоставлен на каркасе pUC57. Все егРНК, используемые в биохимических реакциях, были синтезированы с использованием набора HiScribe™ T7 High Yield RNA Synthesis Kit (NEB). ПЦР-фрагменты, кодирующие молекулы егРНК, с последовательностью T7 на 5'-конце, использовали в качестве матриц для *in vitro* реакции транскрипции. Транскрипцию T7 выполняли в течение 4 часов. Проводили электрофорез молекул егРНК в ПААГ с мочевиной, вырезали полосы из геля и очищали осаждением этанолом.

d. *In vitro* анализ расщепления

In vitro анализы расщепления проводили с очищенной рекомбинантной ThermoCas9. Белок ThermoCas9, *in vitro* транскрибированную егРНК и ДНК-субстраты (полученные методом ПЦР-амплификации с использованием праймеров, описанных в Таблице 2) инкубировали отдельно (если нет иных указаний) при указанной температуре в течение 10 мин, с последующим объединением компонентов и инкубацией их при разных температурах анализа в буфере для расщепления (100 мМ натрий-фосфатный буфер (рН 7), 500 мМ NaCl, 25 мМ MgCl₂, 25% (об/об) глицерина, 5 мМ дитиотреитол (ДТТ)) в течение 1 часа. Каждая реакционная смесь для расщепления содержала 160 нМ белок ThermoCas9, 4 нМ ДНК-субстрат и 150 нМ синтезированную егРНК. Реакции останавливали добавлением 6х красителя для нанесения на гель (NEB), и смеси подвергали электрофорезу в 1,5% агарозных гелях. Гели окрашивали красителем для ДНК SYBR safe DNA stain (Life Technologies) и визуализировали с помощью системы визуализации гелей Gel Doc™ EZ (Bio-rad).

e. Конструирование библиотеки для *in vitro* скрининга РАМ

Для конструирования библиотеки РАМ 122-п.н. фрагмент ДНК, содержащий протоспейсер и 7-п.н. вырожденную последовательность на его 3'-конце, конструировали путем отжига праймеров и удлинения с использованием фрагмента Кленова (экзо-) (NEB). Фрагмент библиотеки РАМ и вектор pNW33n расщепляли BspHI и BamHI (NEB), а затем лигировали (лигаза T4, NEB). Смесь для

лигирования вводили трансформацией в электрокомпетентные клетки *E. coli* DH10B, и из жидких культур выделяли плазмиды. Для определения 7-нт PAM плазмидную библиотеку линейаризовали при помощи SapI (NEB) и использовали в качестве мишени. Для остальных анализов ДНК-субстраты линейаризовали методом ПЦР-амплификации.

f. Скрининг-анализ PAM

Скрининг PAM для ThermoCas9 выполняли с использованием *in vitro* анализов расщепления, смесь для которых состояла из (на реакцию): 160 нМ ThermoCas9, 150 нМ *in vitro* транскрибированной егРНК, 4 нМ ДНК-мишени, 4 мкл буфера для расщепления (100 мМ натрий-фосфатный буфер, рН 7,5, 500 мМ NaCl, 5 мМ ДТТ, 25% глицерина) и MQ воды до конечного объема реакционной смеси 20 мкл. Содержащие PAM расщепленные фрагменты после реакции при 55°C очищали в геле, лигировали с адаптерами секвенирования Illumina и отправляли для секвенирования Illumina HiSeq 2500 (Baseclear). Эквимоллярные количества не обработанной ThermoCas9 библиотеки PAM подвергали тому же процессу и отправляли для секвенирования Illumina HiSeq 2500 в качестве эталона. Результаты HiSeq с идеальным совпадением последовательности с эталонной последовательностью, выбирали для дальнейшего анализа. Из выбранных образцов, те, которые встречались более 1000 раз в обработанной ThermoCas9 библиотеке и по меньшей мере в 10 раз больше в обработанной ThermoCas9 библиотеке, в сравнении с контрольной библиотекой, использовали для анализа WebLogo (Crooks *et al.*, *Genome Res.* 14, 1188-1190 (2004)).

g. Редактирующие и выключающие ген конструкты для *B. smithii* и *P. putida*

Все праймеры и плазмиды, используемые для конструирования плазмид, были спроектированы с соответствующими «липкими» концами для проведения NEBuilder HiFi сборки ДНК (NEB), и они приведены в Таблице 2 и 3, соответственно. Фрагменты для сборки плазмид были получены методом ПЦР с полимеразой Q5 (NEB) или с использованием Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher Scientific), ПЦР-продукты подвергали электрофорезу

в 1% агарозном геле и очищали с использованием набора Zymoogen gel DNA recovery kit (Zymo Research). Собранные плазмиды вводили трансформацией в химически компетентные клетки *E. coli* DH5 α (NEB) или в клетки *E. coli* DH5 α λ pir (Invitrogen) в случае конструкторов для *P. putida*, в последнем случае для облегчения прямой интеграции вектора. Одиночные колонии инокулировали в среду LB, материал плазмиды выделяли с использованием набора GeneJet plasmid miniprep kit (ThermoFisher Scientific), последовательность подтверждали (GATC-biotech), и 1 мкг каждого конструктора вводили трансформацией в электрокомпетентные клетки *B. smithii* ET 138, которые готовили в соответствии с описанным ранее протоколом (Bosma, et al. *Microb. Cell Fact.* 14, 99 (2015)). Набор MasterPure™ Gram Positive DNA Purification Kit (Epicentre) использовали для выделения геномной ДНК из жидких культур *B. smithii* и *P. putida*.

Для конструирования векторов pThermoCas9_ctrl, pThermoCas9_bs Δ pyrF1 и pThermoCas9_bs Δ pyrF2 каркас pNW33n совместно с фланкирующими областями Δ pyrF для гомологичной рекомбинации ПЦР-амплифицировали из вектора pWUR_Cas9sp1_hr (Mougiakos, et al. *ACS Synth. Biol.* 6, 849-861 (2017)) (BG8191and BG8192). Естественный промотор P_{xy1A} был ПЦР-амплифицирован из генома *B. smithii* ET 138 (BG8194 и BG8195). Ген *thermocas9* был ПЦР-амплифицирован из генома *G. thermodenitrificans* T12 (BG8196 и BG8197). Промотор P_{pta} был ПЦР-амплифицирован из вектора pWUR_Cas9sp1_hr (Mougiakos, et al. *ACS Synth. Biol.* 6, 849-861 (2017)) (BG8198 и BG8261_2/BG8263_nc2/BG8317_3). Спейсеры со следующим далее каркасом eгРНК были ПЦР-амплифицированы из вектора pUC57_T7t12sgRNA (BG8266_2/BG8268_nc2/8320_3 и BG8210).

Для конструирования векторов pThermoCas9i_ldhL проектировали и осуществляли сборку четырех фрагментов. Сначала вносили направленные точечные мутации в кодоны каталитических остатков *thermocas9* (мутации D8A и H582A) двухэтапным методом ПЦР с использованием pThermoCas9_ctrl в качестве матрицы. Во время первого этапа ПЦР (BG9075, BG9076) желаемые мутации

вводили на концах полученного ПЦР-фрагмента, и во время второго этапа (BG9091, BG9092) полученный фрагмент использовали в качестве ПЦР-матрицы для введения соответствующих «липких» концов для сборки. Часть *thermocas9* ниже второй мутации, наряду со спейсером для выключения *ldhL*, ПЦР-амплифицировали с использованием pThermoCas9_ctrl в качестве матрицы (BG9077 и BG9267). Каркас eГНК совместно с каркасом pNW33n ПЦР-амплифицировали с использованием pThermoCas9_ctrl в качестве матрицы (BG9263 и BG9088). Промотор совместно с частью *thermocas9* выше первой мутации ПЦР-амплифицировали с использованием pThermoCas9_ctrl в качестве матрицы (BG9089, BG9090).

Для конструирования вектора pThermoCas9i_ctrl проектировали и осуществляли сборку двух фрагментов. Последовательность спейсера в векторе pThermoCas9i_ldhL заменяли случайной последовательностью, содержащей сайты рестрикции BaeI на обоих концах. Каркас eГНК совместно с каркасом pNW33n ПЦР-амплифицировали с использованием pThermoCas9_ctrl в качестве матрицы (BG9548, BG9601). Другую половину конструкта, состоящую из *Thermo-dCas9* и промотора, амплифицировали с использованием pThermoCas9i_ldhL в качестве матрицы (BG9600, BG9549).

Для конструирования вектора pThermoCas9_ppΔpyrF для *P. putida* KT2440 проектировали и осуществляли сборку пяти фрагментов. Репликон из «суицидного» вектора pEMG ПЦР-амплифицировали (BG2365, BG2366). Фланкирующие области *pyrF* амплифицировали из геномной ДНК KT2440 (BG2367, BG2368 для 576-п.н. верхней фланкирующей последовательности и BG2369, BG2370 для 540-п.н. нижней фланкирующей последовательности). Проводили слияние фланкирующих последовательностей методом ПЦР с перекрывающимися праймерами с использованием праймеров BG2367 и BG2370, используя перекрывание праймеров BG2368 и BG2369. eГНК амплифицировали из плазмиды pThermoCas9_ctrl (BG2371, BG2372). Конститутивный промотор P3 амплифицировали из pSW_I-SceI (BG2373, BG2374). Проводили слияние этого фрагмента промотора с фрагментом eГНК методом ПЦР с перекрывающимися праймерами с

использованием праймеров BG2372 и BG2373, используя перекрытие праймеров BG2371 и BG2374. ThermoCas9 амплифицировали из плазмиды pThermoCas9_ctrl (BG2375, BG2376). Индуцируемую систему Pm-XylS, используемую для индукции ThermoCas9 3-метилбензоатом, амплифицировали из pSW_I-SceI (BG2377, BG2378).

Таблица 3. Плазмиды, используемые в настоящем исследовании

Плаزمида	Описание	Используемые сайты рестрикции	Праймеры	Источник
pNW33n	Челночный вектор, клонирующий вектор <i>E. coli-Bacillus</i> , Cam ^R	-	-	BGSC
pUC57_T7sgRNAfull	Вектор pUC57, содержащий ДНК, кодирующую egРНК под контролем промотора T7; служит в качестве матрицы для <i>in vitro</i> транскрипции полноразмерных egРНК с повтором/анти-повтором			Baseclear
pMA2_T7sgRNAtuncated R/AR	Вектор, содержащий ДНК, кодирующую egРНК с укороченной частью повтора/анти-повтора под контролем промотора T7; служит в качестве матрицы для <i>in vitro</i> транскрипции укороченных в части повтора/анти-повтора egРНК	-	-	Gen9
pRARE	Экспрессионный вектор на основе РНК-полимеразы T7, Kan ^R	-	-	EMD Millipore
pML-1B	Плазмида <i>E. coli Rosetta</i> TM (DE3), кодирует редкие tРНК, Cam ^R	-	-	Macrolab, Addgene
pEMG	Суицидный вектор <i>P. putida</i> , использован в качестве матрицы для репликона и Kan ^R		Смотри Таблицу 2	1
pSW_I-SceI	Вектор <i>P. putida</i> , содержащий <i>I-SceI</i> , использован в качестве матрицы для <i>xylS</i> и P _{pm}		Смотри Таблицу 2	1
pWUR_Cas9sp1_hr	pNW33n с spCas9-модулем, содержащая спейсер, направленный на ген <i>pyrF</i> . Эта плазмида была использована в качестве матрицы для создания конструкторов на основе ThermoCas9	-	-	2
pThermo_Cas9	<i>thermocas9</i> с N-конц. His-меткой и сайтом расщепления TEV в pML-1B. Экспрессионный вектор для ThermoCas9	SspI и безлигазное клонирование	BG7886 и BG7887	Настоящее исследование
pThermo_dCas9	<i>cas9dthermocas9</i> с N-конц. His-меткой и сайтом расщепления TEV в pML-1B. Экспрессионный вектор для каталитически неактивной (мертвой) dThermoCas9	SspI и безлигазное клонирование	BG7886 и BG7888	Настоящее исследование

pNW-PAM7nt	Последовательность-мишень в векторе pNW33n, содержащая 7-нт вырожденную PAM для <i>in vitro</i> анализа определения PAM	BamHI BspHI	и	Смотри Таблицу 2	Настоящее исследование
pNW63-pNW78	Последовательность-мишень в векторе pNW33n, содержащая различные нуклеотиды в 6-м и 7-м положениях PAM (CCCCC NNA)	BamHI BspHI	и	Смотри Таблицу 2	Настоящее исследование
pThermoCas9_ctrl	pNW33n с модулем ThermoCas9 ¹ , содержащая не направляющий спейсер. Использована в качестве отрицательного контроля	-	-	Смотри Таблицу 2	Настоящее исследование
pThermoCas9_bsΔpyrF1	pNW33n с модулем ThermoCas9 ¹ , содержащая спейсер 1, направляющий на ген <i>pyrF</i> , и слитые в+н фланкирующие последовательности <i>pyrF</i>	-	-	Смотри Таблицу 2	Настоящее исследование
pThermoCas9_bsΔpyrF2	pNW33n с модулем ThermoCas9 ¹ , содержащая спейсер 2, направляющий на ген <i>pyrF</i> , и слитые в+н фланкирующие последовательности <i>pyrF</i>	-	-	Смотри Таблицу 2	Настоящее исследование
pThermoCas9i_ctrl	pNW33n с модулем Thermo-dCas9 ² , содержащая не направляющий спейсер. Использована в качестве контроля дикого типа	-	-	Смотри Таблицу 2	Настоящее исследование
pThermoCas9i_ldhL	pNW33n с модулем Thermo-dCas9 ² , содержащая спейсер 2, направляющий на ген <i>ldhL</i>	-	-	Смотри Таблицу 2	Настоящее исследование
pThermoCas9_ppΔpyrF	pEMG с модулем ThermoCas9 ³ для <i>Pseudomonas putida</i> , содержащая спейсер, направляющий на ген <i>pyrF</i> , и слитые в+н фланкирующие последовательности <i>pyrF</i>	-	-	Смотри Таблицу 2	Настоящее исследование

¹ Модуль ThermoCas9 содержит *thermocas9* под естественным промотором P_{xylL}, за которым следует eгРНК под контролем промотора P_{pta} *B. coagulans* (фигура 4).

² Подобно модулю ThermoCas9, но с *thermo-dCas9* вместо *thermocas9* (фигура 4).

³ Модуль ThermoCas9 для *Pseudomonas putida* содержит *thermocas9* под транскрипционным контролем индуцируемой системы Pm-XylS, за которым следует eгРНК под контролем конститутивного промотора P3.

h. Протокол редактирования генома для *P. putida*

Трансформацию плазмидой *P. putida* выполняли в соответствии с публикацией Choi *et al.* (Choi *et al.*, *J. Microbiol. Methods* 64, 391–397 (2006)). После трансформации и селекции интегрантов инокулировали ночные культуры. 10 мкл ночной культуры использовали для инокуляции 3 мл свежей селективной среды, и через 2 часа роста при 37°C индуцировали экспрессию ThermoCas9 3-метилбензоатом. Еще через 6 ч разведенную культуру высевали на не селективную среду с добавлением 3-метилбензоата. Для контрольной культуры добавление 3-метилбензоата не использовали на всех этапах. Подтверждение интеграции плазмиды в хромосому *P. putida* выполняли, проводя ПЦР колоний с праймерами BG2381 и BG2135. Подтверждение удаления гена *purF* выполняли, проводя ПЦР колоний с праймерами BG2381 и BG2382.

i. Выделение РНК

Выделение РНК выполняли методом фенольной экстракции, руководствуясь ранее описанным протоколом (van Hijum *et al.* *BMC Genomics* 6, 77 (2005)). 10 мл ночных культур центрифугировали при 4°C и 4816 x g в течение 15 мин и немедленно использовали для выделения РНК. После удаления среды клетки суспендировали в 0,5 мл ледяного буфера TE (pH 8,0) и хранили на льду. Все образцы распределяли в две 2-мл пробирки с завинчивающейся крышкой, содержащие 0,5 г циркониевых гранул, 30 мкл 10% SDS, 30 мкл 3 М ацетата натрия (pH 5,2) и 500 мкл Roti®-Phenol (pH 4,5–5,0, Carl Roth GmbH). Клетки разрушали на приборе FastPrep-24 (MP Biomedicals) при 5500 об/мин в течение 45 с и центрифугировали при 4°C и 10000 об/мин в течение 5 мин. 400 мкл водной фазы из каждой пробирки переносили в новую пробирку, в которую добавляли 400 мкл смеси хлороформ-изоамиловый спирт (Carl Roth GmbH), после чего образцы центрифугировали при 4°C и 18400 x g в течение 3 мин. 300 мкл водной фазы переносили в новую пробирку и смешивали с 300 мкл буфера для лизиса из набора для выделения высокочистой РНК (Roche). Затем выполняли остальную процедуру с данным набором в соответствии с протоколом производителя, за исключением этапа инкубации с ДНКазой, который проводили в

течение 45 мин. Концентрацию и целостность кДНК определяли с использованием Nanodrop-1000, целостность и концентрацию выделенной РНК проверяли с использованием NanoDrop 1000.

ј. Количественное определение мРНК методом ОТ-кПЦР

Синтез первой цепи кДНК выполняли для выделенной РНК с использованием обратной транскриптазы SuperScript™ III (Invitrogen) в соответствии с протоколом производителя. кПЦР проводили с использованием PerfeCTa SYBR Green Supermix for iQ от компании Quanta Biosciences. 40 нг каждой библиотеки кДНК использовали в качестве матрицы для кПЦР. Использовали два набора праймеров; BG9665:BG9666 для амплификации 150-нт области гена *ldhL* и BG9889:BG9890 для амплификации 150-нт последовательности гена *rpoD* (сигма-фактор РНК-полимеразы), который использовали в качестве контроля для кПЦР. кПЦР проводили в термоциклере Bio-Rad C1000.

к. ВЭЖХ

Систему высокоэффективной жидкостной хроматографии (ВЭЖХ) ICS-5000 использовали для количественного определения лактата. Система функционировала с колонкой Aminex HPLC 87H от компании Bio-Rad Laboratories и была оборудована детектором UV1000, функционирующим на длине волны 210 нм, и рефрактометрическим детектором RI-150, 40°C. Подвижная фаза состояла из 0,16 Н H₂SO₄, и скорость потока через колонку составляла 0,8 мл/мин. Все образцы разбавляли 4:1 10 мМ ДМСО в 0,01 Н H₂SO₄.

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. Выделенный белок или полипептид (Cas), связанный с кластером регулярно расположенных группами коротких палиндромных повторов (CRISPR), содержащий:

- a. аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]; и/или
- b. аминокислотный мотив X₁X₂CTX₃X₄ [SEQ ID NO: 3], где X₁ независимо выбирают из изолейцина, метионина или пролина, X₂

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

с. аминокислотный мотив X_5LX_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:

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

б. 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

с. аминокислотный мотив X_5LX_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 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 в человеческой клетке, включающий создание контакта нуклеиновой кислоты с:

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

б. белком или белковым комплексом по любому из параграфов 12-18 и по меньшей мере одной молекулой направляющей РНК по любому из параграфов 6-11; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой; и, необязательно, при этом способ не приводит к модификации генетических особенностей зародышевой линии человека.

26. Способ модификации нуклеиновой кислоты-мишени в человеческой 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; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой; и, необязательно, при этом способ не приводит к модификации генетических особенностей зародышевой линии человека.

27. Способ модификации нуклеиновой кислоты-мишени в человеческой 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; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой; и, необязательно, при этом способ не приводит к модификации генетических особенностей зародышевой линии человека.

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. Клетка-хозяин, трансформированная способом по любому из параграфов 22-36; при этом клетка представляет собой человеческую клетку; и, необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой; и, необязательно, при этом способ не приводит к модификации генетических особенностей зародышевой линии человека.

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'; при этом указанное применение практикуют с клеткой человека; и, необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой.

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 п. 5, при этом полинуклеотид, кодирующий белок Cas, содержит мутации D8A и H582A.

7. Применение по любому из предшествующих пунктов, при этом последовательность PAM представляет собой 5'-NNNNCNNA-3' [SEQ ID NO: 47].

8. Применение по любому из предшествующих пунктов, при этом последовательность PAM представляет собой 5'-NNNNCSAA-3' [SEQ ID NO: 48].

9. Применение по п. 8, при этом последовательность PAM представляет собой 5'-NNNNCCAA-3' [SEQ ID NO: 50].

10. Применение по п. 8 или п. 9, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 70°C.

11. Применение по любому из пунктов 7-10, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 25°C до 65°C.

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

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

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

15. Применение по любому из предшествующих пунктов, при

этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35–200 нуклеотидных остатков.

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

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

18. 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.

19. Применение по п. 17 или п. 18, 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.

20. Применение по любому из пунктов 17–19, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метиلاзы, 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).

21. Применение по п. 20, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

22. Применение по п. 20 или п. 21, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

23. Применение по любому из пунктов 20-22, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

24. Способ связывания, расщепления, маркирования или модификации двухцепочечного полинуклеотида-мишени в человеческой клетке, 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';

б. 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

с. связывание, расщепление, маркирование или модификацию полинуклеотида-мишени рибонуклеопротеидным комплексом; необязательно, при этом указанная человеческая клетка не является эмбриональной стволовой клеткой; и, необязательно, при этом способ не приводит к модификации генетических особенностей зародышевой линии человека.

25. Способ по п. 24, 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.

26. Способ по п. 24 или п. 25, при этом двухцепочечный полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

27. Способ по любому из пунктов 24-26, при этом полинуклеотид-мишень представляет собой двухцепочечную ДНК, и указанное применение приводит к двухцепочечному разрыву в полинуклеотиде.

28. Способ по п. 24 или п. 25, 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.

29. Способ по любому из пунктов 24-28, при этом последовательность PAM представляет собой 5'-NNNNCNNA-3' [SEQ ID NO: 47].

30. Способ по п. 29, при этом последовательность PAM представляет собой 5'-NNNNCSAA-3' [SEQ ID NO: 48].

31. Способ по п. 30, при этом последовательность PAM представляет собой 5'-NNNNCCAA-3' [SEQ ID NO: 50].

32. Способ по п. 30 или п. 31, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 70°C.

33. Способ по любому из пунктов 29-32, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 25°C до 65°C.

34. Способ по любому из пунктов 24-33, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

35. Способ по любому из пунктов 24-34, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

36. Способ по любому из пунктов 24-35, при этом молекула

направляющей РНК содержит crРНК и tracrРНК.

37. Способ по любому из пунктов 24-36, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

38. Способ по любому из пунктов 24-37, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

39. Способ по любому из пунктов 24-39, при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

40. Способ по любому из пунктов 24-40, 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.

41. Способ по п. 39 или 40, 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.

42. Способ по любому из пунктов 39-41, при этом по меньшей мере один функциональный фрагмент представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метиلاзы, 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).

43. Способ по п. 42, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

44. Способ по п. 42 или п. 43, при этом по меньшей мере

один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

45. Способ по любому из пунктов 42-44, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

46. Применение по п. 20 или способ по п. 42, 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.

47. Применение по п. 20 или способ по п. 42, 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.

48. Применение по п. 20 или способ по п. 42, при этом указанное связывание, расщепление, маркирование или модификация происходит *in vivo*.

49. Применение по любому из пунктов 1-4, 7-23 или 46, или способ по любому из пунктов 24-27, 29-46, 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.

50. Трансформированная человеческая 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 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; необязательно, при этом человеческая клетка не является эмбриональной стволовой клеткой.

51. Трансформированная клетка по п. 50, 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.

52. Трансформированная клетка по п. 50 или п. 51, при этом нуклеотидная цепь-мишень, содержащая нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

53. Трансформированная клетка по любому из пунктов 50-52, 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.

54. Трансформированная клетка по п. 50 или п. 51, 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.

55. Трансформированная клетка по любому из пунктов 50-54, при этом последовательность PAM представляет собой 5'-NNNNCNNA-3' [SEQ ID NO: 47].

56. Трансформированная клетка по п. 55, при этом последовательность PAM представляет собой 5'-NNNNCSAA-3' [SEQ ID NO: 48].

57. Трансформированная клетка по п. 56, при этом последовательность PAM представляет собой 5'-NNNNCCAA-3' [SEQ ID NO: 50].

58. Трансформированная клетка по п. 56 или п. 57, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 70°C.

59. Трансформированная клетка по любому из пунктов 55-58, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 25°C до 65°C.

60. Трансформированная клетка по любому из пунктов 50-59, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

61. Трансформированная клетка по любому из пунктов 50-60, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

62. Трансформированная клетка по любому из пунктов 50-61, при этом молекула направляющей РНК содержит crРНК и tracrРНК.

63. Трансформированная клетка по любому из пунктов 50-62, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

64. Трансформированная клетка по любому из пунктов 50-63, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

65. Трансформированная клетка по любому из пунктов 50-64,

при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

66. Трансформированная клетка по любому из пунктов 50-65, 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.

67. Трансформированная клетка по п. 65 или 66, 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, N-концом.

68. Трансформированная клетка по любому из пунктов 65-67, 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).

69. Трансформированная клетка по п. 68, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

70. Трансформированная клетка по любому из пунктов 65-69, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

71. Трансформированная клетка по любому из пунктов 65-69, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

72. Трансформированная клетка по любому из пунктов 65-70, 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.

73. Трансформированная клетка по любому из пунктов 65-69 или способ по п. 42, 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.

74. Трансформированная клетка по любому из пунктов 65-70, при этом белок Cas экспрессируется с экспрессионного вектора.

75. Трансформированная клетка по любому из пунктов 50-74, 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.

76. 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, и 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'; при этом

нуклеопротеидный комплекс находится в человеческой клетке; и, необязательно, при этом указанная клетка не является эмбриональной стволовой клеткой.

77. Нуклеопротеидный комплекс по п. 76, 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.

78. Нуклеопротеидный комплекс по п. 76 или п. 77, при этом двухцепочечный полинуклеотид-мишень, содержащий нуклеотидную последовательность-мишень, расщепляется белком Cas, предпочтительно, при этом указанное расщепление представляет собой расщепление ДНК.

79. Нуклеопротеидный комплекс по любому из пунктов 76-78, 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.

80. Нуклеопротеидный комплекс по п. 76 или п. 77, 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.

81. Нуклеопротеидный комплекс по любому из пунктов 76-80, при этом последовательность PAM представляет собой 5'-NNNNCNNA-3' [SEQ ID NO: 47].

82. Нуклеопротеидный комплекс по п. 81, при этом последовательность PAM представляет собой 5'-NNNNCSAA-3' [SEQ ID NO: 48].

83. Нуклеопротеидный комплекс по п. 82, при этом последовательность PAM представляет собой 5'-NNNNCCAA-3' [SEQ ID NO: 50].

84. Нуклеопротеидный комплекс по п. 82 или п. 83, при этом

связывание, расщепление, маркирование или модификация происходит при температуре от 20°C до 70°C.

85. Нуклеопротеидный комплекс по любому из пунктов 81-84, при этом связывание, расщепление, маркирование или модификация происходит при температуре от 25°C до 65°C.

86. Нуклеопротеидный комплекс по любому из пунктов 76-85, при этом белок Cas может быть получен из бактерии, архея или вируса, предпочтительно, из термофильной бактерии.

87. Нуклеопротеидный комплекс по любому из пунктов 76-86, при этом белок Cas может быть получен из *Geobacillus* sp., предпочтительно, из *Geobacillus thermodenitrificans*.

88. Нуклеопротеидный комплекс по любому из пунктов 76-87, при этом молекула направляющей РНК содержит crРНК и tracrРНК.

89. Нуклеопротеидный комплекс по любому из пунктов 76-88, при этом длина по меньшей мере одной молекулы направляющей РНК находится в диапазоне 35-200 нуклеотидных остатков.

90. Нуклеопротеидный комплекс по любому из пунктов 76-89, при этом нуклеотидная последовательность-мишень имеет длину от 15 до 32 нуклеотидных остатков.

91. Нуклеопротеидный комплекс по любому из пунктов 76-90, при этом белок Cas дополнительно содержит по меньшей мере один функциональный фрагмент.

92. Нуклеопротеидный комплекс по любому из пунктов 76-91, 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.

93. Нуклеопротеидный комплекс по п. 91 или 92, 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.

94. Нуклеопротеидный комплекс по любому из пунктов 91-93, при этом по меньшей мере один функциональный фрагмент

представляет собой белок; необязательно, выбранный из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метиلاзы, 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).

95. Нуклеопротеидный комплекс по п. 94, при этом естественная активность нуклеазы Cas9 инактивирована, и белок Cas связан с по меньшей мере одним функциональным фрагментом.

96. Нуклеопротеидный комплекс по любому из пунктов 91-95, при этом по меньшей мере один функциональный фрагмент представляет собой домен нуклеазы, предпочтительно, домен нуклеазы FokI.

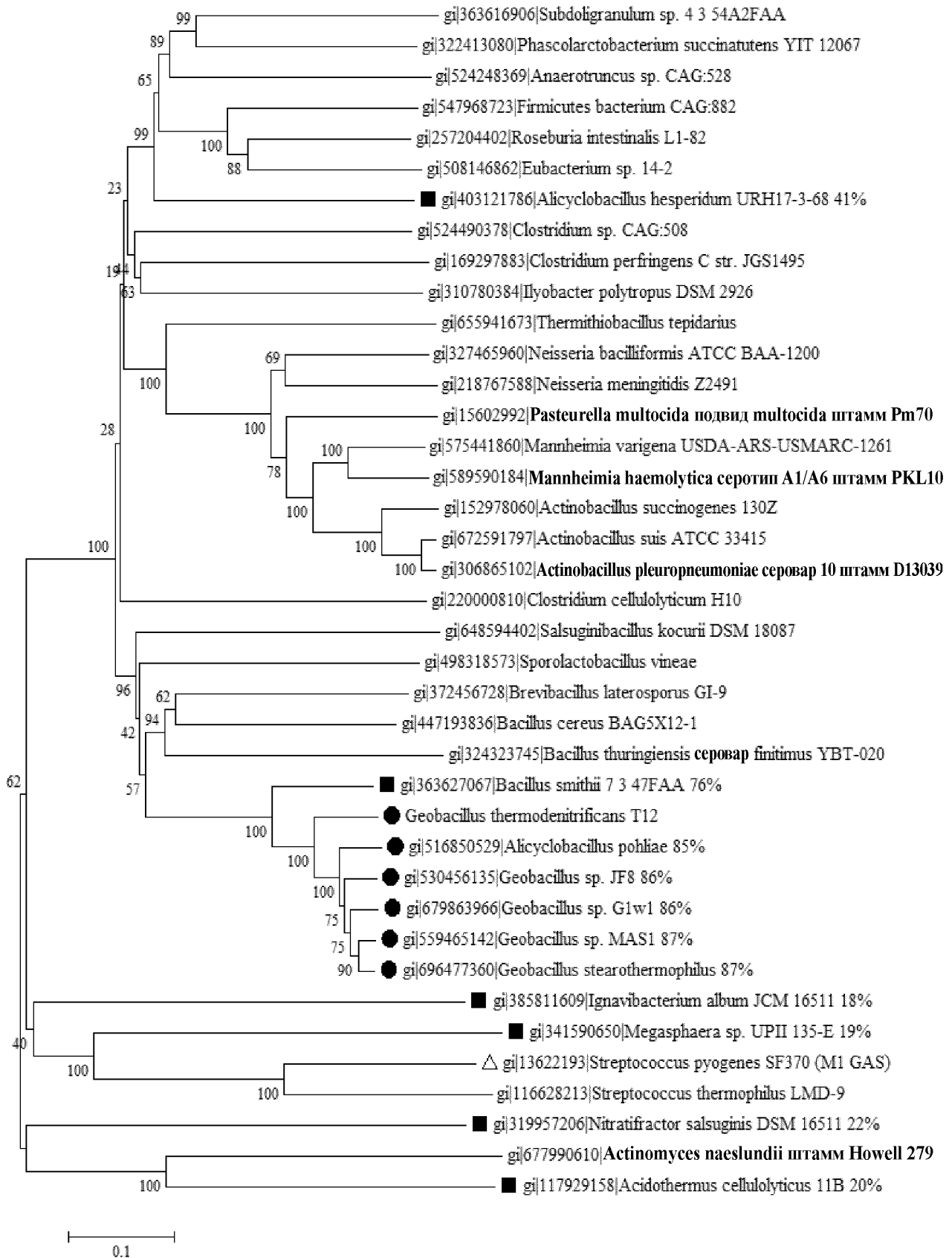
97. Нуклеопротеидный комплекс по любому из пунктов 91-95, при этом по меньшей мере один функциональный фрагмент представляет собой маркерный белок.

98. Нуклеопротеидный комплекс по любому из пунктов 91-96, 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.

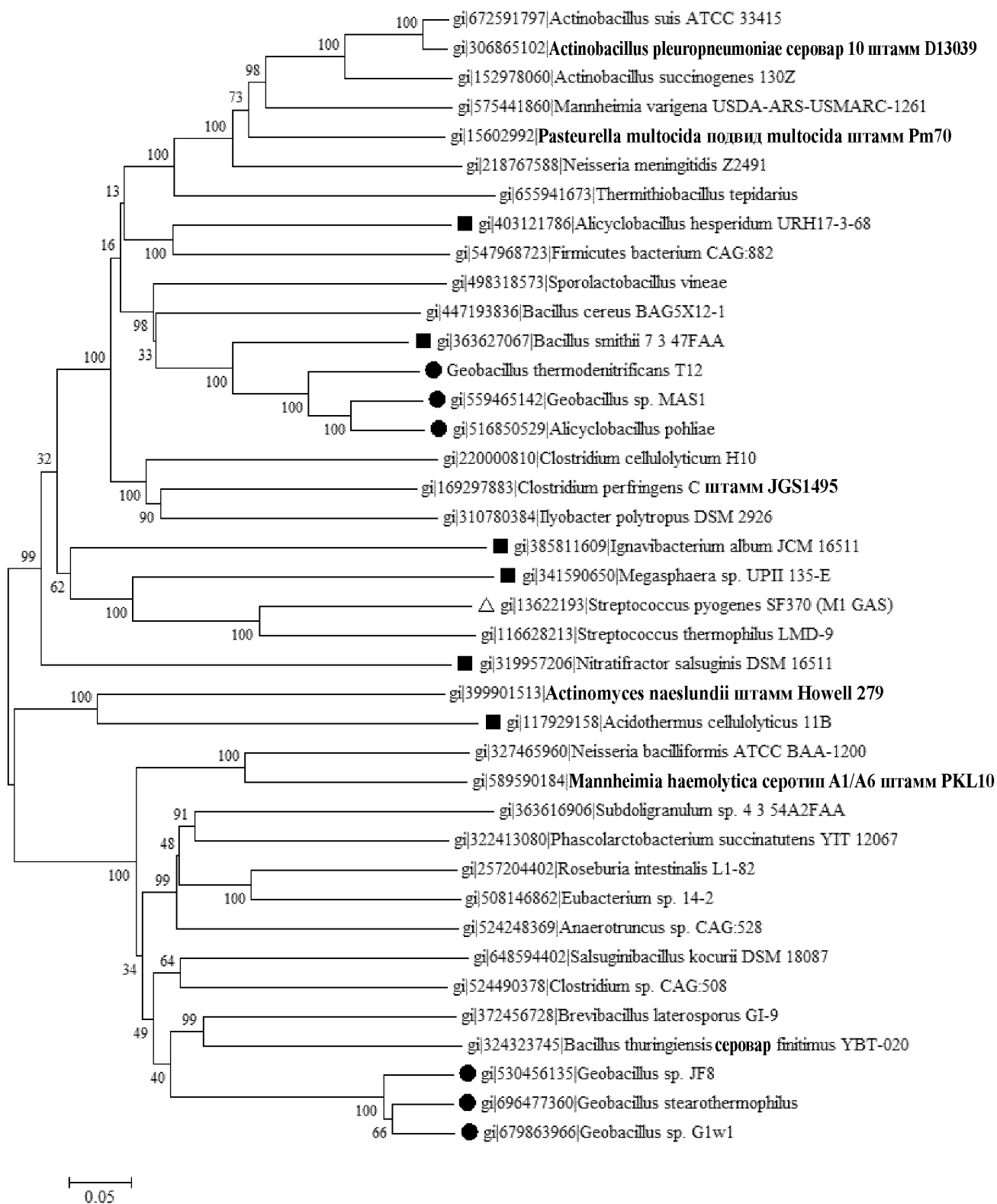
99. Нуклеопротеидный комплекс по любому из пунктов 91-95, при этом нуклеиновая кислота представляет собой дцДНК, функциональный фрагмент выбирают из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции, и образование нуклеопротеидного комплекса приводит к модификации экспрессии гена.

100. Нуклеопротеидный комплекс по любому из пунктов 76-99, 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.

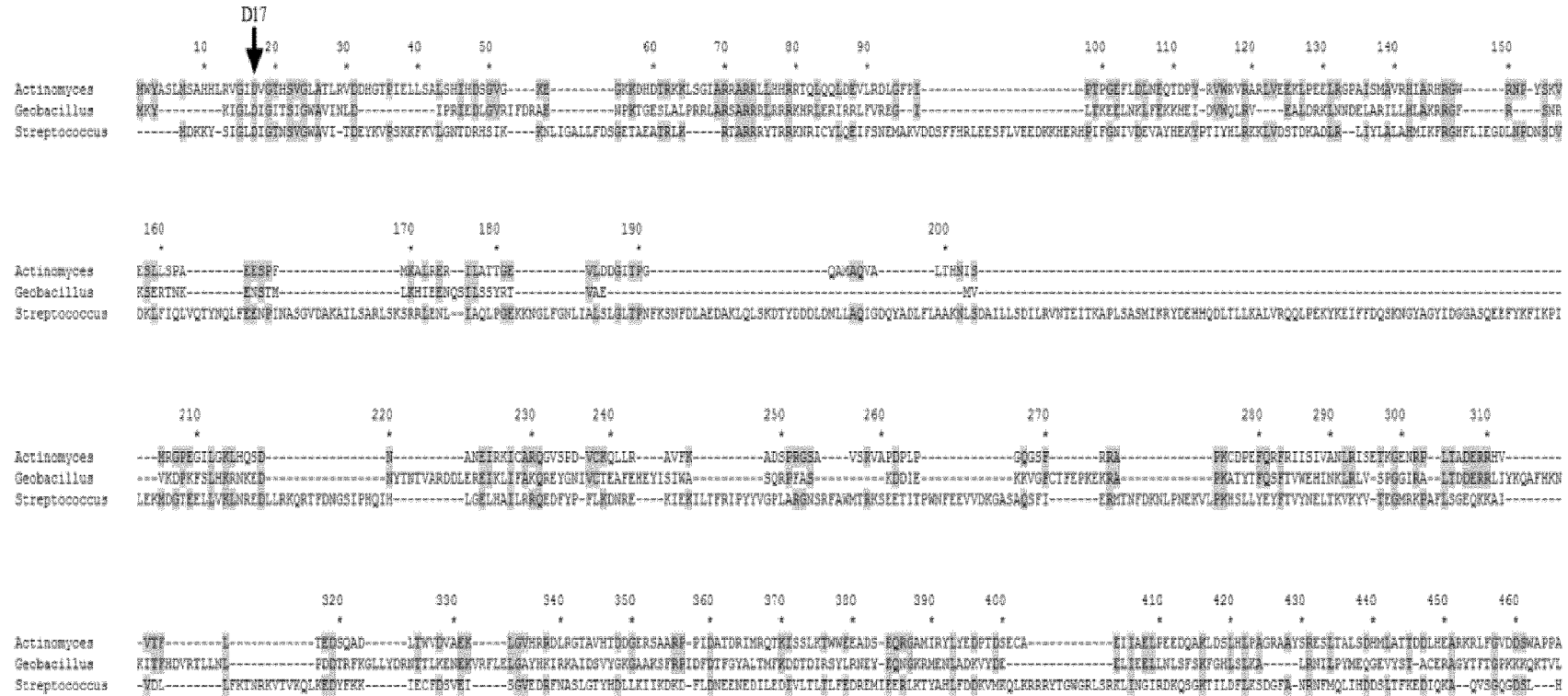
ФИГ. 1



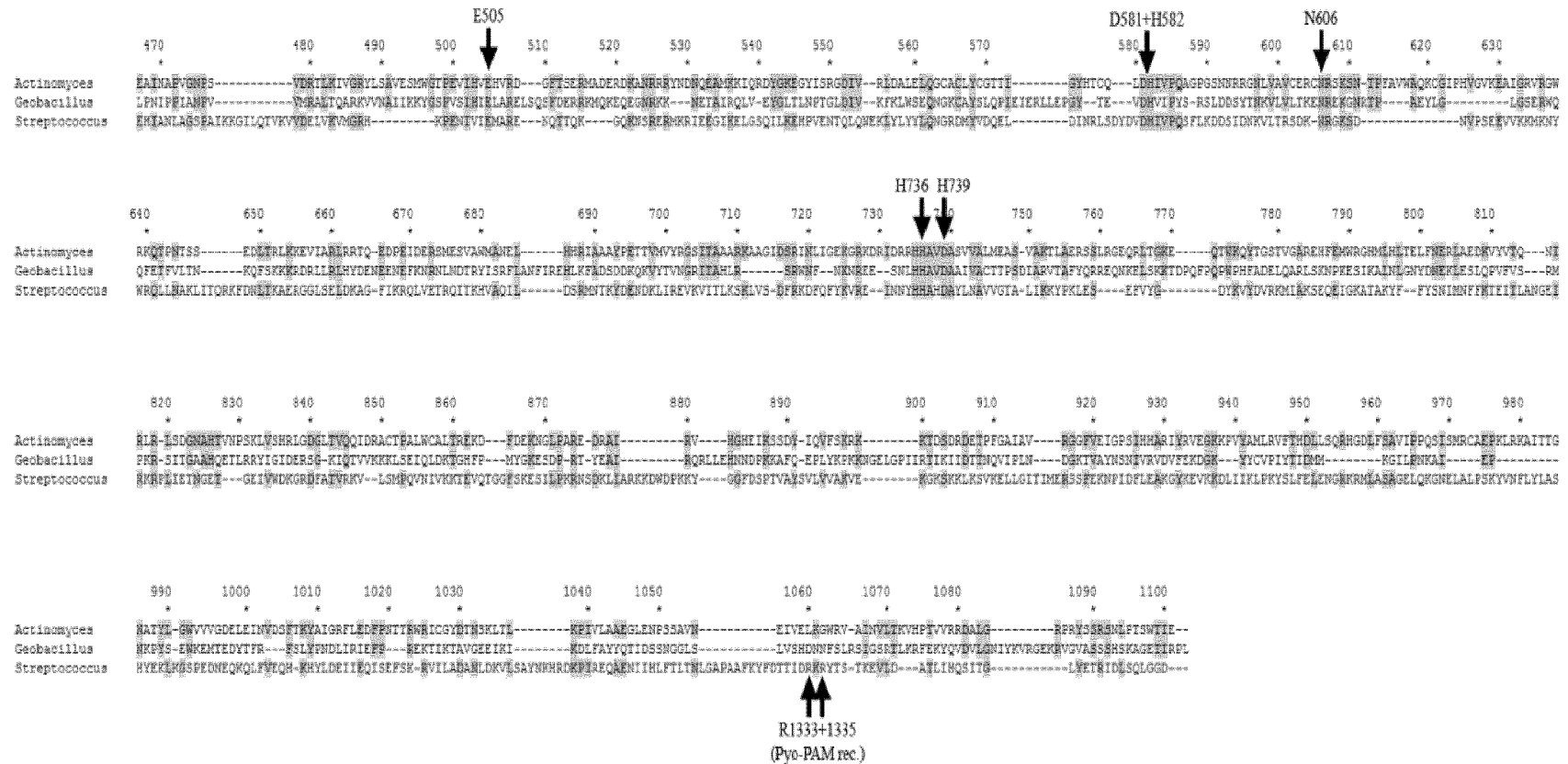
ФИГ. 2



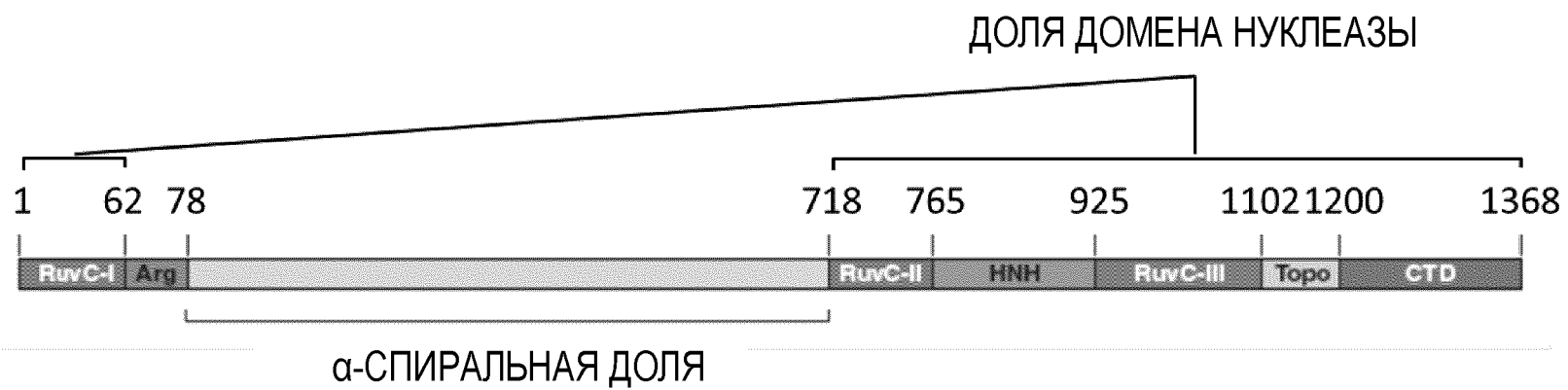
ФИГ. 3



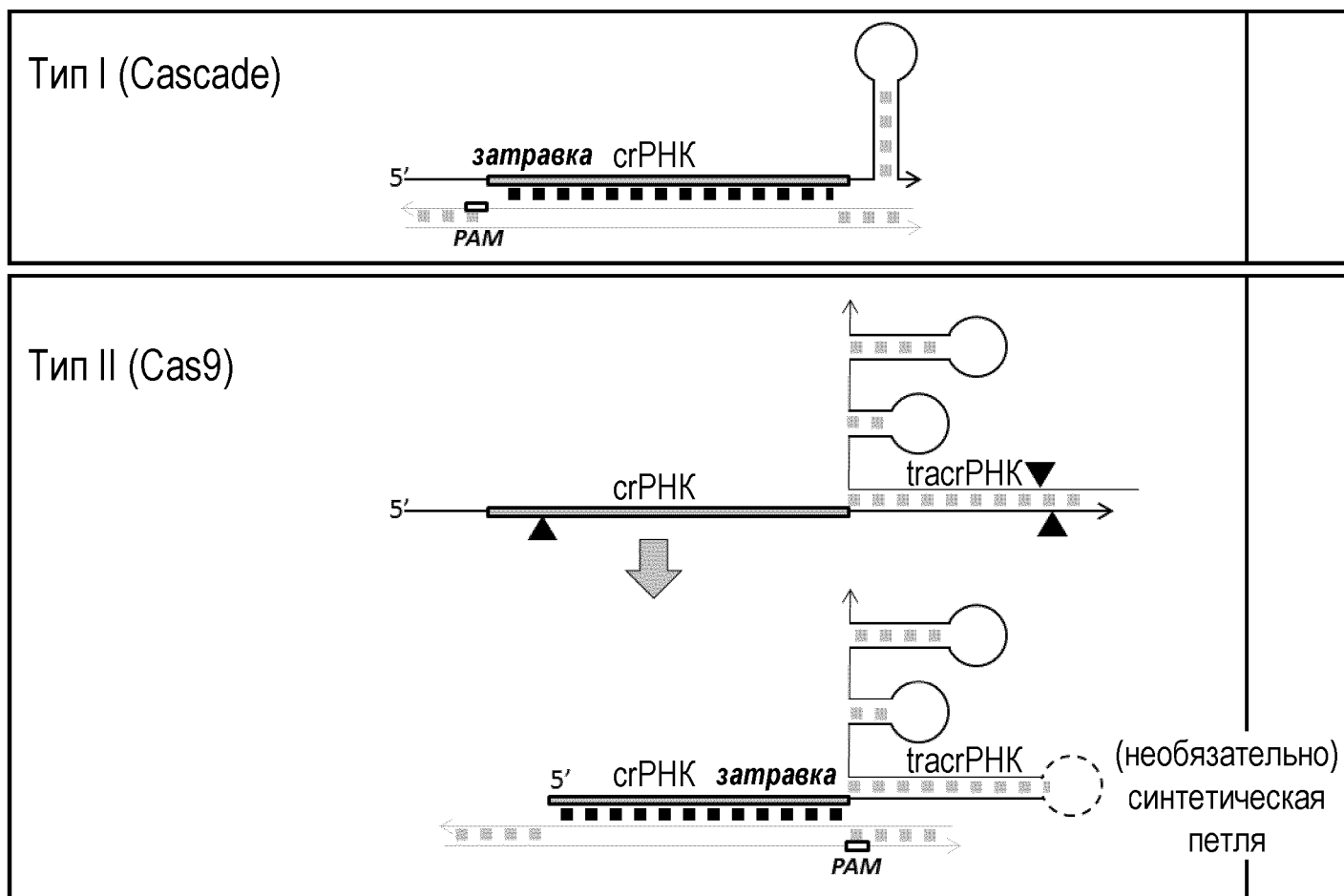
ФИГ. 3 (продолжение)



ФИГ. 4



ФИГ. 5



ФИГ. 6

```
tracrPHK t12      1 attgtttcccctcccatgcacaatagttttatagtaaaaaagacctgacgttttccgccaaggtcttc
повтор gb        -----

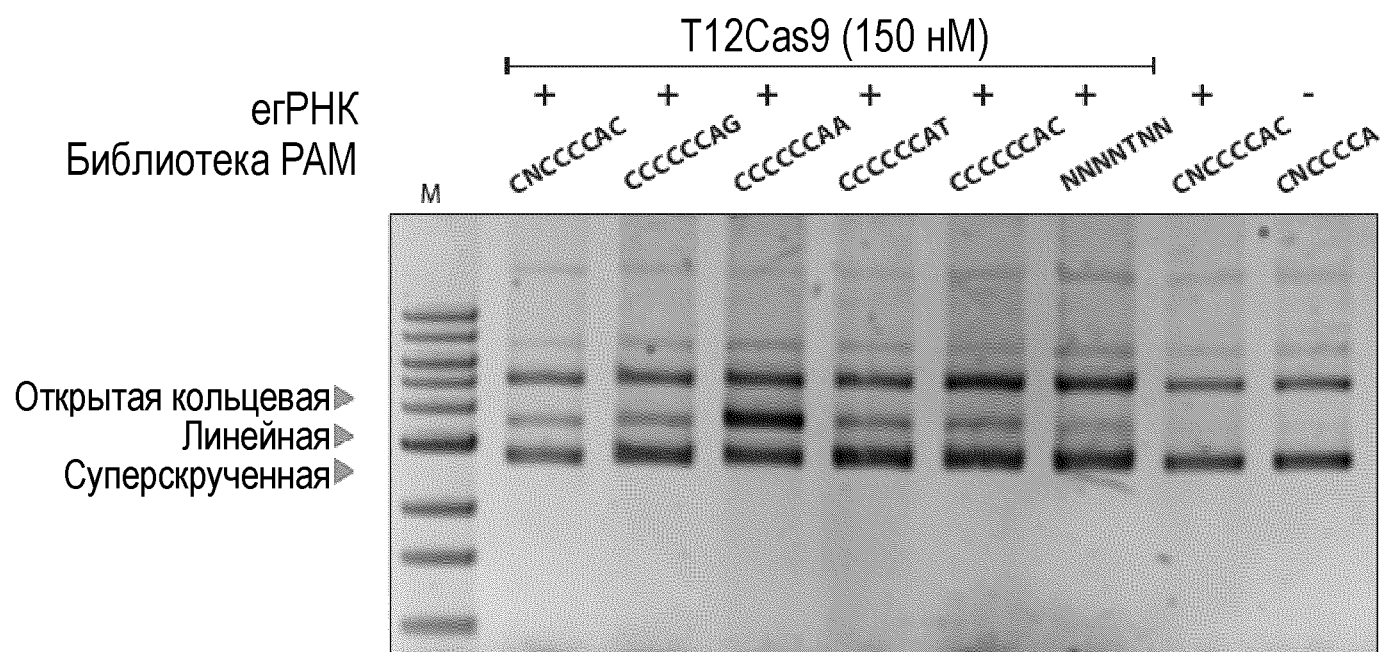
tracrPHK t12     71 gttcgccctaagagtggggaatgccogaagaaagcgggogataggcgatccccaacgccacgggtcagtct
повтор gb        -----

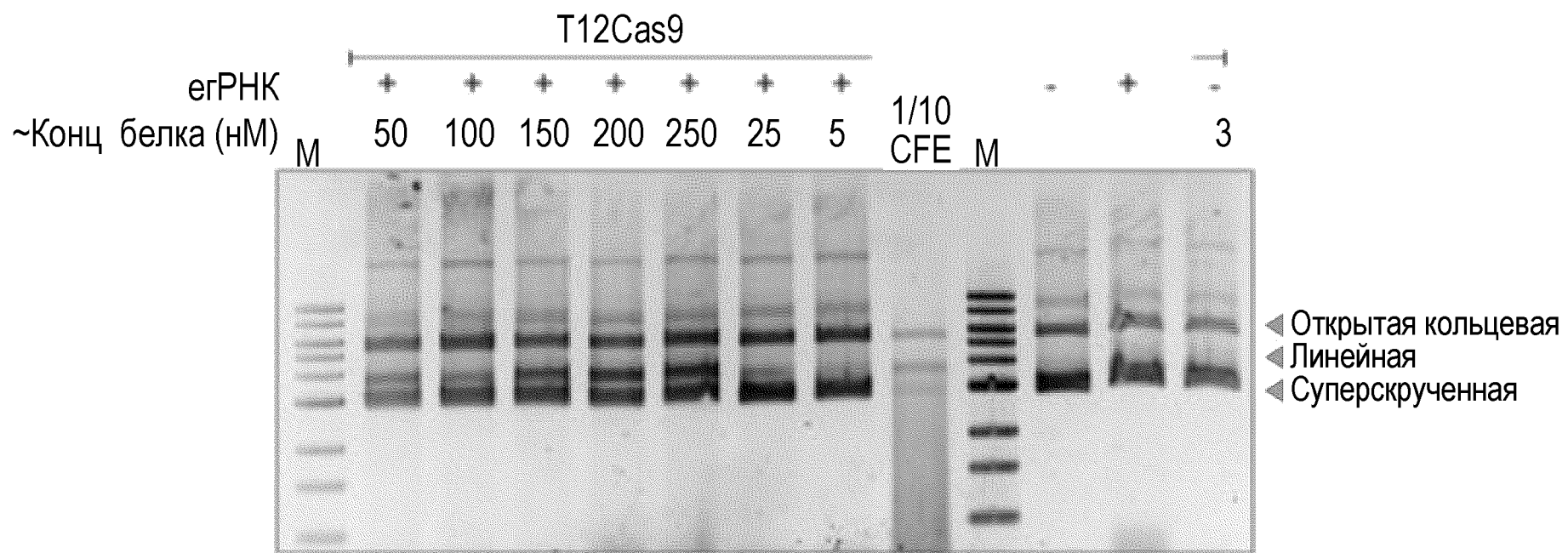
tracrPHK t12    141 gcctataggcagaaagcccttatcatagtaaccctgagatcattgctgtggtataaccctattactataa
повтор gb        1 -----gtcatagttcccctgagattatcgctgtggtataat-----

tracrPHK t12    211 taatgtttatatttgggaaaatcaagtcctttttctatattttttatactttcatttcttcttg cattat
повтор gb        -----

tracrPHK t12    281 gatgatgtgagggaggatagatttctgacaggagggtttcacatcg
повтор gb        -----
```

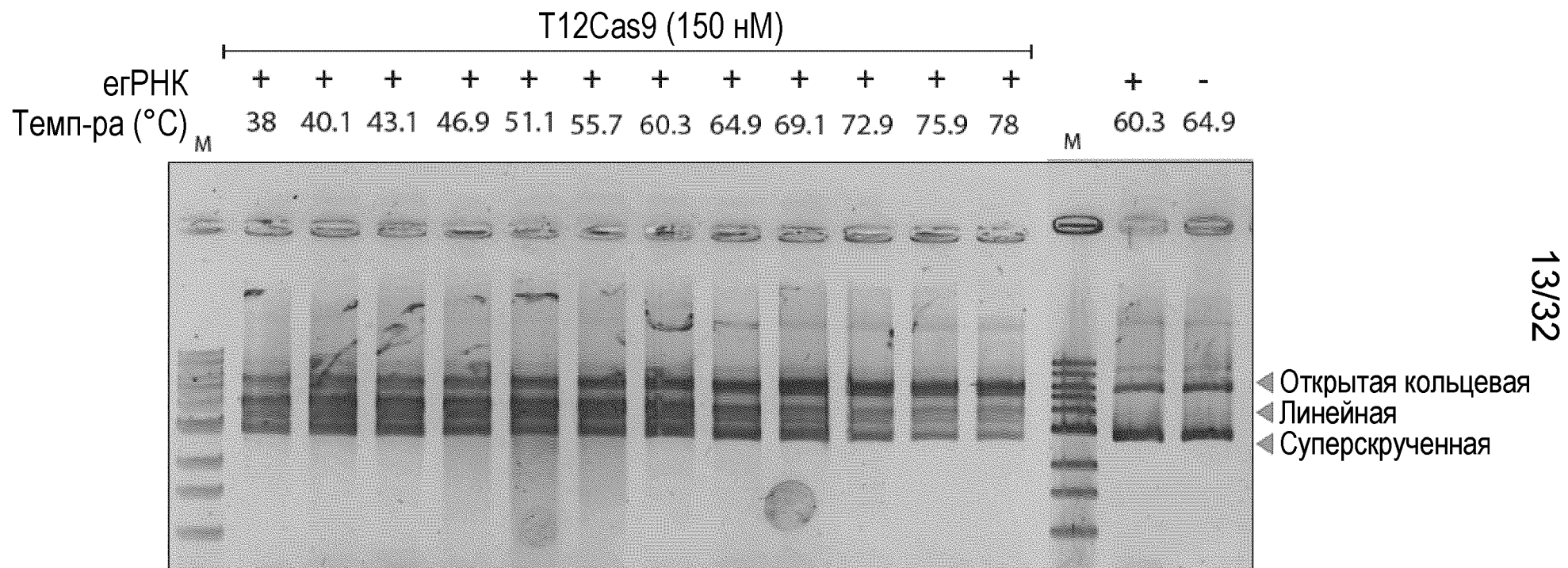

ФИГ. 9





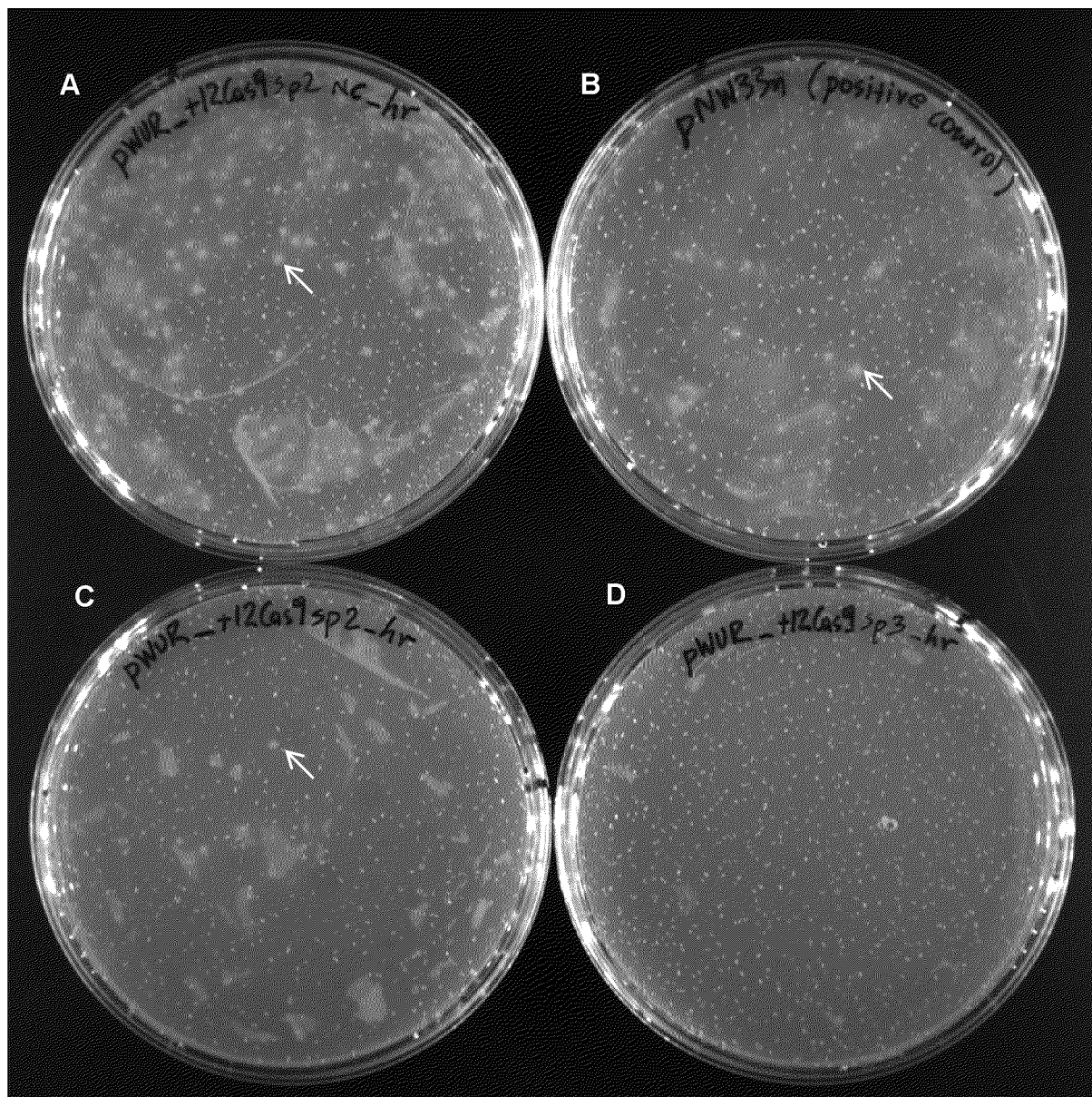
12/32

ФИГ. 10



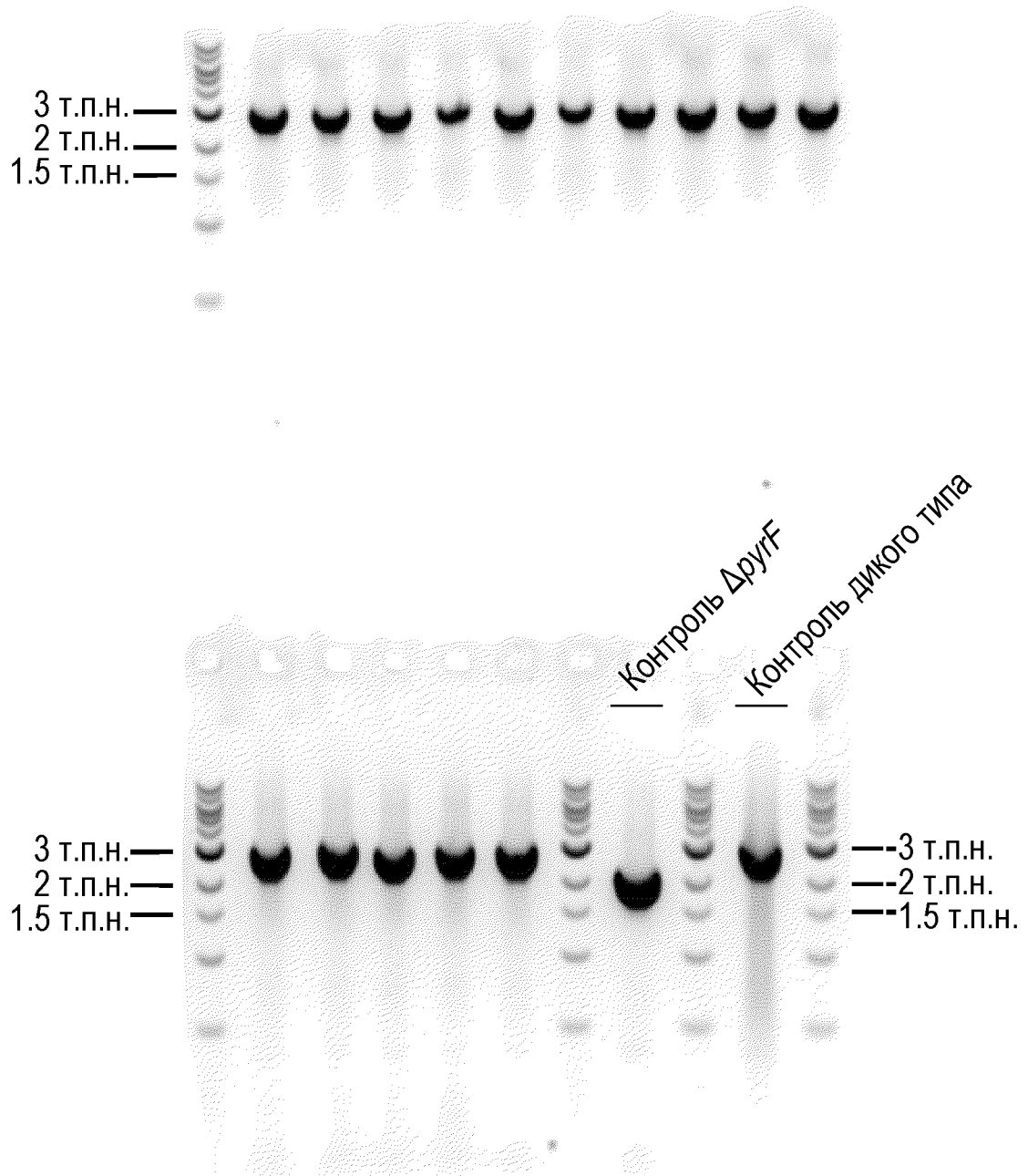
ФИГ. 11

ФИГ. 12

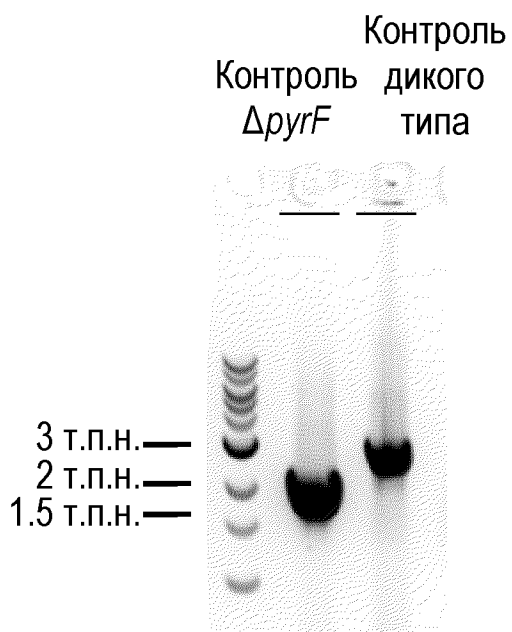
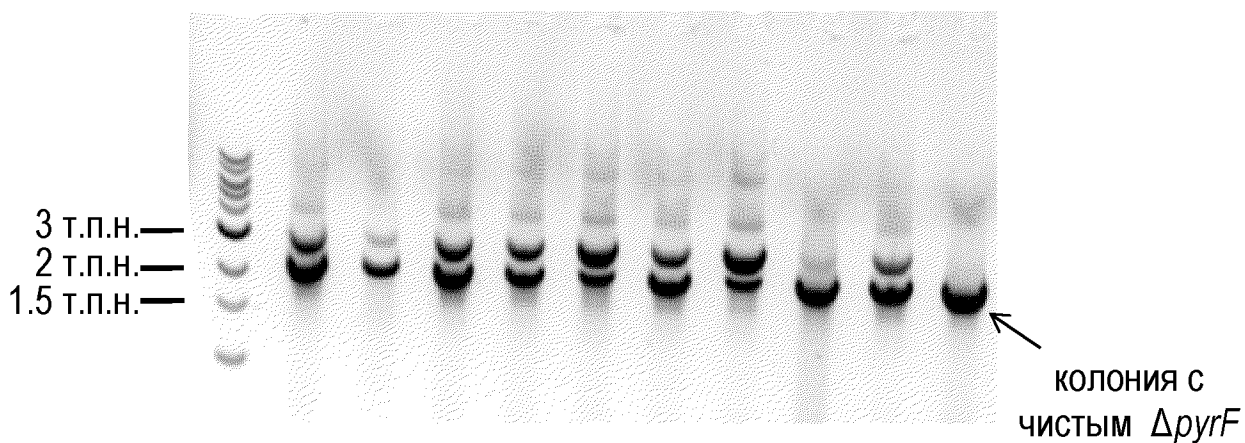
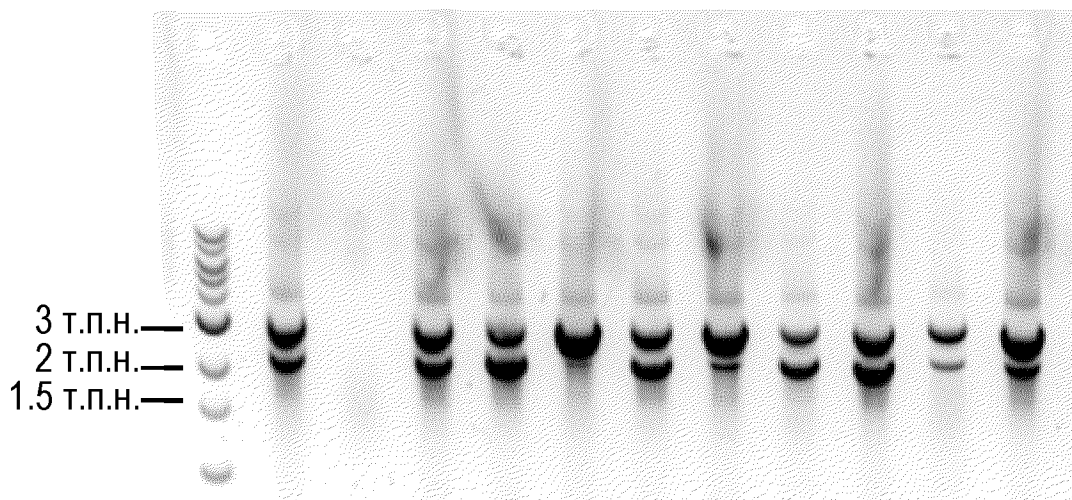


Колонии указаны стрелками

ФИГ. 13

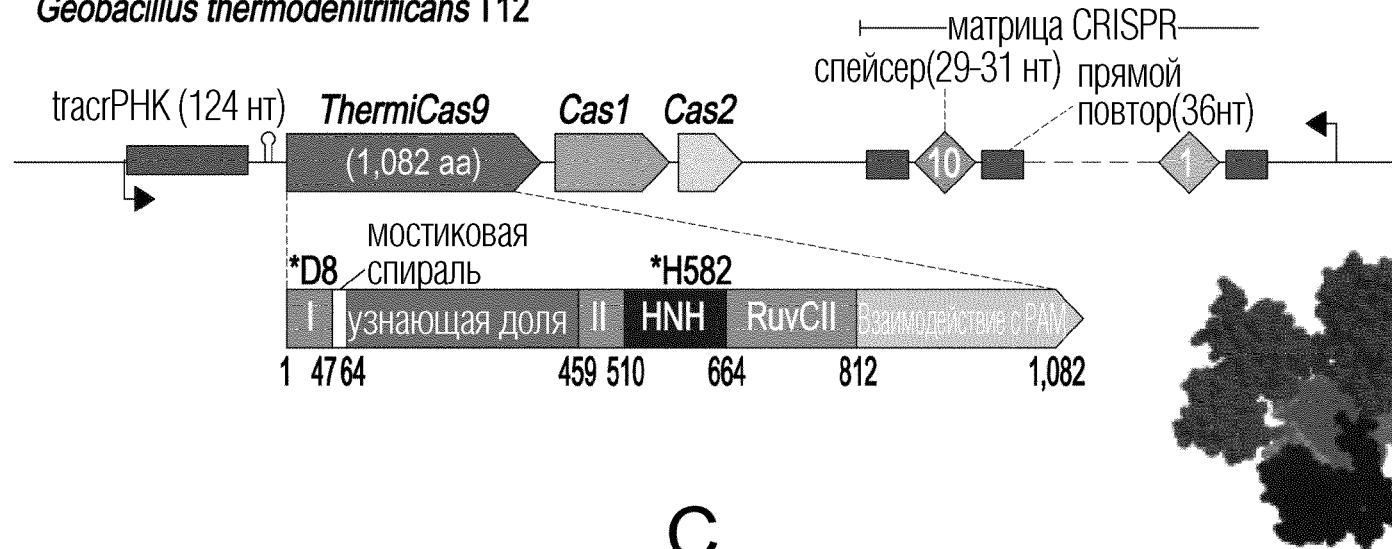


ФИГ. 14

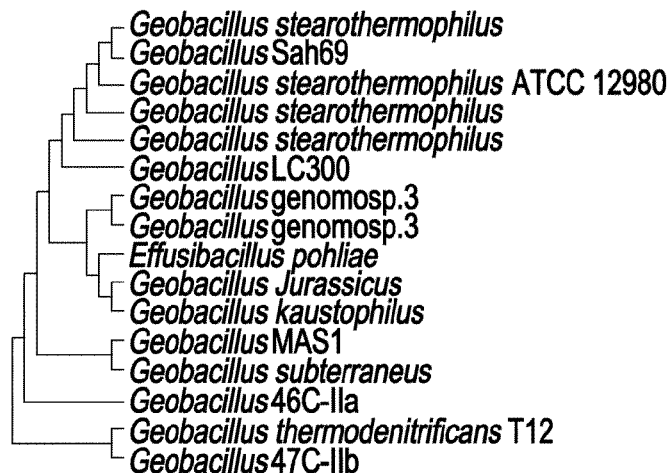


A.

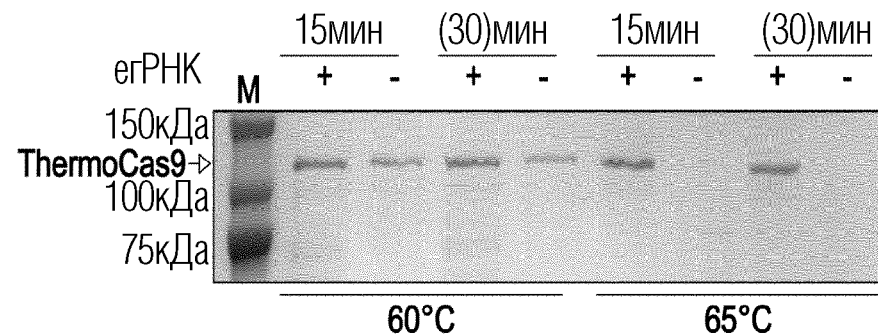
Geobacillus thermodenitrificans T12



B.



C.



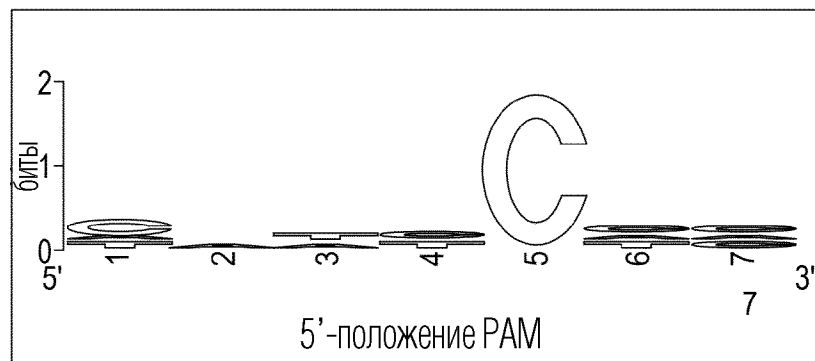
17/32

ФИГ. 15

A.

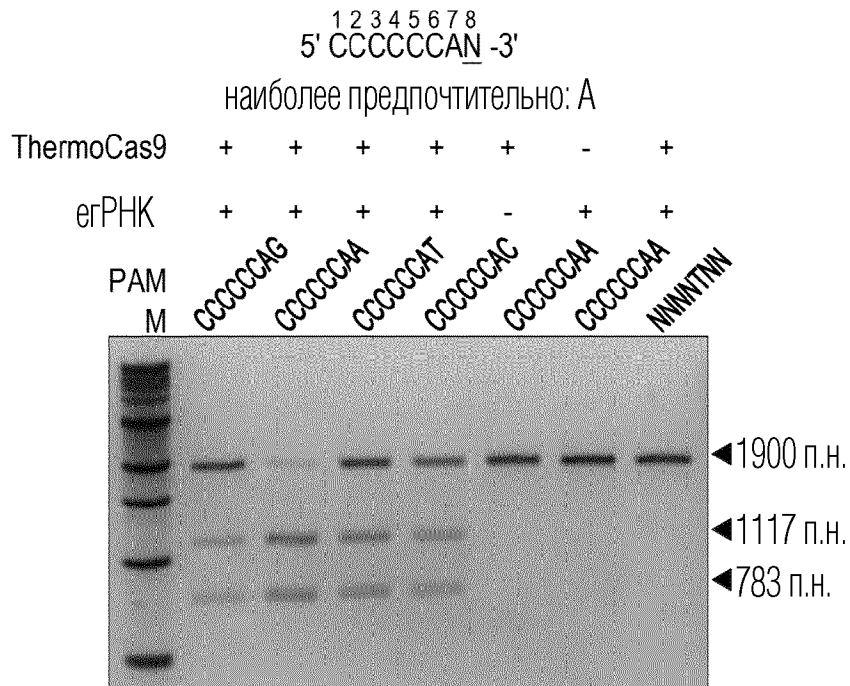


B.

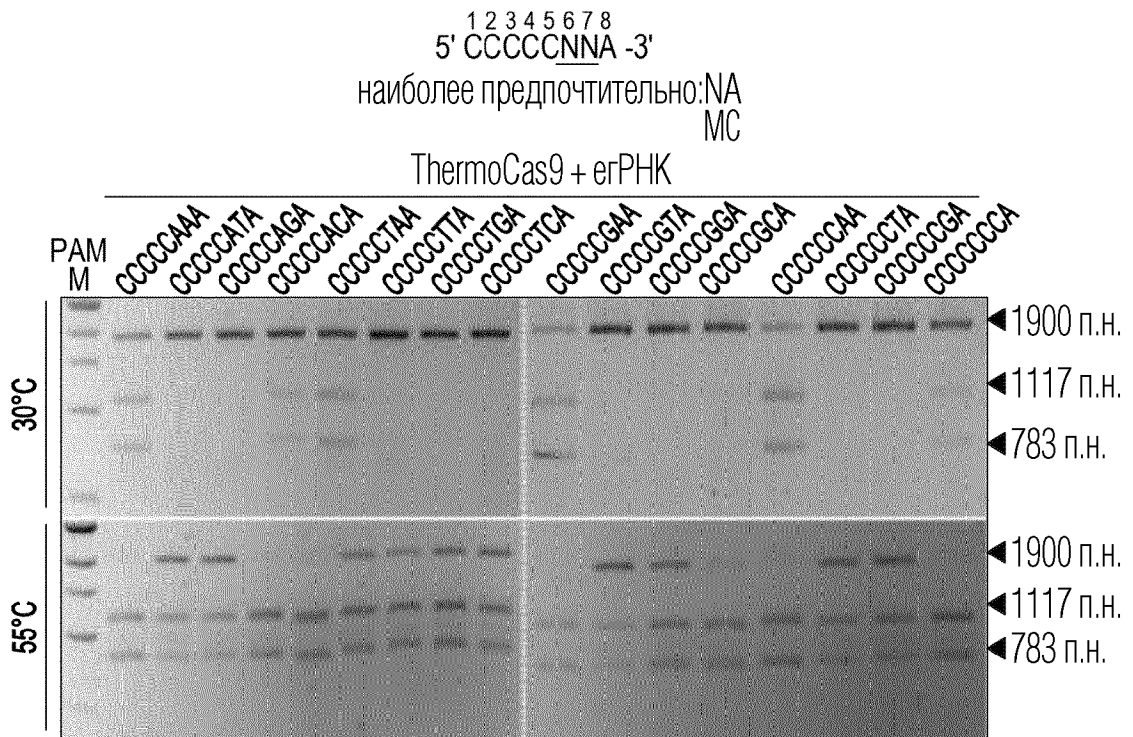


ФИГ. 16

C.



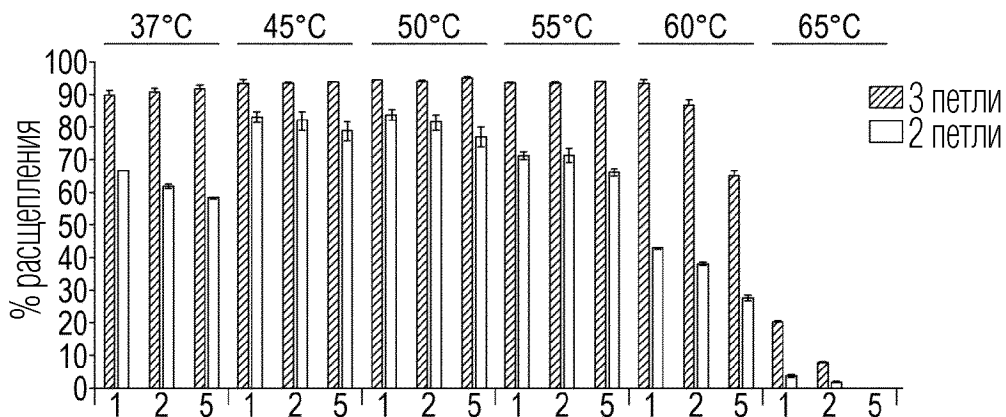
D.



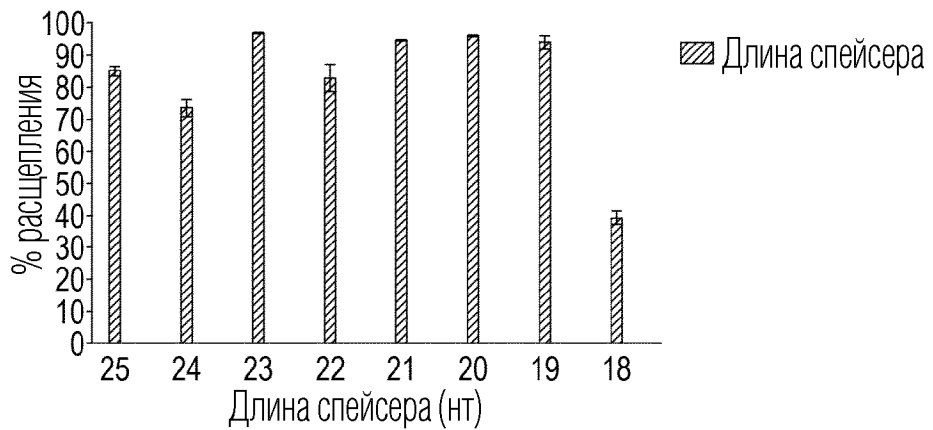
ФИГ. 16(продолжение)



B.

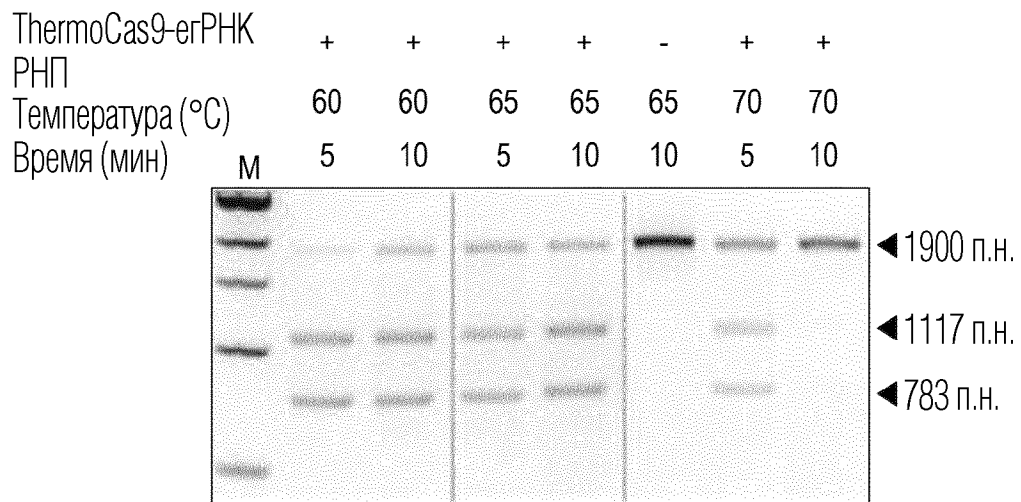


Время предварительного нагрева отдельно ThermoCas9, егРНК и ДНК-субстрата (минуты)

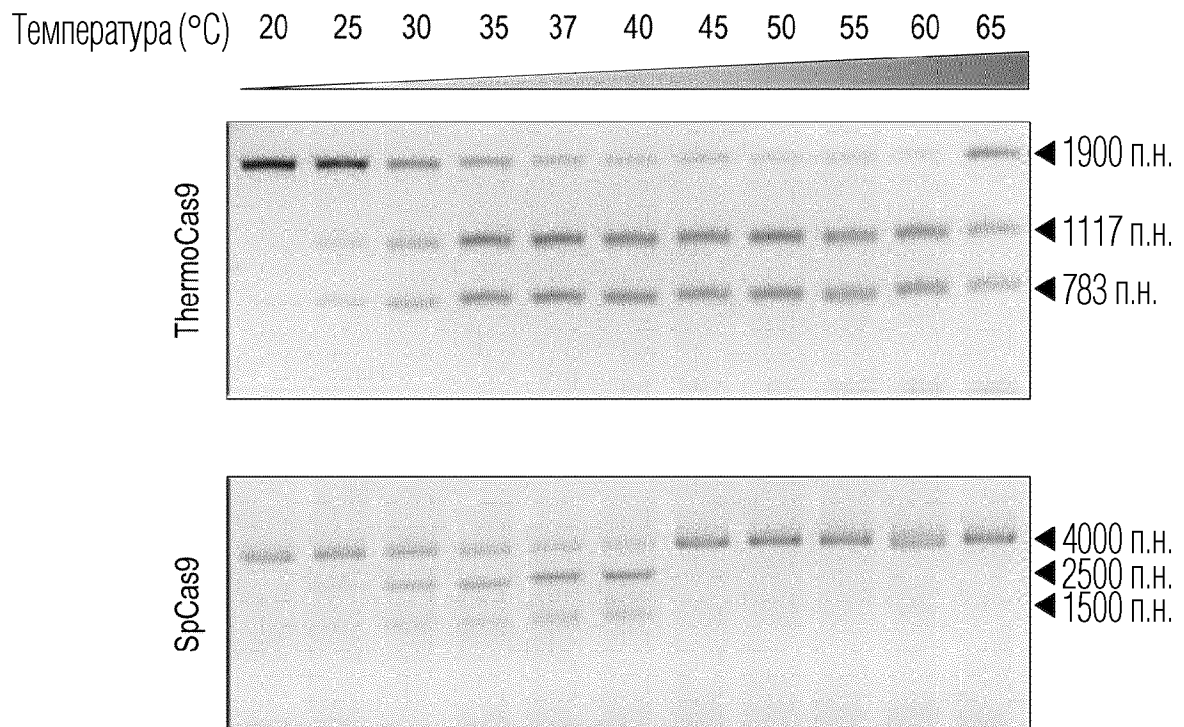


ФИГ. 17

C.



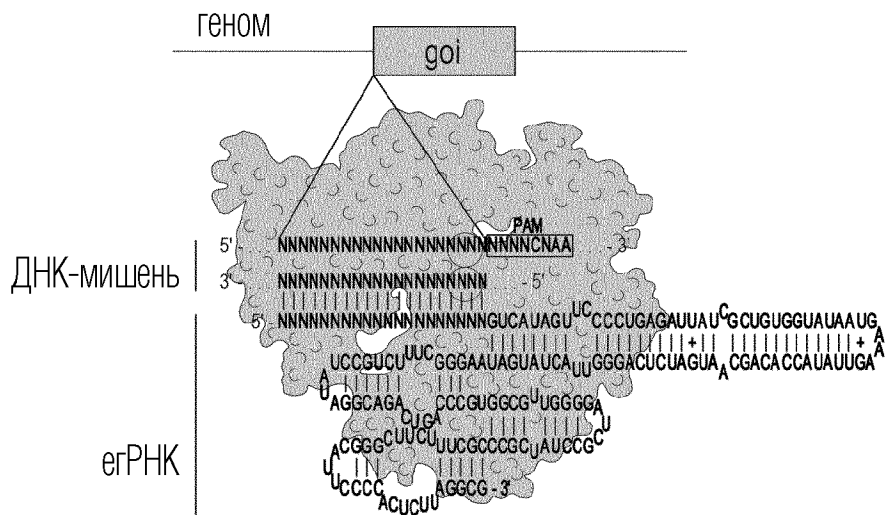
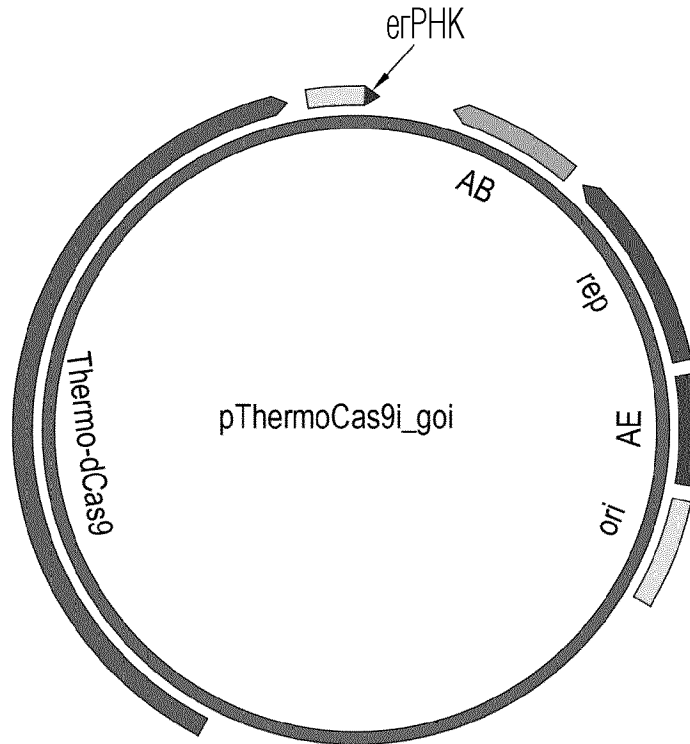
D.



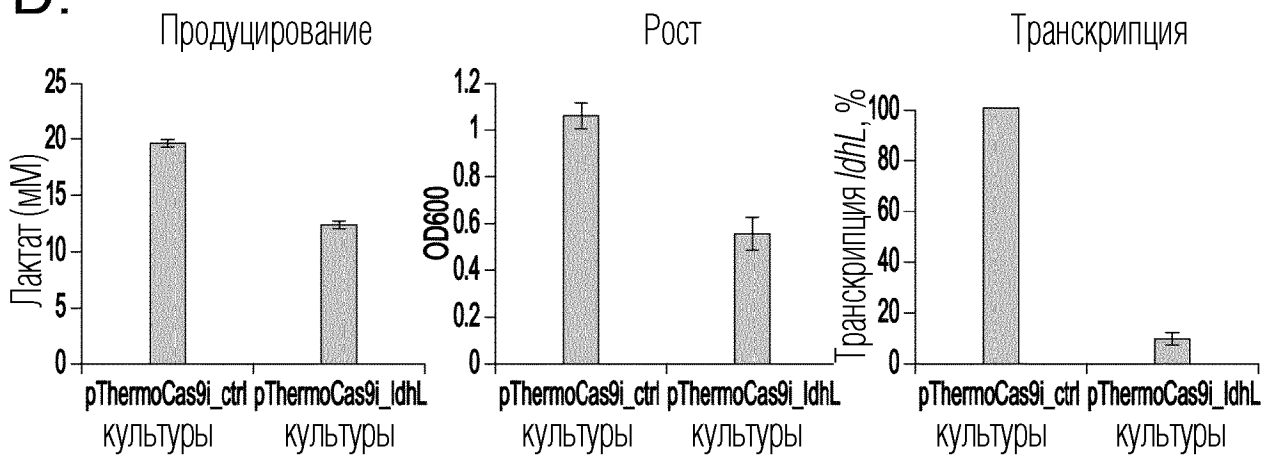
ФИГ. 17 (продолжение)

23/32

C.

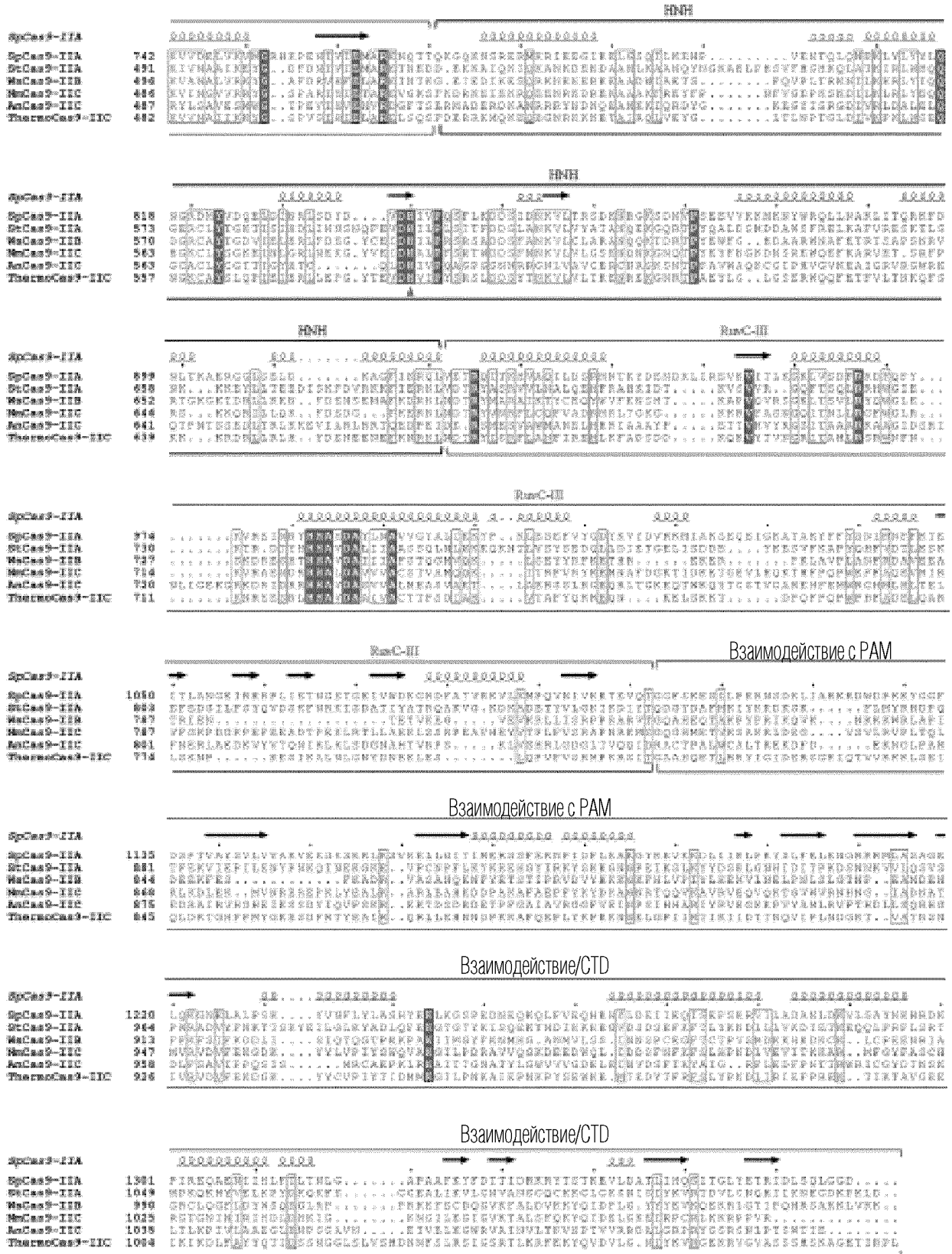


D.



ФИГ. 18 (продолжение)

ФИГ. 19 (продолжение)

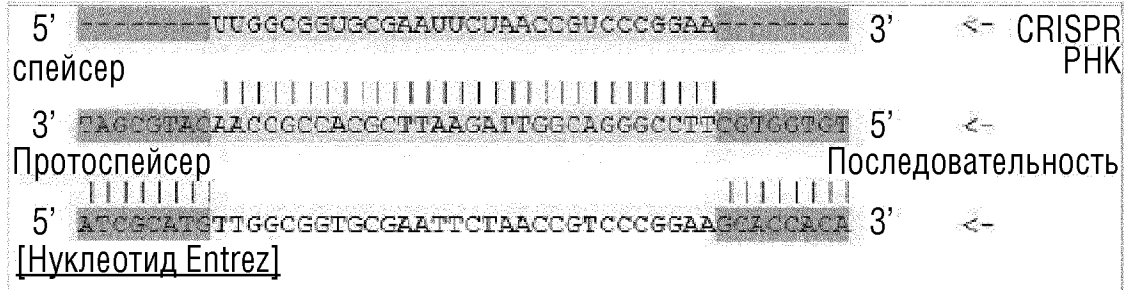


ФИГ. 20

26/32

A

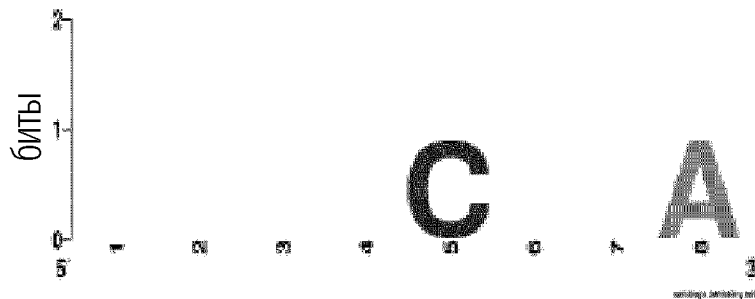
Сопоставление с: **DQ453159 DQ453159 Geobacillus вирус E2, полный геном** (DQ453159)
 положение: 39980-39951, с: **спейсер4** CRISPR №4 Спейсер №1 (спейсер4_4_1)
 положение: 1-30, Цепь: +



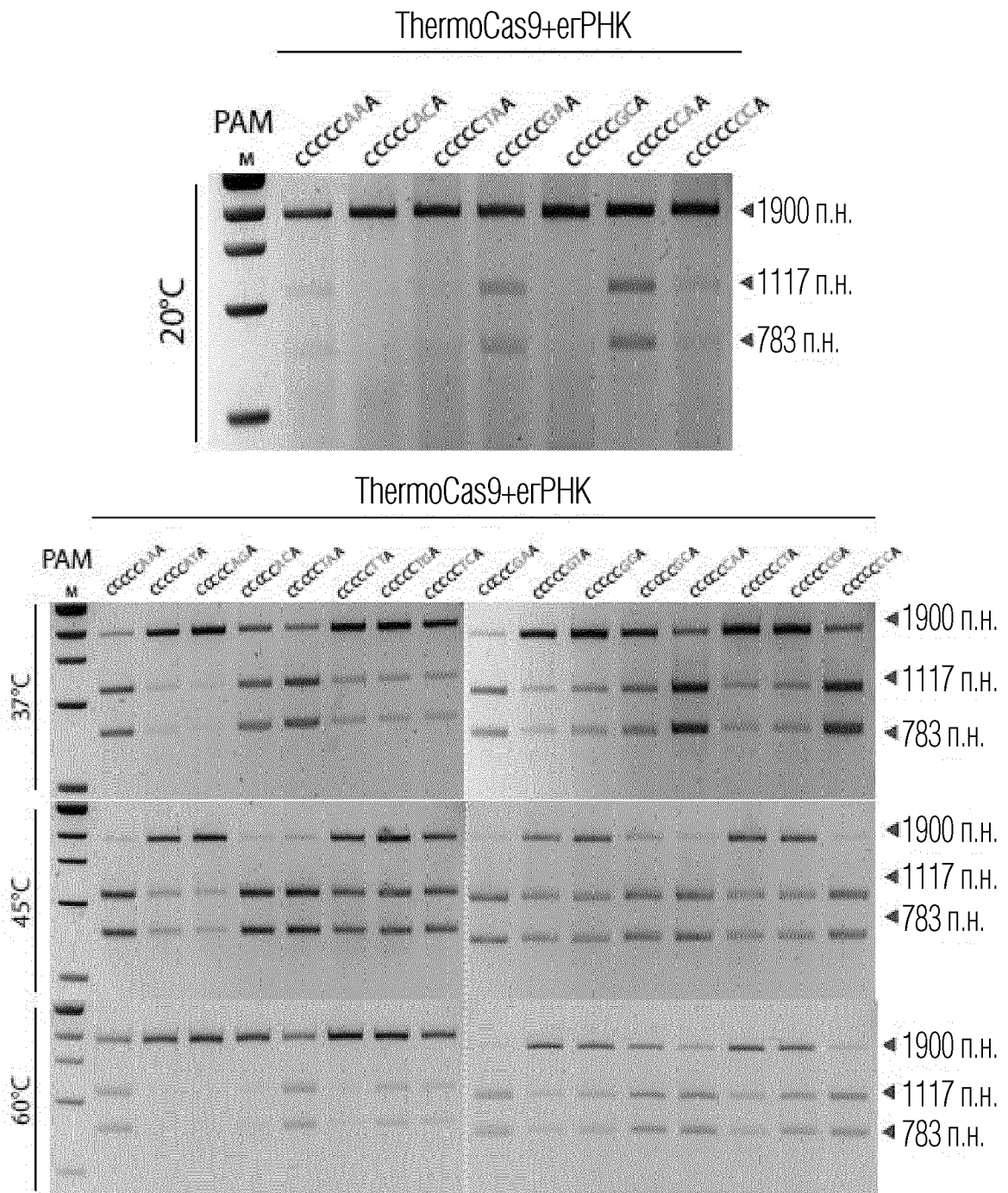
Сопоставление с: **DQ453159 DQ453159 Geobacillus вирус E2, полный геном** (DQ453159)
 положение: 6492-6463, с: **спейсер6** CRISPR №6 Спейсер №1 (спейсер6_6_1) положение:
 1-30, Цепь: +



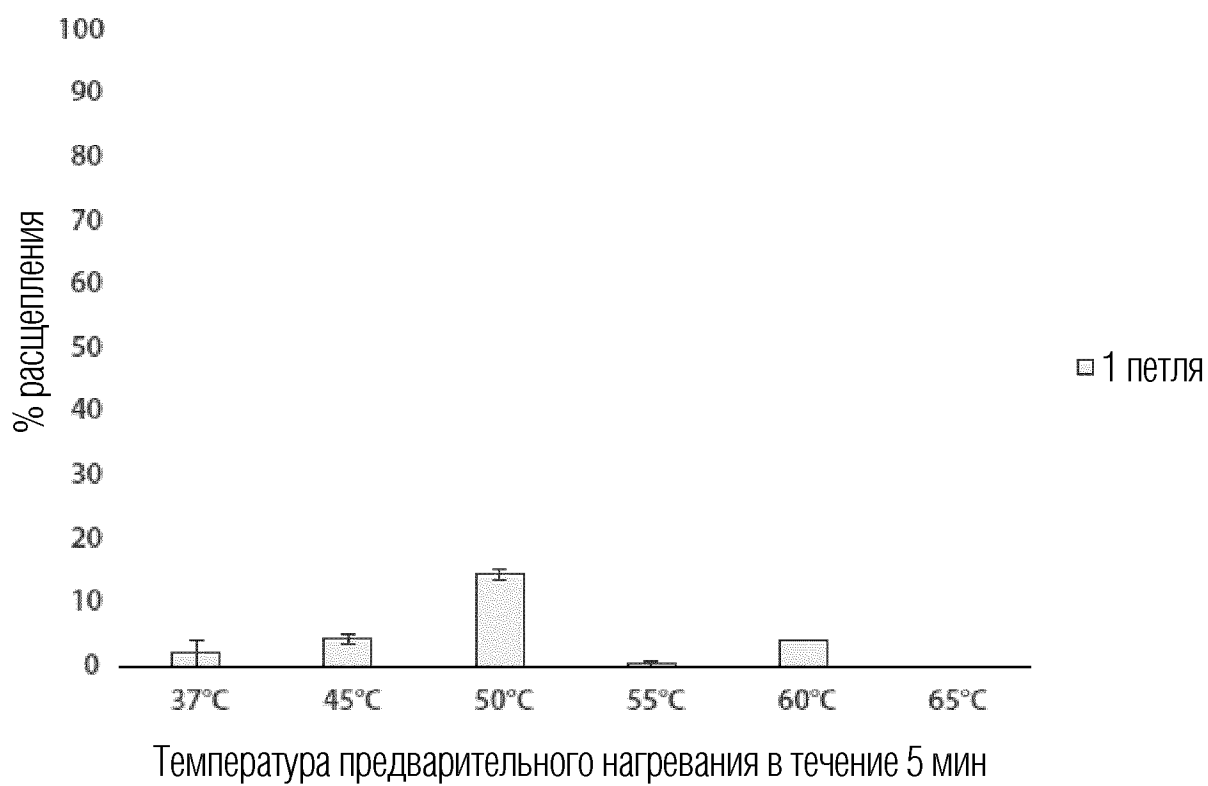
B

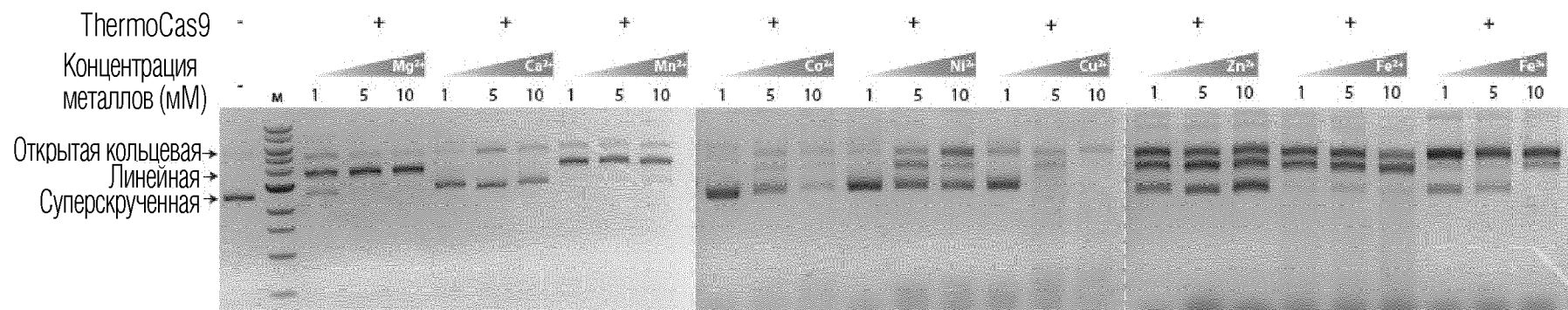
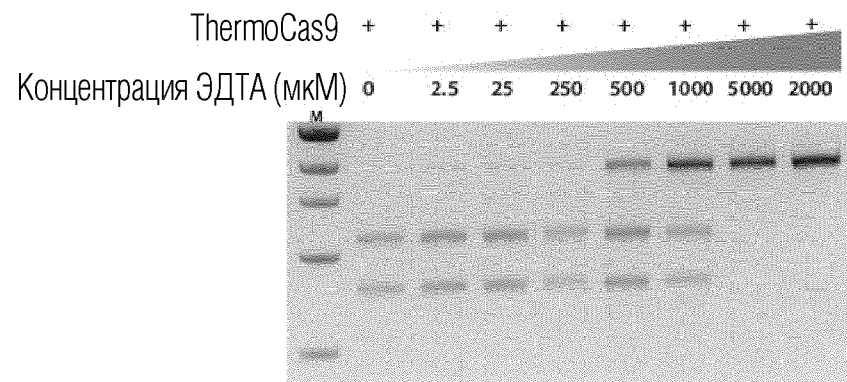


ФИГ. 21



ФИГ. 22

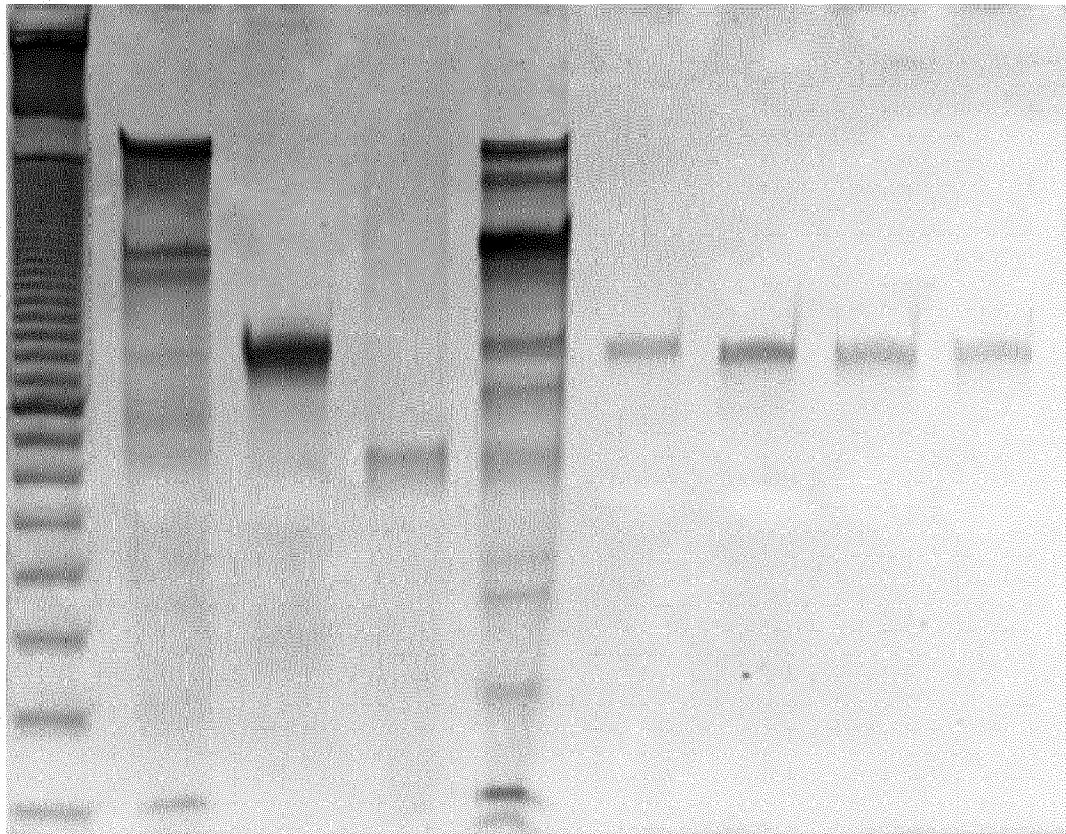




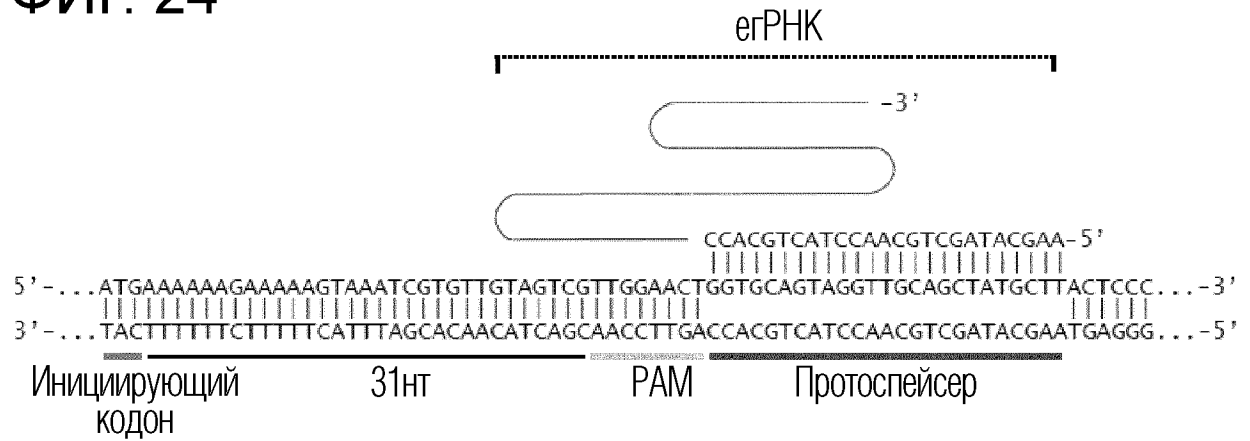
ФИГ. 23

ФИГ. 23(продолжение)

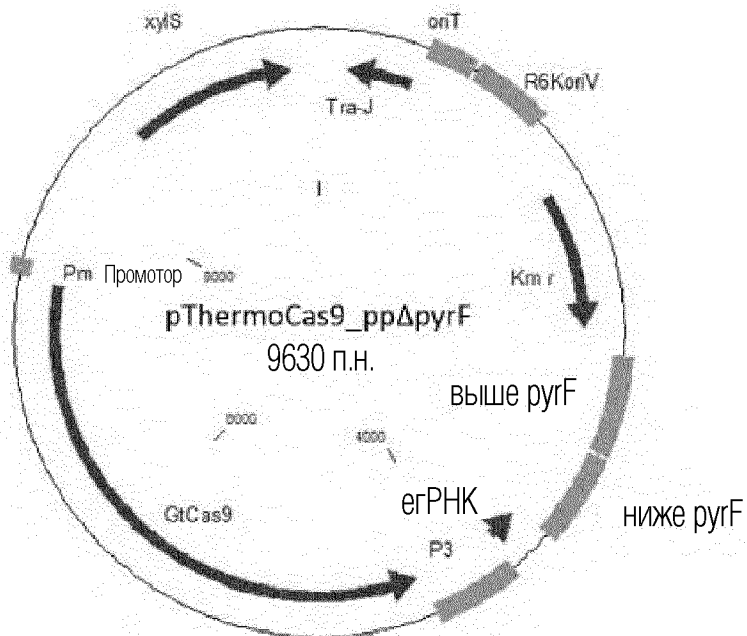
ThermoCas9	-	-	+	+	+	+	+	+
егРНК	+	-	-	+	+	+	+	+
Олигонуклеотиды	-	9301	-	-	9301	9302	9303	9304

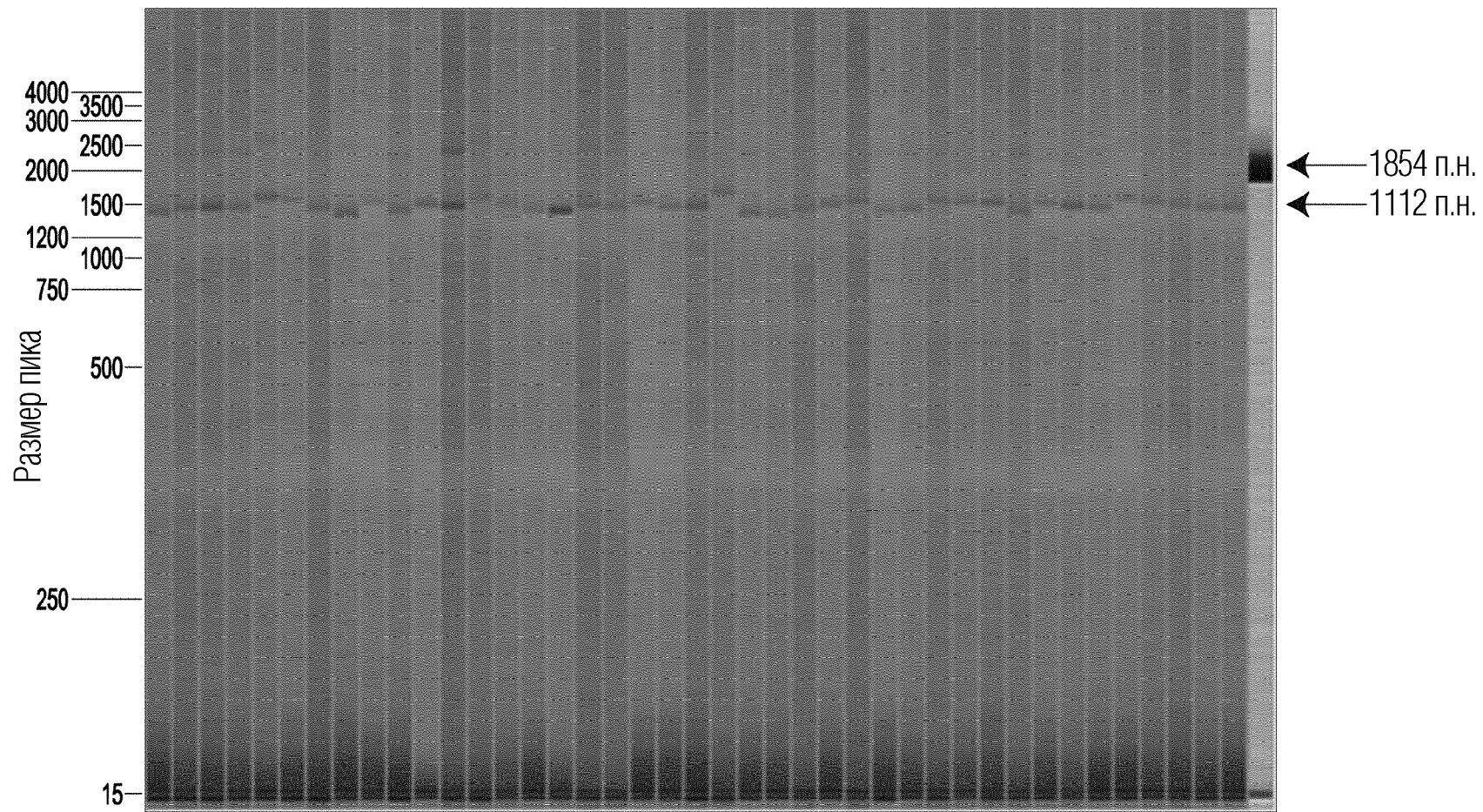


ФИГ. 24



ФИГ. 25





ФИГ. 26