

(19)



**Евразийское  
патентное  
ведомство**

(11) **038500**(13) **B1**(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**(45) Дата публикации и выдачи патента  
**2021.09.07**(51) Int. Cl. **C12N 9/22 (2006.01)**(21) Номер заявки  
**201890032**(22) Дата подачи заявки  
**2016.06.06****(54) ТЕРМОСТАБИЛЬНЫЕ НУКЛЕАЗЫ CAS9**(31) **1510296.5**(32) **2015.06.12**(33) **GB**(43) **2018.07.31**(86) **PCT/EP2016/062817**(87) **WO 2016/198361 2016.12.15**(71)(73) Заявитель и патентовладелец:  
**ВАГЕНИНГЕН ЮНИВЕРСИТЕТ  
(NL)**(72) Изобретатель:  
**Ост Ван Дер Джон, Дас Мартинус  
Йоханнес Арнолдус, Кенген Серватиус  
Вилхелмус Мария, Вос Де Виллем  
Мейндерт (NL)**(74) Представитель:  
**Медведев В.Н. (RU)**

(56) DATABASE UniProt [Online] 19 February 2014 (2014-02-19), "RecName: Full=CRISPR-associated endonuclease Cas9 {ECO:0000256 HAMAP-Rule:MF\_01480}; EC=3.1.-.- {ECO:0000256 HAMAP-Rule:MF\_01480};", XP002760956, retrieved from EBI accession no. UNIPROT:V6VHM9 Database accession no. V6VHM9 the whole document

SIDIQUI MASOOD AHMED ET AL: "Draft Genome Sequence of Geobacillus thermopakistanensis Strain MAS1.", GENOME ANNOUNCEMENTS 2014, vol. 2, no. 3, 2014, XP002760957, ISSN: 2169-8287 the whole document CHYLINSKI KRZYSZTOF ET AL: "Classification and evolution of type II CRISPR-Cas systems", NUCLEIC ACIDS RESEARCH, vol. 42, no. 10, 2014, pages 6091-6105, XP002760958, cited in the application page 6094, column 2

DATABASE Geneseq [Online] 16 January 2014 (2014-01-16), "Bacillus smithii CRISPR polypeptide SEQ: 816.", XP002760959, retrieved from EBI accession no. GSP:BAZ50353 Database accession no. BAZ50353 sequence -&amp; WO 2013/176772 A1 (UNIV CALIFORNIA [US]; UNIV VIENNA [AT]; DOUDNA JENNIFER A [US]; JINEK) 28 November 2013 (2013-11-28) cited in the application

(57) Настоящее изобретение относится к области генной инженерии и более конкретно к редактированию нуклеиновых кислот и модификации генома. В настоящем изобретении предлагается выделенный белок Cas или его полипептидный фрагмент, имеющий аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей по меньшей мере на 77%, где белок или полипептид Cas способен расщеплять ДНК при температуре в диапазоне от 50 до 100°C включительно. В настоящем изобретении дополнительно предлагаются выделенные молекулы нуклеиновой кислоты, кодирующие указанные нуклеазы Cas9, экспрессионные векторы и клетки-хозяева. Нуклеазы Cas9, раскрытые в настоящем изобретении, обеспечивают новыми инструментами для генной инженерии при повышенных температурах и являются особенно ценными для генетических манипуляций с термофильными организмами; особенно микроорганизмами.

**B1****038500****038500****B1**

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

#### **Предпосылки создания изобретения**

В 2007 году впервые было продемонстрировано, что CRISPR-Cas представляет собой адаптивную иммунную систему у многих бактерий и большинства археев (Barrangou et al., 2007, Science 315: 1709-1712), Brouns et al., 2008, Science 321: 960-964). До сих пор были охарактеризованы три типа систем CRISPR-Cas на основе функциональных и структурных критериев, большинство из которых использует малые молекулы РНК в качестве гид-РНК для направления на комплементарные последовательности ДНК-мишеней (Makarova et al., 2011, Nat Rev Microbiol 9: 467-477, Van der Oost et al., 2014, Nat Rev Microbiol 12: 479-492).

В недавнем исследовании лабораторий Doudna/Charpentier была проведена тщательная характеристика эффекторного фермента системы типа II CRISPR-Cas (Cas9), включая демонстрацию того, что введение созданных с помощью CRISPR гид-РНК (со специфическими спейсерными последовательностями) направлено действует на комплементарные последовательности (протоспейсеры) на плазмиде, вызывая разрывы двойной цепи этой плазмиды (Jinek et al., 2012, Science 337: 816-821). Затем Jinek et al., 2012 использовали Cas9 как инструмент для редактирования генома.

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

Кроме того, Cas9 использовали 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) подвергают совместной трансфекции 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 отбирают интеграцию направляющего конструкта в хромосому хозяина. Только тогда, когда происходит дополнительная рекомбинация редактирующего конструкта с целевым сайтом CRISPR в другом месте на хромосоме хозяина, хозяин может избежать проблемы аутоиммунности. Следовательно, в присутствии антибиотика только желаемые (свободные от маркера) мутанты способны выживать и расти. Также представлена сходная стратегия выбора для последующего удаления интегрированного направляющего конструкта из хромосомы, создающая свободный от собственного маркера мутант.

В последние годы было установлено, что редактирование генома, опосредуемое CRISPR-Cas, является полезным инструментом для генной инженерии. Установлено, что прокариотические системы CRISPR служат своим хозяевам как адаптивные иммунные системы (Jinek et al., 2012, Science 337: 816-821) и могут быть использованы для быстрой и эффективной генной инженерии (например, Mali et al., 2013, Nat Methods 10: 957-963), требуя только модификации гид-последовательности для направления на интересующие последовательности.

Тем не менее, существует постоянная потребность в разработке агентов с улучшенным определением специфической для последовательности нуклеиновой кислоты, ее расщеплением и манипуляциями в различных экспериментальных условиях для применения в области генетических исследований и редактирования генома. В частности, имеющиеся в настоящее время инструменты для редактирования генома, специфичного для последовательности, включая Cas9, не пригодны для использования во всех условиях или для всех организмов, например, специфичные для последовательности нуклеазы являются относительно термочувствительными и поэтому не применимы для использования у строго термофильных микроорганизмов (которые способны расти между 41 и 122°C и оптимально растут в диапазоне температур от >60°C до 80°C с гипертермофилами, способными к оптимальному росту при температуре выше 80°C), например, у микроорганизмов, которые используются при промышленной ферментации или в лабораторных методах *in vitro*, проводимых при повышенных температурах.

На сегодняшний день нет экспериментальных доказательств наличия активных белков Cas9 у термофилов. На основе сравнительного скрининга генома, проведенного Chylinski et al. (2014; Nucleic Acids Research 42: 6091-61-05) и на основе присутствия Cas9 у бактерий, было обнаружено, что система типа II-C CRISPR-Cas присутствует только приблизительно у 3,3% от всех бактериальных геномов. Среди термофильных бактерий система типа II недостаточно представлена на основе статистического анализа

( $P=0,0019$ ). Кроме того, у археев не было обнаружено системы типа II, однако это, возможно, могло быть связано с отсутствием белка РНКазы III (вовлеченной в систему типа II) у археев. Chylinski, et al., (2014; Nucleic Acids Research, 42: 6091-6105) действительно описали классификацию и эволюцию систем типа II CRISPR-Cas, в частности, идентифицировали два вида, которые демонстрируют эти системы, однако эти виды максимально растут при 55°C и не демонстрируют строго термофильного роста с оптимальной температурой роста 60-80°C, тогда как 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, изобретатели неожиданно обнаружили несколько термостабильных вариантов Cas9, которые дают возможность осуществления редактирования генома при повышенных температурах. Эти нуклеазы Cas9 предоставляют собой новые инструменты для геномной инженерии при повышенных температурах и имеют особое значение при генетических манипуляциях у термофильных организмов; особенно у микроорганизмов.

### Сущность изобретения

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

- 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 выбран из изолейцина, метионина или пролина,  $X_2$  независимо выбран из валина, серина, аспарагина или изолейцина,  $X_3$  независимо выбран из глутамата или лизина, и  $X_4$  представляет собой одно из аланина, глутамата или аргинина; и/или
- в) аминокислотный мотив  $X_5LXX_6IE$  [SEQ ID NO: 4], где  $X_5$  независимо выбран из метионина или фенилаланина и  $X_6$  независимо выбран из гистидина или аспарагина; и/или
- г) 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.

Полипептид в контексте настоящего изобретения можно рассматривать как фрагмент полноразмерного белка Cas. Такие фрагменты могут быть неактивными и использоваться в способах и с целью, не связанными непосредственно с редактированием генетического материала или его разрезанием, например, для стандартов в анализах или для повышения выработки антител или тому подобного.

В предпочтительных вариантах осуществления, однако, белок или полипептид Cas является функциональным и способен к расщеплению ДНК при температуре в диапазоне от 50 до 100°C включительно, когда он связан по меньшей мере с одной направляющей молекулой РНК и полинуклеотидом, включающим последовательность нуклеиновой кислоты-мишени, распознаваемой молекулой направляющей РНК.

В конкретных вариантах осуществления в изобретении может предлагаться белок или полипептид Cas, включающий аминокислотный мотив EKDGKYYC [SEQ ID NO: 2]. В других вариантах осуществления белки или полипептиды Cas могут дополнительно включать аминокислотный мотив  $X_1X_2CTX_3X_4$  [SEQ ID NO: 3], где  $X_1$  независимо выбран из изолейцина, метионина или пролина,  $X_2$  независимо выбран из валина, серина, аспарагина или изолейцина,  $X_3$  независимо выбран из глутамата или лизина, и  $X_4$  представляет собой одно из аланина, глутамата или аргинина.

В других вариантах осуществления белки или полипептиды Cas, определенные в настоящем документе, могут дополнительно включать еще аминокислотный мотив  $X_5LXX_6IE$  [SEQ ID NO: 4], где  $X_5$  независимо выбран из метионина или фенилаланина и  $X_6$  независимо выбран из гистидина или аспарагина.

В других вариантах осуществления белки или полипептиды Cas, определенные в настоящем документе, могут дополнительно включать еще 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, определенные в настоящем документе, могут дополнительно включать еще 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.

В соответствии с настоящим изобретением должно быть понятно, что белок или полипептид Cas по изобретению может включать любой из мотивов SEQ ID NOs с 2 по 6, либо отдельно, либо в сочетании. Нижеследующее суммирует каждое из сочетаний мотивов, которые могут характеризовать белки или полипептиды Cas по изобретению:

EKDGKYYC [SEQ ID NO: 2].





В другом аспекте настоящее изобретение относится к выделенному белку Cas или его полипептидному фрагменту, имеющему аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную с ней по меньшей мере на 77%, где белок или полипептид Cas способен к расщеплению ДНК при температуре в диапазоне от 50 до 100°C включительно.

Предпочтительно белок или полипептид Cas способен к расщеплению ДНК при температуре в диапазоне от 60 до 80°C включительно. Например, расщепление ДНК происходит при температуре 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79 или 80°C. Более предпочтительно белок или полипептид Cas способен расщеплять ДНК при температуре в интервале от 60 до 65°C. В предпочтительных аспектах белок или полипептидный фрагмент Cas по изобретению может включать 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% SEQ ID NO: 1.

Более конкретно, белок или полипептид Cas по изобретению может включать аминокислотную последовательность со следующей процентной идентичностью 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 75%, 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 аминокислотной последовательности с SEQ ID NO: 1 определяется как функция количества идентичных положений, разделенных последовательностями в выбранном окне сравнения, с учетом количества пробелов и длины каждого пробела, которые необходимо ввести для оптимального выравнивания двух последовательностей.

Белок или полипептидный фрагмент Cas согласно изобретению может быть охарактеризован с точки зрения как референсной последовательности SEQ ID NO: 1, так и любого вышеупомянутого процентного варианта, как определено процентной идентичностью последовательности, отдельно или в сочетании с любым из вышеупомянутых аминокислотных мотивов (т.е. SEQ ID NOS 2 и/или 3, и/или 4, и/или 5, и/или 6) в качестве основных признаков.

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

Во всех вышеупомянутых аспектах настоящего изобретения аминокислотные остатки могут быть заменены консервативно или неконсервативно. Консервативные аминокислотные замены относятся к таким заменам, когда аминокислотные остатки заменены другими аминокислотными остатками со сходными химическими свойствами (например, зарядом или гидрофобностью) и, следовательно, не изменяют функциональных свойств полученного полипептида.

Сходно, специалисту в данной области техники должно быть понятно, что последовательности нуклеиновой кислоты могут быть замещены консервативно или неконсервативно, не влияя на функцию полипептида. Консервативно модифицированные нуклеиновые кислоты представляют собой нуклеиновые кислоты, замещенные нуклеиновыми кислотами, которые кодируют идентичные или функционально идентичные варианты аминокислотных последовательностей.

Специалисту должно быть понятно, что каждый кодон в нуклеиновой кислоте (за исключением AUG и UGG, обычно единственных кодонов для метионина или триптофана, соответственно) может быть модифицирован с получением функционально идентичной молекулы. Соответственно, каждый молчащий вариант (т.е. синонимичный кодон) полинуклеотида или полипептида, который кодирует полипептид по настоящему изобретению, не явно выражен в каждой описанной полипептидной последовательности.

#### Температуры расщепления

Температурный диапазон, включая оптимальный температурный диапазон нуклеазной активности белков Cas по настоящему изобретению, значительно выше, чем у известных белков Cas9. Кроме того, верхняя граница диапазона, в котором он сохраняет нуклеазную активность, намного выше, чем у известных белков Cas9. Более высокая оптимальная температура и функциональный диапазон обеспечивают значительное преимущество в геномной инженерии при высоких температурах и, следовательно, например, при редактировании геномов термофильных организмов, многие из которых применяются в ряде промышленных, сельскохозяйственных и фармацевтических процессов, проводимых при повышенных температурах.

Преимущественно белки или полипептиды Cas согласно изобретению способны расщеплять нук-

леиновые кислоты при температуре от 20 до 100°C, но особенно полезны при повышенных температурах, например, от 41 до 122°C, предпочтительно от 50 до 100°C. Белки и полипептиды Cas по изобретению способны расщеплять ДНК, РНК и синтетические нуклеиновые кислоты. Белки или полипептиды Cas по изобретению могут также обеспечивать действенность нуклеазной активности и применение для редактирования генов, например, при температурах в диапазоне от 20 до 40°C.

Предпочтительно, белки или полипептиды Cas по изобретению при взаимодействии с подходящей gRNA (гид-РНК), которая распознает последовательность-мишень в молекуле(ах) полинуклеотида, подлежащей(их) расщеплению, маркировке или модификации, делают это при температурах в диапазоне от 50 до 100°C, необязательно в диапазоне от 55 до 100°C, от 60 до 100°C, от 65 до 100°C, от 70 до 100°C, от 75 до 100°C, от 80 до 100°C, от 85 до 100°C, от 90 до 100°C, от 95 до 100°C. Более предпочтительно, белки Cas по изобретению расщепляют, маркируют или модифицируют нуклеиновые кислоты при температурах в диапазоне от 51 до 99°C, от 52 до 98°C, от 53 до 97°C, от 54 до 96°C, от 55 до 95°C, от 56 до 94°C, от 57 до 93°C, от 58 до 92°C, от 59 до 91°C, от 60 до 90°C, от 61 до 89°C, от 62 до 88°C, от 63 до 87°C, от 64 до 86°C, от 65 до 85°C, от 66 до 84°C, от 67 до 83°C, от 68 до 82°C, от 69 до 81°C, от 70 до 80°C, от 71 до 79°C, от 72 до 78°C, от 73 до 77°C, от 74 до 76°C, 75°C. Предпочтительно, белки Cas по изобретению расщепляют, маркируют или модифицируют нуклеиновые кислоты при температурах в диапазоне от 60 до 80°C, от 61 до 79°C, от 62 до 78°C, от 63 до 77°C, от 64 до 76°C, от 65 до 75°C, от 66 до 74°C, от 67 до 73°C, от 68 до 72°C, от 69 до 71°C, от 70 до 70°C. Оптимально белки Cas согласно изобретению расщепляют, маркируют или модифицируют нуклеиновые кислоты при температурах в диапазоне от 60 до 65°C, предпочтительно при 65°C.

Во всех аспектах изобретения белки или полипептиды Cas могут быть получены или могут происходить из бактерий, археев или вирусов; или, альтернативно, могут быть синтезированы *de novo*. В предпочтительных вариантах осуществления белок или полипептид Cas по изобретению происходит из термофильного прокариотического организма, который может быть классифицирован как архей или бактерия, но предпочтительно является бактерией. Более предпочтительно белок или полипептид Cas согласно изобретению должен происходить из термофильной бактерии. В настоящем документе термин термофильный следует понимать как обозначающий способность к выживанию и росту при относительно высоких температурах, например, в контексте изобретения, способный расщеплять нуклеиновые кислоты при температуре между 41 и 122°C (106 и 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 согласно изобретению может быть выделен из одной или более термофильных бактерий и будет действовать в диапазоне температур от 60 до 80°C и оптимально между 60 и 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. Гид-РНК и последовательности-мишени Белки Cas по изобретению позволяют проводить специфичное для последовательности расщепление, мечение, маркировку или модификацию нуклеиновых кислот-мишеней при повышенных температурах. Нуклеиновыми кислотами-мишенями могут быть ДНК (одноцепочечные или двухцепочечные), РНК или синтетические нуклеиновые кислоты. Особенно полезным применением настоящего изобретения является специфическая для последовательности направленность и модификация геномной ДНК одним или более белками Cas по изобретению в комплексе с одной или более гид-РНК (gRNA), 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 действия и модификации специфических последовательностей ДНК, расположенных в геномной ДНК клетки. Предполагается, что систему Cas можно использовать для модификации геномов в различных типах клеток и/или у разных организмов.

Соответственно, настоящее изобретение относится к рибонуклеопротеидному комплексу, включающему белок или полипептид Cas по изобретению, как описано в настоящем документе выше, и дополнительно включающему, по меньшей мере, одну молекулу РНК, которая имеет направляющую функцию в том смысле, что она распознает конкретную нуклеотидную последовательность в полинуклеотиде-мишени. Предпочтительно, молекула РНК представляет собой одноцепочечную молекулу РНК, например, РНК CRISPR (crRNA) и связана, например, путем гибридизации с *tracrRNA*. Направляющая РНК может представлять собой химеру crRNA и *tracrRNA*. Вышеупомянутые молекулы РНК могут иметь рибонуклеотидную последовательность по меньшей мере на 90% идентичную нуклеотидной последовательности-мишени или комплементарную ей. Необязательно, молекула РНК имеет рибонуклеотидную последовательность, идентичную 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 или комплементарную ей. Предпочтительной нуклеотидной по-

следовательностью-мишенью является ДНК.

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

Изменение последовательности гид-РНК, которая связывается с белком Cas согласно изобретению, следовательно, позволяет белку Cas программироваться 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 интервале от 35 до 135 остатков, необязательно в интервале 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 110 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 остатка, 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 остатка, 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, где комплементарная часть по меньшей мере одной молекулы РНК имеет по меньшей мере 30 остатков в длину. Альтернативно, комплементарная часть по меньшей мере одной молекулы РНК может иметь 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 в длину.

Предпочтительно, чтобы молекула направляющей РНК удовлетворяла требованиям высокой специфичности и сродства 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, как это может быть определено нативным гель-электрофорезом, или, альтернативно, изотермической калориметрией титрования, поверхностным плазмонным резонансом или методами титрования на основе флуоресценции. Сродство может быть определено с использованием анализа сдвига электрофоретической подвижности (EMSA), также называемого анализом задержки в геле (смотри Semenova E et al. (2011) PNAS 108: 10098-10103).

Молекулу направляющей РНК предпочтительно моделируют на том, что известно в природе у прокариотов в качестве молекул РНК CRISPR (crRNA). Структура молекул crRNA уже установлена и объяснена более подробно в статье Jore et al., 2011, Nature Structural & Molecular Biology 18: 529-537. Вкратце, зрелая crRNA типа I-E часто составляет 61 нуклеотид в длину и состоит из 5'-области "ручки" 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 статье Van der Oost 2014 Nat Rev Microbiol 12: 479-492. В системах типа II (Cas9) существует отличный механизм процессинга, использующий вторую РНК (tracrRNA) и две рибонуклеазы. Зрелая crRNA в системе типа II, вместо шпильки, остается присоединенной к фрагменту tracrRNA (фиг. 5). Однако РНК, используемая в изобретении, не должна быть соз-



дана строго по дизайну природной crRNA, будь то в отношении длины, областей или конкретных последовательностей РНК. Становится понятно, что молекулы РНК для использования в изобретении могут быть сконструированы на основе информации о последовательностях генов, имеющейся в общедоступных базах данных, или недавно открытых, и затем сконструированы искусственно, например, полностью или частично с помощью химического синтеза. Молекулы РНК по изобретению также могут быть сконструированы 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 crRNA в системе типа II (Cas9) также описаны в статье Jinek et al., 2012, там же. В системе типа I имеется так называемая "SEED" часть, образующая 5'-конец спейсерной последовательности, и которая примыкает к этому 5'-концу с помощью 5'-ручки из 8 нуклеотидов. Semenova et al. (2011, PNAS 108: 10098-10103), обнаружили, что все остатки последовательности SEED должны быть комплементарны последовательности-мишени, хотя для остатка в положении 6 может быть допущено несоответствие (фиг. 5). В системе типа II имеется SEED из 10-12 нуклеотидов, которая расположена на 3'-конце спейсера (фиг. 5) (обсуждается Van der Oost 2014, там же). Аналогично, при конструировании и создании компонента РНК рибонуклеопротеидного комплекса по изобретению, направленного на локус-мишень (т.е. на последовательность), могут применяться необходимые правила соответствия и несоответствия для последовательности SEED системы типа II.

Таким образом, изобретение включает способ обнаружения и/или локализации изменения единственного основания в молекуле нуклеиновой кислоты-мишени, включающий контактирование 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 где последовательность направляющей РНК (в том числе, когда она находится в рибонуклеопротеидном комплексе) такова, что она различает нормальный аллель и мутантный аллель в силу изменения единственного основания, например, в положении 6 непрерывной последовательности из 8 нуклеотидных остатков.

Не желая связываться с конкретной теорией, правило создания, которое может быть использовано при получении компонента направляющей РНК рибонуклеопротеидных комплексов по изобретению, включает так называемую последовательность «PAM» (мотив, прилегающий к протоспейсеру) в полинуклеотидной цепи-мишени. Последовательность PAM может быть консервативным триплетом нуклеотидных остатков в системе типа I-E E.coli 5'-CTT-3', 5'-CAT-3', 5'-CCT-3', 5'-CAC-3', 5'-TTT-3', 5'-ATT-3' и 5'-AWG-3', где W представляет собой A, T или U. В системе типа I последовательность PAM, расположенная в цепи-мишени, находится обычно в положении, соответствующем 5'-концу SEED. Однако в системе типа II PAM расположена на другом конце, на цепи, смещенной близко к 3'-концу спейсерной crRNA, в положении, соответствующем 3'-концу SEED (фиг. 5) (Jinek et al., 2012, в цитируемой работе). Для Cas9 *Streptococcus pyogenes* последовательность PAM имеет консервативную пару нуклеотидных остатков 5'-NGG-3'. Недавно были охарактеризованы различные варианты Cas9 (тип IIA и тип IIC) (Ran et al., 2015 Nature 520: 186-191) - фиг. 1A), и были обнаружены PAMs (смотри, Ran et al., 2015, там же. - фиг. 1C). Установленные в настоящее время PAMs Cas9 включают: тип IIA 5'-NGGNNN-3' (*Streptococcus pyogenes*), 5'-NNGTNNN-3' (*Streptococcus pasteurianus*), 5'-NNGGAAN-3' (*Streptococcus thermophilus*), 5'-NNGGGNN-3' (*Staphylococcus aureus*), и тип IIC 5'-NGGNNNN-3' (*Corynebacterium difteriae*), 5'-NNGGGTN-3' (*Campylobacter lari*), 5'-NNNCATN-3' (*Parvobaculum lavamentivorans*), 5'-NNNNGTA-3' (*Neisseria cinerea*). Cas9 *Geobacillus thermodenitrificans* T12 (настоящее изобретение) относится к типу IIC (Ran et al., 2015, там же). Поэтому, хотя и не желая связываться с конкретной теорией, изобретатели ожидают, что последовательности PAM более тесно приближены к последовательностям, установленным для рибонуклеопротеидных комплексов типа IIC Cas9.

В вариантах осуществления изобретения молекула направляющей РНК может иметь длину в диапазоне 35-75 остатков. В предпочтительных вариантах осуществления часть РНК, которая комплементарна и используется для направления на желаемую последовательность нуклеиновой кислоты, составляет 31 или 32 остатка в длину. В контексте природной crRNA это должно соответствовать спейсерной части, как показано, например, на фиг. 1 статьи Semenova et al. (2011, там же).

Рибонуклеопротеидный комплекс по изобретению может иметь направляющий компонент, включающий 8 остатков, происходящих из повторов CRISPR, в направлении 5'-конца последовательности РНК, которая имеет существенную комплементарность последовательности ДНК-мишени. Последовательность РНК, имеющая комплементарность последовательности ДНК-мишени, как должно быть понятно, в соответствии с контекстом crRNA является спейсерной последовательностью. 5'-фланкирующая последовательность РНК должна рассматриваться как соответствующая 5'-ручке crRNA; как показано, например, на фиг. 1 в статье Semenova et al. (2011, там же).

Рибонуклеопротеидный комплекс по изобретению может иметь шпильку и тетрапуклеотидную петлю, образующие последовательность в направлении 3'-конца последовательности направляющей РНК, которая комплементарна последовательности ДНК-мишени, т.е. в направлении 3'-конца относительно того, что должно соответствовать 3'-ручке, примыкающей к спейсерной последовательности в

crRNA; например, как показано на фиг. 1 в статье Semenova et al. (2011, там же).

Не желая связываться с определенной теорией в предпочтительном рибонуклеопротеидном комплексе цепь ДНК последовательности-мишени, которая не спаривается с направляющей РНК рибонуклеопротеидного комплекса, может включать непосредственно прилегающий к 3'-концу триплет, выбранный из 5'-NGGNNNN-3', 5'-NNGTNNN-3', 5'-NNGGAAN-3', 5'-NNGGGNN-3', 5'-NGGNNNN-3', 5'-NNGGGTN-3', 5'-NNNCATN-3', 5'-NNNNGTA-3', и где указанные остатки триплета, каждый на спаривается с соответствующими комплементарными остатками цепи ДНК. Однако должно быть понятно, что в зависимости от желаемого применения могут использоваться другие триплеты или сочетания нуклеотидов. Эти триплеты соответствуют тому, что обозначается как "мотив, прилегающий к протоспейсеру" или "PAM" в контексте природных crRNA. В системах типа ПС CRISPR/Cas эти триплеты PAM требуются для стабильного взаимодействия комплекса Cascade/crRNA с его мишенью дцДНК, для того, чтобы обеспечить высокую степень специфичности crRNA - как в природных системах-мишенях, так и, следовательно, предпочтительно, также в РНК в соответствии с настоящим изобретением - для последовательности-мишени.

#### Функциональные части

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 маркировки, и/или ее модификации. Поэтому следует понимать, что для достижения этого могут быть предоставлены дополнительные белки совместно с белком или полипептидом Cas. Соответственно, белки, полипептиды или рибонуклеопротеидные комплексы Cas по настоящему изобретению могут быть представлены как часть белкового комплекса, включающего по меньшей мере один дополнительный белок. В предпочтительном аспекте настоящее изобретение относится к белку, полипептиду или рибонуклеопротеидному комплексу Cas, где белок Cas или по меньшей мере один дополнительный белок дополнительно включает по меньшей мере одну функциональную часть. По меньшей мере одна функциональная часть может быть слита или связана с белком Cas. Предпочтительно по меньшей мере одна функциональная часть может быть трансляционно слита с белком Cas путем экспрессии в природных или искусственных системах экспрессии белка. Альтернативно по меньшей мере одна функциональная часть может быть ковалентно связана с белком Cas на стадии химического синтеза. Предпочтительно по меньшей мере одна функциональная часть слита или связана с N-концом и/или C-концом белка Cas; предпочтительно с N-концом.

Желательно, чтобы по меньшей мере одна функциональная часть представляла собой белок. Это может быть гетерологичный белок или, альтернативно, это может быть природный белок для бактериальных видов, из которых происходит белок Cas. По меньшей мере одна функциональная часть может быть белком; необязательно выбранным из геликазы, нуклеазы, геликазы-нуклеазы, ДНК-метилазы, гистон-метилазы, ацетилазы, фосфатазы, киназы, (ко-)активатора транскрипции, репрессора транскрипции, ДНК-связывающего белка, ДНК-структурирующего белка, маркерного белка, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антитела или маркера аффинной очистки.

В особенно предпочтительном аспекте настоящее изобретение относится к белку, полипептиду или рибонуклеопротеидному комплексу Cas, где по меньшей мере одна функциональная часть представляет собой маркерный белок, например GFP.

#### Нуклеазная активность

Рибонуклеопротеид Cas по любому аспекту изобретения обладает расщепляющей нуклеиновую кислоту активностью между 50 и 100°C. Рибонуклеопротеиды по изобретению способны расщеплять ДНК, РНК или синтетические нуклеиновые кислоты. В предпочтительных аспектах рибонуклеопротеиды Cas по изобретению способны расщеплять ДНК специфичным для последовательности способом, в частности двухцепочечную ДНК.

В любом аспекте изобретения белки, полипептиды или рибонуклеопротеиды Cas по изобретению могут иметь более одного нуклеазного домена. Сайт-специфические нуклеазы могут дать возможность генерировать двухцепочечные разрывы (DSB) в выбранных положениях вдоль цепи ДНК. В клетке-хозяине мишени это позволяет делать DSBs в определенных заранее выбранных положениях в геноме. Создание таких разрывов сайт-специфическими нуклеазами побуждает эндогенный механизм клеточной репарации к переориентации для того, чтобы вставить, удалить или модифицировать ДНК в желаемых положениях в интересующем геноме.

Один или более сайтов нуклеазной активности молекулы белка или полипептида могут быть инактивированы, например, так, чтобы дать возможность реализоваться активности другой функциональной части, связанной или слитой с белком или полипептидом, например, нуклеазы FokI.

Следовательно, несмотря на тот факт, что белки, полипептиды и рибонуклеопротеиды Cas по изобретению обладают 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 инактиви-

руется, и белок Cas связан, по меньшей мере, с одной функциональной частью. Одним из таких применений является снижение частоты случаев событий неправильной направленности путем комплементации естественной активности нуклеазы Cas9. Это может быть достигнуто желательно путем инактивации природной нуклеазной активности Cas9 белка или рибонуклеопротеидного комплекса Cas и обеспечения гетерологичной нуклеазой, предпочтительно слитой с белком Cas. Соответственно, настоящее изобретение относится к белку или рибонуклеопротеидному комплексу Cas, где по меньшей мере одна функциональная часть является нуклеазным доменом, предпочтительно нуклеазным доменом FokI. В особенно предпочтительном аспекте белок или рибонуклеопротеидный комплекс Cas по изобретению, слитый с нуклеазным доменом FokI, предлагается как часть белкового комплекса, предпочтительно включающего другой белок или рибонуклеопротеидный комплекс Cas по изобретению, слитый с нуклеазным доменом FokI, и где 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 используются для распознавания и модификации конкретной последовательности-мишени в нуклеиновой кислоте, например, для ее метки как части диагностического теста. В таких вариантах применения нуклеазная активность белка Cas может быть инактивирована, и функциональная часть, слитая 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, белка-маркера, репортерного белка, флуоресцентного белка, лигандсвязывающего белка, сигнального пептида, последовательности субклеточной локализации, эпитопа антигена или метки аффинной очистки.

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 одна функциональная часть представляет собой маркерный белок, например GFP. Таким путем становится возможной специфическая направленность на интересующую последовательность нуклеиновой кислоты и ее визуализация с использованием маркера, который генерирует оптический сигнал. Подходящие маркеры могут включать, например, флуоресцентный репортерный белок, например зеленый флуоресцентный белок (GFP), желтый флуоресцентный белок (YFP), красный флуоресцентный белок (RFP), голубой флуоресцентный белок (CFP) или mCherry. Такой флуоресцентный репортерный ген обеспечивает подходящим маркером для визуализации экспрессии белка, поскольку его экспрессия может быть просто и непосредственно определена путем измерения флуоресценции. Альтернативно, репортерная нуклеиновая кислота может кодировать люминесцирующий белок, такой как люцифераза (например, люцифераза светлячков). Альтернативно, репортерный ген может представлять собой хромогенный фермент, который может быть использован для генерирования оптического сигнала, например, хромогенный фермент (такой как бета-галактозидаза (LacZ) или бета-глюкуронидаза (Gus)). Репортеры, используемые для измерения экспрессии, также могут представлять собой пептидные метки антигенов. Другие репортеры или маркеры известны в данной области техники, 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. Поэтому в предпочтительных вариантах осуществления изобретения белки или рибонуклеопротеиды Cas по изобретению могут быть использованы для анализа продуктов экспрессии гена, представляющего интерес.

Везде эталонные последовательности белков Cas по изобретению могут быть определены как нуклеотидная последовательность, кодирующая аминокислотную последовательность. Например, аминокислотная последовательность мотивов, определенных в SEQ ID с 2 по 6, также включает все последовательности нуклеиновой кислоты, которые кодируют эту аминокислотную последовательность.

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

- 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 выбран из изолейцина, метионина или пролина,  $X_2$  независимо выбран из валина, серина, аспарагина или изолейцина,  $X_3$  независимо выбран из глутамата или лизина и  $X_4$  представляет собой одно из аланина, глутамата или аргинина; и/или
- в) аминокислотный мотив  $X_5LXX_6IE$  [SEQ ID NO: 4], где  $X_5$  независимо выбран из метионина или фенилаланина и  $X_6$  независимо выбран 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;

где белок Cas способен расщеплять ДНК между 50 и 100°C при связывании по меньшей мере с одной направляющей молекулой РНК и полинуклеотидом, включающим последовательность нуклеиновой кислоты-мишени, распознаваемой направляющей молекулой РНК.

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

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

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

#### Экспрессионные векторы

Нуклеиновые кислоты по настоящему изобретению могут быть выделены. Однако для того, чтобы экспрессия распознаваемого конструкта нуклеиновой кислоты могла быть осуществлена в выбранной клетке, полинуклеотидная последовательность, кодирующая белок или рибонуклеопротеид Cas, предпочтительно должна быть представлена в экспрессирующемся конструкте. В некоторых вариантах осуществления полинуклеотид, кодирующий белок или рибонуклеопротеид Cas, может быть предоставлен как часть 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 включать нуклеотидную последовательность, кодирующую молекулу направляющей РНК, как определено в настоящем описании выше. Следовательно, такие экспрессионные векторы могут быть использованы в подходящем хозяине для генерации рибонуклеопротеидного комплекса по изобретению, который может быть направлен на желаемую нуклеотидную последовательность. Альтернативно, нуклеотидные последовательности, кодирующие молекулу направляющей РНК, как определено в настоящем описании выше, могут быть представлены в отдельном экспрессионном векторе или, альтернативно, могут быть доставлены в клетку-мишень другими способами.

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

Такие элементы могут включать, например, сильные и/или конститутивные промоторы, 5'- и 3'-UTR's, энхансеры транскрипции и/или трансляции, последовательности транскрипционного фактора или связывающего белка, 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-сигналов, обеспечивающих терминацию транскрипции и стабилизацию транскрипта в клетке-хозяине. Регуляторные последовательности могут происходить от растений, животных, бактерий, грибов или вирусов и предпочтительно могут происходить из того же организма, что и клетка-хозяин. Очевидно, что подходящие регуляторные элементы будут варьироваться в зависимости от интересующей клетки-хозяина. Например, регуляторные элементы, которые способствуют высокому уровню экспрессии в клетках-хозяевах прокариот, таких как *E. coli*, могут включать промоторы pLac, T7, P(Bla), P(Cat), P(Kat), trp или tac. Регуляторные элементы, которые способствуют высокому уровню экспрессии в клетках-хозяевах эукариот, могут включать промотор AOX1 или GAL1 в дрожжах или CMV-или SV40-промоторы, CMV-энхансер, SV40-энхансер, активатор транскрипции вируса простого герпеса VIP или включение интрона глобина в клетках животных. В растениях высокий уровень конститутивной экспрессии может быть получен, например, с помощью промотора убиквитина 1 *Zea mays* или промоторов 35S и 19S вируса мозаики цветной капусты.

Подходящие регуляторные элементы могут быть конститутивными, в результате чего они управляют экспрессией при большинстве условий окружающей среды или стадий развития, специфичных или индуцируемых стадий развития. Предпочтительно, промотор является индуцибельным для управления экспрессией в ответ на сигналы окружающей среды, химические сигналы или сигналы развития, такие

как температура, свет, химикаты, засуха и другие стимулы. Промоторы могут быть выбраны подходящим образом, что разрешает экспрессию интересующего белка на определенных стадиях развития или в ответ на вне- или внутриклеточные условия, сигналы или внешние стимулы. Например, существует ряд промоторов для использования в *E. coli*, которые вызывают высокий уровень экспрессии на определенных стадиях роста (например, промотор *osmY* стационарной фазы) или в ответ на конкретные стимулы (например, промотор *HtrG* теплового шока).

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

Изобретение также включает способ модификации нуклеиновой кислоты-мишени в клетке, включающий трансфекцию, трансформацию или трансдукцию клетки любым из экспрессионных векторов, как описано в настоящем документе выше. Способы трансфекции, трансформации или трансдукции относятся к типам, хорошо известным специалисту в данной области техники. Когда имеется один экспрессионный вектор, используемый для индукции экспрессии рибонуклеопротеидного комплекса по изобретению, и когда направляющую РНК добавляют непосредственно в клетку, то может быть использован тот же самый или другой способ трансфекции, трансформации или трансдукции. Аналогичным образом, когда имеется один экспрессионный вектор, используемый для индукции экспрессии рибонуклеопротеидного комплекса по изобретению, и когда другой экспрессионный вектор используется для генерации направляющей РНК *in situ* посредством экспрессии, то может быть использован тот же самый или другой метод трансфекции, трансформации или трансдукции.

В других вариантах осуществления мРНК, кодирующая белок или полипептид Cas, вводится в клетку, так что комплекс Cascade экспрессируется в клетке. Гид-РНК, которая направляет белковый комплекс Cas к желаемой последовательности-мишени, также вводится в клетку либо одновременно, отдельно, либо последовательно после мРНК, так что в клетке образуется необходимый рибонуклеопротеидный комплекс.

Соответственно, в изобретении также предлагается способ модификации, то есть расщепления, метки, маркирования или связывания, нуклеиновой кислоты-мишени, включающий контактирование нуклеиновой кислоты с рибонуклеопротеидным комплексом, как определено в настоящем документе выше.

Кроме того, изобретение также включает способ модификации нуклеиновой кислоты-мишени, включающий контактирование нуклеиновой кислоты с белком или полипептидом Cas, как определено в настоящем документе выше, в дополнение к молекуле направляющей РНК, как определено в настоящем документе выше.

В соответствии с вышеприведенными способами модификация нуклеиновой кислоты-мишени, следовательно, может проводиться *in vitro* и в бесклеточной среде. В бесклеточной среде добавление каждой нуклеиновой кислоты-мишени, белка Cas и молекулы направляющей РНК может быть одновременным, последовательными (в любом порядке по желанию) или отдельным. Таким образом, нуклеиновая кислота-мишень и направляющая РНК могут быть добавлены к реакционной смеси одновременно, и затем белок или полипептид Cas по изобретению добавляется отдельно на более поздней стадии.

В равной мере модификация нуклеиновой кислоты-мишени может быть осуществлена *in vivo*, то есть в клетке *in situ*, будь то выделенная клетка или часть многоклеточной ткани, органа или организма. В контексте цельной ткани и органов и в контексте организма способ может быть осуществлен желательно *in vivo* или, альтернативно, может быть осуществлен путем выделения клетки из целой ткани, органа или организма, обработки клетки рибонуклеопротеидным комплексом в соответствии со способом и последующим возвратом клетки, обработанной рибонуклеопротеидным комплексом, в ее прежнее местоположение или в другое место, независимо от того, будет ли она находиться в том же или в другом организме.

В этих вариантах осуществления рибонуклеопротеидный комплекс или белок или полипептид Cas требует подходящей формы доставки в клетку. Такие подходящие системы и способы доставки хорошо известны специалистам в данной области техники и включают, но не ограничиваются этим, цитоплазматическую или ядерную микроинъекцию. В предпочтительных способах доставки используется аденоассоциированный вирус (AAV); эта система доставки не вызывает болезни у людей и была одобрена для клинического применения в Европе.

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

- a) с рибонуклеопротеидным комплексом, как определено в настоящем документе выше; или
- b) с белком или белковым комплексом, как определено в настоящем документе выше, и молекулой РНК, как определено в настоящем документе выше.

В следующем аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени в клетке, включающему трансформацию, трансфекцию или трансдукцию клетки экспрессионным вектором, включающим нуклеотидные последовательности, кодирующие рибонуклеопротеидный комплекс, как определено в настоящем документе выше; или альтернативно трансформацию, транс-

фекцию или трансдукцию клетки экспрессионным вектором, включающим нуклеотидные последовательности, кодирующие белок или белковый комплекс, как определено в настоящем документе выше, и дополнительным экспрессионным вектором, включающим нуклеотидную последовательность, кодирующую молекулу направляющей РНК, как определено в настоящем документе выше.

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

В тех вариантах осуществления, когда молекула гид-РНК (gRNA) (т.е. направляющей РНК) и белок или полипептид Cas предоставляются отдельно, а не как часть рибонуклеопротеидного комплекса, молекула gRNA требует подходящей формы доставки в клетку, будь то одновременная, отдельная или последовательная доставка с белком или белковым комплексом Cas. Такие формы введения РНК в клетки хорошо известны специалисту в данной области техники и могут включать доставку *in vitro* или *ex vivo* с помощью обычных методов трансфекции. Может быть использован каждый из физических методов, таких как микроинъекция и электропорация, а также соосаждение кальцием и коммерчески доступные катионные полимеры и липиды, а также проникающие в клетку пептиды, проникающие в клетку (биолитические) частицы. Например, в качестве носителей для доставки будь то в цитоплазму и/или в ядро могут быть использованы вирусы, особенно предпочтительные представляют собой 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, где по меньшей мере одна функциональная часть представляет собой маркерный белок или репортерный белок, и маркерный белок или репортерный белок связан с нуклеиновой кислотой-мишенью; предпочтительно, где маркер представляет собой флуоресцентный белок, например зеленый флуоресцентный белок (GFP).

В вышеупомянутых способах модификации нуклеиновой кислоты-мишени функциональная часть может представлять собой маркер, и маркер связан с нуклеиновой кислотой-мишенью; предпочтительно, где 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 непосредственной визуализации локуса-мишени в молекуле нуклеиновой кислоты, предпочтительно в форме структуры более высокого порядка, такой как суперспирализованная плазида или хромосома, или одноцепочечная нуклеиновая кислота, такая 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 использоваться 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, где нуклеиновая кислота-мишень представляет собой дцДНК, функциональная часть может представлять собой нуклеазу или геликазу-нуклеазу, и модификация предпочтительно представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе. Таким путем может быть сконструировано уникальное специфичное для последовательности разрезание ДНК с использованием подходящей функциональной части, слитой с рибонуклеопротеидным комплексом. Выбранная последовательность компонента РНК конечного рибонуклеопротеидного комплекса обеспечивает желаемую специфичность последовательности для действия функциональной части.

Таким образом, в изобретении также предлагается способ негомологичного соединения концов молекулы дцДНК в клетке в желаемом локусе для удаления, по меньшей мере, части нуклеотидной последовательности из молекулы дцДНК; необязательно для нокаутирования функции гена или генов, где способ включает создание двухцепочечных разрывов с использованием любого из способов модификации нуклеиновой кислоты-мишени, как описано в настоящем документе выше.

В изобретении дополнительно предлагается способ гомологичной рекомбинации нуклеиновой кислоты в молекуле дцДНК в клетке в желаемом локусе для модификации существующей нуклеотидной последовательности или вставки желаемой нуклеотидной последовательности, где способ включает создание двухцепочечного разрыва в желаемом локусе с использованием любого из способов модификации нуклеиновой кислоты-мишени, как описано в настоящем документе выше.

Таким образом, в изобретении также предлагается способ модификации экспрессии генов в организме, включающий модификацию последовательности нуклеиновой кислоты-мишени в соответствии с любым из описанных в настоящем документе выше способов, где нуклеиновая кислота представляет собой дцДНК, и функциональная часть выбрана из фермента, модифицирующего ДНК (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции.

В изобретении дополнительно предлагается способ модификации экспрессии генов в организме,

включающий модификацию последовательности нуклеиновой кислоты-мишени согласно любому из описанных выше в настоящем документе способов, где нуклеиновая кислота представляет собой мРНК, а функциональная часть представляет собой рибонуклеазу; необязательно выбранную из эндонуклеазы, 3'-экзонуклеазы или 5'-экзонуклеазы.

В любом аспекте способов по изобретению, описанных в настоящем документе, 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. Альтернативно, следовательно, в настоящем изобретении также предлагаются способы модификации нуклеиновой кислоты-мишени, где нуклеиновая кислота-мишень представляет собой РНК.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени, где нуклеиновая кислота представляет собой дцДНК, по меньшей мере одна функциональная часть представляет собой нуклеазу или геликазу-нуклеазу, и модификация представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе.

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

- a) создания двухцепочечных разрывов в молекуле дцДНК; и
- b) репарации молекулы дцДНК в клетке путем негомологичного соединения концов (NHEJ).

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

- a) создания двухцепочечного разрыва в желаемом локусе; и
- b) репарации молекулы дцДНК в клетке путем гомологичной рекомбинации.

В другом аспекте настоящее изобретение относится к способу модификации экспрессии генов в клетке, включающему модификацию последовательности нуклеиновой кислоты-мишени, как описано в настоящем документе выше, где нуклеиновая кислота представляет собой дцДНК, и функциональная часть выбрана из модифицирующего ДНК фермента (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции.

В другом аспекте в настоящем изобретении предлагается способ модификации экспрессии генов в клетке, включающий модификацию последовательности нуклеиновой кислоты-мишени, как описано в настоящем документе выше, где нуклеиновая кислота представляет собой мРНК, и функциональная часть представляет собой рибонуклеазу; необязательно выбранную из эндонуклеазы, 3'-экзонуклеазы или 5'-экзонуклеазы.

В другом аспекте настоящее изобретение относится к способу модификации нуклеиновой кислоты-мишени, как описано в настоящем документе выше, где способ осуществляют при температуре от 50 до 100°C. Предпочтительно способ осуществляют при температуре, равной или выше 60°C. Более предпочтительно, способ осуществляют при температуре от 60 до 80°C. Оптимально, способ осуществляют при температуре от 60 до 65°C.

В любом из способов модификации нуклеиновой кислоты-мишени, описанных в настоящем документе выше, клетка может быть прокариотической клеткой или, альтернативно, может быть эукариотической клеткой.

#### Клетки-хозяева

Выгодно, что настоящее изобретение имеет широкую применимость, и клетки-хозяева по настоящему изобретению могут происходить из любого генетически приемлемого организма, который может культивироваться. Соответственно, в настоящем изобретении предлагается клетка-хозяин, трансформированная способом, описанным в настоящем документе выше.

Соответствующие клетки-хозяева могут быть прокариотическими или эукариотическими. В частности, для использования в соответствии с настоящим изобретением могут быть выбраны обычно используемые клетки-хозяева, включая прокариотические или эукариотические клетки, которые являются генетически доступными и которые могут культивироваться, например прокариотические клетки, клетки грибов, растительные клетки и клетки животных, включая клетки человека (но не эмбриональные стволовые клетки). Предпочтительно клетки-хозяева должны быть выбраны из прокариотической клетки, клетки грибов, растительной клетки, клетки простейших или клетки животных. Предпочтительные клетки-хозяева для использования в соответствии с настоящим изобретением обычно происходят от видов, которые обычно демонстрируют высокие темпы роста, легко культивируются и/или трансформируются, характеризуются коротким временем генерации, от видов, которые имеют установленные, связанные с ними генетические ресурсы, или от видов с селекцией, модификацией или синтетическими особенностями для оптимальной экспрессии гетерологичного белка в определенных условиях. В предпочтительных вариантах осуществления изобретения, где интересующий белок в конечном итоге должен использоваться в конкретных промышленных, сельскохозяйственных, химических или терапевтических целях, подхо-

дующая клетка-хозяин может быть выбрана на основе желаемых конкретных условий или клеточного контекста, в которых предполагается использовать интересующий белок. Предпочтительно клетка-хозяин должна представлять собой прокариотическую клетку. В предпочтительных вариантах осуществления клетка-хозяин представляет собой бактериальную клетку. Клеткой-хозяином может быть, например, клетка *Escherichia coli* (*E. coli*). Предпочтительно клетка-хозяин должна представлять собой клетку термофильной бактерии.

#### Краткое описание чертежей

Изобретение будет теперь подробно описано со ссылкой на конкретный вариант осуществления и со ссылкой на прилагаемые фигуры, на которых:

на фиг. 1 показано дерево последовательностей белков Cas9, построенное методом ближайших соседей. Были включены все последовательности, имеющие сходство последовательностей со штаммом T12 выше 40% на основе pBLAST или PSI-BLAST, а также хорошо охарактеризованные в настоящее время последовательности (*S. pyogenes*, *S. thermophilus* и *A. naeslundii*), а также все идентифицированные в настоящее время термофильные последовательности при их идентичности ниже 40%. Для всех термофильных последовательностей процент идентичности с T12 указан после названия штамма. Количество идентификаторов генов (gi) указан перед названием штамма. Надпись: заполненные кружки: термофильные (оптимум при температуре выше 60°C) последовательности Cas9, заполненные квадраты: термотолерантные (оптимум <50°C) последовательности Cas9, пустой треугольник: последовательность Cas9, наиболее часто используемая в настоящее время в целях редактирования генома, мезофильного происхождения; без символа: Cas9 мезофилов. Величины в узлах представляют собой значения бутстрап-анализа 1000-репликатов; измерительная шкала представляет определенные аминокислотные замены на сайт;

на фиг. 2 показано дерево последовательностей генов Cas9, построенное методом ближайших соседей. Идентичность на уровне генов была крайне бедна, для выравнивания генов использовали последовательности из тех же организмов, что и используемые для выравнивания белков. Количество идентификаторов генов (gi) указано перед названием штамма. Надпись: заполненные кружки: термофильные (оптимум при температуре выше 60°C) последовательности Cas9, заполненные квадраты: термотолерантные (оптимум <50°C) последовательности Cas9, пустой треугольник: последовательность Cas9, наиболее часто используемая в настоящее время в целях редактирования генома, мезофильного происхождения, без символа: Cas9 мезофилов. Величины в узлах представляют собой значения бутстрап-анализа 1000-репликатов;

на фиг. 3 показано выравнивание белковой последовательности T12-Cas9 (SEQ ID NO: 1) (тип II-C) с хорошо охарактеризованными последовательностями Cas9 типа II-C (*A. naeslundii*/'ana'; SEQ ID NO: 8) и типа II-A (*S. pyogenes*/'pyo'; SEQ ID NO: 9 и *S. thermophilus*). Важные остатки активного сайта высоко консервативны и указаны черными стрелками. Домены белка, как описано для Ana-Cas9 и Pyo-Cas9 (Jinek, et al., 2014, Science 343: 1247997), указаны заштрихованными боксами и сходно окрашенными буквами. Домен узнавания PAM определен для системы типа II-A *S. pyogenes*, но не для любой системы типа II-C и поэтому указан только в последовательности *S. Pyogenes*;

на фиг. 4 показана белковая структура Cas9 *A. naeslundii* (Cas9-Ana) (Jinek et al., 2014). T12-Cas9 относится к той же самой системе CRISPR типа II-C, и остатки активного сайта могут быть идентифицированы;

на фиг. 5 показано сравнение направленности действия гид-crRNA на комплементарную дцДНК. Пары оснований обозначены пунктирными линиями. РНК изображена черным цветом, ДНК - серым. Пары оснований между спейсером crRNA и протоспейсером-мишенью обозначены толстой черной пунктирной линией, пары оснований между цепями ДНК и между цепями РНК обозначены толстыми серыми пунктирными линиями. Указан 5'-конец crRNA. 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, seed (предсказанная последовательность гида, где начинается спаривание оснований с ДНК-мишенью, и где не допускается никаких ошибочных спариваний) находится близко к PAM и, как таковая, отличается у типов I и II (Van der Oost, 2014 там же). На панели А показана схема системы Cascade типа I *E. coli*. crRNA имеет внутренний спейсер (серый бокс, 31-32 нуклеотида, который позволяет узнавать мишень), к которому примыкают bt из 8 нуклеотидов 5'-ручки, и 3'-ручка из 29 нуклеотидов, которая представляет собой структуру типа стебель-петля (шпилька) (Jore 2011 там же). На панели В схематически показана система Cas9 типа II *S. pyogenes*. crRNA образует пары оснований с tracrRNA, что позволяет осуществлять процессинг RNaseIII (противоположно направленные черные треугольники). Кроме того, 5'-конец crRNA обрезается с помощью РНКазы (черный треугольник), что обычно приводит к 20-нуклеотидному спейсеру. Обратите внимание на то, что синтетическая петля может быть введена для связывания crRNA и tracrRNA, в результате чего образуется единая гид-РНК (sgRNA) (Jinek et al., 2012 там же).

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



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

MKYKIGLDIGITSIGWAVINLDIPRIEDLGVRI FDRAENPKTGESLALPRRLARSARRR  
 LRRRKHRLERIRRLFVREGILTKEELNKLFEKKHEIDVWQLRVEALDRKLNDELARILLHLAK  
 RRGFRSNRKSERTNKENSTMLKHIEENQSILSSYRTVAEMVVKDPKFSLHKRNKEDNYTNTVAR  
 DDLEREIKLIFAKQREYGNIVCTEAFEHEYISIWASQRPFASKDDIEKKVGFCTFEPKEKRAPK  
 ATYTFQSFTVWEHINKLRLVSPGGIRALTDDEERRLIYKQAFHKNKITFHDVRTLLNLPDDTRFK  
 GLLYDRNTTLKENEKVRFLELGAYHKIRKAIDSVYKGGAASFRPIDFDFTFGYALTMFKDDTDI  
 RSYLRNEYEQNGKRMENLADKVYDEELIEELLNLSFSKFGHLSLKALRNILPYMEQGEVYSTAC  
 ERAGYFTFGPKKKQKTVLLPNIPPIANPVVMRALTQARKVVNAI IKKYGSPVSIHIELARELSQ  
 SFDERRKMKEQEGNRKKNETAIRQLVEYGLTLNPTGLDIVKFKLWSEQNGKCAYSLOPIEIER  
 LLEPGYTEVDHVIPYSRSLDDSYTNKVLVLTKENREKGNRTPAEYLGSGSERWQQFETFVLTKN  
 QFSKKKRDRLLRLHYDENEENEFKNRNLNDTRYISRFLANFIREHLKFADSDDKQKVYTVNGRI  
 TAHLRSRWNFNKNREESNLHHAVDAAIVACTTPSDIARVTAIFYQRREQNKELSKKTDPPQFPQW  
 PHFADELQARLSKNPKESIKALNLGNYDNEKLESLQPVFVSRMPKRSITGAAHQETLRRYIGID  
 ERSKGIQTVVKKKLEIQLDKTHGFPMYKESDPRTYEAIRQRLLEHNNDPKKAFQEPYKPKK  
 NGELGPIIRTIKIIDTTNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPYITIDMMKGLPN  
 KAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIEFPREKTIKTAVGEEIKIKDLFAYYQIDSS  
 NGGLSLVSHDNFSLRSIGSRTLKRFEKYQVDVLGNIYKVRGEKRVGVASSSHSKAGETIRPL\*

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

ATGAAGTATAAAATCGGTCTTGATATCGGCATTACGTCTATCGGTTGGGCTGTCATTAA  
 TTTGGACATTCCTCGCATCGAAGATTTAGGTGTCCGCATTTTTGACAGAGCGGAAAACCCGAAA  
 ACCGGGGAGTCACTAGCTCTTCCACGTCGCCTCGCCCGCTCCGCCGACGTCGTCTGCGGCGTC  
 GCAAACATCGACTGGAGCGCATTCGCCGCTGTTCTGTCGCGAAGGAATTTAACGAAGGAAGA  
 GCTGAACAAGCTGTTTGAAGAAAAGCAGCAAAATCGACGCTCGGCAGCTTCGTGTTGAAGCACTG  
 GATCGAAAATAAATAACGATGAATTAGCCCGCATCTTCTTCATCTGGCTAAACGGCGTGGAT  
 TTAGATCCAACCGCAAGAGTGAGCGCACCAACAAGAAAACAGTACGATGCTCAAACATATTGA  
 AGAAAACCAATCCATTTCTTCAAGTTACCGAACGGTTGCAGAAATGGTTGTCAAGGATCCGAAA  
 TTTTCCCTGCACAAGCGTAATAAAGAGGATAATTACACCAACACTGTTGCCCGGACGATCTTG  
 AACGGGAAATCAAACGATTTTCGCCAACAGCGCGAATATGGGAACATCGTTTGCACAGAAGC  
 ATTTGAACACGAGTATATTTCCATTTGGGCATCGCAACGCCCTTTTGCTTCTAAGGATGATATC  
 GAGAAAAAGTCGGTTTCTGTACGTTTGTAGCCTAAAGAAAAACGCGCCAAAAGCAACATACA  
 CATTCAGTCTTCCCGTCTGGGAACATATTAACAACTTCGTCTGTCTCCCGGGAGGCAT  
 CCGGGCACTAACCGATGATGAACGTCGTCTTATATACAAGCAAGCATTCATAAAAAATAAAATC  
 ACCTTCCATGATGTTCGAACATTGCTTAACTTGCCGACGACACCCGTTTTAAAGGCTTTTTAT  
 ATGACCGAAACACCACGCTGAAGGAAAATGAGAAAAGTTCGCTTCTTGAACGCGCCATCA  
 TAAAATACGGAAAGCGATCGACAGCGTCTATGGCAAAGGAGCAGCAAAATCATTTCTGTCGATT  
 GATTTTGATACATTTGGCTACGCATTAACGATGTTAAAGACGACACCGACATTCGCAGTTACT  
 TGCGAAACGAATACGAACAAAATGGAAAACGAATGGAAAATCTAGCGGATAAAGTCTATGATGA  
 AGAATTGATTGAAGAACTTTTAACTTATCGTTTTCTAAGTTGGTCATCTATCCCTTAAAGCG  
 CTTCGCAACATCTTCCATATATGGAACAAGGGAAGTCTACTCAACCGCTGTGAACGAGCAG  
 GATATACATTTACAGGGCCAAAGAAAAACAGAAAACGGTATTGCTGCCGAACATTCGCCGAT  
 CGCCAATCCGGTCGTCATGCGCGCACTGACACAGGCACGCAAGTGGTCAATGCCATTATCAA  
 AAGTACGGCTCACCGTCTCCATCCATATCGAACTGGCCCGGAACTATCACAATCTTTGATG  
 AACGACGTAATGCAGAAAGAACAGGAAGGAAACCGAAAGAAAACGAACTGCCATTCGCCA  
 ACTTGTGTAATATGGGCTGACGCTCAATCCAACGTTGGCTTGCATTTGAAATTCAACTATGG  
 AGCGAACAAAACGGAAAATGTGCCTATTCCTCAACCGATCGAAATCGAGCGGTTGCTCGAAC  
 CAGGCTATACAGAAGTCGACCATGTGATTCATACAGCCGAAGCTTGGACGATAGCTATACCAA

TAAAGTTCCTGTGTTGACAAAGGAGAACCGTGAAAAAGGAAACCGCACCCAGCTGAATATTTA  
GGATTAGGCTCAGAACGTTGGCAACAGTTCGAGACGTTTGTCTTGACAAATAAGCAGTTTTTCGA  
AAAAGAAGCGGGATCGACTCCTTCGGCTTCATTACGATGAAAACGAAGAAAATGAGTTAAAAA  
TCGTAATCTAAATGATACCCGTTATATCTCACGCTTCTTGGCTAACTTTATTCGCGAACATCTC  
AAATTCGCCGACAGCGATGACAAACAAAAAGTATACACGGTCAACGGCCGTATTACCGCCCAT  
TACGCAGCCGTTGGAATTTTAAACAAAAACCGGGAAGAATCGAATTTGCATCATGCCGTGCATGC  
TGCCATCGTCGCCTGCACAACGCCGAGCGATATCGCCCGAGTCACCGCCTTCTATCAACGGCGC  
GAACAAACAAAGAACTGTCCAAAAAGACGGATCCGCAGTTTCCGCAGCCTTGGCCGCACTTTG  
CTGATGAACTGCAGGCGGTTTATCAAAAAATCCAAAGGAGAGTATAAAAGCTCTCAATCTTGG  
AAATATGATAACGAGAAACTCGAATCGTTGCAGCCGTTTTTTGTCTCCCGAATGCCGAAGCGG  
AGCATAACAGGAGCGGCTCATCAAGAAACATTGCGGCGTTATATCGGCATCGACGAACGGAGCG  
GAAAAATACAGACGGTCGTCAAAAAGAACTATCCGAGATCCAACGGATAAAAACAGGTCATTT  
CCCAATGTACGGGAAAGAAAGCGATCCAAGGACATATGAAGCCATTCGCCAACGGTTGCTTGAA  
CATAACAATGACCCAAAAAGGCGTTTTCAAGAGCCTCTGTATAAACCGAAGAAGAACGGAGAAC  
TAGGTCCTATCATCCGAACAATCAAAATCATCGATACGACAAATCAAGTTATTCGCTCAACGA  
TGGCAAAACAGTCGCCTACAACAGCAACATCGTGCAGGTCGACGCTTTTGAGAAAGATGGCAAA  
TATTATTGTGTCCCTATCTATACAATAGATATGATGAAAGGGATCTTGCCAAACAAGCGCATCG  
AGCCGAACAAACCGTACTCTGAGTGGAAAGGAAATGACGGAGGACTATACATTCGGATTCAGTCT  
ATACCCAAATGATCTTATCCGTATCGAATTTCCCGGAGAAAAACAATAAAGACTGCTGTGGGG  
GAAGAAATCAAATTAAGGATCTGTTCCCTATATCAAAACCATCGACTCCTCCAATGGAGGGT  
TAAGTTTGGTTAGCCATGATAACAATTTTCGCTCCGCAGCATCGGTTCAAGAACCCTCAAACG  
ATTTCGAGAAATACCAAGTAGATGTGCTAGGCAACATCTACAAAGTGAGAGGGGAAAAAGAGATT  
GGGGTGGCGTCATCTTCTCATTCGAAAGCCGGGAAACTATCCGTCGGTTATAA

#### Подробное описание

##### Пример 1. Выделение *G. thermodenitrificans*

*G. thermodenitrificans* был неожиданно обнаружен при поиске в библиотеке из  $\pm 500$  изолятов термофила, способного разрушать лигноцеллюлозные субстраты в анаэробных условиях. Сначала была создана библиотека из  $\pm 500$  изолятов, которая после нескольких циклов селекции путем выделения на целлюлозе и ксилане, была урезана до 110 изолятов. Эта библиотека из 110 изолятов состояла исключительно из изолятов *Geobacillus* с *G. thermodenitrificans*, представляющих 79% библиотеки.

Выделенный штамм *G. thermodenitrificans* был назван "T12".

##### Пример 2. Определение основных консенсусных последовательностей Cas9 у *Geobacillus thermodenitrificans*

Были выполнены следующие поисковые запросы и выравнивания баз данных:

rBLAST и pBLAST выполняли на внутреннем сервере BLAST, в котором в качестве запрашиваемой последовательности использовали последовательность либо белка, либо гена *G. thermodenitrificans* T12. Эта база данных была в последний раз обновлена в мае 2014 года и, следовательно, не содержит большинства недавно добавленных геномов *Geobacillus*, но нормальную онлайн-BLAST не использовали для предотвращения публикации последовательности T12. Смотри приложение 1 в отношении результатов собственного pBLAST в формате Excel, наиболее релевантные последовательности для этого (последовательности с идентичностью, превышающей 40%, приведены на фиг. 1).

Для включения более свежих данных о последовательностях, последовательность *Geobacillus* MAS1 (наиболее тесно связанную с Cas9 T12) использовали для выполнения PSI-BLAST на веб-сайте NCBI (Johnson et al., 2008 *Nucleic Acids Res.* 36 (выпуск веб-сервера): W5-9). Было выполнено два последовательных цикла PSI-BLAST, в которых для следующего цикла использовали только последовательности, которые соответствовали следующим критериям: минимальная степень покрытия 96% в первом цикле и 97% во втором и третьем циклах, минимальная идентичность 40%, только один штамм на вид.

Последовательности, полученные из PSI-BLAST, а также последовательности с более чем 40% идентичностью с T12 из внутреннего сервера pBLAST, которые не были показаны в PSI-BLAST, выравнивали вместе с хорошо охарактеризованными в настоящее время мезофильными последовательностями а также со всеми идентифицированными в настоящее время термофильными последовательностями, если они были менее релевантными, из которых было создано дерево с использованием метода ближайших соседей (см. фиг. 1). Выравнивание выполняли в MegaB с использованием ClustalW, после чего создавали дерево с использованием метода ближайших соседей, и бутстрап-анализ осуществляли с использованием 1000 репликатов.

Когда BLASTn выполняли с использованием MAS1 *Geobacillus* sp. в качестве последовательности запроса, только Cas9 *Geobacillus* sp. JF8 идентифицировали с 88% идентичностью, что указывает на

очень низкую гомологию на уровне гена. Фиг. 2 представляет собой дерево последовательностей генов Cas9, выровненных по Clustal, с использованием метода ближайших соседей.

Белковые последовательности *G. thermodenitrificans* T12, *A. naeslundii* и *S. pyogenes* были дополнительно проанализированы на гомологию белковых доменов (см. фиг. 3) путем выравнивания их в Clone-Manager с помощью BLOSUM62 с настройками по умолчанию.

Пример 3. Идентификация коровых аминокислотных мотивов, которые существенны для функции CAS9, и мотивов, которые придают термостабильность термофильным нуклеазам Cas9

Процент идентичности описанных выше выровненных белковых последовательностей представлен на фиг. 1. T12-Cas9 относится к типу II-C. Наиболее изученная и недавно кристаллизованная структура системы типа II-C относится к *Actinomyces naeslundii* (Jinek et al., 2014, Science 343: 1247997). Эта белковая последовательность имеет только 20% идентичности с T12-Cas9, но может быть использована для оценки высококонсервативных остатков. В анализы также были включены две хорошо охарактеризованные системы типа II-A (*S. pyogenes* и *S. thermophilus*) (Jinek et al., 2014, Science 343: 1247997; Nishimasu et al., 2014, Cell 156: 935-949). Выравнивание этих четырех белковых последовательностей показано на фиг. 3; на фиг. 4 показана структура белка, определенная для *A. naeslundii* ("Ana-Cas9") (Jinek et al., 2014, Science 343: 1247997). Длина Cas9 из t12 и *Actinomyces naeslundii* очень сходна (*A. naeslundii* 1101 а/к, T12 1082 а/к), и ожидается, что T12 будет иметь сходную структуру белка, но это еще предстоит определить, так как общая идентичность последовательности с cas9-Ana составляет только 20%. Все активные боковые остатки, описанные Jinek et al. (Jinek et al., 2014, Science 343: 1247997) в Cas9 из *A. naeslundii* и *S. pyogenes* можно идентифицировать в T12-Cas9 (см. фиг. 3). РАМ-связывающий домен определен для системы типа II-A *S. pyogenes*, но не для любой системы типа II-C, и поэтому указан только в последовательности *S. pyogenes*. Кроме того, сайт узнавания РАМ сильно варьируется не только между системами CRISPR, но и между видами, содержащими одну и ту же систему. Для получения дополнительной информации о РАМ см. вопрос 4 и будущие планы.

Пример 4. Определение последовательности РАМ Cas9 *G. thermodenitrificans* T12

Установлено, что прокариотические системы CRISPR служат своим хозяевам как адаптивные иммунные системы (Jinek et al., 2012, Science 337: 816-821) и могут быть использованы для быстрой и эффективной геномной инженерии (Mali et al., 2013, Nat Methods 10: 957-963.).

Белки Cas9 функционируют как специфичные для последовательности нуклеазы для систем типа II CRISPR (Makarova et al., 2011, Nat Rev Micro 9: 467-477). Малые молекулы crRNA, которые состоят из "спейсера" (направляющей части), связанного с областью повторов, представляют собой продукты транскрипции и процессинга локусов CRISPR. "Спейсеры" в природе происходят из генома бактериофагов и мобильных генетических элементов, но они также могут быть разработаны для направленности на конкретную нуклеотидную последовательность в процессе геномной инженерии (Bikard et al., 2013, Nucleic Acids Research, 41: 7429-7437). Молекулы crRNA используются Cas9 в качестве направляющих для идентификации их ДНК-мишеней. Спейсерная область идентична области расщепления ДНК-мишени "протоспейсеру" (Brouns et al., 2012, Science 337: 808-809). Для узнавания мишени с помощью Cas9 требуется РАМ (мотив, прилегающий к протоспейсеру), следующий за протоспейсером (Jinek et al., 2012, Science 337: 816-821).

Пример 5. Направленная генерация с помощью рандомизированного РАМ

Два разных спейсера из локусов II CRISPR штамма *G. thermodenitrificans* T12 амплифицировали с помощью ПЦР с использованием геномной ДНК *G. thermodenitrificans* T12 в качестве матрицы. Для амплификации каждого спейсера использовали две пары вырожденных праймеров:

во-первых, использовали пару, вызывающую введение шести случайных нуклеотидов выше фрагмента "протоспейсера", что приводит к получению пула протоспейсеров со случайными последовательностями РАМ;

во-вторых, использовали пару, вызывающую введение шести случайных нуклеотидов ниже фрагмента "протоспейсера", что приводит к получению пула протоспейсеров со случайными последовательностями РАМ;

полученные фрагменты лигировали с вектором pNW33n с получением 4 пулов конструкторов "протоспейсера" со всеми возможными различными 4096 комбинациями РАМ с 6 нуклеотидами в длину. Скомпонованную ДНК использовали для трансформации клеток *G. thermodenitrificans* T12. Клетки высевали на селекционную среду с хлорамфениколом, и объединяли в пул более  $2 \times 10^6$  клеток из каждого пула протоспейсеров. Плазмидную ДНК экстрагировали из пулов, область-мишень амплифицировали с помощью ПЦР и продукты отправляли для глубокого секвенирования. РАМs с наименьшим количеством считываний следует рассматривать как активные, и процесс должен повторяться только с конструктами pNW33n, которые содержат спейсеры с этими РАМs. Сниженная эффективность трансформации *G. thermodenitrificans* T12 должна подтверждать активность РАМs.

Пример 6. Диапазон рабочих температур для Cas9 *Geobacillus thermodenitrificans*

До сих пор не проводили экспериментов на любом имеющемся в открытом доступе белке Cas9 в отношении температурного диапазона его работы. Белки Cas9, используемые в исследовании, все имеют мезофильное происхождение с максимальной температурой роста 45°C для организма-хозяина *Strepto-*

*coccus pyogenes* (ATCC\_700294).

В настоящем документе экспериментально определен диапазон температур, в пределах которых Cas9 из *Geobacillus thermodenitrificans* является стабильным и функциональным, а также температура, при которой его активность оптимальна. Кроме того, также определены функциональные и предпочтительные диапазоны температур для нуклеаз Cas9 из других видов *Geobacillus*.

Температурные диапазоны нуклеаз Cas9 *Geobacillus thermodenitrificans* T12 определены *in vitro* с использованием единой гид-РНК (sgRNA). sgRNA состоит из эндогенной бактериальной crRNA и tracrRNA в едином химерном транскрипте и создается путем использования клонирующего вектора pT7, содержащего остов crRNA:tracrRNA, которому предшествует последовательность-мишень из 20 пар оснований (п.о.) (Jao et al., 2013, PNAS 110: 13904-13909). Последовательность-мишень начинается с мишени в 20 п.о. и заканчивается последовательностью PAM T12 (полученной из примера 4). Желаемую единую направляющую sgRNA создавали путем транскрипции *in vitro* с промотора T7 после линейаризации плазмиды. Стандартный клонирующий вектор pUC19 использовали в качестве мишени в используемых автографах анализа активности *in vitro*. Сайт-мишень (20 нукл.+PAM) в плазмиде конструировали с помощью двух комплементарных олигонуклеотидов, которые отжигали и лигировали в клонирующий вектор sgRNA. Полученную sgRNA инкубировали совместно с нуклеазой Cas9 T12 при температурном диапазоне от 20°C до 100°C. Активность Cas9 в отношении расщепления анализировали электрофорезом в агарозном геле.

Выделенный и охарактеризованный в настоящем документе организм (T12) имеет оптимальную температуру роста 65°C, которая также является оптимальной температурой для его белка Cas9. Диапазон температур, при котором он активен, составляет от 20 до 100°C, в частности от 20 до 80°C.

Оптимальный температурный диапазон Cas9 у видов *Geobacillus* намного выше, чем у белков Cas9, которые были охарактеризованы до настоящего времени. Аналогично, верхняя граница диапазона, в котором он сохраняет активность нуклеазы, намного выше, чем у известных белков Cas9. Более высокая оптимальная температура и функциональный диапазон обеспечивают значительное преимущество в генной инженерии при высоких температурах и, следовательно, в редактировании геномов термофильных организмов, которые применяются в ряде промышленных, сельскохозяйственных и фармацевтических процессов, проводимых при повышенных температурах.

#### ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Выделенный белок или полипептидный фрагмент Cas, имеющий аминокислотную последовательность SEQ ID NO: 1 или последовательность, идентичную ей, по меньшей мере, на 90%, где белок Cas способен расщеплять полинуклеотид, включающий последовательность нуклеиновой кислоты-мишени, при температуре от 50 до 100°C при связывании по меньшей мере с одной молекулой РНК, которая распознает последовательность-мишень.

2. Белок или полипептидный фрагмент Cas по п.1, где аминокислотная последовательность имеет по меньшей мере 95% идентичность последовательности SEQ ID NO: 1.

3. Белок или полипептидный фрагмент Cas по п.1 или 2, где белок или фрагмент Cas способен расщеплять нуклеиновую кислоту при температуре от 50 до 75°C, предпочтительно выше 60°C, более предпочтительно от 60 до 80°C, еще более предпочтительно от 60 до 65°C.

4. Белок или полипептидный фрагмент Cas по любому из пп.1-3, где расщепление нуклеиновой кислоты представляет собой расщепление ДНК.

5. Белок или полипептидный фрагмент Cas по любому из предшествующих пунктов, где белок Cas получают из бактерии, архея или вируса.

6. Белок или полипептидный фрагмент Cas по любому из предшествующих пунктов, где белок Cas получают из *Geobacillus* sp., предпочтительно из *Geobacillus thermodenitrificans*.

7. Рибонуклеопротеидный комплекс, включающий белок Cas по любому из предшествующих пунктов и включающий по меньшей мере одну молекулу направляющей РНК, которая узнает последовательность в полинуклеотиде-мишени, и где рибонуклеопротеидный комплекс способен расщеплять полинуклеотид, включающий последовательность нуклеиновой кислоты-мишени, при температуре от 50 и 100°C.

8. Рибонуклеопротеидный комплекс по п.7, в котором молекула направляющей РНК включает crRNA и, необязательно, tracrRNA.

9. Рибонуклеопротеидный комплекс по п.7 или 8, где длина по меньшей мере одной молекулы РНК находится в диапазоне 35-135 нуклеотидных остатков.

10. Рибонуклеопротеидный комплекс по п.7 или 8, где последовательность-мишень имеет длину 31 или 32 нуклеотидных остатка.

11. Белок или полипептид Cas по любому из пп.1-6 или рибонуклеопротеидный комплекс по любому из пп.7-10, где белок или полипептид предоставлены как часть белкового комплекса, включающего по меньшей мере один дополнительный функциональный или нефункциональный белок, где белок или полипептид Cas и/или по меньшей мере один дополнительный белок дополнительно включает по меньшей мере одну функциональную часть,

(а) где по меньшей мере одна функциональная часть слита или соединена с N-концом и/или С-концом белка, полипептида или рибонуклеопротеидного комплекса Cas, предпочтительно с N-концом, и/или

(b) где по меньшей мере одна функциональная часть представляет собой белок.

12. Белок или полипептид Cas по п.11, где белок выбирают 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.

13. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по п.12, где природная активность нуклеазы Cas9 инактивирована и белок Cas связан по меньшей мере с одной функциональной частью.

14. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по п.12 или 13, где по меньшей мере одна функциональная часть является нуклеазным доменом, предпочтительно доменом нуклеазы FokI.

15. Белок или полипептид, или рибонуклеопротеидный комплекс Cas по любому из пп.12-14, где по меньшей мере одна функциональная часть представляет собой маркерный белок.

16. Способ модификации нуклеиновой кислоты-мишени, включающий контактирование нуклеиновой кислоты:

а) с рибонуклеопротеидным комплексом по любому из пп.7-10; или

б) с белком или белковым комплексом по любому из пп.11-15 и по меньшей мере одной молекулой направляющей РНК по любому из пп.5-10.

17. Способ модификации нуклеиновой кислоты-мишени по п.16, где по меньшей мере одна функциональная часть представляет собой маркерный белок или репортерный белок, и маркерный белок или репортерный белок связан с нуклеиновой кислотой-мишенью.

18. Способ модификации нуклеиновой кислоты-мишени по п.17, где маркер представляет собой флуоресцентный белок, например зеленый флуоресцентный белок (GFP).

19. Способ модификации нуклеиновой кислоты-мишени по любому из пп.16 или 18, где нуклеиновая кислота-мишень представляет собой ДНК, предпочтительно дцДНК.

20. Способ модификации нуклеиновой кислоты-мишени по любому из пп.16 или 18, где нуклеиновая кислота-мишень представляет собой РНК.

21. Способ модификации нуклеиновой кислоты-мишени по п.19, где нуклеиновая кислота представляет собой дцДНК, по меньшей мере одна функциональная часть представляет собой нуклеазу или геликазу-нуклеазу и модификация представляет собой одноцепочечный или двухцепочечный разрыв в желаемом локусе.

22. Способ модификации экспрессии гена в клетке, включающий модификацию последовательности нуклеиновой кислоты-мишени, как заявлено в способе по любому из пп.16-19, где нуклеиновая кислота представляет собой дцДНК и функциональная часть выбрана из фермента, модифицирующего ДНК (например, метилазы или ацетилазы), активатора транскрипции или репрессора транскрипции.

23. Способ модификации экспрессии гена в клетке, включающий модификацию последовательности нуклеиновой кислоты-мишени в способе по п.20, где нуклеиновая кислота представляет собой мРНК и функциональная часть представляет собой рибонуклеазу.

24. Способ модификации экспрессии гена в клетке по п.22, где рибонуклеазу выбирают из эндонуклеазы, 3'-экзонуклеазы или 5'-экзонуклеазы.

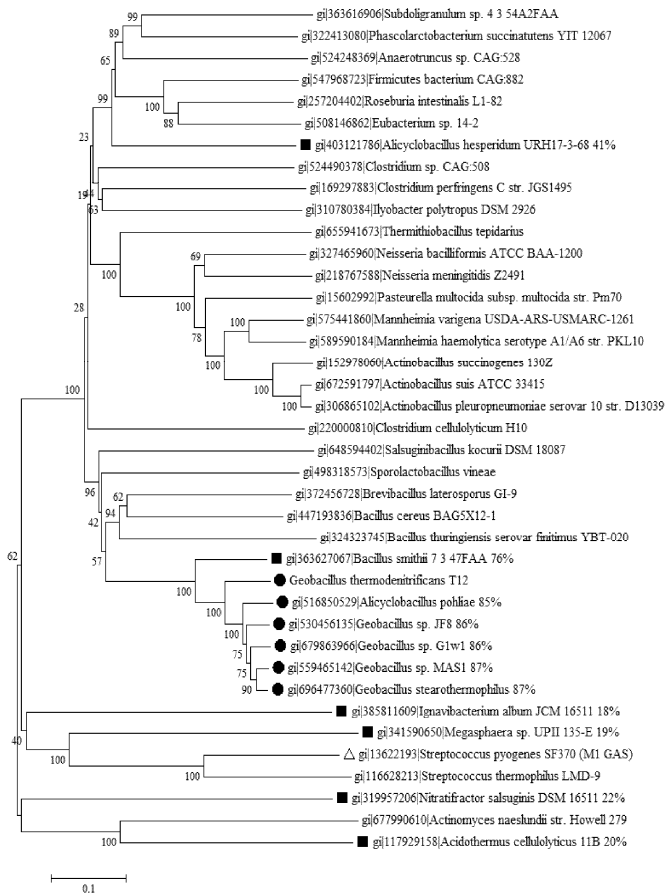
25. Способ модификации нуклеиновой кислоты-мишени по любому из пп.16-22, где способ осуществляют при температуре от 50 до 100°C.

26. Способ модификации нуклеиновой кислоты-мишени по п.25, где способ осуществляют при температуре, равной или выше 60°C, предпочтительно от 60 до 80°C, более предпочтительно от 60 до 65°C.

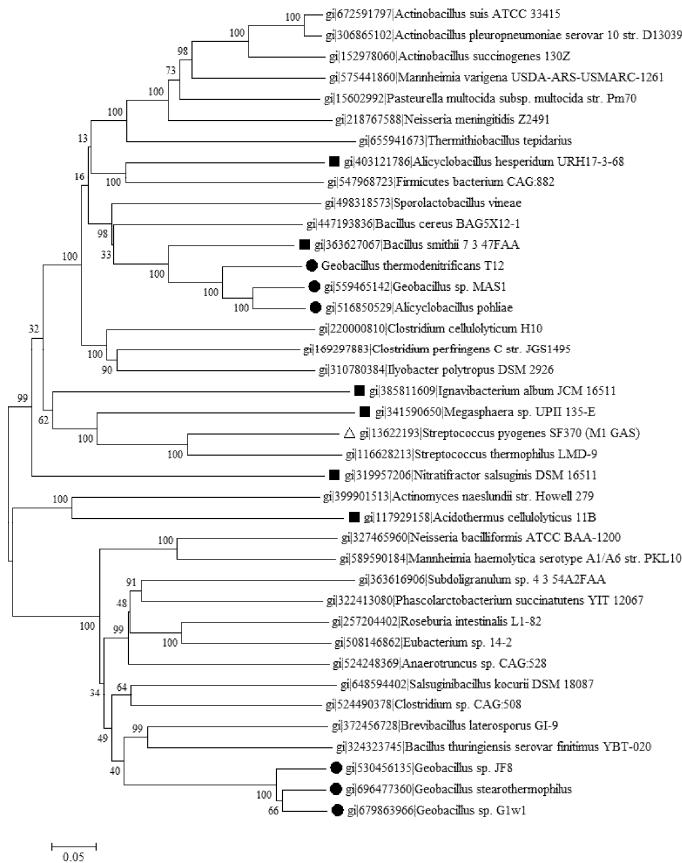
27. Способ по любому из пп.16-26, в котором клетка представляет собой прокариотическую клетку.

28. Способ по любому из пп.16-27, в котором клетка представляет собой эукариотическую клетку.

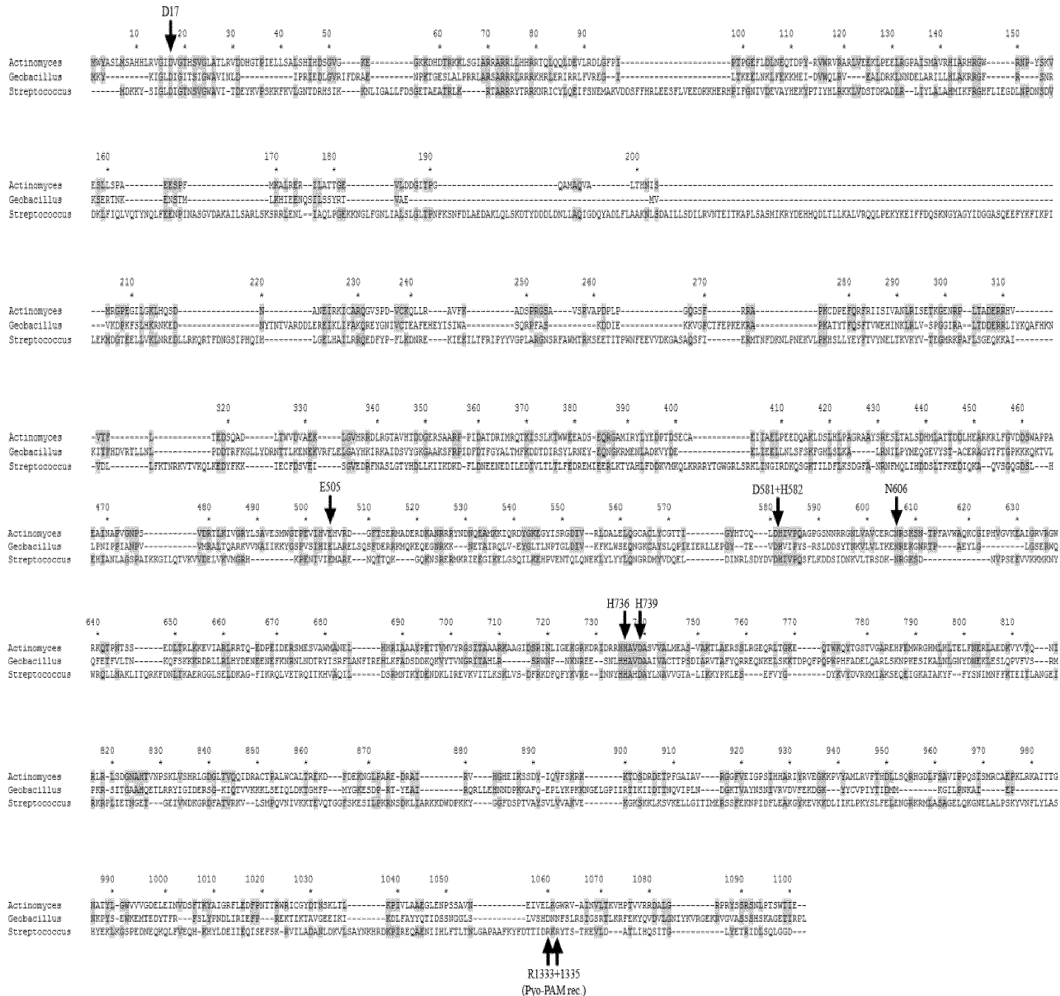
29. Клетка-хозяин, трансформированная экспрессионным вектором, содержащим последовательность нуклеиновой кислоты, кодирующую белок Cas по любому из пп.1-6.



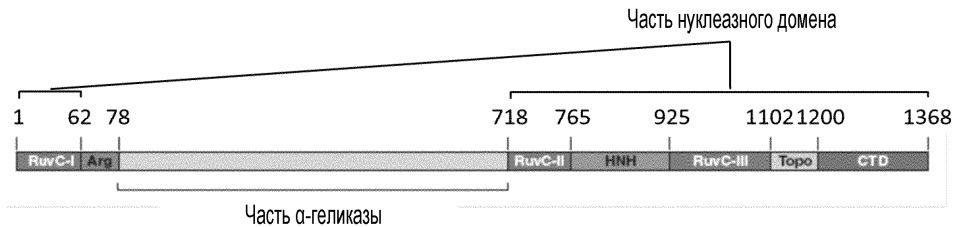
Фиг. 1



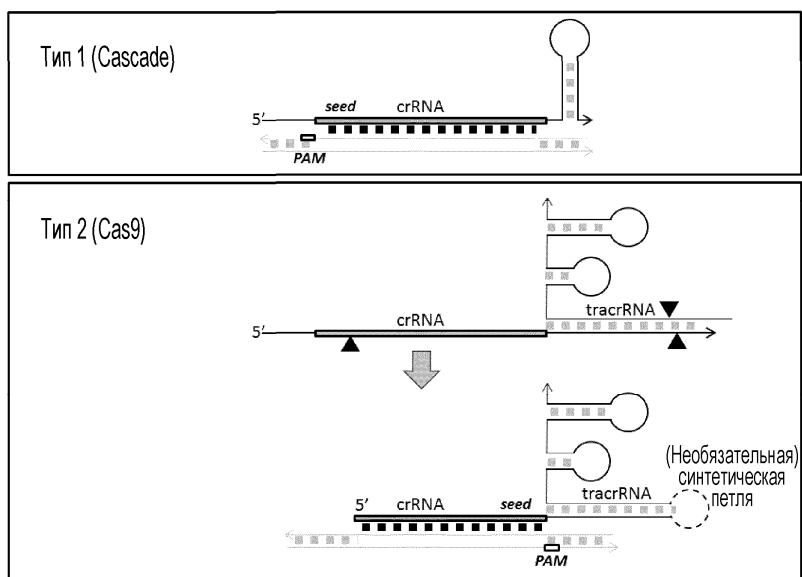
Фиг. 2



Фиг. 3



Фиг. 4



Фиг. 5

