

(19)



Евразийское
патентное
ведомство

(11) 042506

(13) B1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ

(45) Дата публикации и выдачи патента
2023.02.21

(21) Номер заявки
201792663

(22) Дата подачи заявки
2016.05.27

(51) Int. Cl. C12Q 1/02 (2006.01)
C12Q 1/68 (2006.01)
C12N 1/21 (2006.01)
C12N 15/63 (2006.01)

(54) СПОСОБЫ СКРИНИНГА БАКТЕРИЙ, АРХЕЙ, ВОДОРОСЛЕЙ И ДРОЖЖЕЙ С
ИСПОЛЬЗОВАНИЕМ НУКЛЕИНОВЫХ КИСЛОТ CRISPR

(31) 62/168,355; 62/296,853

(32) 2015.05.29; 2016.02.18

(33) US

(43) 2018.04.30

(86) PCT/US2016/034812

(87) WO 2016/196361 2016.12.08

(71)(73) Заявитель и патентовладелец:
НОРТ КАРОЛИНА СТЕЙТ
ЮНИВЕРСИТИ (US)

(72) Изобретатель:
Баррангу Родольф, Селль Курт М.
(US)

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

(56) GOMAA AHMED A. et al.: "Programmable removal of bacterial strains by use of genome-targeting CRISPR-Cas systems", mBio, 2014, vol. 5, issue 1: e00928-13, DOI: 10.1128/mBio.00928-13, p. 1-9, especially, abstract, p. 1-7
WO-A1-2014204727
EP-A1-2860267

(57) Изобретение относится к нуклеиновым кислотам CRISPR для скрининга в отношении жизненно важных и нежизненно важных генов и инородных геномных островов в бактериях, археях, водорослях и/или дрожжах для уничтожения бактерий, архей, водорослей и/или дрожжей, для идентификации фенотипа гена или генов и/или для скрининга в отношении уменьшенного размера генома и/или делеции гена в бактериях, археях, водорослях и/или дрожжах.

042506 B1

042506 B1

042506

B1

Притязания на приоритет

Настоящая заявка притязает на приоритет, согласно 35 U.S.C. 119 (e), предварительной заявки на патент США № 62/168355, поданной 29 мая 2015 г., и предварительной заявки на патент США № 62/296,853, поданной 18 февраля 2016 г., содержание которых включено в настоящее описание полностью в виде ссылки.

Область техники, к которой относится изобретение

Настоящее изобретение относится к нуклеиновым кислотам CRISPR для скрининга в отношении жизненно важных и нежизненно важных генов и инородных геномных островов в бактериях, археях, водорослях и/или дрожжах для уничтожения бактерий, архей, водорослей и/или дрожжей, для идентификации фенотипа гена или генов и/или для скрининга в отношении уменьшенного размера генома и/или делеции гена в бактериях, археях, водорослях и/или дрожжах.

Уровень техники

Короткие палиндромные повторы, регулярно расположенные группами (CRISPR), в комбинации с ассоциированными последовательностями (cas), составляют CRISPR-Cas систему, которая обеспечивает многим бактериям адаптивный иммунитет. CRISPR-опосредованная иммунизация осуществляется путем поглощения ДНК из инвазивных генетических элементов, таких как плазмиды и фаги, в виде новых "спейсеров".

CRISPR-Cas системы состоят из массива коротких ДНК повторов, разделенных гипервариабельными последовательностями, фланкированных cas генами, которые обеспечивают адаптивный иммунитет к инвазивным генетическим элементам, таким как фаги и плазмиды, путем нацеливания на специфические последовательности и интерференции (Barrangou et al., 2007, Science, 315:1709-1712; Brouns et al., 2008, Science, 321:960-4; Horvath and Barrangou, 2010, Science, 327:167-70; Marraffini and Sontheimer, 2008, Science, 322:1843-1845; Bhaya et al., 2011, Annu. Rev. Genet., 45:273-297; Terns and Terns, 2011, Curr. Opin. Microbiol., 14:321-327; Westra et al., 2012, Annu. Rev. Genet., 46:311-339; Barrangou R., 2013, RNA, 4:267-278). Обычно инвазивные ДНК последовательности приобретаются в виде новых "спейсеров" (Barrangou et al., 2007, Science, 315:1709-1712), где каждая последовательность спарена с повтором CRISPR и вставлена в виде новой единицы повтор-спейсер в локус CRISPR. Впоследствии массив повтор-спейсер транскрибируется в виде длинной pre-CRISPR РНК (pre-crPHK) (Brouns et al., 2008, Science, 321:960-4), которая процессируется с образованием коротких интерферирующих CRISPR РНК (crPHK), которые управляют распознаванием специфических последовательностей. По существу crPHK направляют нуклеазы на комплементарные мишени для опосредованного эндонуклеазами Cas расщепления нуклеиновой кислоты со специфической последовательностью (Garneau et al., 2010, Nature, 468:67-71; Haurwitz et al., 2010, Science, 329:1355-1358; Sapranaukas et al., 2011, Nucleic Acid Res., 39:9275-9282; Jinek et al., 2012, Science, 337:816-821; Gasiunas et al., 2012, Proc. Natl. Acad. Sci., 109:E2579-E2586; Magadan et al., 2012, PLoS One, 7:e40913; Karvelis et al., 2013, RNA Biol., 10:841-851). Эти широко распространенные системы встречаются почти у половины бактерий (~46%) и у большинства археев (~90%).

В общих чертах, имеется два класса (Makarova et al., Nat. Rev. Microbiol., 13:722-736 (2015)) CRISPR-Cas систем, которые охватывают пять больших типов и 16 разных подтипов, исходя из содержания cas генов, строения cas оперона, последовательностей белков Cas и этапов процесса (Makarova et al., Biol. Direct., 6:38 (2011); Makarova and Koonin, Methods Mol. Biol., 1311:47-75 (2015); Barrangou, R., Genome Biology, 16:247 (2015)). В типах I и III специализированные эндонуклеазы Cas осуществляют процессинг pre-crPHK, которые впоследствии образуют крупный комплекс мульти-Cas белков, способный распознавать и расщеплять нуклеиновые кислоты, комплементарные crPHK. Системы типа I являются наиболее часто встречающимися и широко распространенными системами, которые нацеливаются на ДНК Cascade-управляемым и РАМ-зависимым способом, разрушая целевые нуклеиновые кислоты при помощи сигнатурного белка Cas3. CRISPR-Cas система типа II имеет другой процесс. В этом случае процессинг pre-CRNA происходит по механизму, при котором транс-активирующая crPHK (tracrPHK) гибридизуется с областями повтора crPHK. Гибридные crPHK-tracrPHK расщепляются RNase III, и после второго события, при котором происходит удаление 5'-конца каждого спейсера, продуцируются зрелые crPHK, которые остаются связанными как с tracrPHK, так и Cas9. Затем зрелый комплекс располагается на dsДНК последовательности-мишени ('протоспейсерная' последовательность), которая является комплементарной спейсерной последовательности в этом комплексе, и разрезает обе нити. Для узнавания и расщепления этим комплексом в системе типа II требуется наличие не только комплементарной последовательности между спейсерной последовательностью комплекса crPHK-tracrPHK и 'протоспейсерной' последовательностью-мишенью, но также требуется последовательность прилегающего к протоспейсеру мотива (РАМ), расположенная на 3'-конце протоспейсерной последовательности.

Сущность изобретения

В одном из аспектов изобретения предоставляется способ скрининга популяции бактериальных клеток в отношении жизненно важных генов, не жизненно важных генов и/или инородных геномных островов, включающий

введение в указанную популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-

спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных клеток указанной популяции, таким образом создавая популяцию трансформированных бактериальных клеток;

определение наличия или отсутствия делеции в популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, причем наличие делеции в популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток означает, что область-мишень содержится в нежизненно важном гене и/или инородном геномном острове, а отсутствие делеции в популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток означает, что область-мишень содержится в жизненно важном гене.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

Во втором аспекте изобретения предоставляется способ скрининга популяции бактериальных, архейных, водорослевых или дрожжевых клеток в отношении жизненно важных генов, нежизненно важных генов и/или инородных геномных островов, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодируемую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, таким образом создавая популяцию трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток; и

определение наличия или отсутствия делеции в популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, причем наличие делеции в популяции бактериальных, архейных, водорослевых или дрожжевых клеток означает, что область-мишень содержится в нежизненно важном гене и/или инородном геномном острове, а отсутствие делеции в популяции бактериальных, архейных, водорослевых или дрожжевых клеток означает, что область-мишень содержится в жизненно важном гене.

В третьем аспекте изобретения предоставляется способ уничтожения одной или более бактериальных клеток в популяции бактериальных клеток, включающий введение в указанную популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (crPHK, crДНК), содержащий (в 5' к 3' направлению) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных клеток указанной популяции, таким образом уничтожая в популяции бактериальных клеток одну или более бактериальных клеток, которые содержат область-мишень. Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В четвертом аспекте изобретения предоставляется способ уничтожения одной или более клеток в популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодируемую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида и/или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, таким образом уничтожая одну или более клеток, которые содержат область-мишень в своем геноме, в популяции бактериальных, архейных, водорослевых или дрожжевых клеток.

В пятом аспекте изобретения предоставляется способ идентификации фенотипа, связанного с бактериальным геном, включающий

введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты,

содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной по меньшей мере одной последовательности повтор-спейсер и последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных клеток указанной популяции, причем эта область-мишень содержит по меньшей мере часть открытой рамки считывания, кодирующей полипептид, или функциональную нуклеиновую кислоту, таким образом уничтожая клетки, содержащие эту область-мишень, и создавая популяцию трансформированных бактериальных клеток, не содержащих область-мишень; и

анализ фенотипа популяции.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В шестом аспекте изобретения предоставляется способ идентификации фенотипа, связанного с бактериальным, ахейным, водорослевым или дрожжевым геном, включающий

введение в популяцию бактериальных, ахейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодируемую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) бактериальных, ахейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, таким образом уничтожая бактериальные, ахейные, водорослевые или дрожжевые клетки, содержащие область-мишень, и создавая популяцию трансформированных бактериальных, ахейных, водорослевых или дрожжевых клеток, не содержащих область-мишень; и

анализ фенотипа популяции трансформированных бактериальных, ахейных, водорослевых или дрожжевых клеток, и/или

(i) выращивание отдельных бактериальных, ахейных, водорослевых или дрожжевых колоний из популяции трансформированных бактериальных, ахейных, водорослевых или дрожжевых клеток, и

(ii) анализ фенотипа отдельных колоний.

В седьмом аспекте изобретения предоставляется способ отбора одной или более бактериальных клеток, имеющих уменьшенный размер генома, из популяции бактериальных клеток, включающий введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более клеток бактерий в указанной популяции, причем клетки, содержащие область-мишень, уничтожаются, таким образом осуществляя отбор одной или более бактериальных клеток, не содержащих эту область-мишень и имеющих уменьшенный размер генома, из популяции бактериальных клеток. Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В восьмом аспекте изобретения предоставляется способ отбора одной или более бактериальных клеток, имеющих уменьшенный размер генома, из популяции бактериальных клеток, включающий введение в популяцию бактериальных клеток

(a) (i) одной или более конструкций гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательными нуклеотидами, присутствующим в геноме указанных бактериальных клеток, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных клеток бактерий, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, причем эта область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере

300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, и клетки, содержащие область-мишень уничтожаются, таким образом осуществляя отбор одной или более бактериальных клеток, не содержащих область-мишень и имеющих уменьшенный размер генома, из популяции трансгенных бактериальных клеток.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В девятом аспекте изобретения предоставляется способ отбора одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих уменьшенный размер генома, из популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включающий введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодируемую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной по меньшей мере одной последовательности и по меньшей мере одной последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида, причем клетки, содержащие эту область-мишень уничтожаются, таким образом осуществляя отбор одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень и имеющих уменьшенный размер генома, из популяции бактериальных, архейных, водорослевых или дрожжевых клеток.

В десятом аспекте изобретения предоставляется способ отбора одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих уменьшенный размер генома, из популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) (i) одной или более конструкций гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательным нуклеотидам, присутствующим в геноме указанных бактериальных, архейных, водорослевых или дрожжевых клеток, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных бактериальных, архейных, водорослевых или дрожжевых клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) (i) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодируемую CRISPR (tracr) нуклеиновую кислоту,

(ii) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной по меньшей мере одной последовательности повтор-спейсер и по меньшей мере одной последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) одной или более бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(iii) Cas9 полипептида и/или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, причем эта область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, и клетки, содержащие область-мишень уничтожаются, таким образом осуществляя отбор одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень и имеющих уменьшенный размер генома, из популяции трансгенных бактериальных, архейных, водорослевых или дрожжевых клеток.

В одиннадцатом аспекте изобретения предоставляется способ идентификации в популяции бактерий по меньшей мере одного изолята, имеющего делецию в своем геноме, включающий

введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток указанной популяции, причем клетки, содержащие область-мишень уничтожаются, таким образом создавая популяцию трансформированных

бактериальных клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных колоний из популяции трансформированных бактериальных клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции трансформированных бактерий имеющих делецию в своем геноме.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В двенадцатом аспекте изобретения предоставляется способ идентификации в популяции бактерий по меньшей мере одного изолята, имеющего делецию в своем геноме, включающий

введение в популяцию бактериальных клеток

(a) (i) одной или более конструкции гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательным нуклеотидам, присутствующим в геноме указанных бактериальных клеток, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных бактериальных клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, причем эта область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, и клетки, содержащие область-мишень уничтожаются, таким образом создавая популяцию трансформированных бактериальных клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных колоний из популяции трансформированных бактериальных клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции бактерий, имеющих делецию в своем геноме.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В тринадцатом аспекте изобретения предоставляется способ идентификации в популяции бактериальных, архейных, водорослевых или дрожжевых клеток по меньшей мере одного изолята, имеющего делецию в своем геноме, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодируемую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, причем клетки, содержащие область-мишень, уничтожаются, таким образом создавая популяцию трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных, архейных, водорослевых или дрожжевых колоний из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих делецию в своем геноме.

В четырнадцатом аспекте изобретения предоставляется способ определения в популяции бактериальных, архейных, водорослевых или дрожжевых клеток по меньшей мере одного изолята, имеющего делецию в своем геноме, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) (i) одной или более конструкции гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательным нуклеотидам, присутствующим в геноме указанных бактериальных, архейных, водорослевых или дрожжевых клеток, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей

мере один транспозон, таким образом создавая популяцию трансгенных бактериальных, архейных, водорослевых или дрожжевых клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) (i) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-кодированную CRISPR (tracr) нуклеиновую кислоту,

(ii) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в направлении от 5'-к 3'-конца) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (хромосомном и/или плазмидном) одной или более бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(iii) Cas9 полипептида и/или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, причем эта область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, и клетки, содержащие область-мишень уничтожаются, таким образом создавая популяцию трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных, архейных, водорослевых или дрожжевых колоний из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции, имеющей делецию в своем геноме.

Настоящее изобретение также относится к экспрессионным кассетам, клеткам и наборам, содержащим конструкции нуклеиновой кислоты, массивы нуклеиновых кислот, молекулы нуклеиновых кислот и/или нуклеотидные последовательности по изобретению.

Эти и другие аспекты изобретения изложены ниже в описании изобретения более подробно.

Краткое описание фигур

На фиг. 1 показана последовательность массивов SthCRISPR1 и SthCRISPR3 для таргетирования (нацеливания) каждого композитного транспозона.

На фиг. 2 показана схематическая диаграмма способа сплайсинга с перекрывающимся удлинением (SOE) для конструкции таргетирования плазмид.

На фиг. 3 показана карта жизненно важных генов, инсерционных последовательностей и геномных островов. (A) Местоположение и распределение предполагаемых OPC (ORF) жизненно важных генов [жизненно важных OPC] (верхний ряд), инсерционных последовательностей (2-й ряд) и предполагаемых геномных островов (3-й ряд). Потенциальные мишени для опосредованных CRISPR-Cas (4-й ряд) делеций определяли путем картирования мобильных элементов различных семейств в геноме *Streptococcus thermophilus* LMD-9. Генетическая структура предполагаемых геномных островов и соответствующих им комбинаций протоспейсер/PAM. (B) Геномный остров 1, кодирующий систему для транспорта олигопептидов. (C) Геномный остров 2, содержащий протеиназу PrtS клеточной оболочки. (D) Геномный остров 3, кодирующий белок АТФазной помпы ионов меди. (E) Геномный остров, кодирующий выбранные OPC, включающие Lac оперон.

На фиг. 4 представлена дендрограмма кодирующих последовательностей транспозонов, распределенных в геноме *S. thermophilus* LMD-9. Выравнивание производилось при помощи компьютерной программы Geneious®. Окончательные обозначения присваивали, используя www-is.biotoul.fr. Буквы соответствуют выравниванию последовательностей, приведенному на фиг. 5 для каждого семейства.

На фиг. 5A-5D показано выравнивание последовательностей транспозов (при помощи компьютерной программы Geneious®) для каждого крупного семейства IS, найденного в геноме *S. thermophilus* LMD-9. Разные семейства имеют разную степень сохранения длины и идентичности нуклеотида. (A) Транспозоны Sth6 были в высокой степени полиморфными и очевидно вырожденными вследствие внутренних делеций в некоторых копиях. IS1167 (B) и IS1191 (C), напротив, имели несколько копий, но сохраняли высокую степень соответствия по длине и идентичности. S1193 (D) имели высокоточные копии, но при этом демонстрировали самое высокое различие внутри семейства по длине.

На фиг. 6A-6C показана структурная основа нацеливания системы на ДНК при помощи CRISPR-Cas9 и явная цитотоксичность этого процесса. Спейсерные последовательности у SthCRISPR1 (A) и SthCRISPR3 (B) для таргетирования *lacZ*. Cas9 опознает ДНК и обратимо связывается с последовательностями PAM, при этом происходит стабилизация Cas9 на мишени за счет образования дуплекса tracrPHK::crPHK. Активация Cas9 вызывает одновременное расщепления каждой нити по доменам RuvC и HNH, указанные черными стрелками. Трансформанты, выделенные после электропорации контрольного

ных и самонацеленных плазмид (С). Средние размеры клонов $\pm$ SD, скринированных в независимых экспериментах по трансформации ($n=4$), для каждой из тестовых плазмид.

На фиг. 7 показаны секвенирование генома и фенотипический анализ Lac-клонов. Данные, полученные для последовательностей, выявили отсутствие хромосомного сегмента, кодирующего *lacZ*, у двух мутантов, полученных независимо путем нацеливания последовательностей *lacZ*, кодирующих (А) 5'-конец и (В) катион-связывающий фрагмент, при помощи CRISPR3 системы. Размер делеций находился в пределах от 101865 до 102146 пар оснований, что составляет примерно 5,5% генома *S. thermophilus*. (С) Рост штаммов с крупными делециями по сравнению с диким типом в полусинтетической среде Элликера, представлен в виде среднего значения $\pm$ SD OD 600 нм для трех независимых биологических повторов. (D) Способность штаммов *S. thermophilus* к сквашиванию обезжиренного молока.

На фиг. 8 показаны события рекомбинации между инсерционными последовательностями (IS). (А) Изображение геля с результатами электрофореза ампликонов с крупными делециями, полученных ПЦР анализом гДНК, выделенной из трансформантов. Скрининг проводили, используя праймеры, фланкирующие IS1193 элементы, расположенные выше и ниже предполагаемого сайта делеций. Линии, отмеченные знаком Δ , амплифицировали из гДНК Lac-клонов, полученных после таргетирования *lacZ*, опосредованного CRISPR-Cas системой, при этом WT происходит от дикого типа. (В) Последовательности предсказанных сайтов рекомбинации определяли путем картирования событий однонуклеотидного полиморфизма, соответствующих либо вышерасположенным (серые), либо нижерасположенным (черные) IS элементам. Три сайта были предсказаны на основе последовательностей, сохранившихся в обоих IS элементах (светло-серые). Изображенные сайты представляют собой генотипы из независимых клонов и являются типичными для феномена Лакса, наблюдаемого в девяти различных сайтах рекомбинации. Следы химерных IS-элементов были обнаружены аналогичным образом в каждом локусе геномного острова на границе делеции. (С) Схематическое изображение IS-элементов, которые согласно прогнозу будут подвергнуты рекомбинации во время хромосомной делеции острова, кодирующего *lacZ*. (D) Ампликоны, полученные из праймеров, фланкирующих геномные острова 1, 2 и 3, для подтверждения делеции. (Е) Ампликоны, полученные из внутренних праймеров, для подтверждения отсутствия последовательностей дикого типа в каждой культуре с делециями, индуцированными CRISPR. Линии, отмеченные знаком Δ , амплифицировали из гДНК клонов, полученных после опосредованного CRISPR-Cas системой нацеливания, при этом WT происходит от дикого типа.

На фиг. 9 показаны мишени летальности, а также использование определенных генетических локусов для оценки летальности, опосредованной CRISPR-Cas системой типа II, путем таргетирования генома *Streptococcus thermophilus* LMD-9. Были протестированы обе ортогональные системы типа II (CRISPR1 и CRISPR3); мишени CRISPR1 показаны темно-серым цветом, а мишени CRISPR3 - светло-серым цветом. Специфические генетические признаки выбирали для тестирования (i) межгенных областей (INT), (ii) мобильных генетических элементов (ISSth7, *oppC*-GEI1, *prtS*-GEI2, *copA*-GEI3, *cI*, *lacZ*-GEI4, *epsU*), (iii) жизненно важных генов (*dltA*, *ltaS*), (iv) полюсов репликаторы (*OriC*, *xerS*) и прямых или обратных нитей ДНК (внешних мишеней по отношению к внутренним мишеням).

На фиг. 10 показана летальность, опосредованная CRISPR, путем нацеливания на области, показанные на фиг. 9. Логарифмическое уменьшение КОЕ (колониеобразующих единиц) рассчитывали в отношении трансформации контроля несамонацеленной плазмиды; *pORI28*. Летальность находилась в пределах уменьшения 2-3 log для всех тестируемых мишеней независимо от местоположения в хромосоме кодирующей последовательности или важности. ISSth7 - элемент инсерционной последовательности, *ltaS* - синтаза липотейхоевых кислот; *prtS* - геномный остров 2; INT - межгенная область; *dltA* - D-аланин лигаза; *gheB* - locus с отсутствующим сайтом *chi*; *oppC* - геномный остров 1; *comS* - плотный locus сайта *chi*; *xerS* - конец репликации; *copA* - геномный остров 3; *cI* - профаговый остаток; *OriC* - начало репликации; *Cas9* - *Cas9* кодирующая последовательность системы CRISPR3; *epsU* - кассета экзополисахаридов.

На фиг. 11 показаны профили транскрипции штаммов с опосредованными CRISPR делециями геномных островов.

На фиг. 12 показано $\log_2$ покрытие прочтениями при секвенировании трансформированной РНК штаммов с делециями геномных островов, GEI1, GEI2, GEI3 и GEI4.

На фиг. 13 показаны XY графики значений экспрессии штаммов с делециями геномных островов (оси X) в сравнении со значениями экспрессии дикого типа (оси Y). Для каждого из штаммов с делециями геномных островов (GEI1-GEI4) экспрессия генов, кодируемых на каждом из островов-мишеней (черные), была минимальной. Гены, кодируемые в GEI1, показаны на верхней левой панели, гены, кодируемые в GEI2, показаны на верхней правой панели, гены, кодируемые в GEI3, показаны на нижней левой панели, а гены, кодируемые в GEI4, показаны на нижней правой панели.

На фиг. 14А, 14В показано введение экзогенного фага, плазмиды или фегемида, кодирующего массив CRISPR (система типа II) для со-орт эндогенных систем для запрограммированной гибели клеток в *Streptococcus thermophilus*.

Фиг. 15. Гиды типа II *Lactobacillus casei*. Верхняя структура - предсказанный гид. Левая нижняя фигура представляет собой правильный двойной гид *crPHK:tracrPHK*, что подтверждается секвенированием

РНК. Правая нижняя фигура является примером предсказанного искусственного одонитевого гида.

На фиг. 16 представлены приведенные в качестве примера гиды типа II *Lactobacillus casei*. Верхняя структура - предсказанный гид. Левая нижняя фигура представляет собой правильный двойной гид $\text{crPНК}:\text{tracrPНК}$, что подтверждается секвенированием РНК. Правая нижняя фигура является примером предсказанного искусственного одонитевого гида.

На фиг. 17 представлены приведенные в качестве примера гиды типа II *Lactobacillus pentosus*. Верхняя структура предсказанный гид. Левая нижняя фигура представляет собой правильный двойной гид $\text{crPНК}:\text{tracrPНК}$, что подтверждается секвенированием РНК. Правая нижняя фигура является примером предсказанного искусственного одонитевого гида.

На фиг. 18 представлены приведенные в качестве примера гиды типа II *Lactobacillus jensenii*. Левая панель представляет собой правильный двойной гид $\text{crPНК}:\text{tracrPНК}$, что подтверждается секвенированием РНК. Правая панель представляет собой пример предсказанного искусственного одонитевого гида.

На фиг. 19 показаны результаты трансформации плазмид, содержащих протоспейсер, который соответствует crPНК с самой высокой степенью транскрипции в нативном массиве CRISPSR типа II *L. gasserii*. В *L. gasserii* были трансформированы четыре разные плазмиды (слева направо): пустой вектор pTRK563, конструкция с правильным протоспейсером, но неверным PAM, с правильным PAM, но с протоспейсером, который не находится в массиве, и с правильными протоспейсером и PAM, которая продемонстрировала самую высокую степень направленности интерференции и гибели клеток. Приведенные значения представляют собой среднее значение $\pm\text{SEM}$ для трех независимых реплик.

На фиг. 20 показаны результаты трансформации плазмид, содержащих протоспейсер, который соответствует crPНК с самой высокой степенью транскрипции в нативном массиве CRISPSR типа II *L. pentosus*. В *L. pentosus* были трансформированы четыре разные плазмиды (слева направо): конструкция с правильным протоспейсером, но неверным PAM (Lpe4 ctGttt), с правильным PAM, но с протоспейсером, который не находится в массиве (Lpe8 noSPCR), пустой вектор pTRK563 (pTRK563) и плазида с правильным протоспейсером и правильным PAM (Lpe1 gtaaat). Приведенные значения представляют собой среднее значение $\pm\text{SEM}$ для трех независимых реплик.

На фиг. 21 представлен приведенный в качестве примера гид CRISPR-Cas типа I *Lactobacillus casei*. Приведенная последовательность представляет собой нативную лидерную последовательность типа I и повтор, который находится в NCK 125 *Lactobacillus casei*. Этот искусственный массив содержит спейсер, который таргетирует ген 16s рДНК в геноме хозяина.

На фиг. 22 показана трансформация плазмид, содержащих протоспейсер, который соответствует crPНК с самой высокой степенью транскрипции в нативном массиве CRISPSR типа II *L. jensenii*. В *L. jensenii* были трансформированы четыре разные плазмиды (слева направо): пустой вектор pTRK563, конструкция с правильным протоспейсером, но неверным PAM, с правильным PAM, но с протоспейсером, который не находится в массиве, и с правильными протоспейсером и PAM, которая продемонстрировала самую высокую степень направленности интерференции и гибели клеток.

На фиг. 23 показано направленное самоуничтожение использованием нативной системы типа I в *Lactobacillus casei* NCK 125. Были сконструированы две мишени в гене 16s рДНК. PAM 5'-YAA-3' была предсказана при помощи нативных спейсерных последовательностей в организме. Искусственный массив, содержащий нативную лидерную последовательность типа I, повторы и выбранные спейсеры, клонировали в pTRK870. Введенные конструкции включали пустой вектор (pTRK563) и два разных искусственных массива: один из которых содержал один спейсер, таргетирующий +цепь в гене 16s (1-2 alt), а другой массив содержал исходный спейсер, таргетирующий +цепь, а также дополнительный спейсер, таргетирующий -цепь в гене 16s (1, 2-3). Приведенные значения представляют собой среднее значение $\pm\text{SEM}$ для трех независимых реплик.

Подробное описание изобретения

Настоящее изобретение описано ниже со ссылкой на прилагаемые чертежи и примеры, в которых показаны варианты осуществления изобретения. Это описание не предназначено в качестве подробного описания всех разных способов реализации изобретения или всех функций, которые могут быть добавлены к настоящему изобретению. Например, признаки, проиллюстрированные в отношении одного варианта осуществления, могут быть включены в другие варианты осуществления, а функции, проиллюстрированные в отношении конкретного варианта осуществления, могут быть удалены из этого варианта осуществления. Таким образом, изобретение предусматривает, что в некоторых вариантах его осуществления любой признак или комбинация признаков, изложенных в настоящем описании, могут быть исключены или опущены. Кроме того, многочисленные варианты и дополнения к предложенным в настоящем описании различным вариантам осуществления будут очевидными для специалистов в данной области исходя из приведенного раскрытия, которые не выходят за рамки настоящего изобретения. Следовательно, приведенное ниже описание предназначено для иллюстрации некоторых конкретных вариантов осуществления изобретения, а не для исчерпывающего определения всех перестановок, комбинаций и их вариантов.

Если не определено иное, то все технические и научные термины, используемые в настоящем описании, имеют те же самые значения, как они обычно понимаются специалистом в соответствующей области техники, к которой относится данное изобретение. Терминология, используемая в описании настоящего изобретения, предназначена только для описания конкретных вариантов осуществления и не предназначена для ограничения изобретения.

Все публикации, заявки на патенты, патенты и другие ссылочные документы, приведенные в настоящем описании, включены в качестве ссылки во всей своей полноте в отношении информации, относящейся к предложению и/или параграфу, в котором приведена эта ссылка.

Если контекстом не определено иное, специально предполагается, что различные признаки изобретения, приведенные в настоящем описании, могут использоваться в любой комбинации. Более того, настоящее изобретение также предусматривает, что в некоторых вариантах осуществления изобретения любой признак или комбинация признаков, изложенных в настоящем описании, могут быть исключены или опущены. С целью иллюстрации, если в описании указано, что композиция содержит компоненты А, В и С, то специально предполагается, что любой компонент из А, В или С или их комбинация могут быть опущены и исключены по отдельности или в любой комбинации.

Как используется в описании изобретения и прилагаемой формуле изобретения, формы единственного числа также включают формы множественного числа, если контекстом в явном виде не определено иное.

Также, как используется в настоящем описании, "и/или" относится к любым и всем возможным сочетаниям одного или нескольких связанных перечисленных элементов, а также к отсутствию их комбинаций при интерпретации в альтернативе ("или").

Используемый в настоящем описании термин "примерно", если он относится к измеряемой величине, такой как доза или период времени и т.п., относится к отклонениям ± 20 , ± 10 , ± 5 , ± 1 , $\pm 0,5\%$ или даже $\pm 0,1\%$ от указанного количества.

Как используется в настоящем описании, фразы, такие как "между X и Y" и "между примерно X и Y", следует интерпретировать как включающие X и Y. Как используется в настоящем описании, фразы, такие как "между примерно X и Y", означают "между примерно X и примерно Y", а фразы, такие как "от примерно X до Y", означают "от примерно X до примерно Y".

Термин "содержать", "содержит" и "содержащий", как используется в настоящем описании, определяет наличие указанных признаков, целых чисел, этапов, операций, элементов и/или компонентов, но не исключает наличия или добавления одной или нескольких других функций, целых чисел, этапов, операций, элементов, компонентов и/или их групп.

Как используется в настоящем описании, переходная фраза "состоящий в основном из" означает, что объем пункта формулы изобретения следует интерпретировать как охватывающий указанные материалы или этапы, перечисленные в этом пункте формулы изобретения, а также те, которые не оказывают существенного влияния на основную и новую характеристику(и) заявленного изобретения. Таким образом, термин "состоящий в основном из" при использовании в пункте формулы настоящего изобретения не следует интерпретировать как эквивалентный термину "содержащий".

"Cas9 нуклеаза" относится к большой группе эндонуклеаз, которые катализируют расщепление двухцепочечной ДНК в CRISPR-Cas системе. Эти полипептиды хорошо известны в уровне техники, и многие их структуры (последовательности) охарактеризованы (см., например, WO 2013/176772; WO 2013/188638). Домены для каталитического расщепления двухцепочечной ДНК представляют собой домен RuvC и домен HNH. Домен RuvC отвечает за разрыв (-)нити, а домен HNH отвечает за разрыв (+)нити (см., например, Gasiunas et al., PNAS, 109 (36):E2579-E2586 (4 сентября, 2012 г.)).

Используемый в настоящем описании термин "химерный" относится к молекуле нуклеиновой кислоты или полипептиду, в котором по меньшей мере два компонента получены из разных источников (например, разных организмов, разных кодирующих областей).

"Комплемент", как используется в настоящем описании, может означать 100% комплементарности или идентичности референсной нуклеотидной последовательности, или может означать менее 100% комплементарности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п. комплементарности).

Термины "комплементарный" или "комплементарность", используемые в настоящем описании, относятся к естественному связыванию полинуклеотидов в пермиссивных солевых и температурных условиях путем спаривания оснований. Например, последовательность "A-G-T" связывается с комплементарной последовательностью "T-C-A". Комплементарность между двумя одонитевыми молекулами может быть "частичной", при которой связываются только некоторые из нуклеотидов, или может быть полной, когда существует полная комплементарность между одноцепочечными молекулами. Степень комплементарности между нитями нуклеиновых кислот оказывает значительное влияние на эффективность и прочность гибридизации между нитями нуклеиновых кислот.

Используемый в настоящем описании термин "контакт", "контактирование", "контактируемый" и его грамматические варианты относятся к помещению компонентов требуемой реакции в условия, подходящие для проведения этой требуемой реакции (например, интеграции, трансформации, скрининга,

отбора, уничтожения, идентификации, амплификации и т.п.). Способы и условия для проведения таких реакций хорошо известны в данной области (см., например, Gasiunas et al. (2012), *Proc. Natl. Acad. Sci.*, 109:E2579-E2586; M.R. Green and J. Sambrook (2012), *Molecular Cloning: A Laboratory Manual*, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

"Делеция", как используется в настоящем описании, может содержать потерю или делецию генетического материала, включая без ограничения делецию части хромосомы или плазмиды, делецию гена или части гена из хромосомы или плазмиды. В некоторых вариантах осуществления делеция может содержать один ген или более одного гена. В некоторых вариантах осуществления делеция может также содержать потерю не кодирующих белок областей, которые могут кодировать небольшие некодирующие РНК. В некоторых вариантах осуществления делеция может содержать потерю всей плазмиды или всего мобильного генетического элемента. В некоторых вариантах осуществления изобретения потеря мобильного генетического элемента может быть определена как, например, отсутствие способности к репликации или персистированию.

В некоторых вариантах осуществления фазмида по изобретению может содержать массив CRISPR из CRISPR-Cas системы типа I, CRISPR-Cas системы типа II, CRISPR-Cas системы типа III, CRISPR-Cas системы типа IV и/или CRISPR-Cas системы типа V (см., Makarova et al., *Nature Reviews Biotechnology*, 13:722736 (2015)).

Таким образом, в некоторых вариантах осуществления, дополнительно к *crPHK* типа I, фазмида по изобретению может содержать полипептиды типа I и/или полипептиды Cascade типа I (т.е. CRISPR-Cas систему типа I).

Используемый в настоящем описании термин "полипептид типа I" относится к любому из полипептида Cas3, полипептида Cas3', полипептида Cas3", их рекомбинантных вариантов, и любому одному или более полипептидам из комплекса противовирусных полипептидов ("Cascade"), ассоциированного с короткими палиндромными повторами типа I (CRISPR). Таким образом, термин "полипептид типа I" относится к полипептидам, которые составляют CRISPR-Cas систему типа I-A, CRISPR-Cas систему типа I-B, CRISPR-Cas систему типа I-C, CRISPR-Cas систему типа I-D, CRISPR-Cas систему типа I-E, CRISPR-Cas систему типа I-F и/или CRISPR-Cas систему типа I-U. Каждая CRISPR-Cas система типа I содержит по меньшей мере один полипептид Cas3. Полипептиды Cas3 обычно содержат как домен геликазы, так и домен HD. Однако в некоторых CRISPR-Cas системах типа I геликаза и домен HD находятся в отдельных полипептидах, Cas3' и Cas3". В частности, Cas3' кодирует домен геликазы, тогда как Cas3" кодирует домен HD. Следовательно, поскольку для функционирования Cas3 необходимы обе области, подтипы типа I либо кодируют Cas3 (I-C, I-D, I-E, I-F, I-U), либо Cas3' и Cas3" (I-A, I-B).

Используемый в настоящем описании термин "полипептиды Cascade типа I" относится к комплексу полипептидов, участвующих в процессинге *pre-crPHK* и последующем связывании с ДНК-мишенью в CRISPR-Cas системах типа I. Эти полипептиды включают без ограничения полипептиды Cascade подтипов I-A, I-B, I-C, I-D, I-E и I-F типа I. Неограничивающие примеры полипептидов Cascade типа I-A включают Cas7 (Csa2), Cas8a1 (Csa13), Cas8a2 (Csa9), Cas5, Csa5, Cas6a, Cas3' и/или Cas3". Неограничивающие примеры полипептидов Cascade типа I-B включают Cas6b, Cas8b (Csh1), Cas7 (Csh2) и/или Cas5. Неограничивающие примеры полипептидов Cascade типа I-C каскада включают Cas5d, Cas8c (Csd1) и/или Cas7 (Csd2). Неограничивающие примеры полипептидов Cascade типа I-D каскада включают Cas10d (Csc3), Csc2, Csc1 и/или Cas6d. Неограничивающие примеры полипептидов Cascade типа I-E включают Cse1 (CasA), Cse2 (CasB), Cas7 (CasC), Cas5 (CasD) и/или Cas6e (CasE). Неограничивающие примеры полипептидов Cascade типа I-F включают Cys1, Cys2, Cas7 (Cys3) и/или Cas6f (Csy4). Неограничивающие примеры полипептидов Cascade типа I-U включают Cas8c, Cas7, Cas5, Cas6 и/или Cas4.

В некоторых вариантах осуществления фазмида по изобретению может содержать CRISPR-Cas систему типа II дополнительно к *crPHK* типа II. CRISPR-Cas системы типа II содержат три подтипа: тип II-A, тип II-B и тип II-C, каждый из которых содержит мультидоменный белок Cas9 дополнительно к адаптивным полипептидам Cas1, Cas2 и необязательно Csn2 и/или Cas4. Большинство локусов типа II также кодируют *tracrPHK*. Организмы, содержащие типичные CRISPR-Cas системы типа II, включают штаммы *Legionella pneumophila* Paris, *Streptococcus thermophilus* CNRZ1066 и *Neisseria lactamica* 020-06.

В дополнительных вариантах осуществления фазмида по изобретению может содержать CRISPR-Cas систему типа III дополнительно к *crPHK* типа III. Подобно CRISPR-Cas системам типа I, в системах типа III процессинг и интерференция осуществляются с участием мультибелковых эффекторных комплексов CRISPR РНК (*crPHK*) (Makarova et al., *Nature Reviews Biotechnology*, 13:722736 (2015)) - "CAS-CADE" в типе I и "Csm" или "Cmr" в типе III. Таким образом, в некоторых вариантах осуществления CRISPR-Cas система типа III может содержать комплекс Csm (например, Csm типа III-A) и/или комплекс Cmr (например, Cmr типа III-B) и необязательно полипептид Cas6. В типичных вариантах осуществления комплекс Csm может содержать полипептиды Cas10 (или Csm1), Csm2, Csm3, Csm4, Csm5 и Csm6, а комплекс Cmr может содержать полипептиды Cmr1, Cas10 (или Cmr2), Cmr3, Cmr4, Cmr5 и Cmr6. В дополнение к комплексу Csm или комплексу Cmr CRISPR-Cas, система типа III может также содержать полипептид Cas7. Охарактеризованы четыре подтипа CRISPR-Cas системы типа III: III-A, III-B, III-C, III-D. В некоторых вариантах осуществления CRISPR-Cas система типа III-A содержит полипептиды

Cas6, Cas10, Csm2, Cas7 (Csm3), Cas5 (Csm4), Cas7 (Csm5) и Csm6. В некоторых вариантах осуществления CRISPR-Cas система типа III-B содержит полипептиды Cas7 (Cmr1), Cas10, Cas5 (Cmr3), Cas7 (Cmr4), Cmr5, Cas6 и Cas7 (Cmr6). В некоторых вариантах осуществления CRISPR-Cas система типа III-C содержит полипептиды Cas7 (Cmr1), Cas7 (Cmr6), Cas10, Cas7 (Cmr4), Cmr5 и Cas5 (Cmr3). В некоторых вариантах осуществления CRISPR-Cas система типа III-D содержит полипептиды Cas10, Cas7 (Csm3), Cas5 (Csm4), Csm2, Cas7 (Csm3) и all1473.

В некоторых вариантах осуществления фазмида по изобретению может содержать CRISPR-Cas систему типа IV дополнительно к crPHK типа IV. CRISPR-Cas системы типа IV могут содержать полипептид Csf4 (dinG) и/или полипептид Csf1, Cas7 (Csf2) и/или Cas5 (csf3). (Makarova et al., *Nature Reviews Microbiology*, 13:722-736 (2015)).

В некоторых вариантах осуществления фазмида по изобретению может дополнительно содержать CRISPR-Cas систему типа V дополнительно к crPHK типа V. CRISPR-Cas системы типа V могут содержать полипептид Cpf1 и/или полипептид Cas1, Cas2 и/или Cas4. (Makarova et al., *Nature Reviews Microbiology*, 13:722-736 (2015)).

Под "фрагментом" или "частью" нуклеотидной последовательности по изобретению следует понимать нуклеотидную последовательность уменьшенной длины (например, уменьшенной на 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 или более нуклеотидов) относительно референсной последовательности нуклеиновой кислоты или нуклеотидной последовательности, содержащую, по существу состоящую и/или состоящую из нуклеотидной последовательности смежных нуклеотидов, идентичных или по существу идентичных (например, идентичных на 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99%) референсной последовательности нуклеиновой кислоты или нуклеотидной последовательности. Такой фрагмент или часть нуклеиновой кислоты в соответствии с изобретением может быть в случае необходимости включенной в более крупный полинуклеотид, частью которого указанный фрагмент или часть является. Таким образом, гибридизующийся (или гибридизуется с, и другие грамматические варианты этого выражения), например, по меньшей мере частью ДНК-мишенью (например, областью-мишенью в геноме) относится к гибридизации с нуклеотидной последовательностью, которая идентична или по существу идентична по длине смежным нуклеотидам ДНК-мишени. В некоторых вариантах повтор последовательности спейсер-повтор или последовательности повтор-спейсер-повтор может содержать фрагмент повторяющейся последовательности локуса CRISPR дикого типа или повторяющейся последовательности синтетического массива CRISPR, причем фрагмент повтора сохраняет функцию повтора в массиве CRISPR - гибридизации с нуклеиновой кислотой *tracr*.

В некоторых вариантах осуществления изобретение может содержать функциональный фрагмент нуклеазы Cas9, Cas3, Cas3', Cas3" или Cpf1. Функциональный фрагмент Cas9 сохраняет одну или более видов активности нативной нуклеазы Cas9, включая без ограничения активность нуклеазы HNH, активность нуклеазы RuvC, активности распознавания и связывания ДНК, РНК и/или РАМ. Функциональный фрагмент нуклеазы Cas9 может быть кодирован фрагментом полинуклеотида Cas9. Функциональный фрагмент Cas3, Cas3' или Cas3" сохраняет одну или более активностей нативной нуклеазы Cas9, включая без ограничения активность нуклеазы, активность экзонуклеазы, связывание ДНК и/или связывание РНК. Функциональный фрагмент нуклеазы Cas3, Cas3' или Cas3" может быть кодирован фрагментом полинуклеотида Cas3, Cas3' или Cas3" соответственно.

Используемый в настоящем описании термин "ген" относится к молекуле нуклеиновой кислоты, которая может быть использована для продуцирования мРНК, антисмысловой РНК, РНКi (miРНК, siРНК, shРНК), анти-микроРНК, антисмысловой олигонуклеотида (АМО) и т.п. Гены могут или не могут использоваться для продуцирования функционального белка или продукта гена. Гены могут включать как кодирующие, так и не кодирующие области (например, интроны, регуляторные элементы, промоторы, энхансеры, терминальные последовательности и/или 5' и 3' нетранслируемые области). Ген может быть "выделен", что подразумевает нуклеиновую кислоту, которая по существу или практически не содержит компонентов, которые обычно связаны с этой нуклеиновой кислотой в ее естественном состоянии. Такие компоненты включают другой клеточный материал, культуральную среду из рекомбинантного продуцирования и/или различные химические вещества, используемые для химического синтеза нуклеиновой кислоты.

Используемый в настоящем описании термин "геном" включает хромосомный/ядерный геном организма, а также любой митохондриальный и/или плазмидный геном.

Используемый в настоящем описании термин "шпильчатая последовательность" представляет собой нуклеотидную последовательность, содержащую шпильки (например, которая образует одну или более шпильчатых структур). Шпилька (например, "петля-на-стебле", конъюгация "в себе") относится к молекуле нуклеиновой кислоты, имеющей вторичную структуру, которая включает область комплементарных нуклеотидов, образующих двойную нить, которая далее с обеих сторон фланкирована одноцепочечными областями. Такие структуры хорошо известны в данной области техники. Как известно в данной области техники, двухнитевая область может содержать некоторые несоответствия в спаривании оснований или может быть абсолютно комплементарной. В некоторых вариантах осуществления настоящего

раскрытия шпильчатая последовательность конструкции нуклеиновой кислоты может быть расположена на 3'-конце *tracg* нуклеиновой кислоты.

"Гетерологичная" или "рекомбинантная" нуклеотидная последовательность представляет собой нуклеотидную последовательность, не связанную естественным образом с клеткой-хозяином, в которую она вводится, включая не встречающиеся в природе многочисленные копии естественной нуклеотидной последовательности.

Различные нуклеиновые кислоты или белки, имеющие гомологию, в настоящем описании именуется "гомологами". Термин гомолог включает гомологичные последовательности из одного и того же и других видов и ортологичные последовательности из одного и того же и других видов. "Гомология" относится к уровню сходства между двумя или более последовательностями нуклеиновой кислоты и/или аминокислоты, выраженному в процентах от позиционной идентичности (т.е. сходства или идентичности последовательностей). Гомология также относится к понятию сходных функциональных свойств между различными нуклеиновыми кислотами или белками. Таким образом, композиции и способы по изобретению дополнительно включают гомологи нуклеотидных последовательностей и полипептидных последовательностей по изобретению. "Ортологичный", как используется в настоящем описании, относится к гомологичным нуклеотидным последовательностям и/или аминокислотным последовательностям у разных видов, которые возникли из общего гена-предка во время видообразования. Гомолог нуклеотидной последовательности по настоящему изобретению обладает значимой идентичностью последовательности (например, по меньшей мере примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 и/или 100%) с указанной нуклеотидной последовательностью по изобретению. Так, например, гомолог полинуклеотида или полипептида типа I, типа II, типа III, типа IV или типа V может быть примерно на 70% или более гомологичным любому известному или позже идентифицированному полинуклеотиду или полипептиду типа I, типа II, типа III, типа IV или типа V.

Как используются в настоящем описании, гибридизация, гибридизуется, гибридизующийся и их грамматические варианты относятся к связыванию двух полностью комплементарных нуклеотидных последовательностей или по существу комплементарных последовательностей, в которых могут присутствовать некоторые несогласованные пары оснований. Условия гибридизации хорошо известны в данной области и изменяются в зависимости от длины нуклеотидных последовательностей и степени комплементарности между нуклеотидными последовательностями. В некоторых вариантах осуществления условия гибридизации могут быть очень жесткими или они могут быть средней жесткости или низкой жесткости в зависимости от степени комплементарности и длины гибридизуемых последовательностей. Условия, которые относятся к условиям низкой, средней и высокой жесткости гибридизации нуклеотидных последовательностей хорошо известны в данной области (см., например, Gasiunas et al. (2012), *Proc. Natl. Acad. Sci.*, 109:E2579-E2586; M.R. Green and J. Sambrook (2012), *Molecular Cloning: A Laboratory Manual*, 4th ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

Используемые в настоящем описании термины "увеличивать", "увеличивающийся", "увеличенный", "усиливать", "усиливающийся", "усиленный" и "улучшение" (и их грамматические варианты) описывают повышение по меньшей мере на примерно 25, 50, 75, 100, 150, 200, 300, 400, 500% или более по сравнению с контролем.

"Нативная" или "дикого типа" нуклеиновая кислота, нуклеотидная последовательность, полипептидная или аминокислотная последовательность относятся к встречающейся в природе или эндогенной нуклеиновой кислоте, нуклеотидной последовательности, полипептидной или аминокислотной последовательности. Так, например, "мРНК дикого типа" представляет собой мРНК, которая в естественных условиях встречается в организме или является эндогенной. "Гомологичная" последовательность нуклеиновой кислоты представляет собой нуклеотидную последовательность, естественным образом связанную с клеткой-хозяином, в которую она введена.

Также используемые в настоящем описании термины "нуклеиновая кислота", "молекула нуклеиновой кислоты", "конструкция нуклеиновой кислоты", "нуклеотидная последовательность" и "полинуклеотид" относятся к РНК или ДНК, которая является линейной или разветвленной, одно- или двухцепочечной или гибридной. Этот термин также включает РНК/ДНК гибриды. Когда dsРНК продуцируется синтетически, менее общие основания, такие как инозин, 5-метилцитозин, 6-метиладенин, гипоксантин и другие, также могут быть использованы для антисмысловой dsРНК и спаривания рибозимов. Например, было показано, что полинуклеотиды, содержащие C-5-пропиновые аналоги уридина и цитидина, связываются с РНК с высоким сродством и являются мощными антисмысловыми ингибиторами экспрессии генов. Другие модификации, такие как модификация фосфодиэфирного скелета, или 2'-гидрокси в группе сахара рибозы РНК также возможны. Конструкции нуклеиновых кислот по настоящему изобретению могут быть ДНК или РНК, но предпочтительно представляют собой ДНК. Таким образом, хотя конструкции нуклеиновых кислот по настоящему изобретению могут быть описаны и использованы в форме ДНК, в зависимости от предполагаемого использования они также могут быть описаны и использованы в форме РНК.

Используемый здесь термин "нуклеотидная последовательность" относится к гетерополимеру нуклеотидов или последовательности этих нуклеотидов в направлении от 5'- к 3'-концу молекулы нуклеино-

вой кислоты и включает молекулы ДНК или РНК, включая кДНК, фрагмент или часть ДНК, геномную ДНК, синтетическую (например, химически синтезированную) ДНК, плазмидную ДНК, мРНК и анти-смысловую РНК, любая из которых может быть одноцепочечной или двухцепочечной. Термины "нуклеотидная последовательность", "нуклеиновая кислота", "молекула нуклеиновой кислоты", "олигонуклеотид" и "полинуклеотид" также используются в настоящем описании взаимозаменяемо для обозначения гетерополимера нуклеотидов. Если не указано иное, молекулы нуклеиновой кислоты и/или нуклеотидные последовательности, представленные в настоящем описании, приведены в 5'-3' направлении слева направо и представлены в виде стандартного кода для представления нуклеотидов в виде символов, как указано в правилах США для последовательностей, 37 CFR 1.821-1.825 и стандарте ST.25 Всемирной организации интеллектуальной собственности (ВОИС).

Используемый в настоящем описании термин "процентная идентичность последовательности" или "процентная идентичность" относится к проценту идентичных нуклеотидов в линейной полинуклеотидной последовательности референсной ("запрашиваемой") полинуклеотидной молекулы (или ее комплементарной цепи) по сравнению с тестовой ("испытуемой") полинуклеотидной молекулой (или ее комплементарной цепью), когда две последовательности выровнены оптимальным образом. В некоторых вариантах осуществления "процентная идентичность" может относиться к проценту идентичных аминокислот в аминокислотной последовательности.

"Протоспейсерная последовательность" относится к двухцепочечной ДНК-мишени и, в частности, к части ДНК-мишени (например, к области-мишени в геноме), которая полностью или по существу комплементарна (и гибридизуется) со спейсерной последовательностью повтор-спейсер CRISPR последовательностей повтор-спейсер, CRISPR последовательностей повтор-спейсер-повтор и/или массива CRISPR.

Используемые в настоящем описании термины "уменьшить", "уменьшенный", "уменьшающий", "уменьшение", "убывать", "подавлять" и "снижать" (и их грамматические варианты) описывают, например, уменьшение на по меньшей мере примерно 5, 10, 15, 20, 25, 35, 50, 75, 80, 85, 90, 95, 97, 98, 99 или 100% по сравнению с контролем. В конкретных вариантах осуществления уменьшение может приводить к отсутствию или по существу отсутствию (т.е. незначительному количеству, например, менее чем примерно 10% или даже 5%) обнаруживаемой активности или количества измеряемого компонента (например, популяции клеток или размера генома). Так, например, уменьшенный размер генома может означать уменьшение размера генома на по меньшей мере примерно 5, 10, 15, 20, 25, 35, 50, 75, 80, 85, 90, 95, 97, 98, 99 или 100% по сравнению с контролем.

Используемый в настоящем описании контроль может представлять собой, например, популяцию бактериальных, архейных, водорослевых или дрожжевых клеток, которые не были трансформированы конструкцией гетерологичной нуклеиновой кислоты по настоящему изобретению. В некоторых вариантах осуществления контроль может представлять собой популяцию дикого типа бактериальных, архейных, водорослевых или дрожжевых клеток или она может быть популяцией бактериальных, архейных или дрожжевых клеток, трансформированных гетерологичной конструкцией, содержащей массив CRISPR, который содержит последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая не является комплементарной области-мишени в геноме бактериальных, архейных или дрожжевых клеток указанной популяции (т.е. несамонацеленный/"скремблированный спейсер"). В дополнительных аспектах контроль может представлять собой, например, популяцию дикого типа бактериальных, архейных, водорослевых или дрожжевых клеток или популяцию бактериальных, архейных, водорослевых или дрожжевых клеток, трансформированных гетерологичной конструкцией, содержащей массив CRISPR, который содержит последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, которая не расположена в непосредственной близости от мотива, примыкающего к протоспейсеру (PAM).

Используемая в настоящем описании "повторяющаяся последовательность" относится, например, к любой повторяющейся последовательности локуса CRISPR дикого типа или повторяющейся последовательности синтетического массива CRISPR, которая отделена "спейсерной последовательностью" (например, последовательность повтор-спейсер или последовательность повтор-спейсер-повтор по изобретению). Повторяющаяся последовательность, используемая в настоящем изобретении, может быть любой известной или позже идентифицированной повторяющейся последовательностью локуса CRISPR. Соответственно, в некоторых вариантах осуществления последовательность повтор-спейсер или повтор-спейсер-повтор содержит повтор, который по существу идентичен (например, идентичен на по меньшей мере примерно 70% (например, идентичен на по меньшей мере примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и более)) повтору из массива CRISPR типа II дикого типа. В некоторых вариантах осуществления повторяющаяся последовательность на 100% идентична повтору из массива CRISPR типа I дикого типа, массива CRISPR типа II дикого типа, массива CRISPR типа III дикого типа, массива CRISPR типа IV дикого типа или массива CRISPR типа V дикого типа. В дополнительных вариантах осуществления повторяющаяся последовательность, исполь-

зуемая в настоящем изобретении, может содержать нуклеотидную последовательность, включающую частичный повтор, который представляет собой фрагмент или часть последовательных нуклеотидов повторяющейся последовательности локуса CRISPR или синтетического массива CRISPR любой из crPHK типа I, crPHK типа II, crPHK типа III, crPHK типа IV или crPHK типа V.

Используемый в настоящем описании термин "массив CRISPR" CRISPR-Cas системы типа I, типа II, типа III, типа IV или типа V относится к конструкции нуклеиновой кислоты, которая содержит в 5'-3' направлении последовательность повтор-спейсер-повтор или содержит в 5'-3' направлении по меньшей мере одну последовательность повтор-спейсер (например, примерно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 или 25 последовательностей повтор-спейсер и любой диапазон или количество в этом диапазоне). Когда в массиве CRISPR содержится более одного повтора-спейсера, спейсер предыдущей (5'-3') последовательности повтор-спейсер может быть связан с повтором следующего повтора-спейсера (например, спейсер первой последовательности повтор-спейсер связан с повтором второй последовательности повтор-спейсер). В некоторых вариантах осуществления массив CRISPR может содержать два повтора (или два частичных повтора), разделенных спейсером (например, последовательность повтор-спейсер-повтор).

Используемый в настоящем описании термин "идентичность последовательности" относится к степени, до которой две оптимально выровненные полинуклеотидные или пептидные последовательности являются инвариантными во всем окне выравнивания компонентов, например нуклеотидов или аминокислот. "Идентичность" можно легко рассчитать известными способами, включая без ограничения способы, которые описаны в "Вычислительная молекулярная биология" [Computational Molecular Biology] (Lesk, A.M., ed.), Oxford University Press, New York (1988); Биокomпьютинг: проекты в области информатики и генома [Biocomputing: Informatics and Genome Projects] (Smith, D.W., ed.), Academic Press, New York (1993); Компьютерный анализ данных последовательности, часть I [Computer Analysis of Sequence Data, Part I] (Griffin, A.M., Griffin, H.G., eds.) Humana Press, New Jersey (1994); Анализ последовательности в молекулярной биологии Sequence [Analysis in Molecular Biology] (von Heinje, G., ed.) Academic Press (1987); и Введение в анализ последовательностей [Sequence Analysis Primer] (Gribskov, M., Devereux, J., eds.), Stockton Press, New York (1991).

"Спейсерная последовательность", как используется в настоящем описании, представляет собой нуклеотидную последовательность, которая комплементарна ДНК-мишени (т.е. области-мишени в геноме или "протоспейсерная последовательность"), которая прилегает к последовательности мотива, примыкающего к протоспейсеру (PAM). Спейсерная последовательность может быть полностью комплементарной или по существу комплементарной (например, на по меньшей мере примерно 70% комплементарной (например, на примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и более)) ДНК-мишени. В типичных вариантах осуществления спейсерная последовательность имеет 100% комплементарность ДНК-мишени. В дополнительных вариантах осуществления комплементарность 3'-области спейсерной последовательности с ДНК-мишенью составляет 100%, но менее 100% с 5'-областью спейсера, и поэтому общая комплементарность спейсерной последовательности с ДНК-мишенью менее 100%. Так, например, первые 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 и т.п. нуклеотиды в 3'-области 20-нуклеотидной спейсерной последовательности (затравочной последовательности) могут быть комплементарными с ДНК-мишенью на 100%, в то время как оставшиеся нуклеотиды в 5'-области спейсерной последовательности являются по существу комплементарными (например, на по меньшей мере, примерно 70% комплементарными) ДНК-мишени. В некоторых вариантах осуществления первые 7-12 нуклеотидов спейсерной последовательности могут быть комплементарными с ДНК-мишенью на 100%, тогда как остальные нуклеотиды в 5'-области спейсерной последовательности являются по существу комплементарными (например, на по меньшей мере, примерно на 70% комплементарными) ДНК-мишени. В других вариантах осуществления первые 7-10 нуклеотидов спейсерной последовательности могут быть на 100% комплементарными с ДНК-мишенью, тогда как остальные нуклеотиды в 5'-области спейсерной последовательности являются по существу комплементарными (например, на по меньшей мере, примерно 70% комплементарными) ДНК-мишени. В типичных вариантах осуществления первые 7 нуклеотидов (внутри затравочной) спейсерной последовательности могут быть комплементарными с ДНК-мишенью на 100%, тогда как остальные нуклеотиды в 5'-области спейсерной последовательности являются по существу комплементарными (например, на по меньшей мере, примерно 70% комплементарными) ДНК-мишени.

Используемый в настоящем описании термин "ДНК-мишень", "область-мишень" или "область-мишень в геноме" относится к области генома в организме, которая является полностью комплементарной или по существу комплементарной (например, на по меньшей мере 70% комплементарной (например, на 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82%, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и более)) спейсерной последовательности в последовательности повтор-спейсер или повтор-спейсер-повтор. В некоторых вариантах осуществления область-мишень может составлять в длину от примерно 10 до примерно 40 последовательных нуклеотидов, непосредственно примыкающих к последовательности PAM (последовательности PAM, расположенной непосредственно после 3'-области-мишени) в геноме организма (например, CRISPR-Cas системы типа I и CRISPR-Cas типа II). В некоторых

вариантах осуществления, например, в системах типа I PAM находится на противоположной стороне от протоспейсера (5'-конце). Для систем типа III PAM не известен. Макарова и др. описывают номенклатуру для всех классов, типов и подтипов CRISPR систем (Nature Reviews Microbiology, 13:722-736 (2015)). Структуры гидов и PAM описаны у R. Barrangou (Genome Biol., 16:247 (2015)).

В некоторых вариантах осуществления область-мишень, используемая в настоящем изобретении, находится в пределах жизненно важного гена гена или нежизненно важного гена.

В типичных вариантах осуществления область-мишень может быть выбрана случайным образом или может быть выбрана специфически. В некоторых вариантах осуществления случайно выбранная область-мишень может быть выбрана из любых по меньшей мере 10 последовательных нуклеотидов (например, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50 и т.п., и любого диапазона или значения в этом диапазоне), расположенных в непосредственной близости от PAM последовательности в геноме бактерий, архей, водорослей или дрожжей. В некоторых вариантах осуществления область-мишень может составлять от примерно 10 до примерно 20 последовательных нуклеотидов, от примерно 10 до примерно 30 последовательных нуклеотидов и/или от примерно 10 до примерно 40 последовательных нуклеотидов и т.п. или в любом диапазоне или при любом значении в этом диапазоне, расположенных в непосредственной близости от последовательности мотива, примыкающего к протоспейсеру (PAM), в геноме бактерий, архей, водорослей или дрожжей. В некоторых вариантах осуществления конкретный выбор области-мишени может включать выбор двух или более областей-мишеней, которые расположены друг от друга через от примерно каждых 100 нуклеотидов до примерно каждых 1000 нуклеотидов, от примерно каждых 100 нуклеотидов до примерно каждых 2000, от примерно каждых 100 нуклеотидов до примерно каждых 3000, от примерно каждых 100 нуклеотидов до примерно каждых 4000 и/или от примерно каждых 100 нуклеотидов до примерно каждых 5000 нуклеотидов и т.п. в геноме одной или нескольких бактерий, архей, водорослей или дрожжевых клеток. В конкретных вариантах осуществления конкретный выбор области-мишени содержит конкретный выбор области-мишени из гена, открытой рамки считывания, предполагаемой открытой рамки считывания или межгенной области, содержащей по меньшей мере от примерно 10 до примерно 40 последовательных нуклеотидов, непосредственно примыкающих к PAM последовательности в геноме бактерий, архей, водорослей или дрожжей.

"Транс-активирующая CRISPR (tracr) нуклеиновая кислота" или "tracr нуклеиновая кислота", как используется в настоящем описании, относится к любой tracr РНК (или кодирующей ее ДНК). Tracr нуклеиновая кислота содержит в 5'-3' направлении нижний стебель, верхний стебель, выпетливание, центральную шпильку и концевые шпильки (см. Briner et al. (2014), Molecular Cell., 56(2):333-339). Транс-активирующая CRISPR (tracr) нуклеиновая кислота функционирует при гибридизации с повторяющейся частью зрелых или незрелых crРНК, рекрутирует белок Cas9 к участку-мишени и может способствовать каталитической активности Cas9 путем индукции структурной перегруппировки. Функциональный состав молекул tracrРНК приведен выше. Последовательности tracrРНК специфичны для CRISPR-Cas системы и могут меняться. В настоящем изобретении можно использовать любую известную или позже идентифицированную нуклеиновую кислоту.

Используемый в настоящем описании термин "по существу идентичный" или "существенная идентичность" в контексте двух молекул нуклеиновых кислот, нуклеотидных последовательностей или белковых последовательностей относится к двум или более последовательностям или подпоследовательностям, которые имеют по меньшей мере примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 и/или 100% идентичности нуклеотидов или аминокислотных остатков при сравнении и выравнивании для максимального соответствия, измеренной при помощи одного из приведенных ниже алгоритмов сравнения последовательностей или путем визуального осмотра. В некоторых вариантах осуществления изобретения существенная идентичность имеется в области последовательностей, которая составляет по меньшей мере от примерно 50 остатков до примерно 150 остатков в длину. Таким образом, в некоторых вариантах осуществления изобретения существенная идентичность имеется в области последовательностей, которая составляет по меньшей мере от 3 до 15 (например, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 остатков в длину и т.п. или любое значение или любой диапазон этих значений), от по меньшей мере примерно 5 до примерно 30, от по меньшей мере примерно 10 до примерно 30, от по меньшей мере примерно 16 до примерно 30, от по меньшей мере 18 до примерно 25, по меньшей мере примерно 18, по меньшей мере примерно 22, по меньшей мере примерно 25, по меньшей мере примерно 30, по меньшей мере примерно 40, по меньшей мере примерно 50, примерно 60, примерно 70, примерно 80, примерно 90, примерно 100, примерно 110, примерно 120, примерно 130, примерно 140, примерно 150 или более остатков в длину и в любом указанном диапазоне. В типичных вариантах осуществления последовательности могут быть по существу идентичными в пределах по меньшей мере примерно 22 нуклеотидов. В некоторых конкретных вариантах осуществления последовательности являются по существу идентичными в пределах по меньшей мере примерно 150 остатков. В некоторых вариантах осуществления последовательности по изобретению могут быть идентичными на от примерно 70% до примерно 100% в пределах по меньшей мере от примерно 16 нуклеотидов до примерно 25 нуклеотидов. В некоторых вариантах осуществления последовательности по изобретению мо-

гут быть идентичными на от примерно 75% до примерно 100% в пределах по меньшей мере от примерно 16 нуклеотидов до примерно 25 нуклеотидов. В некоторых вариантах осуществления последовательности по изобретению могут быть идентичными на от примерно 80% до примерно 100% в пределах по меньшей мере от примерно 16 нуклеотидов до примерно 25 нуклеотидов. В некоторых вариантах осуществления последовательности по изобретению могут быть идентичными на от примерно 80% до примерно 100% в пределах по меньшей мере от примерно 7 нуклеотидов до примерно 25 нуклеотидов. В некоторых вариантах осуществления последовательности по изобретению могут быть идентичными на примерно 70% в пределах по меньшей мере примерно 16 нуклеотидов. В других вариантах осуществления последовательности могут быть идентичными на примерно 85% в пределах примерно 22 нуклеотидов. В других вариантах осуществления последовательности могут быть идентичными на примерно 100% в пределах примерно 16 нуклеотидов. В другом варианте осуществления последовательности являются по существу идентичными по всей длине кодирующей области. Кроме того, в иллюстративных вариантах осуществления по существу идентичные нуклеотидные или полипептидные последовательности выполняют по существу одну и ту же функцию (например, функцию или активность *sgRNA*, *tracrRNA* нуклеиновой кислоты, повторяющейся последовательности, нуклеазы *Cas9* (никазы, распознавания и связывания ДНК, РНК и/или РАР), *Cas3*, *Cas3'*, *Cas3''* или любого другого полинуклеотида или полипептида CRISPR-Cas).

При сравнении последовательностей, как правило, одна последовательность действует как референсная последовательность, с которой сравнивают тестовые последовательности. Когда используется алгоритм сравнения последовательностей, тестовые и референсные последовательности вводят в компьютер, при необходимости обозначают координаты подпоследовательности и задают параметры программы для алгоритма сравнения последовательностей. Затем, используя этот алгоритм сравнения последовательностей, вычисляют процентную идентичность тестовой последовательности(последовательностей) относительно референсной последовательности исходя из заданных параметров программы.

Оптимальное выравнивание последовательностей, окно сравнения для выравнивания хорошо известны специалистам в данной области и может быть выполнено при помощи таких инструментов, как алгоритм локальной гомологии Смита и Уотермана, алгоритм выравнивания областей гомологии Нидмана и Вунша, метода поиска подобий Пирсона и Липмана, а также необязательно при помощи реализации этих алгоритмов в виде компьютерных программ, таких как GAP, BESTFIT, FASTA и TFASTA, которые доступны как часть пакета GCG® Wisconsin Package® (Accelrys Inc., San Diego, CA). "Доля идентичности" для выровненных сегментов тестовой последовательности и референсной последовательности представляет собой количество идентичных компонентов, которые являются общими для двух выровненных последовательностей, деленное на общее количество компонентов в сегменте референсной последовательности, т.е. всей референсной последовательности или меньше определенной части референсной последовательности. Процентная идентификация последовательности представлена в виде доли идентичности, умноженной на 100. Сравнение одной или более полинуклеотидных последовательностей можно выполнять с полноразмерной полинуклеотидной последовательностью или ее частью или более длинной полинуклеотидной последовательностью. Для целей настоящего изобретения "процентная идентичность" также может быть определена при помощи программы BLASTX версии 2.0 для транслируемых нуклеотидных последовательностей и BLASTN версии 2.0 для полинуклеотидных последовательностей.

Программное обеспечение для выполнения BLAST-анализов находится в открытом доступе в Национальном центре биотехнологической информации. Этот алгоритм включает сначала идентификацию пар последовательностей, имеющих высокий показатель (HSP), путем идентификации коротких слов длиной *W* в запрашиваемой последовательности, которые либо совпадают либо удовлетворяют некоторому положительному пороговому значению *T* при выравнивании со словом той же длины в последовательности базы данных. *T* называется порогом показателя сходства слов (Altschul et al., 1990). Эти начальные совпадения исходных слов действуют в качестве затравок для начала поиска более длинных HSP, содержащих их. Затем совпадения слов расширяют в обоих направлениях вдоль каждой последовательности до той степени, до которой можно увеличить совокупный показатель выравнивания. Совокупный показатель рассчитывают, используя для нуклеотидных последовательностей, параметры *M* (оценка вознаграждения для пары совпадающих остатков, всегда >0) и *N* (оценка штрафа за несовпадающие остатки, всегда <0). Для аминокислотных последовательностей для вычисления совокупного показателя используется оценочная матрица. Расширение совпадений слов в каждом направлении прекращается, когда совокупный показатель выравнивания уменьшается на величину *X* от своего максимально достигнутого значения, совокупный показатель сводится к нулю или падает ниже из-за накопления одного или более выравниваний остатков с отрицательной оценкой, или при достижении конца любой последовательности. Параметры алгоритма BLAST *W*, *T* и *X* определяют чувствительность и скорость выравнивания. В программе BLASTN (для нуклеотидных последовательностей) по умолчанию используется длина слова (*W*) 11, ожидание (*E*) 10, отсечение 100, *M*=5, *N*=4, а также сравнение обеих нитей. Для аминокислотных последовательностей в программе BLASTP по умолчанию используется длина слова (*W*) 3, ма-

тематическое ожидание (E) 10 и оценочная матрица BLOSUM62 (см. Henikoff & Henikoff, Proc. Natl. Acad. Sci., USA, 89:10915 (1989)).

Помимо вычисления процентной идентичности последовательностей, алгоритм BLAST также выполняет статистический анализ сходства между двумя последовательностями (см., например, Karlin & Altschul, Proc. Nat'l. Acad. Sci., USA, 90:5873-5787 (1993)). Одной из мер сходства, обеспечиваемой алгоритмом BLAST, является наименьшая сумма вероятностей ($P(N)$), которая указывает на вероятность, с которой совпадение двух нуклеотидных или аминокислотных последовательностей может происходить случайно. Например, тестовую последовательность нуклеиновой кислоты считают совпадающей с референсной последовательностью, если наименьшая сумма вероятностей при сравнении тестовой нуклеотидной последовательности с референсной нуклеотидной последовательностью составляет менее примерно от 0,1 до менее примерно 0,001. Таким образом, в некоторых вариантах осуществления изобретения наименьшая сумма вероятностей при сравнении тестовой нуклеотидной последовательности с референсной нуклеотидной последовательностью составляет менее примерно 0,001.

Две нуклеотидные последовательности также можно считать по существу комплементарными, когда две последовательности гибридизуются друг с другом в жестких условиях. В некоторых типичных вариантах осуществления две нуклеотидные последовательности, которые по существу комплементарны, гибридизуются друг с другом в очень жестких условиях.

"Жесткие условия гибридизации" и "жесткие условия промывки при гибридизации" в контексте экспериментов по гибридизации нуклеиновой кислоты, такой как нозерн-гибридизация и саузерн-гибридизация, зависят от последовательности и различаются по различным параметрам окружающей среды. Подробное руководство по гибридизации нуклеиновых кислот можно найти у Tijssen Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes, part I, chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993). Как правило, очень жесткие условия гибридизации и промывки выбирают таким образом, чтобы температура была на примерно 5°C ниже, чем температура плавления (T_m) для конкретной последовательности при определенной ионной силе и pH.

T_m представляет собой температуру (при определенной ионной силе и pH), при которой 50% последовательности-мишени гибридизуется с идеально подобранным зондом. Очень жесткие условия выбирают равными значению T_m для конкретного зонда. Пример жестких условий гибридизации для гибридизации комплементарных нуклеотидных последовательностей, которые имеют более 100 комплементарных остатков на фильтре в саузерн- или нозерн-блот анализе, составляет 50% формамид с 1 мг гепарина при 42°C, причем гибридизацию проводят в течение ночи. Примером очень жестких условий промывки является 0,15 M NaCl при 72°C в течение примерно 15 мин. Примером жестких условий промывки является промывка 0,2× SSC при 65°C в течение 15 мин (см. Sambrook, *infra*, описание SSC буфера). Зачастую перед промывкой в очень жестких условиях выполняют промывку в условиях низкой жесткости для удаления фонового сигнала зонда. Пример промывки в условиях средней жесткости для дуплекса, например, из более 100 нуклеотидов, составляет 1×SSC при 45°C в течение 15 мин. Пример промывки в условиях низкой жесткости для дуплекса, например, из более 100 нуклеотидов, составляет 4-6×SSC при 40°C в течение 15 мин. Для коротких зондов (например, от примерно 10 до 50 нуклеотидов) жесткие условия обычно включают концентрации солей менее примерно 1,0 M Na-иона, обычно примерно 0,01-1,0 M концентрации Na-ионов (или других солей) при pH 7,0- 8,3, и температура обычно составляет по меньшей мере примерно 30°C. Жесткие условия также могут быть достигнуты путем добавления дестабилизирующих агентов, таких как формамид. В общем случае отношение сигнал/шум, превышающий в 2 (или более) раз отношение, наблюдаемое для несвязанного зонда в конкретном анализе гибридизации, указывает на обнаружение специфической гибридизации. Нуклеотидные последовательности, которые не гибридизуются друг с другом в жестких условиях, по существу идентичны, если белки, которые они кодируют, являются по существу идентичными. Это возможно, например, когда копия нуклеотидной последовательности создается при использовании максимального количества вырожденных кодонов, допускаемого генетическим кодом.

Ниже приведены примеры наборов условий гибридизации/промывки, которые могут быть использованы для клонирования гомологичных нуклеотидных последовательностей, которые по существу идентичны референсным нуклеотидным последовательностям по изобретению. В одном из вариантов осуществления референсная нуклеотидная последовательность гибридизуется с "тестовой" нуклеотидной последовательностью в 7% додецилсульфате натрия (SDS), 0,5 M NaPO₄, 1 mM ЭДТА при 50°C с промывкой в 2× SSC, 0,1% SDS при 50°C. В другом варианте референсная нуклеотидная последовательность гибридизуется с "тестовой" нуклеотидной последовательностью в 7% додецилсульфате натрия (SDS), 0,5 M NaPO₄, 1 mM ЭДТА при 50°C с промывкой в 1× SSC, 0,1% SDS при 50°C или в 7% додецилсульфате натрия (SDS), 0,5 M NaPO₄, 1 mM ЭДТА при 50°C с промывкой в 0,5× SSC, 0,1% SDS при 50°C. В другом варианте осуществления референсная нуклеотидная последовательность гибридизуется с "тестовой" нуклеотидной последовательностью в 7% додецилсульфате натрия (SDS), 0,5 M NaPO₄, 1 mM ЭДТА при 50°C с промывкой в 0,1× SSC, 0,1% SDS при 50°C или в 7% додецилсульфате натрия (SDS),

0,5 М NaPO₄, 1 мМ ЭДТА при 50°C с промывкой в 0.1× SSC, 0.1% SDS при 65°C.

Любая нуклеотидная последовательность и/или гетерологичная конструкция нуклеиновой кислоты по настоящему изобретению может быть оптимизирована по кодонам для экспрессии в любых представляющих интерес видах. Оптимизация по кодонам хорошо известна в данной области и включает модификацию нуклеотидной последовательности для смещения в использовании кодонов при помощи таблиц использования кодонов для конкретных видов. Таблицы использования кодонов генерируются на основе анализа последовательностей генов с самой высокой степенью экспрессии для представляющих интерес видов. Если нуклеотидные последовательности должны экспрессироваться в ядре, таблицы использования кодонов генерируются на основе анализа последовательностей ядерных генов с самой высокой степенью экспрессии для представляющих интерес видов. Модификации нуклеотидных последовательностей определяют путем сравнения таблицы использования кодонов для конкретного вида с кодами, присутствующими в нативных полинуклеотидных последовательностях. Как понятно из уровня техники, оптимизация кодонов нуклеотидной последовательности приводит к получению нуклеотидной последовательности, имеющей менее 100% идентичности (например, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с нативной нуклеотидной последовательностью, но которая по-прежнему кодирует полипептид, имеющий ту же функцию, что и у полипептида, кодируемого исходной нативной нуклеотидной последовательностью. Таким образом, в типичных вариантах осуществления изобретения нуклеотидная последовательность и/или гетерологичная конструкция нуклеиновой кислоты по настоящему изобретению может быть оптимизирована по кодонам для экспрессии в конкретных представляющих интерес видах.

В некоторых вариантах осуществления молекулы гетерологичных или рекомбинантных нуклеиновых кислот, нуклеотидные последовательности и/или полипептиды по изобретению являются "выделенными". "Выделенная" молекула нуклеиновой кислоты, "выделенная" нуклеотидная последовательность или "выделенный" полипептид представляет собой молекулу нуклеиновой кислоты, нуклеотидную последовательность или полипептид, который в результате вмешательства человека существует отдельно от его родной среды и, следовательно, не является продуктом природы.

Выделенная молекула нуклеиновой кислоты, нуклеотидная последовательность или полипептид могут существовать в очищенной форме, которая по меньшей мере частично отделена от по меньшей мере некоторых других компонентов встречающегося в природе организма или вируса, например клеточных или вирусных структурных компонентов или других полипептидов или нуклеиновых кислот, которые обычно связаны с этим полинуклеотидом. В типичных вариантах осуществления выделенная молекула нуклеиновой кислоты, чистота выделенной нуклеотидной последовательности и/или выделенного полипептида составляет по меньшей мере примерно 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 95% или более.

В других вариантах осуществления выделенная молекула нуклеиновой кислоты, нуклеотидная последовательность или полипептид может существовать в неприродной среде, такой как, например, рекомбинантная клетка-хозяин. Так, например, в отношении нуклеотидных последовательностей термин "выделенный" означает, что нуклеотидная последовательность отделена от хромосомы и/или клетки, в которой она встречается в естественных условиях. Полинуклеотид также является выделенным, если он был отделен от хромосомы и/или клетки, в которой он встречается в естественных условиях и затем вставлен в генетический контекст, хромосому и/или клетку, в которой он не встречается в естественных условиях (например, другую клетку-хозяина, разные регуляторные последовательности и/или другое положение в геноме, отличающегося от положения в естественных условиях). Соответственно гетерологичные конструкции нуклеиновых кислот, нуклеотидные последовательности и кодируемые ими полипептиды "выделены" при участии человека в том смысле, что они существуют отдельно от их нативной среды обитания и, следовательно, не являются продуктами природы, однако в некоторых вариантах осуществления они могут быть введены и существовать в рекомбинантной клетке-хозяине.

В некоторых вариантах осуществления гетерологичные или рекомбинантные конструкции нуклеиновых кислот по изобретению являются "синтетическими". "Синтетическая" молекула нуклеиновой кислоты, "синтетическая" нуклеотидная последовательность или "синтетический" полипептид представляет собой молекулу нуклеиновой кислоты, нуклеотидную последовательность или полипептид, который не встречается в природе, но создается руками человека и, следовательно, не является продуктом природы.

В любом из раскрытых в настоящем описании вариантов осуществления нуклеотидные последовательности и/или конструкции гетерологичных нуклеиновых кислот по изобретению могут быть функционально связаны с различными промоторами и другими регуляторными элементами для экспрессии в различных клетках организмов. Таким образом, в типичных вариантах осуществления конструкция нуклеиновой кислоты по настоящему изобретению может дополнительно содержать один или более промоторов, функционально связанных с одной или более нуклеотидными последовательностями.

Под "функционально связанным" или "функционально ассоциированным", как используется в настоящем описании, подразумевается, что указанные элементы функционально связаны друг с другом, а также обычно связаны физически. Таким образом, термин "функционально связанный" или "функционально ассоциированный", как используется в настоящем описании, относится к нуклеотидным последо-

вательностям в одной молекуле нуклеиновой кислоты, которые функционально связаны. Таким образом, первая нуклеотидная последовательность, которая функционально связана со второй нуклеотидной последовательностью, означает ситуацию, когда первая нуклеотидная последовательность помещена в функциональные отношения со второй нуклеотидной последовательностью. Например, промотор функционально связан с нуклеотидной последовательностью, если промотор воздействует на транскрипцию или экспрессию указанной нуклеотидной последовательности. Специалистам в данной области техники будет понятно, что регуляторные последовательности (например, промотор) не обязательно должны быть смежными с нуклеотидной последовательностью, с которой они функционально связаны, при условии, что эти регуляторные последовательности функционируют, чтобы управлять ее экспрессией. Так, например, между промотором и нуклеотидной последовательностью могут находиться промежуточные нетранслируемые, но транскрибируемые последовательности, и промотор при этом можно считать "функционально связанным" с нуклеотидной последовательностью.

"Промотор" представляет собой нуклеотидную последовательность, которая контролирует или регулирует транскрипцию нуклеотидной последовательности (т.е. кодирующей последовательности), которая функционально связана с промотором. Кодирующая последовательность может кодировать полипептид и/или функциональную РНК. Как правило, "промотор" относится к нуклеотидной последовательности, которая содержит сайт связывания для РНК-полимеразы II и управляет инициацией транскрипции. Как правило, промоторы находятся на 5'-конце или выше относительно начала кодирующей области соответствующей кодирующей последовательности. Область промотора может содержать другие элементы, которые функционируют как регуляторы экспрессии генов. Они включают консенсусную последовательность ТАТА-бокс и часто консенсусную последовательность СААТ-бокс (Breathnach and Chambon, (1981), *Annu. Rev. Biochem.*, 50: 349). В растениях СААТ-бокс может быть заменен АГГА-боксом (Messing et al. (1983), in *Genetic Engineering of Plants*, T. Kosuge, C. Meredith and A. Hollaender (eds.), Plenum Press, p. 211-227).

Промоторы могут включать, например, конститутивные, индуцируемые, временно регулируемые, регулируемые в процессе развития, химически регулируемые, тканепредпочтительные и/или тканеспецифические промоторы, которые используются при создании конструкций гетерологичных нуклеиновых кислот, т.е. "химерных генов" или "химерных полинуклеотидов". Эти различные типы промоторов известны в данной области техники.

Выбор промотора будет зависеть от временных и пространственных требований к экспрессии, а также от клетки-хозяина, которую нужно трансформировать. Промоторы для многих разных организмов хорошо известны в данной области. Основываясь на обширных знаниях, представленных в данной области, для конкретного представляющего интерес организма-хозяина можно выбрать подходящий промотор. Так, например, большое количество сведений имеется о промоторах, расположенных выше в высокой степени конститутивно экспрессируемых генов в модельных организмах, и такие сведения являются легко доступными и при необходимости могут использоваться применительно к другим системам.

В некоторых вариантах осуществления конструкция нуклеиновой кислоты по изобретению может представлять собой "экспрессионную кассету" или может быть включена в экспрессионную кассету. Используемый в настоящем описании термин "экспрессионная кассета" означает конструкцию гетерологичной нуклеиновой кислоты, содержащую представляющую интерес нуклеотидную последовательность (например, конструкции нуклеиновых кислот по изобретению (например, конструкцию синтетической *tracr* нуклеиновой кислоты, конструкцию синтетической нуклеиновой кислоты CRISPR, синтетический массив CRISPR, конструкцию химерной нуклеиновой кислоты, нуклеотидную последовательность, кодирующую представляющий интерес полипептид, полипептид типа I, полипептид типа II, полипептид типа III, полипептид типа IV и/или полипептид типа V)), где указанная нуклеотидная последовательность функционально связана с по меньшей мере, регуляторной последовательностью (например, промотором). Таким образом, некоторые аспекты изобретения обеспечивают экспрессионные кассеты, предназначенные для экспрессии последовательностей нуклеотидов по изобретению.

Экспрессионная кассета, содержащая представляющую интерес нуклеотидную последовательность, может быть химерной, что означает, что по меньшей мере один из ее компонентов является гетерологичным по отношению по меньшей мере к одному из других ее компонентов. Экспрессионная кассета также может быть такой, которая встречается в природе, но была получена в рекомбинантной форме, пригодной для гетерологичной экспрессии.

Экспрессионная кассета также может необязательно включать область транскрипции и/или терминирования трансляции (т.е. область терминации), которая является функциональной в выбранной клетке-хозяине. Различные терминаторы транскрипции доступны для использования в экспрессионных кассетах и отвечают за терминацию транскрипции за пределами представляющей интерес гетерологичной нуклеотидной последовательности и правильное полиаденилирование мРНК. Область терминации может быть нативной для области инициации транскрипции, может быть нативной для представляющей интерес функционально связанной нуклеотидной последовательности, может быть нативной для клетки-хозяина или может быть получена из другого источника (т.е. чужеродной или гетерологичной по отношению к промотору, представляющей интерес нуклеотидной последовательности, хозяину или любой их комби-

нации).

Экспрессионная кассета также может включать нуклеотидную последовательность для селективного маркера, который может быть использован для отбора трансформированной клетки-хозяина. Используемый в настоящем описании термин "селективный маркер" означает нуклеотидную последовательность, которая при экспрессии придает экспрессирующий маркер клетке-хозяину конкретный фенотип и, таким образом, позволяет отличать такие трансформированные клетки от тех, которые не имеют маркера. Такая нуклеотидная последовательность может кодировать либо селективный, либо экранируемый маркер в зависимости от того, представляет ли маркер признак, который может быть выбран химическими средствами, например, при помощи селективного агента (например, антибиотика и т.п.), или этот маркер является просто признаком, который можно идентифицировать посредством наблюдения или тестирования, например, путем скрининга (например, флуоресценции). Конечно, многие примеры подходящих селективных маркеров известны в данной области и могут быть использованы в приведенных в настоящем описании экспрессионных кассетах.

В дополнение к экспрессионным кассетам молекулы нуклеиновой кислоты и нуклеотидные последовательности, раскрытые в настоящем описании, могут быть использованы в отношении векторов. Термин "вектор" относится к композиции для переноса, доставки или введения нуклеиновой кислоты (или нуклеиновых кислот) в клетку. Вектор содержит молекулу нуклеиновой кислоты, содержащую нуклеотидную последовательность (последовательности), которая должна быть передана, доставлена или введена. Векторы, предназначенные для использования при трансформации организмов-хозяев, хорошо известны в данной области. Неограничивающие примеры общих классов векторов включают без ограничения вирусный вектор, плазмидный вектор, фаговый вектор, фagemидный вектор, космидный вектор, фосмидный вектор, бактериофаг, искусственную хромосому или бинарный вектор *Agrobacterium* в виде двухнитевой или одонитевой линейной или кольцевой формы, которая может быть или может не быть самостоятельно трансмиссивной или мобилизуемой. Вектор, определенный в настоящем описании, может трансформировать прокариотического или эукариотического хозяина либо путем интеграции в клеточный геном, либо путем внехромосомного существования (например, в виде автономно реплицируемой плазмиды с точкой начала репликации). Дополнительно включаются челночные векторы, под которыми подразумевается носитель ДНК, способный естественным образом или благодаря созданной структуре к репликации в двух разных организмах-хозяевах, которые могут быть выбраны из актиномицетов и родственных видов, бактерий и эукариот (например, клеток высших растений, млекопитающих, дрожжей или грибов). В некоторых типичных вариантах осуществления нуклеиновая кислота в векторе находится под контролем и функционально связана с соответствующим промотором или другими регуляторными элементами для транскрипции в клетке-хозяине. Вектор может быть бифункциональным экспрессионным вектором, который функционирует в нескольких хозяевах. В случае геномной ДНК он может содержать свой собственный промотор или другие регуляторные элементы, а в случае кДНК он может находиться под контролем соответствующего промотора или других регуляторных элементов для экспрессии в клетке-хозяине. Соответственно молекулы нуклеиновой кислоты по настоящему изобретению и/или экспрессионные кассеты могут быть включены в векторы, как описано здесь и как известно в данной области.

"Введение", "вводить", "введенный" (и их грамматические варианты) в контексте представляющего интерес полинуклеотида означают презентацию представляющего интерес полинуклеотида в организме-хозяине или клетке указанного организма (например, клетке-хозяине) таким образом, что полинуклеотид получает доступ к внутренней части клетки. При введении более одного полинуклеотида эти полинуклеотиды могут быть собраны в виде части одной полинуклеотидной конструкции или конструкции нуклеиновой кислоты, или в виде отдельных полинуклеотидных конструкций или конструкций нуклеиновых кислот и могут быть расположены на одних и тех же или разных экспрессионных конструкциях или векторах для трансформации. Соответственно эти полинуклеотиды могут быть введены в клетки одним событием трансформации, отдельными событиями трансформации/трансфекции или, например, они могут быть включены в организм с помощью обычных протоколов размножения. Таким образом, в некоторых аспектах одна или более конструкций нуклеиновых кислот по настоящему изобретению могут быть введены в организм-хозяин или клетку указанного организма-хозяина по отдельности или в комбинации. В контексте популяции клеток "введение" означает контактирование популяции с конструкциями гетерологичных нуклеиновых кислот по изобретению в условиях, когда конструкции гетерологичных нуклеиновых кислот по изобретению получают доступ к внутренней части одной или более клеток популяции, тем самым трансформируя эту одну или более клеток популяции.

Используемый в настоящем описании термин "трансформация" или "трансфекция" относится к введению в клетку гетерологичной нуклеиновой кислоты. Трансформация клетки может быть стабильной или временной. Таким образом, в некоторых вариантах осуществления клетку-хозяин или организм-хозяин трансформируют конструкцией нуклеиновой кислоты по изобретению стабильно. В других вариантах осуществления клетку-хозяин или организм-хозяин трансформируют конструкцией нуклеиновой кислоты по изобретению временно. Таким образом, в типичных вариантах осуществления конструкция гетерологичной нуклеиновой кислоты по изобретению может вводиться в клетку стабильно и/или вре-

менно.

"Временная трансформация" в контексте полинуклеотида означает, что полинуклеотид вводится в клетку и не интегрируется в геном клетки.

Под "стабильным введением" или "стабильно введенным" в контексте полинуклеотида следует понимать, что введенный полинуклеотид стабильно включается в геном клетки и, таким образом, клетка стабильно трансформируется полинуклеотидом.

"Стабильная трансформация" или "стабильно трансформированный", как используется в настоящем описании, означает, что конструкция нуклеиновой кислоты вводится в клетку и интегрируется в геном клетки. Таким образом, интегрированная конструкция нуклеиновой кислоты способна быть унаследованной потомством, в частности потомством нескольких последовательных поколений. "Геном", используемый в настоящем описании, включает ядерный, митохондриальный и плазмидный геном и, следовательно, может включать интеграцию конструкции нуклеиновой кислоты, например, в плазмиду или митохондриальный геном. Стабильная трансформация, используемая в настоящем описании, может также относиться к трансгену, который сохраняется внехромосомно, например, в виде минихромосомы или плазмиды.

Временная трансформация может быть обнаружена, например, при помощи иммуноферментного анализа (ELISA) или Вестерн-блот анализа, который позволяет обнаруживать наличие пептида или полипептида, кодируемого одним или более трансгенами, введенными в организм. Стабильная трансформация клетки может быть обнаружена, например, методом Саузерн-блот-гибридизации геномной ДНК клетки с последовательностями нуклеиновых кислот, которые специфически гибридизуются с нуклеотидной последовательностью вводимого в организм трансгена (например, бактерии, археи, дрожжей, водоросли и т.п.). Стабильная трансформация клетки может быть обнаружена, например, при помощи анализа нозерн-блот гибридизации РНК клетки с последовательностями нуклеиновых кислот, которые специфически гибридизуются с нуклеотидной последовательностью трансгена, введенного в растение или другой организм. Стабильная трансформация клетки также может быть обнаружена, например, при помощи полимеразной цепной реакции (ПЦР) или других реакций амплификации, хорошо известных в данной области, с использованием специфических праймерных последовательностей, которые гибридизуются с последовательностью (последовательностями) - мишенью трансгена, что приводит к амплификации последовательности трансгена, которая может быть обнаружена стандартными методами. Трансформация также может быть обнаружена при помощи протоколов прямого секвенирования и/или гибридизации, хорошо известных в данной области.

Соответственно в некоторых вариантах осуществления нуклеотидные последовательности, конструкции, экспрессионные кассеты могут экспрессироваться временно и/или могут быть стабильно включены в геном организма-хозяина.

Конструкцию гетерологичной нуклеиновой кислоты по изобретению можно вводить в клетку любым способом, известным специалистам в данной области. В некоторых вариантах осуществления изобретения трансформация клетки включает трансформацию ядра. В других вариантах осуществления конструкция (конструкции) гетерологичной нуклеиновой кислоты по изобретению может быть введена в клетку обычными методами размножения.

Способы трансформации как эукариотических, так и прокариотических организмов хорошо известны и являются обычными в данной области и описаны в любом литературном источнике (см., например, Jiang et al., 2013, Nat. Biotechnol., 31:233-239; Ran et al., Nature Protocols, 8:2281-2308 (2013)).

Таким образом, нуклеотидная последовательность может быть введена в организм-хозяин или его клетку любым количеством способов, которые хорошо известны в данной области. Способы по изобретению не зависят от конкретного способа введения одной или более нуклеотидных последовательностей в организм за исключением, что они получают доступ к внутренней части по меньшей мере одной клетки организма. При введении более одной нуклеотидной последовательности они могут быть собраны в виде части одной конструкции нуклеиновой кислоты или в виде отдельных конструкций нуклеиновых кислот и могут быть расположены на одних и тех же или разных конструкциях нуклеиновых кислот. Соответственно нуклеотидные последовательности могут быть введены в представляющую интерес клетку в одном событии трансформации или в отдельных событиях трансформации или в случае необходимости нуклеотидная последовательность может быть включена в организм в рамках протокола размножения.

Мобильные генетические элементы (MGE) доставляют бактериям постоянные проблемы, связанные с геномной стабильностью, способствуя эволюции через горизонтальный перенос генов. Термин MGE охватывает плазмиды, бактериофаги, транспонируемые элементы, геномные острова и многие другие специализированные генетические элементы (1). MGE охватывают гены, обеспечивающие хозяину высокую скорость распространения, адаптивные преимущества и геномную стабильность, что приводит к их практически универсальному присутствию в бактериальных геномах. Чтобы справиться с постоянной угрозой хищных бактериофагов и эгоистичных генетических элементов, бактерии развили как врожденную, так и адаптивную иммунную систему, ориентированную на экзогенные генетические элементы. Врожденный иммунитет включает модификацию клеточной стенки, системы рестрикции/модификации и abortивную фаговую инфекцию (2). Короткие палиндромные повторы, регулярно расположенные груп-

пами (CRISPR) и CRISPR-ассоциированные гены (Cas) являются адаптивной иммунной системой, направленной против инвазивных генетических элементов у бактерий (3). CRISPR-Cas-опосредованный иммунитет основан на разных молекулярных процессах, классифицированных как приобретение, экспрессия и интерференция (3). Приобретение осуществляется через молекулярный "отбор образца (сэмплирование)" чужеродных генетических элементов, из которых короткие последовательности, называемые спейсерами, интегрируются в массив CRISPR в определенной ориентации (4). Экспрессия массивов CRISPR является конститутивной и индуцируемой промоторными элементами в предшествующей лидерной последовательности (5-6). Интерференция возникает из соответствующего транскрипта, который подвергается избирательному процессингу в каждой повторяющейся последовательности, формируя РНК CRISPR (сгРНК), которые направляют белки Cas для распознавания последовательностей и расщепления ДНК-мишени, комплементарной спейсеру (7). Технология CRISPR-Cas применяется для типирования и обнаружения штаммов (8-10), использования естественного/созданного иммунитета против мобильных генетических элементов (11), программируемого редактирования генома в разных условиях (12), контроля транскрипции (13-14) и манипуляции микробными популяциями в определенных консорциумах (15).

В данной области техники известны различные CRISPR системы. Например, см. Макарова и др., где описана номенклатура всех классов, типов и подтипов CRISPR систем (Nature Reviews Microbiology, 13:722-736 (2015)); см. также R. Barrangou (Genome Biol., 16:247 (2015)).

Хотя признаки последовательности, соответствующие массивам CRISPR, были описаны ранее в нескольких организмах (16-17), *Streptococcus thermophilus* был первым микробом, в котором проявились роли специфических генов cas и компонентов массива CRISPR (4). *S. thermophilus* является непатогенной термофильной грамположительной бактерией, используемой в качестве стартовой культуры, которая катаболизирует лактозу в молочную кислоту при синтрофном производстве йогурта и различных сыров (18). *S. thermophilus* кодирует до четырех CRISPR-Cas систем, две из них (SthCRISPR1 и SthCRISPR3) классифицированы как системы типа II-A, которые активны от природы как в случае приобретения, так и интерференции (4, 19). Соответственно геномный анализ *S. thermophilus* и его бактериофагов установил вероятный механизм систем CRISPR-Cas для защиты фага/ДНК. Изучение систем CRISPR-Cas в *S. thermophilus* позволило провести биоинформатический анализ происхождения спейсера (4, 20), открыть последовательности мотива, прилежащего к протоспейсеру (РАМ) (19, 21), понять динамику фаг-хозяин (22-23), продемонстрировать активности эндонуклеазы Cas9 (7, 24-25) и (недавно) определить структурные мотивы tracrRNA, регулирующие функцию и ортогональность систем типа II (26). Геномный анализ *S. thermophilus* выявил эволюционную адаптацию к молоку через потерю углеводного катаболизма и генов вирулентности, обнаруженных у патогенных стрептококках (18). *S. thermophilus* также приобрел существенное количество родственных для этой ниши генов, таких как гены, кодирующие в том числе белки холодового шока, белки устойчивости к меди, протеиназы, бактериоцины и белки катаболизма лактозы (18). Встраиваемые последовательности (IS) широко распространены в геномах *S. thermophilus* и обеспечивают генетическую гетерогенность штаммов, облегчая распространение островов, связанных с генами адаптации к молочным продуктам (18). Сопутствующее присутствие MGE и функциональных CRISPR-Cas систем в *S. thermophilus* предполагает, что гомеостаз генома определяется по меньшей мере частично, взаимодействием этих динамических сил. Таким образом, *S. thermophilus* представляет собой идеального хозяина для исследования генетических результатов нацеливания геномных островов на CRISPR-Cas.

CRISPR-Cas системы недавно стали предметом интенсивных исследований в области применения для редактирования генома (12), но эволюционные роли большинства эндогенных микробных систем остаются неизвестными (27). Еще меньше известно об эволюционных последствиях размещения активных CRISPR-Cas систем, помимо их роли в предотвращении захвата чужеродных ДНК (7), событий приобретения спейсеров (4) и мутации, вызываемой хромосомным самонацеливанием (28-32). Таким образом, авторы изобретения стремились определить результаты таргетинга интегрированных MGE с эндогенными CRISPR-Cas системами типа II. В *S. thermophilus* LMD-9 были идентифицированы четыре острова с длиной от 8 до 102 т.п.о. и общим количеством примерно 132 т.п.о., или 7% генома. Для таргетинга геномных островов плазмиду, используемую для экспрессии сконструированных массивов CRISPR с самонацеленными спейсерами, трансформированы в *S. thermophilus* LMD-9. В совокупности полученные результаты разъясняют результаты фундаментальных генетических исследований событий самонацеливания и показывают, что CRISPR-Cas системы могут направлять эволюцию генома на уровне популяции бактерий.

Используя эти открытия, авторы настоящего изобретения разработали новые способы скрининга популяций бактериальных, архейных, водорослевых или дрожжевых клеток для жизненно важных генов, нежизненно важных генов и/или инородных геномных островов; для уничтожения одной или более клеток в популяции бактериальных, архейных, водорослевых или дрожжевых клеток; для идентификации фенотипа бактериального, архейного, водорослевого или дрожжевого гена; для выбора одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих уменьшенный размер генома из популяции бактериальных, архейных или дрожжевых клеток; и/или для идентификации в популяции

бактериальных, архейных, водорослевых или дрожжевых клеток по меньшей мере одного изолята, имеющего мутацию (например, делецию) в своем геноме.

Таким образом, в одном из аспектов авторы настоящего изобретения разработали методы идентификации генетических вариантов в популяции, которые изменили генетическое содержание, что позволило им избежать таргетирования. В данном случае последовательность-мишень была модифицирована, и был проведен поиск выживших организмов, имеющих эту модификацию. В некоторых аспектах модификация (т.е. мутация) представляет собой делецию. Кроме того, если последовательность-мишень была модифицирована, то генотип дикого типа не является существенным.

Следовательно, в одном из аспектов изобретения предоставляется способ скрининга популяции бактериальных клеток в отношении жизненно важных генов, нежизненно важных генов и/или инородных геномных островов, включающий

введение в указанную популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме клеток бактерий указанной популяции, таким образом создавая популяцию трансформированных бактериальных клеток; и

определение наличия или отсутствия делеции в популяции трансформированных бактериальных клеток, причем наличие делеции в популяции трансформированных бактериальных клеток указывает на то, что эта область-мишень содержится в нежизненно важном гене и/или инородном геномном острове, а отсутствие делеции в популяции бактериальных клеток означает, что эта область-мишень содержится в жизненно важном гене.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В дополнительных аспектах изобретения предоставляется способ скрининга популяции бактериальных, архейных, водорослевых и дрожжевых клеток в отношении жизненно важных генов, нежизненно важных генов и/или инородных геномных островов, включающий

введение в популяцию бактериальных, архейных, водорослевых и дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных, архейных, водорослевых и дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, таким образом создавая популяцию трансформированных бактериальных, архейных, водорослевых и дрожжевых клеток; и

определение наличия или отсутствия делеции в популяции трансформированных бактериальных, архейных, водорослевых и дрожжевых клеток, причем наличие делеции в популяции бактериальных, архейных, водорослевых и дрожжевых клеток означает, что эта область-мишень содержится в нежизненно важном гене и/или инородном геномном острове, а отсутствие делеции в популяции бактериальных, архейных, водорослевых и дрожжевых клеток означает, что эта область-мишень содержится в жизненно важном гене.

В других аспектах предоставляется способ уничтожения одной или более бактериальных клеток в популяции бактериальных клеток, включающий введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных клеток указанной популяции, таким образом уничтожая в популяции одну или более бактериальных клеток, которые содержат область-мишень. Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В дополнительном аспекте предоставляется способ уничтожения одной или более клеток в популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включающий введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последова-

тельность повтор-спейсер, в которой спейсер указанной по меньшей мере одной последовательности повтор-спейсер и последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(с) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, таким образом уничтожая в популяции бактериальных, архейных, водорослевых или дрожжевых клеток.

Трансформация CRISPR РНК для нацеливания на бактериальный геном может быть использована для выборочного уничтожения бактериальных клеток на основе специфичности последовательностей для удаления генетически отличающихся субпопуляций, тем самым обогащая популяции бактерий, не содержащих последовательность-мишень. Такое отличие может возникать на основе гетерогенного распределения ортогональных CRISPR-Cas систем в генетически сходных популяциях. Таким образом, в некоторых вариантах осуществления массив CRISPR, который вводится в популяцию клеток, может оказаться совместимым (т.е. функциональным) с CRISPR-Cas системой в одной или более бактериальных клетках, подлежащих уничтожению, и несовместимым (т.е. нефункциональным) с CRISPR Cas системой по меньшей мере в одной или более бактериальных клетках в популяции. Например, *Escherichia coli* и *Klebsiella pneumoniae* могут иметь CRISPR-Cas системы типа I-E или типа I-F; *Clostridium difficile* кодирует системы типа I-B, а различные штаммы *S. thermophilus* имеют как системы типа II-A, так и типа I-E, или только системы типа II-A. В зависимости от конкретной CRISPR РНК, трансформированной в смесь бактерий, она может специфически нацеливаться на такое подмножество популяции на основе ее функциональной совместимости с родственной ей системой. Это можно применить к различным видам, содержащим эндогенные CRISPR-Cas системы, таким как без ограничения следующие.

Pseudomonas spp. (например: *P. aeruginosa*), *Escherichia* spp. (например: *E. coli*), *Enterobacter* spp. (например: *E. cloacae*), *Staphylococcus* spp. (например: *S. aureus*), *Enterococcus* spp. (например: *E. faecalis*, *E. faecium*), *Streptomyces* spp. (например: *S. somaliensis*), *Streptococcus* spp. (например: *S. pyogenes*), *Vibrio* spp. (например: *V. cholerae*), *Yersinia* spp. (например: *Y. pestis*), *Francisella* spp. (например: *F. tularensis*, *F. novicida*), *Bacillus* spp. (например: *B. anthracis*, *B. cereus*), *Lactobacillus* spp. (например: *L. casei*, *L. reuteri*, *L. acidophilus*, *L. rhamnosus*), *Burkholderia* spp. (например: *B. mallei*, *B. pseudomallei*), *Klebsiella* spp. (например: *K. pneumoniae*), *Shigella* spp. (например: *S. dysenteriae*, *S. sonnei*), *Salmonella* spp. (например: *S. enterica*), *Borrelia* spp. (например: *B. burgdorferi*), *Neisseria* spp. (например: *N. meningitidis*), *Fusobacterium* spp. (например: *F. nucleatum*), *Helicobacter* spp. (например: *H. pylori*), *Chlamydia* spp. (например: *C. trachomatis*), *Bacteroides* spp. (например: *B. fragilis*), *Bartonella* spp. (например: *B. quintana*), *Bordetella* spp. (например: *B. pertussis*), *Brucella* spp. (например: *B. abortus*), *Campylobacter* spp. (например: *C. jejuni*), *Clostridium* spp. (например: *C. difficile*), *Bifidobacterium* spp. (например: *B. infantis*), *Haemophilus* spp. (например: *H. influenzae*), *Listeria* spp. (например: *L. monocytogenes*), *Legionella* spp. (например: *L. pneumophila*), *Mycobacterium* spp. (например: *M. tuberculosis*), *Mycoplasma* spp. (например: *M. pneumoniae*), *Rickettsia* spp. (например: *R. rickettsii*), *Acinetobacter* spp. (например: *A. calcoaceticus*, *A. baumannii*), *Ruminococcus* spp. (например: *R. albus*), *Propionibacterium* spp. (например: *P. freudenreichii*), *Corynebacterium* spp. (например: *C.*

diphtheriae), *Propionibacterium* spp. (например: *P. acnes*), *Brevibacterium* spp. (например: *B. iodinum*), *Micrococcus* spp. (например: *M. luteus*) и/или *Prevotella* spp. (например: *P. histicola*).

Нацеливание CRISPR может привести к удалению специфических подмножеств бактерий на основе различающегося генетического содержания в смешанных популяциях. Подтверждение этого заявления представлено в примерах 4, 5, где Lac^- бактерии отбирают до полного исчезновения в популяции Lac^+ . Генетическое различие между Lac^+ и Lac^- штаммами представлено в примерах 8 и 10, где секвенирование выживших клонов выявило до 5,5% различий в генетическом содержании по сравнению с референсным штаммом *S. thermophilus* дикого типа. Таким образом, спейсеры для нацеливания на CRISPR можно настраивать на разную степень родства между бактериями путем нацеливания на консервативные или дивергентные генетические последовательности. Таким образом, в некоторых вариантах осуществления бактериальные клетки в популяции могут содержать одну и ту же CRISPR-Cas систему, и введенный массив CRISPR, таким образом, может функционировать в популяции бактерий как единое целое, но генетическое содержание других штаммов или видов, которые составляют эту популяцию бактерий, отличается в достаточной степени, для того чтобы область-мишень для введенного массива CRISPR была обнаружена только в одном или более предназначенных для уничтожения видах бактерий в популяции. Это может быть применено к различным видам, содержащим эндогенные CRISPR-Cas системы, такие как без ограничения *Pseudomonas* spp. (например, следующие).

P. aeruginosa), *Escherichia* spp. (например: *E. coli*), *Enterobacter* spp. (например: *E. cloacae*), *Staphylococcus* spp. (например: *S. aureus*), *Enterococcus* spp. (например: *E. faecalis*, *E. faecium*), *Streptomyces* spp. (например: *S. somaliensis*), *Streptococcus* spp. (например: *S. pyogenes*), *Vibrio* spp. (например: *V. cholerae*), *Yersinia* spp. (например: *Y. pestis*), *Francisella* spp. (например: *F. tularensis*, *F. novicida*), *Bacillus* spp. (например: *B. anthracis*, *B. cereus*), *Lactobacillus* spp. (например: *L. casei*, *L. reuteri*, *L. acidophilus*, *L. rhamnosus*), *Burkholderia* spp. (например: *B. mallei*, *B. pseudomallei*), *Klebsiella* spp. (например: *K. pneumoniae*), *Shigella* spp. (например: *S. dysenteriae*, *S. sonnei*), *Salmonella* spp. (например: *S. enterica*), *Borrelia* spp. (например: *B. burgdorferi*), *Neisseria* spp. (например: *N. meningitidis*), *Fusobacterium* spp. (например: *F. nucleatum*), *Helicobacter* spp. (например: *H. pylori*), *Chlamydia* spp. (например: *C. trachomatis*), *Bacteroides* spp. (например: *B. fragilis*), *Bartonella* spp. (например: *B. quintana*), *Bordetella* spp. (например: *B. pertussis*), *Brucella* spp. (например: *B. abortus*), *Campylobacter* spp. (например: *C. jejuni*), *Clostridium* spp. (например: *C. difficile*), *Bifidobacterium* spp. (например: *B. infantis*), *Haemophilus* spp. (например: *H. influenzae*), *Listeria* spp. (например: *L. monocytogenes*), *Legionella* spp. (например: *L. pneumophila*), *Mycobacterium* spp. (например: *M. tuberculosis*), *Mycoplasma* spp. (например: *M. pneumoniae*), *Rickettsia* spp. (например: *R. rickettsii*), *Acinetobacter* spp. (например: *A. calcoaceticus*, *A. baumannii*), *Ruminococcus* spp. (например: *R. albus*), *Propionibacterium* spp. (например: *P. freudenreichii*), *Corynebacterium* spp. (например: *C. diphtheriae*), *Propionibacterium* spp. (например: *P. acnes*), *Brevibacterium* spp. (например: *B. iodinum*), *Micrococcus* spp. (например: *M. luteus*) и/или *Prevotella* spp. (например: *P. histicola*).

Степень уничтожения в популяции способом по настоящему изобретению может зависеть от восприимчивости конкретной популяции к трансформации дополнительно к тому, где находится область-мишень в нежизненно важном гене, жизненно важном гене или инородном острове. Степень уничтоже-

ния в популяции бактериальных, архейных или дрожжевых клеток может варьировать, например, в зависимости от организма, рода и вида. Соответственно используемый в настоящем описании термин "уничтожение" означает удаление 2 log или более клеток в популяции (1% выживаемости или менее). Уничтожение на уровне менее 1 log будет небольшим сокращением численности популяции; в то время как уничтожение на уровне 2-3 log приведет к значительному сокращению численности популяции; а уничтожение на уровне более 3 log указывает на то, что популяция является по существу уничтоженной.

В другом аспекте предоставляется способ идентификации фенотипа, ассоциированного с бактериальным геном, включающий

введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной по меньшей мере одной последовательности повтор-спейсер и последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных клеток указанной популяции, причем эта область-мишень содержит по меньшей мере часть открытой рамки считывания, кодирующей полипептид, или функциональную нуклеиновую кислоту, таким образом уничтожая клетки, содержащие эту область-мишень, и создавая популяцию трансформированных бактериальных клеток, не содержащих эту область-мишень (т.е. выжившие клетки не содержат область-мишень); и

(i) анализ фенотипа популяции клеток; или

(ii) выращивание отдельных бактериальных колоний из популяции трансформированных бактериальных клеток и анализ фенотипа отдельных колоний.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В другом аспекте предоставляется способ идентификации фенотипа бактериального, архейного, водорослевого или дрожжевого гена, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, таким образом уничтожая бактериальные, архейные, водорослевые или дрожжевые клетки, содержащие область-мишень, и создавая популяцию трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень; и

(i) анализ фенотипа популяции клеток; или

(ii) выращивание отдельных бактериальных, архейных, водорослевых или дрожжевых колоний из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток; и анализ фенотипа отдельных колоний.

В некоторых вариантах осуществления анализ включает ПЦР, оптическое картирование генома, секвенирование генома, рестрикционное картирование и/или рестрикционный анализ для идентификации и характеристики мутации, а также анализ комплементарности и/или фенотипические анализы для анализа фенотипа.

В некоторых вариантах осуществления изобретения определение степени уничтожения или сокращения популяции может включать любой способ определения численности популяции, включая (без ограничения)

(1) посев клеток и подсчет колоний,

(2) определение оптической плотности,

(3) подсчет под микроскопом,

(4) определение наиболее вероятной численности, и/или

(5) восстановление метиленового синего.

В некоторых вариантах осуществления для определения профиля композиции смешанных популяций может быть использовано секвенирование 16S рДНК. Это можно сделать, например, путем очистки ДНК от образца в целом, и проведения либо секвенирования всего генома методом дробовика при помощи высокоэффективных технологий, либо, например, путем ПЦР-амплификации 16S гена и секвенирования продуктов таким же образом. Затем путем компьютерных вычислений эти последовательности можно отнести к определенным бактериальным таксонам. В других вариантах осуществления для количественного определения бактерий также можно использовать количественные методы ПЦР. Такие методы хорошо известны в данной области. Например, праймеры для амплификации для количественной

ПЦР могут быть сконструированы, в частности, из вида, рода или группы организмов, которые имеют эту последовательность. Таким образом, для количественного определения указанного организма или группы организмов можно использовать пороговое значение (ct). В дополнительных вариантах осуществления любая бактериальная активность (фенотип), специфичная для целевой популяции, также может быть использована в качестве показателя для определения истощения популяции.

В других вариантах осуществления предоставляется способ отбора одной или более бактериальных клеток, имеющих уменьшенный размер генома, из популяции бактериальных клеток, включающий введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, причем клетки, содержащие область-мишень, уничтожаются, таким образом осуществляя отбор одной или более бактериальных клеток, не содержащих эту область-мишень и имеющих уменьшенный размер генома, из популяции бактериальных клеток. Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В некоторых вариантах осуществления предоставляется способ отбора одной или более бактериальных клеток, имеющих уменьшенный размер генома, из популяции бактериальных клеток, включающий введение в популяцию бактериальных клеток

(a) (i) одну или более конструкций гетерологичной нуклеиновой кислоты, содержащих нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательным нуклеотидам, присутствующим в геноме указанных бактериальных клеток, или (ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных бактериальных клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (crPHK, crДНК), содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, причем эта область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, при этом клетки, содержащие область-мишень уничтожаются, а клетки не содержащие область-мишень выживают, таким образом осуществляя отбор одной или более бактериальных клеток не содержащих эту область-мишень и имеющих уменьшенный размер генома, из популяции трансгенных бактериальных клеток.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В области техники хорошо известно, что транспозоны могут быть созданы при помощи, например, ПЦР амплификации или синтеза созданной ДНК и могут быть введены любым способом трансформации.

В некоторых вариантах осуществления изобретения предоставляется способ отбора одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих уменьшенный размер генома, из популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включающий введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (crPHK, crДНК), содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер по меньшей мере одной последовательности повтор-спейсер и по меньшей мере одной последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, причем клетки, содержащие область-мишень уничтожаются, таким образом осуществляя отбор одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень и имеющих уменьшенный размер генома, из популяции бактериальных, архейных, водорослевых или дрожжевых клеток.

В других вариантах осуществления предоставляется способ отбора одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих уменьшенный размер генома, из популяции

бактериальных, архейных, водорослевых или дрожжевых клеток, включающий введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) (i) одной или более конструкции гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательным нуклеотидам, присутствующим в геноме указанных клеток бактерий, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных бактериальных, архейных, водорослевых или дрожжевых клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) (i) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую CRISPR (tracr) нуклеиновую кислоту,

(ii) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (crPHK, crДНК), содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной по меньшей мере одной последовательности повтор-спейсер и по меньшей мере одной последовательности повтор-спейсер-повтор содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных, архейных, водорослевых или дрожжевых клеток в указанной популяции, и

(iii) Cas9 полипептида и/или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, причем область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, причем клетки, содержащие область-мишень уничтожаются, а клетки, не содержащие область-мишень выживают, таким образом осуществляя отбор одной или более бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень и имеющих уменьшенный размер генома, из популяции трансгенных бактериальных, архейных, водорослевых или дрожжевых клеток.

В некоторых аспектах уменьшенный размер генома может быть уменьшен по сравнению с контролем. В некоторых аспектах контроль может представлять собой популяцию дикого типа бактериальных, архейных, водорослевых или дрожжевых клеток или популяцию бактериальных, архейных, водорослевых или дрожжевых клеток, трансформированных гетерологичной конструкцией, содержащей массив CRISPR (например, CRISPR типа I, типа II, типа III, типа IV или типа V), содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая не является комплементарной области-мишени в геноме бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции (т.е. несамонацеленный/"скремблированный спейсер"). В другом аспекте контроль может представлять собой популяцию бактериальных, архейных, водорослевых или дрожжевых клеток, трансформированных конструкцией гетерологичной нуклеиновой кислоты, содержащей массив CRISPR, содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, но не содержит мотив, прилегающий к протоспейсеру (PAM).

В некоторых вариантах осуществления предоставляется способ идентификации в популяции бактерий по меньшей мере одного изолята, содержащего делецию в своем геноме (например, хромосомную и/или плазмидную делецию), включающий

введение в указанную популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (crPHK, crДНК), содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток указанной популяции, и клетки, содержащие область-мишень, уничтожаются, таким образом создавая популяцию трансформированных бактериальных клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных колоний из популяции трансформированных бактериальных клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции трансформированных бактерий, имеющих делецию в своем геноме.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В других вариантах осуществления изобретения предоставляется способ идентификации в популяции бактерий по меньшей мере одного изолята, содержащего делецию в своем геноме, включающий введение в популяцию бактериальных клеток

(a) (i) одной или более конструкции гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности по меньшей мере 300 последовательными нуклеотидами, присутствующим в геноме указанных бактериальных клеток, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных бактериальных клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (сгРНК, сгДНК), содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, причем эта область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, и клетки, содержащие область-мишень уничтожаются [а клетки, не содержащие область-мишень, выживают], таким образом создавая популяцию трансформированных бактериальных клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных колоний из популяции трансформированных бактериальных клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции бактерий, имеющих делецию в своем геноме.

Массив CRISPR, используемый в этом изобретении, может быть массивом CRISPR типа I, типа II, типа III, типа IV или типа V.

В других вариантах осуществления изобретения предоставляется способ идентификации в популяции бактерий, архей, водорослей или дрожжей по меньшей мере одного изолята, имеющего делецию в своем геноме, включающий

введение в популяцию бактериальных, архейных, водорослевых или дрожжевых клеток

(a) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую CRISPR (tracr) нуклеиновую кислоту,

(b) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (сгРНК, сгДНК), содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме (например, хромосомном, митохондриальном и/или плазмидном геноме) бактериальных, архейных, водорослевых или дрожжевых клеток указанной популяции, и

(c) Cas9 полипептида, причем клетки, содержащие область-мишень уничтожаются, таким образом создавая популяцию трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных, архейных, водорослевых или дрожжевых колоний из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, имеющих делецию в своем геноме.

В других вариантах осуществления изобретения предоставляется способ идентификации в популяции бактерий, архей, водорослей или дрожжей по меньшей мере одного изолята, имеющего делецию в своем геноме, включающий введение в популяцию бактериальных, архейных, водорослевых и дрожжевых клеток

(a) (i) одной или более конструкции гетерологичной нуклеиновой кислоты, содержащей нуклеотидную последовательность, имеющую по меньшей мере 80% идентичности с по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме указанных бактериальных, архейных, водорослевых и дрожжевых клеток, или

(ii) двух или более конструкций гетерологичной нуклеиновой кислоты, содержащих по меньшей мере один транспозон, таким образом создавая популяцию трансгенных бактериальных, архейных, водорослевых и дрожжевых клеток, содержащих неприродный сайт для гомологичной рекомбинации между одной или более конструкциями гетерологичной нуклеиновой кислоты, интегрированными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, или между первым и вторым транспозонами, интегрированными в геном; и

(b) (i) конструкции гетерологичной нуклеиновой кислоты, содержащей транс-активирующую

CRISPR (tracr) нуклеиновую кислоту,

(ii) конструкции гетерологичной нуклеиновой кислоты, содержащей массив CRISPR (crPHK, crДНК), содержащий последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, в которой спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных, архейных, водорослевых или дрожжевых клеток в указанной популяции, и

(iii) Cas9 полипептида и/или конструкции гетерологичной нуклеиновой кислоты, содержащей полинуклеотид, кодирующий Cas9 полипептид, причем область-мишень расположена между одной или более конструкциями гетерологичной нуклеиновой кислоты, введенными в геном, и по меньшей мере 300 последовательными нуклеотидами, присутствующими в геноме, и/или между первым транспозоном и вторым транспозоном, причем клетки, содержащие область-мишень уничтожаются, а клетки, не содержащие область-мишень выживают, таким образом создавая популяцию трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, не содержащих область-мишень; и

выращивание отдельных бактериальных, архейных, водорослевых или дрожжевых колоний из популяции трансформированных бактериальных, архейных, водорослевых или дрожжевых клеток, таким образом осуществляя идентификацию по меньшей мере одного изолята из популяции, имеющей делецию в своем геноме.

В некоторых вариантах осуществления приспособляемость/скорость роста может быть увеличена за счет уменьшения размера генома или путем удаления выбранных генов (кодирующих полипептиды или функциональные нуклеиновые кислоты (например, регуляторы транскрипции)), которые требуют большого количества энергии для транскрипции и трансляции. Таким образом, в некоторых вариантах осуществления предлагается способ увеличения приспособляемости или скорости роста популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включающий выбор уменьшенного размера генома (например, выбор по отсутствию части генома) и/или делеции в геномах бактериальных, архейных, водорослевых или дрожжевых клеток популяций, как раскрыто в настоящем описании. В некоторых вариантах осуществления делеция может содержать один ген или более одного гена. Следовательно, в результате уменьшения размера генома или удаления конкретного гена или генов клетки популяции больше не будут тратить энергию на транскрипцию/трансляцию части отсутствующего генома или удаленного гена или генов; тем самым уменьшается потребность в энергии и улучшается приспособляемость по сравнению с контрольной популяцией, все еще содержащей указанную часть генома и/или указанный ген или гены.

В других вариантах осуществления предлагается способ увеличения количества продукта, полученного из популяции бактериальных, архейных, водорослевых или дрожжевых клеток, включая увеличение приспособляемости или скорости роста клетки путем отбора в отношении делеции в геномах бактериальных, архейных, водорослевых или дрожжевых клеток, как раскрыто в настоящем описании. В некоторых вариантах осуществления продукты могут включать без ограничения антибиотики, вторичные метаболиты, витамины, белки, ферменты, кислоты и фармацевтические препараты.

В некоторых вариантах осуществления массив CRISPR (crPHK, crДНК), используемый в этом изобретении, может представлять собой массив любой из CRISPR-Cas системы типа I, CRISPR-Cas системы типа II, CRISPR-Cas системы типа III, CRISPR-Cas системы типа IV или CRISPR-Cas системы типа V.

Что касается предшествующих вариантов осуществления, то конструкция гетерологичной нуклеиновой кислоты, содержащая tracr нуклеиновую кислоту, и конструкция гетерологичной нуклеиновой кислоты, содержащая массив CRISPR, может быть включена и введена в одну и ту же конструкцию (например, экспрессионную кассету или вектор) или в разные конструкции. В конкретных вариантах осуществления конструкция гетерологичной нуклеиновой кислоты, содержащая tracr нуклеиновую кислоту, и конструкция гетерологичной нуклеиновой кислоты, содержащая массив CRISPR, могут быть включены в одну конструкцию (например, экспрессионную кассету и/или вектор), которая может необязательно дополнительно содержать полинуклеотид, кодирующий полипептид Cas9. В некоторых вариантах осуществления конструкция гетерологичной нуклеиновой кислоты, содержащая tracr нуклеиновую кислоту, и конструкция гетерологичной нуклеиновой кислоты, содержащая массив CRISPR, могут быть функционально связаны с одним промотором и/или отдельными промоторами.

В некоторых вариантах осуществления конструкция гетерологичной нуклеиновой кислоты, содержащая транс-активирующую CRISPR (tracr) нуклеиновую кислоту, и конструкция гетерологичной нуклеиновой кислоты, содержащая массив CRISPR (crPHK, crДНК), могут быть включены в гид CRISPR (gPHK, gДНК). В некоторых вариантах осуществления гид CRISPR может быть функционально связан с промотором.

В некоторых вариантах осуществления полипептид Cas9, используемый в этом изобретении, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью нуклеазы Cas9. Примерами нуклеаз Cas9, применимыми в этом изобретении, могут быть любые нуклеазы Cas9, в отношении которых известно, что они катализируют расщепление ДНК в CRISPR-Cas системе. В данной области известно, что такие нуклеазы Cas9 содержат мотив HNH и мотив RuvC

(см., например, WO 0201/176772, WO 2013/188638). В некоторых вариантах осуществления настоящего изобретения может быть использован функциональный фрагмент нуклеазы Cas9.

CRISPR-Cas системы и группировки нуклеаз Cas9 хорошо известны в данной области и включают, например, *Streptococcus thermophilus* CRISPR1 (Sth CR1) группу нуклеаз Cas9, *Streptococcus thermophilus* CRISPR3 (Sth CR3) группу нуклеаз Cas9, *Lactobacillus buchneri* CD034 (Lb) группу нуклеаз Cas9 и *Lactobacillus rhamnosus* GG (Lrh) группу нуклеаз Cas9. Дополнительные нуклеазы Cas9 включают без ограничения *Lactobacillus curvatus* CRL 705. Кроме того, нуклеазы Cas9, используемые в этом изобретении, включают без ограничения Cas9 из *Lactobacillus animalis* KCTC 3501 и *Lactobacillus farciminis*, WP 010018949.1.

Помимо этого, в конкретных вариантах осуществления нуклеаза Cas9 может кодироваться нуклеотидной последовательностью, которая является оптимизированной по кодонам в отношении организма, содержащего ДНК-мишень. В других вариантах осуществления нуклеаза Cas9 может содержать по меньшей мере одну последовательность ядерной локализации.

В некоторых вариантах осуществления полипептид типа I, используемый в настоящем изобретении, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью Cas3, Cas3' нуклеазы, Cas3" нуклеазы, их рекомбинантными вариантами. В некоторых вариантах осуществления каскадный полипептид типа I, используемый в настоящем изобретении, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью Cas7 (Csa2), Cas8a1 (Csx13), Cas8a2 (Csx9), Cas5, Csa5, Cas6a, Cas6b, Cas8b (Csh1), Cas7 (Csh2), Cas5, Cas5d, Cas8c (Csd11), Cas7 (Csd2), Cas10d (Csc3), Csc2, Csc1, Cas6d, Cse1 (CasA), Cse2 (CasB), Cas7 (CasC), Cas5 (CasD), Cas6e (CasE), Cys1, Cys2, Cas7 (Cys3), Cas6f (Csy4), Cas6 и/или Cas4.

CRISPR-Cas системы типа I хорошо известны в данной области техники и включают, например, *Archaeoglobus fulgidus*, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-A, *Clostridium kluyveri* DSM 555, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-B, *Bacillus halodurans* C-125, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-C, *Cyanothece* sp. PCC 802, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-D, *Escherichia coli* K-12, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-E, *Geobacter sulfurreducens*, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-U, и *Yersinia pseudotuberculosis* YPIII, содержащий приводимую в качестве примера CRISPR-Cas систему типа I-F.

В некоторых вариантах осуществления полипептид типа II, пригодный для данного изобретения, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью Cas9. CRISPR-Cas системы типа II хорошо известны в данной области техники и включают, например, штаммы *Legionella pneumophila* Paris, *Streptococcus thermophilus* CNRZ1066 и *Neisseria lactamica* 020-06.

В некоторых вариантах осуществления полипептид типа III, используемый в этом изобретении, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью Cas6, Cas10 (или Csm1), Csm2, Csm3, Csm4, Csm5, и Csm6, Cmr1, Cas10 (или Cmr2), Cmr3, Cmr4, Cmr5, и Cmr6, Cas7, Cas10, Cas7 (Csm3), Cas5 (Csm4), Cas7 (Csm5), Csm6, Cas7 (Cmr1), Cas5 (Cmr3), Cas7 (Cmr4), Cas7 (Cmr6), Cas7 (Cmr6), Cmr5, Cas5 (Cmr3), Cas5 (Csm10), Csm2, Cas7 (Csm3) и all1473. CRISPR-Cas системы типа III хорошо известны в данной области и включают, например, *Staphylococcus epidermidis* RP62A, содержащий приводимую в качестве примера CRISPR-Cas систему типа III-A, *Rugosoccus furiosus* DSM 3638, содержащий приводимую в качестве примера CRISPR-Cas систему типа III-B, штамм *Metanothermobacter thermautotrophicus* Delta H, содержащий приводимую в качестве примера CRISPR-Cas систему типа III-C, и *Roseiflexis* sp. Rs-1, содержащий приводимую в качестве примера CRISPR-Cas систему типа III-D.

В некоторых вариантах осуществления полипептид типа IV, используемый в настоящем изобретении, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью Csf4 (dinG), Csf1, Cas7 (Csf2) и/или Cas5 (csf3). CRISPR-Cas системы типа IV хорошо известны в данной области, например, *Acidithiobacillus ferrooxidans* ATCC 23270 содержит приводимую в качестве примера CRISPR-Cas систему типа IV.

В некоторых вариантах осуществления полипептид типа V, пригодный для данного изобретения, имеет по меньшей мере 70% идентичности (например, примерно 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% и т.п.) с аминокислотной последовательностью Cpf1, Cas1, Cas2 или Cas4. CRISPR-Cas системы типа V хорошо известны в данной области и включают, например, *Francisella* cf. *novicida* Fx1, содержащий приводимую в качестве примера CRISPR-Cas систему типа V.

В настоящем описании также представлены экспрессионные кассеты и векторы, содержащие кон-

струкции нуклеиновых кислот, массивы нуклеиновых кислот, молекулы нуклеиновых кислот и/или нуклеотидные последовательности по настоящему изобретению, которые могут быть использованы со способами, раскрытыми в настоящем описании.

В других аспектах конструкции нуклеиновых кислот, массивы нуклеиновых кислот, молекулы нуклеиновых кислот и/или нуклеотидные последовательности по настоящему изобретению могут быть введены в клетку организма-хозяина. Можно использовать любую клетку/организм-хозяин, в отношении которого это изобретение применимо. Примеры организмов-хозяев включают бактерии, археи, водорослей и грибы (например, дрожжи).

Далее изобретение описано со ссылкой на приведенные ниже примеры. Следует иметь в виду, что эти примеры не предназначены для ограничения объема формулы изобретения, а скорее предназначены для иллюстрации некоторых вариантов осуществления. Любые вариации в приведенных в качестве примера способах, которые возникают у специалиста в данной области, входят в объем настоящего изобретения.

Примеры

Пример 1. Бактериальные штаммы.

Все бактериальные штаммы перечислены в табл. 1.

Бактериальные культуры были криоконсервированы в подходящей среде для выращивания с 25% глицерином (об./об.) и хранились при -80°C . *S. thermophilus* размножали в среде Элликера (Elliker) (Difco), дополненной 1% говяжьим экстрактом (мас./об.) и 1,9% (мас./об.) β -глицерофосфата (Sigma), в статических аэробных условиях при 37°C или на твердой среде с 1,5% (мас./об.) агаром (Difco), инкубировали в анаэробных условиях при 37°C в течение 48 ч. В случае необходимости для отбора плазмиды в *S. thermophilus* использовали концентрации 2 мкг/мл эритромицина (Em) и 5 мкг/мл хлорамфеникола (Cm) (Sigma). *E. coli* EC1000 размножали в аэробных условиях в бульоне Luria-Bertani (Difco) при 37°C или на твердой среде с сердечно-мозговым экстрактом (BHI) (Difco), дополненной 1,5% агаром. При необходимости отбор *E. coli* с использованием антибиотика выполняли при помощи 40 мкг/мл канамицина (K_n) и 150 мкг/мл Em для рекомбинантной *E. coli*. Оценку скрининга производных *S. thermophilus* в отношении активности β -галактозидазы выполняли на качественном уровне путем дополнения синтетической среды Элликер 1% лактозой, 1,5% агаром и 0,04% бромокрезолом в качестве индикатора pH.

Пример 2. Выделение и клонирование ДНК.

Все наборы, ферменты и реагенты использовались в соответствии с инструкциями производителей. Очистку и клонирование ДНК выполняли, как описано ранее (41). Вкратце, для очистки геномной ДНК из *S. thermophilus* использовали набор ZR Fungal/Bacterial MiniPrep kit (Zymo). Плазмидную ДНК выделяли из *E. coli* при помощи набора Qiagen Spin miniprep kit (Qiagen). Высококачественную ПЦР-амплификацию ДНК выполняли с использованием ДНК-полимеразы PFU HS II (Stratagene). Обычные методы ПЦР проводили с использованием полимеразы Choice-Tag Blue (Denville). Праймеры для ПЦР амплификации приобретали у компании Integrated DNA Technologies (Coralville, IA). Экстракцию ДНК из агарозных гелей проводили при помощи набора Zymoclean DNA gel recovery kit (Zymo). Рестрикционные эндонуклеазы приобретали у компании Roche Molecular Biochemicals. Для лигирования использовали лигазу T4 для быстрой сшивки (New England Biolabs). Секвенирование выполняли в Davis Sequencing Inc. (Davis, CA). Криоконсервированные, обработанные хлоридом рубидия компетентные клетки *E. coli* получали как описано ранее (41). Создавали плазмиды с массивами нацеливания на *lacZ*, каждая из которых содержала в следующем порядке:

- (1) нативную лидерную последовательность, специфичную для SthCRISPR1 или SthCRISPR3;
- (2) нативные повторы, специфичные для CRISPR1 или CRISPR3;
- (3) спейсерную последовательность, специфичную для 5'-конца *lacZ* (4) другой нативный повтор (фиг. 1). Для создания каждой плазмиды перечисленные выше признаки последовательности упорядочивали в виде удлиненных олигомеров (табл. 2), объединяли при помощи сплайсинга методом удлинения перекрывающихся ПЦР-продуктов (42) и клонировали в pORI28 (фиг. 2).

Пример 3. Отбор и разработка CRISPR спейсеров.

Программируемая специфичность хромосомного расщепления зависит от выбора требуемой последовательности спейсера, уникальной для аллеля-мишени. Специфичность дополнительно осложняется требуемым PAM, короткой консервативной последовательностью, которая должна находиться в последовательности-мишени в непосредственной близости от протоспейсера (21, 43). Таким образом, строгими критериями для отбора и создания спейсеров были местоположение консенсусных последовательностей PAM и случайная идентичность последовательности с инородными геномными локусами. Первым ограничением для предполагаемых протоспейсеров было местоположение всех предполагаемых последовательностей PAM в смысловой и антисмысловой нитях *lacZ*. В гене длиной 3081 нуклеотидов было найдено 22 сайта PAM для CRISPR1 (AGAAW) и 39 сайтов PAM для CRISPR3 (GGNG), которые оказались идентичными своим консенсусным последовательностям, полученным биоинформационным методом (21). После идентификации потенциальных спейсеров был проведен BLAST анализ полного протоспейсера, затравки и последовательности PAM относительно генома *S. thermophilus* LMD-9 для предот-

вращения дополнительного таргетинга неспецифических локусов. Спейсеры для CRISPR1 и CRISPR3 имели разные последовательности и не соответствовали сайтам PAM, но были разработаны для нацеливания на 5'-конец *lacZ*, в результате чего предсказанные участки расщепления оказались друг от друга на расстоянии 6 нуклеотидов. Следовательно, лидерные последовательности, повторы и спейсеры на каждой плазмиде представляли собой ортогональные признаки, уникальные для CRISPR1 или CRISPR3, соответственно. Для оценки мутаций, зависящих от локусов-мишеней, создавали дополнительную плазмиду CRISPR3 со спейсером в остатке, связывающем катион металла, который имеет важное значение для активности β -галактозидазы. Плазмиду с массивом CRISPR1, содержащую самонацеленный спейсер, использовали в качестве контроля для количественной оценки летальности в результате самонацеливания.

Пример 4. Трансформация.

Плазмиды электропорировали в компетентный *S. thermophilus*, содержащий чувствительную к температуре хелперную плазмиду pTRK669, в соответствии с методами, описанными ранее (44). Вкратце, ночную культуру *S. thermophilus* инокулировали с концентрацией 1% (об./об.) в 50 мл среды Элликер, дополненной 1% говяжьим экстрактом, 1,9% β -глицерофосфатом и C_m для отбора. Когда культура достигла значения OD_{600} нм, равного 0,3, добавляли пенициллин G до получения конечной концентрации 10 мкг/мл для повышения эффективности электропорации (45). Клетки собирали центрифугированием и промывали 3 раза в 10 мл буфера для холодной электропорации (1 М сахарозы и 3,5 мМ $MgCl_2$). Клетки концентрировали в 100 раз в буфере для электропорации, и 40 мкл суспензии аликвотировали в 0,1 мм кюветы для электропорации. Каждую суспензию объединяли с 700 нг плазмиды. Условия электропорации были следующими: 2500 В, емкость - 25 мкФ и сопротивление - 200 Ом. Временные константы регистрировали; интервал варьирования которых составил от 4,4 до 4,6 мс. Суспензии сразу объединяли с 950 мкл предназначенной для извлечения средой и инкубировали в течение 8 ч при 37°C. Клеточные суспензии высевали на селективную среду, а кюветы для электропорации промывали средой для гарантированного извлечения клеток.

Пример 5. Подтверждение фенотипа β -галактозидазы.

Трансформанты, полученные как из CRISPR1, так и CRISPR3, сначала подвергали скринингу на фенотип, дефицитный по β -галактозидазе, путем посева колоний штрихом на полусинтетической среде Элликер, дополненной 1% лактозой в качестве источника твердых углеводов. Потерю активности β -галактозидазы подтверждали выполнением анализов Миллера (о-нитрофенил- β -D-галактозид (ONPG)) (46). Вкратце, культуры размножали до достижения поздней логарифмической фазы (OD_{600} нм=1,2) в 5 мл среды и собирали центрифугированием (4000×g в течение 10 мин). Клетки промывали и ресуспендировали в 0,5 мл забуференного фосфатом физиологическом растворе (Gibco-Invitrogen). Каждую суспензию объединяли со 100 мкл стеклянных микросфер толщиной 0,1 мм (Biospec), а затем подвергали пяти 60 с циклам гомогенизации в Mini-Beadbeater (Biospec). Затем образцы центрифугировали (15000×g в течение 5 мин) для удаления клеточных остатков и интактных клеток. Клеточные лизаты (10 мкл аликвот) объединяли с 600 мкл раствора субстрата (60 мМ Na_2HPO_4 , 40 мМ NaH_2PO_4 , 1 мг/мл ONPG, 2,7 мкл/мл β -меркаптоэтанол) и инкубировали в течение 5 мин при комнатной температуре, после чего добавляли 700 мкл стоп-раствора (1 М $NaCO_3$). Поглощение регистрировали при 420 нм, а активность β -галактозидазы указывали в единицах Миллера, рассчитанных, как описано ранее (46).

Пример 6. Оценка роста и активности.

Культуры предварительно обрабатывали для оценки роста путем субкультивирования в течение 12 поколений в полусинтетической среде Элликера, дефицитной по лактозе. Свежую среду инокулировали ночной культурой с концентрацией 1% (об./об.) и инкубировали при 37°C статически. OD_{600} регистрировали каждый час до тех пор, пока культуры не достигли стационарной фазы. Сбраживание молока оценивали путем инокуляции ночной культуры в обезжиренное молоко до уровня 10^8 КОЕ/мл и инкубирования при 42°C. Затем следили за значением pH при помощи pH-метра Mettler Toledo Seven Easy и зонда Accumet. Обезжиренное молоко приобретали на молочном заводе NCSU Dairy plant и пастеризовали в течение 30 мин при 80°C.

Пример 7. Идентификация инородных геномных областей.

In silico прогноз инородных мобильных локусов для нацеливания на CRISPR-Cas выполняли на основе

- i) местоположения, ориентации и нуклеотидной идентичности IS элементов, и
- ii) местоположения жизненно важных OPC.

В *Bacillus subtilis* количество идентифицированных жизненно важных OPC составило 271, которое было определено по летальности нокаутов генов в геноме дикого типа (33).

Геном *S. thermophilus* исследовали на наличие гомологов по каждому жизненно важному гену в *B. subtilis* при помощи инструмента для поиска в программе BLASTp с матрицей оценки, используемой по умолчанию для аминокислотных последовательностей. В *S. thermophilus* идентифицировали гомологи для примерно 239 жизненно важных OPC, все из которых кодировались хромосомой (табл. 4). Белки,

участвующие в консервативных клеточных процессах, включая ДНК репликацию/гомеостаз, механизмы трансляции и основные метаболические пути, были легко идентифицированы. Не были найдены гомологи, соответствующие биосинтезу цитохрома/дыханию, в соответствии с метаболическим профилем ферментативных бактерий. Каждую предполагаемую жизненно важную ОРС отображали на референсном геноме при помощи программного обеспечения SnapGene, что облегчало визуализацию их местоположения и распределения в *S. thermophilus* LMD-9 (фиг. 3А).

IS-элементы в геноме *S. thermophilus* группировали путем выравнивания кодирующих транспозоны последовательностей при помощи программного обеспечения Geneious® (фиг. 4). Обозначения семейств определяли в соответствии с анализом BLAST в объеме базы данных IS-элементов (www-is.biotoul.fr/). Для предсказания эффективности опосредованной рекомбинацией эксцизии хромосомных сегментов относительное расположение родственных IS-элементов отображали на геном *S. thermophilus* (фиг. 3А). Семейства IS1193 и Sth6 IS-элементов чаще всего встречались в геноме, и обычно их можно найти в *Streptococcus pneumoniae* и *Streptococcus mutans* (34). Несмотря на преобладание элементов IS1193 было показано, что многие из этих локусов представляли собой небольшие фрагменты, демонстрирующие некоторую степень полиморфизма и вырождения, при этом также имелось несколько копий с высоким уровнем идентичности последовательности (фиг. 5А). Напротив, семейство Sth6 показало значимую степень полиморфизма и высокую вырожденность, причем некоторые копии содержали значительные внутренние делеции (фиг. 5В). Элементы IS1167 и IS1191 были менее частыми, но показали почти идеальную точность воспроизведения копий, идентифицированных в геноме (фиг. 5С и 5D). Исходя из степени сохранности длины и последовательности элементов IS1167 и IS1191 у *S. thermophilus* и их относительной близости к генам адаптации молока была выдвинута гипотеза, что эти консервативные/воспроизводимые с высокой точностью транспозоны были приобретены геномом недавно.

Объединив местоположение предсказанных жизненно важных ОРС и IS-элементов, идентифицировали инородные острова, фланкированные IS-элементами, воспроизводимыми с высокой точностью (фиг. 3А; табл. 3). Первый остров содержал оперон, уникальный для *S. thermophilus* LMD-9, кодирующий предполагаемую АТФ-зависимую транспортную систему олигонуклеотидов с неизвестной специфичностью (фиг. 3В) (35). Второй остров содержал протеиназу клеточной стенки PrtS, которая обеспечивает фенотип быстрого сквашивания у *S. thermophilus* (фиг. 3С) (36).

Примечательно, что хотя prtS не является широко распространенным в геномах *S. thermophilus*, было продемонстрировано, что геномные острова, кодирующие prtS, передаются между штаммами посредством естественной компетентности (36). Третий остров содержал предполагаемый белок АТФ-зависимой помпы ионов меди и присутствовал в каждом секвенированном штамме *S. Thermophiles* (фиг. 3D). Четвертый остров был самым крупным по длине, равной 102 т.п.о., и по содержанию генов с 102 предсказанными ОРС, включая lac оперон (фиг. 3Е). Этот остров был обнаружен во всех штаммах *S. thermophilus*, при этом каждый штамм отличался специфичным содержанием генов и длиной. Чтобы определить результат нацеливания обеих эндогенных систем типа II на крупный геномный остров, для последовательности, кодирующей lacZ, создавали массивы повтор-спейсер (фиг. 3Е) и клонировали в pORI28 (фиг. 2). Четвертый остров выбирали в качестве мишени для CRISPR-Cas из-за его размера, широкой распространенности в штаммах *S. thermophilus* и возможности осуществления скрининга на наличие lacZ-мутаций по отрицательному фенотипу β-галактозидазы.

Пример 8. CRISPR-Cas нацеливание lacZ выбирает события с крупными делециями.

В системах типа II Cas9 распознает ДНК и обратимо связывается с последовательностями PAM, при этом происходит активация Cas9 на мишени через образование дуплекса tracrPНК::crPНК (37), что в конечном итоге приводит к расщеплению dsДНК (фиг. 6А и 6В) (25). Фиг. 14А и 14В представляют собой схематическую диаграмму общего принципа совместного взаимодействия эндогенных CRISPR систем для направленного уничтожения. В частности, на фиг. 14А и 14В показан принцип совместной работы эндогенных систем типа II в *Streptococcus thermophilus* для направленного уничтожения. Таким образом, в *S. thermophilus* запрограммированная гибель клеток достигалась за счет использования системы CRISPR-Sth1 (А) или CRISPR-Sth3 (В) типа II путем разработки направленной на геном последовательности спейсера, фланкированной нативными повторами, экспрессия которых управлялась нативным или синтетическим промотором. Процессинг транскрибированного массива повтор-спейсер осуществлялся при помощи кодированной хозяином РНКазы III и Cas9 с получением зрелых crPНК, которые рекрутируют Cas9 в геном, чтобы вызвать двухцепочечное расщепление ДНК, приводящее к гибели клеток.

Трансформация плазмидами, вызывающими хромосомное самонацеливание посредством CRISPR-Cas систем, оказалась цитотоксичной, согласно результатам измерения уменьшения количества выживших трансформантов по сравнению с несамонацеленными плазмидами (15, 29). Таргетирование гена lacZ в *S. thermophilus* привело к 2,5 log уменьшению количества извлеченных трансформантов (фиг. 6С), таким образом достигая предельного значения эффективности трансформации. Двухнитевые разрывы ДНК (DSB) являются существенной угрозой для выживания организмов. Для соответствующих путей репарации часто требуются резекция концов для восстановления ДНК с тупыми концами. Эндонуклеолиз, индуцированный Cas9, еще более усугубляет давление на мутации, вызванные DSB, так как

для восстановления локуса-мишени в диком типе нельзя избежать последующего CRISPR нацеливания. Определение происхождения спейсеров в молочнокислых бактериях показало, что 22% спейсеров обладают комплементарностью к самим себе и что соответствующие геномные локусы были изменены, что, вероятно, способствовало выживанию встречающихся в природе событий самонацеливания (28).

Чтобы определить, мутировал ли локус-мишень в ответ на вызванное Cas9 расщепление, трансформанты были подвергнуты скринингу сначала в отношении потери активности β -галактозидазы. Клоны, дефицитные по этой активности, были генотипированы в локусе *lacZ*. Ни в одном из секвенированных клонов не наблюдалось ни мутаций из-за классического или альтернативного концевое соединения, ни спонтанных однонуклеотидных полиморфизмов. Отсутствие однонуклеотидных полиморфизмов может быть связано с низкой эффективностью трансформации, усугубленной низкой частотой точечных мутаций, а отсутствие гомологов Ku и LigaseIV коррелирует с отсутствием негомологичного концевое соединения (38). ПЦР-скрининг показал, что *lacZ* дикого типа не присутствовал, но ПЦР ампликоны не соответствовали нативному локусу *lacZ*; вероятнее всего произошла амплификация последовательности, фланкированной IS-элементами, в другом геномном локусе. Чтобы исследовать генотип, ответственный за потерю активности β -галактозидазы, было выполнено одномолекулярное секвенирование в реальном времени в двух клонах: один из которых был получен в результате CRISPR3 нацеливания 5'-конца *lacZ* и другой был получен в результате CRISPR3 нацеливания последовательности, кодирующей ион-связывающий карман, необходимый для катализа β -галактозидазы (фиг. 7А и 7В). Эта стратегия секвенирования использовалась из-за большой длины считывания, чтобы обойти трудности, связанные с надежностью отображения прочтений в правильном локусе, из-за большого количества IS-элементов в геноме (35). Прочтения отображали на референсной геномной последовательности при помощи программного обеспечения Geneious, и было выявлено отсутствие крупного сегмента (примерно 102 т.п.о.), кодирующего открытую рамку считывания *lacZ* (фиг. 7А и 7В). Оба секвенированных штамма подтвердили воспроизводимость границ крупной делеции и показали, что делеция происходит независимо от спейсерной последовательности *lacZ* или CRISPR-Cas системы, используемой для нацеливания. Однако данные секвенирования не давали достоверного отображения точных границ делеции.

Удаленные сегменты размером 102 т.п.о. составляют примерно 5,5% от 1,86 м.п.о. генома *S. thermophilus*. Область содержала 102 предполагаемых ОРС (STER_1278-1379), кодирующие ABC-транспортёры, двухкомпонентные регуляторные системы, гены синтеза бактериоцина, связанные с фагом белки, гены катаболизма лактозы и несколько скрытых генов без аннотированной функции (35). Влияние делеции на фенотип роста оценивали в бульонной культуре путем измерения OD_{600} нм в зависимости от времени (фиг. 7С). Клоны с делециями, по-видимому, имели более длинную лаг-фазу, более низкое конечное значение OD ($p < 0,01$) и показали значительно более длительное время генерации во время логарифмической фазы со средним значением 103 мин, по сравнению с 62 мин для дикого типа ($p < 0,001$). Хотя производные с делециями имели меньший по размеру геном для репликации (на 5,5%) в расчете на одно поколение и не тратили ресурсы на транскрипцию или трансляцию исключенных ОРС, никакого явного улучшения в их приспособляемости относительно дикого типа не наблюдалось. Активность β -галактозидазы является отличительной особенностью для промышленного применения молочнокислых бактерий и имеет важное значение для сохранения пищевых систем путем сквашивания. Таким образом, способность *lacZ*-дефицитных штаммов *S. thermophilus* к сквашиванию молока оценивали путем отслеживания значений pH (фиг. 7D). Как и ожидалось, штамм с делециями не был способен к сквашиванию молока в ходе эксперимента, что резко контрастировало с фенотипом быстрого сквашивания, наблюдаемым у дикого типа.

Пример 9. Геномные делеции происходят путем рекомбинации между гомологичными IS-элементами.

Для изучения механизма делеции были определены нуклеотидные последовательности, фланкирующие сегмент. Единственными гомологичными последовательностями, наблюдаемыми на стыках, были две усеченные инсерционные последовательности IS1193 с 91% идентичностью нуклеотидной последовательности по всей длине, равной 727 п.о. Таким образом, для амплификации геномной ДНК выживших клонов, содержащих делецию, была разработана пара праймеров, фланкирующих два IS-элемента. Каждый из штаммов с делецией имел хорошо различимую полосу предсказанного размера (примерно

1,2 т.п.н.), что является подтверждением события крупной геномной делеции (фиг. 8А). Интересно, что слабый ампликон, соответствующий хромосомной делеции, наблюдался у дикого типа, что указывает на то, что эта область может естественным образом исключаться из генома с низкой скоростью в популяциях дикого типа. Секвенирование ампликона стыка выполняли для 20 клонов, генерируемых хромосомным самонацеливанием при участии CRISPR3. Генотипирование локуса выявило наличие одного химерного IS-элемента в каждом клоне, и, помимо того, у каждого клона была обнаружена транзигция выше-расположенного элемента в расположенную ниже последовательность внутри химеры (фиг. 8В). Наблюдаемый размер делеции варьировал от 101,865 до 102,146 п.о. Точный размер локуса транзигции не был постоянным, но оказался не случайным внутри клонов, что подразумевало потенциальное смещение механизма делеции. *S. thermophilus* имеет типичный рекомбинационный аппарат, кодируемый как RecA

(STER_0077), гомологами AddAB, функционирующими в виде двойных АТФ-зависимых ДНК-экзонуклеаз (STER_1681 и STER_1682), и геликазой (STER_1742) семейства RecD. Высокая нуклеотидная идентичность между фланкирующими IS-элементами и способность *S. thermophilus* к осуществлению сайт-специфической рекомбинации (4) подтверждают эффективность RecA-опосредованной рекомбинации, участвующей в эксцизии геномного сегмента (фиг. 8С).

Затем оценивали возможность использования CRISPR-Cas нацеливания для выделения делеций для каждого локуса с одинаковой генетической структурой. С этой целью создавали три CRISPR3 массива повтор-спейсер, один из которых таргетировал транспортер олигонуклеотидов в первом локусе, *prtS* во втором локусе и ген АТФ-зависимой помпы ионов меди в третьем локусе, и клонировали в pORI28 (фиг. 5). Для скрининга на наличие делеции создавали праймеры, фланкирующие IS-элементы в каждом локусе, для амплификации каждого стыка делеции (фиг. 8D). Отсутствие локусов дикого типа также подтверждали в каждом случае путем разработки внутренних праймеров для каждого геномного острова (фиг. 8E). После трансформаций при помощи направленных плазмид выделяли делеции в каждом локусе и подтверждали их отсутствие у дикого типа. Секвенирование ампликонов стыка делеции подтвердило наличие одного оставшегося следа химерного IS-элемента, что указывает на общий механизм делеции в каждом локусе. Интересно, что праймеры, фланкирующие IS-элементы, также были амплифицированы из гДНК дикого типа, что дополнительно указывает на то, что гетерогенность популяции возникает естественным образом в каждом локусе из-за спонтанных геномных делеций. Эти результаты предполагают, что расщепление Cas9 специфичных в последовательностей происходит избирательно в вариантах, не содержащих комбинации протоспейсера с PAM, необходимые для таргетирования. Таким образом, спонтанные геномные делеции могут быть выделены при помощи CRISPR-Cas нацеливания как эффективного механизма отбора микробных вариантов, потерявших такие геномные острова.

Пример 10. Скрининг популяций.

В этом исследовании нативные системы типа IIA, содержащиеся в *S. thermophilus*, использовали для определения спонтанных делеций крупных геномных островов. Посредством независимого таргетирования четырех островов в *S. thermophilus* создавали стабильных мутантов, у которых в общей сложности отсутствовало 7% генома. Характеристика стыков делеций предполагала, что IS-зависимый механизм рекомбинации вносит вклад в гетерогенность популяции и обеспечивает выявление событий делеции размером от 8 до 102 т.п.о. Точное картирование химерных IS-элементов указывает на то, что события естественной рекомбинации, вероятно, ответственны за большие хромосомные делеции в *S. thermophilus* и потенциально могут быть использованы для направленного редактирования генома.

Наши результаты показывают, что клоны дикого типа были удалены из популяции, в то время как мутанты без признаков направленной CRISPR-Cas выжили. Таким образом, адаптивные острова были идентифицированы и подтверждены, что показывает, что точное нацеливание посредством эндогенного Cas9 может быть использовано для выделения в смешанных популяциях вариантов с крупными делециями.

Эволюция генома бактерий происходит через горизонтальный перенос генов, внутреннюю мутацию и реорганизацию генома. Секвенирование генома и сравнительный анализ штаммов *S. thermophilus* выявили значительное разрушение генома, но также показали, что адаптация к обогащенным питательным средам происходила через специфическое для ниши приобретение генов (18, 35). Наличие MGE, включающих интегрированные и конъюгированные элементы, профаги и IS-элементы, в геномах *S. thermophilus* свидетельствует об их быстрой эволюции в молочной среде (38-39). Мобильные генетические признаки облегчают приобретение генов и, наоборот, инактивацию или потерю нежизненно важных последовательностей. Следовательно, MGE придают геномную пластичность как средство повышения приспособляемости или изменения экологического образа жизни. Полученные результаты четко указывают на то, что CRISPR-Cas таргетирование этих элементов может влиять на хромосомные перегруппировки и гомеостаз. Это контрастирует с экспериментами по нацеливанию на существенные признаки, которые привели к выбору вариантов с инактивированным аппаратом CRISPR-Cas (Jiang 2013). Мутация жизненно важных OPC не является эффективным подходом, для того чтобы избежать CRISPR-Cas нацеливания, и, таким образом, остаются только клоны с инактивированными CRISPR-Cas системами. По замыслу нацеливание на генетические элементы, которые согласно прогнозам являются гипервариабельными и инородными, продемонстрировало, что жизнеспособными оказались варианты с измененными локусами, сохранившими активные CRISPR-Cas системы во время событий самонацеливания.

Несмотря на почти повсеместное распространение IS-элементов в бактериальных геномах они остаются загадочной генетической сущностью в основном из-за их функционального многообразия и пластичности (34). Полученные результаты позволяют спрогнозировать рекомбинацию между родственными IS-элементами путем анализа их местоположения, ориентации и сохранения последовательности (фиг. 4 и 5). CRISPR-Cas нацеливание можно далее использовать для эмпирической проверки гетерогенности популяции в каждом прогнозируемом локусе и одновременно для увеличения извлечения мутантов с низкой степенью встречаемости. Высокая степень встречаемости MGE в молочнокислых бактериях и особенно у *S. thermophilus* соответствует их роли в видообразовании этих гипердаптированных бактерий через эволюцию генома (39-40). Более того, извлечение мутантов с геномными делециями путем

CRISPR-Cas нацеливания может позволить получить фенотипическую характеристику генов с неизвестной функцией. Мутанты, демонстрирующие делецию острова размером 102 т.п.о., кодирующего *lac*-оперон, значительно увеличили время генерации по сравнению с диким типом и достигли более низкого конечного значения OD. При наличии в них 102 предсказанных ОРС вероятно, что возможны дополнительные фенотипы и многие гены не имеют аннотированных функций. Учитывая промышленную значимость генов, специфических для ниши, таких как *prtS*, этот метод позволяет выполнить непосредственную оценку того, каким образом кодируемые островом гены способствуют адаптации к росту в молоке. Более того, она происходит в геномном и экологическом контексте, естественном для этих горизонтально приобретенных признаков, поскольку они, вероятно, были приобретены в виде отдельных островов. Эти результаты обеспечивают новые возможности для применения самонацеленных систем CRISPR-Cas9 в бактериях для изучения транспозиции, механизмов восстановления ДНК и пластичности генома.

CRISPR-Cas системы обычно ограничивают генетическое разнообразие за счет интерференции с генетическими элементами, но приобретенные MGE могут также обеспечивать адаптивные преимущества бактериям-хозяевам. Таким образом, преимущество сохранения интегрированных в геном MGE, несмотря на CRISPR-Cas таргетирование, является важным фактором гомеостаза генома. В совокупности полученные результаты показывают, что *in silico* предсказание GEI можно связать с CRISPR-Cas нацеливанием с целью выделения клонов, демонстрирующих крупные геномные делеции. Следы химерной инсерционной последовательности в каждом стыке делеции указывают на общий механизм удаления всех четырех островов. Высокая степень встречаемости самонацеленных спейсеров, демонстрирующих идентичность с геномными локусами, в сочетании с экспериментальными демонстрациями геномных изменений, позволяет предположить, что CRISPR-Cas самонацеливание может в значительной степени способствовать эволюции генома бактерий (28, 30). В совокупности изучение крупных делеций, индуцированных CRISPR-Cas, подтверждает, что этот подход является быстрым и эффективным средством оценки важности и функциональности генных кластеров, лишенных аннотации, и определения минимальных бактериальных геномов на основе хромосомных делеций, происходящих посредством мобильных элементов.

На фиг. 9 показаны определенные генетические локусы для оценки летальности, основанной на CRISPR-Cas системе типа II, посредством таргетирования генома *Streptococcus thermophilus* LMD-9. Способы выполнения этой оценки известны в уровне техники, см., Selle and Barrangou PNAS. 112 (26):8076-8081 (2015).

Были протестированы обе ортогональные системы типа II (CRISPR1 и CRISPR3); мишени CRISPR1 показаны темно-серым цветом, а мишени CRISPR3 - светло-серым цветом. Конкретные генетические признаки были выбраны для тестирования

- (i) межгенных областей (INT),
- (ii) мобильных генетических элементов (IS, GEI1-GEI3, PRO, *lacZ*, EPS),
- (iii) жизненно важных генов (*dltA*, LTA),
- (iv) полюсов реплихоры (ORI, TER) и прямых или обратных нитей ДНК (внешних мишеней по отношению к внутренним мишеням).

На фиг. 10 показана летальность на основе CRISPR, достигаемая посредством таргетирования областей, показанных на фиг. 9. Логарифмическое уменьшение в единицах КОЕ рассчитывали в отношении трансформации контроля - несамонацеленной плазмиды; *pORI28*; Летальность находилась в пределах уменьшения 2-3 log для всех тестируемых мишеней независимо от местоположения в хромосоме, кодирующей последовательности или важности для выживания.

На фиг. 11 показаны профили транскрипции штаммов с CRISPR-опосредованными делециями геномных островов. Выделение и генотипирование клеток, выживших при CRISPR нацеливании геномных островов 1-4, привело к идентификации стабильных независимых мутантов, не содержащих геномный остров, таргетированный в каждом эксперименте. Затем клетки размножали и их общую РНК выделяли и секвенировали. Используя этот подход, получали профили транскрипции путем отображения прочтений при секвенировании в референсном геноме. В каждом случае отсутствие данных по секвенированию для предсказанного локуса геномного острова далее предполагало потерю генетического объекта-мишени, при этом оказывая минимальное влияние на экспрессию генов ядра во всей остальной части генома.

Кроме того, данные секвенирования РНК подтверждают границы делеций при помощи картирования покрытий прочтений и сравнения значений транскрипции, а также подтверждают распознавание фенотипа при помощи сравнительной транскриптомики, полученной с использованием одного и того же набора данных. В частности, данные высокопроизводительного РНК-секвенирования подтверждают отсутствие транскрипционной активности, присутствующей в ожидаемых областях делеций, как показано на фиг. 12. На фиг. 12 показано Log₂ покрытие прочтений при секвенировании трансформированных РНК штаммов с делециями геномных островов и для каждого штамма с геномным островом (GEI1-GEI4); отсутствие данных секвенирования для предсказанных локусов с геномными островами также подтвердило потерю генетического объекта-мишени с минимальным воздействием на экспрессию генов ядра во всей остальной части генома.

Фиг. 13 дополнительно подтверждает отсутствие транскрипционной активности удаленных генов. Для каждого из штаммов с делециями геномных островов (GEI1-GEI4) экспрессия генов, кодируемых на каждом из островов-мишеней (черные), была минимальной. Гены, кодируемые в GEI1, показаны на верхней левой панели, гены, кодируемые в GEI2, показаны на верхней правой панели, гены, кодируемые в GEI3, показаны на нижней левой панели, а гены, кодируемые в GEI4, показаны на нижней правой панели. В общем случае делеции геномных островов 1 и 2 имели минимальное влияние на транскрипцию других генов (серый), тогда как делеция геномных островов 3 и 4, по-видимому, влияла на транскрипцию других генов, не кодируемых в островах.

Кроме того, данные секвенирования РНК использовались для сравнения уровней транскрипции генов, не кодируемых на удаленном острове (GEI4), т.е. других генов, все еще присутствующих в хромосоме, и идентификации фенотипов, связанных с геномными делециями. Гены, которые дифференциально транскрибируются в штамме с делецией, указывают на то, что на клеточные процессы влияют гены, которые были потеряны, или что существует компенсация потери активности этих генов. Таким образом, можно сделать выводы о путях, к которым относятся эти гены или геномные области. В табл. 5 приведен список дифференциально экспрессируемых генов, идентифицированных в штамме GEI4 с делецией. Многие из генов, которые согласно наблюдениям являются дифференцированно экспрессируемыми, связаны с биосинтетической способностью *Streptococcus thermophilus*, включая биосинтез ароматических аминокислот и пуринов.

Пример 11. Направленное уничтожение *Lactobacillus casei* при помощи CRISPR-Cas системы типа II.

Приведенные в качестве примера гиды CRISPR-Cas типа II для *L. casei* представлены на фиг. 15. На верхней панели представлен предсказанный гид, в то время как на нижней левой панели показана приведенная в качестве примера двойная структура гида, а на нижней правой панели показана приведенная в качестве примера одонитевая структура гида.

Пример 12. Направленное уничтожение *Lactobacillus gasser* при помощи CRISPR-Cas системы типа II.

Приведенные в качестве примера гиды типа II для направленного уничтожения *L. gasser* представлены на фиг. 16. На верхней панели представлен предсказанный гид, а нижней левой панели приведен правильный двойной гид сгРНК:трасРНК (подтвержденный секвенированием РНК), на нижней правой панели показан приведенный в качестве примера предсказанный одонитевой выступающий гид.

Плазмиды трансформировали в *L. gasser*, каждая плазида имела разные конструкции: пустой вектор pTRK563, конструкцию с правильным протоспейсером, но неверным PAM, с правильным PAM, но с протоспейсером, который не находится в массиве, и с правильными протоспейсером и PAM. Результаты показаны в фиг. 19. Плазида, имеющая правильный протоспейсер и правильный PAM, показала существенно более высокую степень направленности интерференции и гибели клеток.

Пример 13. Направленное уничтожение *Lactobacillus pentosus* при помощи CRISPR-Cas системы типа II.

Приведенные в качестве примера гиды типа II для направленного уничтожения *L. pentosus* представлены на фиг. 17. На верхней панели показан предсказанный гид. На левой нижней панели представлен правильный двойной гид сгРНК:трасРНК (подтвержденный секвенированием РНК), а на нижней правой панели показан приведенный в качестве примера предсказанный искусственный одонитевой выступающий гид.

Плазмиды трансформировали в *L. pentosus*, причем каждая плазида имела разные конструкции: конструкцию с правильным протоспейсером, но неверным PAM, конструкцию с правильным PAM, но с протоспейсером, который не находится в массиве, пустой вектор pTRK563, и с правильным протоспейсером и правильным PAM. Результаты показаны в фиг. 20. Плазида, имеющая правильный протоспейсер и правильный PAM (LpeI gttaat), показала существенно более высокую степень направленности интерференции и гибели клеток.

Пример 14. Направленное уничтожение *Lactobacillus jensenii* при помощи CRISPR-Cas системы типа II.

На фиг. 18 представлены в качестве примера гиды CRISPR-Cas типа II. На панели слева показан правильный двойной гид сгРНК:трасРНК, подтвержденный секвенированием РНК, на панели справа представлен в качестве примера предсказанный искусственный одонитевой гид.

Плазмиды были трансформированы в *L. jensenii*, каждая из плазмид имела разные конструкции: конструкцию, содержащую пустой вектор pTRK563, конструкцию с правильным протоспейсером, но неверным PAM, конструкцию с правильным PAM, но с протоспейсером, который не находится в массиве, и конструкцию, имеющую правильный протоспейсер и правильный PAM. Результаты показаны в фиг. 22. Плазида, имеющая правильный протоспейсер и правильный PAM, показала существенно более высокую степень направленности интерференции и гибели клеток.

Пример 15. Направленное уничтожение *Lactobacillus casei* NCK 125 при помощи CRISPR-Cas системы типа I.

На фиг. 21 представлен в качестве примера гид CRISPR-Cas типа I для *L. casei*, который содержит нативную лидерную последовательность типа I и повтор, найденный в *L. casei* NCK 125. PAM 5'-YAA-3' была предсказана с использованием нативных спейсерных последовательностей в организме. Этот ис-

кусственный массив содержит спейсер, который таргетирует ген 16s рДНК в геноме хозяина. Результаты представлены на фиг. 22, где показано значимое уменьшение между пустым вектором и двумя разными искусственными массивами: один из которых содержит единичный спейсер, таргетирующий +цепь в гене 16s (1-2 alt), а другой содержит исходный спейсер, таргетирующий +цепь, и также содержит дополнительный спейсер, таргетирующий -цепь в гене 16s (1, 2, 3).

CRISPR-Cas системы, как раскрыто в настоящем описании, могут быть использованы, например, для

(i) направленного сокращения патогенов в случае медицинского вмешательства (например, патогенов, включая без ограничения грибы, нематоды, простейшие (например, малярию) цестоды, кокцидии (микроспоридии), трематоды, пентастомиды, акантоцефалы, членистоногие и т.п.);

(ii) для защиты потребляемых веществ (пищевые системы, животные, культуры);

(iii) для контроля и/или удаления нежелательных организмов из промышленных ферментативных процессов (сырья, технологического оборудования, стартовых культур); и

(iv) для контроля экологических микробных консорциумов для воздействия на экосистемы и/или химические циклы, а также для рекультивации.

Таблица 1

Бактериальные штаммы и плазмиды

Обозначение штамма	Описание	Источник
<i>E. coli</i> EC1000	Хозяин для pORI плазмид, хромосомных repA ⁺ (pWV01), Km ^R Хозяин для pTRK935	47
<i>S. thermophilus</i> LMD-9 <i>S. thermophilus</i> LMD-9 с pTRK669	Дикий тип Дикий тип, RepA ⁺ и Cm ^R приобретенные посредством pTRK669	40 в этом исследовании
Плазмиды		
pORI28	Широкоиспользуемый нерепликативный вектор, Em ^R	47
pTRK669	Ts-хелперная плаزمида repA ⁺ , Cm ^R	44
pCRISPR1::lacZ	pORI28::CRISPR1-лидер-RSR-lacZ N-терминальный спейсер	This study
pCRISPR3::lacZ	pORI28::CRISPR1-лидер-RSR-lacZ активный сайт спейсер	This study
pCRISPR3::ABC	pORI28::CRISPR1-лидер-RSR-ABC спейсер	This study
pCRISPR3::Cu	pORI28::CRISPR1-лидер-RSR-Cu помпа спейсер	This study
pCRISPR3::prtS	pORI28::CRISPR1-лидер-RSR-prtS спейсер	This study
pCRISPR3::lacZ pCRISPR1::несамо-нацеленная	pORI28::CRISPR3-лидер-RSR-lacZ N-терминальный спейсер pORI28::CRISPR1-лидер-RSR-несамонацеленный спейсер	This study This study

Таблица 2

Праймеры

Название праймера	Последовательность	Функция
C1_N-term_F	CAAGAACAGTTATTGATTTTATAATCACTATGTGGGTATGAAATCTCAAAATCATTG AGGTTTGTGACTCTCAAGATTAAAGTAACGTACAACATTAAGAGATTGTCTTAACTT	Матрица для SOE-ПЦР
C3_N-term	AGCGGATAACAATTTACGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGAA AATCTTCAAGAGATTCAAAATACTGTTTGGAAACCATTCGAAACAACACAGCTCTAAAA CCTCGTAGGATATCTTTCTAC	Матрица для SOE-ПЦР
C1_N-term_R	AGCGGATAACAATTTACGTTGTGACAGTTACTTAAATCTTGAGAGTACAAAAACAGGGGA GATGAAGTTAAGACAATCTCTTAATGT	Матрица для SOE-ПЦР
C3_A-site	AGCGGATAACAATTTACGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC TTGGTCTTCAACACAGCTTGCTGTAGTTTGGAAACCATTCGAAACAACACAGCTCTAAAA CCTCGTAGGATATCTTTCTAC	Матрица для SOE-ПЦР
C3_ABC	AGCGGATAACAATTTACGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC GAGATAAAATCCAGCCACCGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC taggatatctttttctac	
C3_prtS	AGCGGATAACAATTTACGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC TTTGAGGTCTGAGAATACACGCGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC taggatatctttttctac	
C3_Cu	AGCGGATAACAATTTACGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC TCAATCAATCGTTTCAGCTGCTAAGTTTTTGGAAACCATTCGAAACAACACAGCTCTAAACCTCAGTC ctcgtaggatatctttttctac	
C3LF	AGCAGGGATCCTGGTAATAAGTATAGATAGTCTTG	Амплификация Sth3 Лидер из гДНК
C3LR	CTCGTAGGATATCTTTCTAC	Амплификация Sth3 Лидер из гДНК
C1F	AGCAGGGATCCCAAGAACAGTTATTGATTTTATAATC	CRISPR1 SOE-ПЦР Прямой
C3F	AGCAGGGATCCTGGTAATAAGTATAGATAGTCTTG	CRISPR3 SOE-ПЦР Прямой
C1C3R	TGCTGGAGCTCGTGAATTTGTTATCCGCT	CRISPR1 и CRISPR3 SOE-ПЦР Обратный
1193F	TTGAACACATAGGAACCTCATA	Амплификация стыка делеции
1193R	CGTAAGGTTTGTATGACTCAAG	Амплификация стыка делеции
pORI28F	TTGGTTGATAATGAACGTGTGCTG	Секвенирование MCS из pORI28
pORI28R	TTGTTGTTTTTATGATTACAAAGTGA	Секвенирование MCS из pORI28

Таблица 3

Предполагаемые инородные геномные острова и островки

Геномный остров	Область ORC	Длина, т.п.о.	Геномное содержание, %	Известные гены	IS-семейство
1	STER_139- STER_148	7,81	37,1	Транспортеры олигопептидов	IS6
2	STER_840- STER_848	10,29	39,9	Протеиназа PrtS	ISsth16/ IS1167
3	STER_881- STER_888	8,71	39,3	Помпа ионов меди	IS1191
4	STER_1277- STER_1380	101,76	37,2	Катаболизм лактозы, регуляция 2-компонента, синтез бактериоцина, ABC транспортеры	IS1193

Таблица 4

Гомологи до примерно 239 жизненно важных OPC, идентифицированных в *S. thermophilus*

Геном-часть	STER	начало	конец	направление		покрытие	e-значение	aa id	Bacillus subtilis аннотация
CSTER	1	101	1465	+	белок DnaA инициатор хромосомной репликации	98%	4.00E-124	44%	белок DnaA инициатор хромосомной репликации
CSTER	2	1620	2756	+	бета субъединица ДНК полимеразы III	99%	1.00E-88	39%	dnap бета субъединица ДНК полимеразы III
CSTER	6	5818	6387	+	Пептидил-тРНК гидролаза	97%	1.00E-61	51%	
CSTER	9	10331	10702	+	Белок клеточного цикла	94%	1.00E-05	22%	белок DivIC клеточного деления
CSTER	12	12122	13387	+	тРНК(Ile)-лизидин синтаза, MesJ	92%	3.00E-40	32%	тРНК(Ile)-лизидин синтаза tils
CSTER	13	13469	14011	+	гипоксантин-гуанин фосфорибозилтрансфераза	99%	2.00E-82	63%	гипоксантин-гуанин фосфорибозилтрансфераза hprt
CSTER	40	25595	26425	+	MreC белок, определяющий форму клеток	94%	1.00E-22	28%	MreC белок, определяющий форму клеток
CSTER	43	28617	29582	+	рибозо-фосфат пирозофоскиназа	97%	3.00E-145	65%	Рибозо-фосфат пирозофоскиназа prs
CSTER	47	32842	33846	+	Предпол. глицерин-3-фосфат ацилтрансфераза	94%	9.00E-124	54%	Фосфат ацилтрансфераза PlsX
CSTER	48	33846	34091	+	ацил-переносающий белок	76%	1.00E-08	40%	ацил-переносающий белок
CSTER	65	53036	54727	-	аргинил-тРНК синтаза	98%	6.00E-52	28%	
CSTER	95	72798	77192	+	ДНК полимеразы III PolC	99%	0.00E+00	51%	ДНК полимеразы III PolC-тип
CSTER	105	82956	83723	+	30S рибосомный белок S2	96%	1.00E-118	69%	
CSTER	106	83841	84881	+	Фактор элонгации трансляции Ts	97%	8.00E-72	44%	
CSTER	117	97363	98706	+	цистеинил-тРНК синтаза	99%	1.00E-178	54%	
CSTER	127	104406	104852	+	50S рибосомальный белок L13	99%	9.00E-57	57%	
CSTER	128	104880	105272	+	30S рибосомальный белок S9	100%	3.00E-59	68%	
CSTER	193	159363	160967	+	СТР синтаза	99%	0.00E+00	68%	СТР синтаза purg
CSTER	199	166439	166897	-	консерв. гип. белок.	81%	2.00E-07	28%	белок NrdI
CSTER	208	176203	176472	+	30S рибосомальный белок S15	100%	6.00E-38	63%	
CSTER	217	183028	184185	+	ундекапринил пирозофосфатаза	98%	2.00E-68	38%	возможно ундекапринил пирозофосфатаза N-ацетилглюкозоменил 1-фосфат трансфераза tagO
CSTER	218	184346	185116	+	ABC транспортер АТФаза	95%	8.00E-130	71%	Растительный белок 296 sufC
CSTER	219	185153	186415	+	гип. белок	97%	4.00E-111	42%	Белок сборки FeS кластеров SufD
CSTER	220	186469	187701	+	предп. аминотрансфераза (класс V)	98%	0.00E+00	60%	цистеин десульфурара SufS
CSTER	221	187688	188122	+	Семейство белков NifU	95%	3.00E-48	51%	цинк-зависимая сульфотрансфераза SufU
CSTER	245	208199	208993	+	фосфотидат цитидилилтрансфераза	98%	3.00E-67	41%	фосфотидат цитидилилтрансфераза
CSTER	247	210343	212205	+	пролил-тРНК синтаза	98%	0.00E+00	50%	
CSTER	252	215735	216022	+	Ко-шаперон GroES (HSP10)	97%	1.00E-18	43%	
CSTER	253	216071	217690	+	шаперон GroEL (семейство HSP60)	98%	0.00E+00	75%	
CSTER	261	224055	224204	+	50S рибосомальный белок L33	97%	9.00E-12	50%	
CSTER	262	224216	224392	+	Субъединица SccE				Субъединица SccE
CSTER	268	227616	230117	+	препротеин транслоказы	81%	1.40E+00	21%	препротеин транслоказы
CSTER	273	232342	233796	+	лейцил-тРНК синтаза	99%	0.00E+00	70%	
CSTER	274	233808	234629	+	никотинат фосфорибозилтрансфераза	97%	0.00E+00	63%	никотинат фосфорибозилтрансфераза
CSTER	286	248252	249580	+	NAD синтаза	100%	9.00E-109	59%	NH(3)-зависимая NAD(+) синтаза nade
CSTER	302	260849	261664	+	УДФ-N-ацетилмурамат-аланин лигаза	99%	3.00E-170	55%	УДФ-N-ацетилмурамат-аланин лигаза murC
CSTER	307	264328	265041	+	Глутамат рацемаза	94%	7.00E-83	48%	Глутамат рацемаза l racE
CSTER	308	265034	265615	+	Белок A сегрегации и конденсации	93%	5.00E-37	39%	Белок A сегрегации и конденсации
CSTER	313	269788	270297	+	Белок B сегрегации и конденсации	94%	9.00E-36	40%	Белок B сегрегации и конденсации
CSTER	349	303983	305959	+	рРНК метилтрансфераза	96%	3.00E-58	51%	рРНК метилтрансфераза
CSTER	357	312857	314032	+	транскетолаза	98%	0.00E+00	58%	транскетолаза tkt
CSTER	358	314036	314938	+	Белок DnaB инициации репликации хромосомы/прикрепления к мембране	63%	3.00E-02	18%	Белок инициации репликации хромосомы и прикрепления к мембране
CSTER					рибосомальный белок DnaI	99%	7.00E-58	35%	рибосомальный белок DnaI

CSTER	359	315044	316354	+	ГТФ-связывающий белок EngA	100%	0.00E+00	67%	ГТФаза Der
CSTER	368	321054	322394	-	серил-тРНК синтетаза	100%	0.00E+00	63%	
CSTER	376	329673	330116	+	консервативный гип. белок	89%	6.00E-33	43%	Белок биосинтеза тРНК треонилкарбамоиладенозин а TsaE
CSTER	380	332545	333732	+	Фактор элонгации транскрипции	90%	2.00E-113	46%	Белок терминации транскрипции/антитерминации NusA
CSTER	383	334363	337194	+	Фактор инициации транскрипции IF-2	99%	0.00E+00	55%	
CSTER	387	339864	341309	-	УДФ-N-ацетилмуромоилаланил-D-глутамат--2, 6-диаминопимелат лигаза	92%	3.00E-52	29%	УДФ-N-ацетилмуромоилаланил-D-глутамат--2, 6-диаминопимелат лигаза murE
CSTER	419	364962	365894	+	предп. марганец-зависимая неорганическая пирофосфатаза	99%	3.00E-125	58%	марганец-зависимая неорганическая пирофосфатаза ppras
CSTER	430	375355	375579	+	Ацил-переносящий белок	93%	5.00E-14	49%	Ацил-переносящий белок малонил CoA-ацил переносящий белок
CSTER	432	376667	377593	+	Ацил-переносящий белок S-малонилтрансфераза	95%	3.00E-86	47%	малонил CoA-ацил переносящий белок трансфераза
CSTER	433	377606	378340	+	3-кетоацил- (ацил-переносящий-белок) редуктаза	99%	6.00E-79	47%	3-оксоацил- [ацил-переносящий-белок] редуктаза FabG
CSTER	434	378401	379633	+	3-оксоацил- (ацил-переносящий-белок) синтаза II	99%	1.00E-133	48%	3-оксоацил- [ацил-переносящий-белок] синтаза 2
CSTER	435	379637	380125	+	Субъединица биотин карбоксил-переносящего белка ацетил-CoA карбоксилазы	98%	9.00E-26	37%	биотин карбоксил-переносящий белок ацетил-CoA карбоксилазы
CSTER	437	380667	382037	+	Субъединица биотин карбоксилаза ацетил-CoA карбоксилазы	98%	0.00E+00	60%	Биотин карбоксилазы 1
CSTER	438	382043	382909	+	Бета субъединица ацетил-CoA карбоксилазы	94%	8.00E-98	51%	Бета субъединица карбоксил трансферазы ацетил-коэнзим A карбоксилазы
CSTER	439	382906	383676	+	альфа субъединица ацетил-CoA карбоксилазы	76%	2.00E-81	53%	Альфа субъединица карбоксил трансферазы ацетил коэнзим A карбоксилазы
CSTER	442	385819	387417	+	Консервативный гип. белок	98%	0.00E+00	58%	рибонуклеаза Y ymda
CSTER	455	402156	402470	+	50S рибосомальный белок L21	99%	2.00E-44	66%	
CSTER	456	402507	402797	+	50S рибосомальный белок L27	95%	5.00E-47	79%	
CSTER	460	405038	405805	+	дигидродипиколонат редуктаза	98%	7.00E-95	53%	4-гидроксидигидродипиколонат редуктаза dapb
CSTER	461	405802	407010	+	тРНК CCA-пирофосфорилаза	98%	6.00E-86	40%	CCA-добавляющий фермент
CSTER	475	422256	422813	+	Фактор рециклинга рибосомы	100%	2.00E-68	56%	
CSTER	485	430547	432550	+	метионил-тРНК синтетаза	98%	0.00E+00	58%	
CSTER	492	438748	439890	+	Предшественник белка созревания протеазы	91%	6.00E-11	26%	Белок фолдазы PrsA
CSTER	493	440232	442850	+	аланил-тРНК синтетаза	99%	0.00E+00	52%	
CSTER	513	460453	463104	+	валил-тРНК синтетаза	99%	0.00E+00	63%	
CSTER	523	469888	471168	+	белок FtsW клеточного деления	93%	4.00E-59	34%	предполагаемая липид II флиппаза FtsW
CSTER	524	471406	472602	+	Фактор элонгации Tu	99%	0.00E+00	76%	
CSTER	525	472851	473609	+	триосефосфат изомераза	97%	8.00E-109	62%	триосефосфат изомераза tpia
CSTER	526	473848	474477	+	Тимидилат киназа	96%	4.00E-73	55%	Тимидилат киназа tmk
CSTER	527	474486	475361	+	Дельта субъединица ДНК полимеразы III	96%	8.00E-41	33%	Дельта субъединица ДНК полимеразы III
CSTER	539	484827	485744	+	Альфа субъединица глицил-тРНК синтетазы	97%	4.00E-169	74%	
CSTER	540	486028	488064	+	Бета субъединица глицил-тРНК синтетазы	97%	0.00E+00	45%	
CSTER	567	512413	512916	+	50S рибосомальный белок L10	100%	7.00E-61	57%	
CSTER	568	512991	513359	+	50S рибосомальный белок L7/L12	100%	4.00E-38	67%	
CSTER	603	548926	550308	+	N-ацетилглюкозамин-1-фосфат уридилтрансфераза	97%	3.00E-164	52%	Бифункциональный белок GlmU
CSTER	623	566278	566781	+	дигидрофолат редуктаза	94%	8.00E-34	38%	дигидрофолат редуктаза dfra

CSTER	626	568210	568809	+	ГТФаза EngB	96%	7.00E-92	63%	возможный ГТФ-связывающий белок EngB
CSTER	632	573613	574254	-	пред. глицерин-3-фосфат ацилтрансфераза PlsY	90%	3.00E-40	46%	глицерин-3-фосфат ацилтрансфераза plsY
CSTER	633	574392	576341	+	Субъединица В ДНК топоизомеразы IV	97%	0.00E+00	71%	Субъединица В ДНК топоизомеразы 4
CSTER	634	576969	579434	+	Субъединица А ДНК топоизомеразы IV	95%	0.00E+00	54%	Субъединица А ДНК топоизомеразы 4
CSTER	660	598871	599725	+	метилентетрагидрофолат дегидрогеназа/ метилентетрагидрофолат циклогидролаза	99%	2.00E-109	56%	Бифункциональный белок Fold
CSTER	668	605570	606469	+	ГТФаза Era	98%	5.00E-143	64%	ГТФаза Era
CSTER	670	607313	607927	+	дефосфо-CoA киназа	95%	2.00E-50	45%	дефосфо-CoA киназа coae
	672	609192	609338		50S ribosomal prot. L33	97%	1.00E-16	52%	
CSTER	684	620012	621316	+	енолаза	100%	0.00E+00	70%	енолаза epo
CSTER	731	664470	665690	-	предп. цитозин-C5 специф. ДНК метилаза	60%	5.00E-16	31%	Возможная субъединица YdiO метилазы BsuMI модификации
CSTER	733	666939	668429	-	лизил-тРНК синтетаза (класс II)	97%	0.00E+00	64%	
CSTER	761	694548	695582	+	Дельта субъединица ДНК полимеразы III	84%	2.00E-37	33%	Неохарактеризованный белок YqeN
CSTER	773	704704	706056	+	УДФ-N-ацетилмурамоилаланин--D-глутамат лигаза	98%	8.00E-147	49%	УДФ-N-ацетилмурамоилаланин--D-глутамат лигаза murD
CSTER	774	706060	707130	+	N-ацетилглюкозаминил трансфераза	95%	3.00E-30	27%	УДФ-N-ацетилглюкозамин-N-ацетилмурамил-(пентапептид) фосфорил-ундекапренол N-ацетилглюкозамин трансфераза murG
CSTER	775	707140	708264	+	Белок клеточного деления FtsQ	79%	4.00E-15	26%	Белок клеточного деления DivlB
CSTER	776	708388	709764	+	Белок клеточного деления FtsA	84%	4.00E-103	44%	Белок клеточного деления FtsA
CSTER	777	709793	711115	+	Белок клеточного деления FtsZ	92%	2.00E-123	53%	Белок клеточного деления FtsZ
CSTER	783	714803	717592	+	изолейцил-тРНК синтетаза	99%	0.00E+00	58%	
CSTER	787	720193	720459	-	50S рибосомный белок L31	100%	1.00E-22	50%	
CSTER	793	725315	726394	+	Фактор высвобождения пептидной цепи 1	99%	6.00E-158	59%	
CSTER	796	727911	729161	+	Серин гидроксиметилтрансфераза	97%	2.00E-171	60%	Серин гидроксиметилтрансфераза glya
CSTER	833	760678	762396	-	предп. фосфоглюкомутаза	98%	0.00E+00	47%	фосфоглюкомутаза pgm
CSTER	850	783500	783736	-	30S рибосомный белок S20	94%	1.00E-13	46%	
CSTER	864	797961	799307	+	аспаратинил-тРНК синтетаза	99%	4.00E-180	56%	
CSTER	903	833853	835661	+	D-фруктозо-6-фосфат амидотрансфераза	100%	0.00E+00	59%	глутамин--фруктозо-6-фосфат аминотрансфераза [изомеризация] glmS
CSTER	915	845349	846911	+	Белок частицы распознавания сигнала	95%	2.00E-178	57%	Белок частицы распознавания сигнала
CSTER	919	848991	849845	+	Рибосомального биогенезиса ГТФаза	98%	8.00E-101	50%	Рибосомального биогенезиса ГТФаза rbga
CSTER	923	852605	854749	+	ДНК топоизомеразы I	98%	0.00E+00	64%	ДНК топоизомеразы 1
CSTER	994	915311	917623	+	АТФ-зависимая ДНК геликаза PcrA	99%	0.00E+00	54%	АТФ-зависимая ДНК геликаза PcrA
CSTER	1034	959843	961231	-	Субъединица E2 дегидрогеназы разветвленных альфа-кетокислот	99%	5.00E-69	34%	Сукцидилтрансферазный компонент остатка дигидропиолилизина 2-оксоглутарат дегидрогеназного комплекса odhB
CSTER	1036	962412	963383	-	ацетон дегидрогеназный комплекс, E1 компонент, альфа субъединица	81%	3.00E-33	28%	Альфа субъединица компонетна E1 пируват дегидрогеназы pdha
CSTER	1087	1007529	1007888	-	50S рибосомальный белок L20	85%	2.00E-56	80%	
CSTER	1088	1007945	1008145	-	50S рибосомальный белок L35	100%	1.00E-23	62%	
CSTER	1089	1008184	1008714	-	Фактор инициации транскрипции 3 (IF-3)	98%	2.00E-63	60%	
CSTER	1102	1017867	1018880	-	Фактор высвобождения пептидной цепи 2	91%	2.00E-137	57%	
CSTER	1115	1029796	1031136	-	Сенсорная гистидин киназа	97%	4.00E-117	71%	Регуляторный белок транскрипции YucF
CSTER	1116	1031129	1031836	-	Двухкомпонентный регулятор ответа	68%	2.00E-120	47%	Сенсона гистидин киназа YucG

CSTER	1123	1036555	1037826	-	УДФ-N-ацетилглюкозамин 1- карбоксивинилтрансфераза	96%	2.00E-172	59%	УДФ-N-ацетилглюкозамин 1- карбоксивинилтрансфераза 1 muraa
CSTER	1135	1049456	1050649	-	S-аденозилметионин синтетаза	97%	0.00E+00	67%	S-аденозилметионин синтетаза metk
CSTER	1136	1050929	1051867	+	биотин- (ацетил-CoA карбоксилаза) лигаза	92%	4.00E-44	33%	бифункциональный лигаза/репрессор BirA
CSTER	1138	1052266	1053918	-	ДНК полимеразы III, гамма/тау субъединица	93%	3.00E-137	44%	Гамма субъединица ДНК полимеразы III/тау DnaX
CSTER	1144	1056546	1056893	-	50S рибосомальный белок L19	100%	2.00E-61	77%	
CSTER	1164	1073225	1074244	-	6-фосфофруктокиназа	93%	4.00E-124	61%	АТФ-зависимая 6-фосфофруктокиназа pfka
CSTER	1165	1074337	1077447	-	ДНК полимеразы III, альфа субъединица DnaE	96%	0.00E+00	35%	Альфа субъединица ДНК полимеразы III dnae
CSTER	1166	1077600	1078385	-	Фактор трансляции пирувата (SUA5)	42%	3.00E-04	24%	уwlC неизвестный консервативный белок с пируват РНК-связывающим мотивом TW
CSTER	1175	1085950	1086225	-	гистонподобный ДНК-связывающий белок	95%	4.00E-40	72%	ДНК-связывающий белок HU 1
CSTER	1182	1091629	1092501	-	геранилгеранил пирофосфат синтаза	89%	5.00E-58	44%	фарнезил дифосфат синтаза ispa ygid
CSTER	1188	1095764	1096462	-	Белок репликации ДНК DnaD	85%	1.00E-07	22%	Белок репликации ДНК DnaD
CSTER	1193	1101272	1102201	-	рибонуклеаза Z	99%	2.00E-105	50%	Ribonuclease Z rnz
CSTER	1197	1105450	1106877	-	Белок клеточного деления	88%	2.00E-57	38%	Rod форму-определяющий белок RodA
CSTER	1230	1131390	1132742	-	Фосфоглюкозамин мутаза	100%	0.00E+00	63%	Фосфоглюкозамин мутаза ybbt glnM
CSTER	1248	1154457	1156616	+	рибонуклеотид-дифосфат редуктаза, альфа субъединица	98%	0.00E+00	47%	рибонуклеотид-дифосфат редуктаза, альфа субъединица nrde
CSTER	1249	1156775	1157737	+	рибонуклеотид-дифосфат редуктаза, бета субъединица	97%	8.00E-98	49%	рибонуклеотид-дифосфат редуктаза, бета субъединица nrdf
CSTER	1256	1161831	1164284	-	ДНК гираза, А субъединица	98%	0.00E+00	62%	ДНК гираза, А субъединица
CSTER	1267	1174041	1176449	-	фенилаланил-тРНК синтетаза, бета субъединица	99%	0.00E+00	47%	
CSTER	1270	1177285	1178328	-	фенилаланил-тРНК синтетаза, альфа субъединица	98%	3.00E-161	62%	
CSTER	1271	1178950	1182483	-	АТФаза хромосомной сегрегации, SMC белок	99%	0.00E+00	37%	Разделяющий хромосомы белок Smc
CSTER	1272	1182486	1183175	-	рибонуклеаза III	89%	4.00E-68	47%	рибонуклеаза 3
CSTER	1273	1183332	1184267	-	дигидропипиколонат синтаза	93%	5.00E-88	47%	4-гидрокси-тетрагидропипиколонат синтаза dapA
CSTER	1274	1184605	1185681	-	аспартат-семальдегид дегидрогеназа	99%	1.00E-143	58%	аспартат-семальдегид дегидрогеназа asd
CSTER	1382	1293065	1294063	-	тиредоксин редуктаза	97%	7.00E-113	50%	ферредоксин--НАДФ редуктаза 2 yumc
CSTER	1383	1294065	1294784	-	тРНК (гуанин-N(1)-) - метилтрансфераза	100%	6.00E-87	56%	тРНК (гуанин-N(1)-) - метилтрансфераза
CSTER	1395	1303326	1304261	-	метионил-тРНК формилтрансфераза	96%	5.00E-98	49%	
CSTER	1396	1304279	1306675	-	Белок сборки праймосомного комплекса PriA	99%	0.00E+00	48%	Праймосомный белок N PriA
CSTER	1398	1307153	1307782	-	Гуанилат киназа	98%	5.00E-87	59%	Гуанилат киназа gmk
CSTER	1399	1308015	1309406	-	Белок клеточного деления FtsY	97%	8.00E-121	55%	Рецептор частиц, распознающих сигнал FtsY
CSTER	1422	1330503	1331339	-	Неорганический полифосфат киназа /АТФ-НАД	100%	5.00E-78	44%	НАД киназа 1 ppnk
CSTER	1425	1333019	1333990	+	рибозо-фосфат пирофосфокиназа	100%	4.00E-113	54%	рибозо-фосфат пирофосфокиназа prs
CSTER	1426	1333994	1335106	+	предп. цистеин десульфураса	98%	2.00E-115	47%	предп. цистеин десульфураса IscS
CSTER	1448	1355769	1356878	-	Сигма фактор РНК полимеразы RpoD	93%	6.00E-167	70%	Сигма фактор РНК полимеразы sigA
CSTER	1449	1356882	1358693	-	ДНК примаза	68%	9.00E-86	37%	ДНК примаза DnaG
CSTER	1451	1359076	1359252	-	30S рибосомальный белок S21	98%	3.00E-29	89%	
CSTER	1464	1371306	1372619	-	ГТФаза ObgE	100%	0.00E+00	67%	ГТФаза ObgE
CSTER	1480	1387053	1389005	-	субъединица В ДНК гиразы	100%	0.00E+00	68%	субъединица В ДНК гиразы

CSTER	1497	1401241	1402143	-	УДФ-N-ацетиленолпирувоилглюкоз амин редуктаза	98%	1.00E-66	40%	УДФ-N-ацетиленолпирувоилглюкоз амин редуктаза murB
CSTER	1506	1409996	1410268	-	30S рибосомальный белок S16	100%	5.00E-42	67%	
CSTER	1512	1415065	1416078	-	предп. липид киназа	95%	1.00E-96	51%	Диацилглицерин киназа dagK
CSTER	1513	1416080	1418034	-	НАД-зависимая ДНК лигаза	97%	0.00E+00	55%	ДНК лигаза LigA
CSTER	1516	1419659	1420519	-	methionine aminopeptidase	99%	6.00E-56	37%	
CSTER	1519	1422447	1423709	-	УДФ-N-ацетилглюкозамин 1- карбоксивинилтрансфераза	96%	6.00E-125	46%	УДФ-N-ацетилглюкозамин 1- карбоксивинилтрансфераза 1 muraa
CSTER	1522	1426828	1427583	-	1-ацил-sn-глицерин-3-фосфат ацилтрансфераза	92%	1.00E-29	32%	1-ацил-sn-глицерин-3-фосфат ацилтрансфераза plsC
CSTER	1534	1439747	1441120	-	УДФ-N-ацетилмурамоил-трипептид-D-аланил-D-аланил лигаза	98%	2.00E-98	38%	УДФ-N-ацетилмурамоил-трипептид-D-аланил-D-аланил лигаза murF
CSTER	1544	1447537	1448583	-	D-аланил-аланил синтеттаза A	98%	2.00E-95	42%	D-аланил--D-аланил лигаза ddl
CSTER	1583	1483475	1484107	-	Никотиновой кислоты мононуклеотид аденилилтрансфераза	98%	2.00E-62	44%	никотинат-нуклеотид аденилилтрансфераза nadd
CSTER	1584	1484211	1484528	-	предп. РНК-связывающий белок	95%	1.00E-24	50%	возможный РНК-связывающий белок YqeI
CSTER	1585	1484767	1485885	-	ГТФ-связывающий белок YqeH	100%	6.00E-163	59%	Неохарактеризованный белок YqeH
CSTER	1590	1488021	1489463	-	аспартил/глутамил-тРНК амидотрансфераза, субъединица B	99%	0.00E+00	63%	
CSTER	1591	1489463	1490929	-	аспартил/глутамил-тРНК амидотрансфераза, субъединица A	99%	0.00E+00	59%	
CSTER	1592	1490929	1491231	-	аспартил/глутамил-тРНК амидотрансфераза, субъединица C	97%	1.00E-18	41%	
CSTER	1615	1511378	1512298	-	Тиоредоксин редуктаза	95%	3.00E-126	58%	Тиоредоксин редуктаза trxB
CSTER	1665	1554189	1555211	-	фосфо-N-ацетилмурамоил-пентапептид-трансфераза	92%	1.00E-66	41%	фосфо-N-ацетилмурамоил-пентапептид-трансфераза mraY
CSTER	1666	1555213	1557480	-	предп. пенициллин-связывающий белок 2X	90%	9.00E-88	32%	пенициллин-связывающий белок 2B
CSTER	1667	1557484	1557804	-	Белок клеточного деления FtsL				
CSTER	1701	1591284	1592387	-	Аланин рацемаза	93%	1.00E-98	43%	Аланин рацемаза 1 alr
CSTER	1702	1592408	1592767	-	предп. фосфопантетейнил-4- трансфераза	97%	6.00E-29	41%	голо- [ацил-переносящий-белок] синтаза
CSTER	1705	1594955	1597504	-	Субъединица препротин транслоказы SecA	100%	0.00E+00	55%	Субъединица препротин транслоказы SecA
CSTER	1726	1616337	1616576	-	30S рибосомальный белок S18	100%	5.00E-32	65%	
CSTER	1727	1616618	1617136	-	Белок, связывающий одонитевую ДНК	93%	1.00E-41	56%	Белок B, связывающий одонитевую ДНК
CSTER	1728	1617148	1617438	-	30S рибосомальный белок S6	96%	4.00E-40	61%	
CSTER	1745	1633808	1634821	-	предп. O-сиалогликопротеин эндопептидаза	97%	6.00E-130	55%	тРНК N6-аденозин треонилкарбамоилтрансфераза tsad
CSTER	1747	1635236	1635922	-	предп. гликопротеин эндопептидаза	100%	6.00E-44	38%	белок TsaB биосинтеза тРНК треонилкарбамоиладенозин
CSTER	1749	1636421	1638103	+	мРНК деградации рибонуклеазы J1/J2	99%	0.00E+00	62%	рибонуклеаза J1 yqkC
CSTER	1755	1642501	1643700	-	Фосфоглицерат киназа	100%	2.00E-124	50%	Фосфоглицерат киназа pgk
CSTER	1762	1648240	1650321	-	Фактор элонгации транскрипции G	100%	0.00E+00	77%	
CSTER	1763	1650542	1651012	-	30S рибосомальный белок S7	100%	1.00E-86	75%	
CSTER	1764	1651031	1651444	-	30S рибосомальный белок S12	98%	3.00E-80	86%	
CSTER	1770	1656078	1656950	-	рибосома-ассоциированная ГТФаза	99%	5.00E-105	51%	предполагаемая ГТФаза рибосомного биогенеза RsgA
CSTER	1776	1660099	1660413	-	предп. тиоредоксин	92%	1.00E-37	61%	тиоредоксин trxB
CSTER	1787	1670058	1670192	-	50S рибосомальный белок L34	100%	1.00E-15	70%	

CSTER	1790	1673060	1673395	-	рибонуклеаза Р	95%	1.00E-31	47%	Белковый компонент рибонуклеаза Р
CSTER	1793	1677296	1678750	-	глутамил-ТРНК синтетаза	100%	0.00E+00	55%	
CSTER	1797	1681106	1681795	-	50S рибосомальный белок L1	98%	4.00E-99	65%	50S рибосомальный белок L1
CSTER	1809	1690036	1691058	+	глицерин-3-фосфат дегидрогеназа	94%	4.00E-122	53%	глицерин-3-фосфат дегидрогеназа [NAD(P)+] gsrA
CSTER	1813	1693298	1694431	-	металл-зависимая амдаза/аминоацилаза/карбоксипептидаза	95%	6.00E-120	47%	N-ацетилдиаминопимелат диацетилаза ukur
CSTER	1814	1694510	1695208	-	предп. 2,3,4,5-тетрагидропиридин-2-карбоксилат N-сукцинилтрансфераза	98%	4.00E-86	60%	2,3,4,5-тетрагидропиридин-2,6-дикарбоксилат N-ацетилтрансфераза dapH
CSTER	1821	1699209	1699601	-	Белок, связывающий сигнальную нить	93%	1.00E-25	40%	Белок В, связывающий сигнальную нить ДНК
CSTER	1844	1713286	1716924	-	ДНК-зависимая РНК полимераза, бета субъединица	97%	0.00E+00	68%	ДНК-направленная РНК полимераза, бета субъединица
CSTER	1845	1717025	1720606	-	ДНК-зависимая РНК полимераза, бета субъединица	98%	0.00E+00	71%	ДНК-направленная РНК полимераза, бета субъединица
CSTER	1847	1723493	1724749	+	тирозил-ТРНК синтетаза	99%	0.00E+00	58%	
CSTER	1876	1751169	1752050	-	фруктозо-бифосфат альдолаза	100%	4.00E-83	45%	возможная фруктозо-бифосфат альдолаза fbaa
CSTER	1880	1755102	1755488	-	50S рибосомальный белок L17	100%	2.00E-62	73%	
CSTER	1881	1755506	1756444	-	ДНК-зависимая РНК полимераза, альфа субъединица	99%	7.00E-139	62%	ДНК-направленная РНК полимераза, альфа субъединица
CSTER	1882	1756493	1756876	-	30S рибосомальный белок S11	93%	7.00E-69	87%	
CSTER	1883	1756904	1757269	-	30S рибосомальный белок S13	100%	2.00E-58	73%	
CSTER	1884	1757290	1757403	-	50S рибосомальный белок L36	100%	4.00E-16	84%	
CSTER	1885	1757429	1757647	-	Фактор инициации трансляции 1 (IF-1)	100%	1.00E-38	78%	
CSTER	1886	1757765	1758421	-	Аденилат киназа	99%	3.00E-89	59%	Аденилат киназа adk
CSTER	1887	1758553	1759848	-	предп. субъединица транслоказы, SecY	99%	4.00E-140	49%	субъединица транслоказы SecY
CSTER	1888	1759865	1760305	-	50S рибосомальный белок L15	100%	1.00E-70	72%	
CSTER	1889	1760433	1760615	-	50S рибосомальный белок L30	98%	3.00E-20	62%	
CSTER	1890	1760630	1761124	-	30S рибосомальный белок S5	93%	8.00E-70	76%	
CSTER	1891	1761143	1761499	-	50S рибосомальный белок L18	100%	4.00E-55	73%	
CSTER	1892	1761589	1762125	-	50S рибосомальный белок L6	100%	3.00E-69	60%	50S рибосомальный белок L6
CSTER	1893	1762252	1762650	-	30S рибосомальный белок S8	100%	1.00E-73	77%	
CSTER	1894	1762773	1762958	-	30S рибосомальный белок S14	100%	1.00E-31	77%	
CSTER	1895	1762976	1763518	-	50S рибосомальный белок L5	99%	3.00E-100	77%	
CSTER	1896	1763545	1763850	-	50S рибосомальный белок L24	99%	7.00E-38	64%	
CSTER	1897	1763931	1764299	-	50S рибосомальный белок L14	100%	5.00E-74	87%	
CSTER	1898	1764324	1764584	-	30S рибосомальный белок S17	97%	4.00E-46	84%	
CSTER	1899	1764612	1764818	-	50S рибосомальный белок L29	78%	1.00E-14	58%	
CSTER	1900	1764828	1765241	-	50S рибосомальный белок L16	93%	1.00E-76	83%	
CSTER	1901	1765245	1765898	-	30S рибосомальный белок S3	100%	1.00E-119	75%	
CSTER	1902	1765911	1766255	-	50S рибосомальный белок L22	96%	7.00E-49	64%	
CSTER	1903	1766271	1766549	-	30S рибосомальный белок S19	100%	2.00E-54	83%	
CSTER	1904	1766643	1767476	-	50S рибосомальный белок L2	100%	5.00E-159	76%	50S рибосомальный белок L2
CSTER	1905	1767494	1767790	-	50S рибосомальный белок L23	94%	2.00E-31	60%	
CSTER	1906	1767790	1768413	-	50S рибосомальный белок L4	100%	9.00E-92	60%	50S рибосомальный белок L4

CSTER	1907	1768438	1769064	-	50S рибосомальный белок L3	99%	4.00E-115	75%	50S рибосомальный белок L3
CSTER	1908	1769181	1769489	-	30S рибосомальный белок S10	100%	1.00E-58	78%	
CSTER	1936	1795296	1795484	+	50S рибосомальный белок L28	100%	7.00E-26	69%	
CSTER	1948	1805428	1807179	-	аспартил-тРНК синтетаза	98%	0.00E+00	57%	
CSTER	1950	1807719	1808999	-	гистидил-тРНК синтетаза	99%	1.00E-140	49%	
CSTER	1953	1810651	1810833	+	50S рибосомальный белок L32	91%	1.00E-10	48%	
CSTER	1954	1810849	1810998	+	30S рибосомальный белок S4	100%	1.00E-14	55%	
CSTER	1973	1823718	1824329	-	Репликативной ДНК геликазы	100%	1.00E-95	69%	
CSTER	1975	1824852	1826213	-	50S рибосомальный белок L9	95%	8.00E-176	56%	Репликативной ДНК геликазы DnaC
CSTER	1976	1826257	1826715	-	тРНК (5-метиламнометил-2-тиоуридилат)-метилтрансфераза	100%	2.00E-39	55%	
CSTER	1979	1830931	1832052	-		98%	9.00E-177	70%	тРНК-специфическая т-тиоуридилаза MnmA
CSTER	1986	1838183	1838725	-	фосфатидилглицерофосфат синтаза	93%	2.00E-45	52%	CDP-диацилглицеро-глицерин-3-фосфат фосфатидилтрансфераза pgsA
CSTER	1992	1843910	1845391	-	инозин-5'-монофосфат дегидрогеназа	99%	0.00E+00	68%	инозин-5'-монофосфат дегидрогеназа guab
CSTER	1993	1845568	1846590	-	триптофанил-тРНК синтетаза II	98%	8.00E-42	35%	

Таблица 5

Список дифференциально экспрессируемых генов, идентифицированных в штамме GEI4, содержащем делеции

Оперон	Предполагаемый путь/функция	Регулятор-эффектор	Ген	log2 изменение	p-значение
STER_0390-STER_0393	Метаболизм цистеина	CmbR/Homr/Mtar	STER_0390	-1.1	2.7E-23
			STER_0391	-1.3	2.2E-22
			STER_0392	-1.4	1.2E-32
			STER_0393	-1.2	2.8E-15
STER_1016-STER_1017	Метаболизм мальтодекстрина		STER_1016	-1.0	0
			STER_1017	-0.8	7.50E-38
STER_1548-STER_1555	Биосинтез ароматических аминокислот	T-box RNA (Trp)	STER_1548	-1.3	2.3E-37
			STER_1549	-1.0	1.5E-26
			STER_1550	-1.3	7.9E-38
			STER_1551	-1.3	1.2E-35
			STER_1552	-1.3	0
			STER_1553	-1.4	0
			STER_1554	-1.3	0
			STER_1555	-1.6	7.0E-73
STER_1960-STER_1963	Мембранные белки		STER_1960	-0.9	0
			STER_1961	-1.0	0
			STER_1962	-1.0	8.3E-317
			STER_1963	-1.1	3.2E-965
STER_0049-STER_0054	Биосинтез пуринов	PurR	STER_0049	1.2	5.6E-34
			STER_0050	1.1	1.4E-29
			STER_0051	1.1	0
			STER_0052	1.1	0
			STER_0053	1.1	0
			STER_0054	1.1	6.6E-40
STER_0699-STER_0701	Метаболизм этаноламина		STER_0699	1.6	0
			STER_0700	2.1	9.5E-200
			STER_0701	2.2	7.8E-329
STER_1020-STER_1024	Твин аргинин транслоказа	TatA	STER_1020	1.3	5.3E-14
			STER_1021	1.3	8.8E-12
			STER_1022	1.4	3.3E-12
			STER_1023	1.0	9.6E-6
			STER_1024	0.9	0.0022
STER_1025-STER_1028	Гомеостаз железа	PerR	STER_1025	1.1	1.9E-09
			STER_1026	0.9	2.5E-6
			STER_1027	1.2	3E-12
			STER_1028	1.2	1.5E-17
STER_1405-STER_1409	Транспорт с участием ABC белков	Cody	STER_1405	1.6	9.5E-90
			STER_1406	1.6	0
			STER_1407	1.6	3.7E-59
			STER_1408	1.5	0
			STER_1409	1.4	4.2E-54
STER_1821-STER_1823	Стресс		STER_1821	2.0	0
			STER_1822	2.2	5.4E-98
			STER_1823	2.0	0

Литература

1. Darmon E, Leach DF. (2014) Bacterial Genome Instability. *Microbiol. Mol. Biol. Rev.* 78, 1-39.
2. Labrie SJ, Samson JE, and Moineau S. (2010) Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* 8, 317-327.
3. Barrangou R, Marraffini LA (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol Cell* 54(2):234-244.
4. Barrangou R, et al. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315(5819):1709-1712.
5. Brouns SJJ, et al. (2008) Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321(5891):960-964.
6. Young JC et al. (2012) Phage-induced expression of CRISPR-associated proteins is revealed by shotgun proteomics in *Streptococcus thermophilus*. *PLoS ONE* 7(5):e38077.
7. Garneau JE, et al. (2010) The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468(7320):67-71.
8. Groenen PM, Bunschoten AE, Van Soolingen D, Van Embden JD (1993) Nature of DNA polymorphism in the direct repeat cluster of *Mycobacterium tuberculosis*; application for strain differentiation by a novel typing method. *Mol Microbiol* 10(5):1057-1065.
9. Yin S, et al. (2013) The evolutionary divergence of Shiga toxin-producing *Escherichia coli* is reflected in clustered regularly interspaced short palindromic repeat (CRISPR) spacer composition. *Appl Environ Microbiol* 79(18):5710-5720.
10. Liu F, et al. (2011) Novel virulence gene and clustered regularly interspaced short palindromic repeat (CRISPR) multilocus sequence typing scheme for subtyping of the major serovars of *Salmonella enterica* subsp. *enterica*. *Appl Environ Microbiol* 77(6):1946-1956.

11. Barrangou R, Horvath P (2012) CRISPR: new horizons in phage resistance and strain identification. *Annu Rev Food Sci Technol* 3, 143-162.
12. Sander JD, and Joung JK. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat. Biotechnol.* 32, 347-355.
13. Bikard D, et al. (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* 41(15):7429-7437
14. Qi LS, et al. (2013) Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell* 152(5):1173-1183.
15. Gomaa AA, et al. (2014) Programmable Removal of Bacterial Strains by Use of Genome-Targeting CRISPR-Cas Systems. *mBio* 5(1):e00928-13.
16. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169(12):5429-5433.
17. Jansen R, Embden JDA van, Gaastra W, Schouls LM (2002) Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* 43(6):1565-1575.
18. Bolotin A, et al. (2004) Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat Biotechnol* 22(12):1554-1558.
19. Horvath P, et al. (2008) Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1401-1412.
20. Bolotin A, Quinguis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151(Pt 8):2551-2561.
21. Deveau H, et al. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J Bacteriol* 190(4):1390-1400.

22. Paez-Espino D, et al. (2013) Strong bias in the bacterial CRISPR elements that confer immunity to phage. *Nat Commun* 4, 1430.
23. Sun CL, et al. (2013) Phage mutations in response to CRISPR diversification in a bacterial population. *Environ Microbiol* 15(2):463-470.
24. Sapranaukas R, et al. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res* 39(21):9275-9282.
25. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* 109(39):E2579-2586.
26. Briner AE, et al. (2014) Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. *Molecular Cell*. 56(2):333-339
27. Bondy-Denomy J, Davidson AR. (2014) To acquire or resist: the complex biological effects of CRISPR-Cas systems. *Trends Microbiol.* 22, 218-225.
28. Horvath P, et al. (2009) Comparative analysis of CRISPR loci in lactic acid bacteria genomes. *Int J Food Microbiol* 131(1):62-70.
29. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. (2013a) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31(3):233-239.
30. Vercoe RB, et al. (2013) Cytotoxic chromosomal targeting by CRISPR/Cas systems can reshape bacterial genomes and expel or remodel pathogenicity islands. *PLoS Genet* 9(4):e1003454.
31. Oh JH, van Pijkeren JP. (2014) CRISPR-Cas9-assisted recombineering in *Lactobacillus reuteri*. *Nucleic Acids Res* 10.1093/nar/gku623
32. Selle K, Barrangou R. (2015) Harnessing CRISPR-Cas systems for bacterial genome editing. *Trends Microbiol.* In press
33. Kobayashi K, et al. (2003) Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4678-4683.

34. Mahillon J, Chandler M. (1998) Insertion sequences. *Microbiol Mol Biol Rev* 62(3):725-774.
35. Goh YJ, Goin C, O'Flaherty S, Altermann E, Hutkins R (2011) Specialized adaptation of a lactic acid bacterium to the milk environment; the comparative genomics of *Streptococcus thermophilus* LMD-9. *Microbial Cell Factories* 10 (Suppl 1):S22
36. Dandoy D, et al. (2011) The fast milk acidifying phenotype of *Streptococcus thermophilus* can be acquired by natural transformation of the genomic island encoding the cell-envelope proteinase PrtS. *Microb. Cell Fact.* 10 Suppl 1, S21.
37. Deltcheva E, et al. (2011) CRISPR RNA maturation by trans-activating small RNA and host factor RNase III. *Nature* 471, 602-607.
38. Aravind L, Koonin EV (2001) Prokaryotic homologs of the eukaryotic DNA-end-binding protein Ku, novel domains in the Ku protein and prediction of a prokaryotic double-strand break repair system. *Genome Res* 11(8):1365-1374.
39. Koonin EV, Makarova KS (2007) Evolutionary genomics of lactic acid bacteria. *J Bacteriol* 189(4):1199-1208
40. Makarova KS, et al. (2006) Comparative genomics of the lactic acid bacteria. *Proc Natl Acad Sci USA* 103(42):15611-15616.
41. Goh YJ, et al. (2009) Development and application of a upp-based counterselective gene replacement system for the study of the S-layer protein SlpX of *Lactobacillus acidophilus* NCFM. *Appl Environ Microbiol* 75(10):3093-3105.
42. Horton RM, Hunt HD, Ho SN, Pullen JK, Pease LR (1989) Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene* 77(1):61-68.
43. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Almendros C (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* 155(Pt 3):733-740.
44. Russell WM, Klaenhammer TR (2001) Efficient System for Directed Integration into the *Lactobacillus acidophilus* and *Lactobacillus gasseri* Chromosomes via Homologous Recombination. *Appl Environ Microbiol* 67(9):4361-4364.
45. Wei M-Q, et al. (1995) An improved method for the transformation of *Lactobacillus* strains using electroporation. *Journal of Microbiological Methods* 21(1):97-109.
46. Zhang X, Bremer H (1995) Control of the *Escherichia coli* rrnB pl promoter strength by ppGpp. *Journal of Biological Chemistry* 270(19):11181-11189.
47. Law J, et al. (1995) A system to generate chromosomal mutations in *Lactococcus lactis* which allows fast analysis of targeted genes. *J Bacteriol* 177(24):7011-7018.

Приведенное выше описание иллюстрирует настоящее изобретение и не должно быть истолковано как ограничивающее его каким-либо образом. Изобретение определяется приведенной ниже формулой изобретения, которая охватывает все ее эквиваленты.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ уничтожения одной или более бактериальных клеток в популяции бактериальных клеток, включающий

введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей

(a) CRISPR-Cas систему, где CRISPR-Cas система представляет собой CRISPR-Cas систему типа I, где CRISPR-Cas система типа I содержит

(i) конструкцию гетерологичной нуклеиновой кислоты, кодирующей по меньшей мере один каскадный полипептид типа I, и

(ii) полипептид Cas3, Cas3' и Cas3" или конструкцию гетерологичной нуклеиновой кислоты, кодирующей полипептид Cas3, Cas3' и Cas3"; и

(b) массив CRISPR типа I, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, где спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, тем самым уничтожая одну или более бактериальных клеток, которые содержат область-мишень в популяции бактериальных клеток.

2. Способ по п.1, в котором область-мишень находится в жизненно важном гене или нежизненно важном гене.

3. Способ по п.1 или 2, в котором последовательность повтор-спейсер-повтор или по меньшей мере одна последовательность повтор-спейсер содержит повтор, который идентичен повтору из массива CRISPR типа I дикого типа.

4. Способ по любому из пп.1-3, в котором область-мишень выбрана из любых по меньшей мере 10 последовательных нуклеотидов, расположенных рядом с последовательностью PAM в одной или нескольких бактериальных клетках.

5. Способ по любому из предшествующих пунктов, в котором гид CRISPR функционально связан с промотором.

6. Способ по любому из предшествующих пунктов, в котором по меньшей мере одна бактериальная клетка в популяции не содержит область-мишень и, следовательно, по меньшей мере одна бактериальная клетка не уничтожается при введении CRISPR-Cas системы.

7. Способ по любому из предшествующих пунктов, в котором спейсер указанной последовательности повтор-спейсер-повтор или последовательности повтор-спейсер содержит нуклеотидную последовательность, которая на 100% комплементарна области-мишени.

8. Способ по любому из предшествующих пунктов, в котором область-мишень выбрана из гена, открытой рамки считывания, или предполагаемой открытой рамки считывания, или межгенной области.

9. Фармацевтическая композиция для уничтожения одной или более бактериальных клеток в популяции бактериальных клеток, где фармацевтическая композиция содержит конструкцию гетерологичной нуклеиновой кислоты, содержащую

(a) CRISPR-Cas систему, где CRISPR-Cas система представляет собой CRISPR-Cas систему типа I, где CRISPR-Cas система типа I содержит

(i) конструкцию гетерологичной нуклеиновой кислоты, кодирующей по меньшей мере один каскадный полипептид типа I, и

(ii) полипептид Cas3, Cas3' и Cas3" или конструкцию гетерологичной нуклеиновой кислоты, кодирующей полипептид Cas3, Cas3' и Cas3"; и

(b) массив CRISPR типа I, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, где спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме бактериальных клеток указанной популяции, где область-мишень представляет собой генетическую последовательность одной или более бактериальных клеток, где массив CRISPR представляет собой массив CRISPR типа I,

где фармацевтическую композицию вводят в популяцию бактериальных клеток и, таким образом, одна или несколько бактериальных клеток, которые содержат область-мишень в популяции бактериальных клеток, уничтожаются.

10. Фармацевтическая композиция по п.9, в которой последовательность повтор-спейсер-повтор или по меньшей мере одна последовательность повтор-спейсер содержит повтор, который является идентичным повтору из массива CRISPR типа I дикого типа.

11. Фармацевтическая композиция по п.9 или 10, в которой массив CRISPR представляет собой массив CRISPR типа I, а последовательность повтор-спейсер-повтор или по меньшей мере одна последо-

вательность повтор-спейсер содержит повтор, который идентичен повтору из массива CRISPR типа I дикого типа, и CRISPR-Cas система включает

- (i) конструкцию гетерологичной нуклеиновой кислоты, кодирующей по меньшей мере один каскадный полипептид типа I, и
- (ii) полипептид Cas3 или Cas3' и Cas3" или конструкцию гетерологичной нуклеиновой кислоты, кодирующей Cas3 или Cas3' и Cas3" полипептид.

12. Способ отбора из популяции бактериальных клеток по меньшей мере одного изолята, который не содержит в своем геноме области-мишени, которая присутствует в геноме одной или нескольких бактериальных клеток в популяции, включающий введение в популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей

(a) CRISPR-Cas систему, где CRISPR-Cas система представляет собой CRISPR-Cas систему типа I, где CRISPR-Cas система типа I содержит

- (i) конструкцию гетерологичной нуклеиновой кислоты, кодирующей по меньшей мере один каскадный полипептид типа I, и
- (ii) полипептид Cas3, Cas3' и Cas3" или конструкцию гетерологичной нуклеиновой кислоты, кодирующей полипептид Cas3, Cas3' и Cas3"; и

(b) массив CRISPR типа I, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, где спейсер указанной последовательности повтор-спейсер-повтор или указанной по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в геноме одной или более бактериальных клеток в указанной популяции, тем самым отбирая из популяции бактериальных клеток по меньшей мере одного изолята без области-мишени в его геноме, где массив CRISPR представляет собой массив CRISPR типа I.

13. Способ по п.12, дополнительно включающий выращивание отдельных бактериальных колоний по меньшей мере одного изолята без области-мишени в его геноме.

14. Способ по п.12 или 13, в котором последовательность повтор-спейсер-повтор или по меньшей мере одна последовательность повтор-спейсер содержит повтор, который идентичен повтору из массива CRISPR типа I дикого типа.

15. Способ по любому из пп.12-14, в котором массив CRISPR представляет собой массив CRISPR типа I, а последовательность повтор-спейсер-повтор или по меньшей мере одна последовательность повтор-спейсер содержит повтор, который идентичен повтору из массива CRISPR типа I дикого типа, и CRISPR-Cas система содержит

- (i) конструкцию гетерологичной нуклеиновой кислоты, кодирующей по меньшей мере один каскадный полипептид типа I, и
- (ii) полипептид Cas3, Cas3' и Cas3" или конструкцию гетерологичной нуклеиновой кислоты, кодирующей полипептид Cas3, Cas3' и Cas3".

16. Применение конструкции гетерологичной нуклеиновой кислоты, содержащей CRISPR-Cas систему в направленном сокращении патогенов в случае медицинского вмешательства у пациента, причем медицинское вмешательство включает уничтожение специфического подмножества бактерий в смешанной популяции бактериальных клеток на основе различающегося генетического содержания в подмножестве бактерий посредством введения в смешанную популяцию бактериальных клеток конструкции гетерологичной нуклеиновой кислоты, содержащей

(a) CRISPR-Cas систему, где CRISPR-Cas система представляет собой CRISPR-Cas систему типа I, где CRISPR-Cas система типа I содержит

- (i) конструкцию гетерологичной нуклеиновой кислоты, кодирующей по меньшей мере один каскадный полипептид типа I, и
- (ii) полипептид Cas3, Cas3' и Cas3" или конструкцию гетерологичной нуклеиновой кислоты, кодирующей полипептид Cas3, Cas3' и Cas3"; и

(b) массив CRISPR типа I, содержащий (в 5'-3' направлении) последовательность повтор-спейсер-повтор или по меньшей мере одну последовательность повтор-спейсер, где спейсер указанной последовательности повтор-спейсер-повтор или по меньшей мере одной последовательности повтор-спейсер содержит нуклеотидную последовательность, которая по существу комплементарна области-мишени в генетической последовательности специфического подмножества бактерий, причем область-мишень имеет по меньшей мере 10 последовательных нуклеотидов, примыкающих к протоспейсеру мотива (PAM), распознаваемых CRISPR-Cas системой, тем самым уничтожая специфическое подмножество бактерий, которая включает область-мишень.

17. Применение по п.16, где по меньшей мере одна или несколько бактериальных клеток в популяции не содержат области-мишени и, следовательно, по меньшей мере одна бактериальная клетка не погибает при введении CRISPR-Cas системы.

18. Применение по п.16 или 17, где спейсер указанной последовательности повтор-спейсер-повтор или последовательности повтор-спейсер содержит нуклеотидную последовательность, которая на 100% комплементарна области-мишени.

19. Применение по любому из пп.16-18, где область-мишень находится в жизненно важном гене.
20. Применение по любому из пп.16-18, где область-мишень находится в нежизненно важном гене.
21. Применение по любому из пп.16-20, где последовательность повтор-спейсер-повтор или по меньшей мере одна последовательность повтор-спейсер массива CRISPR содержит повтор, который идентичен повтору из массива Cas3 CRISPR дикого типа.
22. Применение по любому из пп.16-21, где область-мишень выбирается из гена, открытой рамки считывания, или предполагаемой открытой рамки считывания, или межгенной области.

CRISPR1

Спейсер

Повтор

LacZ GTTTTGTACTCTCAAGATTTAAGTAAGTGTACAACAAATCCCATAGCAGCAACAATATAGTTGTTGTTTTGTACTCTCAAGATTTAAGTAAGTGTACAAC

CRISPR3

Повтор

Спейсер

Повтор

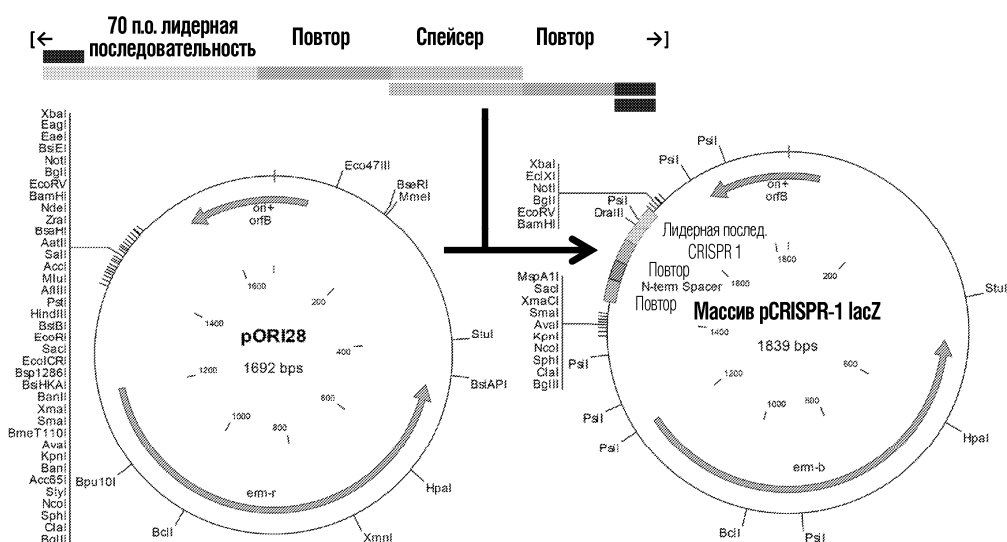
LacZ GTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAACAGTATTTTGAATCTCTTGAAGAATTTTCTGAGGTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAAC

ABC GTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAACGATAACACGAGATAAAACATCCAGCCACCGGTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAAC

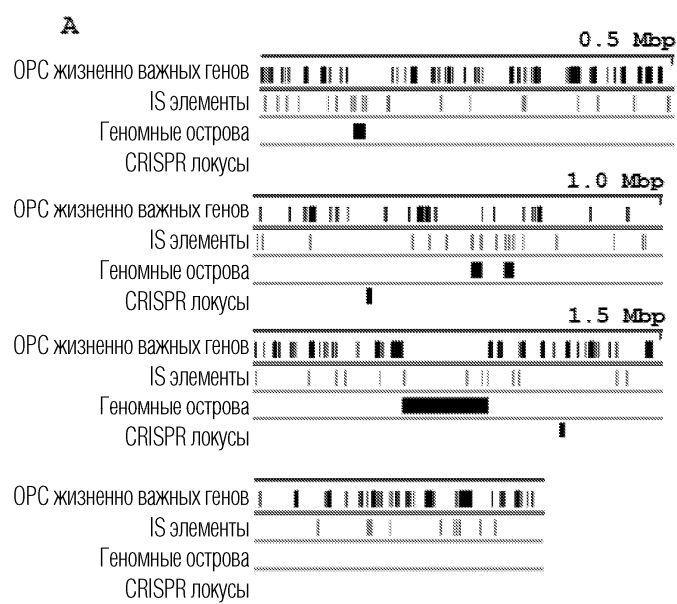
prtS GTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAACGTTGTAGCTTTGAGGTCTGAGAATACACCGGTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAAC

Cu GTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAACGATTGCTCAATCAATCGTTTCAGCTGCTAAGTTTTAGAGCTGTGTGTTTCGAATGGTTCCTAAAAC

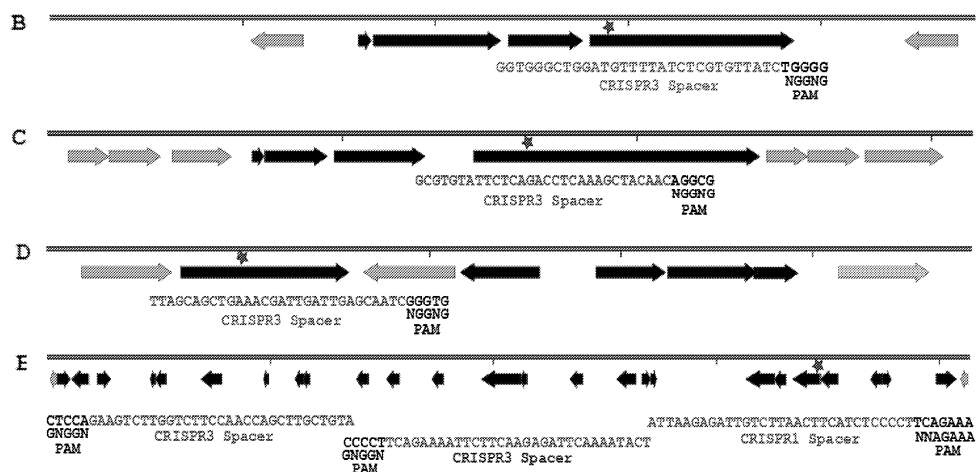
Фиг. 1



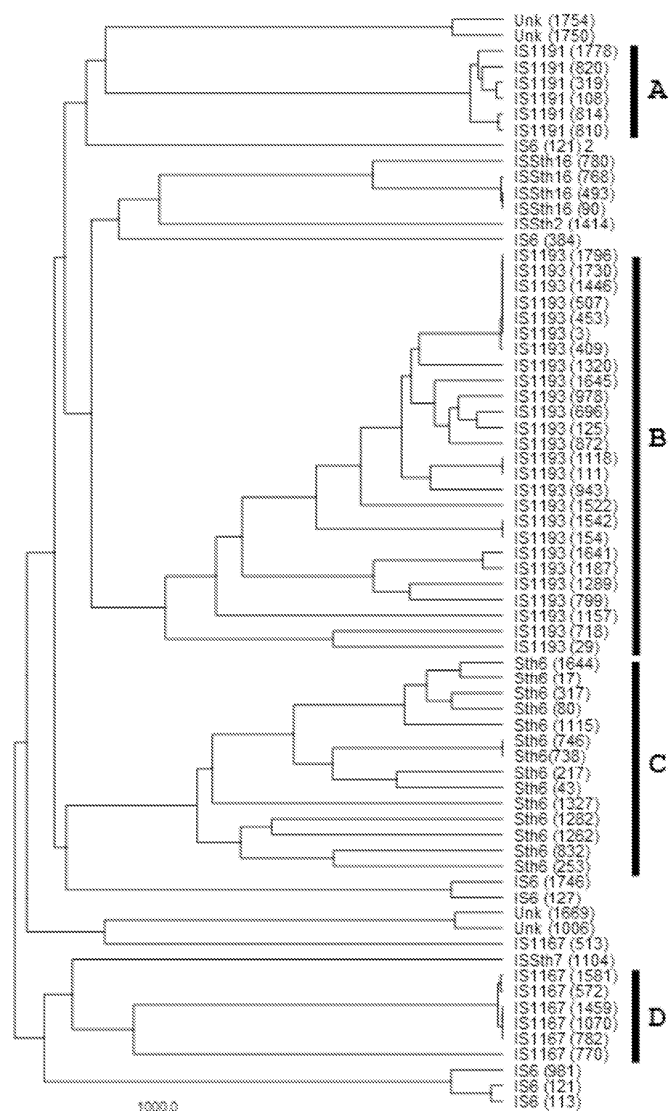
Фиг. 2



Фиг. 3

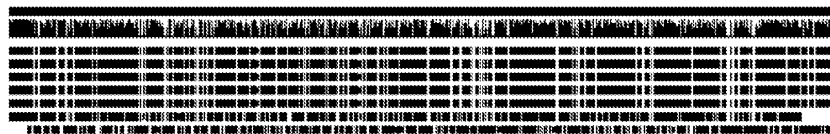


Фиг. 3 (продолжение)



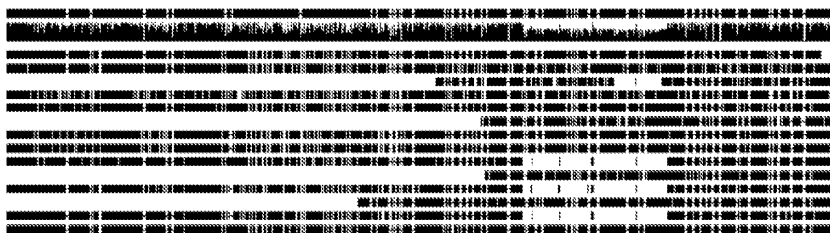
Фиг. 4

IS1191



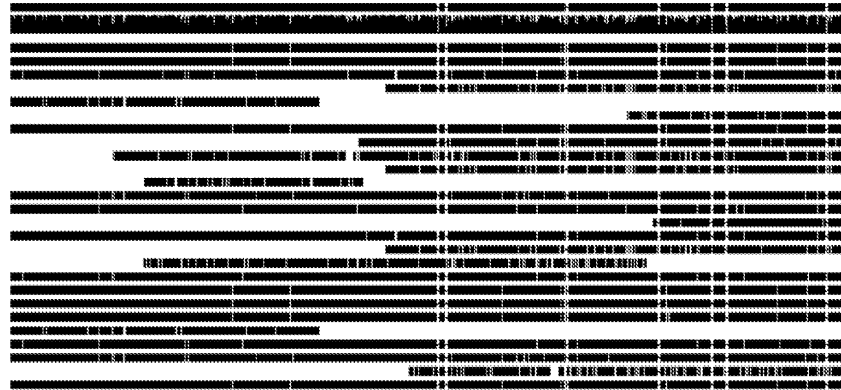
Фиг. 5A

Sth6



Фиг. 5B

IS1193

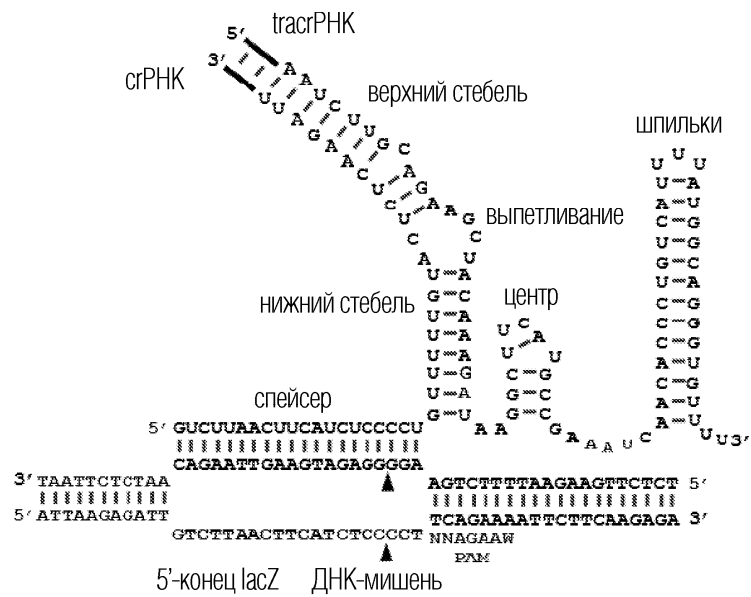


ФИГ. 5С

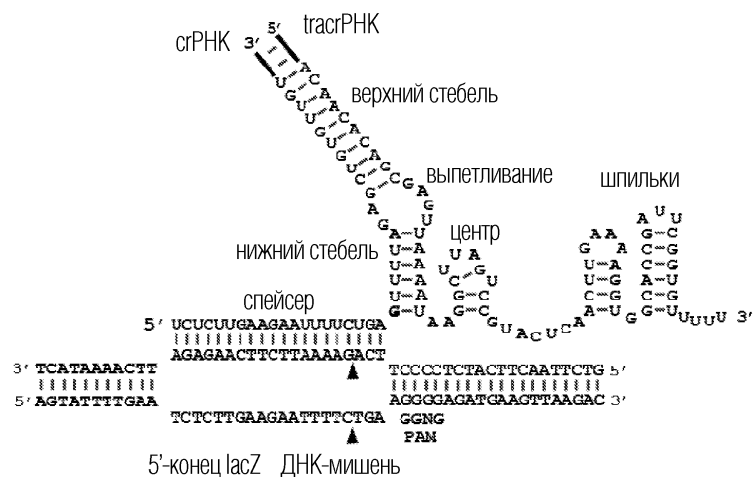
IS1167



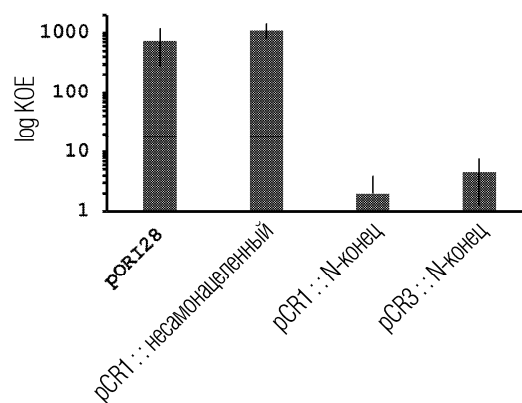
ФИГ. 5D



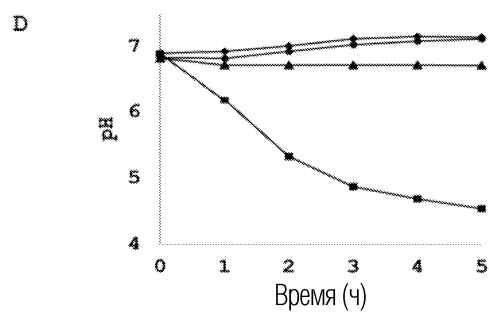
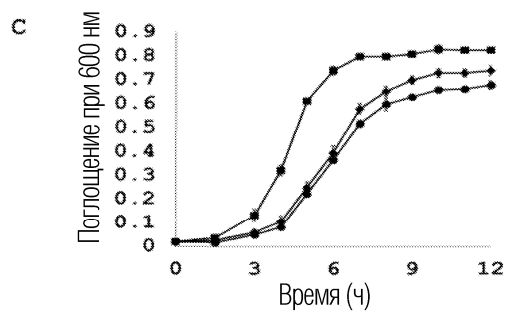
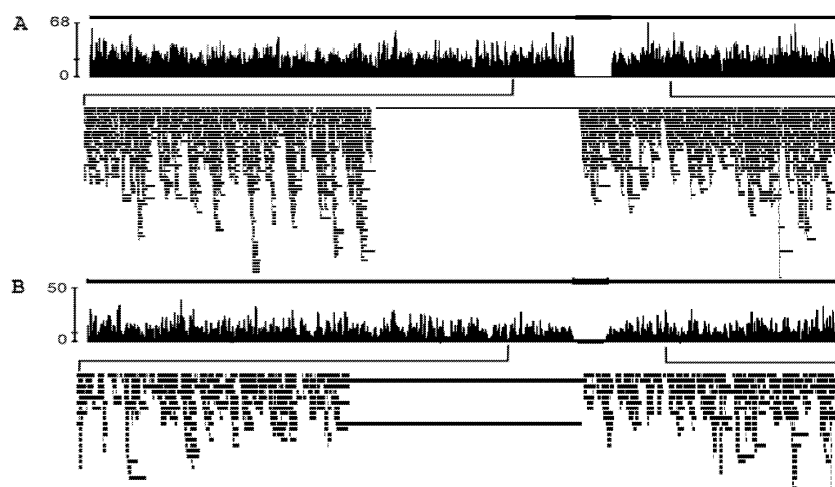
ФИГ. 6А



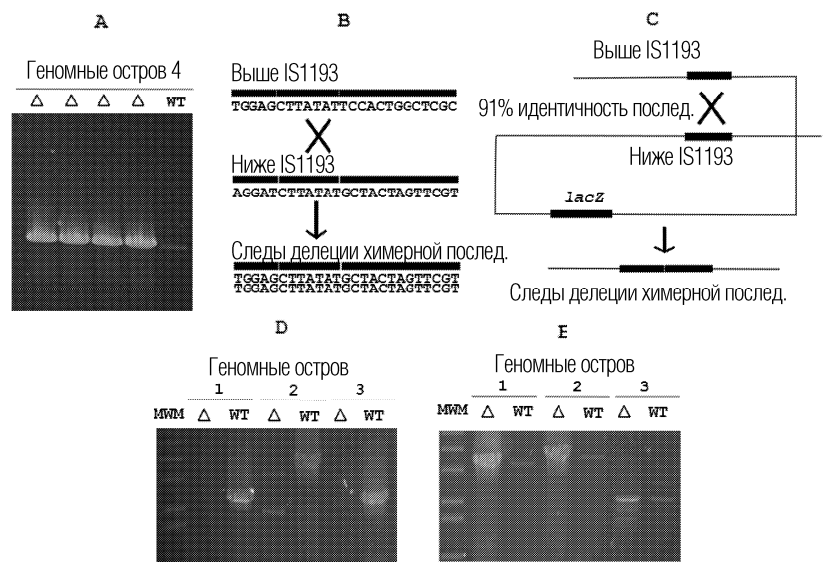
ФИГ. 6В



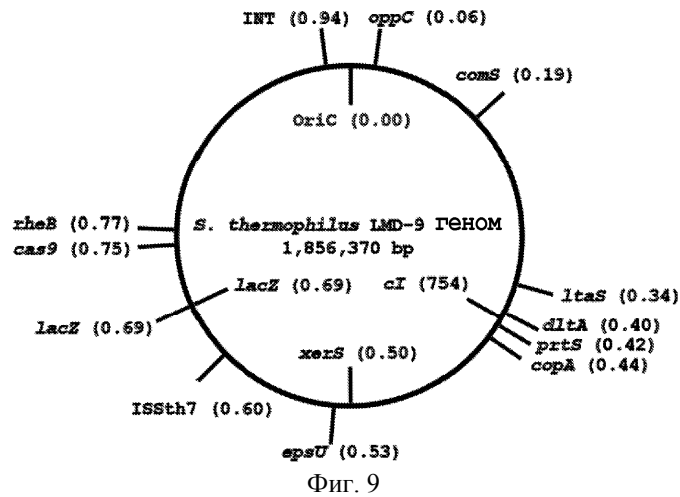
Фиг. 6С



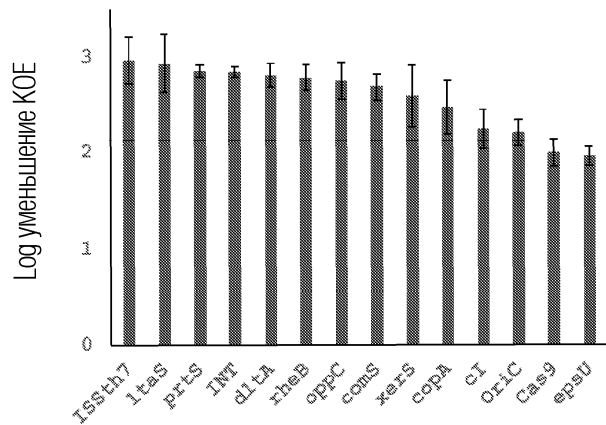
Фиг. 7



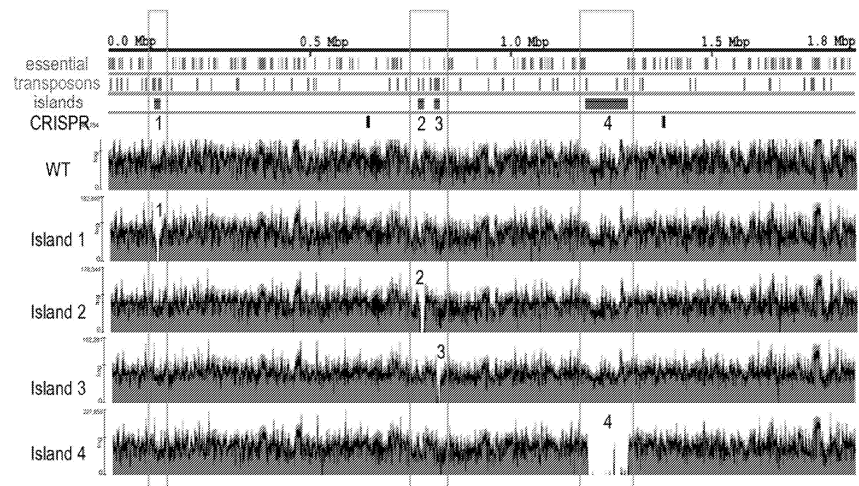
Фиг. 8



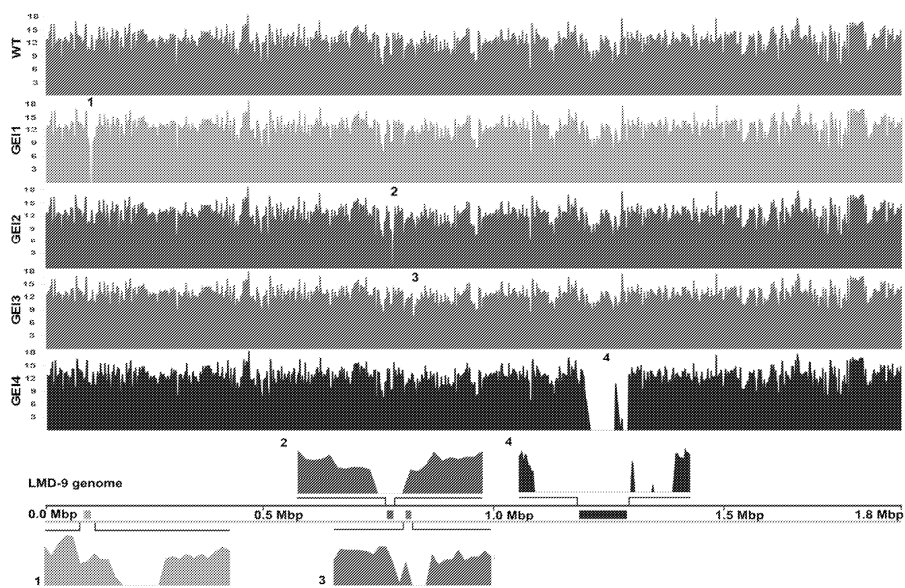
Фиг. 9



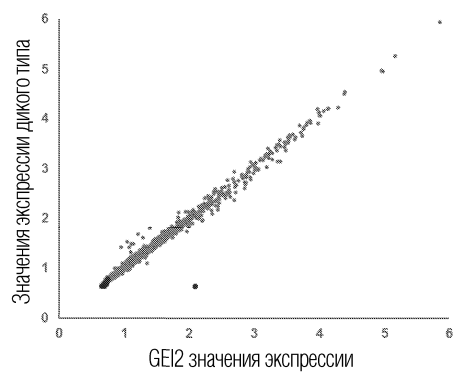
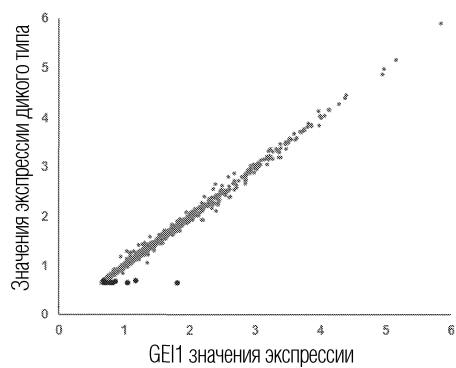
Фиг. 10



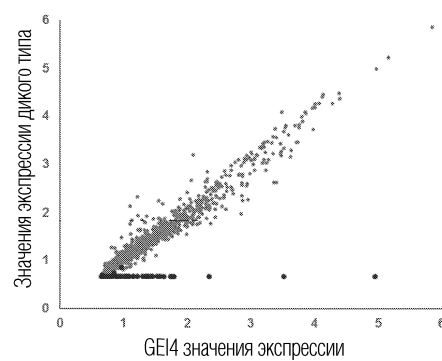
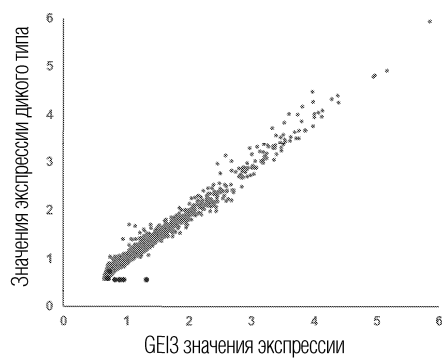
Фиг. 11



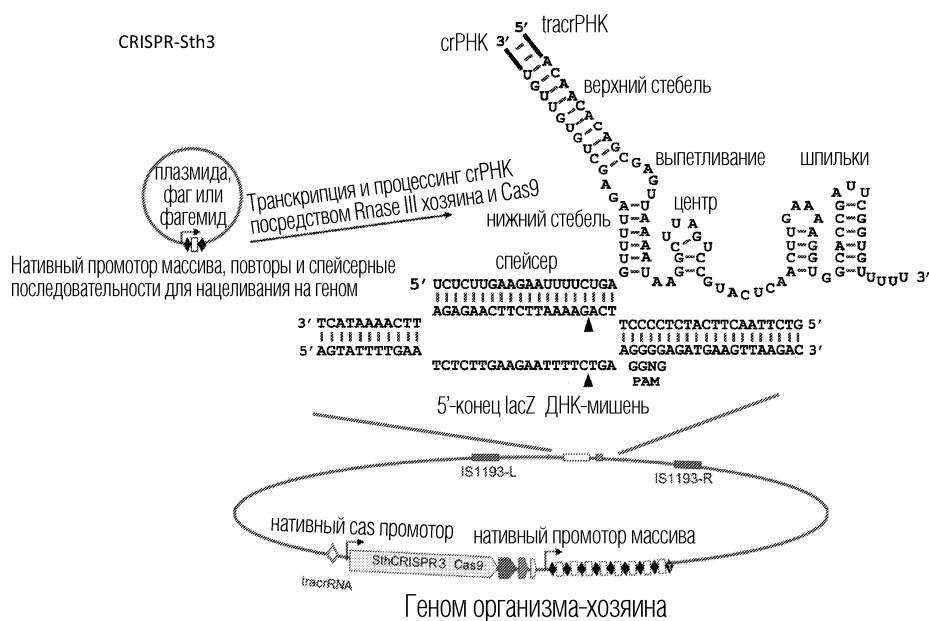
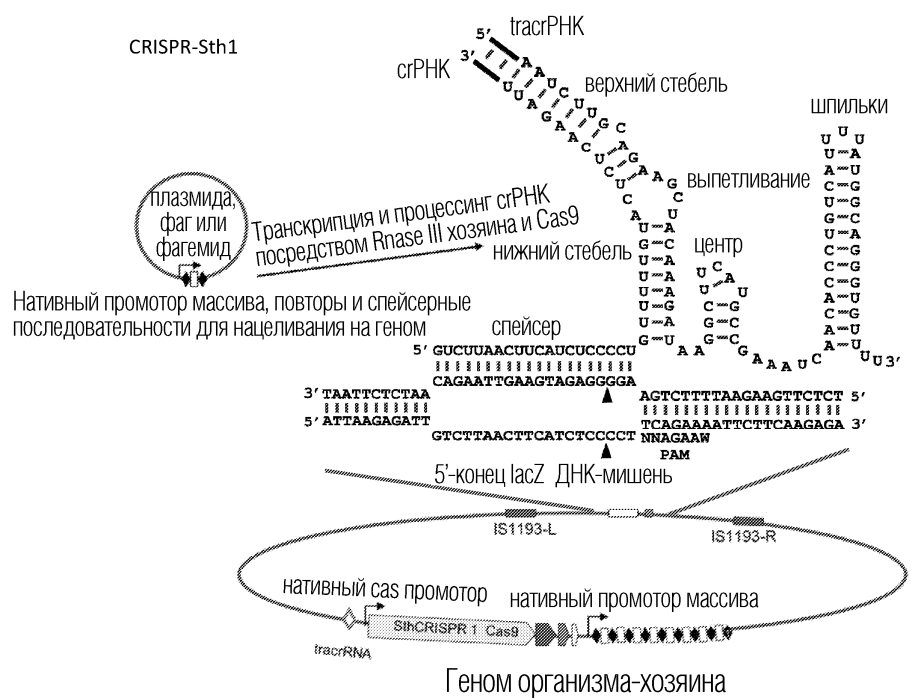
Фиг. 12

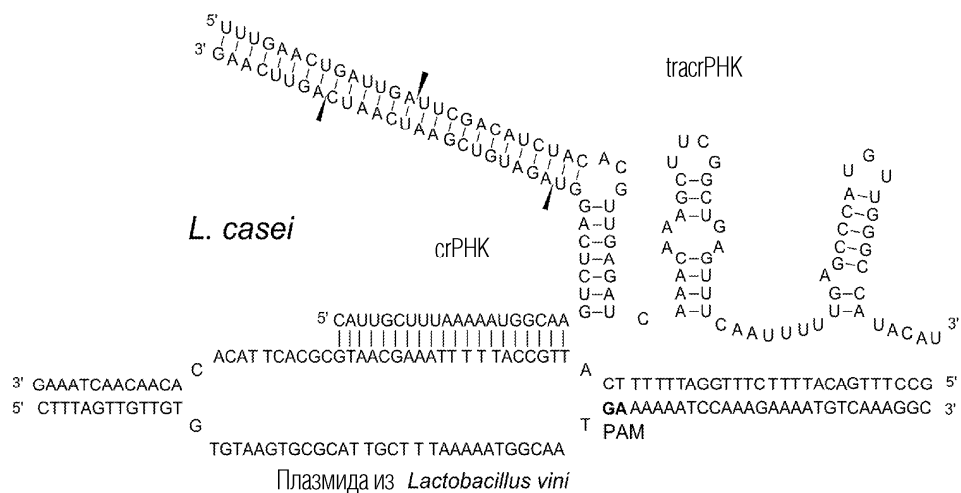
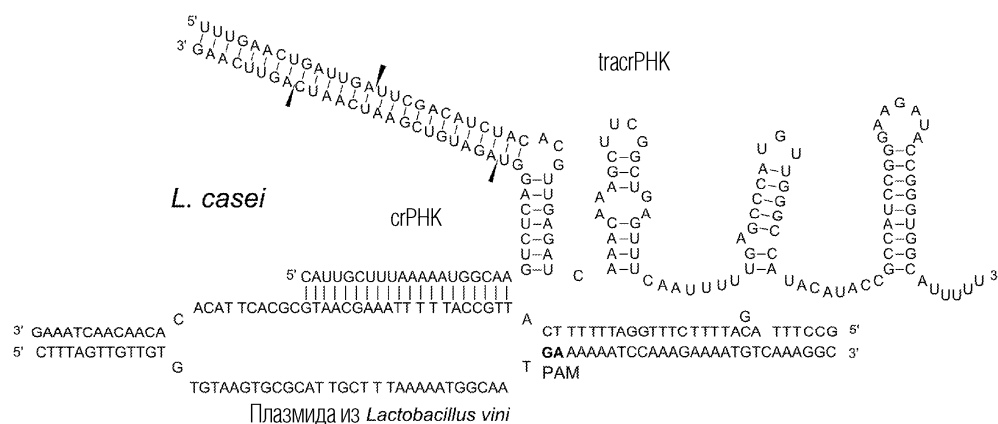


Фиг. 13

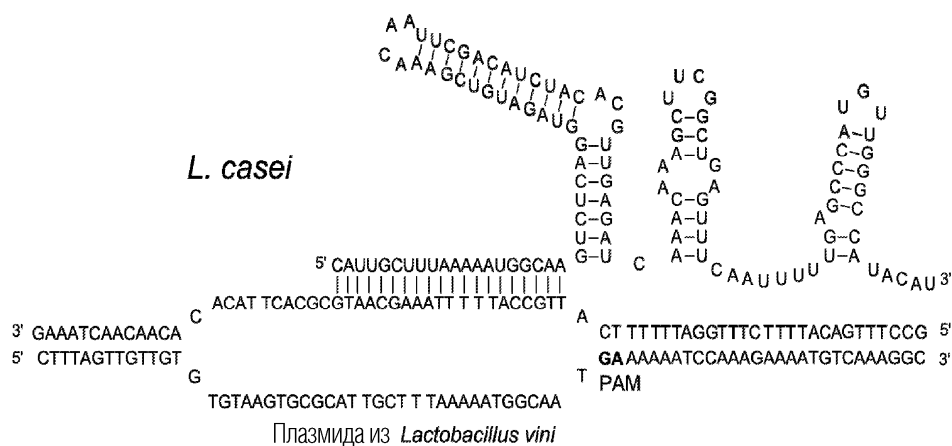


Фиг. 13 (продолжение)

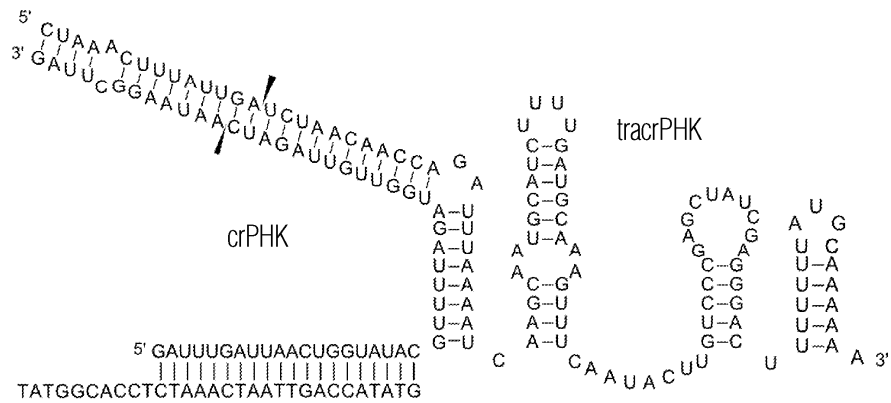
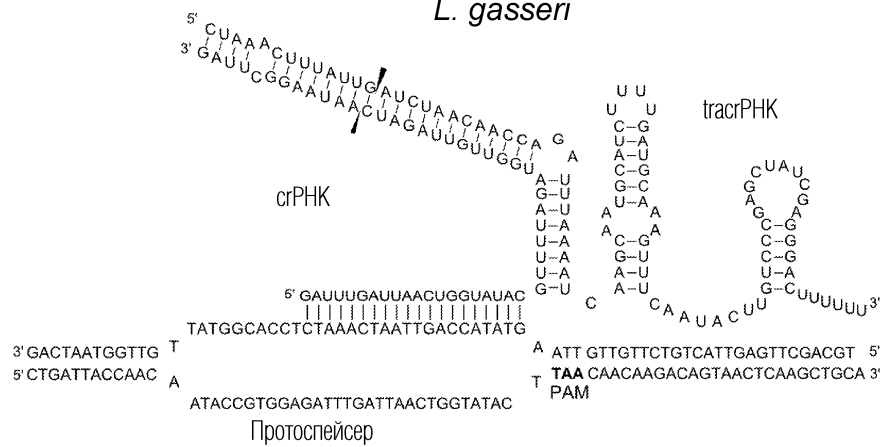




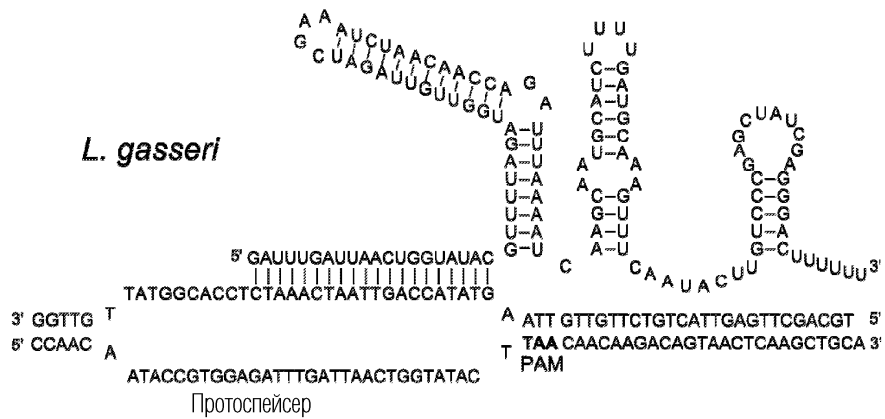
Фиг. 15



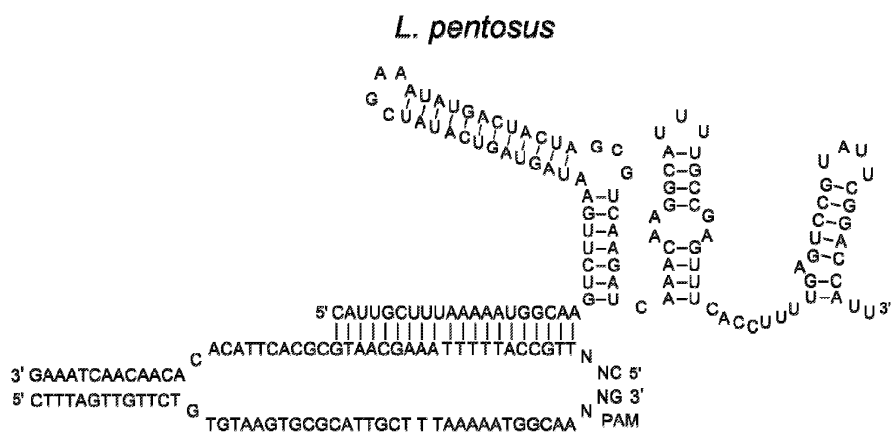
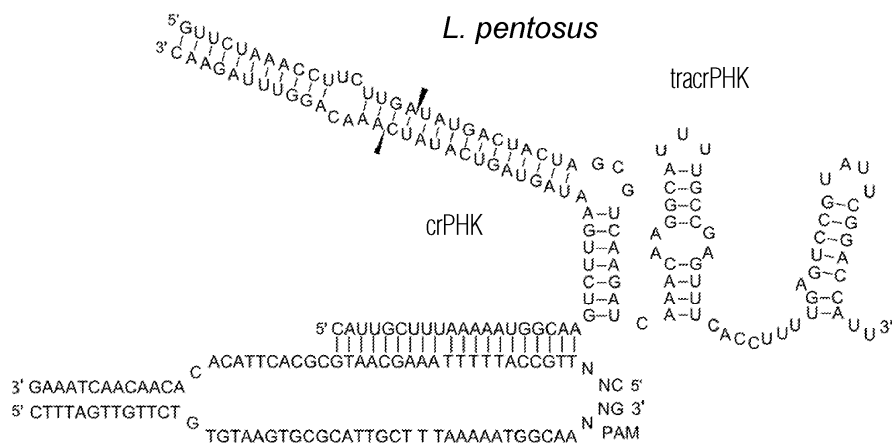
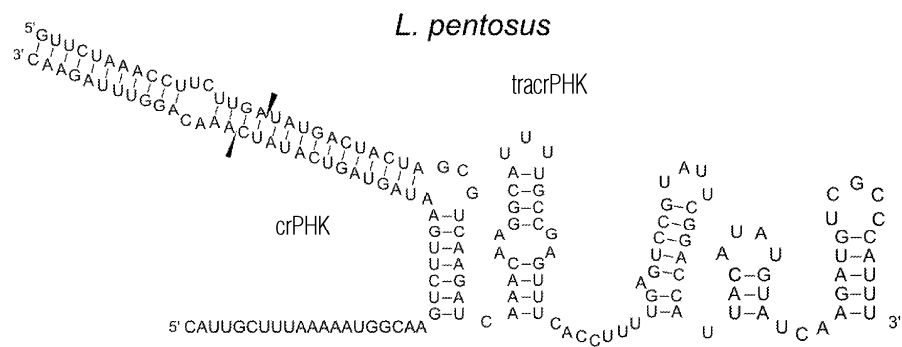
Фиг. 15 (продолжение)

L. gasseri*L. gasseri*

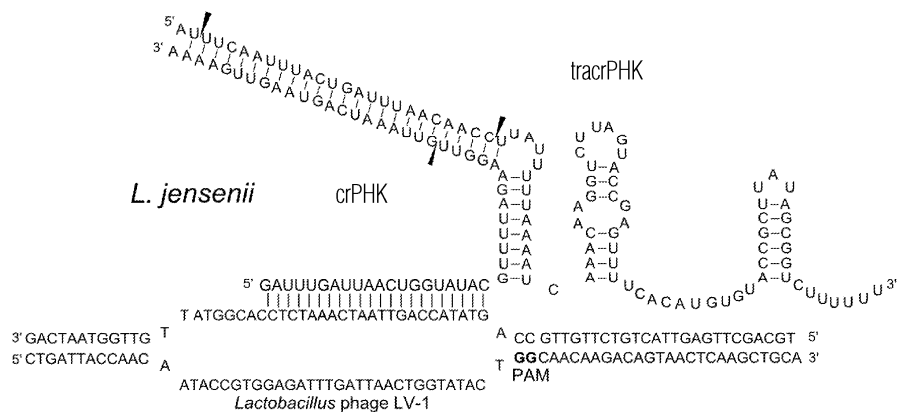
Фиг. 16

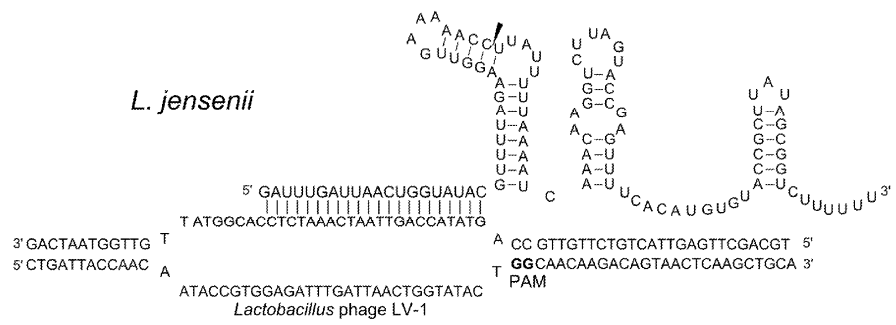


Фиг. 16 (продолжение)

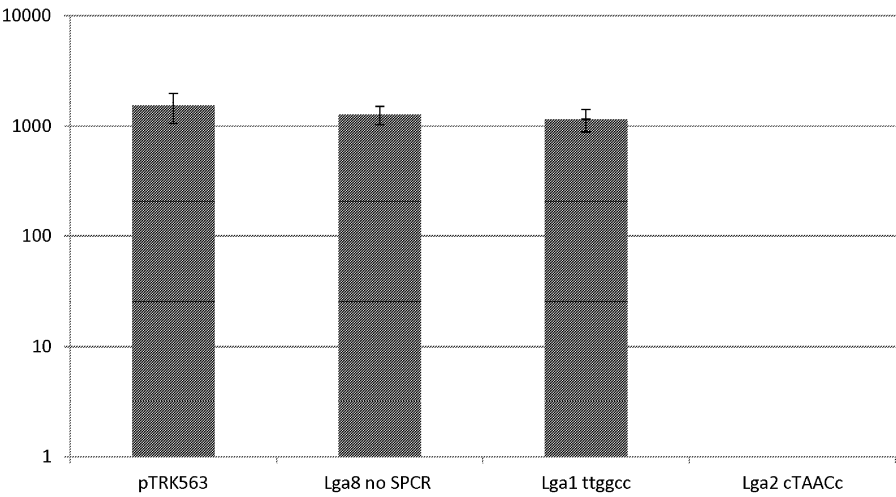


Фиг. 17

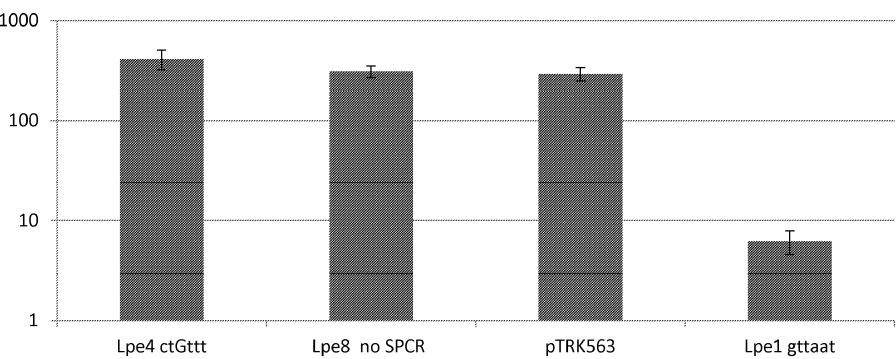




Фиг. 18



Фиг. 19



Фиг. 20

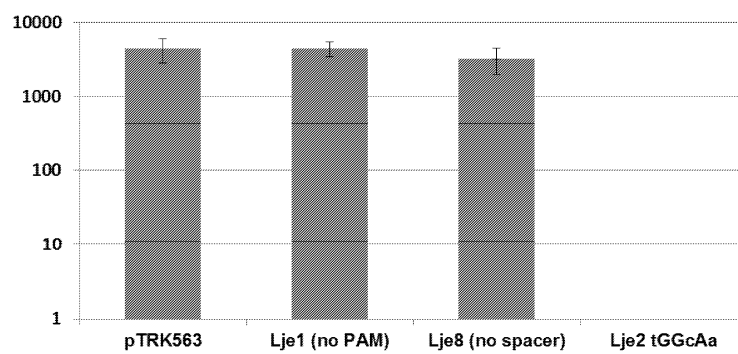
Лидерная последовательность типа I

ATGGGATAGGGATTTTTTAGT

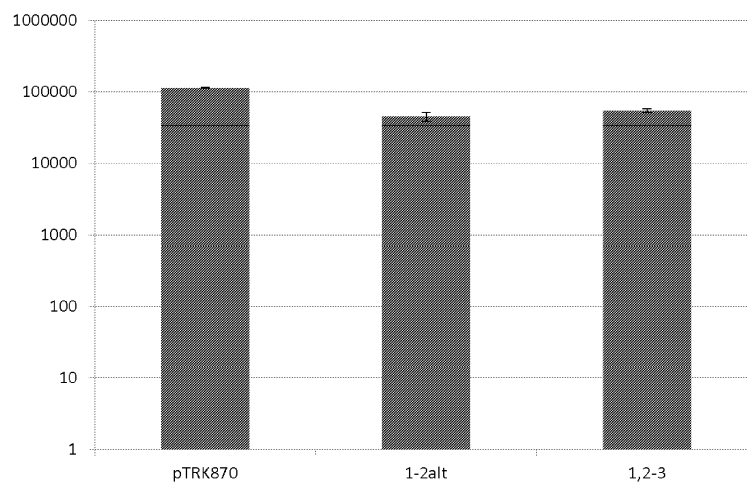
GT TTTCCCGCACATGCGGGGTGATCC GGCATGATACGTAGCCGAAGTGAAGGTTGAT GT TTTCCCGCACATGCGGGGTGATCC

ПОВТОР SPCR 1 ПОВТОР

Фиг. 21



Фиг. 22



Фиг. 23

