

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2025-510586

(P2025-510586A)

(43)公表日 令和7年4月15日(2025.4.15)

(51)国際特許分類		F I		テーマコード(参考)	
C 1 2 N	9/78 (2006.01)	C 1 2 N	9/78	Z N A	4 B 0 6 5
C 0 7 K	19/00 (2006.01)	C 0 7 K	19/00		4 C 0 7 6
C 1 2 N	15/09 (2006.01)	C 1 2 N	15/09	1 0 0	4 C 0 8 4
C 1 2 N	5/10 (2006.01)	C 1 2 N	5/10		4 H 0 4 5
C 1 2 N	15/62 (2006.01)	C 1 2 N	15/62	Z	
		審査請求	未請求	予備審査請求	未請求 (全82頁) 最終頁に続く

(21)出願番号	特願2024-553739(P2024-553739)	(71)出願人	517253632
(86)(22)出願日	令和5年3月7日(2023.3.7)		インスティテュート・オブ・ジェネティクス・アンド・ディヴェロプメンタル・バイオロジー、チャイニーズ・アカデミー・オブ・サイエンス
(85)翻訳文提出日	令和6年11月8日(2024.11.8)		Institute of Genetics and Developmental Biology, Chinese Academy of Sciences
(86)国際出願番号	PCT/CN2023/080052		中華人民共和国 ペキン チャオヤン ウェスト ベイチェン ロード ナンバー 1
(87)国際公開番号	WO2023/169410		No. 1 West Beichen Road, Chaoyang, Beijing 100101, China
(87)国際公開日	令和5年9月14日(2023.9.14)		
(31)優先権主張番号	202210220832.9	(74)代理人	100114890
(32)優先日	令和4年3月8日(2022.3.8)		最終頁に続く
(33)優先権主張国・地域又は機関	中国(CN)		
(81)指定国・地域	AP(BW,CV,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV)		
	最終頁に続く		

(54)【発明の名称】 シトシンデアミナーゼおよび塩基編集におけるその使用

(57)【要約】

本発明は、遺伝子工学の分野に関する。具体的には、本発明は、シトシンデアミナーゼおよび塩基編集におけるその使用に関する。より具体的には、本発明は、デアミナーゼのスクリーニングおよび同定方法、新たに同定されたシトシンデアミナーゼに基づく塩基編集システム、その塩基編集システムを用いて生物(例えば、植物)のゲノム中の標的配列の塩基編集を行う方法、ならびに前記方法によって産生された遺伝子改変生物(例えば、植物)およびその子孫に関する。

【特許請求の範囲】

【請求項 1】

DNA中のデオキシシチジンのシトシン塩基を脱アミノ化することができる、シトシンデアミナーゼ。

【請求項 2】

前記シトシンデアミナーゼは細菌に由来する、請求項 1 に記載のシトシンデアミナーゼ。

【請求項 3】

前記シトシンデアミナーゼと基準シトシンデアミナーゼとのAlphaFold2の三次元構造のTM-スコアが、0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、前記シトシンデアミナーゼが、基準シトシンデアミナーゼのアミノ酸配列と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性を有するアミノ酸配列、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、かつ、前記シトシンデアミナーゼが、DNAのデオキシシチジンのシトシン塩基を脱アミノ化することができる、請求項 1 または 2 に記載のシトシンデアミナーゼ。

10

20

【請求項 4】

前記基準シトシンデアミナーゼは、

(a) 配列番号64に示される配列を有するrAPOBEC1、または

(b) 配列番号65に示される配列を有するDddA、または

(c) 配列番号4に示される配列を有するSdd7である、請求項 3 に記載のシトシンデアミナーゼ。

【請求項 5】

前記シトシンデアミナーゼと配列番号64に示されるrAPOBEC1とのAlphaFold2の三次元構造のTM-スコアが0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、前記シトシンデアミナーゼが、配列番号64と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、かつ、前記シトシンデアミナーゼがDNAのデオキシシチジンのシトシン塩基を脱アミノ化することができる、請求項 4 に記載のシトシンデアミナーゼ。

30

【請求項 6】

前記シトシンデアミナーゼと配列番号65に示されるDddAとのAlphaFold2の三次元構造のTM-スコアが0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、前記シトシンデアミナーゼが、配列番号65と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性を有するアミノ酸配列、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、かつ、前記シトシンデアミナーゼがDNAのデオキシシチジンのシトシン塩基を脱アミノ化することができる、請求項 4 に記載のシトシンデアミナーゼ。

40

50

【請求項 7】

前記シトシンデアミナーゼと配列番号 4 に示される S d d 7 との A l p h a F o l d 2 の三次元構造の T M - スコアが 0 . 6 以上、0 . 7 以上、0 . 7 5 以上、0 . 8 以上、0 . 8 5 以上であり、前記シトシンデアミナーゼが、配列番号 4 と 2 0 ~ 7 0 %、2 0 ~ 6 0 %、2 0 ~ 5 0 %、2 0 ~ 4 5 %、2 0 ~ 4 0 %、2 0 ~ 3 5 % の配列同一性を有するアミノ酸配列、または少なくとも 2 0 %、少なくとも 3 0 %、少なくとも 4 0 %、少なくとも 5 0 %、少なくとも 6 0 %、少なくとも 7 0 %、少なくとも 7 5 %、少なくとも 8 0 %、少なくとも 8 5 %、少なくとも 9 0 %、少なくとも 9 1 %、少なくとも 9 2 %、9 3 %、少なくとも 9 4 %、少なくとも 9 5 %、少なくとも 9 6 %、少なくとも 9 7 %、少なくとも 9 8 %、少なくとも 9 9 % の配列同一性を有するアミノ酸配列を含み、かつ、前記シトシンデアミナーゼが D N A のデオキシシチジンのシトシン塩基を脱アミノ化することができる、請求項 4 に記載のシトシンデアミナーゼ。

10

【請求項 8】

A I D / A P O B E C クレード、S C P 1 . 2 0 1 クレード、M a f B 1 9 クレード、新規 A I D / A P O B E C 様クレード、T M 1 5 0 6 クレード、X O O 2 8 9 7 またはトキシデアミナーゼクレードからのものである、請求項 1 ~ 7 のいずれか 1 項に記載のシトシンデアミナーゼ。

【請求項 9】

A I D / A P O B E C クレードからのものであり、配列番号 1 または 6 3 と少なくとも 5 0 %、少なくとも 6 0 %、少なくとも 7 0 %、少なくとも 7 5 %、少なくとも 8 0 %、少なくとも 8 5 %、少なくとも 9 0 %、少なくとも 9 1 %、少なくとも 9 2 %、9 3 %、少なくとも 9 4 %、少なくとも 9 5 %、少なくとも 9 6 %、少なくとも 9 7 %、少なくとも 9 8 %、少なくとも 9 9 %、さらには 1 0 0 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシンデアミナーゼ。

20

【請求項 10】

S C P 1 . 2 0 1 クレードからのものであり、配列番号 2 8 ~ 4 0 のうちのいずれかと少なくとも 5 0 %、少なくとも 6 0 %、少なくとも 7 0 %、少なくとも 7 5 %、少なくとも 8 0 %、少なくとも 8 5 %、少なくとも 9 0 %、少なくとも 9 1 %、少なくとも 9 2 %、9 3 %、少なくとも 9 4 %、少なくとも 9 5 %、少なくとも 9 6 %、少なくとも 9 7 %、少なくとも 9 8 %、少なくとも 9 9 %、さらには 1 0 0 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシンデアミナーゼ。

30

【請求項 11】

二本鎖 D N A のシトシン塩基を脱アミノ化することができる、請求項 10 に記載のシトシンデアミナーゼ。

【請求項 12】

アミノ酸配列は、配列番号 2 8 ~ 4 0 のうちのいずれかのアミノ酸配列を含む、請求項 10 または 11 に記載のシトシンデアミナーゼ。

【請求項 13】

アミノ酸配列は、配列番号 2 8、3 3、3 4、3 5 のうちのいずれかのアミノ酸配列を含む、請求項 12 に記載のシトシンデアミナーゼ。

40

【請求項 14】

S C P 1 . 2 0 1 クレードからのものであり、配列番号 2 ~ 1 8、4 1 ~ 4 9 のいずれかと少なくとも 5 0 %、少なくとも 6 0 %、少なくとも 7 0 %、少なくとも 7 5 %、少なくとも 8 0 %、少なくとも 8 5 %、少なくとも 9 0 %、少なくとも 9 1 %、少なくとも 9 2 %、9 3 %、少なくとも 9 4 %、少なくとも 9 5 %、少なくとも 9 6 %、少なくとも 9 7 %、少なくとも 9 8 %、少なくとも 9 9 %、さらには 1 0 0 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシンデアミナーゼ。

【請求項 15】

一本鎖 D N A のシトシン塩基を脱アミノ化することができる、請求項 14 に記載のシトシンデアミナーゼ。

50

【請求項 16】

アミノ酸配列は、配列番号 2 ~ 18、41 ~ 49 のうちのいずれかのアミノ酸配列を含む、請求項 14 または 15 に記載のシトシデアミナーゼ。

【請求項 17】

アミノ酸配列は、配列番号 2 ~ 7、12、17 のうちのいずれかのアミノ酸配列を含む、請求項 16 に記載のシトシデアミナーゼ。

【請求項 18】

DNA のデオキシチジンのシトシン塩基を脱アミノ化することができるトランケートシトシデアミナーゼである、請求項 1 ~ 17 のいずれか一項に記載のシトシデアミナーゼ。

10

【請求項 19】

前記トランケートシトシデアミナーゼの長さの範囲が 130 ~ 160 アミノ酸である、請求項 18 に記載のシトシデアミナーゼ。

【請求項 20】

前記トランケートシトシデアミナーゼは、AAV 粒子内に個別にパッケージングされ得る、請求項 19 に記載のシトシデアミナーゼ。

【請求項 21】

前記トランケートシトシデアミナーゼは、配列番号 50 ~ 55 のうちのいずれかと少なくとも 50 %、少なくとも 60 %、少なくとも 70 %、少なくとも 75 %、少なくとも 80 %、少なくとも 85 %、少なくとも 90 %、少なくとも 91 %、少なくとも 92 %、93 %、少なくとも 94 %、少なくとも 95 %、少なくとも 96 %、少なくとも 97 %、少なくとも 98 %、少なくとも 99 %、さらには 100 % の配列同一性を有するアミノ酸配列を含む、請求項 18 ~ 20 のいずれか 1 項に記載のシトシデアミナーゼ。

20

【請求項 22】

前記トランケートシトシデアミナーゼは一本鎖 DNA のシトシン塩基を脱アミノ化することができる、請求項 21 に記載のシトシデアミナーゼ。

【請求項 23】

前記トランケートシトシデアミナーゼは、配列番号 50 ~ 55 のうちのいずれかのアミノ酸配列からなる、請求項 21 または 22 に記載のシトシデアミナーゼ。

【請求項 24】

MafB19 クレードからのものであり、配列番号 19、56、57、58 のうちのいずれかと少なくとも 50 %、少なくとも 60 %、少なくとも 70 %、少なくとも 75 %、少なくとも 80 %、少なくとも 85 %、少なくとも 90 %、少なくとも 91 %、少なくとも 92 %、93 %、少なくとも 94 %、少なくとも 95 %、少なくとも 96 %、少なくとも 97 %、少なくとも 98 %、少なくとも 99 %、さらには 100 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシデアミナーゼ。

30

【請求項 25】

新規 AID / APOBEC 様クレードからのものであり、配列番号 20、21 のうちのいずれかと少なくとも 50 %、少なくとも 60 %、少なくとも 70 %、少なくとも 75 %、少なくとも 80 %、少なくとも 85 %、少なくとも 90 %、少なくとも 91 %、少なくとも 92 %、93 %、少なくとも 94 %、少なくとも 95 %、少なくとも 96 %、少なくとも 97 %、少なくとも 98 %、少なくとも 99 %、さらには 100 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシデアミナーゼ。

40

【請求項 26】

TM1506 クレードからのものであり、配列番号 22 と少なくとも 50 %、少なくとも 60 %、少なくとも 70 %、少なくとも 75 %、少なくとも 80 %、少なくとも 85 %、少なくとも 90 %、少なくとも 91 %、少なくとも 92 %、93 %、少なくとも 94 %、少なくとも 95 %、少なくとも 96 %、少なくとも 97 %、少なくとも 98 %、少なくとも 99 %、さらには 100 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシデアミナーゼ。

50

【請求項 27】

X O O 2 8 9 7 クレードからのものであり、配列番号 2 3、2 4、5 9 ~ 6 2 のうちのいずれかと少なくとも 5 0 %、少なくとも 6 0 %、少なくとも 7 0 %、少なくとも 7 5 %、少なくとも 8 0 %、少なくとも 8 5 %、少なくとも 9 0 %、少なくとも 9 1 %、少なくとも 9 2 %、9 3 %、少なくとも 9 4 %、少なくとも 9 5 %、少なくとも 9 6 %、少なくとも 9 7 %、少なくとも 9 8 %、少なくとも 9 9 %、さらには 1 0 0 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシンデアミナーゼ。

【請求項 28】

トキシデアミナーゼクレードからのものであり、配列番号 7 4 または 7 5 と少なくとも 5 0 %、少なくとも 6 0 %、少なくとも 7 0 %、少なくとも 7 5 %、少なくとも 8 0 %、少なくとも 8 5 %、少なくとも 9 0 %、少なくとも 9 1 %、少なくとも 9 2 %、9 3 %、少なくとも 9 4 %、少なくとも 9 5 %、少なくとも 9 6 %、少なくとも 9 7 %、少なくとも 9 8 %、少なくとも 9 9 %、さらには 1 0 0 % の配列同一性を有するアミノ酸配列を含む、請求項 8 に記載のシトシンデアミナーゼ。

10

【請求項 29】

生物または生物の細胞における塩基編集などの遺伝子編集における、請求項 1 ~ 2 8 のいずれか 1 項に記載のシトシンデアミナーゼの使用。

【請求項 30】

(a) 核酸ターゲティングドメインと、
(b) 少なくとも 1 つの請求項 1 ~ 2 8 のいずれか一項に記載のシトシンデアミナーゼポリペプチドを含むシトシン脱アミノ化ドメインと、を含む、融合タンパク質。

20

【請求項 31】

前記核酸ターゲティングドメインは、T A L E、Z F P または C R I S P R エフェクタータンパク質ドメインである、請求項 3 0 に記載の融合タンパク質。

【請求項 32】

前記 C R I S P R エフェクタータンパク質は、C a s 9、C p f 1、C a s 3、C a s 8 a、C a s 5、C a s 8 b、C a s 8 c、C a s 1 0 d、C s e 1、C s e 2、C s y 1、C s y 2、C s y 3、G S U 0 0 5 4、C a s 1 0、C s m 2、C m r 5、C a s 1 0、C s x 1 1、C s x 1 0、C s f 1、C s n 2、C a s 4、C 2 c 1 (C a s 1 2 b)、C 2 c 3、C 2 c 2、C a s 1 2 c、C a s 1 2 d、C a s 1 2 e、C a s 1 2 f、C a s 1 2 g、C a s 1 2 h、C a s 1 2 i、C a s 1 2 j、C a s 1 2 l、C a s 1 2 m、または他の使用可能な C R I S P R エフェクタータンパク質のうちの少なくとも 1 種である、請求項 3 1 に記載の融合タンパク質。

30

【請求項 33】

前記 C a s 9 は、ヌクレアーゼ不活性 C a s 9、C a s 9 ニッカーゼまたはヌクレアーゼ活性 C a s 9 である、請求項 3 2 に記載の融合タンパク質。

【請求項 34】

前記 C a s 9 は、ヌクレアーゼ不活性 C a s 9 であり、ヌクレアーゼ不活性 C a s 9 は、配列番号 2 6 に示されるアミノ酸配列を含む、請求項 3 3 に記載の融合タンパク質。

【請求項 35】

前記核酸ターゲティングドメインと前記シトシン脱アミノ化ドメインは、リンカーによって融合される、請求項 3 0 ~ 3 4 のいずれか 1 項に記載の融合タンパク質。

40

【請求項 36】

前記融合タンパク質は、ウラシル D N A グリコシラーゼ阻害剤 (U G I) と共発現する、請求項 3 0 ~ 3 5 のいずれか 1 項に記載の融合タンパク質。

【請求項 37】

ウラシル D N A グリコシラーゼ阻害剤 (U G I) をさらに含み、例えば、前記 U G I は、リンカーを介して前記融合タンパク質の他の部分に連結され、または前記 U G I は、「自己切断ペプチド」を介して前記融合タンパク質の他の部分に連結される、請求項 3 0 ~ 3 5 のいずれか 1 項に記載の融合タンパク質。

50

【請求項 38】

1つまたは複数の核局在化配列 (N L S) をさらに含む、請求項 30 ~ 37 のいずれか 1 項に記載の融合タンパク質。

【請求項 39】

核酸標的領域を改変するための塩基編集システムであって、

i) 請求項 1 ~ 28 のいずれか 1 項に記載のシトシンデアミナーゼまたは請求項 30 ~ 38 のいずれか 1 項に記載の融合タンパク質、および / または前記シトシンデアミナーゼまたは融合タンパク質をコードするヌクレオチド配列を含む発現構築物を含む、塩基編集システム。

【請求項 40】

i i) 少なくとも 1 種のガイド RNA および / または前記少なくとも 1 種のガイド RNA をコードするヌクレオチド配列を含む少なくとも 1 種の発現構築物、および / または

i i i) 少なくとも 1 種のウラシル DNA グリコシル化阻害剤 (U G I) および / または前記ウラシル DNA グリコシル化阻害剤 (U G I) をコードするヌクレオチド配列を含む少なくとも 1 種の発現構築物、および / または

i v) 核局在化配列 (N L S) をさらに含む、請求項 39 に記載の塩基編集システム。

【請求項 41】

前記少なくとも 1 種のガイド RNA は、前記融合タンパク質の核酸ターゲティングドメインに結合可能であり、前記ガイド RNA は、前記核酸標的領域内の少なくとも 1 つの標的配列を対象とする、請求項 40 に記載の塩基編集システム。

【請求項 42】

前記ガイド RNA は、長さが 15 ~ 100 ヌクレオチドであり、標的配列に相補的な少なくとも 10 個、少なくとも 15 個または少なくとも 20 個の連続ヌクレオチド配列を含む、請求項 41 に記載の塩基編集システム。

【請求項 43】

前記ガイド RNA は、標的配列に相補的な 15 ~ 40 個の連続ヌクレオチド配列を含む、請求項 42 に記載の塩基編集システム。

【請求項 44】

前記ガイド RNA の長さが 15 ~ 50 ヌクレオチドである、請求項 40 ~ 43 のいずれか 1 項に記載の塩基編集システム。

【請求項 45】

前記標的核酸は DNA である、請求項 40 ~ 44 のいずれか 1 項に記載の塩基編集システム。

【請求項 46】

前記標的核酸は生物のゲノムにある、請求項 40 ~ 45 のいずれか 1 項に記載の塩基編集システム。

【請求項 47】

前記生物は、細菌などの原核生物、植物、真菌または脊椎動物などの真核生物である、請求項 46 に記載の塩基編集システム。

【請求項 48】

前記脊椎動物は、ヒト、マウス、ラット、サル、犬、豚、羊、牛、猫などの哺乳動物である、請求項 47 に記載の塩基編集システム。

【請求項 49】

前記植物は、小麦、イネ、トウモロコシ、大豆、ヒマワリ、ソルガム、アブラナ、アルファルファ、綿、大麦、キビ、サトウキビ、トマト、タバコ、キャッサバまたはジャガイモといった作物植物である、請求項 47 に記載の塩基編集システム。

【請求項 50】

請求項 39 ~ 49 のいずれか 1 項に記載の塩基編集システムを核酸分子の標的配列に接触させることを含む、塩基編集方法。

【請求項 51】

10

20

30

40

50

前記核酸分子は、DNA分子であり、例えば、前記核酸分子は、二本鎖DNA分子または一本鎖DNA分子である、請求項50に記載の塩基編集方法。

【請求項52】

前記核酸分子の標的配列は、植物の形質または発現に関連する配列を含む、請求項50～51のいずれか1項に記載の塩基編集方法。

【請求項53】

前記核酸分子の標的配列は、疾患または障害に関連する配列または点突然変異を含む、請求項50～51のいずれか1項に記載の塩基編集方法。

【請求項54】

前記塩基編集システムは、核酸分子の標的配列に接触して脱アミノ化作用を発揮し、前記脱アミノ化作用により標的配列の1つまたは複数のヌクレオチドの置換が起こる、請求項50～53のいずれか1項に記載の塩基編集方法。 10

【請求項55】

前記標的配列は、DNA配列5'-MCN-3'を含み、MはA、T、CまたはGであり、NはA、T、CまたはGであり、5'-MCN-3'配列の中央にあるCが脱アミノ化される、請求項50～54のいずれか1項に記載の塩基編集方法。

【請求項56】

前記脱アミノ化作用により、スプライス部位が導入または除去される、請求項50～55のいずれか1項に記載の塩基編集方法。

【請求項57】

前記脱アミノ化作用により、遺伝子プロモーターに突然変異が導入され、前記突然変異により、前記遺伝子プロモーターに操作可能に連結された遺伝子の転写が増加または減少する、請求項50～55のいずれか1項に記載の塩基編集方法。 20

【請求項58】

前記脱アミノ化作用により、遺伝子リプレッサーに突然変異が導入され、前記突然変異により、前記遺伝子リプレッサーに操作可能に連結された遺伝子の転写が増加または減少する、請求項50～55のいずれか1項に記載の塩基編集方法。

【請求項59】

前記接触は、生体内または生体外で行われる、請求項50～58のいずれか1項に記載の塩基編集方法。 30

【請求項60】

少なくとも1つの遺伝子改変細胞を産生する方法であって、請求項39～49のいずれか1項に記載の塩基編集システムを少なくとも1つの前記細胞に導入し、それによって前記少なくとも1つの細胞の標的核酸領域内の1つまたは複数のヌクレオチド置換を起こすことを含み、例えば、前記1つまたは複数のヌクレオチド置換は、CからTへの置換である、方法。

【請求項61】

1つまたは複数の所望のヌクレオチド置換を有する細胞を前記少なくとも1つの細胞からスクリーニングするステップをさらに含む、請求項60に記載の方法。

【請求項62】

前記塩基編集システムは、リン酸カルシウムトランスフェクション、プロトプラスト融合、エレクトロポレーション、リポフェクタミントランスフェクション、マイクロインジェクション、ウイルス感染（例えばバキュロウイルス、ワクシニアウイルス、アデノウイルス、アデノ随伴ウイルス、レンチウイルスまたは他のウイルス）、遺伝子銃法、PEG媒介プロトプラスト形質転換、アグロバクテリウム・ツメファシエンス媒介形質転換から選ばれる方法によって細胞に導入される、請求項60または61に記載の方法。 40

【請求項63】

前記細胞は、ヒト、マウス、ラット、サル、犬、豚、羊、牛、猫といった哺乳動物；ニワトリ、アヒル、ガチョウといった家禽；好ましくは、小麦、イネ、トウモロコシ、大豆、ヒマワリ、ソルガム、アブラナ、アルファルファ、綿、大麦、キビ、サトウキビ、トマ 50

ト、タバコ、キャッサバおよびジャガイモといった作物植物である植物に由来する、請求項 60 ~ 62 のいずれか 1 項に記載の方法。

【請求項 64】

データベースから複数の候補タンパク質の配列を取得するステップ(1)と、タンパク質予測プログラムを用いて、前記複数の候補タンパク質のそれぞれの三次元構造を予測するステップ(2)と、

スコアリング関数を用いて前記複数の候補タンパク質の三次元構造の多重構造アライメントを行い、構造類似度マトリックスを取得するステップ(3)と、

系統樹構築方法によって、前記構造類似度マトリックスに基づいて前記複数の候補タンパク質をクラスタリングするステップ(4)と、を含む、タンパク質のクラスタリング方法。

10

【請求項 65】

ステップ(1)において、データベースのアノテーション情報によって前記複数の候補タンパク質の配列を取得し、またはステップ(1)において、基準タンパク質の配列を用い、配列の同一性/類似性に基づいてデータベースを検索することによって前記複数の候補タンパク質の配列を取得する、請求項 64 に記載のタンパク質のクラスタリング方法。

【請求項 66】

前記候補タンパク質は、デアミナーゼ、好ましくはシトシンデアミナーゼである、請求項 64 または 65 に記載のタンパク質のクラスタリング方法。

【請求項 67】

前記データベースは、InterProデータベースである、請求項 64 ~ 66 のいずれか 1 項に記載のタンパク質のクラスタリング方法。

20

【請求項 68】

前記ステップ(2)におけるタンパク質構造予測プログラムは、AlphaFold2、RoseTT、またはタンパク質構造を予測できる他のプログラムから選択される、請求項 64 ~ 67 のいずれか 1 項に記載のタンパク質のクラスタリング方法。

【請求項 69】

前記ステップ(3)で使用されるスコアリング関数には、TM-スコア、RMSD、LDDT、GDT score、QSC、FAPE、またはタンパク質構造の類似度をスコアリングできる他のスコアリング関数が含まれる、請求項 64 ~ 68 のいずれか 1 項に記載のタンパク質のクラスタリング方法。

30

【請求項 70】

前記ステップ(4)の系統樹構築方法は、UPGMAである、請求項 64 ~ 69 のいずれか 1 項に記載のタンパク質のクラスタリング方法。

【請求項 71】

ステップ(4)は、前記複数の候補タンパク質のクラスタリング樹状図を取得する、請求項 64 ~ 70 のいずれか 1 項に記載のタンパク質のクラスタリング方法。

【請求項 72】

三次元構造に基づいてタンパク質の機能を予測する方法であって、請求項 64 ~ 71 のいずれか 1 項に記載のタンパク質のクラスタリング方法に基づいて複数の候補タンパク質をクラスタリングし、次に、クラスタリング結果に基づいて前記候補タンパク質の機能を予測することを含む、方法。

40

【請求項 73】

前記複数の候補タンパク質は、機能が既知の少なくとも 1 つの基準タンパク質を含む、請求項 72 に記載のタンパク質の機能を予測する方法。

【請求項 74】

基準タンパク質が存在するクレードまたはサブクレードの他の候補タンパク質の機能は、機能が既知の基準タンパク質のクラスター(樹状図)での位置によって予測される、請求項 73 に記載のタンパク質の機能を予測する方法。

【請求項 75】

50

前記基準タンパク質は、デアミナーゼ、例えばシトシンデアミナーゼである、請求項 7 3 または 7 4 に記載のタンパク質の機能を予測する方法。

【請求項 7 6】

前記基準タンパク質は、基準シトシンデアミナーゼであり、前記基準シトシンデアミナーゼは、配列が配列番号 6 4 に示される r A P O B E C 1、または配列が配列番号 6 5 に示される D d d A である、請求項 7 5 に記載のタンパク質の機能を予測する方法。

【請求項 7 7】

三次元構造に基づいてタンパク質の最小機能ドメインを同定する方法であって、請求項 6 4 ~ 7 1 のいずれか 1 項に記載のタンパク質のクラスタリング方法によってクラスタリングされた、例えば、同じクレードまたはサブクレードにクラスタリングされた複数の候補タンパク質の構造をアライメントして、保存されたコア構造を決定するステップ a) と、

前記保存されたコア構造を最小機能ドメインとして同定するステップ b) と、を含む、方法。

【請求項 7 8】

前記複数の候補タンパク質は、機能が既知の少なくとも 1 つの基準タンパク質を含む、請求項 7 7 に記載の方法。

【請求項 7 9】

前記基準タンパク質は、デアミナーゼであり、好ましくは、シトシンデアミナーゼである、請求項 7 8 に記載の方法。

【請求項 8 0】

前記基準タンパク質は、基準シトシンデアミナーゼであり、前記基準シトシンデアミナーゼは、配列が配列番号 6 4 に示される r A P O B E C 1、または配列が配列番号 6 5 に示される D d d A である、請求項 7 9 に記載の方法。

【請求項 8 1】

請求項 7 5 ~ 7 6 のいずれか 1 項に記載のタンパク質の機能を予測する方法によって同定されたシトシンデアミナーゼ。

【請求項 8 2】

請求項 7 9 ~ 8 0 のいずれか 1 項に記載の方法によって同定されたシトシンデアミナーゼの最小機能ドメインを含むか、またはそれからなるトランケートシトシンデアミナーゼ

【請求項 8 3】

生物または生物細胞における塩基編集などの遺伝子編集における請求項 8 1 に記載のシトシンデアミナーゼまたは請求項 8 1 に記載のトランケートシトシンデアミナーゼの使用

【請求項 8 4】

請求項 1 ~ 2 8 のいずれか 1 項に記載のシトシンデアミナーゼ、または請求項 2 9 ~ 3 7 のいずれか 1 項に記載の融合タンパク質をコードする、核酸分子。

【請求項 8 5】

請求項 1 ~ 2 8 のいずれか 1 項に記載のシトシンデアミナーゼ、または請求項 3 0 ~ 3 8 のいずれか 1 項に記載の融合タンパク質、または請求項 3 9 ~ 4 9 のいずれか 1 項に記載の塩基編集システム、または請求項 8 4 に記載の核酸分子を含む、細胞。

【請求項 8 6】

(a) 請求項 1 ~ 2 8 のいずれか 1 項に記載のシトシンデアミナーゼをコードする核酸配列と、

(b) (a) の配列の発現を駆動する異種プロモーターと、を含む核酸構築物を含むキット。

【請求項 8 7】

(a) 請求項 3 0 ~ 3 8 のいずれか 1 項に記載の融合タンパク質をコードする核酸配列と、

10

20

30

40

50

(b)(a)の配列の発現を駆動する異種プロモーターと、を含む、核酸構築物を含むキット。

【請求項 88】

ガイドRNA主鎖をコードする発現構築物をさらに含み、前記構築物は、標的配列と同一または相補的な核酸配列の前記ガイドRNA主鎖へのクロニングを可能にするクロニング部位を含む、請求項 86～87のいずれか 1 項に記載のキット。

【請求項 89】

請求項 1～28のいずれか 1 項に記載のシトシンデアミナーゼ、または請求項 30～38のいずれか 1 項に記載の融合タンパク質、または請求項 39～49のいずれか 1 項に記載の塩基編集システム、または請求項 84に記載の核酸分子を含む、組成物。

10

【請求項 90】

前記シトシンデアミナーゼ、融合タンパク質、塩基編集システムまたは核酸分子は、ウイルス、ウイルス様粒子、ピリオン、リポソーム、小胞、エキソソーム、リポソームナノ粒子(LNP)にパッケージングされる、請求項 89に記載の組成物。

【請求項 91】

前記ウイルスはアデノ随伴ウイルス(AAV)または組換えアデノ随伴ウイルス(rAAV)である、請求項 90に記載の組成物。

【請求項 92】

医薬組成物であり、薬学的に許容される担体をさらに含む、請求項 89～91に記載の組成物。

20

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、遺伝子工学の分野に関する。具体的には、本発明は、シトシンデアミナーゼおよび塩基編集におけるその使用に関する。より具体的には、本発明は、新たに同定されたシトシンデアミナーゼに基づく塩基編集システム、その塩基編集システムを用いて生物(例えば、植物)のゲノム中の標的配列の塩基編集を行う方法、および前記方法によって生産された遺伝子改変生物(例えば、植物)およびその子孫に関する。

【0002】

発明の背景

生物のゲノムに特定の配列の改変を加えると、その生物に新しい安定した遺伝的形質を与えることができる。ここで、特定部位の単一ヌクレオチド変異は、遺伝子のアミノ酸配列の変化や早期終結を引き起こしたり、調節配列の変化を引き起こしたりして、優れた形質を生み出す可能性がある。CRISPR/Cas9システムなどのゲノム編集技術は、ゲノム内の標的配列を標的にする機能を実現できる。ゲノム編集システムが標的配列に結合されるという特性を利用して、このゲノム編集システムをデアミナーゼと組み合わせることで開発した塩基編集システムは、ゲノム上の標的ヌクレオチドを正確に脱アミノ化することができる。シトシン塩基編集システムは、APOBEC/AIDファミリーとAPOBEC/AIDファミリー様デアミナーゼを融合することにより、標的部位でシトシン(C)からウラシル(U)への変換を実現することができ、その後、細胞内の関連する修復経路の助けを借りて、シトシンからチミン(T)への変換を実現する。また、反対側の脱アミノ化されていない一本鎖にニックを導入して切断することで、塩基編集の効率を大幅に向上させることができる。

30

40

【0003】

Iyerらは、デアミナーゼの構造比較に基づいて、潜在的な脱アミノ化機能を持つタンパク質を探し、上記のタンパク質を少なくとも20個のクレードに分類した(Iyer, L. M., Zhang, D., Rogozin, I. B., & Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic acids research*, 39(22), 9473-9497.)。異なるクレードからのデアミナーゼは、構造と配列が大きく異な

50

ることを発見した。dCMPをdUMPに変換できる「dCMPデアミナーゼおよびC o m E」クレード、グアニン（G）をキサンチン（I）に変換できる「グアニンデアミナーゼ」クレード、ジアミノヒドロキシホスホリボシルアミノピリミジンデアミナーゼ機能を持つ「R i b D様」クレード、RNAアデニン（A）をキサンチン（I）に変換するRNA編集酵素機能を持つ「T a d 1 / A D A R」クレード、およびホルミルトランスフェラーゼ活性を持つ「P u r H / A I C A Rトランスホルミラーゼ」クレードなどの幾つかのクレードの機能が解析されている。細菌由来のS C P 1 . 2 0 1クレード、X O O 2 8 9 7クレード、M a f B 1 9クレード、P p u t _ 2 6 1 3クレードなどの一部のクレードについては、脱アミノ化活性の有無やどのような基質を脱アミノ化できるかなどの機能が解析、実証されていない。現在、一本鎖DNAに作用できることが証明されており、その

10

【0004】

当該分野では、塩基編集システムを拡張し、標的DNA配列を正確に操作する能力を向上させるために、塩基編集システムで使用できるデアミナーゼをより多く必要としている。

【図面の簡単な説明】

【0005】

【図1】A P O B E C / A I Dクレードの潜在的なデアミナーゼNo. 182（配列番号1）は、レポーターシステムでシトシン塩基編集を実現する。

20

【図2】A P O B E C / A I Dクレードの潜在的なデアミナーゼNo. 182（配列番号1）は、内因性部位でシトシン塩基編集を実現する。

【図3】S C P 1 . 2 0 1クレードの潜在的なデアミナーゼNo. 69（配列番号2）は、内因性部位でシトシン塩基編集を実現する。

【図4】イネO s A C C - T 1の内因性部位で高い編集効率を示す8つのデアミナーゼのシトシン塩基編集効率である。

【図5】イネC D C 4 8 - T 2の内因性部位で高い編集効率を示す8つのデアミナーゼのシトシン塩基編集効率である。

【図6】イネO s A C C - T 1の内因性部位で中程度の編集効率を示す8つのデアミナーゼのシトシン塩基編集効率である。

30

【図7】イネC D C 4 8 - T 2の内因性部位で中程度の編集効率を示す8つのデアミナーゼのシトシン塩基編集効率である。

【図8】A l p h a F o l d 2の予測構造に基づくタンパク質のクラスタリングプロセスであり、A l p h a F o l d 2を用いて候補配列の構造を予測し、構造の類似性に基づいてクラスタリングを行い、続いて、s s D N Aおよびd s D N Aに対する各構造クレードからのタンパク質のシチジン脱アミノ化活性を、植物およびヒトの細胞で実験的にテストした。

【図9】候補デアミナーゼの再アノテーションと合成の流れであり、NCBIデータベースのP r o t e i n B L A S T (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)を用いてデアミナーゼをコードする遺伝子の全長を取得し、次に、h m m s c a nを用いてデアミナーゼドメイン配列に再アノテーションを付け (<https://www.ebi.ac.uk/Tools/hmmer/search/hmmscan>)、得られたドメイン配列は構造分類に使用され、それらがデアミナーゼ活性を有することを確認するために、N末端およびC末端配列が延長された、幾つかの候補デアミナーゼを合成し、レポーターシステムを用いて、または内因性部位でそれらのシチジンデアミナーゼ活性を評価した。

40

【図10】16個のデアミナーゼファミリー（238）と1つの外群J A B（4）からの242個の予測タンパク質構造間の類似性を反映する構造類似性マトリックスであり、異なるファミリーのタンパク質は異なる番号で区別され、ヒートマップの色の濃度は類似性の程度を示す。

50

【図11A】タンパク質の構造に基づいてタンパク質を異なるデアミナーゼファミリーに分類し、異なるファミリーは異なる番号で区別される。

【図11B】16個のデアミナーゼクレードのそれぞれの代表的な予測構造である。

【図12】図11に対応するLmjF365940、APOBEC、dCMP、およびMafB19ファミリーの2つのクレードの代表的な構造アライメントであり、これら4つのファミリーのそれぞれの2つのクレード系は部分的に類似した構造を持っているが、2つのクレード系は、全体的な構造には比較的大きな違いがあるため、異なるクレード系として分類される。

【図13A】タンパク質構造に基づくSCP1.201デアミナーゼの分類であり、JABファミリーは外群とみなされ、テストされたデアミナーゼは機能に従って一本鎖編集(ssDNA)、二本鎖編集(dsDNA)、または非二本鎖/一本鎖編集(non_ds/ss)として表示され、薄灰色の未定義(undefined)のデアミナーゼはさらなる機能解析を必要とし、図中の一本鎖編集を受けたデアミナーゼドメインは、SCP356、SCP020、SCP051、SCP170、SCP014、SCP273、SCP158、SCP013、SCP008、SCP157、SCP315、SCP183、SCP044、SCP012、SCP011、SCP018、SCP038、SCP016、SCP017であり、二本鎖編集を受けたデアミナーゼドメインは、SCP271、SCP103、SCP009、SCP006、SCP004、SCP234、SCP177であり、アノテーションが付けられている残りのデアミナーゼは、編集されていないものである。

10

20

【図13B】AlphaFold2を用いてDddAのコア構造を予測する。

【図13C】Dddタンパク質(二本鎖デアミナーゼ活性を持つタンパク質)の典型的な構造的特徴である。

【図13D】AlphaFold2を用いてSdd7のコア構造を予測する。

【図13E】Sddタンパク質(一本鎖デアミナーゼ活性を持つタンパク質)の典型的な構造的特徴である。

【図14】動物細胞の内因性部位におけるssDNAおよびdsDNAのシトシン脱アミノ化活性の同定であり、(A)は内因性部位編集用のssDNA塩基編集ベクターの概略図であり、(B)はDdCBEベクターおよびその二分子の概略図であり、(C)はdsDNAに対するDdCBE編集活性とssDNAに対するCBE編集活性がそれぞれHEK293T細胞で検出され、ハイスループット配列決定が実行される。

30

【図15】HEK293T細胞の2つの内因性部位におけるDddのdsDNA脱アミノ化活性の実験的評価であり、計算に使用される塩基編集部位では、色の濃度は編集効率のレベルを表す。

【図16A】HEK293T細胞の2つの内因性部位におけるSddのssDNA脱アミノ化活性の実験的評価であり、計算に使用される塩基編集部位では、色の濃度は編集効率のレベルを示す。

【図16B】HsJAK2およびHsSIRT6部位におけるSddのssDNAデアミナーゼ活性の実験的評価であり、データは、3回の反復かつ独立した実験の平均値である。

40

【図17】塩基エディターとして使用するための、新たに発見されたDddタンパク質の編集特性の評価であり、(A)はHEK293T細胞の2つのゲノム標的におけるSCP1.201のdsDNAデアミナーゼDdd1、Ddd7、Ddd8、Ddd9およびDddAの編集効率と編集ウィンドウであり、(B)は哺乳動物細胞における各Dddタンパク質のコンテキスト選好性を分析するためのプラスミドライブラリー分析であり、候補タンパク質は「NC₁₀N」モチーフを標的として編集し、(C)はプラスミドライブラリー分析によるDdd1、Ddd7、Ddd8、Ddd9およびDddAのコンテキスト選好性を要約したモチーフロゴ(logo)プロットであり、プロットでは、ドットは単一の生物学的反復を示し、カラムの高さは編集効率の平均値を示し、エラーバー(error bars)は3回の独立した生物学的実験の標準偏差を示す。

50

【図18】HEK293T細胞における2つの標的におけるSCP1.201のdsDNAデアミナーゼの編集効率と編集ウィンドウのヒートマップである。

【図19】16種類のプラスミドライブラリーにおける様々なDddデアミナーゼのコンテキスト選好性の編集効率のパーセンテージあり、データは3回の独立した実験の平均値によって表される。

【図20】植物の塩基エディターとして使用するための、新たに発見されたSddタンパク質について、イネプロトプラストの6つの内因性標的における10個のSddタンパク質とrAPOBEC1の全体的な編集効率を評価し、各標的でのAPOBEC1の平均編集頻度は1に設定され、各Sddで観察された編集効率はそれに応じて正規化される。

【図21】イネプロトプラストの6つの内因性標的におけるSddデアミナーゼおよびAPOBEC1の編集挙動であり、ヒートマップ(A~F)は、イネプロトプラストにおけるOsAAT(A)、OsACC1(B)、OsCDC48-T1(C)、OsCDC48-T2(D)、OsDEP1(E)およびOsODEV(F)部位での10個のSddデアミナーゼおよびAPOBEC1の編集効率と編集ウィンドウを示し、ヒートマップのセルに指定された値はC-to-Tの編集効率を示し、色の濃度は編集効率のレベルを表し、標的配列はヒートマップの上部に示され、濃い色のボックスはC-to-T編集の位置を示し、最後の3つの薄い色のフォントはPAMsを示し、データは3回の独立した実験の平均値によって表される。

【図22】HEK293T細胞における3つの内因性標的におけるSCP1.201のssDNAデアミナーゼおよびAPOBECデアミナーゼの編集挙動であり、ヒートマップ(A~C)は、HEK293T細胞のHsEMX1(A)、HsHEK2(B)およびHsWFS1(C)部位での4種類のSddデアミナーゼとAPOBEC1、APOBEC3A、APOBEC1-YE1およびAPOBEC1-YEEの編集効率と編集ウィンドウを示し、ヒートマップのセルに指定された値はC-to-Tの編集効率を示し、色の濃度は編集効率のレベルを表し、標的配列はヒートマップの上部に示され、濃い色のボックスはC-to-T編集の位置を示し、最後の3つの薄い色のフォントはPAMを示し、データは3回の独立した実験の平均値によって表される。

【図23】イネプロトプラストの5つの部位におけるSdd7、APOBEC1およびAPOBEC3Aの編集効率の比較であり、(A~E)は5つの内因性標的である(A)OsACTG、(B)OsALS-T1、(C)OsALS-T2、(D)OsCDC48-T3および(E)OsMPK16でのSdd7、APOBEC1およびAPOBEC3A塩基エディターの効率の比較であり、データは3回の独立した実験の代表的なデータであり、カラムの高さは編集効率の平均値を示し、エラーバー(error bars)は3回の独立した生物学的実験の標準偏差を示す。

【図24】イネプロトプラストの5つの内因性標的におけるSddデアミナーゼおよびAPOBEC1の配列選好性であり、積み上げグラフは、5つの内因性標的であるOsAAT、OsACC1、OsCDC48-T1、OsCDC48-T2およびOsDEP1における10種類のSddデアミナーゼとAPOBEC1のコンテキスト選好性であり、棒グラフは、下から上にそれぞれTC、AC、GCおよびCCのC-to-T編集選好性を表し、データは3回の独立した実験の結果である。

【図25A】12K-TRAPseqライブラリーを使用した、HEK293T細胞におけるSddおよびrAPOBEC1の活性および特性のハイスループット定量化の概要であり。

【図25B】12K-TRAPライブラリーによるSddおよびrAPOBEC1編集選好性とパターンの評価であり、左の図はデアミナーゼの編集効率と編集ウィンドウを示し、右の図の配列モチーフのロゴマップは、デアミナーゼのコンテキスト選好性を反映している。

【図26】(A)はイネプロトプラストにおける直交R-loop解析を用いたオフターゲット効果の評価であり、ドットは、イネの6つの標的における各塩基エディターのオンターゲットC-to-Tの変換の平均頻度(図20)、および2つのssDNA(OsD

EP1-SaT1およびOsDEP1-SaT2)におけるsgRNA非依存性のオフターゲットC-to-Tの変換頻度であり、(B)は図26Aにおける各塩基エディターのオンターゲット:オフターゲット(on-target:off-target)編集比であり、(C)はHEK293T細胞の2つのオンターゲット部位と3つのオフターゲット部位でテストされたSdd6、rAPOBEC1-YE1、rAPOBEC1-YEE、rAPOBEC1およびhAPOBEC3Aのオンターゲット:オフターゲット編集比であり、図中のドットは単一の生物学的反復を示し、カラムの高さは平均値を示し、エラーバーは3つの独立した生物学的反復の標準偏差を示す。

【図27】イネプロトプラストの2つの内因性標的におけるSddデアミナーゼおよびAPOBEC1の具体的なオフターゲット頻度(図26Aおよび図26B)であり、オフターゲットは、直交R-loop法を用いて評価され、(A, B)はイネプロトプラストのOsDEP1-SaT1(A)およびOsDEP1-SaT2(B)部位におけるSddデアミナーゼおよびAPOBEC1のオフターゲット頻度であり、データは3回の独立した実験の結果である。

【図28】HEK293T細胞の2つのオンターゲット部位および4つのオフターゲット部位でテストされたSdd6およびAPOBEC塩基エディターのオンターゲットおよびオフターゲットの具体的な編集効率であり(図26C)、HsHEK2のオンターゲット部位に対応するHsJAK2-SaおよびHsSIRT6-Saのオフターゲット部位における、Sdd6、APOBEC1-YE1、APOBEC1-YEE、APOBEC1およびAPOBEC3Aのそれぞれのオンターゲット効率とオフターゲット効率、ならびにHsHEK3オンターゲット部位に対応するHsRNF2-SaおよびHsFANCF-SaT1オフターゲット部位におけるオンターゲットおよびオフターゲット編集効率である。データは3回の独立した実験の結果である。

【図29】AlphaFold2によって予測された高活性Sddデアミナーゼの保存されたタンパク質構造であり、高い脱アミノ化活性を持つSddデアミナーゼのコア構造を示し、一部の活性デアミナーゼでは、4は必須の構造ではない。

【図30A】動物および植物で使用するために工学的にランケットされたSddタンパク質であり、工学的にランケットされたSddタンパク質であり、最上部の図は、AlphaFold2によって予測されたSdd6、Sdd7、Sdd3およびSdd9の構造であり、保存された領域は濃い色で表され、ランケット領域は薄い色で表され、下の図は、2つの内生イネプロトプラストとHEK293T細胞のそれぞれの2つの内因性部位におけるSddとその最小化バージョンのSddの元の長さのタンパク質に対する編集効率であり、ドットは単一の生物学的反復を示し、カラムの高さと折れ線の点は平均値を示し、エラーバーは3回の独立した生物学的実験の標準偏差を示す。

【図30B】動物および植物で使用するために工学的にランケットされたSddタンパク質であり、SaCas9ベースのCBEベクターが理論的には単一のAAVにパッケージングされており、上の図は、APOBEC/AID様デアミナーゼ、Sddの最小化バージョン、およびそのAAVベクターの概略図であり、それらのうち、APOBEC3G、hAPOBEC3B、rAPOBEC1、PmCDA1、APOBEC3AおよびhAIDデアミナーゼは、シングルAAVパッケージを使用するには大きすぎ、下の図は、Sddの最小化miniバージョンに基づくAAVベクターの概略図である。

【図30C】動物および植物で使用するために工学的にランケットされたSddタンパク質であり、マウスN2a細胞におけるMmHPD遺伝子の2つの内因性標的部位でのミニSdd6の編集効率であり、ドットは単一の生物学的反復を示し、カラムの高さと折れ線の点は平均値を示し、エラーバーは3回の独立した生物学的実験の標準偏差を示す。

【図30D】動物および植物で使用するために工学的にランケットされたSddタンパク質であり、大豆毛状根の5つの内因性標的でのミニSdd7、rAPOBEC1、hAPOBEC AおよびヒトAID塩基エディターの編集効率であり、ドットは単一の生物学的反復を示し、カラムの高さと折れ線の点は平均値を示し、エラーバーは3回の独立した生物学的実験の標準偏差を示す。

10

20

30

40

50

【図 3 0 E】動物および植物で使用するために工学的にトランケートされた S d d タンパク質であり、T₀ 世代大豆植物株におけるミニ S d d 7 誘発突然変異の頻度である。

【図 3 0 F】動物および植物で使用するために工学的にトランケートされた S d d タンパク質であり、塩基編集された大豆植物株の遺伝子型である。

【図 3 0 G】動物および植物で使用するために工学的にトランケートされた S d d タンパク質であり、カルフェントラゾンエチルで 10 日間処理した大豆植物株の表現型であり、左の図は野生型大豆植物株 (R 9 8) であり、右の図は、塩基編集された大豆植物株 (C 9 8) である。

【図 3 1】再生イネにおける塩基編集効率であり、(A) はアグロバクテリウムによって形質転換が媒介されたイネ塩基編集バイナリーベクターの概略図であり、(B) は T₀ イネ植物株突然変異の誘導におけるミニ S d d 7 および h A P O B E C 3 A 塩基エディターの効率である。

【図 3 2】大豆におけるアグロバクテリウム媒介形質転換のための塩基編集バイナリーベクターの概略図である。

【0 0 0 6】

発明の詳細

1. 定義

本発明において、特に断りのない限り、本明細書で使用される科学用語および技術用語は、当業者に一般的に理解される意味を有する。さらに、本明細書で使用されているタンパク質および核酸化学、分子生物学、細胞および組織培養、微生物学、免疫学に関連する用語や実験の操作ステップは、対応する分野で広く使用されている用語および通常のステップである。また、本発明をより良く理解できるようにするため、関連する用語の定義と説明を以下に提供する。

【0 0 0 7】

本明細書で使用される「および/または」という用語は、この用語によって接続される項目の全ての組み合わせを包含し、各組み合わせは本明細書で個別に挙げられているものとみなされる。例えば、「A および/または B」には、「A」、「A および B」、「B」が含まれる。例えば、「A、B および/または C」には、「A」、「B」、「C」、「A および B」、「A および C」、「B および C」、「A、B および C」が含まれる。

【0 0 0 8】

「シトシンデアミナーゼ」は、一本鎖 DNA などの核酸を基質として受容することができ、シチジンまたはデオキシシチジンからそれぞれウラシルまたはデオキシウラシルへの脱アミノ化を触媒することができるデアミナーゼを指す。

【0 0 0 9】

本明細書で使用される「ゲノム」は、細胞核に存在する染色体 DNA だけでなく、細胞の細胞内成分 (例えば、ミトコンドリア、色素体) に存在する細胞小器官 DNA も包含する。

【0 0 1 0】

本明細書で使用される場合、「生物」には、ゲノム編集に適した任意の生物、好ましくは真核生物が含まれる。生物の例としては、ヒト、マウス、ラット、サル、犬、豚、羊、牛、猫などの哺乳動物；ニワトリ、アヒル、ガチョウなどの家禽；単子葉植物と双子葉植物、例えば、イネ、トウモロコシ、小麦、ソルガム、大麦、大豆、落花生、シロイヌナズナなどを含む植物が含まれるが、これらに限定されない。

【0 0 1 1】

「遺伝子改変生物」または「遺伝子改変細胞」は、そのゲノム内に外因性ポリヌクレオチド、改変遺伝子または発現調節配列を含む生物または細胞を意味する。例えば、外因性ポリヌクレオチドは、生物または細胞のゲノムに安定して組み込まれ、連続世代に受け継がれ得る。外因性ポリヌクレオチドは、単独で、または組換え DNA 構築物の一部としてゲノムに組み込むことができる。改変遺伝子または発現調節配列とは、その配列が生物または細胞のゲノムにおいて単一または複数のデオキシヌクレオチド置換、欠失および付加

10

20

30

40

50

を含むものである。

【0012】

配列に関する「外因性」とは、外来種に由来する配列、または同じ種に由来する場合には、意図的な人間の介入によってその本来の形態から組成および/または遺伝子座が大幅に変化した配列を意味する。

【0013】

「ポリヌクレオチド」、「核酸配列」、「ヌクレオチド配列」または「核酸断片」は、互換的に使用され、一本鎖または二本鎖のRNAまたはDNAポリマーであり、合成、非天然または改変ヌクレオチド塩基を任意に含む。ヌクレオチドは、次のようにそれらのアルファベット1文字の名前で呼ばれる。「A」はアデノシンまたはデオキシアデノシン（それぞれRNAまたはDNAに対応する）、「C」はシチジンまたはデオキシシチジン、「G」はグアノシンまたはデオキシグアノシン、「U」はウリジン、「T」はデオキシチミジン、「R」はプリン（AまたはG）、「Y」はピリミジン（CまたはT）、「K」はGまたはT、「H」はA、CまたはT、「I」はイノシン、「N」は任意のヌクレオチドを表す。

【0014】

「ポリペプチド」、「ペプチド」および「タンパク質」は、本発明において互換的に使用され、アミノ酸残基のポリマーを指す。この用語は、1つまたは複数のアミノ酸残基が対応する天然アミノ酸の人工化学類似体であるアミノ酸ポリマー、および天然アミノ酸ポリマーに適用される。「ポリペプチド」、「ペプチド」、「アミノ酸配列」および「タンパク質」という用語には、改変形態も含まれ、この形態には、グリコシル化、脂質結合、硫酸化、グルタミン酸残基のガンマカルボキシル化、ヒドロキシル化、ADP-リボシル化が含まれるが、これらに限定されるものではない。

【0015】

配列の「同一性」は、当該技術分野で認識された意味を有し、2つの核酸またはポリペプチドの分子または領域間の配列同一性のパーセンテージは、公開されている技術を用いて計算することができる。配列同一性は、ポリヌクレオチドまたはポリペプチドの全長に沿って、または分子の領域に沿って測定することができる（例えば、Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991参照）。2つのポリヌクレオチドまたはポリペプチド間の同一性を測定する多くの方法があるが、「同一性」という用語は当業者にはよく知られている（Carriello, H. & Lipman, D., SIAM J Applied Math 48:1073(1988)）。

【0016】

「含む」という用語がタンパク質または核酸の配列を説明するために本明細書で使用される場合、前記タンパク質または核酸は、その配列から構成されていてもよいし、前記タンパク質または核酸の一端または両端に追加のアミノ酸またはヌクレオチドを有していてもよいが、本発明に記載の活性を依然として有する。さらに、ポリペプチドのN末端の開始コドンによってコードされるメチオニンが、特定の実際的な状況下（例えば、特定の発現系で発現される場合など）では保持されるが、ポリペプチドの機能に実質的な影響を及ぼさないことは当業者にとって明らかである。したがって、本願の明細書および特許請求の範囲において特定のポリペプチドのアミノ酸配列が記載されている場合、開始コドンによってコードされるN末端メチオニンを含まない場合があるが、その場合においても、当該メチオニンを含む配列が包含される。これに伴って、そのコードヌクレオチド配列に開始コドンが含まれる場合もあり、その逆も同様である。

【0017】

50

ペプチドまたはタンパク質において、適切な保存的アミノ酸置換は当業者に知られており、一般に、得られる分子の生物学的活性を変えなく行うことができる。通常、当業者は、ポリペプチドの非必須領域における単一アミノ酸置換は生物学的活性を実質的に変化させないことを認識している（例えば、Watson et al., Molecular Biology of the Gene, 4th Edition, 1987, The Benjamin/Cummings Pub. Co., p. 224を参照）。

【0018】

本発明で使用される「発現構築物」とは、生体内での標的ヌクレオチド配列の発現に適した、組換えベクターなどのベクターを指す。「発現」とは、機能的な産物を産生することを指す。例えば、ヌクレオチド配列の発現は、ヌクレオチド配列の転写（例えば、mRNAまたは機能性RNAを生成する転写）および/またはRNAの前駆体または成熟タンパク質への翻訳を指し得る。

10

【0019】

本発明の「発現構築物」は、線状核酸断片、環状プラスミド、ウイルスベクター、または幾つかの実施形態では、翻訳可能なRNA（例えば、mRNA）であってもよい。

【0020】

本発明の「発現構築物」は、異なる供給源からの調節配列および標的ヌクレオチド配列、または同じ供給源からのものであるが自然界に通常存在するものとは異なる方法で配置された調節配列および標的ヌクレオチド配列を含んでもよい。

【0021】

「調節配列」および「調節エレメント」は互換的に使用され、コード配列の上流（5'非コード配列）、中間または下流（3'非コード配列）に位置し、関連するコード配列の転写、RNAプロセッシング、安定性、翻訳に影響を与えるヌクレオチド配列を指す。調節配列には、プロモーター、翻訳リーダー配列、イントロンおよびポリアダニル化認識配列が含まれるが、これらに限定されない。

20

【0022】

「プロモーター」とは、別の核酸断片の転写を制御することができる核酸断片を指す。本発明の幾つかの実施形態では、プロモーターは、それが細胞に由来するかどうかに関係なく、前記細胞内の遺伝子の転写を制御することができるプロモーターである。プロモーターは、構成的プロモーター、組織特異的プロモーター、発生調節プロモーターまたは誘導性プロモーターであり得る。

30

【0023】

「構成的プロモーター」とは、一般に、ほとんどの状況下でほとんどの細胞型において遺伝子の発現を引き起こすプロモーターを指す。「組織特異的プロモーター」および「組織優先的プロモーター」は互換的に使用され、主に、必ずしも排他的ではないが、1つの組織または臓器だけでなく、1つの特定の細胞または細胞型においても発現し得るプロモーターを指す。「発生調節プロモーター」とは、その活性が発生事象によって決定されるプロモーターを指す。「誘導性プロモーター」は、内因性または外因性の刺激（環境、ホルモン、化学シグナルなど）に応答して、操作可能に連結されたDNA配列を選択的に発現する。

【0024】

プロモーターの例としては、ポリメラーゼ（pol）I、pol IIまたはpol IIIプロモーターが含まれるが、これらに限定されない。pol Iプロモーターの例としては、ニワトリRNA pol Iプロモーターが含まれる。pol IIプロモーターの例としては、サイトメガロウイルス前初期（CMV）プロモーター、ラウス肉腫ウイルス長末端反復（RSV-LTR）プロモーターおよびサルウイルス40（SV40）前初期プロモーターが含まれるが、これらに限定されない。pol IIIプロモーターの例としては、U6およびH1プロモーターが含まれる。メタロチオネインプロモーターなどの誘導性プロモーターを使用することができる。プロモーターの他の例としては、T7ファージプロモーター、T3ファージプロモーター、 λ -ガラクトシダーゼプロモーターおよびSp6ファージプロモーターが含まれる。植物において使用される場合、プロモ

40

50

ーターは、カリフラワーモザイクウイルス35Sプロモーター、トウモロコシUbi-1プロモーター、小麦U6プロモーター、イネU3プロモーター、トウモロコシU3プロモーター、イネアクチンプロモーターであってよい。

【0025】

本明細書で使用される「操作可能に連結される」という用語は、調節エレメント（例えば、限定されないが、プロモーター配列、転写終結配列など）が核酸配列（例えば、コード配列、オープンリーディングフレームなど）に連結されることにより、ヌクレオチド配列の転写が転写調節エレメントによって制御および調節されるようにすることを意味する。調節エレメント領域を核酸分子に操作可能に連結するための技術は当技術分野で知られている。

10

【0026】

核酸分子（例えば、プラスミド、線状核酸断片、RNAなど）またはタンパク質を生物に「導入する」とは、前記核酸またはタンパク質が細胞内で機能できるように、前記核酸またはタンパク質で生物の細胞を形質転換することを指す。本発明で使用される「形質転換」には、安定形質転換および一過性形質転換が含まれる。

【0027】

「安定形質転換」とは、ゲノムに外因性ヌクレオチド配列を導入し、その結果、外来遺伝子の安定した継承をもたらされることを指す。外因性核酸配列は、安定に形質転換されると、生物およびその後の連続世代のゲノムに安定して組み込まれる。

【0028】

「一過性形質転換」とは、外来遺伝子の安定した継承を伴わずに機能を実行するために、核酸分子またはタンパク質を細胞に導入することを指す。一過性形質転換では、外因性核酸配列はゲノムに組み込まれない。

20

【0029】

2. 三次元構造に基づくタンパク質のクラスタリングと機能予測方法

一態様では、本発明は、タンパク質のクラスタリング方法を提供する。これは、

データベースから複数の候補タンパク質の配列を取得するステップ(1)と、

タンパク質予測プログラムを用いて前記複数の候補タンパク質のそれぞれの三次元構造を予測するステップ(2)と、

スコアリング関数を用いて前記複数の候補タンパク質の三次元構造の多重構造アライメントを行い、構造類似度マトリックスを取得するステップ(3)と、

系統樹構築方法によって、前記構造類似度マトリックスに基づいて前記複数の候補タンパク質をクラスタリングするステップ(4)と、を含む。

30

【0030】

幾つかの実施形態では、ステップ(1)において、データベースのアノテーション情報によって前記複数の候補タンパク質の配列を取得する。例えば、デアミナーゼをクラスタリングする場合、データベースから「デアミナーゼ」のアノテーションが付けられた複数の候補タンパク質の配列を選択してもよい。

【0031】

幾つかの実施形態では、ステップ(1)において、基準タンパク質の配列を用い、配列同一性/類似性に基づいてデータベースを検索することによって前記複数の候補タンパク質の配列を取得する。例えば、機能が既知の基準タンパク質の配列に基づいて、BLASTプログラムを用いてデータベースを検索することにより、前記複数の候補タンパク質の配列を得ることができる。幾つかの実施形態では、前記複数の候補タンパク質は、前記基準タンパク質と少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも80%、少なくとも90%、少なくとも95%の配列同一性を有する。

40

【0032】

幾つかの実施形態では、前記候補タンパク質はデアミナーゼである。幾つかの好ましい実施形態では、前記候補タンパク質はシトシンデアミナーゼである。

50

【 0 0 3 3 】

幾つかの実施形態では、前記データベースは I n t e r P r o データベースである。

【 0 0 3 4 】

幾つかの実施形態では、前記ステップ (2) のタンパク質構造予測プログラムは、A l p h a F o l d 2、R o s e T T、またはタンパク質構造を予測できる他のプログラムから選択される (John Jumper and others, ' Highly Accurate Protein Structure Prediction with AlphaFold ', Nature, 596.7873 (2021), 583 - 89) 。

【 0 0 3 5 】

幾つかの実施形態では、前記ステップ (3) で使用されるスコアリング関数には、T M - スコア、R M S D、L D D T、G D T s c o r e、Q S C、F A P E、またはタンパク質構造の類似度をスコアリングできる他のスコアリング関数が含まれる (John Jumper and others, ' Highly Accurate Protein Structure Prediction with AlphaFold ', Nature, 596.7873 (2021), 583 - 89) 。

【 0 0 3 6 】

幾つかの実施形態では、スコアリング関数が T M - スコアである場合、T M - スコアは少なくとも 0 . 6、少なくとも 0 . 7、少なくとも 0 . 7 5、少なくとも 0 . 8、少なくとも 0 . 8 5、またはそれ以上である。例えば、T M - スコアの計算は、本願の実施例の「材料および方法」の箇所に記載されている式および方法を参照することができる。

【 0 0 3 7 】

幾つかの実施形態では、前記ステップ (4) の系統樹構築方法は、算術平均を用いた重み付けなしペアグループ法である (Unweighted Pair Group Method with Arithmetic Mean, UPGMA) (C. P. Kurtzman, Jack W. Fell, and T. Boekhout, The Yeasts: A Taxonomic Study, 5th ed (Amsterdam: Elsevier, 2011). ; ' A Statistical Method for Evaluating Systematic Relationships - Robert Reuven Sokal, Charles Duncan Michener - Google 図書 ') 。

【 0 0 3 8 】

幾つかの実施形態では、ステップ (4) は、前記複数の候補タンパク質のクラスタリング樹状図を取得する。

【 0 0 3 9 】

一態様において本発明は、三次元構造に基づいてタンパク質の機能を予測する方法を提供する。前記方法は、本発明のタンパク質のクラスタリング方法に基づいて複数の候補タンパク質をクラスタリングし、次に、クラスタリング結果に基づいて前記候補タンパク質の機能を予測することを含む。

【 0 0 4 0 】

幾つかの実施形態では、前記複数の候補タンパク質は、機能が既知の少なくとも 1 つの基準タンパク質を含む。

【 0 0 4 1 】

幾つかの実施形態では、同じクレードまたはサブクレードの他の候補タンパク質の機能は、機能が既知の基準タンパク質のクラスター (樹状図) での位置によって予測される。幾つかの実施形態では、基準タンパク質と同じクレードまたはサブクレードに位置する他の候補タンパク質は、前記基準タンパク質と同じまたは類似の機能を有すると予測される。幾つかの実施形態では、前記同じクレードまたはサブクレード内の異なる候補タンパク質間の T M - スコアは、少なくとも 0 . 6、少なくとも 0 . 7、少なくとも 0 . 7 5、少なくとも 0 . 8、少なくとも 0 . 8 5、またはそれより大きい。幾つかの実施形態では、前記異なるクレードまたはサブクレードの候補タンパク質間の T M - スコアは、0 . 8 5 未満、0 . 8 未満、0 . 7 5 未満、0 . 7 未満、0 . 6 未満、またはそれより小さい。

【 0 0 4 2 】

幾つかの実施形態では、前記基準タンパク質はデアミナーゼである。幾つかの好ましい実施形態では、前記基準タンパク質はシトシンデアミナーゼである。幾つかの実施形態では、前記基準タンパク質は基準シトシンデアミナーゼであり、前記基準シトシンデアミナ

ーゼは、配列が配列番号64に示されるrAPOBEC1、または配列が配列番号65に示されるDddAである。幾つかの実施形態では、基準タンパク質と同じクレードまたはサブクレード内の異なる候補タンパク質間のTM-スコア、または基準タンパク質とのTM-スコアは、少なくとも0.7である。幾つかの実施形態では、基準タンパク質とは異なるクレードまたはサブクレード内の候補タンパク質は、基準タンパク質とのTM-スコアが0.7未満である。

【0043】

別の態様では、本発明は、

本発明の方法によってクラスタリングされた、例えば同じクレードまたはサブクレードにクラスタリングされた複数の候補タンパク質の構造をアライメントして、保存されたコア構造を決定するステップa)と、

前記保存されたコア構造を最小機能ドメインとして同定するステップb)と、を含む、三次元構造に基づいてタンパク質の最小機能ドメインを同定する方法を提供する。

【0044】

本明細書で使用される場合、「最小機能ドメイン」とは、全長タンパク質の機能を実質的に維持することができるタンパク質の最小部分を指す。

【0045】

幾つかの実施形態では、前記複数の候補タンパク質は、機能が既知の少なくとも1つの基準タンパク質を含む。

【0046】

幾つかの実施形態では、前記基準タンパク質はデアミナーゼである。幾つかの好ましい実施形態では、前記基準タンパク質はシトシンデアミナーゼである。幾つかの実施形態では、前記基準タンパク質は基準シトシンデアミナーゼであり、前記基準シトシンデアミナーゼは、配列が配列番号64に示されるrAPOBEC1、または配列が配列番号65に示されるDddAである。

【0047】

別の態様では、本発明は、本発明のタンパク質機能予測方法によって同定されたシトシンデアミナーゼを提供する。

【0048】

別の態様では、本発明は、本発明の方法によって同定されたシトシンデアミナーゼの最小機能ドメインを含む、またはそれからなるトランケートシトシンデアミナーゼを提供する。

【0049】

一態様では、本発明は、生物または生物細胞における塩基編集などの遺伝子編集におけるシトシンデアミナーゼまたはトランケートシトシンデアミナーゼの使用をさらに提供する。

【0050】

3. シトシンデアミナーゼとそれを含む塩基編集融合タンパク質

一態様では、本発明は、DNA中のデオキシシチジンのシトシン塩基を脱アミノ化することができるシトシンデアミナーゼを提供する。幾つかの実施形態では、前記シトシンデアミナーゼは細菌に由来する。

【0051】

幾つかの実施形態では、前記シトシンデアミナーゼは、基準シトシンデアミナーゼの三次元構造のAlphaFold2とのTM-スコアが、0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、基準シトシンデアミナーゼのアミノ酸配列と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少

10

20

30

40

50

なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、当該シトシンデアミナーゼは、DNAのデオキシシチジンのシトシン塩基を脱アミノ化する作用を有する。

【0052】

幾つかの実施形態では、前記基準シトシンデアミナーゼは、

- (a) 配列が配列番号64に示されるrAPOBEC1、または
- (b) 配列が配列番号65に示されるDddA、または
- (c) 配列が配列番号4に示されるSdd7である。

【0053】

幾つかの実施形態では、配列が配列番号64に示されるrAPOBEC1のAlphaFold2の三次元構造とのTM-スコアが、0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、配列番号64と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、当該シトシンデアミナーゼは、DNAのデオキシシチジンのシトシン塩基を脱アミノ化する作用を有する。

【0054】

幾つかの実施形態では、配列が配列番号65に示されるDddAのAlphaFold2の三次元構造とのTM-スコアが、0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、配列番号65と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、当該シトシンデアミナーゼは、DNAのデオキシシチジンのシトシン塩基を脱アミノ化する作用を有する。

【0055】

幾つかの実施形態では、配列が配列番号4に示されるSdd7のAlphaFold2の三次元構造とのTM-スコアが、0.6以上、0.7以上、0.75以上、0.8以上、0.85以上であり、配列番号4と20~70%、20~60%、20~50%、20~45%、20~40%、20~35%の配列同一性、または少なくとも20%、少なくとも30%、少なくとも40%、少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%の配列同一性を有するアミノ酸配列を含み、当該シトシンデアミナーゼは、DNAのデオキシシチジンのシトシン塩基を脱アミノ化する作用を有する。

【0056】

幾つかの実施形態では、前記シトシンデアミナーゼは、AID/APOBECクレード、SCP1.201クレード、MafB19クレード、新規AID/APOBEC様クレード、TM1506クレードまたはXOO2897クレードからのものである。

【0057】

本明細書では、前記シトシンデアミナーゼクレードは、Iyer,L.M.,Zhang,D.,Rogozin,I.B.,&Aravind,L.(2011).Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems.Nucleic acids research,39(22),9473-9497に

記載の内容によって決定される。

【0058】

幾つかの実施形態では、前記シトシデアミナーゼはA I D / A P O B E Cクレードからのものであり、配列番号1と少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。

【0059】

幾つかの実施形態では、前記シトシデアミナーゼはS C P 1 . 2 0 1クレードからのものであり、配列番号28~40のうちのいずれかと少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。幾つかの実施形態では、前記シトシデアミナーゼは、二本鎖DNAのシトシン塩基を脱アミノ化することができる。幾つかの実施形態では、前記シトシデアミナーゼのアミノ酸配列は、配列番号28~40のいずれかのアミノ酸配列からなる。幾つかの実施形態では、前記シトシデアミナーゼのアミノ酸配列は、配列番号28、33、34、35のいずれかのアミノ酸配列からなる。

10

20

【0060】

幾つかの実施形態では、前記シトシデアミナーゼはS C P 1 . 2 0 1クレードからのものであり、配列番号2~18、41~49のいずれかと少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。幾つかの実施形態では、前記シトシデアミナーゼは、一本鎖DNAのシトシン塩基を脱アミノ化することができる。幾つかの実施形態では、前記シトシデアミナーゼのアミノ酸配列は、配列番号2~18、41~49のうちのいずれかのアミノ酸配列からなる。幾つかの実施形態では、前記シトシデアミナーゼのアミノ酸配列は、配列番号2~7、12、17のうちのいずれかのアミノ酸配列からなる。

30

【0061】

幾つかの実施形態では、前記シトシデアミナーゼは、DNAのデオキシシチジンのシトシン塩基を脱アミノ化することができるランケートシトシデアミナーゼである。幾つかの実施形態では、前記ランケートシトシデアミナーゼの長さの範囲は130~160アミノ酸である。幾つかの実施形態では、前記ランケートシトシデアミナーゼは、AAV粒子内に個別にパッケージングされ得る。

【0062】

幾つかの実施形態では、前記ランケートシトシデアミナーゼは、配列番号50~55のうちのいずれかと少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。幾つかの実施形態では、前記ランケートシトシデアミナーゼは、一本鎖DNAのシトシン塩基を脱アミノ化することができる。幾つかの実施形態では、前記ランケートシトシデアミナーゼは、配列番号50~55のうちのいずれかのアミノ酸配列からなる。

40

【0063】

幾つかの実施形態では、前記シトシデアミナーゼはM a f B 1 9クレードからのもの

50

であり、配列番号19、56、57、58のうちのいずれかと少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。

【0064】

幾つかの実施形態では、前記シトシンデアミナーゼは新規AID/APOBEC様クレードからのものであり、配列番号20、21のうちのいずれかと少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。

10

【0065】

幾つかの実施形態では、前記シトシンデアミナーゼはTM1506クレードからのものであり、配列番号22と少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。

【0066】

幾つかの実施形態では、前記シトシンデアミナーゼはXOO2897クレードからのものであり、配列番号23、24、59~62と少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。

20

【0067】

幾つかの実施形態では、前記シトシンデアミナーゼはトキシデアミナーゼクレードからのものであり、配列番号74または75と少なくとも50%、少なくとも60%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%、少なくとも91%、少なくとも92%、93%、少なくとも94%、少なくとも95%、少なくとも96%、少なくとも97%、少なくとも98%、少なくとも99%、さらには100%の配列同一性を有するアミノ酸配列を含む。

30

【0068】

一態様では、本願は、生物または生物の細胞における塩基編集などの遺伝子編集における本発明のシトシンデアミナーゼの使用に関する。

【0069】

幾つかの実施形態では、前記シトシンデアミナーゼは、生物または生物の細胞で塩基編集を行うための塩基編集融合タンパク質または塩基編集システムを製造するために使用される。

40

【0070】

別の態様では、本発明は、核酸ターゲティングドメインおよびシトシン脱アミノ化ドメインを含む塩基編集融合タンパク質を提供し、前記シトシン脱アミノ化ドメインは、少なくとも1つ（例えば、1つまたは2つ）の本発明のシトシンデアミナーゼポリペプチドを含む。

【0071】

本明細書の実施形態では、「融合タンパク質」、「塩基編集融合タンパク質」および「塩基エディター」は、互換的に使用され、配列特異的にゲノム内の標的配列を媒介することができる1つまたは複数のヌクレオチド置換を持つタンパク質を指す。前記1つまたは複数のヌクレオチド置換は、例えば、CからTへの置換である。

50

【0072】

本明細書で使用される場合、「核酸ターゲティングドメイン」とは、配列特異的に（例えば、ガイドRNAによって）ゲノム内の特定の標的配列への塩基編集融合タンパク質の付着を媒介できるドメインを指す。幾つかの実施形態では、前記核酸ターゲティングドメインは、特定の標的配列を対象とする1つまたは複数のジンクフィンガータンパク質ドメイン（ZFP）または転写因子エフェクタードメイン（TALE）を含んでもよい。幾つかの実施形態では、前記核酸ターゲティングドメインは、少なくとも1つ（例えば、1つ）のCRISPRエフェクタータンパク質（CRISPR effector）ポリペプチドを含む。

【0073】

「ジンクフィンガータンパク質ドメイン（ZFP）」は、通常3～6個の個別のジンクフィンガーリピートを含み、それぞれのジンクフィンガーリピートは例えば3bpの独自の配列を認識することができる。異なるジンクフィンガーリピートを組み合わせることで、異なるゲノム配列を標的にすることができる。

【0074】

「転写活性化因子様エフェクタードメイン」は、転写活性化因子様エフェクター（TALE）のDNA結合ドメインである。TALEは、工学的に操作されると、ほぼ任意のDNA配列に結合することができる。

【0075】

本明細書で使用される場合、「CRISPRエフェクタータンパク質」という用語は、通常、天然に存在するCRISPRシステムに存在するヌクレアーゼ（CRISPRヌクレアーゼ）またはその機能的バリエーションを指す。この用語は、細胞内で配列特異的なターゲティングを可能にするCRISPRシステムに基づくあらゆるエフェクタータンパク質を包摂する。

【0076】

本明細書で使用される場合、CRISPRヌクレアーゼに関する「機能的バリエーション」は、ガイドRNAによって媒介される配列特異的なターゲティング能力を少なくとも保持していることを意味する。好ましくは、前記機能的バリエーションはヌクレアーゼ不活性バリエーション、すなわち二本鎖核酸切断活性を欠いている。しかし、二本鎖核酸切断活性を欠いているCRISPRヌクレアーゼには、二本鎖核酸分子にニック（nick）を形成するが二本鎖核酸を完全には切断しないニッカーゼ（nickase）も含まれる。本発明の幾つかの好ましい実施形態では、本発明に記載のCRISPRエフェクタータンパク質はニッカーゼ活性を有する。幾つかの実施形態では、前記機能的バリエーションは、野生型ヌクレアーゼとは異なるPAM（プロトスペーサー隣接モチーフ）配列を認識する。

【0077】

「CRISPRエフェクタータンパク質」は、Cas9ヌクレアーゼまたはその機能的バリエーションを含むCas9ヌクレアーゼに由来してもよい。前記Cas9ヌクレアーゼは、化膿性連鎖球菌（*S. pyogenes*）由来のspCas9または黄色ブドウ球菌（*S. aureus*）由来のSaCas9といった、異なる種に由来するCas9ヌクレアーゼであってもよい。「Cas9ヌクレアーゼ」および「Cas9」は、本明細書では互換的に使用され、Cas9タンパク質またはその断片（例えば、Cas9の活性DNA切断ドメインおよび/またはCas9のgRNA結合ドメインを含むタンパク質）を含むRNA誘導型ヌクレアーゼを指す。Cas9は、CRISPR/Cas（クラスター化された規則的に間隔をあけられた短い回文リピートおよびそれに関連するシステム）ゲノム編集システムのコンポーネントであり、ガイドRNAの誘導の下でDNA標的配列を標的にして切断し、DNA二本鎖切断（DSB）を形成できる。野生型SpCas9の例示的なアミノ酸配列を配列番号25に示す。

【0078】

「CRISPRエフェクタータンパク質」は、Cpf1ヌクレアーゼまたはその機能的バリエーションを含むCpf1ヌクレアーゼに由来してもよい。前記Cpf1ヌクレアーゼはFrancisella novicida U112、Acidaminococcus

10

20

30

40

50

s p . B V 3 L 6 および L a c h n o s p i r a c e a e b a c t e r i u m N D 2 0 0 6 に由来する C p f 1 ヌクレアーゼといった、異なる種に由来する C p f 1 ヌクレアーゼであってもよい。

【0079】

利用可能な「C R I S P R エフェクタータンパク質」は、C a s 3、C a s 8 a、C a s 5、C a s 8 b、C a s 8 c、C a s 1 0 d、C s e 1、C s e 2、C s y 1、C s y 2、C s y 3、G S U 0 0 5 4、C a s 1 0、C s m 2、C m r 5、C a s 1 0、C s x 1 1、C s x 1 0、C s f 1、C s n 2、C a s 4、C 2 c 1 (C a s 1 2 b)、C 2 c 3、C 2 c 2、C a s 1 2 c、C a s 1 2 d (すなわち、C a s Y)、C a s 1 2 e (すなわち、C a s X)、C a s 1 2 f (すなわち、C a s 1 4)、C a s 1 2 g、C a s 1 2 h、C a s 1 2 i、C a s 1 2 j (すなわち、C a s)、C a s 1 2 k、C a s 1 2 l、C a s 1 2 m などのヌクレアーゼに由来してもよく、例えば、これらのヌクレアーゼまたはその機能的バリエーションが含まれる。

10

【0080】

幾つかの実施形態では、前記 C R I S P R エフェクタータンパク質は、ヌクレアーゼ不活性 C a s 9 である。C a s 9 ヌクレアーゼの D N A 切断ドメインには、H N H ヌクレアーゼサブドメインと R u v C サブドメインの2つのサブドメインが含まれることが知られている。H N H サブドメインは g R N A に相補的な鎖を切断するが、R u v C サブドメインは非相補鎖を切断する。これらのサブドメインの突然変異により、C a s 9 のヌクレアーゼ活性が不活性化され、「ヌクレアーゼ不活性 C a s 9」を形成することができる。前記ヌクレアーゼ不活性 C a s 9 は、g R N A 誘導性 D N A 結合能力を保持している。

20

【0081】

本発明に記載のヌクレアーゼ不活性 C a s 9 は、化膿性連鎖球菌 (S . p y o g e n e s) C a s 9 (S p C a s 9)、または黄色ブドウ球菌 (S . a u r e u s) C a s 9 (S a C a s 9) といった、異なる種の C a s 9 に由来してもよい。C a s 9 の H N H ヌクレアーゼサブドメインと R u v C サブドメインを同時に突然変異させると (例えば、突然変異 D 1 0 A と H 8 4 0 A を含む)、C a s 9 ヌクレアーゼは不活性になり、ヌクレアーゼデッド C a s 9 (d C a s 9) になる。一方のサブドメインが突然変異的に不活性化されると、C a s 9 は、ニッカーゼ活性を有し、すなわち、C a s 9 ニッカーゼ (n C a s 9) が得られ、例えば突然変異 D 1 0 A のみを持つ n C a s 9 が得られる。

30

【0082】

したがって、本発明の各態様の幾つかの実施形態では、本発明に記載のヌクレアーゼ不活性 C a s 9 バリエーションは、野生型 C a s 9 と比較してアミノ酸置換 D 1 0 A および / または H 8 4 0 A を含み、アミノ酸番号は配列番号 2 5 に示される。幾つかの好ましい実施形態では、前記ヌクレアーゼ不活性 C a s 9 は、野生型 C a s 9 と比較してアミノ酸置換 D 1 0 A を含み、アミノ酸番号は配列番号 2 5 に示される。幾つかの実施形態では、前記ヌクレアーゼ不活性 C a s 9 は、配列番号 2 6 に記載のアミノ酸配列 (n C a s 9 (D 1 0 A)) を含む。

【0083】

C a s 9 ヌクレアーゼを遺伝子編集に使用する場合、通常、標的配列の 3 ' 末端に 5 ' - N G G - 3 ' の P A M (プロトスペーサー隣接モチーフ) 配列が必要である。しかしながら本発明者らは、驚くべきことに、イネなどの幾つかの種におけるこの P A M 配列の出現頻度が非常に低く、それがイネなどのこれらの種における遺伝子編集を著しく制限することを見出した。そのため本発明において好ましくは、異なる P A M 配列を有する C a s 9 ヌクレアーゼの機能的バリエーションといった、異なる P A M 配列を認識する C R I S P R エフェクタータンパク質が使用される。

40

【0084】

本発明の幾つかの実施形態では、融合タンパク質のシチジン脱アミノ化ドメインは、融合タンパク質 - ガイド R N A - D N A 複合体の形成において生成された一本鎖 D N A のシチジン脱アミノ化を U に変換し、さらに塩基ミスマッチ修復によって C から T への塩基置

50

換を実現する。

【0085】

本発明の幾つかの実施形態では、前記核酸ターゲティングドメインと前記シトシン脱アミノ化ドメインはリンカーによって融合される。

【0086】

本明細書で使用される場合、「リンカー」は、長さが1~50（例えば、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20または20~25、25~50）またはさらに多くのアミノ酸、二次構造以上の構造を持たない非機能的アミノ酸配列であってもよい。例えば、前記リンカーはフレキシブルリンカーであってもよい。

【0087】

幾つかの実施形態では、前記塩基編集融合タンパク質は、N末端からC末端に向かって、シトシン脱アミノ化ドメイン、核酸ターゲティングドメインをこの順で含む。

【0088】

さらに、細胞では、ウラシルDNAグリコシラーゼはDNAからのUの除去を触媒し、塩基除去修復（BER）を開始し、U：GからC：Gへの修復を引き起こす。したがって、いかなる理論にも束縛されるものではないが、本発明の塩基編集融合タンパク質は、ウラシルDNAグリコシラーゼ阻害剤（UGI）と組み合わせて、CからTへの塩基編集の効率を高めることができる。

【0089】

幾つかの実施形態では、前記塩基編集融合タンパク質は、ウラシルDNAグリコシラーゼ阻害剤（UGI）と共発現する。

【0090】

幾つかの実施形態では、前記塩基編集融合タンパク質は、ウラシルDNAグリコシラーゼ阻害剤（UGI）をさらに含む。

【0091】

幾つかの実施形態では、UGIは、リンカーを介して前記塩基編集融合タンパク質の他の部分に連結される。

【0092】

幾つかの実施形態では、UGIは、「自己切断ペプチド」を介して前記塩基編集融合タンパク質の他の部分に連結される。

【0093】

本明細書で使用される「自己切断ペプチド」とは、細胞内で自己切断を実現することができるペプチドを意味する。例えば、前記自己切断ペプチドはプロテアーゼ認識部位を含むことができ、それにより細胞内のプロテアーゼによって認識され、特異的に切断される。あるいは前記自己切断ペプチドは2Aポリペプチドであってもよい。2Aポリペプチドは、翻訳中に自己切断が起こるウイルス由来の短いペプチドである。2Aポリペプチドを用いて2つの異なる標的ポリペプチドを連結し、同じリーディングフレームで発現させると、2つの標的ポリペプチドがほぼ1：1の比率で生成される。一般的に使用される2Aポリペプチドは、豚テコウイルス-1（porcine techovirus-1）由来のP2A、トローセア・アシグナウイルス（*Thosea asigna virus*）由来のT2A、馬鼻炎Aウイルス（equine rhinitis A virus）由来のE2Aおよび口蹄疫ウイルス（foot-and-mouth disease virus）由来のF2Aである。これらの2Aポリペプチドの様々な機能的バリエーションも当技術分野で公知であり、本発明で使用することができる。

【0094】

好ましくは、前記自己切断ペプチドは、前記核酸ターゲティングドメイン、前記シトシン脱アミノ化ドメインの間またはその内部に存在しない。幾つかの実施形態では、UGIは、前記塩基編集融合タンパク質のN末端またはC末端、好ましくはC末端に位置する。

【0095】

幾つかの具体的な実施形態では、前記ウラシルDNAグリコシラーゼ阻害剤（UGI）

10

20

30

40

50

は、配列番号 27 に示されるアミノ酸配列を含む。

【0096】

本発明の幾つかの実施形態では、本発明の融合タンパク質は、核局在化配列 (NLS) をさらに含んでもよい。一般に、前記融合タンパク質中の 1 つまたは複数の NLS は、その塩基編集機能を実現可能にする量の前記融合タンパク質の細胞核への蓄積を促進するのに十分な強度を持っている必要がある。一般に、核局在化活性の強さは、前記融合タンパク質中の NLS の数、位置、使用される 1 つもしくは複数の特定の NLS、またはこれらの要因の組み合わせによって決まる。

【0097】

本発明の幾つかの実施形態では、本発明の融合タンパク質の NLS は、N 末端および / または C 末端に位置してもよい。本発明の幾つかの実施形態では、本発明の融合タンパク質の NLS は、前記アデニン脱アミノ化ドメイン、シトシン脱アミノ化ドメイン、核酸ターゲティングドメインおよび / または UGI の間に位置してもよい。幾つかの実施形態では、前記融合タンパク質は、約 1、2、3、4、5、6、7、8、9、10 またはそれより多くの NLS を含む。幾つかの実施形態では、前記融合タンパク質は、N 末端または N 末端付近に位置する約 1、2、3、4、5、6、7、8、9、10 またはそれより多くの NLS を含む。幾つかの実施形態では、前記融合タンパク質は、C 末端または C 末端付近に位置する約 1、2、3、4、5、6、7、8、9、10 またはそれより多くの NLS を含む。幾つかの実施形態では、前記ポリペプチドはこれらの組み合わせを含み、例えば N 末端に 1 つまたは複数の NLS および C 末端に 1 つまたは複数の NLS を含む。複数の NLS が存在する場合、それぞれを他の NLS から独立して選択できる。

【0098】

一般に、NLS は、タンパク質表面に露出した正に荷電したリジンまたはアルギニンの 1 つまたは複数の短い配列で構成されるが、他のタイプの NLS も知られている。NLS の非限定的な例としては、KKRKKV、PKKKRKKV または KRPAATKKAGQA KKKK が含まれる。

【0099】

さらに、編集を必要とする DNA の位置に応じて、本発明の融合タンパク質は、細胞質局在化配列、葉緑体局在化配列、ミトコンドリア局在化配列などの他の局在化配列を含んでもよい。

【0100】

4. 塩基編集システム

別の態様では、本発明は、i) 本発明のシトシンデアミナーゼまたは塩基編集融合タンパク質、および / または前記シトシンデアミナーゼまたは塩基編集融合タンパク質をコードするヌクレオチド配列を含む発現構築物を含む塩基編集システムを提供する。

【0101】

幾つかの実施形態では、前記塩基編集システムは、核酸標的領域を改変するために使用される。

【0102】

幾つかの実施形態では、前記塩基編集システムは、ii) 少なくとも 1 種のガイド RNA および / または前記少なくとも 1 種のガイド RNA をコードするヌクレオチド配列を含む少なくとも 1 種の発現構築物をさらに含む。しかしながら当業者であれば、前記塩基編集融合タンパク質が CRISPR エフェクタータンパク質に基づいていない場合、前記システムはガイド RNA またはそれをコードする発現構築物を必要としない可能性があることを理解するであろう。

【0103】

幾つかの実施形態では、前記少なくとも 1 種のガイド RNA は、前記融合タンパク質の核酸ターゲティングドメインに結合可能である。幾つかの実施形態では、前記ガイド RNA は、前記核酸標的領域内の少なくとも 1 つの標的配列を対象とする。

【0104】

10

20

30

40

50

本明細書で使用される場合、「塩基編集システム」とは、細胞または生物におけるゲノム配列といった核酸配列の塩基編集に必要なコンポーネントの組み合わせを指す。シトシンデアミナーゼ、塩基編集融合タンパク質、および1種または複数種のガイドRNAといった前記システムの各コンポーネントは、独立して存在することも、組成物として任意の組み合わせで存在することもできる。

【0105】

幾つかの実施形態では、それは、本発明のシトシンデアミナーゼまたは本発明の融合タンパク質と、核酸ターゲティング結合タンパク質に結合することができるガイドRNAと、を含む。

【0106】

本明細書で使用される場合、「ガイドRNA」および「gRNA」は、互換的に使用され、CRISPRエフェクタータンパク質と複合体を形成することができ、かつ標的配列とある程度の同一性を有するため前記複合体を標的配列にターゲティングすることができるRNA分子を指す。ガイドRNAは、標的配列の相補鎖との塩基ペ어링によって前記標的配列を標的にする。例えば、Cas9ヌクレアーゼまたはその機能的バリエーションに使用されるgRNAは、通常、複合体の形成を部分的に相補するcrRNAおよびtracrRNA分子で構成され、ここで、crRNAは、標的配列と十分な同一性を有することで当該標的配列の相補鎖にハイブリダイゼーションし、かつCRISPR複合体(Cas9+crRNA+tracrRNA)が当該標的配列に特異的に結合するように誘導するガイド配列(シード配列とも呼ばれる)を含む。しかしながら、crRNAとtracrRNAの両方の特徴を含むシングルガイドRNA(sgRNA)を設計できることは当該技術分野で知られている。Cpf1ヌクレアーゼまたはその機能的バリエーションに使用されるgRNAは、通常、成熟crRNA分子のみで構成され、sgRNAとも呼ばれる。使用されるCRISPRヌクレアーゼおよび編集対象の標的配列に基づいて適切なgRNAを設計することは、当業者の能力の範囲内にある。

【0107】

幾つかの実施形態では、前記ガイドRNAの長さは15~100ヌクレオチドであり、標的配列に相補的な少なくとも10個、少なくとも15個または少なくとも20個の連続ヌクレオチド配列を含む。

【0108】

幾つかの実施形態では、前記ガイドRNAは、標的配列に相補的な15~40個の連続ヌクレオチド配列を含む。

【0109】

幾つかの実施形態では、前記ガイドRNAの長さは15~50ヌクレオチドである。

【0110】

幾つかの実施形態では、前記標的配列はDNA配列である。

【0111】

幾つかの実施形態では、前記標的配列は生物のゲノム内にある。幾つかの実施形態では、前記生物は原核生物である。幾つかの実施形態では、前記原核生物は細菌である。幾つかの実施形態では、前記生物は真核生物である。幾つかの実施形態では、前記生物は植物または真菌である。幾つかの実施形態では、前記生物は脊椎動物である。幾つかの実施形態では、前記脊椎動物は哺乳動物である。幾つかの実施形態では、前記哺乳動物はマウス、ラットまたはヒトである。幾つかの実施形態では、前記生物は細胞である。幾つかの実施形態では、前記細胞はマウス細胞、ラット細胞またはヒト細胞である。幾つかの実施形態は、前記細胞はHEK-293細胞である。

【0112】

幾つかの実施形態では、本発明の塩基編集システムが細胞に導入された後、前記塩基編集融合タンパク質と前記ガイドRNAは複合体を形成することができ、この複合体はガイドRNAの媒介の下で標的配列を特異的にターゲティングし、標的配列内で1つまたは複数のCがTに置換され、および/または1つまたは複数のAがGに置換される。

10

20

30

40

50

【0113】

幾つかの実施形態では、前記少なくとも1種のガイドRNAは、ゲノム中の標的核酸領域内のセンス鎖（例えば、タンパク質コード鎖）および/またはアンチセンス鎖上に位置する標的配列を対象としてもよい。ガイドRNAがセンス鎖（例えば、タンパク質コード鎖）を標的にする場合、本発明の塩基編集組成物は、センス鎖（例えば、タンパク質コード鎖）上の標的配列内の1つまたは複数のCをTで置換させ、および/または1つまたは複数のAをGで置換させることができる。ガイドRNAがアンチセンス鎖を標的にする場合、本発明の塩基編集組成物は、センス鎖（例えば、タンパク質コード鎖）上の標的配列内の1つまたは複数のGをAで置換させ、および/または1つまたは複数のTをCで置換させることができる。

10

【0114】

細胞内で効率的な発現を得るために、本発明の幾つかの実施形態では、前記シトシンデアミナーゼまたは塩基編集融合タンパク質をコードするヌクレオチド配列は、ゲノムが改変される生物に対してコドン最適化される。

【0115】

コドン最適化は、天然配列の少なくとも1つのコドン（例えば、約1、2、3、4、5、10、15、20、25、50またはそれより多くのコドン）を、宿主細胞の遺伝子においてより頻繁に、または最も頻繁に使用されるコドンで置換するとともに、この天然アミノ酸配列を維持し、標的宿主細胞における発現を増強するために核酸配列を改変することを指す。種が異なれば、特定のアミノ酸の特定のコドンに対して特定の選好性が示される。コドンの選好性（生物間のコドン使用の違い）はメッセンジャーRNA（mRNA）の翻訳効率に関係していることが多く、該翻訳効率は翻訳されるコドンの性質と特定のトランスファーRNA（tRNA）分子の利用可能性に依存すると考えられている。細胞内の選択されたtRNAの優勢は、通常、ペプチド合成に最も頻繁に使用されるコドンを反映する。したがって、コドン最適化に基づく特定の生体内で最適な遺伝子発現が得られるように遺伝子を調整できる。コドン使用頻度表は、例えばwww.kazusa.or.jp/codon/で入手可能なコドン使用頻度データベース（「Codon Usage Database」）で入手するなど、容易に入手することができ、これらの表は様々な方法で調整してもよい（Nakamura Y.ら, "Codon usage tabulated from the international DNA sequence databases: status for the year 2000. Nucl. Acids Res., 28:292(2000)参照）。

20

30

【0116】

本発明の塩基編集システムによってゲノムを改変できる生物には、塩基編集に適した任意の生物、好ましくは真核生物が含まれる。生物の例としては、ヒト、マウス、ラット、サル、犬、豚、羊、牛、猫などの哺乳動物；ニワトリ、アヒル、ガチョウなどの家禽；および単子葉植物と双子葉植物を含む植物、例えば、小麦、イネ、トウモロコシ、大豆、ヒマワリ、ソルガム、アブラナ、アルファルファ、綿、大麦、キビ、サトウキビ、トマト、タバコ、キャッサバおよびジャガイモを含むがこれらに限定されない作物植物が含まれるが、これらに限定されない。

【0117】

5. 塩基編集方法

別の態様では、本発明は、本発明の塩基編集システムを核酸分子の標的配列に接触させることを含む、塩基編集方法を提供する。

40

【0118】

幾つかの実施形態では、前記核酸分子はDNA分子である。幾つかの好ましい実施形態では、前記核酸分子は二本鎖DNA分子または一本鎖DNA分子である。

【0119】

幾つかの実施形態では、前記核酸分子の標的配列は、植物の形質または発現に関連する配列を含む。

【0120】

幾つかの実施形態では、前記核酸分子の標的配列は、疾患または障害に関連する配列ま

50

たは点突然変異を含む。

【0121】

幾つかの実施形態では、前記塩基編集システムは、核酸分子の標的配列と接触して脱アミノ化作用を発揮し、前記脱アミノ化作用により標的配列の1つまたは複数のヌクレオチドの置換が起こる。

【0122】

幾つかの実施形態では、前記標的配列は、DNA配列5'-MCN-3'を含み、MはA、T、CまたはGであり、NはA、T、CまたはGであり、5'-MCN-3'配列の中央にあるCが脱アミノ化される。

【0123】

幾つかの実施形態では、前記脱アミノ化作用によりスプライス部位が導入または除去される。

【0124】

幾つかの実施形態では、前記脱アミノ化作用により、遺伝子プロモーターに突然変異が導入され、前記突然変異により前記遺伝子プロモーターに操作可能に連結された遺伝子の転写が増加または減少する。

【0125】

幾つかの実施形態では、前記脱アミノ化作用により、遺伝子リプレッサーに突然変異が導入され、前記突然変異により、前記遺伝子リプレッサーに操作可能に連結された遺伝子の転写が増加または減少する。

【0126】

幾つかの実施形態では、前記接触は生体内で行われる。

【0127】

幾つかの実施形態では、前記接触は生体外で行われる。

【0128】

6. 遺伝子改変細胞の産生方法

別の態様では、本発明は、また、少なくとも1つの遺伝子改変細胞を産生する方法であって、本発明の塩基編集システムを少なくとも1つの前記細胞に導入し、それによって前記少なくとも1つの細胞の標的核酸領域内の1つまたは複数のヌクレオチド置換を起こすことを含む、方法を提供する。幾つかの実施形態では、前記1つまたは複数のヌクレオチド置換は、CからTへの置換である。

【0129】

幾つかの実施形態では、前記方法は、1つまたは複数の所望のヌクレオチド置換を有する細胞を前記少なくとも1つの細胞からスクリーニングすることをさらに含む。

【0130】

幾つかの実施形態では、本発明の方法は生体外で行われる。例えば、前記細胞は、単離された細胞、または単離された組織もしくは器官内の細胞である。

【0131】

別の態様では、本発明は、また、本発明の方法によって産生される遺伝子改変細胞またはその子孫細胞を含む遺伝子改変生物を提供する。好ましくは、前記遺伝子改変細胞またはその子孫細胞は、1つまたは複数の所望のヌクレオチド置換を有する。

【0132】

本発明では、改変される標的核酸領域は、タンパク質をコードする遺伝子などの機能遺伝子内など、ゲノム内のどこに位置してもよいし、プロモーター領域またはエンハンサー領域などの遺伝子発現調節領域に位置してもよい。それによって、前記遺伝子機能の改変または遺伝子発現の改変を実現する。幾つかの実施形態では、前記所望のヌクレオチド置換により、遺伝子機能または遺伝子発現の所望の改変がもたらされる。

【0133】

幾つかの実施形態では、前記標的核酸領域は、前記細胞または生物の形質に関連する。幾つかの実施形態では、前記標的核酸領域における突然変異は、前記細胞または生物の形

10

20

30

40

50

質の変化をもたらす。幾つかの実施形態では、前記標的核酸領域はタンパク質のコード領域に位置する。幾つかの実施形態では、前記標的核酸領域は、タンパク質の機能に関連するモチーフまたはドメインをコードする。幾つかの好ましい実施形態では、前記標的核酸領域における1つまたは複数のヌクレオチド置換は、前記タンパク質のアミノ酸配列におけるアミノ酸置換をもたらす。幾つかの実施形態では、前記1つまたは複数のヌクレオチド置換は、タンパク質の機能の変化をもたらす。

【0134】

本発明の方法では、前記塩基編集システムは、当業者に周知の様々な方法によって細胞に導入することができる。

【0135】

本発明の塩基編集システムを細胞に導入するために使用できる方法には、リン酸カルシウムトランスフェクション、プロトプラスト融合、エレクトロポレーション、リポフェクタミントランスフェクション、マイクロインジェクション、ウイルス感染（例えば、バキュロウイルス、ワクシニアウイルス、アデノウイルス、アデノ随伴ウイルス、レンチウイルスおよびその他のウイルス）、遺伝子銃法、PEG媒介プロトプラスト形質転換、アグロバクテリウム・ツメファシエンス媒介形質転換などが含まれるが、これらに限定されない。

【0136】

本発明の方法によって塩基編集が可能な細胞は、例えば、ヒト、マウス、ラット、サル、犬、豚、羊、牛、猫といった哺乳動物；ニワトリ、アヒル、ガチョウといった家禽；および単子葉植物と双子葉植物を含む植物、好ましくは、小麦、イネ、トウモロコシ、大豆、ヒマワリ、ソルガム、アブラナ、アルファルファ、綿、大麦、キビ、サトウキビ、トマト、タバコ、キャッサバおよびジャガイモを含むがこれらに限定されない作物植物に由来してもよい。

【0137】**7. 植物における使用**

本発明の塩基編集融合タンパク質、塩基編集システム、および遺伝子改変細胞を産生する方法は、植物の遺伝子改変に特に適している。好ましくは、前記植物は、小麦、イネ、トウモロコシ、大豆、ヒマワリ、ソルガム、アブラナ、アルファルファ、綿、大麦、キビ、サトウキビ、トマト、タバコ、キャッサバおよびジャガイモを含むがこれらに限定されない作物植物である。より好ましくは、前記植物はイネである。

【0138】

別の態様では、本発明は、本発明の塩基編集システムを少なくとも1つの前記植物に導入し、それによって前記少なくとも1つの植物のゲノム中の標的核酸領域内の1つまたは複数のヌクレオチド置換を起こすことを含む、遺伝子改変植物を産生する方法を提供する。

【0139】

幾つかの実施形態では、前記方法は、1つまたは複数の所望のヌクレオチド置換を有する植物を前記少なくとも1つの植物からスクリーニングすることをさらに含む。

【0140】

本発明の方法では、前記塩基編集組成物は、当業者に周知の様々な方法によって植物に導入することができる。本発明の塩基編集システムを植物に導入するために使用することができる方法には、遺伝子銃法、PEG媒介プロトプラスト形質転換、アグロバクテリウム・ツメファシエンス媒介形質転換、植物ウイルス媒介形質転換、花粉管チャンネル法および子房注射法が含まれるが、これらに限定されない。好ましくは、前記塩基編集組成物は一過性形質転換によって植物に導入される。

【0141】

本発明の方法では、前記塩基編集融合タンパク質とガイドRNAを植物細胞に導入または産生するだけで標的配列の改変を実現することができ、また、前記塩基編集システムのコンポーネントをコードする外因性ポリヌクレオチドで植物を安定的に形質転換せずとも

10

20

30

40

50

、改変は安定して継承され得る。それによって、安定して存在する（継続的に生成される）塩基編集組成物の潜在的なオフターゲット効果が回避され、植物ゲノムへの外因性ヌクレオチド配列の組み込みも回避され、それによってより高い生物学的安全性が得られる。

【0142】

幾つかの好ましい実施形態では、前記導入は選択圧の非存在下で行われ、それによって外因性ヌクレオチド配列の植物ゲノムへの組み込みが回避される。

【0143】

幾つかの実施形態では、前記導入は、本発明の塩基編集システムを単離された植物細胞または組織に形質転換し、次いで形質転換された前記植物細胞または組織を無傷の植物に再生することを含む。好ましくは、前記再生は、選択圧の非存在下で行われ、すなわち、組織培養中に発現ベクター上に担持される選択遺伝子に対するいかなる選択剤も使用せずに行われる。選択剤を使用しないことにより、植物の再生効率が向上し、外因性ヌクレオチド配列を含まない改変植物を得ることができる。

【0144】

他の幾つかの実施形態では、本発明の塩基編集システムは、葉、茎頂、花粉管、若い穂または胚軸などの無傷の植物の特定の部分に形質転換することができる。これは、組織培養では再生が難しい植物の形質転換に特に適している。

【0145】

本発明の幾つかの実施形態では、生体外発現タンパク質および/または生体外転写RNA分子（例えば、前記発現構築物は生体外転写RNA分子である）を前記植物に直接形質転換する。前記タンパク質および/またはRNA分子は植物細胞内で塩基編集を行うことができ、その後細胞によって分解され、植物ゲノムへの外因性ヌクレオチド配列の組み込みを回避する。

【0146】

したがって、幾つかの実施形態では、本発明の方法を用いた植物の遺伝子改変および育種は、そのゲノムに外因性ポリヌクレオチドの組み込みがない植物、すなわち導入遺伝子フリー（transgene-free）改変植物を得ることができる。

【0147】

本発明の幾つかの実施形態では、前記改変された標的核酸領域は、農業形質などの植物形質に関連しており、それにより、前記1つまたは複数のヌクレオチド置換により、前記植物は、野生型植物と比較して変化した（好ましくは、改善した）形質、例えば農学的形質を有する。

【0148】

幾つかの実施形態では、前記方法は、1つまたは複数の所望のヌクレオチド置換および/または農業形質などの所望の形質を有する植物をスクリーニングするステップをさらに含む。

【0149】

本発明の幾つかの実施形態では、前記方法は、前記遺伝子改変植物の子孫を取得することをさらに含む。好ましくは、前記遺伝子改変植物またはその子孫は、1つまたは複数の所望のヌクレオチド置換および/または農業形質などの所望の形質を有する。

【0150】

別の態様では、本発明は、本発明の上記の方法によって得られる、遺伝子改変植物またはその子孫もしくはその一部をさらに提供する。幾つかの実施形態では、前記遺伝子改変植物またはその子孫もしくはその一部は、導入遺伝子フリーである。好ましくは、前記遺伝子改変植物またはその子孫は、所望の遺伝子改変および/または農業形質などの所望の形質を有する。

【0151】

別の態様では、本発明は、上記の本発明の方法によって得られた標的核酸領域に1つまたは複数のヌクレオチド置換を含む遺伝子改変された第1植物を、前記1つまたは複数のヌクレオチド置換を含まない第2植物とハイブリダイゼーションし、それによって前記1

10

20

30

40

50

つまたは複数のヌクレオチド置換を第2植物に導入することを含む、植物育種方法をさらに提供する。好ましくは、前記遺伝子改変された第1植物は、農業形質などの所望の形質を有する。

【0152】

8. 治療における使用

本発明は、疾患治療における本発明の塩基編集システムの使用をさらに包含する。

【0153】

本発明の塩基編集システムによって疾患関連遺伝子を改変することにより、疾患関連遺伝子の上方制御、下方制御、不活性化、活性化、突然変異の修正などを実現することができる。例えば、本発明に記載の標的核酸領域は、疾患関連遺伝子のタンパク質コード領域に位置するか、またはプロモーター領域やエンハンサー領域といった遺伝子発現調節領域に位置することによって前記疾患関連遺伝子の機能または疾患関連遺伝子の発現を改変することができる。したがって、本明細書に記載の疾患関連遺伝子の改変には、疾患関連遺伝子自体（例えば、タンパク質コード領域）の改変だけでなく、その発現調節領域（例えば、プロモーター、エンハンサー、イントロンなど）の改変も含まれる。

【0154】

「疾患関連」遺伝子は、非疾患対照組織または細胞と比較して、疾患影響を受けた組織に由来する細胞において異常なレベルまたは異常な形態で転写または翻訳産物を産生する任意の遺伝子を指す。発現の変化が疾患の出現および/または進行に関連する場合、それは、異常に高いレベルで発現される遺伝子である可能性があり、異常に低いレベルで発現される遺伝子である可能性がある。疾患関連遺伝子は、また、1つまたは複数の突然変異を有する、または疾患の病因に直接関与する、または疾患の病因に関与する1つまたは複数の遺伝子と連鎖不均衡にある遺伝的変異を有する遺伝子を指す。前記突然変異または遺伝的変異は、例えば、単一ヌクレオチド変異(SNV)である。転写または翻訳の産物は、既知の場合も未知の場合もあり、また、レベルが正常な場合も異常な場合もある。

【0155】

したがって、本発明は、それを必要とする対象において疾患を治療する方法であって、有効量の本発明の塩基編集システムを前記対象に送達して、前記疾患に関連する遺伝子を改変すること（例えば、融合タンパク質または複数の融合タンパク質でミトコンドリアDNAを脱アミノ化する）を含む、方法をさらに提供する。本発明は、これを必要とする対象において疾患を治療するための医薬組成物の製造における塩基編集システムの使用をさらに提供し、前記塩基編集システムは、前記疾患に関連する遺伝子を改変するために使用される。本発明は、本発明の塩基編集システムと、任意に選択される、薬学的に許容される担体とを含む、それを必要とする対象において疾患を治療するための医薬組成物をさらに提供し、前記塩基編集システムは、前記疾患に関連する遺伝子を改変するために使用される。

【0156】

幾つかの実施形態では、本発明に記載の融合タンパク質または塩基編集システムは、標的核酸塩基（例えば、C残基）を脱アミノ化することによって点突然変異を核酸に導入するために使用される。幾つかの実施形態では、標的核酸塩基の脱アミノ化により、遺伝子産物の機能喪失をもたらす点突然変異の修正といった、遺伝的欠陥の修正が起こる。幾つかの実施形態では、遺伝的欠陥は、疾患または障害（例えば、リソソーム蓄積症または代謝性疾患、例えば、I型糖尿病）に関連している。幾つかの実施形態では、本明細書で提供される方法を用いて、疾患または障害に関連する遺伝子産物をコードする遺伝子または対立遺伝子に不活性点突然変異を導入することができる。

【0157】

幾つかの実施形態では、本明細書に記載の形態の目的は、ゲノム編集によって機能不全遺伝子の機能を回復することである。本明細書で提供される核酸塩基編集タンパク質は、ヒト細胞培養物における疾患関連突然変異の修正といった、ヒト細胞の生体外遺伝子編集

10

20

30

40

50

に使用することができる。核酸編集可能なDNAタンパク質（例えば、CRISPRエフェクタータンパク質Cas9）およびシトシンデアミナーゼドメインを含む融合タンパク質といった、本明細書で提供される核酸塩基編集タンパク質は、任意のTからCまたはAからGへの単一点突然変異を修正するために使用することができる。前者の場合、突然変異体のCは、脱アミノ化によって突然変異を修正するが、後者の場合、突然変異体Aと対になったCは、脱アミノ化とその後の複製によって突然変異を修正する。

【0158】

幾つかの実施形態では、本発明に記載の形態の目的は、本明細書に提供されるDNA塩基編集融合タンパク質によって修正することができる点突然変異に関連する、またはそれによって引き起こされる疾患を有する患者を治療することである。幾つかの実施形態では、疾患は、増殖性疾患である。幾つかの実施形態では、疾患は、遺伝性疾患である。幾つかの実施形態では、疾患は、新生疾患である。幾つかの実施形態では、疾患は代謝性疾患である。幾つかの実施形態では、疾患はリソソーム蓄積症である。

10

【0159】

幾つかの実施形態では、本明細書に記載の形態の目的は、ミトコンドリア疾患または障害を治療することである。本明細書で使用される場合、「ミトコンドリア疾患」とは、ミトコンドリア遺伝子突然変異、酵素経路などの異常なミトコンドリアによって引き起こされる疾患を指す。疾患の例としては、神経疾患、運動制御の喪失、筋力低下と痛み、胃腸疾患と嚥下困難、成長不良、心臓病、肝臓病、糖尿病、呼吸器合併症、てんかん、視覚/聴覚の問題、乳酸アシドーシス、発達遅延、感染症に対する高感受性が含まれるが、これらに限定されない。

20

【0160】

本発明に記載される疾患の例としては、遺伝性疾患、循環器系疾患、筋肉疾患、脳、中枢神経系疾患、免疫系疾患、アルツハイマー病、セクレターゼ障害、筋萎縮性側索硬化症（ALS）、自閉スペクトラム症、トリヌクレオチド反復伸長疾患、聴覚障害、非分裂細胞（ニューロン、筋肉）の遺伝子標的療法、肝臓および腎臓の疾患、上皮細胞および肺の疾患、癌、アッシュー症候群または網膜色素変性症 - 39、嚢胞性線維症、HIVとエイズ、ベータサラセミア、鎌状赤血球症、単純ヘルペスウイルス、自閉症、薬物中毒、加齢黄斑変性症、および統合失調症が含まれるが、これらに限定されない。点突然変異を修正するか、または疾患関連遺伝子に不活性突然変異を導入することによって治療される他の疾患は当業者に知られており、本開示はこの点において限定されない。本発明で例示的に記載される疾患に加えて、本発明によって提供される戦略および融合タンパク質は、他の関連疾患の治療にも使用することができ、この使用は当業者には明らかである。本発明に適用可能な疾患または標的については、国際公開第2015089465号（PCT/US2014/070135）、国際公開第2016205711号（PCT/US2016/038181）、国際公開第2018141835号（PCT/EP2018/052491）、国際公開第2020191234号（PCT/US2020/023713）、国際公開第2020191233号（PCT/US2020/023712）、国際公開第2019079347号（PCT/US2018/056146）、国際公開第2021155065号（PCT/US2021/015580）に記載の塩基編集システムが適用する関連疾患を参照する。

30

40

【0161】

本発明の塩基編集システムまたは医薬組成物の投与は、患者または被験者の体重および種に合わせて調整することができる。投与頻度は、医学または獣医学で許可されている範囲内である。それは、患者または被験者の年齢、性別、一般的な健康状態、その他の状態、および対処される特定の病状または症状を含む一般的な要因によって異なる。

【0162】

9. アデノ随伴ウイルス（AAV）

本発明による塩基編集融合タンパク質および/または前記塩基編集融合タンパク質をコードするヌクレオチド配列を含む発現構築物、または本発明の塩基編集システムを含む1

50

つまたは複数の gRNA は、アデノ随伴ウイルス (AAV)、レンチウイルス、アデノウイルス、またはその他のプラスミドもしくはウイルスベクタータイプを用いて送達することができる。AAVには4.5~4.75 Kbのパッケージ制限があるため。これは、プロモーターと転写ターミネーターの両方を同じウイルスベクターに組み込む必要があることを示している。4.5~4.75 Kbを超える構築物は、ウイルス送達効率の著しい低下をもたらす。シトシンデアミナーゼはサイズが大きいため、AAVにパッケージングすることが困難である。したがって、本発明の実施形態は、塩基編集を実現するためにAAVにパッケージングするためのトランケートシトシンデアミナーゼの使用を提供する。

【0163】

10. 核酸、細胞および組成物

別の態様では、本発明は、本発明のシトシンデアミナーゼまたは本発明の融合タンパク質をコードする核酸分子を提供する。

【0164】

別の態様では、本発明は、本発明のシトシンデアミナーゼ、または本発明の融合タンパク質、または本発明の塩基編集システム、または本発明の核酸分子を含む細胞を提供する。

【0165】

別の態様では、本発明は、本発明のシトシンデアミナーゼ、または本発明の融合タンパク質、または本発明の塩基編集システム、または本発明の核酸分子を含む組成物を提供する。

【0166】

幾つかの実施形態では、前記シトシンデアミナーゼ、融合タンパク質、塩基編集システムまたは核酸分子は、ウイルス、ウイルス様粒子、ビリオン、リポソーム、小胞、エキソソーム、リポソームナノ粒子 (LNP) にパッケージングされる。

【0167】

幾つかの実施形態では、前記ウイルスはアデノ随伴ウイルス (AAV) または組換えアデノ随伴ウイルス (rAAV) である。

【0168】

11. キット

本発明は、本発明の方法で使用するためのキットをさらに含み、このキットは、本発明の塩基編集融合タンパク質および/または前記塩基編集融合タンパク質をコードするヌクレオチド配列を含む発現構築物、または本発明の塩基編集システムを含む。キットには、通常、キットの内容物の所望の使用目的および/または使用方法を示すラベルが含まれる。ラベルという用語には、キットに記載されるか、またはキットとともに提供されるか、またはその他の方法でキットとともに提供されるあらゆる書面または記録された資料が含まれる。本発明の前記キットは、本発明の塩基編集システムにおける発現ベクターを構築するための適切な材料をさらに含んでもよい。本発明の前記キットは、本発明の塩基編集融合タンパク質または塩基編集組成物を細胞に形質転換するのに適した試薬をさらに含んでもよい。

【0169】

一態様では、本発明は、

(a) 本発明のシトシンデアミナーゼをコードする核酸配列と、

(b) (a) の配列の発現を駆動する異種プロモーターと、を含む核酸構築物を含むキットを提供する。

【0170】

一態様では、本発明は、

(a) 本発明の融合タンパク質をコードする核酸配列と、

(b) (a) の配列の発現を駆動する異種プロモーターと、を含む、核酸構築物を含むキットを提供する。

【0171】

10

20

30

40

50

幾つかの実施形態では、ガイドRNA骨格をコードする発現構築物をさらに含み、前記構築物は、標的配列と同一または相補的な核酸配列の前記ガイドRNA骨格へのクローニングを可能にするクローニング部位を含む。

【0172】

実施例

本発明の理解を容易にするために以下では、関連する特定の実施例および添付図面を参照し、本発明をさらに詳細に説明する。本発明の好ましい実施例は添付図面に示されている。しかしながら、本発明は多くの異なる形態で具体化することができ、本明細書に記載した実施例に限定されない。むしろ、これらの実施例を提供する目的は、本発明に開示される内容を完全に理解させるためである。

10

【0173】

材料および方法

1. ベクターの構築

同定された新しいデアミナーゼの配列は、南京金斯瑞公司によるイネと小麦の二重コード最適化を受けて、pJIT63-nCas9-PBE骨格(Addgene No. = #98164)に構築された。実施例で使用したレポーターシステムのプラスミドは、本研究室で構築された。

【0174】

sgRNAの場合、発現にはpOsU3ベクター(Addgene No. = #170132)を使用した。

20

【0175】

2. プロトプラストの単離および形質転換

本発明で使用されるプロトプラストは、イネ中花11品種に由来する。

【0176】

2.1 イネ苗栽培

まず、イネ種子を75%エタノールで1分間洗浄し、次に4%次亜塩素酸ナトリウムで30分間処理し、滅菌水で5回以上洗浄した。M6培地上に置いて、26℃で遮光して3~4週間培養した。

【0177】

2.2 プロトプラストの単離

(1) イネの茎を切り、中央部分を刃物で0.5~1mmの千切りにし、0.6Mマニトール溶液に入れて、遮光して10分間処理し、その後、フィルターで濾過し、酵素加水分解液50mL(0.45μm濾過膜で濾過)に入れ、30分間真空吸引し(圧力は約15kPa)、取り出してシェーカー(10rpm)上に置き、室温で5時間酵素加水分解した。

30

【0178】

(2) 30~50mLのW5を加えて酵素加水分解産物を希釈し、酵素加水分解液を75μmナイロン濾過膜で丸底遠沈管(50mL)に濾過した。

【0179】

(3) 23℃、250g(rcf)で、速度上昇を3レベル、速度低下を3レベルにして、3分間遠心分離し、上清を捨てた。

40

【0180】

(4) 細胞をW5 20mLで静かに懸濁し、ステップ(3)を繰り返した。

【0181】

(5) 適量のMMGを加えて懸濁させ、形質転換した。

【0182】

2.3 イネプロトプラスト形質転換

(1) 必要な形質転換ベクター各10μgをそれぞれ2mL遠心分離管に加え、均一に混合した後、鋭利な先端を取り外したピペットチップを用いてプロトプラスト200μLを吸い取り、ピペッティングして均一に混合し、PEG4000溶液220μLを加えて

50

、ピペティングして均一に混合し、室温で遮光して20～30分間形質転換を誘導した。

【0183】

(2) W5 880 μ Lを加えて、静かに反転して均一に混合し、250 g (rcf)で、速度上昇を3レベル、速度低下を3レベルにして、3分間遠心分離し、上清を捨てた。

【0184】

(3) WI溶液1 mLを加えて、静かに反転して均一に混合し、静かにフローチューブに移し、暗所、室温で48時間インキュベートした。

【0185】

3. フローサイトメータによる細胞の蛍光の観察

プロトプラストGFP陰性および陽性集団を、FACS Aria III (BD Biosciences) 機器でフローサイトメトリーによって分析した。

【0186】

4. プロトプラストおよび植物DNAの抽出とアンプリコン配列決定解析

プロトプラストを2 mL遠心分離管に集め、CTAB法を用いてプロトプラストDNA (約30 μ L)を抽出し、NanoDrop超微量分光光度計を用いてその濃度(30～60 ng/ μ L)を測定し、-20 で保存した。

【0187】

標的部位に特異的なゲノムプライマーを用いてプロトプラストDNAテンプレートのPCR増幅を実行した。20 μ L増幅システムには、4 μ L 5x Fastpfu buffer、1.6 μ L dNTPs (2.5 mM)、0.4 μ L Forward primer (10 μ M)、0.4 μ L Reverse primer (10 μ M)、0.4 μ L FastPfu polymerase (2.5 U/ μ L)、および2 μ L DNA template (約60 ng)が含まれている。増幅条件: 95 で5分間予備変性、95 で30秒間変性、50～64 で30秒間アニーリング、72 で30秒間伸長、35サイクル; 72 で5分間十分に伸長し、12 で保存した。

【0188】

上記の増幅産物を10倍希釈し、1 μ Lを2回目のPCR増幅のテンプレートとして使用した。増幅プライマーは、Barcodeを含む配列決定プライマーであった。50 μ L増幅システムには、10 μ L 5x Fastpfu buffer、4 μ L dNTPs (2.5 mM)、1 μ L Forward primer (10 μ M)、1 μ L Reverse primer (10 μ M)、1 μ L FastPfu polymerase (2.5 U/ μ L)および1 μ L DNA templateが含まれている。増幅条件は上記の通りであり、増幅サイクル数は35サイクルであった。

【0189】

PCR産物を2%アガロースゲル電気泳動で分離し、AxyPrep DNA Gel Extraction kitを用いて標的断片からゲルを回収し、回収された産物をNanoDrop超微量分光光度計で定量的に分析した。回収された産物を100 ngずつ取って混合し、アンプリコン配列決定ライブラリーの構築およびアンプリコン配列決定解析のために、諾禾致源会社に送った。

【0190】

5. ヒトおよび動物細胞のトランスフェクション

ヒトHEK293T細胞(ATCC、CRL-3216)およびマウスN2a細胞(ATCC、CCL131)を、10% (vol/vol) ウシ胎児血清(FBS、Gibco)および1% (vol/vol) ペニシリン-ストレプトミセス(Gibco)を添加したダルベッコ改変イーグル培地(DMEM、Gibco)を用いて、37、5% CO₂の加湿インキュベーター内で培養した。全ての細胞に対し、マイコプラズマ検出キット(Transgen Biotech)を用いて、マイコプラズマ汚染について通常の検出を行った。

細胞を、抗生物質を使用せずに、ポリ-D-リジンでコートされた48ウェルプレート(

10

20

30

40

50

Corning) に接種した。16 ~ 24 時間後、細胞を 1 uL の Lipofectamine 2000 (ThermoFisher Scientific)、300 ng のデアミナーゼベクターおよび 100 ng の sgRNA 発現ベクターとともにインキュベートした。シトシン塩基編集システムでトランスフェクションする際、細胞を 1 uL の Lipofectamine 2000、300 ng TALE-L および 300 ng TALE-R とともにインキュベートした。72 時間後、細胞を PBS で洗浄し、DNA を抽出した。R-loop 法を用いてオフターゲット効果を検出するには、BE4max ベクター、SaCas9BE4max ベクターおよび対応する sgRNA ベクターの全てで細胞を同時トランスフェクションした (Koblan, L.W., Doman, J.L., Wilson, C., Levy, J.M., Tay, T., Newby, G.A., Maianti, J.P., Raguram, A., & Liu, D.R. (2018). Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.*, 36, 843 - 846.)。 10

【0191】

6. TRAPseq ライブラリー

sgRNA 12K-TRAPseq ライブラリーを用いて、デアミナーゼ塩基編集システムのパフォーマンスを評価した。ウイルス形質導入の 20 時間前に、 2×10^6 個の細胞を 100 mm 培養皿に接種した。sgRNA レンチウイルス 500 μ L を形質導入した。安定して組み込まれた細胞に対し、1 μ g/mL のプロマイシン (Gibco) を使用してスクリーニングした。各塩基エディターに対し、トランスフェクションの 24 時間前に、 2×10^6 個の細胞を 6 つの培養皿に接種した。Lipofectamine 2000 60 μ L を用いて、各 CBE メンバーにプラスミド DNA 15 μ g と Tol2 DNA 15 μ g をトランスフェクションした。トランスフェクションの 24 時間後、10 μ g/mL のプラスタジジン (Gibco) を含む新しい培地に交換した。3 日後、細胞を洗浄し、再懸濁し、10 μ g/mL の濃度のプラスタジジン培地に接種した。6 日後、PBS で洗浄して全ての細胞を収集し、遠心分離し、Cell/Tissue DNA Isolation Mini Kit (Vazyme) を用いて DNA を抽出した。各デアミナーゼ塩基エディターサンプルに対し、次世代配列決定によって配列を同定した。 20

【0192】

7. DNA 抽出

HEK293T 細胞および N2a 細胞は、それらのゲノム DNA を、Lysis Buffer および Proteinase K 処理および Triumfi マウス組織直接増幅キット (北京金沙生物) による処理によって抽出を行った。 30

【0193】

植物プロトプラストは、そのゲノム DNA を 72 時間培養した後、植物ゲノム DNA キット (天根生化科技) を用いてゲノム DNA を抽出した。全ての DNA サンプルは、NanoDrop 2000 分光光度計 (Thermo Fisher) を用いて定量化した。

【0194】

8. タンパク質の構造解析とクラスタリング

AlphaFold v2.2.0 を用いてタンパク質の構造解析を行った (John Jumper and others, 'Highly Accurate Protein Structure Prediction with AlphaFold', *Nature*, 596.7873 (2021), 583 - 89)。 40

【0195】

解析結果に対して TM-align ソフトウェアを用いて TM-score を計算した。TM-score の具体的な計算式は以下の通りである (参考文献: Zhang, Yang, and Jeffrey Skolnick. (2004). Scoring function for automated assessment of protein structure template quality. *Proteins* 57(4), 702-710.)。 50

【0196】

【数1】

$$\text{TM-score} = \text{Max} \left[\frac{1}{L_N} \sum_{i=1}^{L_T} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2} \right]$$

ここで、 L_N は標的タンパク質のアミノ酸配列の長さ、 L_T はテンプレートと標的構造の両方に現れるアミノ酸配列の長さ、 d_i はテンプレートと標的構造の*i*番目の残基ペア間の距離、 d_0 は正規化されたマッチング差のスケールである。「Max」は最適空間を重ね合わせた後の最大値を表す。 10

【0197】

TM-スコアを変換後、R言語のAPEおよびphangornパッケージを使用し、UPGMA法を使用してクラスタリング計算を行った(C.P.Kurtzman, Jack W. Fell, and T. Boekhout, *The Yeasts: A Taxonomic Study*, 5th ed (Amsterdam: Elsevier, 2011).; 'A Statistical Method for Evaluating Systematic Relationships-Robert Reuven Sokal, Charles Duncan Michener-Google 図書')。まず、次の式を用いて、任意の2つの間の距離を取得した。

【0198】

【数2】

20

$$d_{(ABX)} = \frac{1}{2}(d_{ax} + d_{bx})$$

ここで、 $d_{(ABX)}$ は2点間の距離を表す。

【0199】

クラスタリングプロセスで使用される平均距離の計算式は次の通りである。 C_1 、 C_2 が、それぞれ新しいセットCにマージされるセット n_1 および n_2 を含む末端分類クラスターである場合、任意の他のクラスターDまでの平均距離は次の式により算出される。 30

【0200】

【数3】

$$d_{(C_1C_2)D} = \frac{n_1}{n_1 + n_2} d_{C_1D} + \frac{n_2}{n_1 + n_2} d_{C_2D}$$

【0201】

実施例1. 塩基編集に使用できるAPOBEC/AIDクレードの新規デアミナーゼの同定 40

既存の塩基編集システムで使用されているデアミナーゼとは異なる新規デアミナーゼを見つけるために、まず、既存のデアミナーゼとの配列類似性が低いAPOBEC/AIDクレードのデアミナーゼを、Iyerらの研究に列挙されている代表的なデアミナーゼのリストでテストした(Iyer, L.M., Zhang, D., Rogozin, I.B., & Aravind, L. (2011). Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic acids research*, 39(22), 9473-9497.)。このうち、デアミナーゼNo. 182(配列番号1)は既存のデアミナーゼとの類似性が非常に低く、そのアミノ酸配列は最も類似するマウスrAPOBEC1とも34%の配列同一性しかない。デアミナーゼNo. 182をpJIT163-nCas9-PBE骨格に構築した。つまり、rAPOBE 50

C 1 の代わりにデアミナーゼ No . 1 8 2 を用いて n C a s 9 と融合させた。レポーターシステムの評価の結果、1 8 2 - P B E が細胞内で塩基編集を受けることができることが判明した (図 1 、 図 6 および 図 7) 。

【 0 2 0 2 】

その編集能力をさらに確認するために、1 8 2 - P B E 構築物とターゲティング内因性 s g R N A 構築物をイネプロトプラストに同時形質転換した。6 つの内因性部位の編集結果の分析により、1 8 2 - P B E は、塩基編集を効果的に実現でき、その編集ウィンドウは一般的に使用されている r A P O B E C 1 に基づくシトシン塩基編集システムよりも著しく大きいことがわかった (図 2 、 図 6 および 図 7) 。したがって、タンパク質 No . 1 8 2 は、一本鎖 D N A 上のシトシンを脱アミノ化する機能を有しており、このタンパク質を基にして新規シトシン塩基編集システムを確立することができる。

10

【 0 2 0 3 】

実施例 2 . 異なるクレードにおけるデアミナーゼのシトシン脱アミノ化活性の検出

I y e r らは、既知のデアミナーゼと類似したフォールディングモードを持つタンパク質をデータベースで検索し、ドメインに基づいて上記のタンパク質を少なくとも 2 1 のクレード (c l a d e) に分割した (表 1) 。現在塩基編集に広く使用されているシトシンデアミナーゼ A P O B E C 1 、 A P O B E C 3 、 A I D および C D A 1 は、いずれも A P O B E C / A I D に似たクレードに分類される。上記のクレード以外にも、d C M P を d U M P に変換できる「 d C M P デアミナーゼおよび C o m E 」クレード、グアニン (G) をキサンチン (I) に変換できる「グアニンデアミナーゼ」クレード、ジアミノヒドロキシホスホリボシルアミノピリミジンデアミナーゼ機能を有する「 R i b D 様」クレード、R N A アデニン (A) をキサンチン (I) に変換する R N A 編集酵素機能を有する「 T a d 1 / A D A R 」クレード、ホルミルトランスフェラーゼ活性を有する「 P u r H / A I C A R トランスホルミラーゼ」クレードなど、機能が実証されているクレードもある。さらに、S C P 1 . 2 0 1 クレード、X O O 2 8 9 7 クレード、M a f B 1 9 クレード、P p u t _ 2 6 1 3 クレードなど、細菌由来でタンパク質ドメインに基づいて命名されたクレードの一部については、脱アミノ化機能がまだ解明されていない。

20

【 0 2 0 4 】

表 1 . デアミナーゼ分類ファミリー (I y e r ら , 2 0 1 1)

【 表 1 】

30

デアミナーゼが存在するファミリーの名称		
ボルバキア_B3gp45 (AID/APOBEC デアミナーゼ)	DYW様 クレード	RibD様デアミナーゼ
新規AID/APOBEC様 クレード	FdhD クレード	SCP1.201 クレード
CDD/CDA様シチジン デアミナーゼ	MafB19 クレード	TM1506 クレード
グアニンデアミナーゼクレード	OTT1508 クレード	XOO2897 クレード
Tad1/ADAR クレード	Pput_2613 クレード	YwqJ クレード
Bd3614 クレード	PurH/AICAR トランスホルミラーゼクレード	JAB クレード
トキシンデアミナーゼ (BURPS668_1122) クレード	TadA-Tad2(ADAT2), Tad3(ADAT3) クレード	SCP1. 201デアミナーゼに関連した Imm1免疫タンパク質 クレード

40

【 0 2 0 5 】

上記のクレードがシトシンデアミナーゼ活性を有するか否かを検出するために、A P O B E C / A I D クレードを除く、B d 3 6 1 4 、 C D D / C D A 様、D Y W 様、F d h D 、 M a f B 1 9 、 新規 A I D / A P O B E C 様、O T T 1 5 0 8 、 P u r H / A I C A R トランスホルミラーゼ、R i b D 様、T M 1 5 0 6 、 S C P 1 . 2 0 1 、 S C P 1 . 2 0

50

1デアミナーゼに関連したImm1免疫タンパク質、YwqJおよびXOO2897の14個のクレードに分布する合計48個のデアミナーゼタンパク質を、Iyerらに列挙された代表的なデアミナーゼリストから選択した。全てのタンパク質をpJIT163-nCas9-PBE骨格に構築し、BFP-to-GFPレポーターシステム(Zong, Y.ら.Nat.Biotechnol.35,438-440(2017))によって一本鎖DNAに結合して脱アミノ化する活性を評価した。その結果、5つのクレードからの合計23種類のタンパク質がシトシンデアミナーゼ活性を有し、これらは、新規AID/APOBEC様クレード(No.2~1479、およびNo.2~1478)、ならびに細菌由来のSCP1.201クレード(No.69、No.55、No.57、No.64、No.76、No.2~1146、No.2~1160、No.54、No.56、No.59、No.60、No.61、No.72、No.74、No.75、No.63、No.2~1158)、XOO2897クレード(No.2~1429、No.2~1442)、TM1506クレード(No.2~39)およびMafB19クレード(No.101m)に由来することが判明した。特にSCP1.201クレードでは、テストした19種類のタンパク質のうち18種類においてシトシンデアミナーゼ活性が検出された。イネの2つの内因性部位のテストにより、シトシンデアミナーゼ活性を有する上記の合計23種類のタンパク質は、8つの高編集効率デアミナーゼ(図4および図5)、8つの中等編集効率デアミナーゼ(図6および図7)、7つの低編集効率デアミナーゼに分類できる。

【0206】

新たに発見されたデアミナーゼの編集能力をさらに確認するために、レポーターシステムを光らせることができるグループの中から候補デアミナーゼNo.69を選択してテストした。このタンパク質はSCP1.201クレードに属する。その編集能力をさらに確認するために、69-PBE構築物とターゲティング内因性sgRNA構築物をイネプロトプラストに同時形質転換した。6つの内因性部位の編集結果を解析した結果、69-PBEは、塩基編集を効果的に実現でき、その編集効率は一般的に使用されているrAPOBEC1に基づくシトシン塩基編集システムよりも著しく高いことが判明した(図3)。したがって、新たに同定されたタンパク質は、一本鎖DNA上のシトシンを脱アミノ化する機能を有することが可能であり、これらのタンパク質に基づいて新たなシトシン塩基編集システムを確立することができる。

【0207】

実施例3.タンパク質の構造解析、クラスタリングおよび新規シトシンデアミナーゼの発見

以上の実施例を踏まえ、タンパク質の機能を効率的に解析するためには、効果的なプロテアーゼの機能同定およびスクリーニング方法を提案する必要がある。タンパク質の三次元構造はその機能に対し決定的な作用を有するため、既知または予測されるタンパク質構造の比較分析および分類クラスタリングは、デアミナーゼを機能クレードに分類するのに有効な方法である可能性がある。そこで、AI支援タンパク質構造予測、構造キャリブレーション、クラスタリングを組み合わせ、デアミナーゼ間の新しいタンパク質分類関係を生成した(図8)。

【0208】

InterProデータベースから、デアミナーゼドメイン含有のアノテーションが付けられた238個のタンパク質配列と、JABドメインファミリーからの4つの外群候補タンパク質配列を選択した(図9)。具体的には、16個のデアミナーゼファミリーから、少なくとも100アミノ酸の長さを持つ15個の候補遺伝子をそれぞれ選択し、AlphaFold2を用いてそれらのタンパク質構造を予測した。標準化されたスコアリングモデルTM-スコアを用いて、全ての候補タンパク質に対して多重構造アライメント(MSA:multiple structural alignments)を実行した。TM-スコアの具体的な計算式は次の通りである(参考文献:Zhang,Yang,and Jeffrey Skolnick.(2004).Scoring function for automated assessment of protein structure template qual

10

20

30

40

50

ity. Proteins 57(4), 702-710.)。

【0209】

【数4】

$$\text{TM-score} = \text{Max} \left[\frac{1}{L_N} \sum_{i=1}^{L_T} \frac{1}{1 + \left(\frac{d_i}{d_0}\right)^2} \right]$$

ここで、 L_N は標的タンパク質のアミノ酸配列の長さ、 L_T はテンプレートと標的構造の両方に現れるアミノ酸配列の長さ、 d_i はテンプレートと標的構造の*i*番目の残基ペア間の距離、 d_0 は正規化されたマッチング差のスケールである。「Max」は、最適空間を重ね合わせた後の最大値を表す。

【0210】

M S Aの結果に基づいて、タンパク質間の全体的な構造の関連性を反映する構造類似マトリックスを生成した。次に、算術平均を用いた重み付けなしペアグループ法 (UPGMA) を用いて、これらの類似マトリックスを構造に基づく樹状図 (図10) にした。樹状図では、238個のタンパク質が20個の独自の構造クレードにクラスタリングされ、各クレードのデアミナーゼは、異なる保存されたタンパク質ドメインを有する (図11Aおよび図11B)。保存された遺伝子近傍やドメインのアーキテクチャなどのコンテキスト情報を使用しなくても、タンパク質構造に基づいて正確なタンパク質クラスター分類を生成できることを発見した。構造に基づく階層的クラスタリングを使用した場合、異なるクレードは独自の構造を反映し、これは異なる触媒機能および特性を有することを意味している (図11Aおよび図11B)。興味深いことに、この構造に基づくクラスタリング方法は、機能的類似性のランク付けにおいて、従来的一次元アミノ酸配列クラスタリング方法よりも効果的であることもわかった。例えば、アミノ酸配列に基づくクラスタリング方法では、プリン代謝に関与するアデニンデアミナーゼ (A_d e a m i n、I n t e r P r o データベースでは P F O 2 1 3 7) が異なるクレードに分類されるのに対し、構造に基づくクラスタリング方法では、同一のデアミナーゼクレードに分類される。

【0211】

さらに、構造に基づくクラスタリング方法を用いて、4つのデアミナーゼファミリー (d C M P、M a f B 1 9、L m j F 3 6 5 9 4 0 および A P O B E C (I n t e r P r o によるアノテーション)) を2つの独立したクレードに分類した (図11Aおよび図11B)。タンパク質構造を比較した結果、これら4つのデアミナーゼファミリーの2つのクレードは、全く異なる構造を有し、これは、これらの I n t e r P r o の命名および配列に基づく分類とは反対になる可能性がある (図11Bおよび図12)。つまり、人工知能支援によるタンパク質の三次元構造に基づくタンパク質のクラスタリング分類は、信頼性の高いクラスタリング結果を提供し、必要なアミノ酸配列が1つだけであり、他のゲノム推論を必要としないため、他の方法よりも便利で効果的なタンパク質関係生成戦略となる。

【0212】

実施例4：三次元構造樹を使用した、塩基編集に使用できる S C P 1 . 2 0 1 クレードのデアミナーゼの機能の同定

実施例2において各クレードにおけるデアミナーゼの機能を評価した結果、驚くべきことに、S C P 1 . 2 0 1 クレードの一部のデアミナーゼが一本鎖DNA基質の脱アミノ化を触媒する能力を有することを発見した。以前は、これらのデアミナーゼは、I n t e r P r o データベース (P F 1 4 4 2 8) において二本鎖DNAデアミナーゼ毒素A様 (D d d A 様) デアミナーゼとしてアノテーションが付けられていた。このうち、D d d A 酵素は、最近非 C R I S P R 二本鎖DNAシトシン塩基エディター (D d C B E : C R I S P R - f r e

10

20

30

40

50

e double-stranded DNA cytosine base editor) で使用されるようになったデアミナーゼで、二本鎖DNAシトシン塩基の脱アミノ化に使用可能である (NCBI Reference Sequence: WP_006498588.1) (B.Y.Mok, M.H.de Moraes, J.Zeng, D.E.Bosch, A.V.Kotrys, A.Raguram, F.Hsu, M.C.Radey, S.B.Peterson, V.K.Mootha, J.D.Mougous, D.R.Liu, A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature 583, 631-637 (2020).)。まさに DddA の存在により、それが属する SCP1.201 クレード内の全てのタンパク質が二本鎖DNAデアミナーゼ (Ddd: Double-stranded DNA deaminases) としてアノテーションが付けられていた。

【0213】

この問題を踏まえ、実施例3におけるタンパク質の三次元構造に基づいた機能予測を利用して、以下の作業を行った。この SCP1.201 クレードを再分析するために、InterPro データベースから 489 個の SCP1.201 デアミナーゼを全て選択した。また、BLAST アライメントによって DddA と 35% ~ 50% の類似度を有することが判明したが、InterPro では個別に記載されている他の 7 つのタンパク質も含めた。認識およびカバレッジスクリーニングの後、332 個の SCP1.201 デアミナーゼに対して新しい人工知能支援タンパク質構造分類を実行した。構造クラスター分析の結果、SCP1.201 デアミナーゼはそれぞれ独自のコアドメインモチーフを有する異なるサブクレードにクラスタリングされている (図 13A ~ 図 13E)。

【0214】

ここで重要なのは、DddA と他の 10 個のタンパク質が同じ SCP1.201 の 1 つのサブクレードにクラスタリングされていることである。このサブクレードの 11 個のタンパク質全ての 3D 予測構造を分析することにより、それらが DddA と類似したコア構造を持っていることがわかった。DddA との構造類似性を考慮して、このサブクレードの他のタンパク質も二本鎖DNAシトシン脱アミノ化機能を有するものと予測した。

【0215】

実施例5. 動物細胞の塩基編集における DddA サブクレードを含むデアミナーゼの検証

実施例4において本発明の予測方法を用いて得られた、DddA が存在するサブクレードの SCP1.201 候補タンパク質が DddA と機能的類似性を有するか否か、すなわち dsDNA に対する脱アミノ化作用を有するか否かを評価するために、このサブクレードの各デアミナーゼによって個別に形成されるか、またはデアミナーゼ様 DddA 構造を 2 つの残基の対立遺伝子部位で 2 つの部分に分割し、それらをデュアルTALE システムで連結することによって形成される DdCBE を設計した。(方法については、B.Y.Mok, M.H.de Moraes, J.Zeng, D.E.Bosch, A.V.Kotrys, A.Raguram, F.Hsu, M.C.Radey, S.B.Peterson, V.K.Mootha, J.D.Mougous, D.R.Liu, A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature 583, 631-637 (2020) を参照) (図 14、表 2)。HEK293T 細胞の JAK2 および SIRT6 部位でこの Ddd サブクレードからのタンパク質を評価し、dsDNA 塩基編集が可能な 13 個のタンパク質を観察した (表 2)。以下、これらのデアミナーゼを二本鎖DNAデアミナーゼ (Ddd) と命名し、この新たに発見された Ddd サブクレードに分類した。

【0216】

表 2. DddA が存在するサブクレードのタンパク質の触媒活性

10

20

30

40

50

【表 2】

命名	遺伝子 ID	配列番号	dsDNA 触媒活性
Ddd1	SCP177	28	++
Ddd2	/	29	+
Ddd3	SCP001	30	++
Ddd4	/	31	+
Ddd6	SCP009	32	+
Ddd7	SCP103	33	++
Ddd8	SCP234	34	++
Ddd9	SCP003	35	++
Ddd10	SCP004	36	+
Ddd11	SCP271	37	+
Ddd12	SCP005	38	+
Ddd13	SCP006	39	++
Ddd14	SCP007	40	+

10

記号 ++ は触媒活性が高いことを表し、+ は触媒活性が低いことを表し、- は触媒活性がないことを表す。

【0217】

20

実施例 6 . 植物細胞および動物細胞の塩基編集における D d d A サブクレードを含まないデアミナーゼの検証

比較のために、この実験では、D d d A サブクレードを含まない他の S C P 1 . 2 0 1 候補タンパク質の脱アミノ化作用をさらに評価した。そこから 2 4 個のタンパク質をランダムに選択し、C B E 蛍光レポーターシステムに配置した。その結果、これらのうち 2 2 個のタンパク質が検出可能な蛍光を示した。これらのうち 1 3 個のタンパク質を選択して、哺乳動物細胞における C B E 条件下での内因性部位での塩基編集を評価した (図 1 6 A 、表 3) 。これらのタンパク質は、以前に D d d A 様タンパク質としてアノテーションが付けられていたが、実験結果では、これらのタンパク質が s s D N A に対してのみシトシン塩基編集活性を示し (図 1 3 A 、図 1 6 A および表 3) 、d s D N A に対しては活性を示さなかった (図 1 6 B) 。これらの機能と作用に基づいて、今後の作業において、s s D N A を標的とする作用を持つ S C P 1 . 2 0 1 クレードからのこれらのタンパク質を一本鎖 D N A デアミナーゼ (S d d : Single-stranded DNA deaminases) と命名した。

30

【0218】

上記の実験結果によれば、驚くべきことに、S C P 1 . 2 0 1 クレードからのタンパク質メンバーのほとんどが、I n t e r P r o データベース (P F 1 4 4 2 8) においてアノテーションが付けられている D d d A 様タンパク質ではなく、S d d タンパク質であることを発見した。また、例えば S d d 7 の構造に示されるように (図 1 3 D 、図 1 3 E) 、これらの S d d タンパク質は互いに類似しており、D d d タンパク質の構造とは明確に区別されることも見出した。S d d 7 は、s s D N A 編集効率が最も高いシトシン塩基エディターの 1 つである。したがって、本発明の方法により、I n t e r P r o データベース (P F 1 4 4 2 8) においてアノテーションが付けられている D d d A 様デアミナーゼは、さらに細分化され、適切なアノテーションが改めて付けられるべきことが示される。

40

【0219】

対照として、さらに一次元アミノ酸配列に基づいて S C P 1 . 2 0 1 クレードからのタンパク質をクラスタリングし、J A B 外群を用いて構造樹を検証したところ、J A B 外群のメンバーが樹全体に分散していることがわかった。これらの結果は、タンパク質の関係を比較および評価するためにタンパク質の構造に基づく分類を使用することの有効性と重要性を示している。

50

【 0 2 2 0 】

表 3 . D d d A が存在しないサブクレードのタンパク質の触媒活性

【表 3】

命名	遺伝子ID	配列番号	dsDNA 触媒活性 (内因性部位)	ssDNA 触媒活性 (内因性部位)	ssDNA 触媒活性 (レポーターシステム)	配列選好性			
						CC	AC	GC	TC
Sdd9	SCP044	2	x	√	√	++	++	++	++
Sdd5	SCP017	3			√	++	++	++	+
Sdd7	SCP016	4	x	√	√	++	++	++	++
Sdd4	SCP014	5			√	++	++	++	++
Sdd76	SCP012	6			√	++	++	++	++
Sdd6	SCP273	7			√	++	++	+	+
/	SCP021	8			√	++	+	+	-
/	SCP038	9			√	+	+	-	-
/	SCP051	10			√	-	-	-	+
Sdd59	SCP183	11	x	√	√	+	-	-	-
Sdd10	SCP018	12			√	+	+	-	+
/	SCP157	14			√	-	+	-	+
Sdd3	SCP170	17			√	-	+	+	-
/	2-1158	18			√	+	-	-	-
/	SCP158	42			√	+	+	-	+
/	SCP315	43	x	√	√	++	+	+	++
/	SCP020	44	x	√	√	++	+	++	+
/	2-1156	45			√	+	+	+	+
Sdd2	Sdd2	46			√	+	+	-	+
/	SCP008	47			√	+	+	-	-
/	SCP011	48			√	+	-	-	-
/	SCP013	49			√	+	-	-	-
/	SCP090	66	x	x	x	-	-	-	-
/	SCP015	67	x	x	x	-	-	-	-
/	SCP278	68	x	x	x	-	-	-	-
/	SCP287	69	x	x	x	-	-	-	-
/	SCP341	70	x	x	x	-	-	-	-
/	SCP353	71	x	x	x	-	-	-	-
/	SCP357	72	x	x	x	-	-	-	-
/	SCP372	73	x	x	x	-	-	-	-

10

20

30

40

記号 ++ は編集選好性が高いことを表し、 + は編集選好性が低いことを表し、 - は編集選好性がないことを表す。

【 0 2 2 1 】

また、実施例 5、6 におけるタンパク質の機能の検証結果を総合的に解析した。この結果は、人工知能支援によるタンパク質の三次元構造に基づくタンパク質のクラスタリング分類が信頼性の高いクラスタリング結果を提供し、本発明の方法を用いて構築された三次元構造樹がタンパク質の詳細な機能を正確に同定および予測できることを示している。ま

50

た、必要なアミノ酸配列が1つだけであり、他のゲノム推論を必要としないため、他の方法よりも便利で効果的なタンパク質関係生成戦略となる。三次元構造樹において、TMスコアが0.7以上であることをクラスタリング条件とした場合、そのクラスタリング結果によるタンパク質の触媒機能の予測結果は、実験検証の結論と一致する(表4)。すなわち、基準タンパク質とのTMスコアリングによるTMスコアが0.7以上であることをアノテーションとしてクラスタリングが行われ、得られたサブクレードは、基準タンパク質と同一または類似の触媒機能を有し、本発明の方法は、同定および予測の効率を大幅に向上させている。

【0222】

表3. DddAが存在しないサブクレードのタンパク質の触媒活性

10

【表4】

命名	遺伝子ID	配列番号	dsDNA触媒活性 (内因性部位)	ssDNA触媒活性 (内因性部位)	ssDNA触媒活性 (レポーターシステム)
Sdd9	SCP044	2	x	√	√
Sdd5	SCP017	3			√
Sdd7	SCP016	4	x	√	√
Sdd4	SCP014	5			√
Sdd76	SCP012	6			√
Sdd6	SCP273	7			√
/	SCP021	8			√
/	SCP038	9			√
/	SCP051	10			√
Sdd59	SCP183	11	x	√	√
Sdd10	SCP018	12			√
/	SCP157	14			√
Sdd3	SCP170	17			√
/	2-1158	18			√
/	SCP158	42			√
/	SCP315	43	x	√	√
/	SCP020	44	x	√	√
/	2-1156	45			√

20

30

【0223】

実施例7. 新しいDddタンパク質とDddAの有する異なる編集選好性

DddAは5'-TCモチーフに対して厳密な選好性を有するため、DddAに基づくdsDNA塩基エディターの使用は、主にTC標的に限定される(B.Y.Mok, M.H.de Moraes, J.Zeng, D.E.Bosch, A.V.Kotrys, A.Raguram, F.Hsu, M.C.Radey, S.B.Peter son, V.K.Mootha, J.D.Mougous, D.R.Liu, A bacterial cytidine deaminase toxin enables CRISPR-free mitochondrial base editing. Nature 583, 631-637 (2020).)。最近進化したDddA11は、より一般的な適用性を示しており、5'-HC (H=A, CまたはT)モチーフを脱アミノ化してシトシン塩基編集を実現するために使用できるが、AC、CCおよびGCの標的に対する編集効率にはまだ改善する必要がある(B.Y.Mok, A.V.Kotrys, A.Raguram, T.P.Huang, V.K.Mootha, D.R.Liu, CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. Nat. Biotechnol. 40, 1378-1387)。本発明の新たに発見されたDddタンパク質を評価することにより、それらがDdCBEの効率を高め、ターゲティング範囲を拡張できるか否かを確定した。Dddサブクレードに属する13個のデアミナーゼをDdCBEに構築し、HEK293T細胞の内因性JAK2およびSIRT6部位におけるdsDNA塩基編集の状況を評価した(図15、図18および表2)。興味深いことに、Ddd1、Ddd7、Ddd8およびDdd9はDddAと

40

50

比較して同等またはより高い編集効率を有することがわかった(図17Aおよび図18)。ここで重要なのは、D d d 1およびD d d 9がD d d Aよりも5' - GCモチーフに対してはるかに高い編集活性を有することである(図17Aおよび図18)。ここで注目すべきは、J A K 2のC 1 0 (5' - GC)残基とS I R T 6のC 1 1 (5' - GC)残基では、D d d Aの編集比はそれぞれ21.1%と0.6%にすぎないが、D d d 9の編集比はそれぞれ65.7%と45.7%であったことである(図17A)。

【0224】

一部のD d dタンパク質はD d d Aと比較して異なる編集パターンを示すようである。そのため、これらのD d dタンパク質の全てのモチーフの選好性を評価することにした。まず、J A K 2標的配列をコードする複数のプラスミド(B. Y. Mok, A. V. Kotrys, A. Raguram, T. P. Huang, V. K. Mootha, D. R. Liu, CRISPR-free base editors with enhanced activity and expanded targeting scope in mitochondrial and nuclear DNA. Nat. Biotechnol. 40, 1378-1387)を構築した。また、9~11位のG C CをM C N (M/N = A, T, CおよびG)の9~11位に変更し、16個の異なるプラスミドを得て、各プラスミドをそれぞれD d C B Eバリエーションとともに同時トランスフェクションした(図17B)。各M C NのC・G - to - T・A塩基変換頻度の比較分析後、各d s DNAデアミナーゼの配列のコンテキスト選好性を反映する、対応するモチーフロゴ(log o)プロットを生成した(図17C)。前述したように、D d d Aおよびその構造ホモログD d d 7が5' - T Cモチーフを強く好むことを見出した(図17C、図19)。一方、D d d 1およびD d d 9は5' - G Cモチーフの基質を編集する傾向があり、D d d 8は、5' - W C (W = AまたはT)モチーフの基質を編集する傾向があることがわかった。したがって、D d dサブクレードのより深い分析を通じて、異なるモチーフの編集に使用できる一連の新しいD d dタンパク質を発見した。これらのタンパク質は、D d C B Eのターゲティング範囲および有用性を大幅に拡大し、大きな応用の可能性を示す(図17C、図19)。

【0225】

実施例8. ヒト細胞および植物におけるS d dデアミナーゼの塩基編集

次に、新たに発見されたS d dタンパク質がより正確またはより効率的な塩基編集にも使用できるか否かを確認することにした。そのため、6つの最も活性のあるS d dと4つの弱いS d dを選択して評価し、蛍光レポーターシステムを用いてこれらの活性を比較した(表3)。10個のS d dのそれぞれについて植物C B Eを設計し、イネプロトプラストの6部位におけるそれらの内因性塩基編集を評価した(図20および図21)。そのうち7種類のデアミナーゼ(S d d 7、S d d 9、S d d 5、S d d 6、S d d 4、S d d 7 6およびS d d 10)がラットA P O B E C 1 (r A P O B E C 1)ベースのC B Eよりも高い活性を有することを発見した。最も活性のあるS d d 7塩基エディターのシトシン塩基編集比は55.6%と高く、これはr A P O B E C 1の3.5倍以上である。

【0226】

これらのデアミナーゼの汎用性を検証するために、対応するヒト細胞を標的にするB E 4 m a xベクター(L. W. Koblan, J. L. Doman, C. Wilson, J. M. Levy, T. Tay, G. A. Newby, J. P. Maianti, A. Raguram, D. R. Liu, Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. Nat. Biotechnol. 36, 843-846(2018).)も構築し、H E K 2 9 3 T細胞の3つの内因性標的におけるこれらの編集効率を評価した。H E K 2 9 3 T細胞における結果はイネにおける結果と一致しており、S d d 7が最も高い編集活性を有することが判明した(図22)。

【0227】

以前、ヒトA P O B E C 3 A (A 3 A)が植物において大きな編集ウィンドウを有し、高い編集活性を発揮することを発見した(Y. Zong, Q. Song, C. Li, S. Jin, D. Zhang, Y. Wang, J.-L. Qiu, C. Gao, Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. Nat. Biotechnol. 36, 950-953(2018)., Q. Lin, Z. Zhu, G. Liu, C. Sun, D. Lin, C. Xue, S. Li, D. Zhang, C. Gao, Y. Wang, J.-L. Qiu, Genom

e editing in plants with MAD7 nuclease. *J. Genet. Genomics* 48, 444-451 (2021).)。したがって、ヒト細胞 (図 22) および植物 (図 23) における A3A および Sdd7 の編集活性を比較した。興味深いことに、Sdd7 は、HEK293T 細胞の 3 つの標的部 10 位全て (図 22)、およびイネプロトプラストの 5 つの内因性部位 (図 23) において、A3A と同等の編集活性を有していた。これらの結果は、Sdd7 が植物およびヒトの細胞に普遍的に適用できる強力なシトシン塩基エディターであることを裏付けている。

【0228】

実施例 9 . Sdd タンパク質の有する独自の塩基編集特性

内因性塩基編集を評価する際、ヒトおよびイネの細胞でテストされたゲノム標的部 10 位で、異なる Sdd-CBE の異なる編集パターンが観察された。例えば、Sdd7、Sdd9 および Sdd6 は特定のモチーフ編集選好性を示さなかったが、Sdd3 は、5'-GC および 5'-AC モチーフの編集を好み、5'-TC および 5'-CC モチーフの編集を著しく嫌っているようであった (図 24)。各デアミナーゼの編集パターンをより適切に 20 分析するため、塩基編集結果の並行定量化のためのハイスループット方法である標的レポーターアンカー位置シーケンシング (TRAP-seq: Targeted Reporter Anchored Positional Sequencing) を使用した (Xi Xiang, Kunli Qu, Xue Liang, Xiaoguang Pan, Jun Wang, Peng Han, Zhanying Dong, Lijun Liu, Jiayan Zhong, Tao Ma, Yiqing Wang, Jiaying Yu, Xiaoying Zhao, Siyuan Li, Zhe Xu, Jinbao Wang, Xiuqing Zhang, Hui Jiang, Fengping Xu, Lijin Zou, Huajing Teng, Xin Liu, Xun Xu, Jian Wang, Huanming Yang, Lars Bolund, George M. Church, Lin Lin, Yongliun Luo. (2020). Massively parallel quantification of CRISPR editing in cells by TRAP-seq enables better design of Cas9, ABE, CBE gRNAs of high efficiency and accuracy. *bioRxiv* 2020.05.20.103614)。それぞれが独自の gRNA 発現カセットと対応する代替標的部 30 位を含む 12,000 個の TRAP 構造を含む 12K TRAP-seq ライブラリーは、レンチウイルス形質導入によって HEK293T 細胞に安定して組み込まれた。細胞培養および抗体選択に続いて、塩基エディターをこの 12K-TRAP 細胞株に一過性トランスフェクションし、続いて 10 日間プロマイシンおよびプラストサイジン選択を行った (図 25A)。トランスフェクション後 11 日目に、ゲノム DNA を抽出し、ディープアンプリコン配列決定を実行して、各デアミナーゼの編集産物を評価した (図 25A)。Sdd7 および Sdd6 は高い配列コンテキスト選好性を示さなかったが、rAPOBEC1 は 5'-TC および 5'-CC 塩基に対して高い選好性を有する一方、5'-GC および 5'-AC 塩基に対しては関心を示さないことを発見した (図 25B)。一方、Sdd3 は完全に相補的なパターンを示し、5'-GC および 5'-AC 塩基の編集を好むが、5'-TC および 5'-CC 塩基に対してはほとんど 40 活性を示さなかった (図 25B)。興味深いことに、Sdd6 および Sdd3 が rAPOBEC1 および Sdd7 と比較して異なる編集ウィンドウを有し、遠方の PAM 側の +1 から +3 位置をより一層好んで編集することを発見した (図 25B)。要するに、新たに認識された Sdd 塩基エディターは、編集効率の向上、異なる脱アミノ化の選好性、編集ウィンドウの変更など、従来のシトシン塩基エディターと比較して独自の塩基編集特性を示す。 40

【0229】

実施例 10 . Sdd タンパク質の高忠実度編集特性

CBE がゲノム全体の Cas9 に基づくオフターゲット編集結果をもたらす可能性があるという以前の報告により、臨床応用におけるこれらの高精度ゲノム編集技術の安全性について懸念が生じている。これらのオフターゲット突然変異はシチジンデアミナーゼの過剰発現の結果である可能性があると考えている。新たに発見された Sdd タンパク質がオフターゲット編集とオンターゲット編集間のより好ましいバランスを提供できるか否かについて確認することにした。したがって、イネプロトプラストにおける直交 R-loop アッセイを用いて、10 個の Sdd の Cas9 非依存性オフターゲット効果を評価した。 50

10種類のデアミナーゼのうち6種類(Sdd2、Sdd3、Sdd4、Sdd6、Sdd10およびSdd59)のオフターゲット活性がrAPOBEC1よりも低いことがわかった。興味深いことに、Sdd6はオフターゲット編集活性をほとんど示さなかったが、イネおよびヒト細胞の6つの内因性部位でテストしたところ、依然として強力なオフターゲット塩基編集能力を有していた(図26Aおよび図27)。これら10種類のデアミナーゼのオフターゲット：オフターゲット比を分析したところ、Sdd6は、最も高いオフターゲット：オフターゲット編集比を示し、これはAPOBEC1の37.6倍であった(図26B)。さらに、HEK293T細胞における、Sdd6とrAPOBEC1、およびその2つの高忠実度デアミナーゼバリエーションであるYE1とYEEのオフターゲット編集とオフターゲット編集を比較した。ここで重要なのは、Sdd6が最も高いオフターゲット：オフターゲット編集比を有することである。計算した結果、rAPOBEC1、YE1およびYEEよりもそれぞれ2.8倍、2.1倍および2.5倍高く(図26B、図26Cおよび図28)、hA3Aよりも10.4倍高かった(図26Cおよび図28)。なお、Sdd6のオフターゲット活性はrAPOBEC1の活性に匹敵し、YE1およびYEEよりもはるかに高かった(図28)。したがって、SCP1.201クレードには、高忠実度の塩基エディターとして使用できる、独自でより正確なSddタンパク質が含まれていると判断した。

10

【0230】

実施例11. AlphaFold2構造予測支援によるSddタンパク質の合理的な設計

20

CBE送達のためのウイルスの使用は、疾患治療において巨大な潜在力を有するが、APOBEC/AID様デアミナーゼのサイズが大きいため、生体内編集用途として1つのアデノ随伴ウイルス(AAV)粒子にパッケージングされる能力が制限される(31)。他の研究者によってデュアルAAV戦略送達方法が開発されており、CBEをアミノ末端断片とカルボキシル末端断片に分割し、それらを単独のAAV粒子にパッケージングするといったものである。しかしながら、このような送達の手段は、量産能力やより高いウイルス量が課題であり、使用するうえでの安全性に潜在的な懸念がある。最近、CDA-1に基づくトランケートヤツメウナギCBEを用いてシングルAAVパッケージCBEが開発されたが、これらのベクターはHEK293T細胞においてほとんど編集活性を示さなかった。SCP1.201デアミナーゼの標準的な緻密性と保存性により、これらはシングルAAVパッケージCBEの開発にとって理想的なタンパク質である可能性があると考えられる。本発明は、人工知能支援による三次元構造タンパク質モデリングにより、新たに発見されたSddタンパク質をさらに設計し、そのサイズを短縮することを試みた。

30

【0231】

まず、全ての活性SddデアミナーゼについてAlphaFold2の予測構造を比較し、それらが保存されたコア構造を有することを発見した(図13D、図13Eおよび図29)。次に、Sdd7、Sdd6、Sdd3、Sdd9、Sdd10およびSdd4の複数のトランケートバリエーションをそれぞれ生成し、イネプロトプラストの2つの部位でこれらバリエーションの内因性塩基編集をテストした。その結果、ミニSdd7、ミニSdd6、ミニSdd3、ミニSdd9、ミニSdd10およびミニSdd4が新しい最小化デアミナーゼであることを発見した。これらはいずれも非常に小さく(約130~160aa)、イネプロトプラストやヒト細胞の全長タンパク質と同等以上の編集効率を有する(図30A)。ここで注目すべきは、全てのミニデアミナーゼは、シングルAAVパッケージのScas9に基づくCBE(<4.7kb)の構築を可能にした(図30B)。ミニSdd6を用いてシングルAAVにパッケージングされたScas9ベクターを構築し、一過性トランスフェクションにより、マウス神経芽腫N2a細胞のHPD遺伝子(ハツカネズミの4-ヒドロキシフェニルピルビン酸ジオキシゲナーゼ)の2つの部位でその編集効率が約60%であることを発見した(図30C)。これらの結果は、AAVに基づくCRISPR塩基編集送達において、Sddタンパク質がAPOBEC/AIDデアミナーゼよりも大きな優位性を有することを示している。AAVパッケージングのためにSd

40

50

dタンパク質をさらに短縮することに成功したことは、本発明の三次元構造に基づくタンパク質の機能予測方法にとって非常に有利であることを示している。

【0232】

実施例12. 新しいS d dに基づくC B Eの塩基編集能力

次に、植物の塩基編集における新しいS d d工学的タンパク質の使用を検討した。まず、アグロバクテリウム媒介によるイネのゲノム編集でのミニS d d 7の使用可能性を評価した。農業用途で最も一般的に使用されるヒトA 3 A (h A 3 A)に基づくC B Eに対し、ミニS d d 7に基づくC B Eでは、より多くのイネ陽性植物株が観察された。さらに、編集された植物の数が多く、編集効率が高く、これはh A 3 AよりもC B E効率が高く、毒性が低いことを反映している(図31)。

10

【0233】

大豆は世界中で栽培されている最も重要な作物の1つであり、植物油とタンパク質の重要な供給源である。塩基編集は大豆で実証されているが、大豆作物のほとんどのテスト部位では依然として編集が難しく、編集効率が低いという問題が存在する。新たに開発したS d dに基づくC B Eが大豆においてより優れたシトシン塩基編集効果を生み出すか否かを把握するために、s g R N A発現を駆動するA t U 6プロモーター、およびC B E発現を駆動するC a M V 2 x 3 5 Sプロモーターを用いてベクターを構築し、アグロバクテリウム媒介形質転換後の遺伝子組換え大豆の毛状根を評価した(図32)。A P O B E C / A I Dデアミナーゼが、大豆において他のC B Eによる編集が特に困難であるG m A L S 1 - T 2およびG m P P O 2の2つの部位を含む、評価した5つの部位全てで低い編集活性を有することを発見した(図30D)。なお、ミニS d d 7の5つの部位でのシトシン塩基編集レベルは、他のデアミナーゼr A P O B E C 1、h A 3 Aおよびh A I Dのそれぞれの26.3倍、28.2倍、10.8倍であり、編集効率は67.4%と高かった(図30D)。したがって、新たに発見されたこれらのS d dタンパク質を重点的に利用することで、大豆作物での効率的なシトシン塩基編集の限界を突破することができる。

20

【0234】

次に、ミニS d d 7を塩基編集に使用して、アグロバクテリウム媒介遺伝子組換え大豆植物株を取得することを試みた。内因性G m P P O 2遺伝子を編集してR 9 8 C突然変異を生成し、これにより、カルフェントラゾンエチルに対する耐性を持つ大豆植物株が得られた。3つの独立した形質転換実験から77本の遺伝子組換え大豆苗を取得し、そのうち21本は塩基編集用のヘテロ接合であった(図30E、図30F)。カルフェントラゾンエチルによる10日間の処理後、野生型植物株は、しおれに敏感で根を張ることができなかったが、突然変異型植物株は良好かつ正常に成長したことが顕著に観察できた(図30G)。大豆植物で使用するための効率的なシトシン塩基エディターの開発により、将来的には様々な応用が可能になる。

30

【0235】

実施例13. 他のファミリーのタンパク質の塩基編集特性

S C P 1 . 2 0 1ファミリーのタンパク質の詳細な分類と検証に加えて、I y e rデアミナーゼ分類ファミリー(表1)の他のファミリーのデアミナーゼ機能と選好性についても検証した。分析および検証方法によれば、M a f B 1 9、A I D / A P O B E C、新規A I D / A P O B E C様、T M 1 5 0 6、トキシンデアミナーゼ、X O O 2 8 9 7など他のファミリーでも、類似するS d d活性を持つ一連のデアミナーゼ(具体的なデアミナーゼは表5に示される)が見つかった。M a f B 1 9ファミリーを例に挙げると、実施例2では、M a f B 1 9クレードのタンパク質の一部(No. 101m)が一本鎖デアミナーゼの機能を有することを発見している。また実施例3では、人工知能支援によるタンパク質の三次元構造のクラスタリング分類に基づいて、M a f B 1 9デアミナーゼファミリーには全く異なる構造を有する2つのクレードが存在することが判明している(図11Bおよび図12)。本発明のデアミナーゼスクリーニングおよび同定方法を適用することにより、M a f B 1 9ファミリーの3つのタンパク質(No. 2~1241、No. 2~1231およびNo. 99)もS d d触媒活性を有することを発見し、その塩基編集配列のモ

40

50

チーフの選好性を得た(表5)。本発明の方法によって選定した、異なる編集特性を有する様々な新規シトシン塩基デアミナーゼは、塩基編集の手段を豊富にし、塩基編集システムを拡充し、標的DNA配列を正確に操作する能力を高めた。

【0236】

表5. 異なるファミリーのタンパク質の触媒および編集特性

【表5】

遺伝子ID	配列番号	ファミリー	ssDNA触媒活性	配列選好性			
				CC	AC	GC	TC
2-1241	56	MafB19	√	++	++	++	++
2-1231	57	MafB19	√	++	+	++	+
99	58	MafB19	√	++	+	+	+
101m	19	MafB19	√	-	-	-	+
2-1223	76	MafB19	x	-	-	-	-
2-1224	77	MafB19	x	-	-	-	-
2-1228	78	MafB19	x	-	-	-	-
2-1229	79	MafB19	x	-	-	-	-
2-1230	80	MafB19	x	-	-	-	-
2-1234	81	MafB19	x	-	-	-	-
2-1235	82	MafB19	x	-	-	-	-
3-2107A_c121x07	83	MafB19	x	-	-	-	-
4-2130A_c121x30	84	MafB19	x	-	-	-	-
181	63	AID/APOBEC	√	++	++	++	++
182	1	AID/APOBEC	√	++	++	++	++
2-1479	20	新規AID/APOBEC様	√	+	+	-	+
2-1478	21	新規AID/APOBEC様	√	+	-	-	-
2-39	22	TM1506	√	+	-	-	-
2-1128	74	トキシングデアミナーゼ	√	++	+	++	++
2-1114	75	トキシングデアミナーゼ	√	++	+	+	+
2-1430	59	XOO2897	√	++	++	++	+
2-1429	23	XOO2897	√	++	++	++	++
2-1440	60	XOO2897	√	+	+	+	+
2-1432	61	XOO2897	√	+	+	+	+
2-1437	62	XOO2897	√	+	+	+	+
2-1442	24	XOO2897	√	+	+	-	+

記号 ++ は編集選好性が高いことを表し、+ は編集選好性が低いことを表し、- は編集選好性がないことを表す。

【0237】

実験の結論

10

20

30

40

50

従来のデアミナーゼに基づくCBEには、編集効率が低く、編集ウィンドウが小さく、選好性が顕著であるなどの欠点があった。本発明の三次元構造に基づくタンパク質の機能予測方法を用いて、一連の新規シトシンデアミナーゼが得られた。

【0238】

これらのシトシンデアミナーゼは、優れた応用の可能性と用途を備えていることが証明されている。例えば、アグロバクテリウム媒介形質転換の遺伝子組換え大豆の毛状根では、APOBEC/AIDデアミナーゼは、大豆での他のCBEによる編集が特に困難であるGmALS1-T2およびGmPPO2の2つの部位を含む、評価した5つの部位全てで編集活性が低いことがわかった。なお、rAPOBEC1、hA3AおよびhAIDと比較して、MiniSdd7は、5つの部位でそれぞれ26.3倍、28.2倍、10.8倍のシトシン塩基編集レベルを示し、編集効率は67.4%と高かった。したがって、大豆作物における効率的なシトシン塩基編集の限界を突破するために、新たに発見されたこれらのSddタンパク質を重点的に利用した。次に、MiniSdd7を塩基編集に使用して、アグロバクテリウム媒介遺伝子組換え大豆植物株を取得することを試みた。内因性GmPPO2遺伝子を編集してR98C突然変異を生成し、これにより、カルフェントラゾンエチルに対する耐性を持つ大豆植物株が得られた。30個の遺伝子組換え大豆苗から2つの塩基編集用のヘテロ接合体を得た。カルフェントラゾンエチルによる10日間の処理後、野生型植物株は、しおれに敏感で根を張ることができなかったが、突然変異型植物株は、良好かつ正常に成長したことが顕著に観察できた。大豆植物で使用するための効率的なシトシン塩基エディターの開発により、将来的には様々な応用が可能になる。将来的には、構造予測と並行して配列決定の取り組みが行われることにより、機能性タンパク質の発見、追跡、分類、設計が大きく進歩すると信じている。現在、シトシン塩基エディターとして使用されているシトシンデアミナーゼはわずか数種類のみである。タンパク質工学と指向性進化のみに基づく標準的な取り組みは、編集特性を多様化するのに役立つが、これらの取り組みを確立するのは多くの場合困難である。三次元構造に基づくクラスタリング予測方法により、異なる特性を持つ一連のデアミナーゼを発見し、分析した。例えば、新たに発見されたデアミナーゼのうち、Sdd7とSdd6はいずれも治療および農業への応用に大きな可能性を有することがわかった。Sdd7は、テストした全ての種において強力な塩基編集能力を有し、最も一般的に使用されるAPOBEC/AID様デアミナーゼよりも高い編集活性を有している。驚いたことに、シトシン塩基の編集がこれまで困難であった大豆植物株において（植物遺伝子は通常、GC配列含有量が高い）、Sdd7は効率的な編集が可能であることを発見した。哺乳動物のAPOBEC/AIDデアミナーゼと比較して、細菌であるアクチノシンネマ・ミルム（*Actinosynnema mirum*）由来のSdd7は、大豆の生育に適した温度下でより高い活性を有する可能性があるとして推測している。Sdd6を分析したところ、このデアミナーゼは、高い標的編集活性を維持しながら、デフォルトで他のデアミナーゼより特異的であることがわかった。興味深いことに、AlphaFold2に基づくモデリングにより、タンパク質工学の取り組みでタンパク質のサイズを最小化することがさらに可能になることがわかった。これは、生体内治療用途でのウイルス送達にこれらの編集技術を使用するうえで非常に重要である。

【0239】

上記の内容は本発明の好ましい実施形態にすぎないが、当業者であれば本発明の方法を逸脱することなく幾つかの改良および補足を行うことが可能である。それらの改良および補足も当然ながら本発明の保護範囲にあるものとみなされる。

【0240】

関連する配列および簡単な説明

【0241】

10

20

30

40

50

【化 1 - 1】

> 配列番号:1 No.182 (AID/APOBEC クレード)

MICKLDSVLMTQKKFIFHYKNVRWARGRHETYLCFVVKRRIGPDSLSFDFGHLNRNRSGCHVELLF
LRHLGALCPGLSASSVDGARLCYSVTWFCSWSPCSKCAQQLAHFLSQTPNLRRLRIFVSRLYFCDEE
DSVEREGLRHLKRAGVQISVMTYKDFFYCWQTFVARRERSFKAWDGLHENSURLVRKLNRLQP
CETEDLRDVFALLGL

> 配列番号:2 Sdd9/SCP044/No.69 (SCP1.201 クレード)

MLDVEDDPHTFYVLGKTPVLVHNSECPWVKNALQELVGRKETSQKVFVDVVGKPIGPDIOQSGYKD
RELRRGVYETLRKSPFFQKHFPSTATWYVSLHVEAQYAVWMLRNRIKHATVVINNTYVCSDMNR
LHDNCMTAVPHILPEGYTMTVWKADRTEVTLRGKAPKE

10

> 配列番号:3 Sdd5/SCP017/No.55 (SCP1.201 クレード)

MGPLLDGIAARLEAVRAALLGEGGGAGDDEPPAVPWDRVERLRRELPPPVPNTGQKTHGRWIG
PDGQARPIVSGRDDKSVLVNPLLRRGKGAPGPTRRSDVEMKLAHMAARGIRHATVVINNTPCR
GPLRCDTLVPILLPEGSTLTVHGINENGTTRIRYTGARPWWS

> 配列番号:4 Sdd7/SCP016/No.57 (SCP1.201 クレード)

MLEAVRARLIGEGGGPGAVPEGGDGPPAVPAEEVERLRGELPPPVPVPGTGQKTHGRWIGPDGRVR
AIVSGRDEDAALVHAQLAAKGPDEPTRNSDVEQKLAHMAANGIRHVTLVINHRPCRGFDDSCD
TLVPIILPEGCTLTVHGQTDKGMVRVRVRYTGARPWWS

> 配列番号:5 Sdd4/SCP014/No.64 (SCP1.201 クレード)

MLDAMDAYLSEIAGNAPARAGPKAPEKQPGGSSSPRARDGRIDFRALLERLKAQGVVGLGRS
DDPIPDPFKKQNPACYQGLAPRQKQKPVVGNLFFPDGRRWINDVALESSRGEPFDLNIKPEYRS
LSPARGHLEGNVAAWMRSTFHQEMVLYINESPCRKHGKGLYTLHFVLPVLRGYSRNRDRGE
WRGNTFRGSGEAFTEGA

20

> 配列番号:6 Sdd76/SCP012/No.76 (SCP1.201 クレード)

MYVLAGNTPVLVHNTGPGCGEPGFVSDAANSLSGRRITGQIFDASGNPIGPEITSGGGSLADRAQ
SYLADSPNIRNLPKARYASADHVEAQYAVWMRENGVTDASVVINQNYVCGLPLGCQAAVPAIL
PRGSTMTVWYPGSGSPIVLRGVG

> 配列番号:7 Sdd6/SCP273/No.2-1146 (SCP1.201 クレード)

MVETRDKIIAAKSRSDAGLLAFQQATNGSIDSRAEAIANLQRAKTHLDEAQRVANSDDAAVDNY
INAILGGASAATAQPSA VIPASKPSRFKPMRTDPAKADEIRPHVKGDRAVATLWDADGNRVLGLH
SADDDGPAATAAWKPPWRDYVRLRRHVEAHAAARMHQDGHKTMVMIYINLPPCKYFDGCKLNL
EDILPKGSTLWMHRVFNQGGTKIYQFNGTGRAYV

30

> 配列番号:8 SCP021/No.2-1160 (SCP1.201 クレード)

【 0 2 4 2 】

40

50

【化 1 - 2】

MPIRLSGGLNLYQYAPETNNWIDPLGCSGHRRRHEKMPPEGARLTTGNLFRHAQDEGISPILSSKN
DEFYRRLIKIYAGTGMLNANIRGIASHVEPKAGLILNEDGSGWKIGSLYINYPNGPCLDCRITLMPFI
LNDGSILYVTFPTLGLDGYSGHFHGREPGFFREGTPCNLPHPE

> 配列番号:9 SCP038/No.54 (SCP1.201 クレード)

MADTPQPDNSNPLPARDGQRDLREQAEAIRAQLPPTVKPGTGKTHGCVWDEQGGPQSVTSGQDN
SAAAVWARLQALGIPLSGPPTATADVEQKVAIQMIQQGRQHVDVVINNEPCRGRFSCDTLVPIILP
EGSSLTVHGTNGFRKTYTGGAKPWSR

> 配列番号:10 SCP051/No.56 (SCP1.201 クレード)

MRAGCGVPDGRTRSASGKDEGFALALRTIRELGMTRGFPLRAADVEMKVASSEMRANGITSATLV
INHVPCCDGMFSCDRMVPVLLPAGSTLTVFGAGGFRMTYHGGEQLPCPTP

10

> 配列番号:11 Sdd59/SCP183/No.59 (SCP1.201 クレード)

MLLTPPPRPAAPTTRPKPLVARTGDAYPPGTEWALPLIVQPHPPVGGTVPVEGHVRLALRPESQISH
VFHPGGGHWTEQARARLRLVLPFGFWAVNLGHHVELQIAAWMTACQIHHAEVLNRPCCGERYG
LGCHQALPVLLPRGYRLTVSSTRGGPQPYQHHEGKA

> 配列番号:12 Sdd10/SCP018/No.60 (SCP1.201 クレード)

MLDAALGAVRRHIAALGTSGAERASPGANGSERVDELAERLPPTVVPNTSAKTHGWWFTGQGAA
QELISGEGPDARAAYEALREEGYPRPGMPFVAMHVEIKLAAHMRRNDIEHATVVINNIPCLVWG
CENLIGVVLP EGSSLTVHGSNGYERTFTTGGGRKPPWR

> 配列番号:13 SCP011/No.61 (SCP1.201 クレード)

MLGGVLPARSVMFPGHVEPDAHFGNPERHHPALVEVPIVWAGRQEDRTSTWARRVQRGFPRYT
VGAKTAGMFYNAGSQSWELLSGVDHRGGLTRKASQHSRMLSSGFFDGGKPLDTKSDHLRMLNYT
STHVETKAAIWARDSDQETIDVVTRNRYVCGESYDPPDDVDEPPGCYQAVESVLRREGQTMVWTT
DPENRVITIHGKGM

20

> 配列番号:14 SCP157/No.72 (SCP1.201 クレード)

MAAGGGVSRPPATRDGAANPTRVNPPEWLPWGLTEAARDLPRRQAKDPTSGVALINGERIPMRS
GRDPAADLKAAYKLIATTTDHLAKLAARMRRDQVMHAEVLTNNPPCDYEPYGCEKILSRLL
PAGAQLSVYVRDDDDQVRLWRTYIGNKAIA

> 配列番号:15 SCP008/No.74 (SCP1.201 クレード)

MTTGGGSDISRPPATRDTSATATEAPAQPELVPEGLTDAARDLPRRQAKDPTSGVALIGGERIPMR
SGRDPDAAADLKPAYKLIATTTDHLAKLAARMRRDHITQAAVVTTNNPPCDYTPYGCEKILSRLL
PAGARLAVYVRDDDDQVRHWRTYTGNGKAIA

> 配列番号:16 SCP013/No.75 (SCP1.201 クレード)

MIRARDRLTAVTASSRHPLVDQALQHVTAAIERLQVADRDAALAASALVAYGRTLGISLPVPPVVS
APTRGAAPVPSWIRQTGQDLPTRPDDHGPTHGQAFDSTGRPLSAEPWRSRNIASDRLRPIPLK
GFPWTLTDHVESRAAQMRPGAPREVSLLVNKEPCTDDPYGCDRILRHIIIPAGSRLTIYVRDPDA
PAGVRTVGQYEG

30

> 配列番号:17 Sdd3/SCP170/No.63 (SCP1.201 クレード)

MSASAQLNTYLAAIGNSTTTVEAQPEAAPPAAAAESLDSTPRLPDGGIDFHALAKRLGLLEARPTE
QPPFDPRRFNPACWQGLKPYDQAGTAEGNLFAPGKRWNTRPMQASKLEVGPQSDLHPQWRSRK
APWHIEGKIAAYMRQKGFDTGCVYLNARPCSGPDGCARNLPDLLPVGSTLHVHARYIDRTGETRF
YYREYRGTGKALT

> 配列番号:18 No.2-1158 (SCP1.201 クレード)

MPIRLSGGLNLYQYAPETNNWIDPLGCSGYRRRHEKMPPEGARLTTGNLFRHAQDEGIPPIFSEN
DEFYHRLIEIYAGTGVNLNAYIRGIASHVEPKAGLILNEDGSGWKIGSLYINYPDGPCCLGCRTLMPFIL
NDGSILYVTFPTLGLDGYSGHFHGGVSGFFREGTPCNLRHPE

> 配列番号:19 No.101m (MafB19 クレード)

MCGWSELASYRAREGMPARGSADDTFTAARLQIDGQVFFGRNAHGRPVDIRVNAQTKTHAEAD
VFQAKDAGATGTRAVLHVVRDFCRSCGATGGVGSMLMRGLGVEELLVHSPSGIFTINAVRRPSTP
RPLG

40

【 0 2 4 3 】

【化 1 - 3】

> 配列番号:20 No.2-1479 (新規 AID/APOBEC様クレード)

MRTRPAFGRCDGASDRGWVAVAGGYTSEASHVRRSPTPDGNSLLGGVAQLCHAFFHCSPTPELSS
HPDELCLRVACDPAGRKCETKGVVVVAALRDRAGDLRFLSRYSNCPSSSHAEYVVRDEELVRA
VMEMAPEDDARSSTKTPGSAGTLTYQRLQPCGSSDNRGPLWCS DALVAGLHRELLGPRGVSL
RVAVSYTYRAHWVDRGFESERERRWWGPKVEAAREGIRVFAAAKDGVTLEALNAEDWAFVLS
LCDEEDVARDYAAAFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEFGEF
AG

> 配列番号:21 No.2-1478 (新規 AID/APOBEC様クレード)

MADLDDVQDPLDLDALDSTKDEADDSVLSEIAVNDTSVDDGVEDPDHEHKKIAKAGDKVLGNKK
EFCGAFYHVPRSKSGCLDKQSCAIAKRGHDA TPLTAVALVKEYEQESESSEWAIKSVRRYTNCSDKM
KHAEFFLMDIDCQLEARHKGEEGFLDFWNKKKWKQITMYLTMQPCHLSTDTGGTKEDQSCCEV
MIKAKEKLGDNVEIVIKPTHLCQVWGYKGPKEPKNAEKGVKLFKTTGIELECMKEGDWKYL
LQYAQPEVENKLPDYDTSRRKTEDEKIGEELHNQQLAPELLAQLSVNEKRRK

10

> 配列番号:22 No.2-39 (TM1506 クレード)

MVYSMDQKNKVTTEKLKEGGYSFVLYKDGWSTSEKRGIAPIPELLKENKELLRGAYVADKVIK
AAALLLIEGGISYLHAEIIEHAIEVLQNSNIEYEQELVPYIVNRSRSGDMCPMEETVLDVTDTKIAF
ELLQEKIKKMQAAMQAQNMK

> 配列番号:23 No.2-1429 (XOO2897様クレード)

MISDAAVAGIASKMAEKYYSACKLRSRPISTLGVIGKPVPEYSCDGIVPYNSTDLGRMAYKARV
EAGFGIFGGRNVAVARVPGWDDPKTGDLVVGFSGNGFHAEDHVLEQLTKKDISPKKITELYSER
QPCAACGNLENHLSPGTEITWSVQWGSDEMNSAFTELLGKLIQQQ

> 配列番号:24 No.2-1442 (XOO2897様クレード)

MAPDSLWVFDPLGLIVLQQVPYNDHPLFGAVSEFIQKRSRDLRGRNVAAVLLDDGTVIVRASEG
GGNHAERVLMLGLSEVPAKVAVVYTERSPCTGRINCHDLLDSSLGADVPPYYTHEMIRGQEGKT
AQIEADRNQFCRGG

20

> 配列番号:25 SpCas9

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRNRYCYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVSDTKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVDFLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD
LTLKALVROQLPEKYKEIFFDQSKNGYAGYIDGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTFRIPYVVGPLARGNSRFAW
MTRKSEETIPWNFEVVDKASASQSFIERMTNFDKNLPNEKVLPHSLLYEYFTVYNELTKVKY
VTEGMRKPAFLSGEQKKAIVDILLFKTRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH
DLLKIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGR
LSRKLINGIRDKQSGKTLDFLKSDFANRNFMLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLA
GSPAIAKKGILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKQKNSRERMKRIEIEGKELGSQ
ILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLLKDDSIDNKVLR
SDKNRGKSDNPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRLVE
TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSLSVDFRKFDFQYKVRINNYHHAY
LNAVVGTAIIKYPKLESEFVYGDYKVDVRKMIKSEIQEKATAYFFYSNIMNFFKTEITLAN
GEIRKRPLIETNGETGIVWDKGRDFATVRKVLSPQVNVKKEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKEKGGKSKLKSVKELLGITIMERSSEKPNIDFLEAKG
YKEVKKDLIILPKYSLFELENKRMLASAGELQKGNELALPSKYVNFYLAHYEKLKGPED
NEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHRDKPIREQAENIIHLFTLNLG
APAAFKYFDTTIDRKYRSTSTKEVLDTLIHQSTGLYETRIDLSQLGGD

30

> 配列番号:26 nCas9(D10A)

MDKKYSIGLAIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKR
TARRRYTRRNRYCYLQEIFSNEMAKVDDSFHRLSEESFLVEEDKKHERHPIFGNIVDEVAYHEKY
PTIYHLRKKLVSDTKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSDVDFLFIQLVQTYNQLFEE
NPINASGVDAKAILSARLSKSRLENLIAQLPGEKKNLFGNLIASLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQD

40

【 0 2 4 4 】

【化 1 - 4】

LTLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
 LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPLARGNSRFAW
 MTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLYEYFTVYNELTKVKY
 VTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYH
 DLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWGR
 LSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQDLSLHEHIANLA
 GSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIIEEGIKELGSQ
 ILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTR
 SDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGSELKAGFIKRQLVE
 TRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSCLVSDFRKDFQFYKVVREINNYHHAHDAY
 LNAVVGITALIKKYPKLESEFVYGDYKVDVRKMIKSEQEIGKATAKYFFYSNIMNFFKTEITLAN
 GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLVSMQVNVKTEVQTGGFSKESILPKRNSDKLIA
 RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGGKSKLKSVKELLGITIMERSSEKPNIDFLEAKG
 YKEVKKDLIILPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFYLAASHYEKLGKSPED
 NEQKQLFVEQHKHYLDEIIEQISEFSKRVLADANLKVLSAYNKHHRDKPIREQAENIHLFTLTNLG
 APAAFKYFDTIDRKRYSSTKEVLDATLIHQISITGLYETRIDLSQLGGD

10

>配列番号:27 UGI
 MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDENVMLLTSDAPEYKP
 WALVIQDSNGENKIKML

>配列番号:28 Ddd1/SCP177
 MSLPEYDGTTHGVLVDDGTQIGFTSGNGDPRYTNYRNNGHVEQKSALYMRENNISNATVYHN
 NTNGTCGYCNTMTATFLPEGATLTVVPPENAVANNSRAIDYVKTYTGTSPNDPKISPRYKGN

>配列番号:29 Ddd2
 MEDFHYYHVGKCRLLVHNANCNQEKPVLPKYDGKTTGVMVTPDGKQISFKSGNSSTPSYPQYK
 AQSASHVEGKAALYMRENGINEATVFHNNPNGTCGFCDRQVPALLPKGAKLTVVPPSNSVANNV
 RAIPVPKTYIGNSTVPKIK

20

>配列番号:30 Ddd3/SCP001
 MLSSSYNAFALTYGVLILDDGKQYSFNSGKPDPIYRNYIPASHVEGKAAIYMRENKIQSGTVYHNN
 TDGTCPCYCDKMLPTLLEKDSLKVVPQATSSKKGWITNEKIYIGNDKIPKTAR

>配列番号:31 Ddd4
 MGGDEEEENLTSNNEKKNANKQKIELPPYDGKTTYGVLILDDGKQYSFNSGKPAPIYRNYIPASH
 VEGKAAIYMRENKIQSGTVYHNNTDGTCPYCDKMLPTLLEKDSLKVVPQATSSKKGWITNEK
 IYIGNDKIPKTAR

>配列番号:32 Ddd6/SCP009
 MALLREAYPSMEGATLPPFDGKTTIGLMFYTDASGQYQVKLFSGEKVLVSNYDATGHVEGKAALI
 MRNEKITEAVVMHNHPSGTCNYCDKQVETLLPKNATLRVIPPENAKAPTSYWNQPTTYRGDGK
 DPKAPSKK

30

>配列番号:33 Ddd7/SCP103
 MIGLMGGLNLYQYAPNSIAWTDWWGLAGSYTLGYSQISAPQLPAYNGQTVGTFYYVNGAGGLE
 SRTFSSGGPTYPNYANAGHVEGQSALFMRDNGISDGLVFHNNPEGTCGFCVNMTEITLLPENSKL
 TVVPEGAIPVKGATGETRFTFTGNSKSPKSPVKGEC

>配列番号:34 Ddd8/SCP234
 MQDNTNIIDNRPKLPDYDGKTHGILVTPNSEHIPFSSGNPNPNYKNIYIPASHVEGKSAIYMRENGI
 TSGTIYNNNTDGTCPYCDKMLSTLLEEGSVLEVIPPINAKAPKPSWVDKPKTYIGNNKVPKPNK

>配列番号:35 Ddd9/SCP003
 MGKLSSESQATLSVAQRLLATIGEEGKTAGVLELDGELIPLVSGKSSLPNYAASGHVEGQAALIMR
 DRGATSGRLIDNPSGICGYCKSQVATLLPENATLQVGTPLGTVTPSSRWSASRTFTGNDRDPKPV
 PR

>配列番号:36 Ddd10/SCP004

40

【 0 2 4 5 】

【化 1 - 5】

MASPAVGTNAAGSSGKNVRMPRDYASELPEYDGKTTTHGVLVTNEGKVIQLRSGGKEEPYTGKYA
VSASHVEGKAAIWIRENGSSGGTVYHNNTTGTGCGYCNQVVKALLPEGVELKIVPPTNAVAKNAQA
RAVPTINVGNGTQPRKQK

>配列番号:37 Ddd11/SCP271
MQGTSSDTIAEMLNSASQPRTAGVLDIDGELTPLTSGRPSLPNYIASGHVEGQAAMIMRQQQVQS
ATVYHDNPNGTGCGYCYSQLPTLLPEGAALDVVPPAGTVPPSNRWHNGGSPFIGNSSEPKPWPR

>配列番号:38 Ddd12/SCP005
MGVAGGAATNADAQALLGSIRQAGKTAGVLDNIDGDLMLPLVSRKSSLPNYAASGHVEGQAALIMR
ERGVSSAELLIDNPNGICSYCTSQVPTLLPEGAQLMVRPPLGTVPQTQWWFNGRTFLGNAANPKPSP
W

10

>配列番号:39 Ddd13/SCP006
MHNINGCGPSAVQQLSANGEPGKTAGVLDLNGELTSLVSGKGELPNYAASGHVEGQAAMMMR
AEKATSATLYIDNPNGICGYCRSQIATLLPEGATLEVVTPLGTVEPTARWSSKVFTGNERYPKGW
VE

>配列番号:40 Ddd14/SCP007
MGSVVGGGIAATGAKALTTGKKLTESPGTLNAAQRLLASIGEEGKTAGVLEVDGALFPLVSGKS
VLPNYAASGHVEGQAALLMQGMGATNGRLLIDNPNGICGYCTSQVPTLLPENAVLEVGTPLGTVT
PSARWSASKPFIGNDREPKPWPR

>配列番号:41 No.2-1157 (SCP1.201 クレード)
MDPIRLSGGLNLYQYAPETNNWIDPLGCSGHRRRHEKMPPEGAPLTTGNLFRHAQDEGIPPIFSRK
DDEFYHRLIEIYAGTGVLNAYIRGIASHVEPKAGLILNEDGNGWKIGSLYINYPDGPCPCRRLLMPF
ILNDGSILYVTFPTLGLDGYSYGHFHGGVSGFFREGTPCNLRHPE

20

>配列番号:42 No.73/SCP158(SCP1.201 クレード)
MPGGGEINRPPTTQADDVPQPVGSEWERTEPDALPGTVRAAVERLQPRPAGSTRPTLGVFNAGEEIT
SGGGDRSLAADLDHDLRGPVTFYDHVESKAAARMRRTGSTESDLAIDNTVCGTNDRDQSYWP
TCDKILPAILPNGSRLRVWVTRDGGVTWWHRVYIGTGERITK

>配列番号:43 No.2-1145/SCP315 (SCP1.201 クレード)
MSPKKPTASSDLKAIGERLGLKPCGELLGTLPAAMKPNQGRTRGRWHKHPDRELTSAGANDRDWE
HVKDFWHHNIWSGTAEDTEPRWLAHLELKFAMTMRRTKSEPVQVHEEITINHPDGPCPCQCL
LLPYFLEEGSSLTIHWPAGSATYIGRPYFDRPLRDVVKPYINEEQQ

>配列番号:44 No.2-1157/SCP020 (SCP1.201 クレード)
MDPIRLSGGLNLYQYAPETNNWIDPLGCSGHRRRHEKMPPEGAPLTTGNLFRHAQDEGIPPIFSRK
DDEFYHRLIEIYAGTGVLNAYIRGIASHVEPKAGLILNEDGNGWKIGSLYINYPDGPCPCRRLLMPF
ILNDGSILYVTFPTLGLDGYSYGHFHGGVSGFFREGTPCNLRHPE

30

>配列番号:45 No.2-1156 (SCP1.201 クレード)
MDPIRLSGGLNLYQYAPETNNWIDPLGCSGYRRRHEKMPPEGARLTTGNLFRHAQDEGIPPIFSSE
NDEFYRLIKIYAGTGILNANIRGIASHVEPKAGLILNEDGSGWKIGSLYINYPNGPCLDCRRLMPFI
LNEGSILYVTFPTLGLDGYSYGHFHGREPGFFREGTPCNLRHPE

>配列番号:46 Sdd2 (SCP1.201 クレード)
MAPDSLWVFDPLGLIVLQVPYNDHPLFGAVSEFIQKSRSDLRGRNVA AVLDDGTVIVRASEG
GGNHAERVLMLSEVPAKVVAVYTERSPCTGRINCHDLLDSSLGADVVPVYTHEMIRQEGKT
AQQIEADRNQFCRGG

>配列番号:47 SCP008 (SCP1.201 クレード)
MTTGGGSDISRPPATRDTSATATEAPAQPELVPEGLTDAARDLPRRQAKDPTSGVALIGGERIPMR
SGRDPDAAAADLKPAYKLIATTTDHLEAKLAARMRRDHTQAAVVTTNNPPCDYTPYGCEKILSRL
PAGARLAVYVRDDDGQVRHWRTYTGNGKAIA

40

>配列番号:48 SCP011 (SCP1.201 クレード)

【 0 2 4 6 】

【化 1 - 6】

MLGGVLPARSVMFPGHVEPDAHFGNPNERHHPALVEVPIVWAGRQEDRTSTWARRVQRGFPRYT
VGAKTAGMFYNAGSQSWELLSGVDHRGGLTRKASQHISRMSSGFFDGKPLDTKSDHLRMLNYT
STHVETKAAIWARDSDQETIDVVTRNRYVCGESYDPPDDVDEPPGCYQAVESVLREGQTMRVWTT
DPENRVITIHGKGM

>配列番号:49 SCP013 (SCP1.201 クレード)
MIRARDRLTAVTASSRHPLVDQALQHVTAAIERLQVADRDAALAASALVAYGRTLGISLPVPPVVS
APTRGAAPVPSWIRQTGQDLPTRPDDHGPETHGQAFDSTGRPLSAEPWRSRNIASDRLRPIPGLK
GFPWTLTDHVESRAAQMMRRPGAPREVSLVVNKEPCTDDPYGCDRILRHIIIPAGSRLTIYVRDPDA
PAGVRTVGGQYEG

>配列番号:50 ミニSdd7
MEGGGPGAVPEGGDGPAPVPAEEVERLRGELPPPVVPGTGQKTHGRWIGPDGRVRAIVSGRDEDA
ALVHAQLAAKGPDEPTRNSDVEQKLAAHMVANGIRHVTLVINHRPCRGFDDSCDTLVPILPEGC
TLTVHGQTDKGMRRVRYTGGARPWWS

10

>配列番号:51 ミニSdd4
MDPKKQNPACYQGLAPRQKPKVVRGNLFFPDGRRWNDVALESSRGEPFDLNIKPEYRSLSPAR
GHLEGNVAAWMRSTFHQEMVLYINESPCRKHGKGLYTLHEFLPRGYVLHVWSRNDRGWEWRGN
TFRSGEAFTEGA

>配列番号:52 ミニSdd9
MCPiWVKNALQELVGRKETSgKvFDVDGKPIGPDIOGSYKdRELRRGVYETLRKSPFFQKHFPSTA
TWYVSLHVEAQYAVWMLRNRIKHATVVINNTYVCSDMNRLHDNCMTAVPHILPEGYTMTVWK
ADRTEVTLRGKAPKE

>配列番号:53 ミニSdd6
MPASKPSRFKPMRTDPAKAEIRPHVgKDRAVATLWDADGNRVLGLHSADDDGPAATAAWKPP
WRDYVRLRRHVEAHAAARMHQDGHKTMVMYINLPPCKYFDGCKLNLEDILPKGSTLWMHRVF
QNGGTKIYQFNGTGRAYV

20

>配列番号:54 ミニSdd10
MPGANGSERVDELAERLPPTVVPNTSAKTHGWWFTGQAAQELISGEGPDARAAAYEALREEGYP
RPGMPFVAMHVEIKLAAHMRRNDIEHATVVINNIPCLVWGCENLIGVVLPEGSSLTVHGSNGYE
RTFTGGRKPPWR

>配列番号:55 ミニSdd3
MPRRFPACWQGLKPYDQAGTAEGNLFiAPGKRWNTRPMQASKLEVGPQSDLHPQWRSRKAPW
HIEGKIAAYMRQKGFDTGCVYLNARPCSGPDGCARNLPDLLPVGSTLHVHARYIDRTGETRFYR
EYRGTGKALT

>配列番号:56 No.2-1241 (MafB19 クレード)
MGLEGTPCDGFALAAARRKSLGLPAAGSEGDSTLSLLRINGQSFEGINSSDQNPKITLDRVNAQ
TKTHAEAEAVQKAVNAGMAGKASHAEMWVDRDPCRACGIPGAGGLRSLARNLGCPIVHSPSGT
QVYTPTK

30

>配列番号:57 No.2-1231 (MafB19 クレード)
MLGPPLDLNPANRAPEFGRCdGTSWIDSYRTINNATDLfGRPVWPNHRGTVAVARIDGDIYFGVN
SKAPGYSDADWNLAAGLRDQMALEHPELIRGESRGRPLDAVFHAEANLLIRASRYVGLVKRSI
EVQVDRPVCWSCEQALPKVGLELGDpYVTIREVRSGRASVMWQGEWLVRKK

>配列番号:58 No.99 (MafB19 クレード)
MGWVDPLGLVSGGAWDAISFFRDQNSLLSVVDEDLAASGAKNAQNTVALLRVGDREFIGVNSR
IQNPKNPFTAGPINNITKFHAEGNAQQAIDAGMVGKHRIAEMWVDRDLCHACGPSNGVGLS
LGLDAIIVHTPAGTRKFNAPCAG

>配列番号:59 No.2-1430 (XOO2897様クレード)
MVKGLDGFVKTCTRTRGKLTARASTAVGGCPVGLVAYNSEEMSHWAYRYRTESEYFEGDHVNA
VAKVPGWNDPRTGDFIANSKFSGHSETEILGKLEAKGFTPGQITALYTERQPCPACASVLTGSLKE
GTPVTWSVPYHPDYAKESRSLDSYVRQANGQQRARPTTTQRLTEGNEAHD

40

【 0 2 4 7 】

【化 1 - 7】

> 配列番号:60 No.2-1440 (XOO2897様クレード)

MRLGQSVDPRLLEMAKEARVTQAGISREAFASYNVATARVRVGTETIRYLDAGNSPGRMLHSEWDW
LITQVEELRRVHGRESVALEQLFSERIPCCECLPKLERLFNAEVFYAVAKRGTRATDLMKAYGLR

> 配列番号:61 No.2-1432 (XOO2897様クレード)

MPDPLGLAPAANDRAYVPNPLTWADPYGLACTGTTEPGSTDLSSQAVIQERLRLGKKGNNFAAAR
YIDDNGVEQIAVAASSKGFQFMHAERKLRVQYGDKITEVYSEFEPCIGTNQCRKTLGDMGIKYTYS
WAWTLSKDGVAANAARKAYVDQIFDDAEAGNWAAPWAD

> 配列番号:62 No.2-1437 (XOO2897様クレード)

MNIDGDSYLVPGALAAAMFLHKPGRGGKGVGYGTTDLGQSVRLQRLIDKNRGMNTNYAAARLDD
GDVIVGKSKKHVHAEHLFQQAGKRKIVELYSEREPSCNKCEDLVKDIPFVSVSFKWNHPDRIKQ
DAIRDKANADLKDAVRSLFNSP

10

> 配列番号:63 No.181 (AID/APOBEC クレード)

MQGYIVDESGRVLNANGLPIASLPPADDLSKWANYTVESGLHDDLAALENRLDLLYRQQFGLPM
APPWHLETQLAYRVATREVALRDSTLRLVMNPNPGVCDVPLTGTGPDQRQAVAGCIQAVK
MLLPAGTTMIIYYPDPDDPAELLEITVRGVGRWLD

> 配列番号:64 rAPOBEC

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHISWRHTSQNTNKHVEVNF
EKFTTERYFCNTRCSITWFLSWPCGECRAITEFLSRYPHVTLFIYIARLYHHADPRNRQGLRDLI
SSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPYPHYLWVRLYVLELYCILGLPCLNLRKQP
QLTFFTIALQSCHYQRLPPHILWATGLK

> 配列番号:65 DddA

GSYALGPYQISAPLPAAYNGQTVGTFYYVNDAGGLESKVFSGGPTYPNYANAGHVEGQSALF
MRDNGISEGLVFHNNPEGTCGFCVNMTELLPENAKMTVVPPEGAIPVKRGATGETKVFTGNSNS
PKSPTKGGC

20

> 配列番号:66 SCP090 (SCP1.201 クレード)

MDGPHGTPVLDRIAALREELPPPAPVPGKQKTDGRWFDGNGAVRDSVSGKDVDSSEAWRLLRES
GIPLPRPPVVAHAEMKVAAMRRLNVRHAVLVITNVPCDERWSCENLLPAVLPVGCSSLVHGGP
YQRTFHGRTPKW

> 配列番号:67 SCP015 (SCP1.201 クレード)

MRPTTPPGPHARWRPDPSPAPHVAAIRRVGWPKKQSDDDVRRARGQLYHRDGPWNASMLIASR
RGPASQRTDLKEPWASDPGYTTGWHIEGNTAALMVKHQQRDAVLYINQAVCGAEEPPQDPKRCHS
NIVAMLSVGYALYVHSVQESGWLRRRVYKGTGEAIR

> 配列番号:68 SCP278(SCP1.201 クレード)

MQRHGGRVRLLSGENDDPHSWQQQAARFLRETFFDKGGLAVLSRHVEIQLAVRLRHRPTNEVV
HEVLVIDRVVCGRDPRTQGREYTCDTVLPFVLDEGATLTVVEHDGARVTVYRGRGR

30

> 配列番号:69 SCP287 (SCP1.201 クレード)

MKILYIKSAEGYPSMLKNNPRIPTNARSYTHVEGRAASIMRQSGIKSAKLTINNTNGVCDPCRGN
MEKSLLPDGGKLVNRYPDGQGGYTQRILI

> 配列番号:70 SCP341 (SCP1.201 クレード)

MREFGLPIEPPAWHLETQLAYRVSKREVALRENTLRLVMNPNPGVCDVPAKDRGPDGQRQVVA
GCVQAIQMLLPAGTTMIIYYPDPANPAKLLQVTVRGVGRWLD

> 配列番号:71 SCP353 (SCP1.201 クレード)

MLQGWRLAVDGDSSDRELASGTKLASGQTDPSYTAAVQRARELGLARGGFVPDIARHIEIKEAST
MTAGETRITIVIGKDPGIDPVTNVSCHPFLRYFLPPGATLIVYGPRGEPYRYEGKRTS

> 配列番号:72 SCP357 (SCP1.201 クレード)

40

【 0 2 4 8 】

【化 1 - 8】

MVGATTDVPTQPGGRARSAPRPPRPKTHGRWCSDGNAVVLSESGKGGEYEAATRARGVAL
GLAKGIPNAEPSIARHVETQFVSRMIDQIEYAEIEINRPVCGTTPKDQQ

> 配列番号:73 SCP372 (SCP1.201 クレード)
MQGYIVDESGRVLNANGLPIASLPPADDLSKWANYTVESGLHDDLAALLENRLDLLYRQQFGLPM
APPAWHLETQLAYRVATREVALRDSTLRVMNPNPQGVCDVPLTGTGPDQRQAVAGCIQAVK
MLLPAGTTMIIYYPDDPAELLEITVRGVGRWLD

> 配列番号:74 No.2-1128 (トキシシ デアミナーゼ クレード)
MQGYIVDESGRVLNANGLPIASLPPADDLSKWANYTVESGLHDDLAALLENRLDLLYRQQFGLPM
APPAWHLETQLAYRVATREVALRDSTLRVMNPNPQGVCDVPLTGTGPDQRQAVAGCIQAVK
MLLPAGTTMIIYYPDDPAELLEITVRGVGRWLD

10

> 配列番号:75 No.2-1114 (トキシシ デアミナーゼ クレード)
MGAPRTMGNMGVAQISIPGVQSKMAASSQIPDPTAAQRALGFVGEVNETFPSASVWTGGDTPYLL
NRKVDSEAKILNNIAAQLGDNTSASGTINLFTERPPCESCSNTIIFQEKYPNIKINVMDSNGVIRPS
KR

> 配列番号:76 No. 2-1223 (MafB19 クレード)
MQGQFDGVPANEFIKAWGEAMV AEAAGLGIVAGLGRFGLWGAKGATPVVTAEGRIGNSVFTD
VNQTARPAQANPNQPTLIADRVDAKIAAKGTPHPNGNMADAHAEIGVIQKAFNEGKTVGSDMT
MNVVVGKDVCGYCRGDIAAAASKSGLKSLTIQAKDDITGLPKTYWVEVGMKSIREKKI

> 配列番号:77 No.2-1224 (MafB19 クレード)
MYLPRGTSVTSKETVAKDPVSLAQADNEAGILVDRNVIVGGTKGVSPVVTAEKGIGKTFDFN
QTARPAEANASQPTLISDRVTAKADASGKVL PNGNMADAHAEIGVIQQA YTAGKTMGASMELT
VSGKAVCGYCRGDIAAAMAEKSGLTSLEVKEVATGKTLYWQPMRALRERN

20

> 配列番号:78 No.2-1228 (MafB19 クレード)
MKKPTGSIVSPETTIVQESSKILDKKTHTSIPKVEAELIDKETGKIFKDTNQGPNRDPYFLGDKSRPTLI
NDRIEAKVEKNPSKYL PNGNMASAHAEVGTIQAFEDGITVGRDMNMKVTK EAVCGYCRGDIAA
MADKAGLKSLTVYEESTGKTLYWNP GMKSLKEKK

> 配列番号:79 No.2-1229 (MafB19 クレード)
MAGVLAPEVYLAKTPRVGDNSARGVGDGRSTTPKVTA EAEVDGVKFNDTNQNARPSEAA NPNI
TLISDDIQVKIDKNPDKFPNGNMATAHA EVGAIQQA YDAGKTQGKNMTRVTGEDVCDYCRSD
LRKAADKSGLNSLSVYEETTGR TLTWTRREDGTIGVKIIEPEG

> 配列番号:80 No.2-1230 (MafB19 クレード)
MGVDRKTAQGYAETKQGM DTIVASVTPILGAAA AKQLSKVVDANIKVVAEGNVNGAKFSDTNQ
GARPSNLADV NKP TLIDGRIQAKIDKQNKPL PNGNMATAHA EVGVIQQA FEKGMSQGREMTMSV
SKEPVCYCRSDIAAMADKAGLKSLTIYEETTGSVLYWQPMKSLKIRD

30

> 配列番号:81 No.2-1234 (MafB19 クレード)
MDRQTAESYTETKQGLEIIAASVTPILGSVAAKQLSKIVDANLKV VARGNV D GARFSDTNQV RP
SQLADFNKPTLINDVVQAKIDKRPDKNY PNGNMGT AHA EVGVIQQA FDKGMTQGREMAMSVGG
KEVCNYCLSDVRIMAEKAGLKSLTIYEATGNVLFWQQGMKKIENRGP AK

> 配列番号:82 No.2-1235 (MafB19 クレード)
MTKSALGEIVIVVSDLVIP TNYVEILPVGKLSKVAKILKIGEDG TKSAGRLAEELAE LQKVDIKFGK
TLPGAKAPITVTAESNIGGKHMFD TNQTARPEVNRTNPTLAAGNAKIDPSNPNLTMKNAHAEIAL
IQRAYDAGLTKGETMQV LVRGKEVCDHCGQVMKTM YERSGLSKLIHDTTSGTTTYYKVIDAK
TKIATTKIEV

> 配列番号:83 No.3-2107_cl21x07 (MafB19 クレード)

40

【 0 2 4 9 】

50

【化 1 - 9】

MDDSYMYMKQALLEAQKAGERGEVVPVAVVCKDRIIARAHNLTETLTDVTAHAEMQAITAAAST
 LGGKYLNECALYVTVPEPCVMCAGAIAWAQTGKLVFGAEDEKRGYQRYAPQALHPKTMVVKGV
 LADECAALMKNFFAAKRK

>配列番号:84 No.4-2130A_cl21x30 (MafB19 クレード)

MTKSALGEIVIVVSDLVIPNTNYVEILPVGKLSKVAKILKIGEDGTKSAGRLAEELAELOKVDIKFGK
 TLPGAKAPITVTAESNIGGKHMFDTNQTARPEVNRNTPTLAAGNAKIDPSNPNTMKNAHAEIAL
 IQRAYDAGLTKGETMQVLVRGKEVCDHCGQVMKTMYERSGLSKLIHDTTSGTTTTYYKVIDAK
 TKIATTKIEV

>配列番号:85 HsJAK2 (5'-3') 図 15, 17, 18
 AGGCAGGCCATTCCCA

>配列番号:86 HsJAK2 (3'-5') 図 15, 17, 18
 TCCGTCCGGTAAGGGT

>配列番号:87 HsSIRT6 (5'-3') 図 15, 17, 18
 TCGCCGTACGCGGACAAGGG

>配列番号:88 HsSIRT6 (3'-5') 図 15, 17, 18
 AGCGGCATGCGCCTGTTCCC

>配列番号:89 HsHEK2 (5'-3') 図 16A
 GAACACAAAGCATAGACTGCGGG

>配列番号:90 HsHEK2 (3'-5') 図 16A
 CTTGTGTTTCGTATCTGACGCC

>配列番号:91 HsWFS1 (5'-3') 図 16A
 CAGCAGTATGGTGCCTGTGCGG

>配列番号:92 HsWFS1 (3'-5') 図 16A
 GTCGTCATACCACGCGACACGCC

>配列番号:93 OsAAT 図 21
 ACAAGGATCCCAGCCCCGTGAAGG

>配列番号:94 OsACC1または OsACC-T1 図 21
 TCTCAGCATAGCACTCAATGCGGTCTGGG

>配列番号:95 OsCDC48-T1 図 21
 TAGCACCCATGACAATGACATGG

>配列番号:96 OsCDC48-T2 図 21
 GACCAGCCAGCGTCTGGCGCCGG

>配列番号:97 OsDEP1 図 21
 CTAGCACATGAGAGAACAATATTGGG

>配列番号:98 OsODEV 図 21
 GCACACACACACTAGTACCTCTGG

>配列番号:99 HsEMX1 図 22
 ACAAAGTACAAACGGCAGAAGCTGGAGG

>配列番号:100 HsHEK2 図 22
 ACTGGAACACAAAGCATAGACTGCGGG

>配列番号:101 HsWFS1 図 22
 CCTGGCAGCAGTATGGTGCCTGTGCGG

【 0 2 5 0 】

10

20

30

40

【化 1 - 1 0】

> 配列番号:102 MmHPD-T1 図 30C
GCAACCAACCCGACCAAGAAATGCAGT

> 配列番号:103 MmHPD-T2 図 30C
AGTCATTCAACGTCACAACCACCAGGT

> 配列番号:104 GmALS1-T1 図 30D
TCTCCATCGACGCACCGCCGGGG

> 配列番号:105 GmALS1-T2 図 30D
CAGGTCCCCCGCCGGATGATCGG

> 配列番号:106 GmALS1-T3 図 30D
GATCCATTACTGGGAATCATCGG

> 配列番号:107 GmPPO2 図 30D
AAGCGCTATATTGTGAAAAATGG

> 配列番号:108 GmEPSPS 図 30D
CAATGCGTCTTTGACAGCAGCTGTGG

> 配列番号:109 GmPPO2 野生型 図 30F
CATAAGCGCTATATTGTGAAAAATGGGGCA

> 配列番号:110 GmPPO2 編集済み 図 30F
CATAAGTGCTATATTGTGAAAAATGGGGCA

10

20

【 図 面 】

【 図 1 】

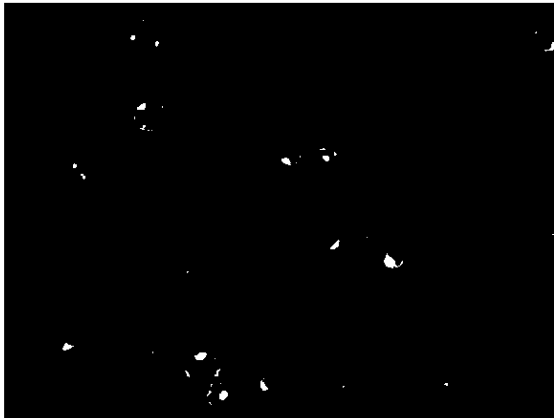
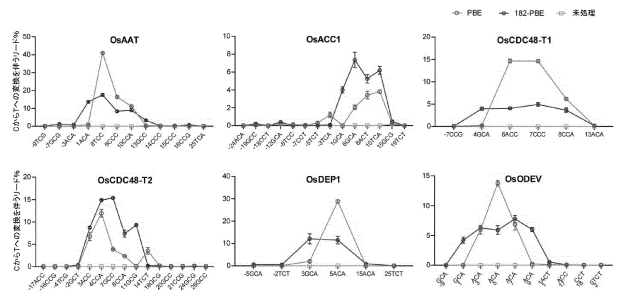


图 1

30

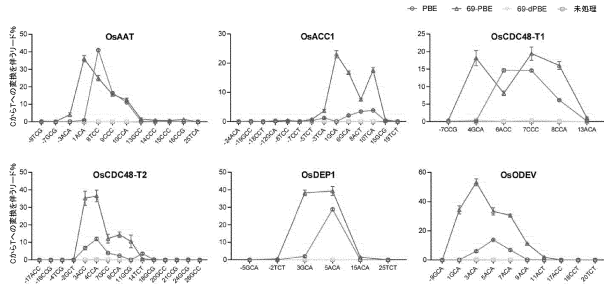
【 図 2 】



40

50

【 図 3 】



【 図 4 】

OsACC-T1

PBE	0.06	0	0.21	0	0	0.38	1.26	0	1.04	2.42	5.46	0.17	0
NO.2-1479	0.06	0	0.25	0	0	0	0	0	15.25	0	0.40	0.15	0
NO.76	0.06	0	0.19	0	0	0	0	5.67	29.58	2.99	5.96	0.61	0
NO.2-1146	0.06	0	0.24	0	0	0	0.22	18.95	8.06	6.08	0.20	0	0
NO.64	0.06	0	0.21	0	0	0	0	10.47	31.46	3.33	3.50	0.14	0
NO.55	0.20	0	0.47	2.00	1.34	3.84	9.76	28.30	38.44	19.17	40.17	0.31	0
NO.69	0.08	0	0.23	0	0	0.28	4.66	28.65	19.61	18.52	32.56	0.88	0
NO.57	0.05	0	0.26	1.16	0.19	5.09	4.05	30.64	26.71	22.86	1.11	0	0
NO.2-1429	0.07	0	0.21	0	0.02	0.58	2.83	23.11	11.50	0.15	0	0	0
A3A	0.05	0	0.45	0.51	0.66	3.51	6.68	8.10	13.73	12.51	14.57	0.60	0
WT	0.07	0	0.23	0	0	0	0	0	0.01	0	0	0.12	0

-19GCC -18CCT -12GCA -8TCC -7CCT -5TCT -3TCA 1GCA 6GCA 8ACT 10TCA 15GCG 19TCT

图 4

10

【 図 5 】

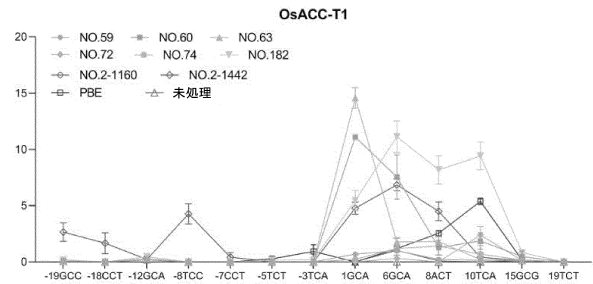
OsCDC48-T2

PBE	0	0	4.47	10.68	1.94	0.87	0.01	5.17	0.12	0
NO.2-1479	0	0	0.01	18.78	0.01	2.06	6.36	0	0.06	0
NO.76	0	0	19.98	24.09	5.86	1.51	0.71	1.46	0.10	0.02
NO.2-1146	0	0	28.16	38.31	12.50	22.22	0	0	0.12	0
NO.64	0	0	40.13	42.18	8.70	2.15	0.01	0.02	0.19	0.03
NO.55	1.45	1.45	43.18	35.91	32.61	7.99	18.59	1.51	0.03	0
NO.69	0.02	0.32	24.22	56.48	24.73	2.59	0.10	0.01	0	0
NO.57	0.68	0.01	3.64	3.64	0.98	0.01	0	0	0	0
NO.2-1429	0	0.55	23.23	10.41	0.85	0.14	0.01	0	0	0
A3A	0.35	0.01	0.31	3.41	3.44	3.04	3.31	2.38	0.14	0.01
WT	0	0	0	0	0	0.01	0.01	0.12	0.02	0

-4TCC -3GCT 3ACC 4CCA 7GCC 8CCA 10GCG 13TCT 17GCG 19GCC

图 5

【 図 6 】



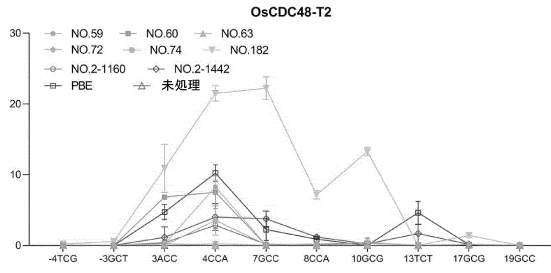
20

30

40

50

【 図 7 】

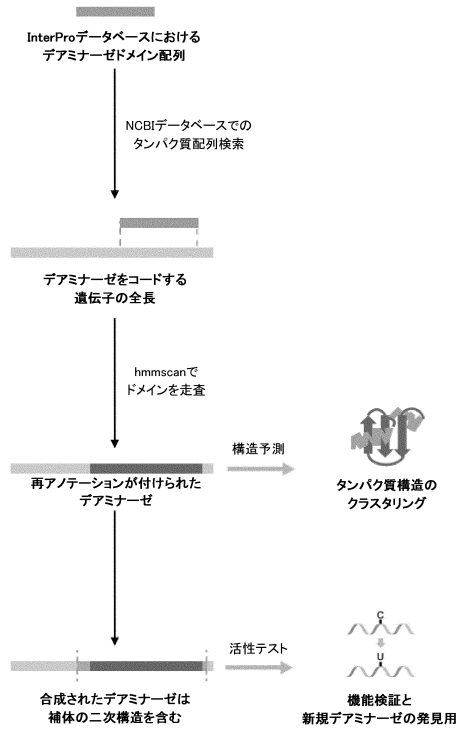


【 図 8 】

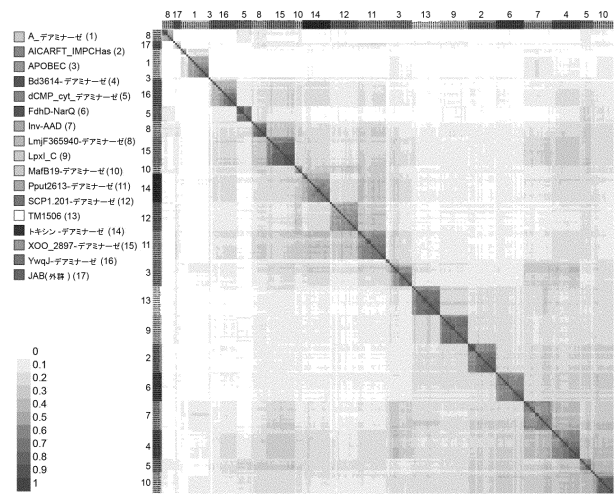


10

【 図 9 】



【 図 10 】



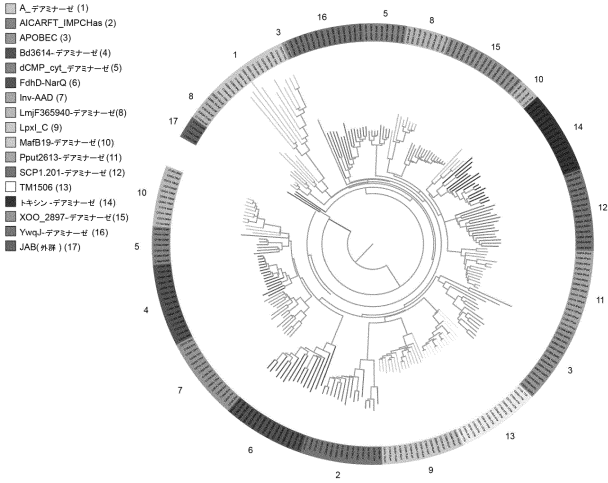
20

30

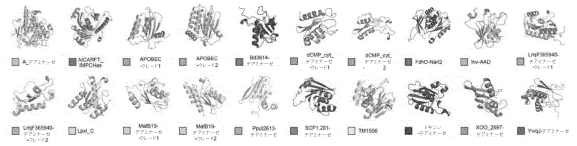
40

50

【 1 1 A 】

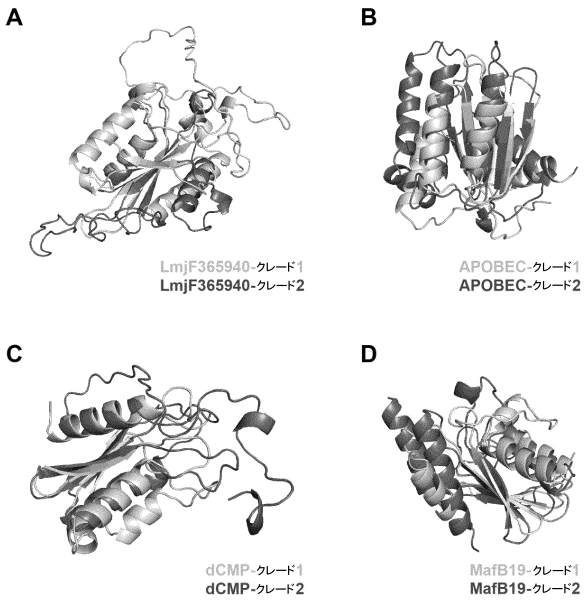


【 1 1 B 】

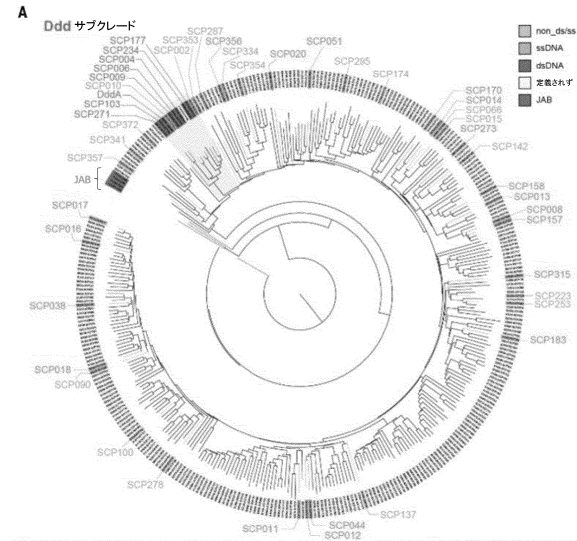


10

【 1 2 】



【 1 3 A 】



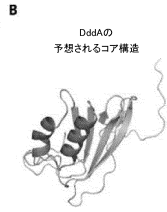
20

30

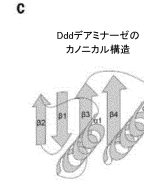
40

50

【 図 1 3 B 】



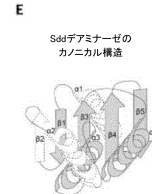
【 図 1 3 C 】



【 図 1 3 D 】

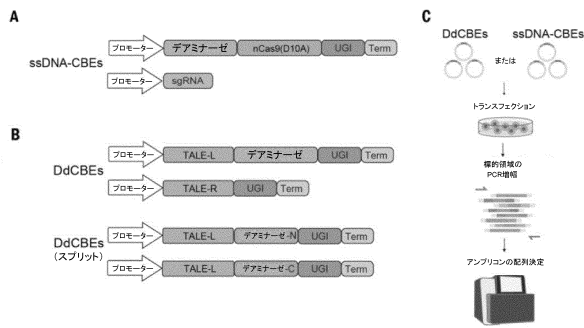


【 図 1 3 E 】

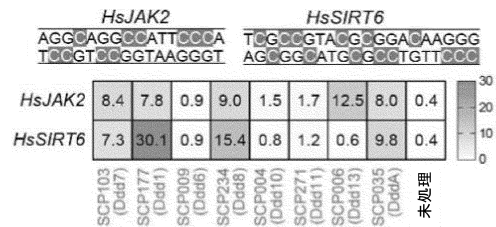


10

【 図 1 4 】

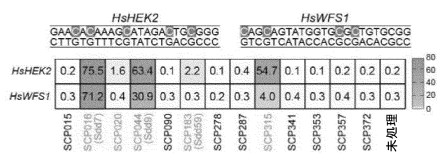


【 図 1 5 】



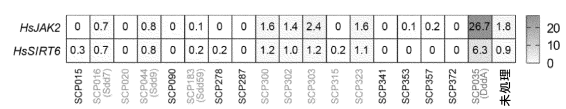
20

【 図 1 6 A 】



30

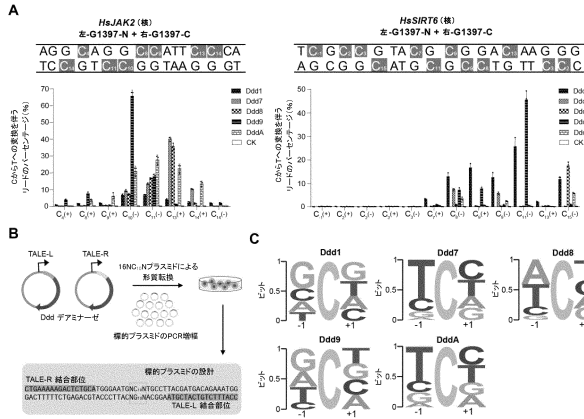
【 図 1 6 B 】



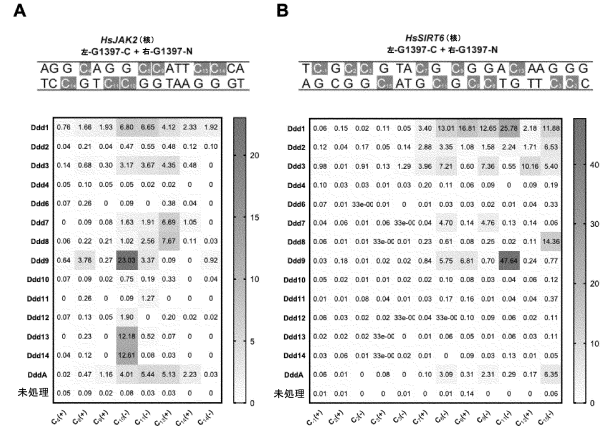
40

50

【 図 17 】



【 図 18 】



10

【 図 19 】

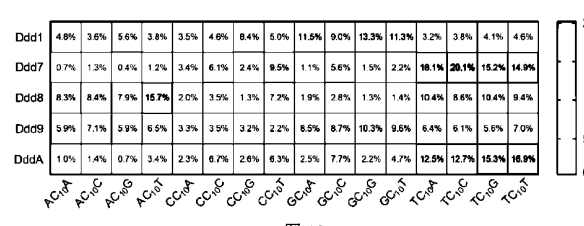
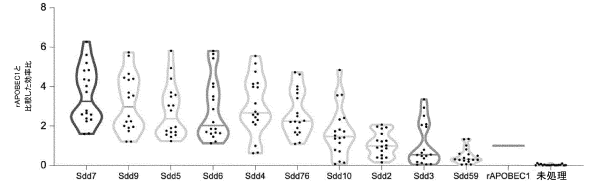


図 19

【 図 20 】



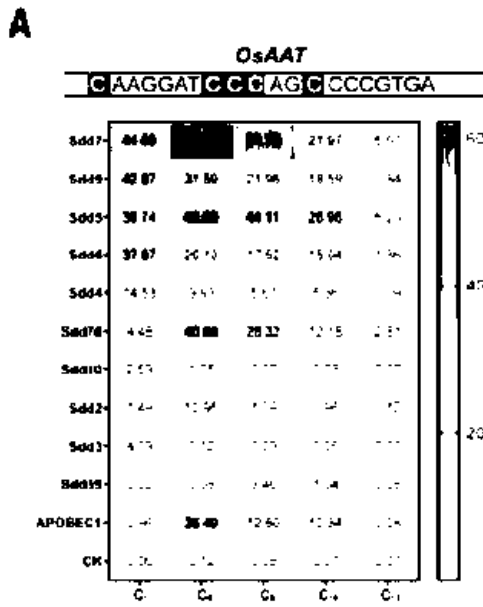
20

30

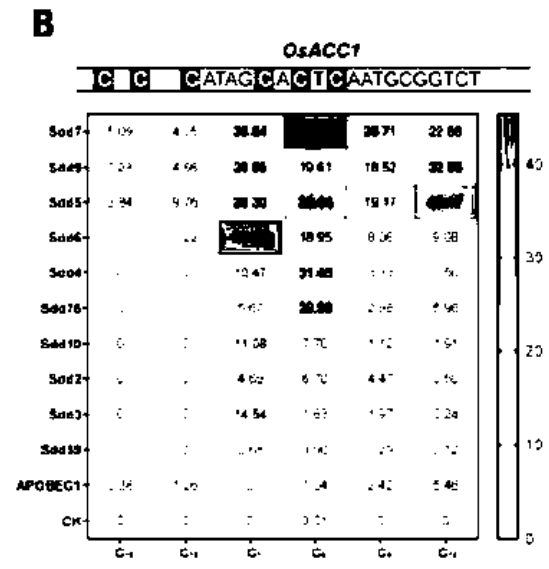
40

50

【 2 1 A 】



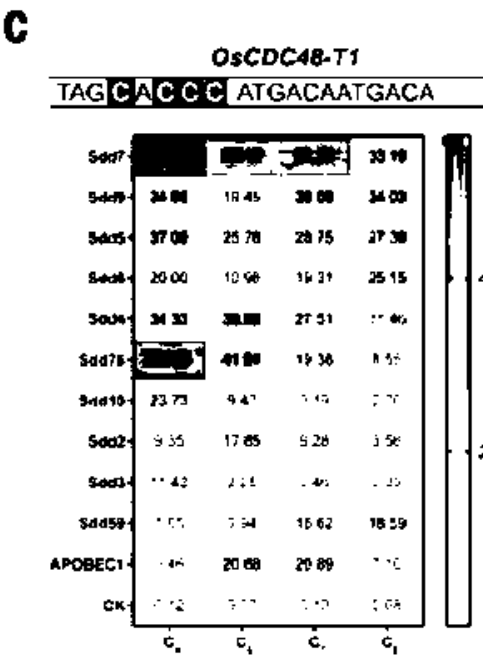
【 2 1 B 】



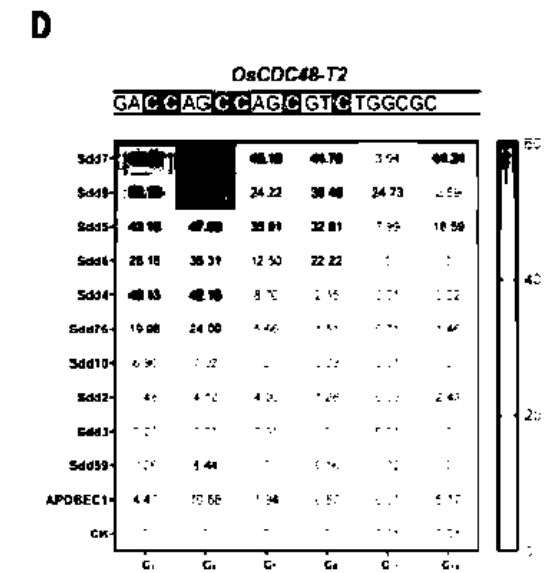
10

20

【 2 1 C 】



【 2 1 D 】



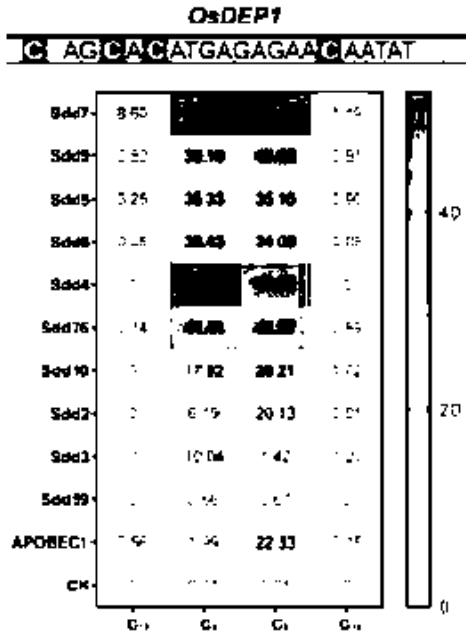
30

40

50

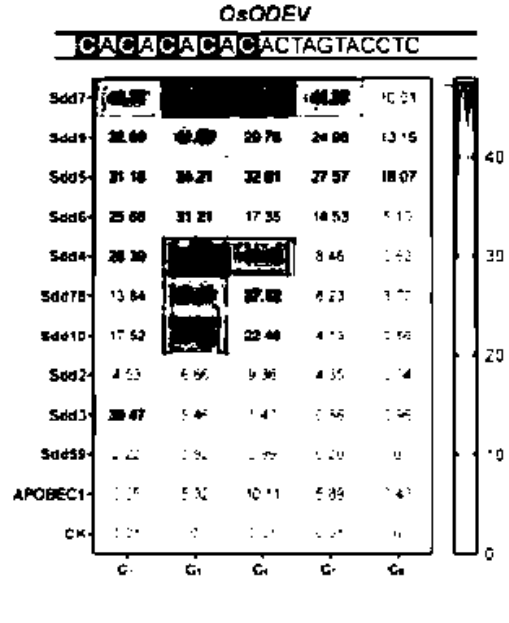
【 2 1 E 】

E



【 2 1 F 】

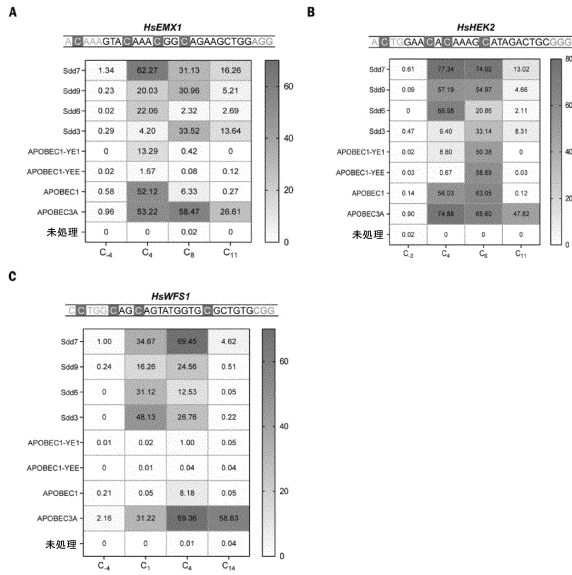
F



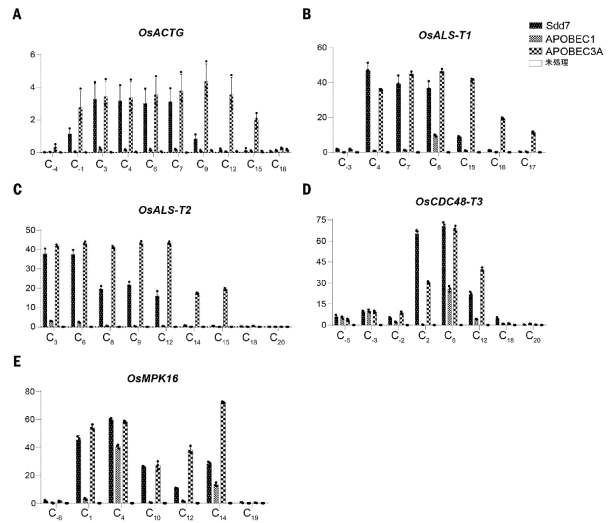
10

20

【 2 2 】



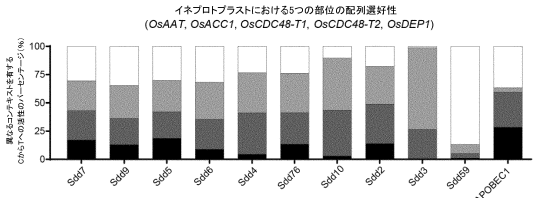
【 2 3 】



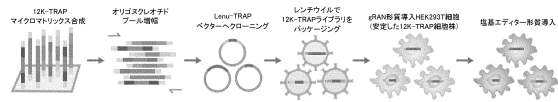
30

40

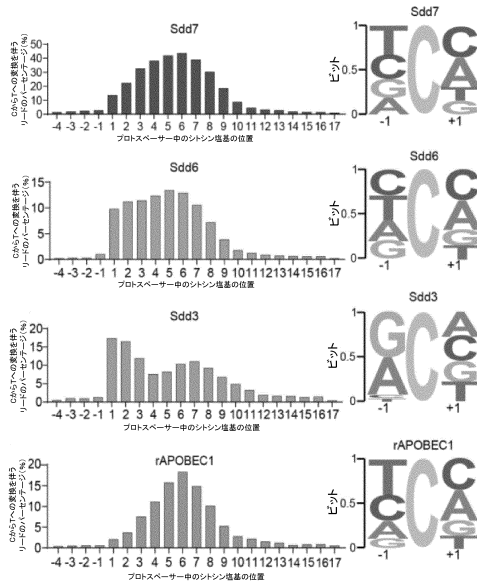
【 2 4 】



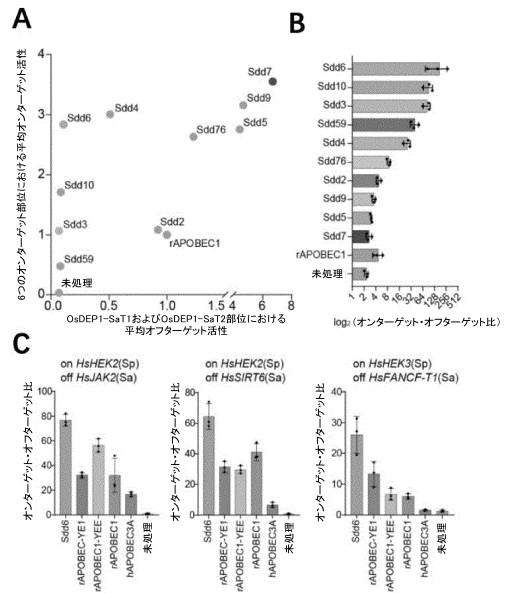
【 2 5 A 】



【 2 5 B 】



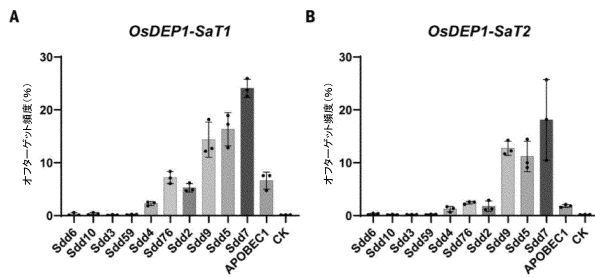
【 2 6 】



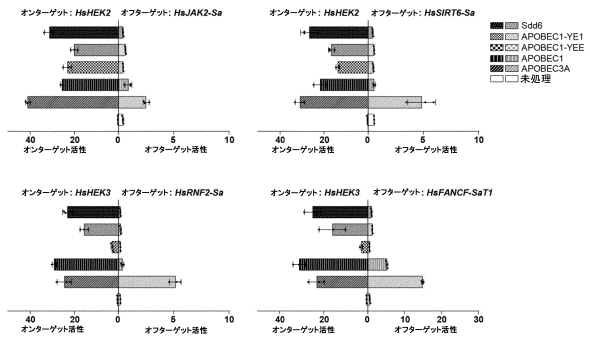
10

20

【 2 7 】



【 2 8 】



30

40

50

【 図 2 9 】

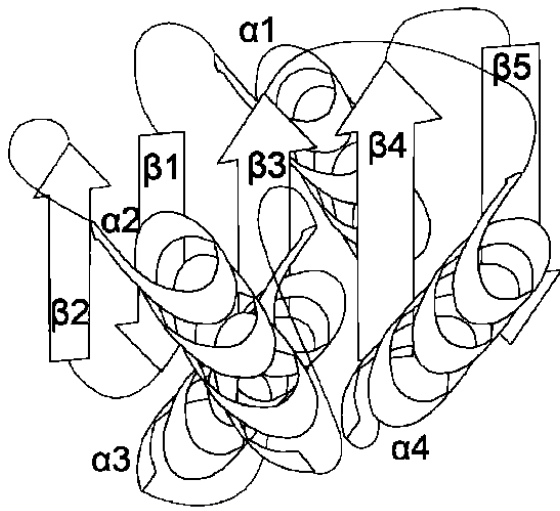
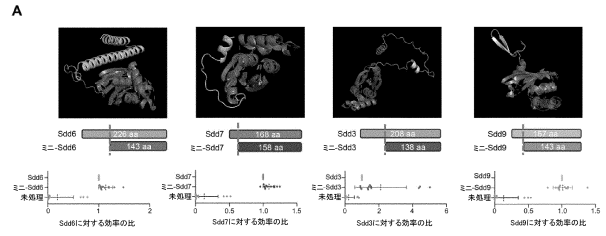


图 29

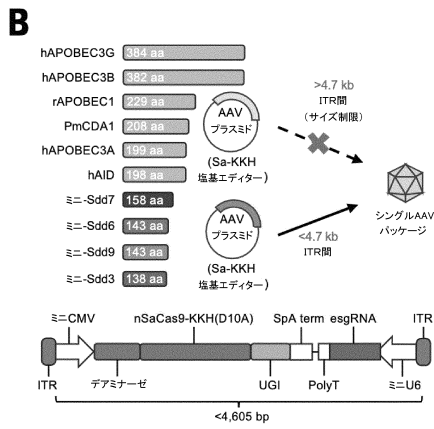
【 図 3 0 A 】



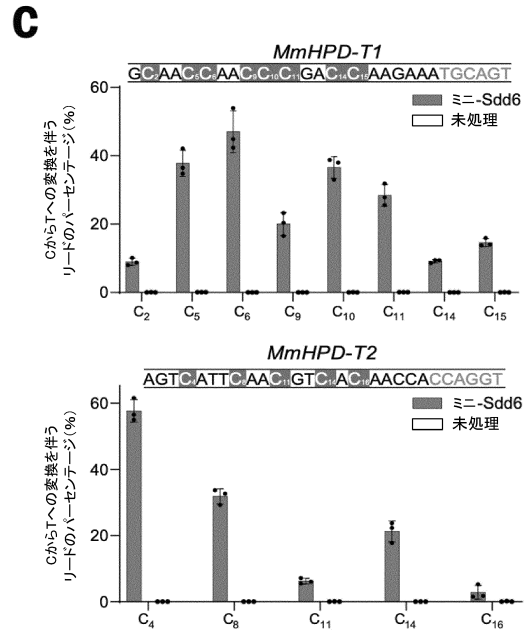
10

20

【 図 3 0 B 】



【 図 3 0 C 】

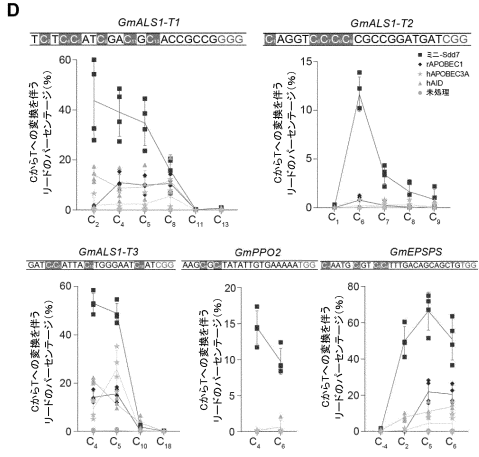


30

40

50

【 図 3 0 D 】



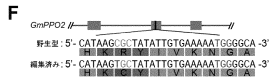
【 図 3 0 E 】

E

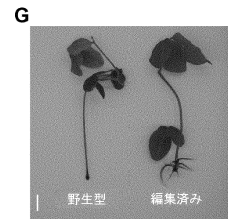
株選し	デアミナーゼ	標的遺伝子	遺伝子組換え植物の数	変異体の数	編集効率
例 1	ミニ-Sdd7 APOBEC3A	GmPPO2	40	7	17.5%
			31	0	0.0%
例 2	ミニ-Sdd7 APOBEC3A	GmPPO2	30	11	36.7%
			45	0	0.0%
例 3	ミニ-Sdd7 APOBEC3A	GmPPO2	7	3	42.9%
			8	0	0.0%

10

【 図 3 0 F 】



【 図 3 0 G 】



20

【 図 3 1 】

A

B

デアミナーゼ	標的遺伝子	変異体の数	遺伝子組換え植物の数	CからTへの頻度 (%)	ヘテロ接合 / ホモ接合
ミニ-Sdd7	OsODEV	222	298	74.50	64/145
		127	279	45.52	28/75
ミニ-Sdd7	OsACC1	161	246	65.45	37/117
		10	52	19.23	2/7
ミニ-Sdd7	OsAAT	185	220	84.09	10/170
		82	102	80.39	9/70

30

【 図 3 2 】



40

【 配列表 】

2025510586000001.xml

50

【 国际调查报告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/CN2023/080052
A. CLASSIFICATION OF SUBJECT MATTER		
C12N9/78(2006.01)i; C07K19/00(2006.01)i; C12N15/113(2010.01)i; C12N7/00(2006.01)i; C12N5/10(2006.01)i; C12N15/79(2006.01)i		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) IPC: C12N C07K		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNABS; CNTXT; CJFD; WPABS; WPABSC; ENTXTC; ENTXT; DWPI; CNKI; 万方, WANFANG; ISI Web of Science; STN; GenBank; EMBL; 中国专利生物序列检索系统, China Patent Biological Sequence Search System; 中国科学院遗传与发育生物学研究所, 高彩霞, 胞嘧啶脱氨酶, CRISPR效应蛋白, 尿嘧啶DNA糖基化酶抑制, 核定位序列, 向导RNA, 腺相关病毒, 接头, 基于序列1, 4, 26, 64, 65的检索, search based on sequence 1, 4, 26, 64, 65, INSTITUTE OF GENETICS AND DEVELOPMENTAL BIOLOGY, CHINESE ACADEMY OF SCIENCES, AID, CAS9, UGL, NLS, sgRNA, AAV, rAAV, linker		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 102482639 A (MEDICAL RESEARCH COUNCIL) 30 May 2012 (2012-05-30) claims 100-109, description paragraphs [0053], [0085]	1, 3-9
Y	CN 102482639 A (MEDICAL RESEARCH COUNCIL) 30 May 2012 (2012-05-30) claims 100-109, description paragraphs [0053], [0085]	29-63, 84-92
Y	CN 109295053 A (SHANGHAI INSTITUTES FOR BIOLOGICAL SCIENCES, CHINESE ACADEMY OF SCIENCES) 01 February 2019 (2019-02-01) claims 10-15, description paragraphs [0077], [0081]-[0082]	29-63, 84-92
<input type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "D" document cited by the applicant in the international application "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 23 May 2023		Date of mailing of the international search report 05 June 2023
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) China No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088		Authorized officer Telephone No.

Form PCT/ISA/210 (second sheet) (July 2022)

10

20

30

40

50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/080052

Box No. I	Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)
1.	With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
a.	<input checked="" type="checkbox"/> forming part of the international application as filed.
b.	<input type="checkbox"/> furnished subsequent to the international filing date for the purposes of international search (Rule 13ter.1(a)), <input type="checkbox"/> accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.
2.	<input type="checkbox"/> With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.
3.	Additional comments:

10

20

30

40

50

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2023/080052

Box No. III	Observations where unity of invention is lacking (Continuation of item 3 of first sheet)
	<p>This International Searching Authority found multiple inventions in this international application, as follows:</p> <p>Invention 1: the solutions relating to the cytosine deaminase shown in SEQ ID NO: 1 in claims 1, 3-9, 29-63 and 84-92</p> <p>Invention 2: the solutions relating to the cytosine deaminase shown in SEQ ID NO: 63 in claims 1, 3-9, 29-63 and 84-92</p> <p>Inventions 3-15: the solutions relating to the cytosine deaminase shown in SEQ ID NOs: 28-40 in claims 1-8, 10-13, 29-63 and 84-92</p> <p>Inventions 16-41: the solutions relating to the cytosine deaminase shown in SEQ ID NOs: 2-18 and 41-49 in claims 1-8, 14-17, 29-63 and 84-92</p> <p>Inventions 42-47: the solutions relating to the cytosine deaminase shown in SEQ ID NOs: 50-55 in claims 1-8, 18-23, 29-63 and 84-92</p> <p>Inventions 48-51: the solutions relating to the cytosine deaminase shown in SEQ ID NOs: 19, 56, 57 and 58 in claims 1-8, 24, 29-63 and 84-92</p> <p>Inventions 52-53: the solutions relating to the cytosine deaminase shown in SEQ ID NOs: 20 and 21 in claims 1-8, 25, 29-63 and 84-92</p> <p>Invention 54: the solutions relating to the cytosine deaminase shown in SEQ ID NO: 22 in claims 1-8, 26, 29-63 and 84-92</p> <p>Inventions 55-60: the solutions relating to the cytosine deaminase shown in SEQ ID NOs: 23, 24 and 59-62 in claims 1-8, 27, 29-63 and 84-92</p> <p>Inventions 61-62: the solutions relating to the cytosine deaminase shown in SEQ ID NO: 74 or 75 in claims 1-8, 28-63 and 84-92</p> <p>Invention 63: the solutions relating to a protein clustering method in claims 64-71</p> <p>Invention 64: the solutions relating to a protein function prediction method based on a three-dimensional structure in claims 72-76</p> <p>Invention 65: the solutions relating to a method for identifying a minimal functional domain of a protein on the basis of a three-dimensional structure in claims 77-80</p> <p>Invention 66: the solution relating to a cytosine deaminase identified using the protein function prediction method of either of claims 75-76 in claim 81</p> <p>Invention 67: the solutions relating to a truncated cytosine deaminase in claims 82-83</p> <p>The present International Authority is of the opinion that the claims comprise 67 inventions. Inventions 1-62 and inventions 63-67 do not have a same or corresponding special technical feature that defines a contribution which the inventions make over the prior art. The same or corresponding technical feature of the 67 inventions is cytosine deaminase. However, the prior art (e.g., CN 102482639 A) discloses an activation-induced cytidine deaminase (AID) mutant. Therefore, the 67 inventions do not have a same or corresponding special technical feature that defines a contribution which the inventions make over the prior art, do not have any technical relationship therebetween, do not fall within a single general inventive concept, and thus do not meet the requirement of unity of invention and do not comply with PCT Rule 13.1.</p>

10

20

30

40

50

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2023/080052

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

- 1. As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
- 2. As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
- 3. As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
- 4. No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: **The solutions relating to the cytosine deaminase shown in SEQ ID NO: 1 in claims 1, 3-9, 29-63 and 84-92**

10

- Remark on Protest**
- The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
 - The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
 - No protest accompanied the payment of additional search fees.

20

30

40

50

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2023/080052

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	102482639	A	30 May 2012	JP	2012522501	A	27 September 2012
				JP	5683566	B2	11 March 2015
				US	2012151613	A1	14 June 2012
				US	9683226	B2	20 June 2017
				EP	2414507	A1	08 February 2012
				EP	2414507	B1	02 July 2014
				WO	2010113039	A1	07 October 2010
				CA	2757178	A1	07 October 2010
				CA	2757178	C	19 May 2020
				RU	2011144569	A	10 May 2013
				RU	2537264	C2	27 December 2014
				AU	2010230985	A1	06 October 2011
				AU	2010230985	B2	24 September 2015
				CN	102482639	B	06 January 2016
CN	109295053	A	01 February 2019	EP	3712272	A1	23 September 2020
				EP	3712272	A4	13 October 2021
				US	2021355508	A1	18 November 2021
				WO	2019020007	A1	31 January 2019
				CA	3106738	A1	31 January 2019

10

20

30

40

50

国际检索报告		国际申请号 PCT/CN2023/080052
A. 主题的分类 C12N9/78(2006.01)i; C07K19/00(2006.01)i; C12N15/113(2010.01)i; C12N7/00(2006.01)i; C12N5/10(2006.01)i; C12N15/79(2006.01)i 按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类		
B. 检索领域 检索的最低限度文献(标明分类系统和分类号) IPC; C12N C07K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用)) CNABS;CNTXT;CJFD;WPABS;WPABSC;ENTXTC;ENTXT;DWPI;CNKI;万方;ISI Web of Science;STN;GenBank;EMBL;中国专利生物序列检索系统;中国科学院遗传与发育生物学研究所, 高彩霞, 胞嘧啶脱氨酶, CRISPR效应蛋白, 尿嘧啶DNA糖基化酶抑制, 核定位序列, 向导RNA, 腺相关病毒, 接头, 基于序列1,4,26,64,65的检索, INSTITUTE OF GENETICS AND DEVELOPMENTAL BIOLOGY, CHINESE ACADEMY OF SCIENCES, AID, CAS9, UGI, NLS, sgRNA, AAV, rAAV, linker		
C. 相关文件		
类型*	引用文件, 必要时, 指明相关段落	相关的权利要求
X	CN 102482639 A (医学研究会) 2012年5月30日 (2012-05-30) 权利要求100-109, 说明书第[0053]、[0085]段	1, 3-9
Y	CN 102482639 A (医学研究会) 2012年5月30日 (2012-05-30) 权利要求100-109, 说明书第[0053]、[0085]段	29-63, 84-92
Y	CN 109295053 A (中国科学院上海生命科学研究院) 2019年2月1日 (2019-02-01) 权利要求10-15, 说明书第[0077]、[0081]-[0082]段	29-63, 84-92
<input type="checkbox"/> 其余文件在C栏的续页中列出。 <input checked="" type="checkbox"/> 见同族专利附件。		
* 引用文件的具体类型: "A" 认为不特别相关的表示了现有技术一般状态的文件 "D" 申请人在国际申请中引证的文件 "E" 在国际申请日的当天或之后公布的在先申请或专利 "L" 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的) "O" 涉及口头公开、使用、展览或其他方式公开的文件 "P" 公布日先于国际申请日但迟于所要求的优先权日的文件 "T" 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 "X" 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 "Y" 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 "&" 同族专利的文件		
国际检索实际完成的日期 2023年5月23日		国际检索报告邮寄日期 2023年6月5日
ISA/CN的名称和邮寄地址 中国国家知识产权局 中国北京市海淀区蓟门桥西土城路6号 100088		受权官员 张智贤 电话号码 (+86) 0512-88996583

PCT/ISA/210 表(第2页) (2022年7月)

10

20

30

40

50

国际检索报告

国际申请号

PCT/CN2023/080052

第1栏 核苷酸和/或氨基酸序列(续第1页第1.c项)

1. 关于国际申请中所公开的任何核苷酸和/或氨基酸序列, 国际检索是基于下列序列表进行的:

- a. 作为国际申请的一部分提交的;
- b. 为国际检索的目的在国际申请日之后提交(细则13之三.1(a)),
 附有说明序列表不超出所提交国际申请公开范围的声明。

10

2. 本报告是在没有收到符合WIPO ST. 26标准的序列表的情况下, 考虑了国际申请中披露的任何核苷酸和/或氨基酸序列, 在可进行有意义检索的范围内做出的。

3. 补充意见:

20

30

40

50

国际检索报告

国际申请号

PCT/CN2023/080052

第III栏 缺乏发明单一性的意见(续第1页第3项)

本国际检索单位在该国际申请中发现多项发明,即:

- 发明1: 权利要求1, 3-9, 29-63, 84-92中涉及SEQ ID NO:1所示胞嘧啶脱氨酶的方案
- 发明2: 权利要求1, 3-9, 29-63, 84-92中涉及SEQ ID NO:63所示胞嘧啶脱氨酶的方案
- 发明3-15: 权利要求1-8, 10-13, 29-63, 84-92中涉及28-40所示胞嘧啶脱氨酶的方案
- 发明16-41: 权利要求1-8, 14-17, 29-63, 84-92中涉及2-18、41-49所示胞嘧啶脱氨酶的方案
- 发明42-47: 权利要求1-8, 18-23, 29-63, 84-92中涉及50-55所示胞嘧啶脱氨酶的方案
- 发明48-51: 权利要求1-8, 24, 29-63, 84-92中涉及19、56、57、58所示胞嘧啶脱氨酶的方案
- 发明52-53: 权利要求1-8, 25, 29-63, 84-92中涉及20、21所示胞嘧啶脱氨酶的方案
- 发明54: 权利要求1-8, 26, 29-63, 84-92中涉及22所示胞嘧啶脱氨酶的方案
- 发明55-60: 权利要求1-8, 27, 29-63, 84-92中涉及23、24、59-62所示胞嘧啶脱氨酶的方案
- 发明61-62: 权利要求1-8, 28-63, 84-92中涉及74 或75所示胞嘧啶脱氨酶的方案
- 发明63: 权利要求64-71中涉及一种蛋白质聚类方法的方案
- 发明64: 权利要求72-76中涉及一种基于三维结构的蛋白质功能预测方法的方案
- 发明65: 权利要求77-80中涉及一种基于三维结构的鉴定蛋白质最小功能结构域的方法的方案
- 发明66: 权利要求81中涉及通过权利要求75-76 中任一项的蛋白质功能预测方法鉴定的胞嘧啶脱氨酶的方案
- 发明67: 权利要求82-83中涉及截短的胞嘧啶脱氨酶的方案

本国际单位认为, 权利要求书中包括67项发明。其中发明1-62与发明63-67不具有相同或相应的体现发明对现有技术做出贡献的特定技术特征。上述67项发明的相同或相应技术特征为均是胞嘧啶脱氨酶。然而, 现有技术(例如CN102482639A)已经公开了胞苷脱氨酶(AID)突变体。因此, 上述67项发明不具有相同或相应的体现发明对现有技术做出贡献的特定技术特征, 不存在技术关联, 不属于一个总的发明构思, 因而不满足发明单一性的要求, 不符合PCT细则13.1的规定。

- 1. 由于申请人按时缴纳了被要求缴纳的全部附加检索费, 本国际检索报告涉及全部可作检索的权利要求。
- 2. 由于无需付出有理由要求附加费的劳动即能对全部可检索的权利要求进行检索, 本单位未通知缴纳任何加费。
- 3. 由于申请人仅按时缴纳了部分被要求缴纳的附加检索费, 本国际检索报告仅涉及已缴费的那些权利要求, 具体地说, 是权利要求:
- 4. 申请人未按时缴纳被要求缴纳的附加检索费。因此, 本国际检索报告仅涉及权利要求书中首先提及的发明; 包含该发明的权利要求是: 权利要求1, 3-9, 29-63, 84-92中涉及SEQ ID NO:1所示胞嘧啶脱氨酶的方案。

- 对异议的意见
- 申请人缴纳了附加检索费, 同时提交了异议书, 适用时, 缴纳了异议费。
 - 申请人缴纳了附加检索费, 同时提交了异议书, 但未在通知书规定的时间期限内缴纳异议费。
 - 缴纳附加检索费时未提交异议书。

10

20

30

40

50

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2023/080052

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	102482639	A	2012年5月30日	JP	2012522501	A	2012年9月27日
				JP	5683566	B2	2015年3月11日
				US	2012151613	A1	2012年6月14日
				US	9683226	B2	2017年6月20日
				EP	2414507	A1	2012年2月8日
				EP	2414507	B1	2014年7月2日
				WO	2010113039	A1	2010年10月7日
				CA	2757178	A1	2010年10月7日
				CA	2757178	C	2020年5月19日
				RU	2011144569	A	2013年5月10日
				RU	2537264	C2	2014年12月27日
				AU	2010230985	A1	2011年10月6日
				AU	2010230985	B2	2015年9月24日
				CN	102482639	B	2016年1月6日
CN	109295053	A	2019年2月1日	EP	3712272	A1	2020年9月23日
				EP	3712272	A4	2021年10月13日
				US	2021355508	A1	2021年11月18日
				WO	2019020007	A1	2019年1月31日
				CA	3106738	A1	2019年1月31日

10

20

30

40

50

フロントページの続き

(51)国際特許分類	F I	テーマコード (参考)
C 1 2 N 1/15 (2006.01)	C 1 2 N 1/15	
C 1 2 N 1/19 (2006.01)	C 1 2 N 1/19	
C 1 2 N 1/21 (2006.01)	C 1 2 N 1/21	
C 1 2 N 15/864 (2006.01)	C 1 2 N 15/864	1 0 0 Z
A 6 1 K 47/64 (2017.01)	A 6 1 K 47/64	
A 6 1 K 38/46 (2006.01)	A 6 1 K 38/46	
G 1 6 B 30/00 (2019.01)	G 1 6 B 30/00	
G 1 6 B 15/00 (2019.01)	G 1 6 B 15/00	

,MC,ME,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,IT,JM,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,MG,MK,MN,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

弁理士 アイゼル・フェリックス = ラインハルト

(74)代理人 100098501

弁理士 森田 拓

(74)代理人 100116403

弁理士 前川 純一

(74)代理人 100134315

弁理士 永島 秀郎

(74)代理人 100162880

弁理士 上島 類

(74)代理人 100210099

弁理士 遠藤 太介

(72)発明者 ツァイシア ガオ

中華人民共和国 ペキン チャオヤン ディストリクト ウェスト ベイチェン ロード ナンバー 1
 インスティテュート・オブ・ジェネティクス・アンド・ディヴェロプメンタル・バイオロジー ビルディング ナンバー 1

(72)発明者 キウベン リン

中華人民共和国 ペキン チャオヤン ディストリクト ウェスト ベイチェン ロード ナンバー 1
 インスティテュート・オブ・ジェネティクス・アンド・ディヴェロプメンタル・バイオロジー ビルディング ナンバー 1

(72)発明者 ジアンイン ファン

中華人民共和国 ペキン チャオヤン ディストリクト ウェスト ベイチェン ロード ナンバー 1
 インスティテュート・オブ・ジェネティクス・アンド・ディヴェロプメンタル・バイオロジー ビルディング ナンバー 1

(72)発明者 ケヴィン ティー・ ツァオ

中華人民共和国 ペキン チャンピン ディストリクト ライフ サイエンス パーク ライフ サイエンス パーク ロード ナンバー 2 7 ビルディング ナンバー 1

F ターム (参考) 4B065 AA01X AA01Y AA57X AA83X AA87X AA87Y AA88X AA88Y AB01 AC20
 BA01 CA60

4C076 AA95 CC41 EE59

4C084 AA03 AA07 DC22 NA13 NA14

4H045 AA10 AA20 AA30 BA10 BA41 CA11 DA89 FA74