



등록특허 10-2718353



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2024년10월15일  
(11) 등록번호 10-2718353  
(24) 등록일자 2024년10월11일

(51) 국제특허분류(Int. Cl.)  
*C12N 15/86* (2006.01) *A61K 39/00* (2006.01)  
*A61K 48/00* (2006.01) *C12N 15/67* (2006.01)

(52) CPC특허분류  
*C12N 15/86* (2013.01)  
*A61K 48/0066* (2013.01)

(21) 출원번호 10-2019-7013847

(22) 출원일자(국제) 2017년10월03일  
심사청구일자 2020년09월18일

(85) 번역문제출일자 2019년05월14일

(65) 공개번호 10-2019-0082226

(43) 공개일자 2019년07월09일

(86) 국제출원번호 PCT/US2017/054928

(87) 국제공개번호 WO 2018/075235  
국제공개일자 2018년04월26일

(30) 우선권주장  
62/409,228 2016년10월17일 미국(US)

(56) 선행기술조사문헌

WO2014170493 A2\*

J. Virol., Vol.83, No.17,  
pp.8327-8339(2009.06.10.)\*

US20140079734 A1

\*는 심사관에 의하여 인용된 문헌

전체 청구항 수 : 총 25 항

심사관 : 전명숙

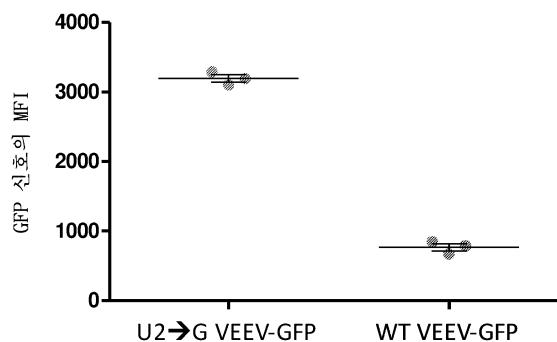
(54) 발명의 명칭 재조합 바이러스 레플리콘 시스템 및 그의 용도

**(57) 요약**

본 발명은 전반적으로 관심 대상 분자의 생성에 적합한 바이러스 기반 발현 시스템에 관한 것이다. 본 발명은 변형된 5'-비번역 영역(5'-UTR)을 포함하는 변형된 레플리콘 RNA를 함유하고, 선택적으로, 구조 단백질을 인코딩하는 그의 본래의 바이러스 서열의 적어도 일부가 결실된, 발현 백터와 같은 핵산 작제물에 관한 것이다. 또한, 관심 대상 폴리펩티드를 생성하는 방법이 개시된다.

**대 표 도** - 도6

평균 형광 강도



(52) CPC특허분류

*C12N 15/67* (2013.01)

*A61K 2039/53* (2013.01)

*C12N 2770/36143* (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

변형된 알파바이러스 레플리콘 RNA를 포함하는 핵산 분자로서, 상기 변형된 레플리콘 RNA는 변형된 5'-UTR을 포함하고 바이러스 구조 단백질을 인코딩하는 핵산 서열의 적어도 일부가 없고, 여기서 상기 변형된 5'-UTR은 위치 2에서 뉴클레오티드 치환이 있는 서열 번호 2 내지 서열 번호 18 중 적어도 하나에 상응하는 RNA 서열을 포함하고, 여기서 상기 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→G 치환인 것인, 핵산 분자.

#### 청구항 2

제1항에 있어서, 상기 변형된 레플리콘 RNA는 알파바이러스 구조 단백질에서 제공된 10개 이상의 연속 아미노산을 인코딩하는 서열이 없는 것인, 핵산 분자.

#### 청구항 3

제1항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함하지 않는, 핵산 분자.

#### 청구항 4

제1항에 있어서, 하나 이상의 발현 카세트를 추가로 포함하며, 상기 발현 카세트 각각은 이종성 핵산 서열에 작동 가능하게 연결된 프로모터를 포함하는, 핵산 분자.

#### 청구항 5

제1항에 있어서, 상기 변형된 레플리콘 RNA는 VEEV/EEEV 군, 또는 SF 군, 또는 SIN 군에 속하는 알파바이러스의 것인, 핵산 분자.

#### 청구항 6

제1항에 있어서, 상기 알파바이러스는 베네수엘라 말 뇌염 바이러스(VEEV)인, 핵산 분자.

#### 청구항 7

제1항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 이종성 조절 요소에 작동 가능하게 연결되는, 핵산 분자.

#### 청구항 8

제1항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 비구조 폴리펩티드 nsp1, nsp2, nsp3, 및 nsp4를 코딩하는 서열을 포함하는, 핵산 분자.

#### 청구항 9

변형된 알파바이러스 레플리콘 RNA를 인코딩하는 서열을 포함하는 핵산 분자로서, 상기 핵산 분자는 서열 번호 1을 포함하고, 상기 변형된 알파바이러스 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없으며, 여기서 상기 변형된 레플리콘 RNA는 변형된 5'-UTR을 포함하고, 여기서 서열 번호 1은 상기 5'-UTR의 위치 2에서의 U→G 뉴클레오티드 치환을 포함하는 레플리콘을 인코딩하는 것인, 핵산 분자.

#### 청구항 10

제9항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 알파바이러스 구조 단백질에서 제공된 10개 이상의 연속 아미노산을 인코딩하는 서열이 없는 것인, 핵산 분자.

#### 청구항 11

제9항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함하지 않는, 핵산 분자.

#### 청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, 상기 레플리콘은 관심 대상 폴리펩티드를 인코딩하는 발현 카세트를 포함하는, 핵산 분자.

#### 청구항 13

제12항의 핵산 분자를 포함하는 재조합 세포로서, 여기서 상기 세포는 인체 내의 세포가 아닌 것인, 재조합 세포.

#### 청구항 14

제13항에 있어서, 상기 재조합 세포는 원핵 세포 또는 진핵 세포인, 재조합 세포.

#### 청구항 15

제13항에 있어서, 상기 재조합 세포는 동물 세포인, 재조합 세포.

#### 청구항 16

관심 대상 폴리펩티드를 생성하는 방법으로서, 제12항의 핵산 분자를 포함하는 숙주 세포를 배양하는 단계를 포함하고, 여기서 상기 숙주세포가 인체 내에서의 세포가 아닌 것인, 방법.

#### 청구항 17

인간이 아닌 대상체에서 관심 대상 폴리펩티드를 생성하는 방법으로서, 제12항의 핵산 분자를 상기 대상체에게 투여하는 단계를 포함하는 것인, 방법.

#### 청구항 18

제17항에 있어서, 상기 대상체는 말, 피그(pig), 영장류, 마우스, 소, 돼지, 양, 토끼, 고양이, 개, 조류, 어류, 염소, 당나귀, 햄스터, 또는 벼팔로인, 방법.

#### 청구항 19

제16항에 있어서, 상기 변형된 레플리콘 RNA가 알파바이러스 구조 단백질에서 제공된 10 이상의 연속 아미노산을 인코딩하는 서열이 없는 것인, 방법.

#### 청구항 20

제16항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함하지 않는 것인, 방법.

#### 청구항 21

제16항에 있어서, 상기 레플리콘 RNA가 비구조 폴리펩티드 nsp1, nsp2, nsp3, 및 nsp4를 코딩하는 서열을 포함하는 것인, 방법.

#### 청구항 22

제16항에 있어서, 하나 이상의 발현 카세트를 추가로 포함하며, 상기 발현 카세트 각각은 이종성 핵산 서열에 작동 가능하게 연결된 프로모터를 포함하는 것인, 방법.

#### 청구항 23

제16항에 있어서, 상기 변형된 레플리콘 RNA는 VEEV/EEEV 군, 또는 SF 군, 또는 SIN 군에 속하는 알파바이러스의 것인, 방법.

청구항 24

제23항에 있어서, 상기 알파바이러스는 베네수엘라 말 뇌염 바이러스(VEEV)인, 방법.

청구항 25

제16항에 있어서, 상기 변형된 알파바이러스 레플리콘 RNA는 이종성 조절 요소에 작동 가능하게 연결되는 것인, 방법.

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

## 발명의 설명

### 기술 분야

관련 출원

[0002] 본 출원은 2016년 10월 17일에 출원된 미국 가출원 제62/409228호에 대해 35 U.S.C. § 119(e) 하에 우선권을 주장하며, 이 가출원은 전체적으로 본 명세서에 참고로 포함된다.

[0003] 서열 목록의 포함

[0004] 첨부된 서열 목록 자료는 이로써 본 출원에 참고로 포함된다. 파일명이 SGI011WO\_SeqListing.txt인 첨부된 서열 목록 텍스트 파일은 2017년 10월 3일에 생성되었고, 크기는 34 KB이다.

기술분야

[0006] 본 발명은 변형된 바이러스 레플리콘을 포함하는 핵산 분자, 및 세포 배양물 내 또는 생체 내에서 적합한 숙주 세포에서의 원하는 생성물의 생성을 위한 그러한 핵산 분자의 용도를 포함하는 분자 생물학 분야에 관한 것이다.

### 배경 기술

[0007] 최근 수 년간, 몇몇 상이한 동물 바이러스 군이 상동 재조합에 의해 또는 그들의 계놈의 직접적인 조작에 의해 유전자 조작되었다. DNA 및 RNA 바이러스 둘 모두에 대한 역유전학 시스템의 이용 가능성은, 예를 들어 백신, 발현 벡터, 항종양제, 유전자 요법 벡터, 및 약물 전달 비히클로서의 재조합 바이러스의 사용에 대한 새로운 관점을 생성하였다.

[0008] 예를 들어, 배양된 재조합 세포에서의 이종성 단백질의 발현을 위해 많은 바이러스 기반 발현 벡터가 활용되었다. 특히, 숙주 세포에서의 유전자 발현을 위한 변형된 바이러스 벡터의 적용은 계속 확장된다. 이와 관련된 최근의 진보는 다중 서브유닛 단백질 복합체의 생성을 위한 기법 및 시스템의 추가 개발, 및 이종성 단백질 생성을 개선하기 위한 단백질 변형 효소의 동시 발현을 포함한다. 바이러스 발현 벡터 기술에 관한 다른 최근의 진전은 유전자 발현, 바이러스 벡터의 제조, 생체내 유전자 요법 적용, 및 백신 전달 벡터의 생성을 제어하기 위한 많은 진보된 계놈 조작 적용을 포함한다.

[0009] 그러나, 재조합 발현 시스템에서 관심 대상 유전자를 발현하기 위한 더욱 효율적인 방법 및 시스템에 대한 필요성이 여전히 존재한다.

## 발명의 내용

[0010] 이 섹션은 본 출원의 전반적인 개요를 제공하며, 그의 전체 범주 또는 모든 특징을 포괄하는 것은 아니다.

[0011] 일 태양에서, 변형된 레플리콘 RNA를 포함하는 핵산 분자가 본 명세서에 개시되며, 변형된 레플리콘 RNA는 변형된 5'-UTR을 포함하고 바이러스 구조 단백질을 인코딩하는 핵산 서열의 적어도 일부가 없다. 본 발명의 이러한 태양 및 다른 태양의 다양한 실시 형태에서, 본 명세서에 개시된 바와 같은 핵산 분자는 다음과 같은 특징들 중 하나 이상을 포함할 수 있다. 일부 실시 형태에서, 변형된 레플리콘 RNA는 변형된 알파바이러스 레플리콘 RNA이다. 일부 실시 형태에서, 변형된 알파바이러스 레플리콘 RNA는 변형된 알파바이러스 게놈을 포함한다. 일부 실시 형태에서, 변형된 5'-UTR은 위치 1, 2, 4, 또는 이들의 조합에서 하나 이상의 뉴클레오티드 치환을 포함한다. 일부 실시 형태에서, 뉴클레오티드 치환 중 적어도 하나는 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환이다. 일부 실시 형태에서, 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→G 치환이다.

[0012] 일부 실시 형태에서, 핵산 분자는 바이러스 구조 단백질을 인코딩하는 핵산 서열의 실질적인 부분이 없는 변형된 레플리콘 RNA를 포함한다. 일부 실시 형태에서, 본 명세서에 개시된 바와 같은 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함하지 않는다.

[0013] 본 발명의 이러한 태양 및 다른 태양의 다양한 실시 형태에서, 핵산 분자는 하나 이상의 발현 카세트를 추가로 포함하며, 발현 카세트 각각은 이종성 핵산 서열에 작동 가능하게 연결된 프로모터를 포함한다. 일부 실시 형태에서, 핵산 분자는 적어도 2개, 3개, 4개, 5개, 또는 6개의 발현 카세트를 포함한다. 일부 실시 형태에서, 발현 카세트 중 적어도 하나의 프로모터는 26S 서브게놈(subgenomic) 프로모터이거나 이를 포함한다.

[0014] 일부 실시 형태에서, 본 명세서에 개시된 바와 같은 발현 카세트 중 적어도 하나의 이종성 핵산 서열은 관심 대상 유전자(GOI)의 코딩 서열을 포함한다. 일부 실시 형태에서, GOI는 치료용 폴리펩티드, 예방용 폴리펩티드, 진단용 폴리펩티드, 뉴트라슈티컬(neutraceutical) 폴리펩티드, 산업용 효소, 및 리포터 폴리펩티드로 이루어진 군으로부터 선택되는 폴리펩티드를 인코딩한다. 일부 실시 형태에서, GOI는 항체, 항원, 면역 조절인자, 및 사이토카인으로 이루어진 군으로부터 선택되는 폴리펩티드를 인코딩한다. 일부 특정 실시 형태에서, GOI의 코딩 서열은 기준 코딩 서열의 발현 수준보다 높은 수준의 발현을 위해 최적화된다.

[0015] 일부 실시 형태에서, 핵산 분자는 토가비리대(*Togaviridae*) 과의 알파바이러스(*Alphavirus*) 속에 속하는 바이러스의 변형된 게놈 또는 레플리콘 RNA를 포함하는 변형된 레플리콘 RNA를 포함한다. 일부 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 VEEV/EEEV 군, 또는 SF 군, 또는 SIN 군에 속하는 알파바이러스의 것이다. 일부 실시 형태에서, 알파바이러스는 동부 말 뇌염 바이러스(EEEV), 베네수엘라 말 뇌염 바이러스(VEEV), 에버글레이즈 바이러스(Everglades, EVEV), 무캄보 바이러스(Mucambo virus, MUCV), 픽수나 바이러스(Pixuna virus, PIXV), 미들버그 바이러스(Middleburg virus, MIDV), 치쿤구니아 바이러스(Chikungunya virus, CHIKV), 오니옹-니옹 바이러스(O'Nyong-Nyong virus, ONNV), 로스 리버 바이러스(Ross River virus, RRV), 바마 포레스트 바이러스(Barmah Forest virus, BF), 게타 바이러스(Getah virus, GET), 사기야마 바이러스(Sagiyama virus, SAGV), 베바루 바이러스(Bebaru virus, BEBV), 마야로 바이러스(Mayaro virus, MAYV), 우나 바이러스(Una virus, UNAV), 신드비스 바이러스(Sindbis virus, SINV), 오라 바이러스(Aura virus, AURAV), 와타로아 바이러스(Whataroa virus, WHAV), 바반키 바이러스(Babanki virus, BABV), 키질라가크 바이러스(Kyzylagach virus, KYZV), 서부 말 뇌염 바이러스(WEEV), 하이랜드 제이 바이러스(Highland J virus, HJV), 포트 모건 바이러스(Fort Morgan virus, FMV), 두무(Ndumu, NDUV), 및 부기 크릭 바이러스(Buggy Creek virus)로 이루어진 군으로부터 선택된다. 일부 실시 형태에서, 알파바이러스는 베네수엘라 말 뇌염 바이러스(VEEV)이다.

[0016] 일부 실시 형태는 이종성 조절 요소에 작동 가능하게 연결된 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하는 핵산 분자를 제공한다. 일부 실시 형태에서, 이종성 조절 요소는 프로모터 서열을 포함한다. 일부 실시 형태에서, 프로모터 서열은 T7 프로모터 서열을 포함한다. 일부 실시 형태에서, 이종성 조절 요소는 전사 종결 서열을 포함한다. 일부 실시 형태에서, 전사 종결 서열은 T7 종결 서열이거나 이를 포함한다.

[0017] 일부 실시 형태에서, 본 명세서에 개시된 바와 같은 핵산 분자는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하며, 핵산 분자는 서열 번호 1의 핵산 서열에 대해 적어도 80%의 서열 동일성을 나타내고, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 5'-비번역 영역(5'-UTR)의 위치 2에서 U→G 치환을 포함하고 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없다. 일부 실시 형태에서, 핵산 분자는 서열 번호 1의 핵산 서열에 대해 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%의 서열 동일성을 나타낸다.

[0018] 일부 실시 형태에서, 본 명세서에 개시된 바와 같은 핵산 분자는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하며, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 2 내지 서열 번호 18 중 적어도 하나의 핵산 서열에 대해 적어도 80%의 서열 동일성을 나타내는 5'-UTR 및 5'-UTR의 위치 2에서의 U→G 치환을 포함하고, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 2 내지 서열 번호 18 중 적어도 하나의 핵산 서열에 대해 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%의 서열 동일성을 나타내는 5'-UTR을 포함한다. 특정 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열의 실질적인 부분이 없다. 특정 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함하지 않는다.

[0019] 일 태양에서, 본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 핵산 분자를 포함하는 재조합 세포에 관한 것이다. 일부 실시 형태에서, 재조합 세포는 원핵 세포 또는 진핵 세포이다. 일부 실시 형태에서, 재조합 세포는 동물 세포이다. 일부 실시 형태에서, 재조합 세포는 척추 동물 세포 또는 무척추 동물 세포이다. 일부 실시 형태에서, 재조합 세포는 폐 말 동맥 내피 세포, 말 진피 세포, 베이비 햄스터 신장(baby hamster kidney, BHK) 세포, 토끼 신장 세포, 마우스 근육 세포, 마우스 결합 조직 세포, 인간 자궁경부 세포, 인간 표피양 후두 세포(human epidermoid larynx cell), 차이니즈 햄스터 난소 세포(CHO), 인간 HEK-293 세포, 마우스 3T3 세포, Vero 세포, 마딘-다비 개 신장 상피 세포(Madin-Darby Canine Kidney Epithelial Cell, MDCK), 1차 닭 섬유모세포, HuT78 세포, A549 폐 세포, HeLa 세포, PER.C6® 세포, WI-38 세포, MRC-5 세포, FRhL-2, 및 CEM T-세포로 이루어진 군으로부터 선택된다. 본 명세서에 개시된 일부 실시 형태는 본 명세서에 개시된 바와 같은 적어도 하나의 재조합 세포를 포함하는 세포 배양물을 제공한다.

[0020] 일 태양에서, 일부 실시 형태는 변형된 5'-UTR을 포함하고 바이러스 구조 단백질을 인코딩하는 핵산 서열의 적어도 일부가 없는 핵산 분자를 포함하는 숙주 세포를 배양하는 단계를 포함하는, 관심 대상 폴리펩티드를 생성하는 방법을 제공한다. 일부 실시 형태에서, 재조합 세포는 척추 동물 세포 또는 무척추 동물 세포이다.

[0021] 추가의 태양에서, 일부 실시 형태는 변형된 5'-UTR을 포함하고 바이러스 구조 단백질을 인코딩하는 핵산 서열의 적어도 일부가 없는 핵산 분자를 대상체에게 투여하는 단계를 포함하는, 대상체에서 관심 대상 폴리펩티드를 생성하는 방법을 제공한다. 일부 실시 형태에서, 대상체는 인간, 말, 피그(pig), 영장류, 마우스, 소, 돼지, 양, 토끼, 고양이, 개, 조류, 어류, 염소, 당나귀, 햄스터, 또는 베팔로이다.

[0022] 본 발명에 따른 방법의 실시 형태의 구현에는 다음과 같은 특징들 중 하나 이상을 포함할 수 있다. 일부 실시 형태에서, 변형된 레플리콘 RNA는 변형된 알파바이러스 레플리콘 RNA이다. 일부 실시 형태에서, 변형된 알파바이러스 레플리콘 RNA는 변형된 알파바이러스 게놈을 포함한다. 일부 실시 형태에서, 변형된 5'-UTR은 위치 1, 2, 4, 또는 이들의 조합에서 하나 이상의 뉴클레오티드 치환을 포함한다. 일부 실시 형태에서, 뉴클레오티드 치환 중 적어도 하나는 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환이다. 일부 실시 형태에서, 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→G 치환이다. 특정 실시 형태에 있어서, 변형된 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열의 실질적인 부분이 없다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함하지 않는다.

[0023] 일부 실시 형태에서, 본 명세서에 기재된 바와 같은 핵산 분자는 하나 이상의 발현 카세트를 추가로 포함하며, 발현 카세트 각각은 이종성 핵산 서열에 작동 가능하게 연결된 프로모터를 포함한다. 일부 실시 형태에서, 핵산 분자는 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 또는 적어도 6개의 발현 카세트를 포함한다. 일부 실시 형태에서, 발현 카세트 중 적어도 하나의 프로모터는 26S 서브게놈 프로모터를 포함한다. 일부 특정 실시 형태에서, 발현 카세트 중 적어도 하나의 프로모터는 알파바이러스 26S 서브게놈 프로모터를 포함한다. 바람직하게는, 프로모터는 베네수엘라 말 뇌염(VEEV) 26S 서브게놈 프로모터를 포함한다. 특정 실시 형태에서, 발현 카세트 중 적어도 하나의 이종성 핵산 서열은 관심 대상 유전자(GOI)의 코딩 서열을 포함한다. 일부 실시 형태에서, GOI의 코딩 서열은 기준 코딩 서열의 발현 수준보다 높은 수준의 발현을 위해 최적화된다. 일부 실시 형태에서, 이종성 핵산 서열에 작동 가능하게 연결된 프로모터는 이종성 프로모터 서열을 포함한다. 적합한 이종성 프로모터는 뇌심근염 바이러스(EMCV), 소 바이러스성 설사 바이러스(BVDV), 폴리오바이러스, 구제역 바이러스(FMD), 엔테로바이러스 71, 또는 C 형 간염 바이러스로부터 유래된 내부 리보솜 진입 부위(IRES)로부터의 조절 요소를 포함하지만 이로 한정되지 않는다.

[0024] 일부 실시 형태에서, 변형된 레플리콘 RNA는 토가비리대 과의 알파바이러스 속하는 바이러스의 변형된 계

놈 또는 레플리콘 RNA를 포함한다. 일부 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 VEEV/EEEV 군, 또는 SF 군, 또는 SIN 군에 속하는 알파바이러스의 것이다. 일부 실시 형태에서, 알파바이러스는 동부 말 뇌염 바이러스(EEEV), 베네수엘라 말 뇌염 바이러스(VEEV), 에버글레이즈 바이러스(EVEV), 무캄보 바이러스(MUCV), 핵수나 바이러스(PIXV), 미들버그 바이러스(MIDV), 치쿤구니아 바이러스(CHIKV), 오니옹-니옹 바이러스(ONNV), 로스리버 바이러스(RRV), 바마 포레스트 바이러스(BF), 게타 바이러스(GET), 사기야마 바이러스(SAGV), 베바루 바이러스(BEBV), 마야로 바이러스(MAYV), 우나 바이러스(UNAV), 신드비스 바이러스(SINV), 오라 바이러스(AURAV), 와타로아 바이러스(WHAV), 바반키 바이러스(BABV), 키질라가크 바이러스(KYZV), 서부 말 뇌염 바이러스(WEEV), 하이랜드 제이 바이러스(HJV), 포트 모건 바이러스(FMV), 두무(NDUV), 및 부기 크릭 바이러스로 이루어진 군으로부터 선택된다. 일부 실시 형태에서, 알파바이러스는 베네수엘라 말 뇌염 바이러스(VEEV)이다.

[0025] 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 이종성 조절 요소에 작동 가능하게 연결된다. 일부 실시 형태에서, 이종성 조절 요소는 프로모터 서열을 포함한다. 일부 실시 형태에서, 프로모터 서열은 T7 프로모터 서열을 포함한다. 일부 실시 형태에서, 이종성 조절 요소는 전사 종결 서열을 포함한다. 일부 실시 형태에서, 전사 종결 서열은 T7 종결 서열이거나 이를 포함한다.

[0026] 일부 실시 형태에서, 핵산 분자는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하며, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 1의 핵산 서열에 대해 적어도 80%의 서열 동일성을 나타내는 5'-UTR 및 5'-UTR의 위치 2에서의 U→G 치환을 포함하고, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없다. 일부 실시 형태에서, 핵산 분자는 서열 번호 1의 핵산 서열에 대해 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%의 서열 동일성을 나타낸다. 일부 실시 형태에서, 핵산 분자는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하며, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 2 내지 서열 번호 18 중 적어도 하나의 핵산 서열에 대해 적어도 80%의 서열 동일성을 나타내는 5'-UTR 및 5'-UTR의 위치 2에서의 U→G 치환을 포함하고, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 2 내지 서열 번호 18 중 적어도 하나의 핵산 서열에 대해 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%의 서열 동일성을 나타내는 5'-UTR을 포함한다.

[0027] 추가의 태양에서, 본 명세서에 개시된 일부 실시 형태는 본 명세서에 개시된 하나 이상의 실시 형태에 따른 방법에 의해 생성된 재조합 폴리펩티드를 제공한다.

[0028] 일 태양에서, 본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 바와 같은 재조합 폴리펩티드를 포함하는 조성물에 관한 것이다. 일부 실시 형태에서, 조성물은 약제학적으로 허용 가능한 담체를 포함하는 약제학적 조성물이다. 일부 실시 형태에서, 조성물은 예방학적 조성물, 뉴트라슈티컬 조성물, 약제학적 조성물, 또는 이들의 조합이다.

[0029] 추가의 태양에서, 본 명세서에 개시된 일부 실시 형태는 본 명세서에 개시된 바와 같은 핵산 분자를 포함하는 조성물에 관한 것이다. 일부 실시 형태에서, 조성물은 약제학적으로 허용 가능한 담체를 포함하는 약제학적 조성물이다. 일부 실시 형태에서, 조성물은 예방학적 조성물, 뉴트라슈티컬 조성물, 약제학적 조성물, 또는 이들의 조합이다.

[0030] 또 다른 태양에서, 본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 바와 같은 재조합 세포를 포함하는 조성물에 관한 것이다. 일부 실시 형태에서, 조성물은 약제학적으로 허용 가능한 담체를 포함하는 약제학적 조성물이다. 일부 실시 형태에서, 조성물은 예방학적 조성물, 뉴트라슈티컬 조성물, 약제학적 조성물, 또는 이들의 조합이다.

[0031] 전술한 개요는 단지 예시적이며, 어떠한 방식으로든 제한하고자 하는 것이 아니다. 본 명세서에 기술된 예시적인 실시 형태 및 특징에 더하여, 본 출원의 추가의 태양, 실시 형태, 목적 및 특징은 도면, 및 상세한 설명 및 청구범위로부터 충분히 명백해질 것이다.

### 도면의 간단한 설명

[0032] 도 1a는 다음과 같은 대표적인 알파바이러스들로부터의 5' 비번역 영역(5'-UTR)들의 서열 정렬을 도시한다: 오라 바이러스(AURAV; 서열 번호 2), 치쿤구니아 바이러스(CHIKV, 서열 번호 3), 오니옹-니옹 바이러스(서열 번호 4, ONNV), 베바루 바이러스(서열 번호 5, BEBV), 셈리키 포레스트 바이러스(Semliki Forest virus, 서열 번호 6, SFV), 마야로 바이러스(서열 번호 7, MAYV), 게타 바이러스(서열 번호 8, GETV), 사기야마 바이러스(서열 번

호 9, SAGV), 두무 바이러스(서열 번호 10, NDUV), 미들버그 바이러스(서열 번호 11, MIDV), 동부 말 뇌염 바이러스(서열 번호 12, EEEV), 포트 모건 바이러스(서열 번호 13, FMV), 부기 크릭 바이러스(서열 번호 14, Buggy), 베네주엘라 말 뇌염 바이러스(서열 번호 15, VEEV), 와타로아 바이러스(서열 번호 16, WHAV), 신드비스 바이러스(서열 번호 17, SINV), 및 베바루 바이러스(서열 번호 18, BEBV). **도 1a**의 서열 정렬은 디폴트 설정으로 프로그램 MUSCLE 3.6을 사용하여 생성되었다. 본 명세서에 나타낸 서열 정렬에서, 최적 정렬을 위해 프로그램 MUSCLE 3.6에 의해 생성되는 정렬된 서열 중의 대시(–)는 캡, 즉, 그 위치에서의 뉴클레오티드의 결여를 나타낸다. 하기에 상세히 논의되는 바와 같이, 몇몇 보존된 뉴클레오티드 잔기가 이러한 서열 비교 분석에서 확인되었다. 별표는 정렬된 서열들 중에서 동일한 뉴클레오티드 잔기를 확인한다. **도 1b**는 컨센서스(consensus) 서열을 각종 컨센서스로서 그래프로 표현한 것이며, 주어진 아미노산을 지정하는 문자의 크기는 모티프를 생성하는 데 사용된 상이한 서열들 중의 잔기의 보존에 비례한다(문자의 크기는 정렬된 서열들 사이에서 그 위치에서의 잔기의 상대 빈도를 나타냄). 글자의 크기는 비트(bit)로 측정된 정보량을 반영한다.

**도 2**는 비제한적이고 예시적인 기본 1가 VEEV 레플리콘 설계인 Rep-Alpha의 구조의 개략도를 도시하며, Rep-Alpha는 T7 프로모터 서열, 본 명세서에 기재된 바와 같은 U2→G 치환을 갖는 VEEV로부터의 5' UTR 서열, 알파 바이러스 계놈의 비구조 폴리펩티드인 nsp1, nsp2, nsp3, 및 nsp4의 코딩 서열을 포함한다. 기본 1가 VEEV 레플리콘인 Rep-Alpha는 또한 26S 서브계놈 프로모터 서열, 3'UTR 서열, T7 종결 서열, 폴리아데닐화 서열인 PolyA, 및 레플리콘 내로의 추가 성분의 삽입을 용이하게 하도록 조작된 다수의 독특한 제한 부위를 함유한다.

**도 3 및 도 4**는 2개의 비제한적이고 예시적인 1가 VEEV 레플리콘 설계의 구조를 그래프로 도시한 것이며, 벡터 내로 작동 가능하게 포함된 관심 대상 유전자(GOI)는 각각 A Vietnam 1203 HA 유전자(**도 3**) 및 향상된 녹색 형광 단백질(eGFP) 리포터 유전자(**도 4**)였다.

**도 5**는 변형된 VEEV-HA 레플리콘에서 5' UTR의 위치 2에서의 U2→G 변형이 변형된 레플리콘의 생물학적 활성에 영향을 주지 않음을 나타내는 예시적인 실험의 결과를 그래프로 요약한 것이다. 인플루엔자 헤마글루터닌-HA 유전자를 발현하는 변형된 U2→G VEEV-HA 레플리콘으로 전기천공된 세포 상에서 유세포측정 분석(FACS)을 수행하였다(레플리콘의 구조 체계에 대해서는 **도 3**을 또한 참조함). 야생형 VEEV-HA 레플리콘, 즉, 위치 2에 U 잔기를 함유하는 VEEV-HA 레플리콘을 대조군으로서 사용하였다.

**도 6**은 5' UTR에서 U2→G 변형을 지닌 알파바이러스 레플리콘으로부터의 발현이 임의의 특정한 관심 대상 유전자(GOI)로 제한되지 않음을 입증하기 위해 수행된 예시적인 유세포측정 분석의 결과를 그래프로 요약한 것이다. 이 실험에서, 변형된 알파바이러스 레플리콘이 녹색 형광 단백질(GFP) 리포터 유전자를 발현시키도록 조작되었다(레플리콘의 구조 체계에 대해서는 **도 4**를 또한 참조함). 변형된 VEEV-GFP 레플리콘의 5' UTR에서의 U2→G 변형은 야생형 대조 레플리콘으로부터 검출된 발현에 비해 GFP 유전자의 발현을 3배 향상시키는 것으로 나타났다.

**도 7**은 변형된 VEEV 레플리콘으로부터의 레드 파이어플라이(red Firefly) 리포터 유전자의 발현을 평가하는 또 다른 예시적인 유세포측정 분석의 결과를 그래프로 요약한 것이다. 이 실험에서, 변형된 VEEV-rFFF 레플리콘의 5' UTR에서의 U2→G 변형은 야생형 대조 레플리콘으로부터 검출된 발현에 비해 레드 파이어플라이 유전자의 발현을 2배 향상시키는 것으로 나타났다.

**도 8**은 비제한적이고 예시적인 알파바이러스 계놈 구조 및 계놈 발현을 개략적으로 도시한다(문헌[Strauss *et al.*, Microbiological Reviews, pp. 491–562, September 1994]으로부터 조정이 이루어진 것임). SIN 바이러스의 계놈 체계가 나타나 있다. 비구조 유전자 및 구조 단백질 유전자의 명칭이 주어져 있다. 유전자 및 단백질의 명명법에 대한 언급은 상기 문헌[Strauss *et al.*, 1994]에서 찾아볼 수 있다. 49S 계놈 RNA가 중심에 개략적으로 도시되어 있으며, 이때 그의 번역 ORF는 비어있는 박스로 나타난다. 작은 검은색 박스들은 보존된 서열 요소들이고; 비어있는 마름모는 누설(leaky) 오팔(opal) 종결 코돈을 표시한다. 비구조 폴리프로테인(polyprotein) 및 그의 처리 생성물은 위에 나타나 있다. 오팔 코돈에서의 종결은 P123을 생성하며, 복제에서 그의 주요 기능은 활성 RNA 레플리카제에서 폴리프로테인을 처리하기 위해 트랜스(*trans*)로 작용하는 프로테인 나아제로서 존재하는 것으로 믿어지고; 이러한 프로테이나아제 도메인은 nsP2 영역에서 발견된다. 오팔 정지 코돈의 리드쓰루(readthrough)는 P1234를 생성하며, 이는 활성 레플리카제를 형성할 수 있다. 26S 서브계놈 mRNA가 아래에서 확대되어 구조 ORF 및 그의 번역 생성물을 나타낸다. 비리온에 존재하는 폴리펩티드는 음영 처리되어 있다. vcRNA는 계놈 RNA의 마이너스 가닥 상보체이다.

본 발명의 전술한 및 다른 특징은 첨부 도면과 함께 취해지는 하기의 설명 및 첨부된 청구범위로부터 더욱 충분히 명백해질 것이다. 이들 도면이 본 발명에 따른 몇몇 실시 형태만을 도시하는 것이고 본 발명의 범위를 한정

하는 것으로 간주되지 않아야 함을 이해한다면; 본 발명은 첨부 도면의 사용을 통해 추가로 구체적으로 그리고 상세히 설명될 것이다.

### 발명을 실시하기 위한 구체적인 내용

[0033] 본 발명은 전반적으로, 재조합 세포에서, 예를 들어 백신 및 치료용 폴리펩티드와 같은 이종성 분자를 발현하기에 적합한 뛰어난 발현 잠재력을 갖는 바이러스 발현 시스템에 관한 것이다. 예를 들어, 본 발명의 일부 실시 형태는, 변형된 5'-비변역 영역(5'-UTR)을 포함하는 변형된 레플리콘 RNA를 함유하고, 선택적으로, 구조 단백질을 인코딩하는 그의 본래의 바이러스 서열의 적어도 일부가 결실된, 예를 들어 발현 작제물 및 벡터와 같은 핵산 분자에 관한 것이다. 또한, 이종성 폴리펩티드를 인코딩하는 하나 이상의 발현 카세트를 포함하는 바이러스 기반 발현 벡터가 본 발명의 일부 실시 형태에 따라 포함된다. 따라서, 본 명세서에 개시된 핵산 분자들 중 하나 이상을 포함하도록 유전자 변형된 재조합 세포뿐만 아니라, 그러한 세포로부터 유래된 생체 물질 및 재조합 생성물이 또한 본 출원의 범위 내에 있다. 본 발명의 특정 태양에서, 본 명세서에 개시된 분자 및/또는 재조합 세포 중 하나 이상을 포함하는 조성물이 추가로 제공된다.

[0034] RNA 바이러스(예를 들어, 알파바이러스)에 기반한 자가 증폭 RNA(레플리콘)는 강력한 발현 시스템으로서 사용될 수 있다. 예를 들어, 야생형 바이러스(예를 들어, 알파바이러스)의 5' 비변역 영역(UTR)에 대한 변형은 RNA 복제 및 RNA 전사 둘 모두에 관여하는 프로모터 요소를 포함하는 핵심 RNA 뉴클레오티드의 절제를 가능하게 할 수 있다. 5' UTR 서열의 조작에 의한 향상된 바이러스(예를 들어, 알파바이러스) 발현 시스템의 개발은 레플리콘 플랫폼 개발에서 중요한 진보를 나타낸다. 임의의 특정 이론에 의해 제한됨이 없이, 바이러스 발현 벡터로서 알파바이러스를 사용하는 비제한적인 이점은 이러한 알파바이러스가 재조합 숙주 세포에서 다량의 이종성 단백질의 합성을 지시할 수 있는 것이라고 믿어진다. 특히, 다른 이점들 중에서, 본 명세서에 개시된 알파바이러스 레플리콘 플랫폼 시스템은, 일부 실시 형태에서, 알파바이러스 레플리콘으로부터 정상적으로 발현되는 단백질의 양의 최대 3배를 발현할 수 있다. 이러한 개선은 표준 알파바이러스 레플리콘 시스템과 관련하여 주목되는 이미 천연적으로 높은 발현 수준을 고려하고 이 능력을 부여하는 5' UTR 돌연변이가 알파바이러스 복제 및 전사에 대한 거의 치명적인 돌연변이인 것으로 이전에 간주되었음을 고려할 때 유의한 것이다. 예를 들어, 치료용 단일 사슬 항체와 같은 폴리펩티드는 생체내에서 높은 수준으로 발현되는 경우 가장 효과적일 수 있다. 또한, 배양물에서(생체외에서) 세포로부터 정제된 재조합 항체를 생성하기 위하여, 레플리콘 RNA로부터의 높은 단백질 발현은 항체 생성물의 전체 수율을 증가시킬 수 있다. 더욱이, 발현되는 단백질이 백신 항원인 경우, 높은 수준의 발현은 생체내에서 가장 강력한 면역 반응을 유도할 수 있다.

[0035] 하기의 상세한 설명에서, 그의 일부를 형성하는 첨부 도면을 참조한다. 도면에서, 문맥상 달리 지시되지 않는 한, 유사한 기호는 전형적으로 유사한 성분을 식별한다. 상세한 설명, 도면, 및 청구범위에 기술된 예시적인 대안은 제한적인 것으로 의도되지 않는다. 다른 대안이 사용될 수 있고, 본 명세서에 제공된 대상 내용의 사상 또는 범주로부터 벗어남이 없이 다른 변경이 이루어질 수 있다. 본 명세서에 전반적으로 기술되고 도면에 예시된 바와 같은 태양들이 매우 다양한 상이한 구성으로 배열, 치환, 조합, 및 설계될 수 있으며, 이들 모두는 명백하게 고려되고 본 출원의 일부를 이룬다는 것이 용이하게 이해될 것이다.

[0036] 달리 정의되지 않는 한, 본 명세서에 사용되는 모든 기술 용어, 표기 및 다른 학술 용어 또는 전문 용어는 본 출원이 속하는 기술 분야의 숙련자에 의해 통상적으로 이해되는 의미를 갖는 것으로 의도된다. 일부 경우에, 통상적으로 이해되는 의미를 갖는 용어는 명확성을 위해 그리고/또는 용이한 참조를 위해 본 명세서에서 정의되고, 본 명세서 내의 그러한 정의의 포함은 당업계에서 일반적으로 이해되는 것에 비해 실질적인 차이를 나타내는 것으로 반드시 해석되어야 하는 것은 아니다. 본 명세서에 기술되거나 언급된 많은 기법 및 절차는 당업자에 의해 잘 이해되고 종래의 방법론을 사용하여 통상적으로 이용된다.

### 일부 정의

[0038] 단수 형태는 문맥이 명확하게 달리 지시하지 않는 한 복수의 지시 대상을 포함한다. 예를 들어, 용어 "세포"는 하나 이상의 세포를 포함하며, 그 혼합물을 포함한다. "A 및/또는 B"는 다음과 같은 대안을 모두 포함하도록 본 명세서에서 사용된다: "A", "B", "A 또는 B", 및 "A 및 B".

[0039] 본 명세서에 사용되는 바와 같은 용어 "약"은 대략이라는 그의 보통의 의미를 갖는다. 근사의 정도가 문맥으로부터 달리 명확하지 않은 경우, "약"은 제공된 값을 포함하는 모든 경우에 제공된 값의  $\pm 10\%$  이내이거나 가장 가까운 유효 숫자로 반올림됨을 의미한다. 범위가 제공되는 경우, 범위는 경계값을 포함한다.

[0040] 본 명세서에 사용되는 바와 같은 용어 "세포", "세포 배양물", "세포주", "재조합 숙주 세포", "수용자 세포"

및 "숙주 세포"는, 전달 횟수에 관계없이, 1차 대상 세포 및 그의 임의의 자손을 포함한다. ㅋ모든 자손이 (의도적인 또는 의도하지 않은 돌연변이 또는 환경 차이로 인해) 모 세포(parental cell)와 정확히 동일한 것은 아니지만, 그러한 변형된 자손은 자손이 본래 형질전환된 세포와 동일한 기능을 보유하는 한 이를 용어에 포함되는 것으로 이해되어야 한다.

[0041] 본 명세서에 사용되는 바와 같이, 용어 "작제물"은, 임의의 공급원으로부터 유래되고, 게놈 통합 또는 자율 복제가 가능하고, 하나 이상의 핵산 서열이 기능적으로 작동하는 방식으로 연결된, 예를 들어 작동 가능하게 연결된 핵산 분자를 포함하는, 임의의 재조합 핵산 분자, 예컨대 발현 카세트, 플라스미드, 코스미드, 바이러스, 자율 복제성 폴리뉴클레오티드 분자, 파지, 또는 선형 또는 원형, 단일 가닥, 또는 이중 가닥 DNA 또는 RNA 폴리뉴클레오티드 분자를 의미하도록 의도된다.

[0042] 용어 "유전자"는 단백질을 인코딩하거나 기능적 RNA로 전사될 수 있는 핵산 분자의 임의의 세그먼트를 지칭하기 위해 광범위하게 사용된다. 유전자는 전사되지만 최종, 성숙, 및/또는 기능적 RNA 전사체의 일부가 아닌 서열을 포함할 수 있고, 단백질을 인코딩하는 유전자는 전사되지만 번역되지 않는 서열, 예를 들어 5' 비번역 영역, 3' 비번역 영역, 인트론 등을 추가로 포함할 수 있다. 또한, 유전자는 선택적으로 그의 발현에 필요한 조절 서열을 추가로 포함할 수 있고, 그러한 서열은, 예를 들어, 전사되지 않거나 번역되지 않는 서열일 수 있다. 유전자는 관심 대상 공급원으로부터 클로닝하거나 공지된 또는 예측된 서열 정보로부터 합성하는 것을 포함하여 다양한 공급원으로부터 얻어질 수 있고, 원하는 파라미터를 갖도록 설계된 서열을 포함할 수 있다.

[0043] 폴리뉴클레오티드, 유전자, 또는 핵산 분자와 관련하여 사용될 때, 용어 "이종성"은 숙주 종으로부터 유래되지 않는 폴리뉴클레오티드, 유전자, 또는 핵산 분자를 지칭한다. 예를 들어, 본 명세서에 사용되는 바와 같은 "이종성 유전자" 또는 "이종성 핵산 서열"은 이것이 도입되는 숙주 유기체의 종과 상이한 종으로부터의 유전자 또는 핵산 서열을 지칭한다. 유전자 조절 서열 또는 유전자 서열의 발현을 조작하는 데 사용되는 보조 핵산 서열(예를 들어, 5' 비번역 영역, 3' 비번역 영역, 폴리 A 부가 서열 등) 또는 단백질 도메인 또는 단백질 국제화 서열을 인코딩하는 핵산 서열을 지칭할 때, "이종성"은 조절 또는 보조 서열, 또는 단백질 도메인 또는 국제화 서열을 인코딩하는 서열이 게놈에서 조절 또는 보조 핵산 서열, 또는 단백질 도메인 또는 국제화 서열을 인코딩하는 핵산 서열과 병치되어 있는 유전자와는 상이한 공급원으로부터의 것임을 의미한다. 따라서, (예를 들어, 유전자 조작되지 않은 유기체의 게놈에서) 천연 상태에서 작동 가능하게 연결되어 있지 않은 유전자에 작동 가능하게 연결된 프로모터는, 프로모터가 이것이 연결된 유전자와 동일한 종(또는 일부 경우에, 동일한 유기체)으로부터 유래될 수 있다고 하더라도 본 명세서에서 "이종성 프로모터"로 지칭된다. 예를 들어, 본 명세서에 개시된 일부 실시 형태에서, 이종성 관심 대상 유전자(GOI)의 코딩 서열은 그의 천연 상태에서 EAV 레플리콘 서열에 연결되어 있지 않다. 일부 실시 형태에서, 코딩 GOI 서열은 다른 유기체, 예를 들어 다른 바이러스, 세균, 진균, 인간 세포(예를 들어, 종양 Ag), 기생충(예를 들어, 말라리아) 등으로부터 유래된다.

[0044] 용어 "핵산 분자" 및 "폴리뉴클레오티드"는 본 명세서에서 상호 교환 가능하게 사용되고, cDNA, 게놈 DNA, 합성 DNA, 및 핵산 유사체를 함유하는 DNA 또는 RNA 분자를 포함하는 핵산 분자를 포함하는 RNA 및 DNA 분자 둘 모두를 지칭한다. 핵산 분자는 임의의 3차원 구조를 가질 수 있다. 핵산 분자는 이중 가닥 또는 단일 가닥(예를 들어, 센스 가닥 또는 안티센스 가닥)일 수 있다. 핵산 분자의 비제한적인 예로는 유전자, 유전자 단편, 엑손, 인트론, 메신저 RNA(mRNA), 전사 RNA, 리보솜 RNA, siRNA, 마이크로 RNA, tracrRNA, crRNA, 가이드 RNA, 리보자임, cDNA, 재조합 폴리뉴클레오티드, 분지형 폴리뉴클레오티드, 핵산 프로브 및 핵산 프라이머가 포함된다. 핵산 분자는 비통상적인 또는 변형된 뉴클레오티드를 함유할 수 있다. 본 명세서에서 상호 교환 가능하게 사용되는 바와 같은 용어 "폴리뉴클레오티드 서열" 및 "핵산 서열"은 폴리뉴클레오티드 분자의 서열을 지칭한다. 37 CFR § 1.822에 제시된 바와 같은 뉴클레오티드 염기에 대한 명명법이 본 명세서에서 사용된다.

[0045] 핵산 분자는 약 3 Kb 내지 약 50 Kb, 예를 들어 약 3 Kb 내지 약 40 Kb, 약 3 Kb 내지 약 40 Kb, 약 3 Kb 내지 약 30 Kb, 약 3 Kb 내지 약 20 Kb, 5 Kb 내지 약 40 Kb, 약 5 Kb 내지 약 40 Kb, 약 5 Kb 내지 약 30 Kb, 약 5 Kb 내지 약 20 Kb, 또는 약 10 Kb 내지 약 50 Kb, 예를 들어, 약 15 Kb 내지 30 Kb, 약 20 Kb 내지 약 50 Kb, 약 20 Kb 내지 약 40 Kb, 약 5 Kb 및 약 25 Kb, 또는 약 30 Kb 내지 약 50 Kb 인 핵산 분자를 포함하지만 이로 한정되지 않는 임의의 길이의 핵산 분자일 수 있다. 핵산 분자는 또한 예를 들어 50 kb를 초과할 수 있다.

[0046] 본 발명의 폴리뉴클레오티드는 다른 핵산에 하이브리드화하는 핵산의 능력, 또는 전사 인자 및/또는 핵산 중합 효소에 의해 인식되고 결합되는 폴리뉴클레오티드 서열의 능력과 같은 구조적 속성과 관련하여 "생물학적 활성"일 수 있다.

[0047] 본 명세서에 사용되는 바와 같은 용어 "재조합" 또는 "조작된" 핵산 분자는 인간 개입을 통해 변경된 핵산 분자

를 지칭한다. 비제한적인 예로서, cDNA는 재조합 DNA 분자이며, 이는 시험관내 중합효소 반응(들)에 의해 생성되거나, 링커가 부착되거나, 벡터, 예컨대 클로닝 벡터 또는 발현 벡터 내로 통합된 임의의 핵산 분자이다. 비제한적인 예로서, 재조합 핵산 분자는 1) 예를 들어, 화학적 또는 효소적 기법을 사용하여 (예를 들어, 화학적 핵산 합성의 사용에 의해, 또는 핵산 분자의 복제, 중합, 엑소뉴클레오리티ك(exonucleolytic) 분해, 엔도뉴클레오리티ك(endonucleolytic) 분해, 라이게이션, 역전사, 전사, (예를 들어, 메틸화를 포함하는) 염기 변형, 또는 (상동 및 부위 특이적 재조합을 포함하는) 재조합을 위한 효소의 사용에 의해) 시험관내에서 합성 또는 변형된 것이고/거나; 2) 자연 상태에서 결합되어 있지 않은 결합된 뉴클레오티드 서열들을 포함하고/거나, 3) 자연 발생적 핵산 분자 서열과 관련하여 하나 이상의 뉴클레오티드를 결여하도록 분자 클로닝 기법을 사용하여 조작된 것이고/거나, 4) 자연 발생적 핵산 서열과 관련하여 하나 이상의 서열 변화 또는 재배열을 갖도록 분자 클로닝 기법을 사용하여 조작된 것이다. 비제한적인 예로서, cDNA는 재조합 DNA 분자이며, 이는 시험관내 중합효소 반응(들)에 의해 생성되거나, 링커가 부착되거나, 벡터, 예컨대 클로닝 벡터 또는 발현 벡터 내로 통합된 임의의 핵산 분자이다.

[0048] 본 명세서에 사용되는 바와 같이, 바이러스 구조 폴리펩티드를 인코딩하는 핵산 서열의 "실질적인 부분"은, 당업자에 의한 서열의 수동 평가에 의해, 또는 BLAST(예를 들어, 문헌[Basic Local Alignment Search Tool]; Altschul SF et al., J. Mol. Biol. 215:403-410, 1993] 참조)와 같은 알고리듬을 이용한 컴퓨터 자동화 서열 비교 및 확인에 의해 그 바이러스 구조 폴리펩티드의 추정적 확인을 제공하기에 충분한 바이러스 구조 폴리펩티드를 인코딩하는 핵산 서열을 포함할 수 있다. 일반적으로, 당업자는 폴리펩티드 또는 핵산 서열을 공지된 단백질 또는 유전자와 상동성인 것으로 추정적으로 확인하기 위하여 10개 이상의 연속 아미노산 또는 30개 이상의 뉴클레오티드의 서열이 필요하다는 것을 인지할 것이다. 더욱이, 뉴클레오티드 서열과 관련하여, 15개 내지 30개의 연속 뉴클레오티드를 포함하는 유전자 특이적 올리고뉴클레오티드 프로브가 서열의 존적 유전자 확인 방법(예를 들어, 서던 하이브리드화) 및 단리(예를 들어, 세균 콜로니 또는 박테리오파지 플라크의 원위치 하이브리드화)에 사용될 수 있다. 또한, 10개 내지 15개 염기의 짧은 올리고뉴클레오티드가 PCR에서 증폭 프라이머로 사용되어, 프라이머를 포함하는 특정 핵산 단편을 얻을 수 있다. 따라서, 뉴클레오티드 서열의 "실질적인 부분"은 서열을 포함하는 핵산 단편의 특이적 확인 및/또는 단리를 제공하기에 충분한 서열을 포함한다. 본 발명은 하나 이상의 바이러스 구조 폴리펩티드를 인코딩하는 부분적 또는 완전한 핵산 서열이 없는 핵산 분자를 제공한다. 본 명세서에 개시된 바와 같은 서열의 이익을 갖는 당업자는 당업자에게 공지된 목적을 위해 개시된 서열의 전부 또는 실질적인 부분을 용이하게 사용할 수 있다. 따라서, 본 출원은 본 명세서에 개시된 바와 같은 완전한 서열, 예를 들어, 침부된 서열 목록에 제시된 것뿐만 아니라 상기 정의된 바와 같은 서열의 실질적인 부분을 포함한다.

[0049] 당업자에 의해 이해되는 바와 같이, 기재된 설명을 제공한다는 측면에서와 같은 임의의 그리고 모든 목적을 위해, 본 명세서에 개시된 모든 범위는 임의의 및 모든 가능한 하위 범위, 및 하위 범위의 조합을 또한 포함한다. 임의의 열거된 범위는 동일한 범위가 적어도 동일한 1/2, 1/3, 1/4, 1/5, 1/10 등으로 나뉘어짐을 충분히 설명하고 이를 가능하게 하는 것으로 용이하게 인식될 수 있다. 비제한적인 예로서, 본 명세서에 논의된 각각의 범위는 하위 1/3, 중간 1/3 및 상위 1/3 등으로 용이하게 나뉘어질 수 있다. 당업자에 의해 또한 이해되는 바와 같이, "최대", "적어도", "초과", "미만" 등과 같은 모든 표현은 언급된 수를 포함하고, 앞서 논의된 바와 같은 하위 범위로 후속하여 나뉘어질 수 있는 범위를 지칭한다. 마지막으로, 당업자에 의해 이해되는 바와 같이, 범위는 각각의 개별 구성원을 포함한다. 따라서, 예를 들어, 1개 내지 3개의 물품을 갖는 군은 1개, 2개, 또는 3개의 물품을 갖는 군들을 지칭한다. 유사하게, 1개 내지 5개의 물품을 갖는 군은 1개, 2개, 3개, 4개, 또는 5개의 물품을 갖는 군들 등을 지칭한다.

[0050] 본 명세서에 제공된 일반적인 방법의 논의는 단지 예시적인 목적을 위해 의도된다. 다른 대안적인 방법 및 대안이 본 출원의 검토시 당업자에게 명백할 것이고, 본 출원의 사상 및 범위 내에 포함되어야 한다.

#### 알파바이러스

[0052] 알파바이러스는 적어도 30개의 구성원을 포함하는 군 IV 토가비리대 과의 유전적으로, 구조적으로, 그리고 혈청학적으로 관련된 바이러스의 속이며, 이를 각각은 바이러스 스파이크 단백질을 함유하는 외피에 의해 둘러싸인 뉴클레오팩시드에 봉입된 양의 극성의 단일 가닥 RNA 게놈을 갖는다. 현재, 알파바이러스 속은 특히 신드비스 바이러스(SIN), 셈리키 포레스트 바이러스(SFV), 로스 리버 바이러스(RRV), 베네수엘라 말 뇌염 바이러스(VEEV), 및 동부 말 뇌염 바이러스(EEEV)를 포함하며, 이들은 모두 밀접하게 관련되고, 다양한 척추 동물, 예컨대 포유류, 설치류, 어류, 조류 종, 및 대형 포유류, 예컨대 인간 및 말뿐만 아니라 곤충과 같은 무척추 동물을 감염시킬 수 있다. 종간 및 개체간 전파는 모기를 통해 주로 발생하며, 이는 알파바이러스를 아보바이러스

(Arbovirus) 또는 아트로포드-본 바이러스(Arthropod-Borne Virus)의 수집에 대한 기여인자가 되게 한다. 특히, 신드비스 및 셈리키 포레스트 바이러스는 널리 연구되어 왔고, 이들 바이러스의 생활 주기, 복제 방식 등은 잘 특성 규명되어 있다. 특히, 알파바이러스는 동물 세포에서 매우 효율적으로 복제되는 것으로 밝혀졌으며, 이는 알파바이러스를 그러한 세포에서의 단백질 및 핵산의 생성을 위한 벡터로서 가치가 있게 한다.

[0053]

알파바이러스 입자는 외피가 있고, 직경이 70 nm이고, (다소 다형성이긴 하지만) 구형이 되려는 경향이 있고, 대략 40 nm의 등적성(isometric) 뉴클레오캡시드를 갖는다. **도 8**은 전형적인 알파바이러스 게놈 구조 및 게놈 발현을 도시한다(문헌[Strauss *et al.*, *Microbiological Reviews*, pp. 491-562, September 1994]으로부터 조정이 이루어진 것임). 알파바이러스 게놈은 대략 11 내지 12 kb 길이의 양의 극성의 단일 가닥 RNA로서, 5' 캡, 3' 폴리-A 테일, 및 효소 기능을 갖는 비구조 단백질을 인코딩하는 제1 프레임과 바이러스 구조 단백질(예를 들어, 캡시드 단백질 C, E1 당단백질, E2 당단백질, E3 단백질 및 6K 단백질)을 인코딩하는 제2 프레임을 갖는 2 개의 개방 해독 프레임을 포함한다.

[0054]

알파바이러스 게놈의 5'의 2/3은 바이러스 RNA의 전사 및 복제에 필요한 다수의 비구조 단백질을 인코딩한다. 이들 단백질은 RNA로부터 직접 번역되고, 세포 단백질과 함께 바이러스 게놈 복제 및 서브게놈 RNA의 전사에 필수적인 RNA 의존성 RNA 중합효소를 형성한다. 4개의 비구조 단백질(nsP1-4)은 단일 폴리프로테인으로서 생성되고, 바이러스의 복제 기구를 구성한다. 폴리프로테인의 처리는 고도로 조절된 방식으로 일어나며, 이때 P2/3 접합부에서의 절단은 게놈 복제 동안 RNA 주형 사용에 영향을 준다. 이 부위는 좁은 틈의 기저부에 위치하고 쉽게 접근할 수 없다. 일단 절단되면, nsP3는 nsP2를 둘러싸는 고리 구조를 생성한다. 이러한 두 단백질은 광범위한 계면을 갖는다. P2/P3 계면 영역에서 비세포변성 바이러스 또는 온도 민감성 표현형을 생성하는 nsP2에서의 돌연변이는 P2/P3 계면 영역에 밀집되어 있다. nsP2 비세포변성 돌연변이의 위치와 반대쪽에 있는 P3 돌연변이는 P2/3의 효율적인 절단을 방지한다. 이는 이어서 RNA 감염성에 영향을 주어 바이러스 RNA 생성 수준을 변경할 수 있다.

[0055]

게놈의 3'의 1/3은 바이러스 입자를 형성하는 데 필요한 모든 구조 단백질, 즉, 코어 뉴클레오캡시드 단백질 C, 및 이종이량체로서 회합하는 외피 단백질인 p62 및 E1의 번역을 위한 주형으로 작용하는 서브게놈 RNA를 포함한다. 바이러스 막 고정 표면 당단백질은 수용체 인식 및 막 융합을 통한 표적 세포 내로의 진입을 담당한다. 서브게놈 RNA는 nsp4 단백질을 인코딩하는 RNA 서열의 3' 단부에 존재하는 p26S 서브게놈 프로모터로부터 전사된다. E2 및 E3로의 P62의 단백분해 성숙은 바이러스 표면의 변화를 가져온다. E1, E2, 및 때로는 E3 당단백질 "스파이크(spike)"는 함께 E1/E2 이량체 또는 E1/E2/E3 삼량체를 형성하며, 여기서 E2는 중심부으로부터 정점까지 연장되고, E1은 정점들 사이의 공간을 채우고, E3은, 존재하는 경우, 스파이크의 원위 단부에 있다. 엔도솜의 산성으로의 바이러스의 노출 시에, E1은 E2로부터 해리되어 E1 동종삼량체를 형성하며, 이는 세포 및 바이러스 막을 함께 구동시키기 위한 융합 단계에 필요하다. 알파바이러스 당단백질 E1은 클래스 II 바이러스 융합 단백질이며, 이는 인플루엔자 바이러스 및 HIV에서 발견되는 클래스 I 융합 단백질과 구조적으로 상이하다. E2 당단백질은 그의 세포질 도메인을 통해 뉴클레오캡시드와 상호 작용하는 기능을 하는 한편, 그의 엑토도메인(ectodomain)은 세포 수용체와의 결합을 담당한다. 대부분의 알파바이러스는 말초 단백질 E3를 손실하는 반면, 셈리키 바이러스에서 이는 바이러스 표면과 회합된 상태로 유지된다.

[0056]

알파바이러스 복제는 세포의 세포질 내에서 일어나는 것으로 보고되었다. 감염 주기의 제1 단계에서, 게놈 RNA의 5' 단부는 게놈 RNA에 상보적인 음성 가닥을 생성하는 RNA 중합효소 활성을 갖는 폴리프로테인(nsP1-4)으로 번역된다. 제2 단계에서, 음성 가닥은 다음과 같은 2개의 RNA 각각의 생성을 위한 주형으로서 사용된다: (1) 번역에 의해 다른 nsp 단백질을 생성하는 2차 바이러스의 게놈에 상응하고 바이러스에 대한 게놈으로서 작용하는 양성 게놈 RNA; 및 (2) 감염성 입자를 형성하는 바이러스의 구조 단백질을 인코딩하는 서브게놈 RNA. 양성 게놈 RNA/서브게놈 RNA의 비는 nsp 1, nsp 2, nsp 3 및 nsp 4로의 폴리프로테인의 단백분해 자가절단에 의해 조절된다. 실제로, 바이러스 유전자 발현은 두 단계로 일어난다. 제1 단계에서는, 양성 게놈 가닥 및 음성 가닥의 주된 합성이 존재한다. 제2 단계 동안, 서브게놈 RNA의 합성은 사실상 배타적이며, 이에 따라 다양한 구조 단백질의 생성을 가져온다.

[0057]

알파바이러스의 5'-비번역 영역(5'-UTR)의 이전의 상세한 분석은 RNA 복제 및 전사를 지지하는 극단의 5' 뉴클레오티드의 절대적인 중요성을 밝혀내었다. 특히, **도 1**에 도시된 바와 같이, 5' UTR 서열의 뉴클레오티드 위치 1 및 2 각각에서의 AU 다이뉴클레오티드의 보존이 모든 알파바이러스 사이에서 주목되며, 이는 이러한 뉴클레오티드의 중요성을 암시한다. 본 명세서에 사용되는 바와 같이, "A1"은 5'-UTR(예를 들어, 알파바이러스 5'-UTR)의 뉴클레오티드 위치 1에서의 보존된 A 뉴클레오티드를 지칭하고, "U2"는 5'-UTR(예를 들어, 알파바이러스 5'-UTR)의 뉴클레오티드 위치 2에서의 보존된 U 뉴클레오티드를 지칭한다. 또한, 베네수엘라 말 뇌염 바이러스

(VEEV)의 경우, 5'의 처음(5' most) 3개의 뉴클레오티드뿐만 아니라 이 서열 바로 다음에 발견되는 스템 루프 영역의 상세한 분석이 수행되었다. 특히, 5' UTR의 위치 2에서 U 잔기를 유지하는 것의 중요성이 이미 확정되었다(문헌[Kulasegara-Shylini *et al.*, *J. Virol.* 83:17 p 8327-8339, 2009a]; 및 문헌[Kulasegara-Shylini *et al.*, *J. Virol.* 83:17 p 8327-8339, 2009b]). 구체적으로, 5' UTR에서 단일 U2→G 변화를 함유하는 (G2)VEE/SINV로 명명된 전장 감염성 클론으로부터의 시험관내 전사된 RNA는 야생형 VEE/SINV 감염성 클론으로부터의 시험관내 전사된 RNA와 비교하여 거의 3 로그의 감염성의 손실을 나타내었다. 이 보고서는 위치 2에 있는 U가 RNA 복제에 중요하고, G로 대체될 수 없음을 강하게 암시한다. 그러나, 본 명세서에 상세히 기술된 바와 같이, 5' UTR에서 U2→G 변화를 갖는 VEEV 레플리콘은 놀랍게도 그리고 이러한 이전의 보고서와 정반대로 강력한 복제가 전적으로 가능할 뿐만 아니라 위치 2에 U 잔기를 함유하는 야생형 5' UTR과 비교하여 3배의 VEEV 레플리콘의 발현 잠재력을 가져온다.

[0058]

대부분의 RNA 바이러스의 극단의 5' 및 3' 서열은 고도로 제약을 받고, 임의의 변화가 허용되는 경우 거의 존재하지 않으며; 대부분의 변형은 RNA 복제에 대해 고도로 손상된 또는 치명적인 결과를 초래한다. 쿠라세가란-실리니(Kulasegara-Shylini) 등은 모든 알파바이러스를 대표하는 키메라 VEEV/SINV 감염성 클론에 대한 RNA 복제에 중요한 5' 뉴클레오티드 서열의 심층 분석을 완료하였다(상기 문헌[Kulasegaran-Shylini *et al.* 2009a] 참조). 이 보고서는 임의의 특정 알파바이러스의 5' 단부에서 일어날 수 있는 RNA 서열 변이에서의 제한을 명확하게 지지하는 많은 연구자에 의해 25년 동안 수행된 분석을 기반으로 한다. 쿠라세가란-실리니 등의 2009b 논문(문헌[J. Virol. 83:17 p 8327-8339, 2009])은 5' UTR에서의 U 잔기로부터 G 잔기로의 뉴클레오티드 2의 변화(U2→G)가 그 감염성 클론 RNA의 생존력을 유의하게 감소시킨다는 것을 구체적으로 명시/제시한다. 즉, 5'-UTR에서의 그 특정 변화는 감염성 클론 RNA의 생물학적 활성을 거의 1,000배 감소시켰다. 본 명세서에 개시된 바와 같이, VEEV(균주 TC-83) 레플리콘 RNA 내로 포함된 5'-UTR에서의 변화(예를 들어, U2→G 변화)는 레플리콘의 복제를 손상시키지 않을 뿐만 아니라 레플리콘의 생물학적 활성을 실제로 증가시킬 수 있다. 예를 들어, U2→G 치환을 포함하는 레플리콘은, 일부 실시 형태에서, 동일한 단백질을 발현하는 야생형 레플리콘보다 3배 더 많은 관심 대상 단백질의 발현을 가져올 수 있다. 이러한 결과는 놀라운 것이고, U2→G 변화를 지닌 레플리콘의 증가된 생물학적 활성을 예측될 수 없었다. 이러한 변형된 레플리콘은 백신 및 치료적 적용 둘 모두를 지지하기 위한 뛰어난 RNA 발현 플랫폼이 될 잠재력을 갖는다.

[0059]

5'의 처음 2개의 뉴클레오티드의 보존은 알파바이러스 아형의 모든 계놈 RNA에 걸쳐 관찰되었다. 보존된 AU 다이뉴클레오티드(A1 및 U2)는 RNA 복제를 위해 결정적으로 필요한 것으로 또한 밝혀졌다(상기 문헌[Kulasegaran-Shylini *et al.* 2009a and 2009b]). 극단의 5' 단부에 AG 다이뉴클레오티드를 지닌 알파바이러스 레플리콘 RNA가 전적으로 기능적일 뿐만 아니라 향상된 생물학적 활성을 나타낸다는 입증은 놀라운 것이고, 현장의 정설과 완전히 반대되는 것이다.

[0060]

본 명세서에 개시된 바와 같이, 단일유전자(monogenic) 또는 다유전자(multigenic) 알파바이러스 발현 시스템은, 예를 들어, 변형된 5'-UTR을 함유하는 레플리콘 RNA와 같은 발현/번역 향상 활성을 갖는 변형된 레플리콘 RNA를 사용함으로써 생성될 수 있다. 일부 실시 형태에서, 본 명세서에 기재된 바와 같은 바이러스(예를 들어, 알파바이러스) 발현 시스템은 하나 이상의 바이러스 구조 단백질에 대한 일부 또는 전체 코딩 영역이 추가로 없다. 예를 들어, 알파바이러스 발현 시스템은 바이러스 캡시드 단백질 C, E1 당단백질, E2 당단백질, E3 단백질 및 6K 단백질 중 하나 이상에 대한 전체 코딩 서열 또는 그의 일부가 없을 수 있다. 일부 실시 형태에서, 레플리콘 RNA와 관련된, 베네수엘라 말 뇌염 바이러스(VEEV)의 5' UTR 서열의 cDNA 카페에서의 티민(T) 뉴클레오티드로부터 구아닌(G) 뉴클레오티드로의 위치 2에서의 뉴클레오티드의 변형(T2→G 돌연변이)은 야생형 5' UTR 서열을 갖는 VEEV 레플리콘과 비교하여 유의하게 높은 단백질 발현 잠재력을 레플리콘에 부여한다.

[0061]

일부 실시 형태에서, 본 명세서에 개시된 바와 같은 변형된 레플리콘 RNA의 발현 및/또는 번역 향상 활성의 수준은, 상응하는 비변형된 레플리콘, 예를 들어 야생형 5' UTR을 갖는 레플리콘으로부터 검출된 발현 수준에 비해, 적어도 1.1배, 1.2배, 1.3배, 1.4배, 1.5배, 1.6배, 1.7배, 1.8배, 1.9배, 2배, 3배, 4배, 5배, 6배, 7배, 8배, 또는 그 초과의 배수이다. 임의의 특정 이론에 의해 제한됨이 없이, 향상된 번역은 번역에 이용 가능한 전사체의 증가된 수준을 가져오는 전사의 향상에 기인할 수 있고/거나 전사와 독립적일 수 있고, 예를 들어, 향상된 리보솜 결합에 기인할 수 있다. 향상 활성의 수준은 전사체 수준, 단백질의 양, 단백질 활성 등을 포함하지만 이로 한정되지 않는 당업계에 공지된 임의의 편리한 방법 및 기법에 의해 측정될 수 있다(예를 들어, 하기 실시예 1, 실시예 3 내지 실시예 5 참조).

[0062]

핵산 분자

[0063]

일 태양에서, 변형된 레플리콘 RNA를 포함하는 신규한 핵산 분자가 본 명세서에 개시되어 있다. 예를 들어, 변형된 레플리콘 RNA는 모(parent) 레플리콘 RNA의 본래의 게놈 영역(예를 들어, 개방 해독 프레임(ORF) 및/또는 비코딩 영역(예를 들어, 프로모터 서열)) 중 하나 이상에서 돌연변이(들), 결실(들), 치환(들), 및/또는 삽입(들)을 포함할 수 있다. 일부 실시 형태에서, 변형된 레플리콘 RNA는 변형된 5'-비번역 영역(5'-UTR)을 포함한다. 일부 실시 형태에서, 변형된 5'-UTR은 위치 1, 2, 4, 또는 이들의 조합에서 하나 이상의 뉴클레오티드 치환을 포함한다. 일부 실시 형태에서, 뉴클레오티드 치환 중 적어도 하나는 변형된 5'-UTR의 위치 1에서의 뉴클레오티드 치환이다. 일부 실시 형태에서, 뉴클레오티드 치환 중 적어도 하나는 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환이다. 일부 실시 형태에서, 뉴클레오티드 치환 중 적어도 하나는 변형된 5'-UTR의 위치 4에서의 뉴클레오티드 치환이다. 일부 실시 형태에서, 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→G 치환이다. 일부 실시 형태에서, 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→A 치환이다. 일부 실시 형태에서, 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→C 치환이다.

[0064]

본 명세서에 사용되는 바와 같이, 용어 "레플리콘 RNA"는 허용 세포(permissive cell) 내에서 그 자신의 증폭 또는 자가 복제를 지시하는 데 필요한 모든 유전 정보를 함유하는 RNA를 지칭한다. 그 자신의 복제를 지시하기 위하여, RNA 분자는 1) RNA 증폭 과정을 촉매하기 위해 바이러스 또는 숙주 세포 유래 단백질, 핵산 또는 리보뉴클레오단백질과 상호 작용할 수 있는 중합효소, 레플리카제, 또는 다른 단백질을 인코딩하고; 2) 서브게놈 레플리콘 인코딩 RNA의 복제 및 전사에 필요한 시스(cis) 작용성 RNA 서열을 함유한다. 이러한 서열은 복제 과정 동안 그의 자가 인코딩 단백질, 또는 비-자가 인코딩 세포 유래 단백질, 핵산 또는 리보뉴클레오단백질, 또는 이들 성분 중 임의의 것 사이의 복합체에 결합될 수 있다. 본 발명의 목적을 위해, 알파바이러스 유래 레플리콘 RNA 분자는 전형적으로 다음과 같은 순서의 요소들을 함유한다: 복제를 위해 시스로 필요한 5' 바이러스 또는 결손-간섭 RNA 서열(들), 생물학적으로 활성인 알파바이러스 비구조 단백질(예를 들어, nsP1, nsP2, nsP3, 및 nsP4)을 코딩하는 서열, 서브게놈 RNA에 대한 프로모터, 복제를 위해 시스로 필요한 3' 바이러스 서열, 및 폴리아데닐레이트 트랙(polyadenylate tract). 또한, 용어 레플리콘 RNA는 일반적으로 양의 극성 또는 "메시지(message)" 센스의 분자를 지칭하고, 레플리콘 RNA는 임의의 공지된 자연 발생적 알파바이러스와는 상이한 길이를 가질 수 있다. 본 발명의 일부 실시 형태에서, 레플리콘 RNA는 구조 바이러스 단백질 중 적어도 하나의 서열을 함유하지 않으며; 구조 유전자를 인코딩하는 서열은 이종성 서열로 치환될 수 있다. 이러한 경우에, 레플리콘 RNA가 재조합 알파바이러스 입자로 패키징되는 경우, 이는 입자 형성으로 이어지는 알파바이러스 구조 단백질과의 상호 작용을 개시하는 역할을 하는 하나 이상의 서열, 소위 패키징 신호를 함유하여야 한다.

[0065]

본 명세서에 사용되는 바와 같이, "서브게놈 RNA"는 그것이 유래된 게놈 RNA보다 작은 길이 또는 크기의 RNA 분자를 지칭한다. 알파바이러스 서브게놈 RNA는, 서열이 게놈 RNA 또는 그의 상보체 내에 존재하는 내부 프로모터로부터 전사되어야 한다. 알파바이러스 서브게놈 RNA의 전사는 숙주 세포 인코딩 단백질, 리보뉴클레오단백질(들), 또는 이들의 조합과 관련된 바이러스 인코딩 중합효소(들)에 의해 매개될 수 있다. 본 발명의 일부 실시 형태에서, 서브게놈 RNA는 본 명세서에 개시된 바와 같은 변형된 레플리콘 RNA로부터 생성되고, 하나 이상의 관심 대상 유전자(GOI)를 인코딩 또는 발현한다. 천연 서브게놈 프로모터 대신에, 서브게놈 RNA는 뇌심근염 바이러스(EMCV), 소 바이러스성 설사 바이러스(BVDV), 폴리오바이러스, 구제역 바이러스(FMD), 엔테로바이러스 71, 또는 C 형 간염 바이러스로부터 유래된 내부 리보솜 진입 부위(IRES)의 제어 하에 놓여질 수 있다.

[0066]

따라서, 일부 실시 형태에서, 하나 이상의 바이러스 구조 단백질에 대한 일부 또는 전체 코딩 서열은 본 명세서에 개시된 핵산 분자에 부재하고/하거나 그러한 핵산 분자에서 변형된다. 따라서, 일부 특정 실시 형태에서, 본 명세서에 개시된 바와 같은 변형된 레플리콘 RNA는 변형된 5'-UTR을 포함하고, 하나 이상의 바이러스 구조 단백질을 인코딩하는 핵산 서열의 적어도 일부가 없는데, 예를 들어, 바이러스 구조 단백질을 인코딩하는 핵산 서열의 처음 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개 또는 그 초과의 뉴클레오티드가 없다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 구조 폴리펩티드인 E1, E2, E3, 6K, 및 캡시드 단백질 C 중 하나 이상을 인코딩하는 서열의 약 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 또는 그 초과가 없을 수 있다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 구조 폴리펩티드인 E1, E2, E3, 6K, 및 캡시드 단백질 C 중 하나 이상을 인코딩하는 전체 서열 또는 그의 실질적인 부분이 없다. 본 명세서에 사용되는 바와 같이, 바이러스 구조 단백질을 인코딩하는 핵산 서열의 "실질적인 부분"은, 당업자에 의한 서열의 수동 평가에 의해, 또는 BLAST(예를 들어, 상기 문헌[Altschul SF et al. 1993] 참조)와 같은 알고리듬을 이용한 컴퓨터 자동화 서열 비교 및 확인에 의해 그 바이러스 구조 단백질의 추정적 확인을 제공하기에 충분한 바이러스 구조 단백질을 인코딩하는 핵산 서열을 포함한다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 구조 폴리펩티드인 E1, E2, E3, 6K, 및 캡시드 단백질 C 중 하나 이상을 인코딩

하는 전체 서열이 없다.

[0067] 본 출원의 일부 특정 실시 형태에서, 본 명세서에 개시된 바와 같은 핵산 분자는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하며, 핵산 분자는 서열 번호 1의 핵산 서열에 대해 적어도 80%의 서열 동일성을 나타내고, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 5'-비번역 영역(5'-UTR)의 위치 2에서 U→G 치환을 포함하고 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없다. 일부 실시 형태에서, 핵산 분자는 서열 번호 1의 핵산 서열에 대해 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%의 서열 동일성을 나타낸다. 일부 실시 형태에서, 핵산 분자는 서열 번호 1의 핵산 서열에 대해 100%의 서열 동일성을 나타낸다.

[0068] 일부 실시 형태에서, 본 명세서에 개시된 바와 같은 핵산 분자는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함하며, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 본 명세서에 개시된 적어도 하나의 5'-UTR의 핵산 서열에 대해 적어도 80%의 서열 동일성을 나타내는 5'-UTR 및 5'-UTR의 위치 2에서의 U→G 치환을 포함하고, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 바이러스 구조 단백질을 인코딩하는 서열의 적어도 일부가 없다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 2 내지 서열 번호 18에 제시된 서열들 중 적어도 하나에 대해 적어도 80%의 서열 동일성을 나타내는 5'-UTR을 포함한다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 번호 2 내지 서열 번호 18에 제시된 서열들 중 적어도 하나에 대해 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%의 서열 동일성을 나타내는 5'-UTR을 포함한다. 일부 실시 형태에서, 변형된 알파바이러스 게놈 또는 레플리콘 RNA는 서열 목록의 서열 번호 2 내지 서열 번호 18에 제시된 서열들 중 적어도 하나에 대해 100%의 서열 동일성을 나타내는 5'-UTR을 포함한다.

[0069] 본 명세서에 개시된 5' UTR의 핵산 서열에 대해 고도의 서열 동일성(예를 들어, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%)을 갖는 핵산 분자는, 본 명세서에서 확인된 서열(예를 들어, 서열 번호 1 내지 서열 번호 18) 또는 당업계에 공지된 바와 같은 임의의 다른 알파바이러스 5' UTR, 예를 들어 GenBank/NCBI 수탁 번호 J02363, NC\_001547, U38305, L04653, NC\_001449, U38304, X04129, NC\_003215, 및 TR339 게놈 서열을 갖는 서열(문헌[Klimstra *et al.*, J. Virol. 72:7357, 1988]; 문헌 [McKnight *et al.*, J. Virol. 70:1981, 1996])을 사용함으로써 확인 및/또는 단리될 수 있고, 각각의 알파바이러스 게놈에서 확인된 서열로부터의 축퇴 프라이머 또는 유전자 특이적 프라이머를 이용한 게놈 서열 분석, 하이브리드화, 및/또는 PCR에 의해 확인 및/또는 단리될 수 있다. 본 명세서에 사용되는 바와 같이, "서열 동일성"은 2개의 최적으로 정렬된 폴리뉴클레오티드가 성분들, 예를 들어, 뉴클레오티드들의 정렬 윈도우 전체에 걸쳐 불변인 정도를 지칭한다. 시험 서열과 기준 서열의 정렬된 세그먼트에 대한 "동일성 분율"은 2개의 정렬된 서열이 공유하는 동일한 성분 수를 기준 서열 세그먼트, 예를 들어 전체 기준 서열 또는 기준 서열의 정의된 더 작은 부분 내의 전체 성분 수로 나눈 것이다.

[0070] 본 명세서에 개시된 일부 실시 형태는 변형된 레플리콘 RNA를 포함하는 핵산 분자에 관한 것이며, 변형된 레플리콘 RNA는 번역 향상 활성을 갖는, 예를 들어 5' UTR과 같은 변형된 알파바이러스 염기 서열을 포함한다. 그러한 변형된 레플리콘 RNA는 원하는 생성물을 코딩하는 이종성 핵산 서열(예를 들어, DNA 또는 cDNA)의 발현의 향상된 수준을 달성하는 데 사용될 수 있다. 일부 실시 형태에서, 변형된 레플리콘 RNA는, 예를 들어 세포 배양물 내의 세포이거나 생체 내의 세포일 수 있는 세포에서 변형된 레플리콘의 도입 후에 원하는 생성물을 코딩하는 이종성 핵산 서열(예를 들어, DNA 또는 cDNA)의 발현의 향상된 수준을 달성하는 데 사용된다.

[0071] 또한, 일부 실시 형태에서, 핵산 분자는 바이러스 조작 및/또는 투여의 안전성을 증가시키도록 하나 이상의 약화 돌연변이(attenuating mutation)를 함유하는 변형된 알파바이러스 게놈 또는 레플리콘 RNA를 포함할 수 있다. 본 명세서에 사용되는 바와 같은 문구 "약화 돌연변이"는, 돌연변이가 치환 돌연변이이든지 또는 프레임 내 결실 또는 삽입 돌연변이이든지 간에, 당업계의 표준 전문 용어에 따라, 숙주에서 질병을 야기할 확률의 감소(즉, 병독성의 손실)를 가져오는 뉴클레오티드 돌연변이 또는 그러한 돌연변이로 인해 인코딩된 아미노산을 의미한다. 약화 돌연변이는 알파바이러스 게놈의 코딩 영역 또는 비코딩 영역(예를 들어, 5' UTR)에 있을 수 있다. 당업자에게 공지된 바와 같이, 문구 "약화 돌연변이"는 바이러스에 치명적인 돌연변이 또는 돌연변이들의 조합을 배제한다. 또한, 당업자는 일부 약화 돌연변이가 제2 부위 억제 돌연변이(들)의 부재 하에서 치명적일 수 있음을 인지할 것이다.

[0072] 이러한 신규 핵산 분자를 삭제하고 특성 규명하는 문자적 기법 및 방법은 본 출원의 실시예에 더욱 충분히 기재되어 있다. 실시예 섹션에서, 베네수엘라 말 뇌염 바이러스(VEEV)가 본 명세서에 개시된 조성물 및 방법을 예

시하기 위해 사용되었다.

[0073] 일부 실시 형태에서, 핵산 분자는 재조합 핵산 분자이다. 본 명세서에 사용되는 바와 같이, 재조합이라는 용어는 임의의 문자(예를 들어, DNA, RNA, 폴리펩티드), 즉 인간 조작으로부터의 결과이지만 간접적인 것을 의미한다. 비제한적인 예로서, cDNA는 재조합 DNA 분자이며, 이는 시험관내 중합효소 반응(들)에 의해 생성되거나, 링커가 부착되거나, 백터, 예컨대 클로닝 베터 또는 발현 베터 내로 통합된 임의의 핵산 분자이다. 비제한적인 예로서, 재조합 핵산 분자는 1) 예를 들어, 화학적 또는 효소적 기법을 사용하여 (예를 들어, 화학적 핵산 합성의 사용에 의해, 또는 핵산 분자의 복제, 중합, 엑소뉴클레오리티 분해, 엔도뉴클레오리티 분해, 라이게이션, 역전사, 전사, (예를 들어, 메틸화를 포함하는) 염기 변형, 또는 (상동 및 부위 특이적 재조합을 포함하는) 재조합을 위한 효소의 사용에 의해) 시험관내에서 합성 또는 변형된 것이고/거나; 2) 자연 상태에서 결합되어 있지 않은 결합된 뉴클레오티드 서열들을 포함하고/거나; 3) 자연 발생적 뉴클레오티드 서열과 관련하여 하나 이상의 뉴클레오티드를 결여하도록 문자 클로닝 기법을 사용하여 조작된 것이고/거나, 4) 자연 발생적 뉴클레오티드 서열과 관련하여 하나 이상의 서열 변화 또는 재배열을 갖도록 문자 클로닝 기법을 사용하여 조작된 것이다.

[0074] 일부 실시 형태에서, 본 명세서에 개시된 핵산 분자는 재조합 DNA 기술(예를 들어, 중합효소 연쇄 반응(PCR) 증폭, 클로닝 등) 또는 화학적 합성을 이용하여 생성된다. 본 명세서에 개시된 바와 같은 핵산 분자는, 천연 대립 유전자 변이체 및 하나 이상의 뉴클레오티드 잔기가 삽입, 결실, 및/또는 치환된 변형된 핵산 분자로서, 그러한 변형이 본 명세서에 기재된 바와 같은 생물학적 활성을 달성하는 데 있어서 원하는 특성을 제공하는 방식으로 이루어진 변형된 핵산 분자를 포함하지만 이로 한정되지 않는, 천연 핵산 분자 및 그의 상동체를 포함한다.

[0075] 자연 발생적 핵산 서열의 변이체를 포함하는 핵산 분자는 당업자에게 공지된 다수의 방법을 이용하여 생성될 수 있다(예를 들어, 문헌[Sambrook *et al.*, In: Molecular Cloning, A Laboratory Manual, 2nd Edition, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1989)] 참조). 핵산 분자의 서열은, 이것이 유래된 자연 발생적 서열과 관련하여, 부위 특이적 돌연변이유발, 돌연변이를 유도하기 위한 핵산 분자의 화학적 처리, 핵산 단편의 제한 효소 절단, 핵산 단편들의 라이게이션, 핵산 서열의 선택된 영역의 PCR 증폭 및/또는 돌연변이유발, 재조합적 클로닝, 및 올리고뉴클레오티드 혼합물의 화학 합성 및 핵산 분자들의 혼합물, 및 이들의 조합을 "구축"하기 위한 혼합물 군의 라이게이션을 포함하는 화학 합성과 같은 그러나 이로 한정되지 않는 고전적 돌연변이유발 기법 및 재조합 DNA 기법을 포함하지만 이로 한정되지 않는 다양한 기법을 이용하여 변형될 수 있다. 핵산 분자 상동체는 핵산 분자에 의해 인코딩되는 단백질 또는 레플리콘의 기능에 대한 스크리닝에 의해 및/또는 야생형 유전자 또는 이의 단편과의 하이브리드화에 의해, 또는 표적 또는 야생형 핵산 분자 또는 서열에 대해 상동성을 갖는 프라이머를 사용하는 PCR에 의해, 변형된 핵산 분자들의 혼합물로부터 선택될 수 있다.

[0076] 본 명세서에 개시된 다양한 실시 형태에서, 본 명세서에 개시된 핵산 분자는 다음과 같은 특징들 중 하나 이상을 포함할 수 있다. 일부 실시 형태에서, 변형된 레플리콘 RNA는 변형된 알파바이러스 레플리콘 RNA이다. 일부 실시 형태에서, 변형된 알파바이러스 레플리콘 RNA는 변형된 알파바이러스 계놈을 포함한다. 일부 실시 형태에서, 변형된 5'-UTR은 위치 1, 2, 4, 또는 이들의 조합에서 하나 이상의 뉴클레오티드 치환을 포함한다. 특정 실시 형태에서, 뉴클레오티드 치환 중 적어도 하나는 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환이다. 일부 특정 실시 형태에서, 변형된 5'-UTR의 위치 2에서의 뉴클레오티드 치환은 U→G 치환이다.

[0077] 본 명세서에 개시된 일부 실시 형태에서, 변형된 알파바이러스 계놈 또는 레플리콘 RNA는 이종성 조절 요소에 작동 가능하게 연결된다. 본 명세서에 사용되는 바와 같이, "조절 요소", "조절 서열", 또는 "조절 요소 서열"은, 예를 들어, 폴리펩티드 인코딩 서열 또는 기능적 RNA 인코딩 서열과 같은 코딩 서열의 상류(5'), 내부, 또는 하류(3')에 위치하는 뉴클레오티드 서열을 지칭한다. 코딩 서열의 전사 및/또는 코딩 서열의 전사로부터 생성되는 RNA 분자의 번역은 전형적으로 조절 요소의 존재 또는 부재에 의해 영향을 받는다. 이러한 조절 요소는 프로모터, 시스 요소, 인핸서, 종결자, 또는 인트론을 포함할 수 있다. 당업자는 본 명세서에 기재된 조절 요소가 키메라 또는 하이브리드 조절 발현 요소에 존재할 수 있음을 인지할 것이다. 일부 실시 형태에서, 이종성 조절 요소는 프로모터 서열이거나 이를 포함한다. 이종성 프로모터 서열은 임의의 이종성 프로모터 서열, 예를 들어 SP6 프로모터, T3 프로모터, 또는 T7 프로모터, 또는 이들의 조합일 수 있다. 일부 특정 실시 형태에서, 프로모터 서열은 T7 프로모터 서열을 포함한다.

[0078] 또한, 일부 실시 형태에서, 변형된 알파바이러스 계놈 또는 레플리콘 RNA는 하나 이상의 이종성 전사 종결 신호 서열을 포함할 수 있다. 본 명세서에서 상호 교환 가능하게 사용되는 바와 같이, 용어 "전사 종결 신호", "종

결자" 또는 "종결자 서열" 또는 "전사 종결자"는 RNA 중합효소가 전사를 중지하게 하는 유전자 서열의 조절 색션을 지칭한다. 이종성 전사 종결 신호 서열은 일반적으로 임의의 이종성 전사 종결 신호 서열, 예를 들어 SP6 종결 신호 서열, T3 종결 신호 서열, T7 종결 신호 서열, 또는 그의 변이체일 수 있다. 따라서, 본 발명의 일부 실시 형태에 따른 핵산 분자는 SP6 종결 신호 서열, T3 종결 신호 서열, T7 종결 신호 서열, 또는 그의 변이체로 이루어진 군으로부터 선택되는 하나 이상의 이종성 전사 종결 신호 서열 중 적어도 하나를 포함할 수 있다. 일부 특정 실시 형태에서, 전사 종결 서열은 T7 종결 신호 서열을 포함한다.

[0079] 일부 실시 형태에서, 본원에 개시된 핵산 분자는 하나 이상의 발현 카세트를 포함할 수 있다. 원칙적으로, 본 명세서에 개시된 핵산 분자는 일반적으로 임의의 수의 발현 카세트를 포함할 수 있다. 일부 특정 실시 형태에서, 본 명세서에 개시된 핵산 분자는 적어도 2개, 적어도 3개, 적어도 4개, 적어도 5개, 또는 적어도 6개의 발현 카세트를 포함할 수 있다. 본 명세서에 사용되는 바와 같이, 용어 "발현 카세트"는, 코딩 서열 및 생체내 및/또는 생체외에서 수용자 세포에서 코딩 서열의 적절한 전사 및/또는 번역을 지시하기에 충분한 조절 정보를 함유하는 유전 물질의 작제물을 지칭한다. 발현 카세트는 원하는 숙주 세포로의 표적화를 위해 벡터 내로 및/또는 대상체 내로 삽입될 수 있다. 또한, 발현 카세트라는 용어는 용어 "발현 작제물"과 상호 교환 가능하게 사용될 수 있다. 일부 실시 형태에서, 용어 "발현 카세트"는, 예를 들어 프로모터 및/또는 종결 신호와 같은 조절 요소에 작동 가능하게 연결된, 단백질 또는 기능적 RNA를 인코딩하는 유전자, 및 선택적으로, 유전자의 전사 또는 번역에 영향을 주는 다른 핵산 서열들 중 임의의 것 또는 이들의 조합을 포함하는 핵산 작제물을 지칭한다.

[0080] 본 명세서에 사용되는 바와 같은 용어 "작동 가능하게 연결된"은 둘 이상의 서열 사이의 기능적 연결을 나타낸다. 예를 들어, 관심 대상 폴리뉴클레오티드와 조절 서열(예를 들어, 프로모터) 사이의 작동 가능한 연결은 관심 대상 폴리뉴클레오티드의 발현을 가능하게 하는 기능적 결합이다. 이러한 의미에서, 용어 "작동 가능하게 연결된"은 조절 영역이 관심 대상 코딩 서열의 전사 또는 번역을 조절하는 데 효과적이 되도록 하는 전사될 코딩 서열 및 조절 영역의 위치 설정을 지칭한다. 본 명세서에 개시된 일부 실시 형태에서, 용어 "작동 가능하게 연결된"은, 제어 서열이 폴리펩티드를 인코딩하는 mRNA, 폴리펩티드, 또는 기능적 RNA의 발현 또는 세포 국재화를 지시 또는 조절하도록 조절 서열이 폴리펩티드 또는 기능적 RNA를 인코딩하는 서열에 대하여 적절한 위치에 배치되는 구성을 나타낸다. 따라서, 프로모터가 핵산 서열의 전사를 매개할 수 있는 경우 프로모터는 핵산 서열과 작동 가능한 연결 상태로 있다. 작동 가능하게 연결된 요소는 연속적이거나 비연속적일 수 있다.

[0081] DNA의 둘 이상의 서열을 함께 작동 가능하게 연결하기 위한 기본적 기법은 당업자에게 친숙하고, 그러한 방법은 표준 분자 생물학적 조작을 위한 다수의 교재에 기재되어 있다(예를 들어, 문헌[Maniatis *et al.*, "Molecular Cloning: A Laboratory Manual" 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.]; 및 문헌[Gibson *et al.*, *Nature Methods* 6:343-45, 2009] 참조).

[0082] 따라서, 본 명세서에 제공된 바와 같은 핵산 분자는, 예를 들어, 이종성 핵산 서열에 작동 가능하게 연결된 조절 요소를 포함할 때 이종성 핵산 서열의 발현에 영향을 미칠 수 있는 발현 벡터로서 사용될 수 있다. 일부 실시 형태에서, 이종성 뉴클레오티드 서열은 관심 대상 유전자(GOI)의 코딩 서열을 포함한다. 일부 실시 형태에서, GOI의 코딩 서열은 기준 코딩 서열의 발현 수준보다 높은 수준의 발현을 위해 최적화된다. 일부 실시 형태에서, 기준 코딩 서열은 최적화되지 않은 서열이다. 일부 실시 형태에서, GOI 코딩 서열의 최적화는 서열 최적화를 포함할 수 있다. 뉴클레오티드 서열의 서열 최적화와 관련하여, 유전자 코드의 축퇴성은 유전자로부터 생성된 폴리펩티드의 아미노산 서열이 변화하지 않게 하면서 유전자의 단백질 인코딩 서열의 적어도 하나의 염기를 상이한 염기로 치환할 가능성을 제공한다. 따라서, 본 출원의 핵산 분자는 유전자 코드의 축퇴성에 따른 치환에 의해 본 명세서에 개시된 임의의 폴리뉴클레오티드 서열로부터 변화된 임의의 염기 서열을 또한 가질 수 있다. 코돈 사용을 설명하는 참고 문헌은 용이하게 공개적으로 이용 가능하다. 일부 실시 형태에서, 폴리뉴클레오티드 서열 변이체는 다양한 이유로, 예를 들어, (알파 바이러스 mRNA에서의 코돈 사용을 인간, 햄스터, 마우스, 또는 원숭이와 같은 다른 유기체에 의해 바람직한 것으로 변화시키는 것과 같이) 특정 숙주에 대해 발현을 최적화하기 위해 생성될 수 있다.

[0083] GOI에 의해 인코딩되는 폴리펩티드는 일반적으로 임의의 폴리펩티드일 수 있고, 예를 들어 치료용 폴리펩티드, 예방용 폴리펩티드, 진단용 폴리펩티드, 뉴트라슈티컬 폴리펩티드, 또는 산업용 효소일 수 있다. 일부 실시 형태에서, GOI는 항체, 항원, 면역 조절인자, 및 사이토카인으로 이루어진 군으로부터 선택되는 폴리펩티드를 인코딩한다.

[0084] 일부 실시 형태에서, GOI의 코딩 서열은 원하는 특성을 위해 최적화된다. 일부 실시 형태에서, GOI의 코딩 서

열은 기준 코딩 서열의 발현 수준보다 높은 수준의 발현을 위해 최적화된다.

[0085] 일부 실시 형태에서, 본 명세서에 개시된 변형된 게놈 또는 레플리콘 RNA는 알파바이러스의 게놈 또는 레플리콘 RNA, 예컨대 토가비리대 과의 알파바이러스 속에 속하는 바이러스 종의 게놈 또는 레플리콘 RNA이다. 일부 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 VEEV/EEEV 군, 또는 SF 군, 또는 SIN 군에 속하는 알파바이러스의 것이다(검토를 위해, 예를 들어 문헌[Strauss and Strauss. *Microbiol. Rev.* 58:3 p 492-562, 1994]을 참조함). SF 군의 알파바이러스의 비제한적인 예로는 셈리키 포레스트 바이러스, 오니옹-니옹 바이러스, 로스 리버 바이러스, 미들버그 바이러스, 치쿤구니아 바이러스, 바마 포레스트 바이러스, 게타 바이러스, 마야로 바이러스, 사기야마 바이러스, 베바루 바이러스, 및 우나 바이러스가 포함된다. SIN 군의 알파바이러스의 비제한적인 예로는 신드비스 바이러스, 거드우드 에스.에이. 바이러스(Girdwood S.A. virus), 사우스 아프리카 아보바이러스 No. 86, 오켈보 바이러스(Ockelbo virus), 오라 바이러스, 바반키 바이러스, 와타로아 바이러스, 및 키질라가크 바이러스가 포함된다. VEEV/EEEV 군의 알파바이러스의 비제한적 예로는 동부 말 뇌염 바이러스(EEEV), 베네수엘라 말 뇌염 바이러스(VEEV), 에버글레이즈 바이러스(EVEV), 무캄보 바이러스(MUCV), 핵수나 바이러스(PIXV), 미들버그 바이러스(MIDV), 치쿤구니아 바이러스(CHIKV), 오니옹-니옹 바이러스(ONNV), 로스 리버 바이러스(RRV), 바마 포레스트 바이러스(BF), 게타 바이러스(GET), 사기야마 바이러스(SAGV), 베바루 바이러스(BEBV), 마야로 바이러스(MAYV), 우나 바이러스(UNAV), 신드비스 바이러스(SINV), 오라 바이러스(AURAV), 와타로아 바이러스(WHAV), 바반키 바이러스(BABV), 키질라가크 바이러스(KYZV), 서부 말 뇌염 바이러스(WEEV), 하이랜드 제이 바이러스(HJV), 포트 모건 바이러스(FMV), 두무(NDUV), 및 부기 크릭 바이러스가 포함된다. 병독성 및 무병독성 알파바이러스 군주 둘 모두가 적합하다. 일부 특정 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 신드비스 바이러스(SIN), 셈리키 포레스트 바이러스(SFV), 로스 리버 바이러스(RRV), 베네수엘라 말 뇌염 바이러스(VEEV), 또는 동부 말 뇌염 바이러스(EEEV)의 것이다. 일부 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 베네수엘라 말 뇌염 바이러스(VEEV)의 것이다.

#### [0086] 재조합 세포

알파바이러스 종의 비제한적인 예로는 동부 말 뇌염 바이러스(EEEV), 베네수엘라 말 뇌염 바이러스(VEEV), 에버글레이즈 바이러스(EVEV), 무캄보 바이러스(MUCV), 핵수나 바이러스(PIXV), 미들버그 바이러스(MIDV), 치쿤구니아 바이러스(CHIKV), 오니옹-니옹 바이러스(ONNV), 로스 리버 바이러스(RRV), 바마 포레스트 바이러스(BF), 게타 바이러스(GET), 사기야마 바이러스(SAGV), 베바루 바이러스(BEBV), 마야로 바이러스(MAYV), 우나 바이러스(UNAV), 신드비스 바이러스(SINV), 오라 바이러스(AURAV), 와타로아 바이러스(WHAV), 바반키 바이러스(BABV), 키질라가크 바이러스(KYZV), 서부 말 뇌염 바이러스(WEEV), 하이랜드 제이 바이러스(HJV), 포트 모건 바이러스(FMV), 두무(NDUV), 및 부기 크릭 바이러스가 포함된다. 병독성 및 무병독성 알파바이러스 군주 둘 모두가 적합하다. 일부 특정 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 신드비스 바이러스(SIN), 셈리키 포레스트 바이러스(SFV), 로스 리버 바이러스(RRV), 베네수엘라 말 뇌염 바이러스(VEEV), 또는 동부 말 뇌염 바이러스(EEEV)의 것이다. 일부 실시 형태에서, 변형된 게놈 또는 레플리콘 RNA는 베네수엘라 말 뇌염 바이러스(VEEV)의 것이다.

#### [0087] 재조합 세포

[0088] 일 태양에서, 본 명세서에 개시된 일부 실시 형태는 동물 세포와 같은 숙주 세포 내로 본 명세서에 제공된 바와 같은 핵산 분자를 도입하는 단계, 및 형질전환된 세포를 선택하거나 스크리닝하는 단계를 포함하는, 세포를 형질전환시키는 방법에 관한 것이다. 용어 "숙주 세포" 및 "재조합 숙주 세포"는 본 명세서에서 상호 교환 가능하게 사용된다. 그러한 용어들은 특정 대상 세포뿐만 아니라 그러한 세포의 자손 또는 잠재적 자손도 지칭하는 것으로 이해된다. 돌연변이 또는 환경적 영향으로 인해 후속 세대에서 특정 변형이 일어날 수 있기 때문에, 그러한 자손은 실제로 모 세포와 동일하지 않을 수 있지만, 본 명세서에 사용되는 바와 같은 용어의 범주 내에 여전히 포함된다. 일부 실시 형태에서, 핵산 분자는 전기천공 절차 또는 바이오리스틱(biolistic) 절차에 의해 숙주 세포 내로 도입된다.

#### [0089] 관련 태양

관련 태양에서, 일부 실시 형태는 재조합 숙주 세포, 예를 들어 본 명세서에 기재된 핵산 분자를 포함하는 재조합 동물 세포에 관한 것이다. 핵산 분자는 숙주 게놈에 안정하게 통합될 수 있거나, 에피솜으로 복제될 수 있거나, 안정한 또는 일시적 발현을 위한 미니-서클(mini-circle) 발현 벡터로서 재조합 숙주 세포에 존재할 수 있다. 따라서, 본 명세서에 개시된 일부 실시 형태에서, 핵산 분자는 에피솜 유닛으로서 재조합 숙주 세포에서 유지 및 복제된다. 일부 실시 형태에서, 핵산 분자는 재조합 세포의 게놈 내로 안정하게 통합된다. 안정한 통합은 고전적인 무작위 게놈 재조합 기법을 이용하여 또는 가이드 RNA 지시 CRISPR/Cas9, 또는 DNA 가이드 엔도뉴클레아제 게놈 편집 NgAgo(*Natronobacterium gregoryi* Argonaute), 또는 TALEN 게놈 편집(전사 활성화인자 유사 이펙터 뉴클레아제)을 사용하는 것과 같은 더욱 정확한 게놈 편집 기법으로 완료될 수 있다. 일부 실시 형태에서, 핵산 분자는 안정한 또는 일시적 발현을 위한 미니-서클 발현 벡터로서 재조합 숙주 세포에 존재한다.

#### [0090] 일부 실시 형태

일부 실시 형태에서, 숙주 세포는, 예를 들어, 본 출원의 벡터 작제물로 유전자 조작(예를 들어, 형질도입 또는 형질전환 또는 형질감염)될 수 있으며, 이 벡터 작제물은 예를 들어 숙주 세포의 게놈의 일부에 상동인 핵산 서열을 포함하는 상동 재조합을 위한 벡터일 수 있거나 관심 대상 유전자들 중 임의의 것 또는 이들의 조합의 발

현을 위한 발현 벡터일 수 있다. 벡터는, 예를 들어,プラス미드, 바이러스 입자, 파지 등의 형태일 수 있다. 일부 실시 형태에서, 관심 대상 폴리펩티드의 발현을 위한 벡터는, 예를 들어, 상동 재조합에 의한 숙주 내로의 통합을 위해 또한 설계될 수 있다. 본 명세서에 기재된 바와 같은 폴리뉴클레오티드 서열을 함유하는 벡터, 예를 들어 변형된 알파바이러스 게놈 또는 레플리콘 RNA뿐만 아니라, 선택적으로, 선택 가능한 마커 또는 리포터 유전자를 포함하는 핵산 분자가 사용되어 적절한 숙주 세포를 형질전환시킬 수 있다.

[0091] 본 명세서에 개시된 방법 및 조성물은 원핵 생물 및 진핵 생물 종을 포함하지만 이로 한정되지 않는 임의의 종의 유전자 조작을 위해 활용될 수 있다. 본 발명에 따른 조성물 및 방법을 사용하여 변형시키려는 적합한 숙주 세포는 조류(algal) 세포, 세균 세포, 헤테로콘트(heterokont), 진균 세포, 키트리드(chytrid) 세포, 미세진균(microfungi), 미세조류(microalgae), 및 동물 세포를 포함할 수 있지만 이로 한정되지 않는다. 일부 실시 형태에서, 동물 세포는 무척추 동물 세포이다. 일부 실시 형태에서, 척추 동물 세포는 포유류 세포이다. 숙주 세포는 형질전환되지 않은 세포 또는 적어도 하나의 핵산 분자로 이미 형질감염된 세포일 수 있다.

[0092] 본 명세서에 개시된 방법 및 조성물은, 예를 들어, 수경 재배, 농업, 축산업을 위해 및/또는 백신, 약제학적 제품, 산업용 제품, 화학 물질 등의 제조에 사용되는 폴리펩티드의 생성을 포함하는 치료적 및 의학적 적용을 위해 중요하거나 관심을 끄는 대상체 및/또는 숙주 세포와 함께 사용될 수 있다. 일부 실시 형태에서, 본 명세서에 개시된 조성물 및 방법은 알파바이러스의 천연 숙주, 예컨대 설치류, 마우스, 어류, 조류, 및 대형 포유 동물, 예컨대 인간, 말, 피그, 원숭이, 및 유인원뿐만 아니라 무척추 동물인 종으로부터의 숙주 세포와 함께 사용될 수 있다. 특히 바람직한 종은, 본 출원의 일부 실시 형태에서, 척추 동물 종 및 무척추 동물 종이다. 원칙적으로, 임의의 동물 종이 일반적으로 사용될 수 있고, 예를 들어 인간, 개, 조류, 어류, 말, 피그, 영장류, 마우스, 소, 돼지, 양, 토끼, 고양이, 염소, 당나귀, 햄스터, 또는 벼팔로일 수 있다. 적합한 조류 종의 비제한적인 예로는 닭, 오리, 거위, 칠면조, 타조, 애뮤, 백조, 공작, 꿩, 자고새(partridge), 및 뿔닭(guinea fowl)이 포함된다. 일부 특정 실시 형태에서, 어류 종은 연어 종이다. 1차 포유류 세포 및 연속/불멸화 세포 유형이 또한 적합하다. 적합한 동물 숙주 세포의 비제한적인 예로는 폐 말 동맥 내피 세포, 말 진피 세포, 베이비 햄스터 신장(BHK) 세포, 토끼 신장 세포, 마우스 근육 세포, 마우스 결합 조직 세포, 인간 자궁경부 세포, 인간 표피양 후두 세포, 차이니즈 햄스터 난소 세포(CHO), 인간 HEK-293 세포, 마우스 3T3 세포, Vero 세포, 마린-다비 개 신장 상피 세포(MDCK), 1차 닭 섬유모세포, HuT78 세포, A549 폐 세포, HeLa 세포, PER.C6® 세포, WI-38 세포, MRC-5 세포, FRhL-2, 및 CEM T-세포가 포함되지만 이로 한정되지 않는다. 일부 실시 형태에서, 숙주 세포는 베이비 햄스터 신장 세포이다. 일부 실시 형태에서, 베이비 햄스터 신장 세포는 BHK-21 세포이다.

[0093] 매우 다양한 상기 언급된 숙주 세포 및 종을 형질전환시키는 기법은 당업계에 공지되어 있고, 기술 및 학술 문헌에 기재되어 있다. 따라서, 본 명세서에 개시된 바와 같은 적어도 하나의 재조합 세포를 포함하는 세포 배양물이 또한 본 출원의 범주 내에 있다. 세포 배양물을 생성하고 유지하기에 적합한 방법 및 시스템은 당업계에 공지되어 있다.

#### 폴리펩티드를 생성하는 방법

[0095] 원핵 또는 진핵 숙주 세포와 같은 본 발명의 숙주 세포는, 예를 들어, 본 명세서에 개시된 바와 같은 관심 대상 유전자(GOI)의 개방 해독 프레임에서 인코딩되는 폴리펩티드와 같은 관심 대상 분자를 생성(즉, 발현)하는 데 사용될 수 있다. 따라서, 본 출원은, 예를 들어 세포 배양물 내의 세포 또는 생체 내의 세포일 수 있는 본 발명의 숙주 세포를 사용하여, 예를 들어 폴리펩티드와 같은 관심 대상 분자를 생성하는 방법을 추가로 제공한다.

[0096] 따라서, 본 명세서에 개시된 일부 실시 형태는 숙주 세포에서 관심 대상 폴리펩티드를 생성하는 방법을 제공한다. 그러한 방법은 전술한 태양들 및 실시 형태들 중 어느 하나에 따른 핵산 분자를 포함하는 재조합 숙주 세포의 배양을 포함한다. 일부 실시 형태에서, 방법은 관심 대상 분자가 생성되도록 적합한 배지에서 (관심 대상 분자를 인코딩하는 재조합 발현 벡터가 도입된) 본 발명의 숙주 세포를 배양하는 단계를 포함한다. 일부 실시 형태에서, 방법은 배지 또는 숙주 세포로부터 관심 대상 분자를 단리하는 단계를 추가로 포함한다.

[0097] 다른 태양에서, 일부 실시 형태는 전술한 태양들 및 실시 형태들 중 어느 하나에 따른 핵산 분자를 대상체에게 투여하는 단계를 포함하는, 대상체에서 관심 대상 폴리펩티드를 생성하는 방법에 관한 것이다.

[0098] 본 명세서에 개시된 방법 및 조성물에 사용하기 위한 적합한 숙주 세포 및/또는 대상체로는 원핵 생물 및 진핵 생물 종이 포함되지만 이로 한정되지 않는다. 본 발명에 따른 조성물 및 방법을 사용하여 변형시키려는 적합한 숙주 세포는 조류 세포, 세균 세포, 헤테로콘트, 진균 세포, 키트리드 세포, 미세진균, 미세조류, 및 동물 세포를 포함할 수 있지만 이로 한정되지 않는다. 일부 실시 형태에서, 동물 세포는 무척추 동물 세포이다. 일부

실시 형태에서, 척추 동물 세포는 포유류 세포이다. 숙주 세포는 형질전환되지 않은 세포 또는 적어도 하나의 핵산 분자로 이미 형질감염된 세포일 수 있다. 따라서, 전술한 태양들 및 실시 형태들 중 어느 하나에 따른 재조합 세포의 생물학적 샘플, 바이오매스, 및 자손도 또한 본 출원의 범주 내에 있다. 따라서, 하기에 더욱 상세히 논의되는 바와 같이, 본 출원의 이러한 태양에 따른 방법에 의해 생성된 폴리펩티드는 또한 본 출원의 범주 내에 있다.

[0099]

일부 실시 형태에서, 재조합 세포는 동물 세포이다. 소규모 및 대규모로의 치료용 단백질 생성은 제약 산업에서 중요한 개발 분야인데, 그 이유는 동물 세포에서 생성된 단백질이 일반적으로 적절한 처리, 번역후 변형을 가지며, 이에 따라 생리학적 질환의 치료를 위한 적절한 활성을 갖는 것으로 믿어지기 때문이다. 원칙적으로, 임의의 동물 종이 일반적으로 사용될 수 있고, 예를 들어 인간, 개, 조류, 어류, 말, 퍼그, 영장류, 마우스, 소, 돼지, 양, 토끼, 고양이, 염소, 당나귀, 햄스터, 또는 벼팔로일 수 있다. 적합한 조류 종의 비제한적인 예로는 닭, 오리, 거위, 칠면조, 타조, 애류, 백조, 공작, 꿩, 자고새, 및 뿔닭이 포함된다. 일부 특정 실시 형태에서, 어류 종은 연어 종이다. 1차 포유류 세포 및 연속/불멸화 세포 유형이 또한 적합하다. 적합한 동물 숙주 세포의 비제한적인 예로는 폐 말 동맥 내피 세포, 말 진피 세포, 베이비 햄스터 신장(BHK) 세포, 토끼 신장 세포, 마우스 근육 세포, 마우스 결합 조직 세포, 인간 자궁경부 세포, 인간 표피양 후두 세포, 차이니즈 햄스터 난소 세포(CHO), 인간 HEK-293 세포, 마우스 3T3 세포, Vero 세포, 마틴-다비 개 신장 상피 세포(MDCK), 1차 닭 섬유모세포, HuT78 세포, A549 폐 세포, HeLa 세포, PER.C6® 세포, WI-38 세포, MRC-5 세포, FRhL-2, 및 CEM T-세포가 포함되지만 이로 한정되지 않는다. 일부 실시 형태에서, 숙주 세포는 베이비 햄스터 신장 세포이다. 일부 실시 형태에서, 베이비 햄스터 신장 세포는 BHK-21 세포이다.

[0100]

### 재조합 폴리펩티드

[0101]

본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 하나 이상의 실시 형태에 따른 방법에 의해 생성된 재조합 폴리펩티드에 관한 것이다. 본 출원의 재조합 폴리펩티드는 일반적으로 임의의 재조합 폴리펩티드일 수 있고, 예를 들어 치료용 폴리펩티드, 예방용 폴리펩티드, 진단용 폴리펩티드, 뉴트라슈티컬 폴리펩티드, 산업용 효소, 및 리포터 폴리펩티드 중 하나 이상일 수 있다. 일부 실시 형태에서, 재조합 폴리펩티드는 항체, 항원, 면역 조절인자, 및 사이토카인 중 하나 이상일 수 있다. 일부 실시 형태에서, 관심 대상 폴리펩티드는 치료적 또는 예방적 활성을 가질 수 있다.

[0102]

### 조성물

[0103]

본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 임의의 재조합 폴리펩티드를 포함하는 조성물에 관한 것이다. 조성물은, 예를 들어, 뉴트라슈티컬 조성물, 예방적 조성물, 약제학적으로 허용 가능한 담체를 포함하는 약제학적 조성물, 또는 이들의 혼합물일 수 있다. 일부 실시 형태에서, 본 출원의 조성물은 백신으로서 사용될 수 있다.

[0104]

본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 임의의 핵산 분자를 포함하는 조성물에 관한 것이다. 조성물은, 예를 들어, 뉴트라슈티컬 조성물, 예방적 조성물, 약제학적으로 허용 가능한 담체를 포함하는 약제학적 조성물, 또는 이들의 혼합물일 수 있다. 일부 실시 형태에서, 본 출원의 조성물은 백신으로서 사용될 수 있다.

[0105]

본 명세서에 개시된 일부 실시 형태는 본 명세서에 기재된 임의의 재조합 세포를 포함하는 조성물에 관한 것이다. 조성물은, 예를 들어, 뉴트라슈티컬 조성물, 예방적 조성물, 약제학적으로 허용 가능한 담체를 포함하는 약제학적 조성물, 또는 이들의 혼합물일 수 있다. 일부 실시 형태에서, 본 출원의 조성물은 백신으로서 사용될 수 있다.

[0106]

본 명세서에 사용되는 바와 같이, 용어 "약학적으로 허용 가능한 담체"는 일반적으로 안전하고, 비독성이고, 생물학적으로 또는 달리 바람직하지 못한 것이 아닌 약제학적 조성물 또는 제형을 제조하는 데 유용한 담체를 의미하고, 수의학적 용도뿐만 아니라 인간용 약제학적 용도를 위해 허용 가능한 담체를 포함한다. 일부 실시 형태에서, 약제학적으로 허용 가능한 담체는 물처럼 간단하지만, 이는 또한 예를 들어 생리학적 염 농도의 용액을 포함할 수 있다. 일부 실시 형태에서, 약제학적으로 허용 가능한 담체는 안정제, 희석제 및 완충제일 수 있거나, 이를 포함할 수 있다. 적합한 안정제는 예를 들어 SPGA, 탄수화물(예컨대, 분유, 혈청 알부민 또는 카세인) 또는 이들의 분해 생성물이다. 적합한 완충제는 예를 들어 알칼리 금속 포스페이트이다. 희석제는 물, 수성 완충제(예컨대, 완충 식염수), 알코올 및 폴리올(예컨대, 글리세롤)을 포함한다. 동물 또는 사람에게 투여하기 위하여, 본 출원에 따른 조성물은 특히 비강내로, 분무에 의해, 피내로, 피하로, 경구로, 에어로졸에

의해 또는 근육내로 주어질 수 있다.

[0107] 본 발명에 언급된 모든 간행물 및 특허 출원은 각각의 개별 간행물 또는 특허 출원이 구체적으로 그리고 개별적으로 참고로 포함되도록 표시되어 있는 것처럼 그와 동일한 정도로 본 명세서에 참고로 포함된다.

[0108] 본 명세서에 인용된 임의의 참고 문헌은 종래 기술을 구성하는 것으로 인정되지 않는다. 참고 문헌의 논의는 그 저자들이 주장한 바를 명시하고, 본 발명자들은 인용된 문헌의 정확성 및 적절성에 이의를 제기할 권리를 유보한다. 학술 저널 소논문, 특히 문헌, 및 교재를 비롯한 다수의 정보 출처가 본 명세서에서 언급되지만, 이러한 참조는 이들 문헌 중 임의의 것이 당업계에서의 통상적인 일반 지식의 일부를 형성한다는 인정을 구성하는 것이 아님이 명확하게 이해될 것이다.

[0109] 본 명세서에 주어진 일반적인 방법의 논의는 단지 예시적인 목적을 위해 의도된다. 다른 대안적인 방법 및 대안은 본 발명의 검토시 당업자에게 명백할 것이고, 본 출원의 사상 및 범위 내에 포함되어야 한다.

[0110] 실시예

[0111] 추가의 대안이 하기의 실시예에서 더욱 상세히 개시되며, 이는 어떠한 방식으로든 청구범위의 범주를 제한하고자 하는 것이 아니다.

[0112] 실시예 1

[0113] 일반적 실험 절차

[0114] 시험관내 전사

[0115] 시험관내 전사(IVT) 반응은 37°C에서의 1 시간 인큐베이션에 걸쳐 20  $\mu$ l의 반응물로 상기에 기재된 바와 같이 제조된 1  $\mu$ g의 DNA 주형을 사용하여 수행하였다(NEB 카탈로그 번호 E2065S). 이어서, 공급업체에 의해 제공된 1 유닛의 DNase I을 IVT 반응에 직접 첨가하고, 37°C에서 추가로 15 분 동안 인큐베이션하였다. 이어서, 반응물을 얼음 상에 놓고, 제조업체의 제안된 방법(Qiagen 카탈로그 번호 74104)을 사용하여 정제하였다. 이어서, 정제된 RNA를 NanoDrop 2000c UV-Vis 분광광도계를 사용하여 정량하였다. RNA를 0.8% 아가로스 겔(Life Technologies 카탈로그 번호 G5018-08)을 통해 전기영동에 의해 가시화하고, 전기천공을 진행하기 전에 밀레니엄 RNA 마커(Millennium RNA Marker)(Ambion 카탈로그 번호 AM7150)와 비교하였다.

[0116] 50 ng/ml의 카르바미실린(carbamicilin)(Teknova 카탈로그 번호 NC9730116)이 보충된 LB 브로쓰 배지(Teknova 카탈로그 번호: L8000 06)에서 성장시킨 300 mL의 포화 대장균 트랜스포맥스(TransformMax) Epi300(Epicentre 카탈로그 번호 EC300105) 배양물로부터 플라스미드 DNA 주형을 정제하였다(Qiagen 카탈로그 번호 12163). 플라스미드 DNA를 37°C에서 1 시간 동안 Not-I 분해(New England Biolabs NEB 카탈로그 번호 R3189S)에 의해 선형화하였다. 이어서, 선형화된 주형 DNA를 재정제하고(Zymo 카탈로그 번호 D4003), 시판되는 2-log DNA 래더(ladder)(New England Biolabs, NEB 카탈로그 번호 N3200S)에 대해 0.8% 아가로스 겔(Life Technologies 카탈로그 번호 G5018-08)에 의해 분석하였다. 단일 랜드의 존재는, 시험관내 전사를 진행하기 전에, 선형 DNA 주형의 예상 단편 크기에 상응하는 각각의 샘플에서 확인하였다.

[0117] 형질감염 및 분석

[0118] 전형적인 세포 형질감염 실험에서, 레플리콘 RNA를 4D-뉴클레오팩터 시스템(4D-Nucleofector™ System)(Lonza)을 위한 SF 셀라인 뉴클레오팩터 키트(SF Cell Line Nucleofector™ kit)를 사용한 전기천공에 의해 BHK-21 세포 내로 도입하였다. 0.25% 트립신을 사용하여 BHK-21 세포를 수거하고, 냉 PBS로 1 회 세척하였다. 세포를 20  $\mu$ L의 전기천공 반응물당 1  $\times$  10<sup>6</sup>개 세포의 세포 밀도로 SF 버퍼(Buffer) 중에 재현탁시켰다. 3 마이크로그램의 RNA를 16-웰 큐벳 스트립(16-well cuvette strip)에서 3 중으로 세포 내로 전기천공하고, 실온에서 10 분 동안 인큐베이션하였다. 전기천공된 세포를 10% 소 태아 혈청을 함유하는 둘베코 변형 이글 배지(Dulbecco's Modified Eagle Medium)를 함유하는 플레이트 내로 회수한 후, 표준 세포 배양 조건에서 16 시간 내지 18 시간 동안 인큐베이션하였다.

[0119] 레플리콘 형질감염 효율 및 단백질 생성의 세포내 분석을 유세포 측정에 의해 수행하였다. 형질감염된 BHK-21 세포를 고정시키고, fix/perm 농축물 및 투과 완충제(eBioscience)를 사용하여 투과시켰다. 세포를 R-피코에리트린(R-Phycoerythrin)(Innova Biosciences)과 접합된 이중 가닥 RNA 생성(J2 항-dsRNA IgG2A 모노클로날 항체, English & Scientific Company)용 항체와 함께 인큐베이션하였다. 항원 생성은 PE-Cy5(Innova Biosciences)와 접합된 항원 특이적 항체(예를 들어, 그린 레닐라(green Renilla), 레드 파이어플라이, HA 또는

RSV-F0에 대한 항체(Abcam))와의 추가 인큐베이션에 의해 평가하였다. 이어서, 세포를 1 회 세척하고, FACS Aria™ 퓨전 셀 소터(Fusion Cell Sorter)(BD Biosciences) 또는 FACS Aria™ II 셀 소터(BD Biosciences)를 사용하여 분석하였다. 보상 제어를 위해 단일 색으로 염색된 형질감염된 BHK-21 세포를 샘플 수집 전에 러닝(running)시켰다. 데이터를 FACSDiva(BD Biosciences)를 사용하여 수집하고, FlowJo 소프트웨어를 사용하여 추가로 분석하였다. 초기 게이팅(gating)을 수행하여, 전방 및 측면 산란 플롯들을 사용하여 죽은 세포 및 잔해물을 배제시켰다. 추가의 게이팅을 실행하여 dsRNA(R-PE 양성) 및 단백질 발현(GFP 발현에 대하여 PE-Cy5 양성 또는 FITC 양성) 둘 모두에 대해 양성인 세포 집단을 확인하였다. 빈도 및 평균 형광 강도를 수집하고, 각 제물 비교 및 최적화를 위해 이용하였다.

[0120] 실시예 2

[0121] 5'-UTR 서열 변형

[0122] 본 실시예는 VEEV 레플리콘 RNA 벡터에서 인코딩된 관심 대상 유전자의 발현을 향상시키기 위해 5'-UTR 서열을 변형시킨 실험의 결과를 설명한다. 부위 특이적 돌연변이유발(SDM)을 VEEV 레플리콘 벡터의 cDNA 카피를 함유하는 플라스미드 상에서 수행하였다. SDM 프라이머는 5' UTR의 위치 2에서 티민(T) 잔기를 G 잔기로 변화시키도록 설계하였다.

[0123] 이들 실험에서, 각각 rFF 루시퍼라아제 유전자, A Vietnam 1203 인플루엔자 헤마글루티닌 유전자(HA) 또는 녹색 형광 단백질(GFP) 유전자를 각각 발현하는 VEEV 레플리콘을 부위 특이적 돌연변이유발을 수행하기 위한 주형으로서 사용하였다. 야생형 VEEV(균주 TC-83) 서열의 5' UTR에서 위치 2에서의 뉴클레오티드 T를 G로 변화시켰다. 프라이머 "VEE 5' T→G nt 2 F" 및 "VEE 5' T→G nt 2 R"을 사용하여 Agilent로부터의 부위 특이적 돌연변이유발 키트를 통해 뉴클레오티드 2 변화를 도입하였다. 프라이머를 Agilent의 웹사이트를 통해 설계하였다. 양성 클론을 확인하고, rFF, HA 및 GFP를 발현하는 대표적인 클론을 서열 확인하여 완전히 정확하게 되도록 하였다.

[0124] Alpha-R-T2G 레플리콘을 제조하기 위한 SDM 프라이머는 다음과 같다.

[0125] 정방향 프라이머: VEE 5' T→G nt 2 F(서열 번호 23): cgactcactatagaGaggcgccatgag. 역방향 프라이머: VEE 5' T→G nt 2 R(서열 번호 24): ctcatgcgcgcctCtctatagtgagtgc.

[0126] VEEV 레플리콘 cDNA에서의 T2→G 변화의 서열 확인 후에, 선형화된 플라스미드 DNA 상에서의 T7 RNA 중합효소를 사용한 시험관내 전사에 의해 RNA를 생성하였다. 시험관내 전사된 RNA를 정제하고, 이를 사용하여 BHK-21 세포를 전기천공하였다. 복제 및 GOI 발현 둘 모두를, 각각 항-dsRNA 특이적 항체 및 GOI 특이적 항체를 사용하여 FACS에 의해 모니터링하였다. U2→G VEEV 레플리콘으로부터의 복제 효율 및 GOI 발현을 동일한 GOI를 발현하는 야생형 VEEV 레플리콘과 직접 비교하였다. 레드 파이어플라이 리포터 유전자를 포함하는 Alpha-R-rFF-T2G 레플리콘의 서열은 서열 목록 중의 T7 프로모터 및 40 A 잔기를 갖는 폴리A 테일을 갖는 서열 번호 19로서 제공된다. T7 프로모터 서열 다음의 위치 2에서의 돌연변이된 뉴클레오티드가 또한 표시된다.

[0127] 실시예 3

[0128] 5' UTR에서의 U2→G 치환은 변형된 VEEV-HA 레플리콘의 생물학적 활성에 영향을 주지 않음

[0129] 본 실시예는 A Vietnam 1203 인플루엔자 헤마글루티닌 유전자(HA) 리포터 유전자의 발현에 대한 변형된 알파바이러스 레플리콘의 5' UTR에서의 U2→G 치환의 영향을 평가하는 실험의 결과를 설명한다.

[0130] 5' UTR에서 U2→G 변화를 함유하는 알파바이러스 레플리콘이 실제로 단백질을 발현할 수 있음을 입증하기 위하여, 레플리콘 RNA를 5' UTR에서 U2→G 변화(U2→G VEEV-HA)를 지닌 벡터로부터 시험관내에서 전사시켰다. 발현에 대한 양성 대조군으로서, 야생형 5' UTR(WT VEEV-HA)을 지닌 벡터로부터 RNA를 시험관내에서 전사시켰다. 베이비 햄스터 신장(BHK-21) 세포를 3  $\mu$ g의 U2→G VEEV-HA RNA 또는 야생형 VEEV-HA RNA로 전기천공하였다. 이러한 비교를 위한 유세포 측정 분석의 예가 도 1에 도시되어 있다. 세포를 HA 특이적 항체를 사용하여 FACS에 의해 분석하여, 발현된 HA 단백질의 존재 및 세포당 기준으로 표현되는 HA의 상대적인 양(평균 형광 강도 - MFI) 둘 모두를 나타내었다. 야생형 5' UTR을 갖는 레플리콘으로부터 검출된 활성에 비해 U2→G 5' UTR 변화를 지닌 레플리콘 RNA로부터의 생물학적 활성의 손실은 없다.

[0131] 본 실시예에서 제시된 실험 데이터는 U2→G VEEV-HA 레플리콘이 HA 단백질을 발현할 수 있을뿐만 아니라 발현 수준이 야생형 VEEV-HA 레플리콘과 동등하다는 것을 나타낸다. 이러한 결과는 동일한 U2→G 5' UTR 변화를 지닌 VEEV/SINV 감염성 클론이 거의 3 로그의 생물학적 활성의 손실을 가졌음을 고려할 때 예상치 못한 것인데;

본 발명자들은 U2→G 5' UTR 변화를 지닌 레플리콘의 경우 생물학적 활성의 감소가 전혀 없음을 본 명세서에서 보여주고 있다.

[0132] 실시예 4

[0133] 5' UTR에서의 U2→G 치환은 GFP 리포터의 발현을 3배 향상시킴

[0134] 본 실시예는 녹색 형광 단백질(GFP) 리포터 유전자의 발현에 대한 변형된 알파바이러스 레플리콘의 5' UTR에서의 U2→G 치환의 영향을 평가하는 실험의 결과를 설명한다.

[0135] 5' UTR에서 U2→G 변화를 함유하는 알파바이러스 레플리콘으로부터의 발현이 임의의 특정 GOI로 제한되지 않는다는 것을 입증하기 위하여, GFP 유전자를 발현하는 레플리콘 벡터를 유사한 방식으로 비교하였다. 레플리콘 RNA를 5' UTR에서 U2→G 변화(U2→G VEEV-GFP)를 지닌 벡터로부터 시험관내에서 전사시켰다. 발현에 대한 양성 대조군으로서, 야생형 5' UTR(WT VEEV-GFP)을 지닌 벡터로부터 RNA를 시험관내에서 전사시켰다. BHK 세포를 3  $\mu$ g의 U2→G VEEV-GFP RNA 또는 야생형 VEEV-GFP RNA로 전기천공하였다. 이러한 비교를 위한 유세포 측정 분석의 예가 도 2에 도시되어 있다. 세포를 FACS에 의해 GFP 발현에 대해 분석하여, 발현된 GFP 단백질의 존재 및 세포 당 기준으로 발현되는 GFP의 상대적인 양(평균 형광 강도 - MFI) 둘 모두를 나타내었다. U2→G 5' UTR을 지닌 레플리콘 RNA로부터의 생물학적 활성의 손실이 없을 뿐만 아니라, U2→G 5' UTR 변화는 야생형 5' UTR을 갖는 레플리콘으로부터 검출된 발현에 비해 발현을 실제로 3배 향상시켰다.

[0136] 다시 한번, 본 실시예에서 제시된 실험 데이터는 U2→G 5' UTR 변화를 지닌 레플리콘 벡터가 단백질(이 경우는 GFP)을 발현할 수 있음을 나타낸다. 아마도, 단순한 단백질 발현보다 더욱더 예기치 못한 것은 U2→G VEEV-GFP 레플리콘이 야생형 VEEV-GFP 레플리콘보다 GFP를 3배 더 많이 발현했다는 것이다. 한번 더, 5' UTR에서의 U2→G 변화로 인해 기대되는 생물학적 활성의 예상 손실은 실현되지 않았고, 이러한 결과는 5' UTR에서의 U2→G 변화가 GOI 레플리콘 발현을 실제로 유의하게 향상시킬 수 있음을 나타내었다.

[0137] 실시예 5

[0138] 5' UTR에서의 U2→G 치환은 rFF 리포터의 발현을 2배 향상시킴

[0139] 본 실시예는 레드 파이어플라이(rFF) 리포터 유전자의 발현에 대한 변형된 알파바이러스 레플리콘의 5' UTR에서의 U2→G 치환의 영향을 평가하는 실험의 결과를 설명한다.

[0140] 이러한 실험에서, 5' UTR에서 U2→G 변화를 함유하는 알파바이러스 레플리콘으로부터의 발현의 다른 예로서, rFF 유전자를 발현하는 레플리콘 벡터들을 유사한 방식으로 비교하였다. 레플리콘 RNA를 5' UTR에서 U2→G 변화(U2→G VEEV-rFF)를 지닌 벡터로부터 시험관내에서 전사시켰다. 발현에 대한 양성 대조군으로서, 야생형 5' UTR(WT VEEV-rFF)을 지닌 벡터로부터 RNA를 시험관내에서 전사시켰다. BHK 세포를 3  $\mu$ g의 U2→G VEEV-rFF RNA 또는 야생형 VEEV-rFF RNA로 전기천공하였다. 루시페라아제 단백질 발현의 예가 도 3에 도시되어 있다. 이 실험에서, 5' UTR에서 U2→G 변화(U2→G alpha rFF)를 지니도록 변형된 알파바이러스 레플리콘의 능력을 야생형 5' UTR를 갖는 알파바이러스 레플리콘(Alpha rFF)과 비교하였다. BHK 세포를 어느 하나의 레플리콘으로부터의 동등한 양의 시험관내 전사된 RNA로 전기천공하고, 이어서 세포를 rFF 루시페라아제 발현에 대해 분석하였다. 세포당 기준으로 표현되는(상대 광 단위(RLU)로 표현되는) 루시페라아제의 양이 제시된다. U2→G 5' UTR을 지닌 레플리콘 RNA로부터의 생물학적 활성의 손실이 없을 뿐만 아니라, U2→G 5' UTR 변화는 야생형 5' UTR을 갖는 레플리콘으로부터 검출된 발현에 비해 발현을 실제로 대략 2배 향상시켰다.

[0141] 다시 한번, 본 실시예에서 제시된 실험 데이터는 U2→G 5' UTR 변화를 지닌 레플리콘 벡터가 단백질(이 경우는 rFF)을 발현할 수 있음을 나타낸다. 아마도, 단순한 단백질 발현보다 더욱더 예기치 못한 것은 U2→G VEEV-rFF 레플리콘이 야생형 VEEV-rFF 레플리콘보다 rFF를 약 2배 더 많이 발현했다는 것이다. 한번 더, 5' UTR에서의 U2→G 변화로 인해 기대되는 생물학적 활성의 예상 손실은 실현되지 않았고, 이러한 결과는 5' UTR에서의 U2→G 변화가 GOI 레플리콘 발현을 실제로 유의하게 향상시킬 수 있음을 나타내었다.

[0142] 실시예 6

[0143] 다가 VEEV 레플리콘 설계

[0144] 본 실시예는 다가 VEEV 레플리콘을 제작하고 평가하기 위해 수행된 실험을 설명하며, 이러한 레플리콘은 후속적으로 재조합 세포에서의 적어도 2개의 상이한 폴리펩티드의 발현을 위해 활용된다. 일부 실험에서, 다가 VEEV 레플리콘은 5'에서 3' 순서로 (i) 비구조 단백질 매개 증폭에 필요한 5' 서열, (ii) VEEV 비구조 단백질인

nsP1, nsP2, nsP3, 및 nsP4를 인코딩하는 뉴클레오티드 서열, (iii) 각각이 이종성 핵산 서열에 작동 가능하게 연결된 적어도 2개의 프로모터로서, 이종성 핵산 서열은 VEEV 구조 단백질 유전자 중 하나 또는 전부를 대체하는, 적어도 2개의 프로모터, (iv) 비구조 단백질 매개 증폭에 필요한 3' 서열, 및 (v) 폴리아데닐레이트 트랙을 포함한다.

[0145]

본 발명의 특정 대안이 개시되었지만, 다양한 변형 및 조합이 첨부된 청구범위의 진정한 사상 및 범주 내에서 가능하고 고려된다는 것이 이해되어야 한다. 따라서, 본 명세서에 제시된 정확한 초록 및 개시 내용에 대한 제한의 의도는 없다.

## 도면

### 도면 1a

```

서열 번호: 2 (AURAV) ATAGCGGACGGACTAGTACTTGTACTACAGAATTAAAC-TGCCGTGTGCCG-----  

서열 번호: 3 (CHIKV) ATGGCTG-CGTGAGA-----CACACGTAGCCTACCAAGTTCTTACTGCTCTACTCT-----  

서열 번호: 4 (ONNV) ATAGCTG-CGTGATA-----CACACACGCAGCTTACGGGTTTCAACTGCTCTACT-----  

서열 번호: 5 (BEBV) ATGGCGGCTGTGTGA-CACACGAGCCGTGATTCAA-CCTTCTTGCTCCCT-----  

서열 번호: 6 (SFV) ATGGCGGATGTGTGA-CATACACGACGCCAAAGATT-TTGGTTCCAGCTCCT-----  

서열 번호: 7 (MAYV) ATGGCGGCAAGTGA-CACTTGTCCGCCGGTCGTCT-----CTAACGCTCTCCTC-----  

서열 번호: 8 (GETV) ATGGCGGACGTGTGACATCACCGTTCGCTCTTTCTAG-GATCCTTTGCTAC-----  

서열 번호: 9 (SAGV) ATGGCGGACGTGTGACATCACCGTTCGCTCTTTCTAG-GATCCTTTGCTAC-----  

서열 번호: 10 (NDUV) ATGG-TGCGGAGTT-GA---GAGACGA-AGCACCAA-ACAACCTACGCCGCTCACC-AT-----  

서열 번호: 11 (MIDV) ATGGGTGGTTACGTA---CACGTGCCACCACCCCCCA-CCCTCCAAGCGATCCA-----  

서열 번호: 12 (EEEV) ATAGGGTACGGTGT-----GAGGCAACCACCTAT-----TTCCACCTATCCAAAATGG-----  

서열 번호: 13 (FMV) ATAGGGTATGGTTA-----GAGGCCTACCCCTAC-----TTAACCGATCCAAAATGG-----  

서열 번호: 14 (Buggy) ATAGGGTATGGTTA-----GAGGCCTACCCCTAC-----TTAACCGATCCAAAATGG-----  

서열 번호: 15 (VEEV) ATGGCGGCGCAAGA-----GAGAAGCCAAACCAA-----TTACCTACCCAAAATGGAG---  

서열 번호: 16 (WHAV) ATGGCGGCGCATAGTA-----CATACTATATAAAAGAAAAGCCGACCAATTGCAC-  

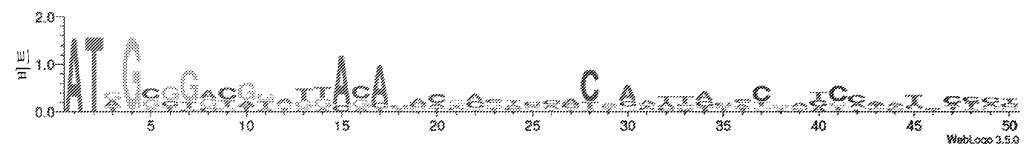
서열 번호: 17 (SINV) ATTGACGGCGTAGTA-----CACACTAT-TGAATCAAACAGCCGACCAATTGCACT  

서열 번호: 18 (BABV) ATTGGCGGCGTAGTA-----CACACTAT-TGAATCAAACAGCCGACCAATTGCACT  

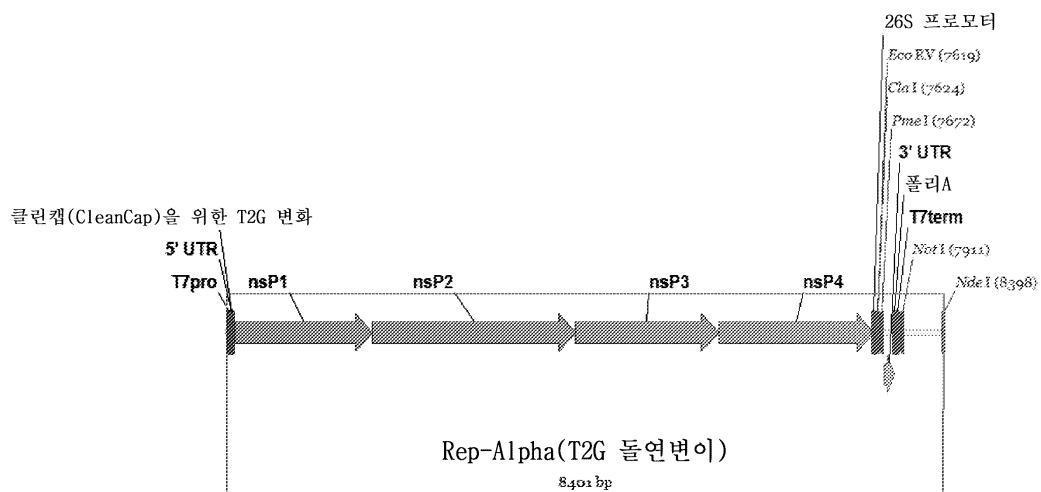
*** *

```

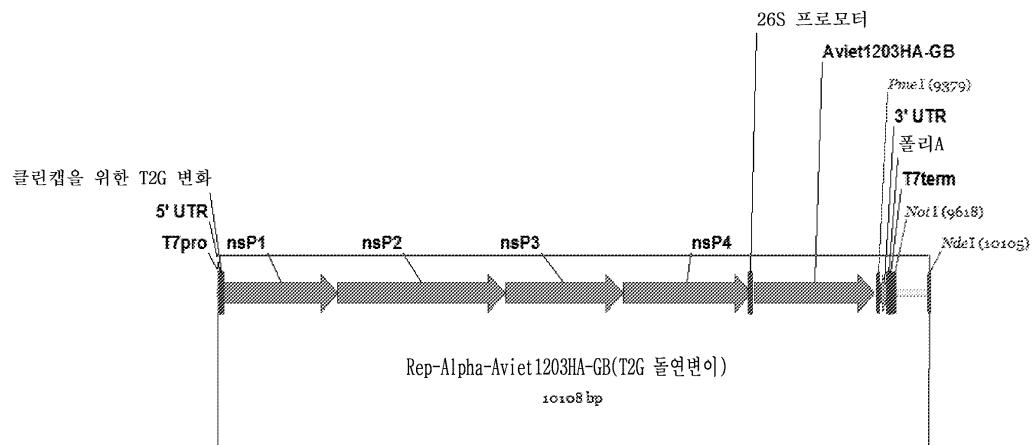
### 도면 1b



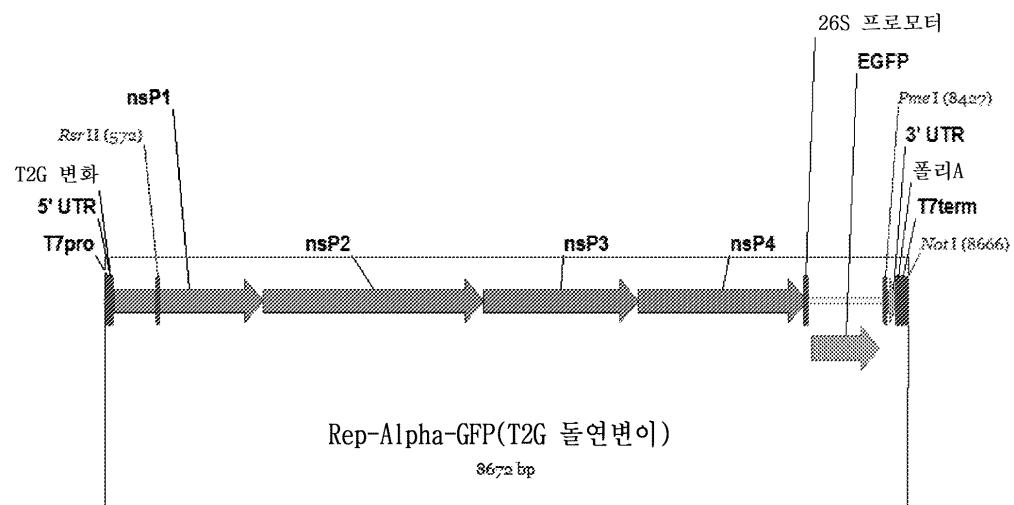
### 도면2



## 도면3

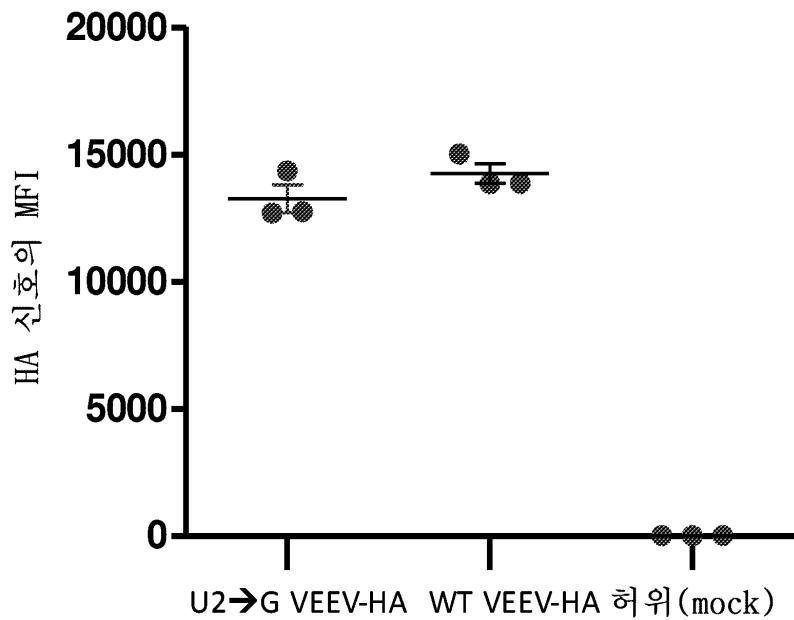


## 도면4



도면5

## 평균 형광 강도



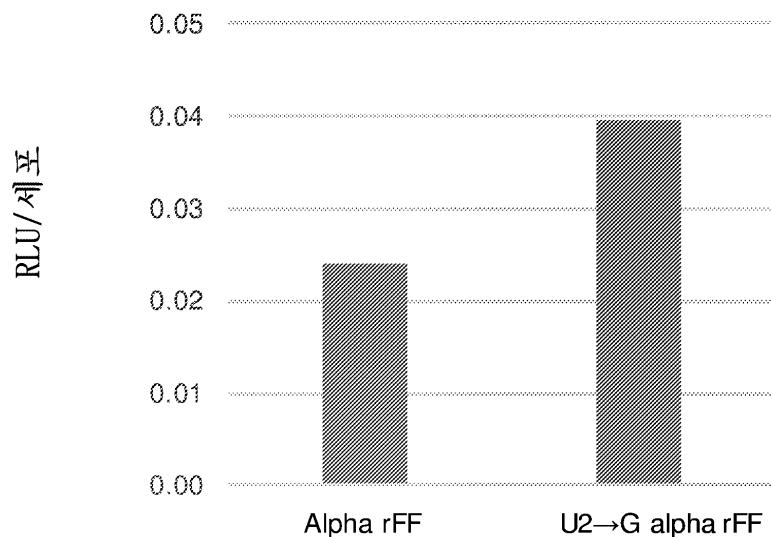
도면6

## 평균 형광 강도

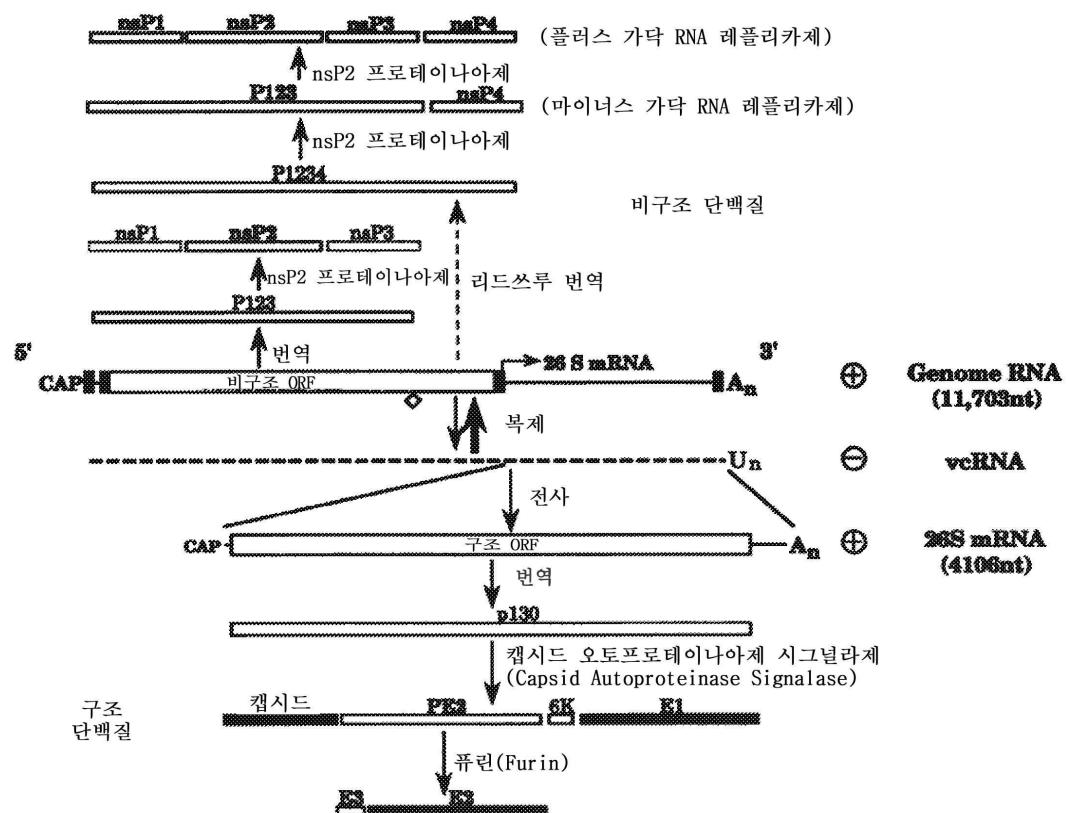


## 도면7

## 세포당 루시퍼라아제 발현



## 도면8



## 서 열 목록

<110> SYNTHETIC GENOMICS INC  
 <120> RECOMBINANT VIRUS REPLICON SYSTEMS AND USES THEREOF

<130> SGI.011W0  
 <150> PCT/US  
 <151> 2016-10-17  
 <160> 24  
 <170> KoPatent In 3.0  
 <210> 1  
 <211> 8364  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Synthetic polynucleotide  
 <220><221> misc\_feature  
 <223> Rep-Alpha replicon  
 <220><221> misc\_feature  
 <223> T7 promoter, modified VEEV 5'UTR and nonstructural genes  
 <400> 1

taatacact cactatagag aggccgcga tgagagaagc ccagaccaat tacctacca	60
aaatggagaa agttcacgtt gacatcgagg aagacagccc attcctcaga gcttgcagc	120
ggagctccc gcagtttagt gtagaagcca agcaggtcac tgataatgac catgtaatg	180
ccagagcggtt ttgcacatctg gttcaaaac tgatcgaaac ggaggtggac ccatccgaca	240
cgatccttga catttggaaat ggcggccccc gcagaatgta ttctaaagcac aagtatcatt	300
gtatctgtcc gaatggatgt ggcggaaatc cggacagatt gtataatgtat gcaactaagc	360
tgaagaaaaa ctgttggaaataactgata aggaatttggca aagaaaaatg aaggagctcg	420
ccggccgtcat gagcggaccct gaccttggaaa ctggactat gtgcctccac gacgacgagt	480
cgtgtcgcta cgaaggccaa gtcgtgttt accaggatgt atacgcgggtt gacggaccga	540
caagtctcta tcaccaagcc aataaggag tttagatcgctactggata ggcttigaca	600
ccaccccttt tatgtttaag aacttggctg gagcatatcc atcatactt accaactggg	660
ccgacgaaac cgtgttaacg gtcgttaaca taggcctatg cagctctgac gttatggagc	720
ggtcacgtt agggatgtcc attcttagaa agaagtattt gaaaccatcc aacaatgttc	780
tattctctgt tggctcgacc atctaccacg agaagaggaa cttactgagg agctggcacc	840
tgcgtctgt atttactta cgtggcaagc aaaattacac atgtcggtgt gagactatag	900
ttagttgcga cgggtacgtc gttaaaagaa tagctatcag tccaggcctg tatggaaagc	960
cttcaggccta tgctgctacg atgcaccgcg agggattctt gtgctgaaa gtgacagaca	1020

cattgaacgg ggagagggtc tctttcccg tgtgcacgt aatgtgtg	1080
accaaatgac tggcatactg gcaacagatg tcagtgcgga cgacgcgcaa aaactgtgg	1140
ttgggctaa ccagcgtata gtcgtcaacg gtcgcaccca gagaaacacc aataccatga	1200
aaaatcacct ttigccgta gtggccagg catttgcgt gttggcaag gaatataagg	1260
aagatcaaga agatgaaagg ccactaggac tacgagatag acagtttagtc atgggtgtt	1320
gttggcttt tagaaggcac aagataacat ctattataa gcgcggat acccaaacca	1380
tcatcaaagt gaacagcgat ttccactcat tcgtgtgcc caggataggc agtaacacat	1440
tggagatcg gctgagaaca agaatcagga aaatgttaga ggagcacaag gagccgtcac	1500
ctctcattac cgccgaggac gtacaagaag ctaagtgcgc agccgtatgag gctaaggagg	1560
tgcgtgaagc cgaggagtt cgccgcagtc taccacctt ggcagctgtat gttgaggagc	1620
ccactctgga aecgcgtatgc gacttgcgt tacaagaggc tggggccggc tcagtggaga	1680
cacctcggtt ctgtataaag gttaccagct acgtggcga ggacaagatc ggctcttacg	1740
ctgtgtttc tcccgaggct gtactcaaga gtgaaaatt atcttgcac caccctctcg	1800
ctgaacaagt catagtgata acacactctg gccgaaaagg gcgttatgcc gtggaaaccat	1860
accatggtaa agtagtggtg ccagagggac atgcaatacc cgtccaggac tttcaagctc	1920
tgagtgaaag tgcaccatt gtgtacaacg aacgtgagtt cgtaaacagg tacgtgcacc	1980
atattgccac acatggagga gcgctgaaca ctgtatgaa atattacaaa actgtcaagc	2040
ccagcgagca cgacggcgaa tacctgtacg acatcgacag gaaacagtgc gtcaagaaag	2100
aactagtac tgggctaggg ctcacaggcg agctggtgga tcctcccttc catgaattcg	2160
cctacgagag tctgagaaca cgaccagecg ctccttacca agtaccaacc ataggggtgt	2220
atggcgtgcc aggtcaggc aagtctggca tcattaaaag cgcgttccacc aaaaaagatc	2280
tagtggtgag cgccaagaaa gaaaactgtg cagaattat aaggacgtc aagaaaatga	2340
aaggcgtgga cgtcaatgcc agaactgtgg actcagtgtctt cttgtatggc tgcaacacc	2400
ccgttagagac cctgtatatt gacgaagtt ttgttgcgtca tgcaggact ctcagagcgc	2460
tcatagccat tataagaccc aaaaaggcag tgctctgccc ggatccaaa cagtgcgtt	2520
tttttaacat gatgtgcctg aaagtgcatt ttaaccacga gattgcaca caagtcttcc	2580
acaaaagcat ctctcgccgt tgcactaaat ctgtgacttc ggtcgctca accttgttt	2640
acgacaaaaaa aatgagaacg acgaatccga aagagactaa gattgtgatt gacactaccg	2700
gcagttaccaa acctaaggcag gacgttccca ttctacttg ttccagaggg tgggtgaagc	2760
agttgcaaat agattacaaa ggcaacgaaa taatgacggc agctgcctct caagggctga	2820
cccgtaaagg tgtgtatgcc gttcggtaca aggtgaatga aaatcctctg tacgcaccca	2880

cctctgaaca tgtgaacgtc ctactgaccc gcacggagga ccgcacgtg tgaaaacac	2940
tagccggcga cccatggata aaaacactga ctgccaagta ccctggaat ttcactgcca	3000
cgatagagga gtggcaagca gagcatgtg ccatcatgag gcacatctg gagagaccgg	3060
accctaccga cgttccag aataaggca acgtgtgtg ggccaaggct ttagtgcgg	3120
tgctgaagac cgctggata gacatgacca ctgaacaatg gaacactgtg gattatgg	3180
aaacggacaa agtcactca gcagagatag tattgaacca actatgcgtg aggttcttg	3240
gactcgatct ggactccggt ctatttctg cacccactgt tccgttatcc attaggaata	3300
atcaactggga taactccccg tcgcctaaca tgtacggct gaataaagaa gtggccgtc	3360
agctctctcg caggtaccca caactgcctc gggcagttgc cactgaaaga gtctatgaca	3420
tgaacactgg tacactgcgc aattatgatc cgccataaaa cctagtaccc gtaaacagaa	3480
gactgcctca tgcttagtc ctccaccata atgaacaccc acagagtgc ttttcttcat	3540
tcgtcagcaa attgaaggc agaactgtcc tgggtgtcg ggaaaagttg tccgtccag	3600
gcaaaatggt tgactggttg tcagaccggc ctgaggctac cttagagct cggctggatt	3660
taggcattcc aggtgatgtc cccaaatatg acataatatt tgtaatgtg aggacccat	3720
ataaataccca tcaactatcg cagtgtgaag accatgccat taagcttagc atgttgacca	3780
agaaagcttgc tctgcattcg aatccggcg gaaacctgtt cagcataggt tatggttacg	3840
ctgacagggc cagcgaaagc atcattgggt ctatagcgcg gcagttcaag ttttccggg	3900
tatgcaaaacc gaaatcctca cttgaagaga cgaaagtct gttgtattc attgggtacg	3960
atcgcaaggc ccgtacgcac aatccttaca agcttcatc aaccttgacc aacattata	4020
caggttccag actccacgaa gccggatgtg caccctata tcatgtgggt cgagggata	4080
ttgccacggc caccgaagga gtgattataa atgctgctaa cagcaaagga caacctggcg	4140
gaggggtgtg cggagcgctg tataagaaat tcccgaaag ctgcattta cagccgatcg	4200
aagtaggaaa agcgcgactg gtcaaagggt cagctaaaca tatcattcat gccgtaggac	4260
caaacttcaa caaagtttcg gaggttgaag gtgacaaaca gttggcagag gcttatgagt	4320
ccatcgctaa gattgtcaac gataacaatt acaagtcaatc agcgattcca ctgtgttcca	4380
ccggcatctt ttccggaaac aaagatcgac taacccaatc attgaaccat ttgctgacag	4440
cttttagacac cactgatgca gatgtagcca tatactgcag ggacaagaaa tggaaatga	4500
ctctcaagga agcagtggct aggagagaag cagtggagga gatatgcata tccgacgact	4560
cttcagtgac agaacctgtat gcagagctgg tgagggtgca tccgaagagt tctttggctg	4620

gaaggaaggg ctacagcaca agcgatggca aaactttctc atatttggaa gggaccaagt	4680
ttcaccaggc ggccaaggat atagcagaaa ttaatgccat gtggccgtt gcaacggagg	4740
ccaatgagca ggtatgcattatcctcg gagaaagcat gagcagtatt aggtcgaaat	4800
gccccgtcga agagtcggaa gcctccacac cacctagcac gctgccttgc ttgtgcattcc	4860
atgcattgac tccagaaaga gtacagcgcc taaaagccctc acgtccagaa caaattactg	4920
tgtgctatc cttccatttgcgaagtata gaatcactgg tgtgcagaag atccaatgct	4980
cccagcctat attttctca ccgaaagtgc ctgcgtatata tcatccaagg aagtatctcg	5040
tggaaacacc accggtagac gagactccgg agccatggc agagaaccaa tccacagagg	5100
ggacacctga acaaccacca cttataaccg aggatgagac caggactaga acgcctgagc	5160
cgatcatcat cgaagaggaa gaagaggata gcataagttt gctgtcagat ggcccgaccc	5220
accagggtct gcaagtcgag gcagacattc acggggccgc ctctgtatct agctcatct	5280
gttccattcc tcatgcattcc gactttgatg tggacagttt atccatactt gacaccctgg	5340
agggagctag cgtgaccagc ggggcaacgt cagccgagac taactttac ttcgcaaaga	5400
gtatggagtt tctggcgca cgggtgcctg cgcctcgaac agtattcagg aaccctccac	5460
atcccgctcc ggcacaaga acaccgtcac ttgcacccag cagggcctgc tcgagaacca	5520
gcctagtttccacccggcca ggcgtgaata ggggtatcac tagagaggag ctgcaggcgc	5580
ttaccccgctc acgcactcct agcaggtcg tctcgagaaac cagcctggc tccaaccgc	5640
caggcgtaaa tagggtgatt acaagagagg agtttgggc gttcgtagca caacaacaat	5700
gacggttga tgcgggtgca tacatcttt cctccgacac cggtaaggg cattacaac	5760
aaaaatcagt aaggcaaacg gtgctatccg aagtgggtt ggagaggacc gaattggaga	5820
tttctgtatgc ccccgccctc gaccaagaaa aagaagaatt actacgcaag aaattacagt	5880
taaatcccac acctgctaagc agaagcagat accagtccag gaaggtggag aacatgaaag	5940
ccataacagc tagacgtatt ctgcaaggcc tagggcatta tttgaaggca gaaggaaaag	6000
tggagtgcata cggaaacctg catcctgttc ctttgtattt atctagtgt aaccgtgcct	6060
tttcaagccc caaggtcgca gtggaaaccc gtaacgccc gttgaaagag aactttccga	6120
ctgtggcttc ttactgtattt attccagagt acgatgccta tttggacatg gttgacggag	6180
cttcatgcctg ctttagacact gccagttttt gcccgtcaaa gctgcgcagc tttccaaaga	6240
aacactctta ttggaaaccc acaatacgt cggcagtgcc tttagcgtac cagaacacgc	6300
tccagaacgt cctggcagct gccacaaaaa gaaattgcaa tgtcagcata atgagagaat	6360

tgcccgatt ggattcggcg gccttaatg tggaatgctt caagaaatat gcgtgtataa	6420
atgaatattg gaaacgttt aaagaaaacc ccatcaggtt tactgaagaa aacgtggtaa	6480
attacattac caaattaaaa ggacccaaag ctgctgctt tttgcgaag acacataatt	6540
tgaatatgtt gcaggacata ccaatggaca ggttgtaat ggacttaaag agagacgtga	6600
aagtgactcc aggaacaaaa catactgaag aacggccaa ggtacaggtg atccaggctg	6660
ccgatccgct agcaacagcg tatctgtcg gaatccaccg agagctggtt aggagattaa	6720
atgcggtcct gctccgaac attcatacac tgtttgat gtcggctgaa gactttgacg	6780
ctattatagc cgagcacttc cagcctgggg attgtgttct ggaaactgac atcgcgtcgt	6840
ttgataaaag tgaggacgac gccatggctc tgaccgcgtt aatgattctg gaagacttag	6900
gtgtggacgc agagctgttg acgctgattt aggcggctt cggcggaaatt tcatcaatac	6960
atttgccac taaaactaaa tttaaattcg gagccatgtat gaaatctgga atgttctca	7020
cactgtttgtt gaacacagtc attaacatttgaatcgcaag cagagtgtt agagaacggc	7080
taaccggatc accatgtgca gcattcattt gagaatgacaa tatcgtgaaa ggagtcaa	7140
cgacaaattt aatggcagac aggtgcgcac cctgggtt gaaatggaaatc aagattatag	7200
atgcgtgtt gggcgagaaa ggccttatt tctgtggagg gtttattttt tttgtactccg	7260
tgaccggcac agcgtgcgt gtggcagacc ccctaaaaag gctgttaag cttggcaaac	7320
ctctggcagc agacgatgaa catgatgtt acaggagaag ggcattgtcat gaagagtcaa	7380
cacgctggaa ccgagtgggtt attcttcag agctgtgcaat ggcagttagaa tcaaggtatg	7440
aaaccgtagg aacttccatc atagttatgg ccatgactac tctagcttagt agtggtaat	7500
cattcagcta cctggagggg gcccctataa ctctctacgg ctaacctgaa tggactacga	7560
catagtcgttccgcaaga tatcatgtt acagcagcaa ttggcaagct gcttacatag	7620
aaggcgcgcc gtttaaacgg ccggcctttaa ttaagtaacg atacagcagc aattggcaag	7680
ctgccttacat agaactcgcg gcgattggca tgccgtttaa aaattttttt tttttttt	7740
ttttttttt cgaatcgat ttttttttta atatttcaaa aaaaaaaaaa aaaaaaaaaa	7800
aaaaaaaaaa aaaaaaaaccctctctaaac ggaggggttt tttcagcgt aactggactg	7860
gccacagtta ggcggccgcg catgttcatc atcgttaccg cgtatcgta gcatccttc	7920
tcgttcatc ggtatcatta cctccatgaa cagaaatccc ctttacacgg aggcatcgt	7980
gaccaaacag gaaaaaacgg cccttaacat gcccgcctt atcagaagcc agacattaac	8040
gcttctggag aaactcaacg agctggacgc ggttgcacag gcagacatct gtgtatcgct	8100
tcacgaccac gctgtatgacg tttaccgcag ctgcctcgctt cggttgcgtt atgacgggtga	8160
aaacctctga cacatgcagc tccggagac ggtcacagct tgtctgtaa gggatgccgg	8220

gaggcagacaa	gcccgtcagg	gcccgtcagg	gcccgtcagg	gggtgttggc	gggtgtcggg	gcccgtcagg	gcccgtcagg	8280
gaccaggatca	cgttagcgata	gcggaggtgt	gcggaggtgt	tactggctta	actatgcggc	atcagagcag	atcagagcag	8340
attgtactga	gagtgcacca	tatg						8364
<210>	2							
<211>	50							
<212>	DNA							
<213>	Aura virus							
<220><221>	misc_feature							
<223>	5'UTR							
<400>	2							
atagcggacg	gactagta	tgtactacag	aattaactgc	cgtgtgccgc				50
<210>	3							
<211>	50							
<212>	DNA							
<213>	Chikungunya virus							
<220><221>	misc_feature							
<223>	5'UTR							
<400>	3							
atggctgcgt	gagacacacg	tagcctacca	gtttcttact	gctctactct				50
<210>	4							
<211>	50							
<212>	DNA							
<213>	O'nyong-nyong virus							
<220><221>	misc_feature							
<223>	5'UTR							
<400>	4							
atagctgcgt	gatacacaca	cgcagttac	gggtttcata	ctgctctact				50
<210>	5							
<211>	50							
<212>	DNA							
<213>	bebaru virus							
<220><221>	misc_feature							
<223>	5'UTR							

&lt;400&gt; 5

atggcggctg tgtgacacac gagccgtcga tttcaacctt cttgctccct 50

&lt;210&gt; 6

&lt;211&gt; 50

&lt;212&gt; DNA

&lt;213&gt; Semliki Forest virus

&lt;220&gt;&lt;221&gt; misc\_feature

&lt;223&gt; 5'UTR

&lt;400&gt; 6

atggcggatg tgtgacatac acgacgccaa aagatttgt tccagctcct 50

&lt;210&gt; 7

&lt;211&gt; 50

&lt;212&gt; DNA

&lt;213&gt; Mayaro virus

&lt;220&gt;&lt;221&gt; misc\_feature

&lt;223&gt;

&gt; 5'UTR

&lt;400&gt; 7

atggcggca agtgacactt gttccggccgg tcgtctctaa gctttcctc 50

&lt;210&gt; 8

&lt;211&gt; 50

&lt;212&gt; DNA

&lt;213&gt; getah virus

&lt;220&gt;&lt;221&gt; misc\_feature

&lt;223&gt; 5'UTR

&lt;400&gt; 8

atggcggacg tgtgacatca ccgttcgtc tttcttaggt cctttgtac 50

&lt;210&gt; 9

&lt;211&gt; 50

&lt;212&gt; DNA

&lt;213&gt; Sagiymama virus

&lt;220&gt;&lt;221&gt; misc\_feature

&lt;223&gt; 5'UTR

&lt;400&gt; 9

atggcggacg tgtgacatca ccgttcgctc tttcttaggat cctttgtac	50
<210> 10	
<211> 50	
<212> DNA	
<213> Ndumu virus	
<220><221> misc_feature	
<223> 5'UTR	
<400> 10	
atggtgccga gttagagac gaagcaccaa acaactacgc ggctcaccat	50
<210> 11	
<211> 50	
<212> DNA	
<213> Middelburg virus	
<220><221> misc_feature	
<223> 5'UTR	
<400> 11	
atggtggtt acgtacacgt gccaccaccc cccaccctcc aagcgatcca	50
<210> 12	
<211> 50	
<212> DNA	
<213> Eastern equine encephalomyelitis virus	
<220>	
<221> misc_feature	
<223> 5'UTR	
<400> 12	
ataggtagcgtgttagaggc aaccacccta ttccacccta tccaaaatgg	50
<210> 13	
<211> 50	
<212> DNA	
<213> Fort Morgan virus	
<220><221> misc_feature	
<223> 5'UTR	
<400> 13	
ataggtagcgttttagaggc gcctacccta cttaaccgat ccaaacatgg	50

<210> 14  
 <211> 50  
 <212> DNA  
 <213> Buggy Creek virus  
 <220><221> misc\_feature  
 <223> 5'UTR  
 <400> 14  
 ataggtagatg gtttagaggc gcctacccta ctttaaccgat ccaaacatgg 50

<210> 15  
 <211> 50  
 <212> DNA  
 <213> Venezuelan equine encephalitis virus  
 <220><221> misc\_feature  
 <223> 5'UTR  
 <400> 15  
 atggcggcg caagagagaa gcccaaacca attacctacc caaaatggag 50

<210> 16  
 <211> 50  
 <212> DNA  
 <213> Whataroa virus  
 <220><221> misc\_feature  
 <223> 5'UTR  
 <400> 16  
 attggcggca tagtacatac tatataaaag aaacagccga ccaattgcac 50

<210> 17  
 <211> 50  
 <212> DNA  
 <213> Sindbis virus  
 <220><221> misc\_feature  
 <223> 5'UTR  
 <400> 17  
 attgacggcg tagtacacac tattgaatca aacagccgac caattgcact 50

<210> 18

<211> 50  
 <212> DNA  
 <213> bebaru virus  
 <220><221> misc\_feature  
 <223> 5'UTR  
 <400> 18  
 attggcggcg tagtacacac tattgaatca aacagccgac caattgcact 50  
 <210> 19  
 <211> 9476  
 <212> DNA  
 <213> Artificial Sequence  
 <220><223> Alpha-R-rFF-T2G  
 <220><221> misc\_feature  
 <222> (1)..(18)  
 <223> T7 promoter

<220><221> misc\_feature  
 <222> (20)  
 <223> T to G substitution  
 <220><221> misc\_feature  
 <222> (9437)..(9476)  
 <223> Poly A  
 <400> 19  
 taatacgact cactatagag aggccgcga tgagagaagc ccagaccaat tacctacca 60  
 aaatggagaa agttcacgtt gacatcgagg aagacagccc attcctcaga gcttgcgc 120  
 ggagctccc gcagtttagt gtagaagcca agcaggtcac tgataatgac catgtaatg 180  
 ccagagcggtt ttgcgcattcgt gcttcaaaac tgatcgaaac ggaggtggac ccatccgaca 240  
 cgatccttga catttggaaatgtt ggcggccccc gcagaatgta ttctaaagcac aagtatcatt 300

gtatctgtcc gatgagatgt gcgaaagatc cggacagatt gtataagtat gcaactaagc 360  
 tgaagaaaaa ctgttggaa ataactgata aggaatttggaa caagaaaaatg aaggagctcg 420  
 ccgcgcgtcat gagcgcaccct gacctggaaa ctgagactat gtgcctccac gacgcgcgt 480  
 cgtgtcgcta cgaaggccaa gtcgtgtttt accaggatgt atacgcgggtt gacggaccga 540  
 caagtctcta tcaccaagcc aataagggag ttagagtgcgctactggata ggcttgaca 600  
 ccaccccttt tatgtttaag aacttggctg gagcatatcc atcatactt accaactgg 660

ccgacgaaac cgtgttaacg gtcgtaaca taggcctatg cagctctgac gttatggac	720
ggtcacgtag agggatgtcc attcttagaa agaagtattt gaaaccatcc aacaatgttc	780
tattctctgt tggctcgacc atctaccacg agaagaggga cttaactgagg agctggcacc	840
tgccgtctgt atttactta cgtggcaagc aaaattacac atgtcggtgt gagactatag	900
ttagttcgca cgggtacgct gttaaaagaa tagctatcag tccaggcctg tatggaaagc	960
cttcaggcata tgctgctacg atgcaccgcg agggattctt gtgctgaaa gtgacagaca	1020
cattgaacgg ggagagggtc tctttcccg tgtgcacgta tgtgccagct acattgtgt	1080
accaaatgac tggcatactg gcaacagatg tcagtgcgga cgacgcgcaa aaactgctgg	1140
ttgggctcaa ccagcgtata gtcgtcaacg gtcgcaccca gagaacacacc aataccatga	1200
aaaattacct ttgcggcgtaa gtggccagg catttgctag gtggccaaag gaatataagg	1260
aagatcaaga agatgaaagg ccactaggac tacgagatag acagttagtc atgggtgtt	1320
gttggcctt tagaaggcac aagataacat ctattataa gcgcggat acccaaacc	1380
tcatcaaagt gaacagcgat ttccactcat tcgtgctgcc caggataggc agtaacacat	1440
tggagatcg gctgagaaca agaatcagga aaatgttaga ggagcacaag gagccgtcac	1500
ctctcattac cgccgaggac gtacaagaag ctaagtgcgc agccgatgag gctaaggagg	1560
tgcgtgaagc cgaggagttg cgccgcagctc taccacctt ggcagctgat gttgaggagc	1620
ccactctgga agccgatgtc gacttgatgt tacaagagc tggggccggc tcagtggaga	1680
cacctgtgg ctgtataaag gttaccagct acgatggcga ggacaagatc ggctttagc	1740
ctgtgcttc tcccgaggct gtactcaaga gtgaaaaatt atttgcate caccctctcg	1800
ctgaacaagt catagtgata acacactctg gccgaaaagg gcgttatgcc gtggaaaccat	1860
accatggtaa agtagtggtg ccagagggac atgcaatacc cgtccaggac tttcaagctc	1920
tgagtgaaag tgccaccatt gtgtacaacg aacgtgagtt cgtaaacagg tacctgcacc	1980
atattgccac acatggagga gcgcgtgaaca ctgtatgaaga atattacaaa actgtcaagc	2040
ccagcgagca cgacggcgaa tacctgtacg acatcgacag gaaacagtgc gtcaagaaag	2100
aactagtac tgggctaggg ctcacaggcg agctggtgaa tcctcccttc catgaattcg	2160
cctacgagag tctgagaaca cgaccagccg ctccttacca agtaccaacc ataggggtgt	2220
atggcgtgcc aggtcaggc aagtctggca tcattaaaag cgcgtcacc aaaaaagatc	2280
tagtggtgag cgccaagaaa gaaaactgtg cagaattat aaggacgtc aagaaaatga	2340
aaggcgtgga cgtcaatgcc agaactgtgg actcagtgtcttgaaatgga tgcaaacacc	2400



caaactcaa caaagttcg gaggttgaag gtgacaaaca gttggcagag gcttatgagt	4320
ccatcgctaa gattgtcaac gataacaatt acaagtcagt agcgattcca ctgttgtcca	4380
ccggcatctt ttccggAAC aaagatcgac taacccaatc attgaaccat ttgctgacag	4440
cttagacac cactgatgca gatgttagcca tatactgcag ggacaagaaa tggaaatga	4500
ctctcaagga agcagtggct aggagagaag cagtgaggaa gatatgcata tccgacgact	4560
cttcagtgac agaacctgat gcagagctgg tgagggtgca tccgaagagt tctttggctg	4620
gaagggaggc ctacagcaca agcgatggca aaacttctc atattggaa gggaccaagt	4680
ttcaccaggc gccaaggat atagcagaaa ttaatgccat gtggccgtt gcaacggagg	4740
ccaatgagca ggtatgcatg tatacctcg gagaagcat gagcagtatt aggtcgaat	4800
gccccgtcga agagtggaa gcctccacac cacctagcac gctgccttgc ttgtgcatcc	4860
atgccatgac tccagaaaga gtacagcgcc taaaagccctc acgtccagaa caaattactg	4920
tgtgctcatc cttccattt ccgaagtata gaatcactgg tgtgcagaag atccaatgct	4980
cccagcctat atgttctca ccgaaagtgc ctgcgtatata tcatccaagg aagtatctcg	5040
tggaaacacc accggtagac gagactccgg agccatcgcc agagaaccaa tccacagagg	5100
ggacacctga acaaccacca cttataaccg aggatgagac caggactaga acgcctgagc	5160
cgatcatcat cgaagaggaa gaagaggata gcataagtt gctgtcagat gggccgaccc	5220
accagggtct gcaagtgcag gcagacattc acggggccccc ctctgtatct agctcatct	5280
ggtcattcc tcatgcatcc gactttgtat tggacagttt atccatactt gacaccctgg	5340
aggagactag cgtgaccaggc ggggcaacgt cagccgagac taactttac ttgcacaaga	5400
gtatggagtt tctggcgcga ccgggcctg cgcctcgaaac agtattcagg aaccctccac	5460
atcccgctcc ggcacaaga acaccgtcac ttgcacccag cagggcctgc tcgagaacca	5520
gcctagttt caccggcca ggcgtgaata gggtgatcac tagagaggag ctcgaggcgc	5580
ttacccgctc acgcactctt acgcaggctgg tctcgagaaac cagcctggc tccaaacccgc	5640
caggcgtaaa taggtgatt acaagagagg agtttgggc gttcgttagca caacaacaat	5700
gacggttga tgccggtgca tacatcttt cctccgacac cggtaaggg catttacaac	5760
aaaaatcaat aaggcaaaac gtgctatccg aagtgggtt ggagaggacc gaattggaga	5820
tttcgtatgc cccgcgcctc gaccaagaaa aagaagaatt actacgcaag aaattacagt	5880
taaatccac acctgtcaac agaagcagat accagtccag gaaggtggag aacatgaaag	5940
ccataacagc tagacgtatt ctgcaaggcc tagggcatta tttgaaggca gaaggaaaag	6000
tggagtgcta cggaaacctg catcctgttc ctttgtatcc atctgtgt aaccgtgcct	6060
tttcaagccc caaggctcgca gtggaaagcc gtaacgcatt gttgaaagag aactttccga	6120

ctgtggcttc ttactgtatt attccagagt acgatgccta tttggacatg gttgacggag	6180
cttcatgctg cttagacact gccagtttt gccctgcaaa gctgcgcagc tttccaaaga aacactccta ttggaaccc acaatacgat cggcagtgcc ttcagcgatc cagaacacgc tccagaacgt cctggcagct gccacaaaaa gaaattgcaa tgcacgcaa atgagagaat tgcccgatt ggattcgccg gccttaatg tggaatgctt caagaatat gcgtgtaata atgaatattg gaaacgttt aaagaaaacc ccatcaggct tactgaagaa aacgtggtaa attacattac caaattaaaa ggaccaaaag ctgctgctt tttgcgaag acacataatt tgaatatgtt gcaggacata ccaatggaca ggttgtaat ggacttaaag agagacgtga	6240 6300 6360 6420 6480 6540 6600
aagtgactcc aggaacaaaa catactgaag aacggccaa ggtacaggtg atccaggctg ccgatccgct agcaacagcg tatctgtcg gaatccaccc agagctggtt aggagattaa atgcggctt gctccgaac attcatacac tggatgtat gtcggctgaa gactttgacg ctattatgc cgagcacttc cagcctgggg attgtgttct ggaaactgac atcgcgtcgt ttgataaaag tgaggacgac gccatggctc tgaccgcgtt aatgattctg gaagacttag gtgtggacgc agagctgtt acgctgattt aggccgttt cggcgaaatt tcatcaatac atttgcccac taaaactaaa tttaattcg gagccatgat gaaatctgga atgttcctca	6660 6720 6780 6840 6900 6960 7020
cactgtttgt gaacacagtc attaacattt gatcgcaag cagagtgtt agagaacggc taaccggatc accatgtgca gatttcattt gagaatgacaa tatcgtgaaa ggagtcaaat cgacaaatt aatggcagac aggtgcgcca cctgggttgaat tatggaagtc aagattatag atgcgttgtt gggcgagaaa ggccttatt tctgtggagg gttatttt gtcgtactcg tgaccggcac agcgtccgt gtggcagacc ccctaaaaag gctgtttaag cttggcaaac ctctggcagc agacgatgaa catgatgatg acaggagaag ggcattgcat gaagagtcaa cacgctggaa ccgagtggtt attcttcag agctgtgcaag ggcagttagaa tcaaggtatg	7080 7140 7200 7260 7320 7380 7440
aaaccgtagg aactccatc atagttatgg ccatgactac tctagctac agtgttaat cattcagcta cctgagaggg gcccctataa ctctctacgg ctaacctgaa tggactacga catagtctag tccgccaaga tatgcacca tggaaaat gaaaaacgc gagaacatcg tggggccc caagcccttc taccctatcg aggaaggcag cgccggcacc cagctgcgga agtacatgga aagatacgcc aagctggcg ccattgcctt caccaacgcc gtgaccggcg tggactacag ctacgcccag tacctggaaa agagctgtcg cctggcaag gctctgcaga actacggctt ggtgggtggac ggccggatcg ccctgtgcag cgagaactgc gaggaattct	7500 7560 7620 7680 7740 7800 7860

tcatccccgt gatcgccggc ctgttcatcg gcgtggcggt ggctcccacc aacgagatct	7920
acaccctgcg ggagctgggt cacagcctgg gcatcagcaa gcccaccatc gtgtttagca	7980
gcaagaaggc cctggacaaa gtcataccg tgcagaaaac cgtgaccacc atcaagacca	8040
tcgtgatcct ggacagcaag gtggactacc ggggctacca gtgcctggac accttcaica	8100
agcggAACAC cccccctggc ttccaggcca gcagttcaa gaccgtggag gtggaccgga	8160
aagaacaggt gcccctgatc atgaacagca gcggcagcac cggcctgccc aaggcgtgc	8220
agctgaccca cgagaacacc gtgacccggt tcagccacgc cagggacccc atctacggca	8280
accaggtgtc ccccgccacc gccgtgtga ccgtggtgcc cttccaccac ggcttcggca	8340
tgttcaccac cctggctac ctgatctcg gcttcgggt ggtgatgtc accaagttcg	8400
acgagggaaac cttctgaaa accctgcagg actacaagt cacctacgtt attctggtgc	8460
ccaccctgtt cgcacatcctg aacaagagcg agctgtgaa caagtacgac ctgagcaacc	8520
tggtgagat cggcagcggc ggagcccccc tgagcaaaga agtggagag ggcgtcgcca	8580
ggcggttcaa tctgccccggc gtgcggcagg gctacggctt gaccgagaca accagcgcca	8640
tcatcatcac ccccgagggc gacgacaaggc ctggagccag cggcaaggtg gtggccctgt	8700
tcaaggccaa agtgatcgac ctggacacca agaagagcct gggcccaac agacggggcg	8760
aagtgtgcgt gaaggcccccc atgctgtatga agggctacgt gaacaacccc gagggccacca	8820
aagagctgat cgacgaagag ggctggctgc acaccggcga catcggtac tacgacgaag	8880
agaagcactt ctcatcggt gaccggctga agagctgtat caagtacaag ggctatcagg	8940
tgccccctgc cgagctggaa agcgtctgc tgcagcaccc cagcatcttc gacggccggcg	9000
tggccggggt gccagatctt gtggccggcg agctgcctgg cggcgtgggt gtgcgtggaa	9060
ccggcaagaa catgaccgag aaagaagtga tggactacgt cggcagccag gtgtccaacg	9120
ccaagcggct gagaggcggc gtgagattcg tggacgaagt gccaaggc ctgaccggca	9180
agatcgacgg cagggccatc cgggagatcc tgaagaaacc cgtggccaag atgtgattaa	9240
ttgatcgata cagcagcat tggcaagctg cttacataga aggcgcggc tttaaacggc	9300
cggccttaat taagtaacga tacagcagca attggcaagc tgcttacata gaactcgccg	9360
cgattggcat gcccgtttaa aatttttatt ttattttct tttttttcc gaatcggtt	9420
ttgttttaa tattcaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa	9476
<210> 20	
<211> 1647	
<212> DNA	
<213> Artificial Sequence	

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;220&gt;&lt;221&gt; misc\_feature

&lt;223&gt; Red Firefly reporter gene - rFF

&lt;400&gt; 20

atggaaaata	tgaaaaacga	cgagaacatc	gtggtggcc	ccaagccctt	ctacccatc	60
gaggaaggca	gcgcggcac	ccagctgcgg	aagtacatgg	aaagataacgc	caagctggc	120
gccattgcct	tcaccaacgc	cgtgaccggc	gtggactaca	gctacggca	gtacctggaa	180
aagagctgct	gcctggcaa	ggctctgcag	aactacggcc	tgggtggta	cggccggatc	240
gccctgtgca	gcgagaactg	cgaggaattc	ttcatcccc	tatcgccgg	cctgttcatc	300
ggcgtggcg	tggctccac	caacgagatc	tacaccctgc	gggagctgg	gcacagcctg	360
ggcatcagca	ageccaccat	cgtgttcagc	agcaagaagg	gcctggacaa	agtcatcacc	420
gtgcagaaaa	ccgtgaccac	catcaagacc	atcgtgatcc	tggacagcaa	ggtggactac	480
cgggctacc	agtgcctgga	cacccatc	aagcggaaaca	ccccctgg	cttcaggcc	540
agcagcttca	agaccgtgga	ggtggaccgg	aaagaacagg	tggccctgat	catgaacagc	600
agcggcagca	ccggcttgcc	caagggcg	cagctgaccc	acgagaacac	cgtgaccgg	660
ttcagccacg	ccagggaccc	catctacggc	aaccaggtt	ccccggcac	cgccgtgctg	720
accgtggtgc	cttccacca	cggttcggc	atgttacca	ccctggcta	cctgatctgc	780
ggctccggg	tgtgtatgct	gaccaagttc	gacgaggaaa	cttctgaa	aaccctgcag	840
gactacaagt	gcacctacgt	gattctggtg	cccaccctgt	tcgcacatct	gaacaagagc	900
gagctgctga	acaagtacga	cctgagcaac	ctggggaga	tcgcacgcgg	cggagcccc	960
ctgagcaaaag	aagtggaga	ggccgtcgcc	aggcggttca	atctgcccgg	cgtgcggcag	1020
ggctacggcc	tgaccgagac	aaccagcgcc	atcatcatca	ccccgaggg	cgacgacaag	1080
cctggagcca	gcccgaaggt	ggtgcccctg	ttcaaggcca	aagtgtatcga	cctggacacc	1140
aagaagagcc	tggcccaa	cagacggggc	gaagtgtgcg	tgaaggccc	catgtgtat	1200
aaggctacg	tgaacaaccc	cgaggccacc	aaagagctga	tcgacgaaga	gggctggctg	1260
cacaccggcg	acatcggtca	ctacgacgaa	gagaagcact	tcttcatctg	ggaccggctg	1320
aagagctgta	tcaagtacaa	ggctatcag	gtgccccctg	ccgagctgg	aagcgtctg	1380
ctgcagcacc	ccagcatctt	cgacggccgc	gtggccgggg	tgcacatcc	tgtggccggc	1440
gagctgctg	gcccgtggt	ggtgtggaa	tccggcaaga	acatgaccga	gaaagaagt	1500
atggactacg	tcgcaccca	ggtgttcaac	gccaaggccgc	ttagaggccg	cgtgagattc	1560
gtggacgaag	tgcacaaagg	cctgaccggc	aagatcgacg	gcaggccat	ccggagatc	1620

ctgaagaaac ccgtggccaa gatgtga	1647
<210> 21	
<211> 729	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic polynucleotide	
<220><221> misc_feature	
<223> EGFP	
<400> 21	
atggaagag ccggcgtgag caagggcgag gagctgttca cgggggttgt gcccattctg	60
gtcgagctgg acggcgtacgt aaacggccac aagtccagcg tgcggcgtt gggcgaggc	120
gatgccacct acggcaagct gaccctgaag ctgatctgca ccaccggcaa gctgcccgtg	180
ccctggccca ccctcggttacccctggc tacggcctgc agtgcttcgc ccgttacccc	240
gaccacatga agcagcacgttcaag tccggcatgc ccgaaggcta cgtccaggag	300
cgcacatct tcttcaaggttcaaggttcaacgttccggcaac tacaagaccc ggcggaggttcaaggttccggcaac	360
ggcgacaccc tggtaaccgttcatcg aaggcatcg acttcaaggttcaacgttccggcaac	420
atcctgggc acaagctggttacaactac aacagccaca acgtctat caccggccac	480
aaggcagaaga acggcatcaa ggccaaacttca aagatccgttcaacatcgaa ggacggccgttcaacgttccggcaac	540
gtgcagctgttccgaccacta ccagcagaac accccatcg ggcggccgttcaacgttccggcaac	600
cccgacaacc actacctgag ctaccgttcc gcccgttca aagacccaa cgagaagcgcc	660
gatcacatgg tcctgttca gttcggttcc gccggccggatcaactcttcatggacgttca	720
ctgttacaag	729
<210> 22	
<211> 1707	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic polynucleotide	
<220><221> misc_feature	
<223> HA Influenza A virus	
A-VietNam-1203-2004 (H5N1) - GenBank	
<400> 22	
atggagaaaa tagtgcgttcttttgcata gtcgtcttgc taaaagtgttgc	60

atggttacc atgcaaaca ctcgacagag caggttgaca caataatgga aaagaacgtt	120
actgttacac atgccaaga catactggaa aagaacaca acggaaagct ctgcgatcta	180
gatggagtga agcctcta at ttgagagat tgttagctag ctggatggct cctcgaaac	240
ccaaatgtgt acgaattcat caatgtgcgc gaatgtt acatagtggaa gaaggccat	300
ccagtcatttgc acctctgtt cccagggat ttcaatgact atgaagaatt gaaacaccta	360
ttgagcagaa taaaccattt tgagaaaatt cagatcatcc cccaaagtcc ttggccagt	420
catgaaggct cattaggggt gagctcagca tgtccatacc agggaaagtc ctccttttc	480
agaatgtgg tatggctt caaaaagaac agtacatacc caacaataaa gaggagctac	540
aataatacca accaagaaga tctttggta ctgtgggaa ttccatcc taatgtgcg	600
gcagagcaga caaagctcta taaaaccca accacctata ttccgttgg gacatcaaca	660
ctaaaccaga gattggtacc aagaatagct actagatcca aagtaaacgg gcaaagtgg	720
aggatggagt tcttctggac aattttaaag ccgaatgatg caatcaactt cgagagtaat	780
ggaaatttca ttgctccaga atatgcatac aaaattgtca agaaaggaa ctcaacaatt	840
atgaaaagtg aatttggata tggtaactgc aacaccaagt gtcaaactcc aatggggcgc	900
ataaactcta gcatgccatt ccacaatata caccctctca ccattgggaa atgccccaaa	960
tatgtgaaat caaacagatt agtccttgcg actgggctca gaaatagccc tcaaagagag	1020
agaagaagaa aaaagagagg attatttggaa gctatagcag gtttataga gggaggatgg	1080
cagggaatgg tagatggttt gtaggggtac caccatagca atgagcagg gtaggggtac	1140
gctcagaca aagaatccac taaaaggca atagatggg tcaccaataa ggtcaactcg	1200
atcattgaca aatgaacac tcagtttgg gccgttggaa gggaaattaa caacttagaa	1260
aggagaatag agaattttaa caagaagatg gaagacgggt tcctagatgt ctggacttat	1320
aatgctgaac ttctggttt catggaaaat gagagaactc tagacttca tgactcaa	1380
gtcaagaacc ttacgacaa ggtccgacta cagcttaggg ataatgcaaa ggagctgggt	1440
aacggttgc ttccatgtc tccataatgt gataatgaat gtagggaaat tgtaagaaat	1500
ggaaacgtatg actacccgca gtattcagaa gaagcggac taaaagaga gggaaataat	1560
ggagtaaaat tggaaatcaat aggaatttac caaatactgt caatttattc tacagtggcg	1620
atgtccctag cactggcaat catggtagct ggtctatcct tatggatgtc ctccaaatgg	1680
tcgttacaat gcagaatttgcattttaa	1707

&lt;210&gt; 23

&lt;211&gt; 29

&lt;212&gt; DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<220><223> VEE 5' T-G nt 2 F - Forward Primer

<400> 23

cgactcacta tagagaggcg gcgcatgag

29

<210> 24

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic polynucleotide

<220><223> VEE 5' T-G nt 2 R - Reverse Primer

<400> 24

ctcatgcgcc gcctcttat agtgagtgc

29