



- (12) BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ
(19) Cộng hòa xã hội chủ nghĩa Việt Nam (VN) (11) 
CỤC SỞ HỮU TRÍ TUỆ
1-0048957
(51)^{2020.01} C12N 15/82; C12N 15/74; A01H 5/00; (13) B
C12N 15/32
-

- (21) 1-2022-02143 (22) 26/08/2020
(86) PCT/US2020/047899 26/08/2020 (87) WO 2021/045942 11/03/2021
(30) 62/896,735 06/09/2019 US
(45) 25/07/2025 448 (43) 25/08/2022 413A
(73) SYNGENTA CROP PROTECTION AG (CH)
Rosentalstrasse 67, 4058, Basel, Switzerland
(72) AZHAKANADAM, Kasimalai (US); ZHOU, Ailing (US); CONVILLE, Jared (US);
CLARKE V, Joseph Dallas (US).
(74) Công ty TNHH Ban Ca (BANCA)
-

- (54) CATXET BIỂU HIỆN BAO GỒM VÙNG KHỞI ĐỘNG, VECTƠ TÁI TỔ HỢP VÀ
PHƯƠNG PHÁP BIỂU HIỆN PROTEIN HOẶC POLYNUCLEOTIT ĐƯỢC
QUAN TÂM

(21) 1-2022-02143

(57) Sáng chế đề cập đến các vùng khởi động mà có việc sử dụng cụ thể trong việc dẫn hướng việc biểu hiện đặc hiệu theo gốc của các gen dị loại mà truyền các đặc tính nông học, làm vườn và/hoặc diệt sinh vật gây hại gia tăng cho thực vật chuyển gen nhất định. Sáng chế cũng đề cập đến các phân tử ADN bao gồm các vùng khởi động theo sáng chế và các mô thực vật được biến nạp chứa các phân tử ADN bao gồm vùng khởi động theo sáng chế được liên kết hoạt động với gen hoặc các gen dị loại, và các hạt của chúng. Cụ thể, sáng chế đề cập đến catxet biểu hiện bao gồm vùng khởi động, vector tái tổ hợp. Sáng chế còn đề cập đến phương pháp biểu hiện protein hoặc polynucleotit được quan tâm trong thực vật hoặc tế bào thực vật và phương pháp biểu hiện đặc hiệu trình tự mã hóa dị loại trong mô thực vật chuyển gen.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập chung đến lĩnh vực sinh học phân tử thực vật và việc điều hòa của sự biểu hiện gen ở các thực vật. Sáng chế cũng đề cập đến các sinh vật nhân thực chuyển gen, bao gồm các tế bào thực vật, các thực vật, và các hạt chuyển gen, mà hệ gen của chúng bao gồm các cấu trúc phân tử để kiểm soát sự biểu hiện gen dị loại.

Tình trạng kỹ thuật của sáng chế

Các cây trồng chuyển gen chứa các cải biến di truyền ngày càng phức tạp bao gồm nhiều gen chuyển mà tạo nên nhiều tính trạng khác nhau, cũng được gọi là “các nhóm gen” hoặc “các nhóm tính trạng”. Ví dụ, nhiều sản phẩm ngô chuyển gen hiện nay trên thị trường chứa trong cùng thực vật nhiều gen mã hóa các protein diệt côn trùng để kiểm soát phổ rộng của các côn trùng gây hại, nhiều gen mã hóa các protein giúp thực vật có khả năng chống chịu với phổ rộng của các chất diệt cỏ hóa học và nhiều gen mã hóa các protein mà được sử dụng làm các đánh dấu chọn lọc trong suốt quá trình biến nạp thực vật. Nhiều trong số các protein chuyển gen được sử dụng để kiểm soát các côn trùng gây hại, ví dụ các nội độc tố tinh thể từ *Bacillus thuringiensis* (được gọi là các protein Cry) có hoạt tính chống lại các côn trùng gây hại bộ cánh vảy hoặc bộ cánh cứng. Các ví dụ của các protein Cry hoạt tính bộ cánh vảy bao gồm Cry1A, Cry1B, Cry1C, Cry1D, Cry1E, Cry1F và Cry9. Các ví dụ của các protein Cry hoạt tính bộ cánh cứng bao gồm, Cry3A, Cry3B, Cry3C, Cry8, Cry23-Cry37 nhị phân và Cry34-Cry35 nhị phân. Hầu hết các protein Cry riêng lẻ có hoạt tính sinh học chống lại phổ hẹp của các loài côn trùng trong bộ côn trùng đã biết. Ngay cả với phổ hoạt tính hẹp này, các protein Cry nhất định có thể có hoạt tính thấp đến trung bình chống lại các loài không phải sinh vật gây hại nhất định trong cùng bộ của các côn trùng như các côn trùng là sinh vật gây hại mục tiêu. Ví dụ, Hellmich et al. (2001) Proc. Natl. Aca. Sci. 98:11925-11930, đã phát hiện rằng các protein Cry được tinh sạch nhất định mà có hoạt tính chống lại sinh vật gây hại bộ cánh vảy, ví dụ sâu đục thân ngô châu Âu (*Ostrinia nubilalis*), cũng có một số hoạt tính chống lại giai đoạn đầu tiên của côn trùng bộ cánh vảy không phải sinh

vật gây hại, bướm chúa (*Danaus plexippus*). Tuy nhiên, các giai đoạn ấu trùng sau đó *D. plexippus* là ít nhạy cảm hơn nhiều.

Hiện nay, việc biểu hiện của hầu hết các gen chuyển mã hóa các protein diệt côn trùng trong các cây trồng chuyển gen thương mại được thúc đẩy bởi các vùng khởi động cấu thành, nghĩa là các vùng khởi động mà hoạt động chức năng trong toàn bộ thực vật trong tất cả hoặc phần lớn các loại mô, bao gồm hạt phấn, xuyên suốt toàn bộ chu trình sinh trưởng của thực vật. Do hạt phấn thực vật có thể là nguồn thức ăn cho một số loài côn trùng không phải sinh vật gây hại hoặc được giả thiết rằng hạt phấn thực vật có thể được mang đi bởi gió để ký gửi trên các thực vật chủ của côn trùng không phải sinh vật gây hại, có một số mối quan tâm trong các cơ quan quản lý mà quản lý việc thương mại hóa cây trồng chuyển gen mà các mức cao của việc biểu hiện của các protein diệt côn trùng nhất định, ví dụ các protein Cry nhất định, trong hạt phấn có thể có các tác dụng bất lợi lên các quần thể côn trùng không phải sinh vật gây hại cục bộ. Ngoài ra, đã quan sát thấy rằng việc biểu hiện của các protein diệt côn trùng nhất định trong hạt phấn có các tác động bất lợi lên khả năng sinh sản giống đực của thực vật chuyển gen. Ví dụ, các mức cao của protein diệt côn trùng Vip3 được biểu hiện trong hạt phấn ngô có thể gây ra sự suy giảm về khả năng sinh sản giống đực hoặc tình trạng vô sinh hoàn toàn ở một số nền tảng di truyền cùng dòng nhất định (Patent Mỹ số 10,214,784; được kết hợp ở đây bằng việc viện dẫn). Do đó, cũng là có lợi khi điều hòa việc biểu hiện của các protein diệt côn trùng trong các thực vật chuyển gen, ví dụ, để có các mức cao của sự biểu hiện ở các mô sinh dưỡng, ví dụ mô lá, khi phần lớn các côn trùng là sinh vật gây hại tiêu thụ ban đầu, nhưng có sự biểu hiện giảm ở hạt phấn, mô thực vật mà một số côn trùng không phải sinh vật gây hại có thể tiêu thụ.

Do đó, mong muốn là cung cấp các thực vật, cụ thể là các cây ngô mà loại trừ việc biểu hiện của gen chuyển trong các mô của các cấu trúc sinh sản của thực vật chẳng hạn như hạt phấn và/hoặc cụm hoa. Điều này có thể đạt được trong phạm vi của sáng chế bằng việc cung cấp trình tự nucleotit điều hòa, ít nhất một phần của nó có chức năng khởi đầu phiên mã định hướng việc biểu hiện của polynucleotit được liên kết hoạt động mã hóa protein được quan tâm tới cơ bản tất cả các mô thực vật, nhưng gần như loại bỏ việc biểu hiện trong các mô của các cấu trúc sinh sản đực của thực vật, cụ thể là trong hạt phấn và/hoặc các mô cụm hoa sao cho có rất ít hoặc không có sản phẩm biểu hiện nào hiện diện trong các mô này. Trình tự nucleotit điều hòa sau đó có thể được sử dụng

để phát triển các hệ thống biểu hiện mà cho phép sự tích lũy hiệu quả của protein được quan tâm chẳng hạn như, ví dụ, protein diệt côn trùng, trong các mô mà các sinh vật gây hại mục tiêu thường ăn, và hạn chế hoặc giảm sự tích lũy của protein diệt côn trùng trong các mô hoặc các cơ quan không phải mục tiêu và/hoặc trong các mô mà có thể bị tổn hại bởi protein được quan tâm.

Bản chất kỹ thuật của sáng chế

Sáng chế đề xuất các chế phẩm và các phương pháp điều khiển một cách chọn lọc việc biểu hiện gen chuyển trong các mô thực vật chuyển gen. Cụ thể là, sáng chế đề xuất các polynucleotit vùng khởi động cải tiến có khả năng khởi đầu và/hoặc điều hòa việc phiên mã của polynucleotit ADN mà chúng được liên kết hoạt động. Các vùng khởi động theo sáng chế được đặc trưng bởi khả năng của chúng để điều hòa một cách chọn lọc việc biểu hiện của polynucleotit ADN được liên kết hoạt động bất kỳ, chẳng hạn như polynucleotit mã hóa protein diệt côn trùng, tốt hơn là trong nhiều mô thực vật chẳng hạn như lá, thân, rễ, và tương tự, nhưng không trong các mô sinh sản được chẳng hạn như hạt phấn.

Trong một số khía cạnh, vùng khởi động polynucleotide theo sáng chế bao gồm bất kỳ trong số các SEQ ID NO:1-11 hoặc phần bổ sung của chúng, hoặc đoạn, vùng, thành phần cis, hoặc các polynucleotit liên quan tới bất kỳ trong số các SEQ ID NO:1-11, mà hoạt động chức năng như các vùng khởi động ở các thực vật. Các vùng khởi động theo sáng chế định hướng việc phiên mã chức năng trong mô thực vật bất kỳ ngoài mô sinh sản được, ví dụ hạt phấn, trong đó việc phiên mã bị giới hạn hoặc được giảm đáng kể so sánh với các mô không phải sinh sản được khác của thực vật.

Sáng chế còn đề cập đến các cấu trúc khảm, các catxet biểu hiện và các vectơ bao gồm vùng khởi động theo sáng chế được liên kết hoạt động với dị loại polynucleotit ADN mà mã hóa protein mà được mong muốn là được biểu hiện trong các mô thực vật ngoài hạt phấn. Ví dụ, phân tử ADN tái tổ hợp theo sáng chế có thể bao gồm, theo chiều 5' đến 3', vùng khởi động theo sáng chế được liên kết hoạt động với trình tự ADN mà mã hóa protein diệt côn trùng mà có hoạt tính chống lại các côn trùng gây hại mà ăn các mô thực vật và trình tự vùng kết thúc được liên kết hoạt động với đầu 3' của gen được quan tâm. Trong một số khía cạnh, trình tự vùng kết thúc theo sáng chế bao gồm bất kỳ trong số các SEQ ID NO:12-20 hoặc phần bổ sung của chúng, hoặc đoạn, vùng, hoặc

các trình tự nucleotit liên quan tới bất kỳ trong số các SEQ ID NO:12-20. Trong các khía cạnh khác, vectơ theo sáng chế là vectơ nhị phân hữu dụng cho việc biến nạp thực vật. Trong các khía cạnh khác, vectơ nhị phân bao gồm bất kỳ trong số các SEQ ID NO:21-28 hoặc các trình tự liên quan tới bất kỳ trong số các SEQ ID NO:21-28.

Trong các khía cạnh khác, các trình tự điều hòa theo sáng chế được liên kết hoạt động với trình tự nucleotit mã hóa protein diệt côn trùng. Do đó, khi chèn catxet biểu hiện bao gồm vùng khởi động theo sáng chế được liên kết hoạt động với trình tự mã hóa protein diệt côn trùng vào hệ gen của thực vật, thực vật chuyển gen tạo thành sẽ được bảo vệ khỏi sự tấn công bởi các côn trùng là sinh vật gây hại mà ăn các mô không phải hạt phần, chẳng hạn như lá và thân, nhưng sẽ không ảnh hưởng đến các côn trùng có lợi mà ăn hạt phần.

Sáng chế còn đề cập tới các thực vật chuyển gen, chẳng hạn như các cây ngô chuyển gen, bao gồm vùng khởi động, cấu trúc khảm, catxet biểu hiện hoặc vectơ theo sáng chế.

Sáng chế cũng đề xuất các phương pháp để biểu hiện đặc hiệu trình tự mã hóa di loại trong các mô thực vật chuyển gen ngoại trừ các mô sinh sản định, ví dụ hạt phần, bằng việc kết hợp vào hệ gen của thực vật phân tử ADN tái tổ hợp chứa vùng khởi động theo sáng chế được liên kết hoạt động với polynucleotit ADN mã hóa protein được quan tâm, chẳng hạn như protein diệt côn trùng.

Sáng chế còn đề cập đến phương pháp giảm thiểu sự bất thụ đực trong thực vật chuyển gen, ví dụ cây ngô lai cùng dòng, bằng việc kết hợp vào hệ gen của thực vật phân tử ADN tái tổ hợp chứa vùng khởi động theo sáng chế được liên kết hoạt động với polynucleotit ADN mã hóa protein mà gây ra sự bất thụ đực khi được biểu hiện trong mô sinh sản chẳng hạn như hạt phần. Trong một số khía cạnh, protein là protein diệt côn trùng Vip3 và thực vật chuyển gen là cây ngô lai cùng dòng.

Các khía cạnh trên và các khía cạnh khác theo sáng chế sẽ trở nên rõ ràng từ phần mô tả chi tiết sau đây, các hình vẽ kèm theo và các danh mục trình tự.

Mô tả vắn tắt các trình tự trong danh sách trình tự

SEQ ID NO:1 là trình tự nucleotit của vùng khởi động PMP370-3.

SEQ ID NO:2 là trình tự nucleotit của vùng khởi động PMP393-1.

SEQ ID NO:3 là trình tự nucleotit của vùng khởi động PMP393-2.
SEQ ID NO:4 là trình tự nucleotit của vùng khởi động PMP393-3.
SEQ ID NO:5 là trình tự nucleotit của vùng khởi động PMP393-4.
SEQ ID NO:6 là trình tự nucleotit của vùng khởi động PMP855-1.
SEQ ID NO:7 là trình tự nucleotit của vùng khởi động PMP747-1.
SEQ ID NO:8 là trình tự nucleotit của vùng khởi động PMP004-1.
SEQ ID NO:9 là trình tự nucleotit của vùng khởi động PMP335-1.
SEQ ID NO:10 là trình tự nucleotit của vùng khởi động PMP722-1.
SEQ ID NO:11 là trình tự nucleotit của vùng khởi động PMP948-2.
SEQ ID NO:12 là trình tự nucleotit của vùng kết thúc t370-2.
SEQ ID NO:13 là trình tự nucleotit của vùng kết thúc t393-1.
SEQ ID NO:14 là trình tự nucleotit của vùng kết thúc t393-2.
SEQ ID NO:15 là trình tự nucleotit của vùng kết thúc t855-1.
SEQ ID NO:16 là trình tự nucleotit của vùng kết thúc t747-1.
SEQ ID NO:17 là trình tự nucleotit của vùng kết thúc t004-1.
SEQ ID NO:18 là trình tự nucleotit của vùng kết thúc t335-1.
SEQ ID NO:19 là trình tự nucleotit của vùng kết thúc t722-1.
SEQ ID NO:20 là trình tự nucleotit của vùng kết thúc t948-2.
SEQ ID NO:21 là trình tự nucleotit của vector nhị phân pSYN18499.
SEQ ID NO:22 là trình tự nucleotit của vector nhị phân pSYN18500.
SEQ ID NO:23 là trình tự nucleotit của vector nhị phân pSYN18501.
SEQ ID NO:24 là trình tự nucleotit của vector nhị phân pSYN18498.
SEQ ID NO:25 là trình tự nucleotit của vector nhị phân pSYN18617.
SEQ ID NO:26 là trình tự nucleotit của vector nhị phân pSYN18618.
SEQ ID NO:27 là trình tự nucleotit của vector nhị phân pSYN18619.
SEQ ID NO:28 là trình tự nucleotit của vector nhị phân pSYN18705.

SEQ ID NO:29 là trình tự nucleotit của catxet biểu hiện PMP393-4:Cry1Ai.

Mô tả chi tiết sáng chế

Phần mô tả này không nhằm để trở nên một danh mục chi tiết về tất cả các cách khác nhau mà sáng chế có thể được thực hiện, hoặc tất cả các dấu hiệu mà có thể được bổ sung vào sáng chế này. Ví dụ, các dấu hiệu được minh họa theo một phương án có thể được kết hợp vào các phương án khác, và các dấu hiệu được mô tả theo phương án cụ thể có thể được xóa khỏi phương án đó. Do đó, sáng chế dự kiến rằng theo một số phương án theo sáng chế, dấu hiệu bất kỳ hoặc kết hợp của các dấu hiệu được quy định ở đây có thể được loại trừ hoặc được bỏ qua. Ngoài ra, nhiều cải biến và bổ sung cho các phương án khác nhau được đề nghị ở đây sẽ là rõ ràng với người có trình độ trung bình trong lĩnh vực dựa trên phần bộc lộ này, mà không lệch khỏi sáng chế này. Do đó, các phần mô tả sau đây được nhằm để minh họa một số phương án cụ thể theo sáng chế, và không chỉ định đầy đủ tất cả các hoán vị, kết hợp và biến thể của chúng.

Trừ khi được định nghĩa khác đi, tất cả các thuật ngữ kỹ thuật và khoa học được sử dụng ở đây có cùng ý nghĩa như được hiểu thông thường bởi người có trình độ trung bình trong lĩnh vực kỹ thuật mà sáng chế đề cập đến. Thuật ngữ được sử dụng trong bản mô tả theo sáng chế ở đây là chỉ nhằm mục đích mô tả các phương án cụ thể và không nhằm mục đích giới hạn sáng chế. Các định nghĩa của các thuật ngữ thông thường trong sinh học phân tử có thể được tìm thấy trong Benjamin Lewin, Genes V, được xuất bản bởi Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew et al. (eds.), The Encyclopedia of Molecular Biology, được xuất bản bởi Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); và Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, được xuất bản bởi VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Tất cả các công bố, các đơn sáng chế, các patent và các tham chiếu khác được trích dẫn ở đây được kết hợp bằng việc tham chiếu toàn bộ các hướng dẫn liên quan tới câu và/hoặc đoạn mà tài liệu tham chiếu được trình bày.

Các trình tự nucleotit được cung cấp ở đây được thể hiện theo chiều 5' đến 3', từ trái sang phải và được thể hiện sử dụng mã tiêu chuẩn để thể hiện các bazơ nucleotit như được quy định trong 37 CFR §§1.821 - 1.825 và tiêu chuẩn cơ quan sở hữu trí tuệ thế giới (WIPO) ST.25, ví dụ: adenin (A), cytosin (C), thymine (T), và guanin (G).

Các axit amin được chỉ định tương tự bằng cách sử dụng tiêu chuẩn WIPO ST.25, ví dụ: alanin (Ala; A), arginin (Arg; R), asparagin (Asn; N), axit aspartic (Asp; D), cystein (Cys; C), glutamin (Gln; Q), axit glutamic (Glu; E), glyxin (Gly; G), histidin (His; H), isoleuxin (Ile; I), leuxin (Leu; L), lysin (Lys; K), methionin (Met; M), phenylalanin (Phe; F), prolin (Pro; P), serin (Ser; S), threonin (Thr; T), tryptophan (Trp; W), tyrosin (Tyr; Y), và valin (Val; V).

Trừ khi ngữ cảnh nêu rõ khác đi, được dự định cụ thể rằng các dấu hiệu khác nhau theo sáng chế được mô tả ở đây có thể được sử dụng theo sự kết hợp bất kỳ. Hơn thế nữa, sáng chế cũng dự kiến rằng theo một số phương án của sáng chế, dấu hiệu bất kỳ hoặc kết hợp của các dấu hiệu được quy định ở đây có thể được loại trừ hoặc được bỏ qua. Để minh họa, nếu phần mô tả nêu rằng chế phẩm bao gồm các thành phần A, B và C, thì mục đích cụ thể là bất kỳ thành phần nào trong số A, B hoặc C, hoặc tổ hợp của chúng, có thể được bỏ qua và được từ chối một cách đơn lẻ hoặc theo sự kết hợp bất kỳ.

Để rõ ràng, các thuật ngữ nhất định được sử dụng trong bản mô tả được định nghĩa và trình bày như sau:

Như được sử dụng ở đây và trong các yêu cầu bảo hộ kèm theo, các mạo từ dạng số ít bao gồm cả dạng số nhiều trừ khi ngữ cảnh nêu rõ khác đi. Do đó, ví dụ, việc tham chiếu tới "thực vật" là tham chiếu tới một hoặc nhiều thực vật và bao gồm các tương đương của nó đã biết với người có trình độ trung bình trong lĩnh vực, và tương tự.

Như được sử dụng ở đây, thuật ngữ "và/hoặc" là để chỉ và bao gồm bất kỳ và tất cả các kết hợp có thể của một hoặc nhiều các mục được liệt kê có liên quan, cũng như việc thiếu các kết hợp khi được diễn giải theo cách thay thế, "hoặc."

Thuật ngữ "khoảng" được sử dụng ở đây có nghĩa xấp xỉ, khoảng, xung quanh, hoặc trong khoảng. Khi thuật ngữ "khoảng" được sử dụng kết hợp với khoảng số học, nó cải biến khoảng đó bằng việc mở rộng các biên trên và dưới các giá trị số học được quy định. Nói chung, thuật ngữ "khoảng" được sử dụng ở đây để cải biến giá trị số học trên và dưới giá trị được nêu theo phương sai 20 phần trăm, tốt hơn là 10 phần trăm trên hoặc dưới (cao hơn hoặc thấp hơn). Liên quan đến nhiệt độ, thuật ngữ "khoảng" có nghĩa là $\pm 1^{\circ}\text{C}$, tốt hơn là $\pm 0,5^{\circ}\text{C}$. Khi thuật ngữ "khoảng" được sử dụng trong bối cảnh của

sáng chế này (ví dụ, trong sự kết hợp với nhiệt độ hoặc các giá trị khối lượng phân tử) giá trị chính xác (nghĩa là, không có “khoảng”) được ưu tiên.

Như được sử dụng ở đây, thuật ngữ "được khuếch đại" có nghĩa việc cấu trúc nhiều bản sao của phân tử axit nucleic hoặc nhiều bản sao bổ sung cho phân tử axit nucleic sử dụng ít nhất một trong các phân tử axit nucleic làm khuôn. Các hệ thống khuếch đại bao gồm hệ thống phản ứng chuỗi polymeraza (PCR), hệ thống phản ứng chuỗi ligaza (LCR), khuếch đại dựa trên trình tự axit nucleic (NASBA, Cangene, Mississauga, Ontario), các hệ thống Q-Beta Replicaza, hệ thống khuếch đại trên cơ sở phiên mã (TAS), và việc khuếch đại thay thế chuỗi (SDA). Xem, ví dụ, *Diagnostic Molecular Microbiology: Principles and Applications*, PERSING et al., Ed., American Society for Microbiology, Washington, D.C. (1993). Sản phẩm khuếch đại được đặt tên với thuật ngữ “đơn vị siêu sao chép (amplicon).”

"Hoạt tính" của protein diệt côn trùng có nghĩa là protein diệt côn có chức năng như một chất kiểm soát côn trùng hoạt tính đường miệng, có tác dụng gây độc, và/hoặc có thể làm gián đoạn hoặc ngăn chặn việc ăn của côn trùng, mà có thể hoặc có thể không gây ra sự chết của côn trùng. Khi protein diệt côn trùng được cung cấp cho côn trùng, kết quả thường là sự chết của côn trùng, hoặc côn trùng không ăn nguồn cung cấp protein diệt côn trùng sẵn có với côn trùng. “Diệt sinh vật gây hại” được xác định là hoạt tính sinh vật gây độc có khả năng kiểm soát sinh vật gây hại, chẳng hạn như côn trùng, giun tròn, nấm, vi khuẩn, hoặc virus, tốt hơn bằng việc giết hoặc tiêu diệt chúng. "Diệt côn trùng" được định nghĩa là hoạt tính sinh vật gây độc có khả năng kiểm soát côn trùng, tốt hơn là bằng việc tiêu diệt chúng. “Chất diệt sinh vật gây hại” là chất mà có hoạt tính diệt sinh vật gây hại. “Chất diệt côn trùng gây hại” là chất mà có hoạt tính diệt côn trùng gây hại.

Như được sử dụng ở đây, các thuật ngữ "lai ngược" và "sự lai ngược" là để chỉ quá trình bởi đó thực vật thể hệ con được lai ngược với một trong các bố mẹ của nó trong một hoặc nhiều thế hệ (ví dụ, 2 hoặc nhiều hơn, 3 hoặc nhiều hơn, 4 hoặc nhiều hơn, 5 hoặc nhiều hơn, 6 hoặc nhiều hơn, hoặc 7 hoặc nhiều lần, v.v.). Trong sơ đồ lai ngược, bố mẹ "đối tượng cho" là để chỉ cây bố mẹ với gen mong muốn hoặc cấu trúc ADN hoặc locus cần được thâm nhập. Bố mẹ "đối tượng nhận" (được sử dụng một hoặc nhiều lần) hoặc bố mẹ "lặp lại" (được sử dụng hai hoặc nhiều lần) là để chỉ cây bố mẹ mà vào đó gen hoặc cấu trúc ADN hoặc locus được thâm nhập. Ví dụ, xem Ragot et al.

Marker-assisted Backcrossing: A Practical Example, trong TECHNIQUES ET UTILISATIONS DES MARQUEURS MOLECULAIRES LES COLLOQUES, Vol. 72, pp. 45-56 (1995); và Openshaw et al., Marker-assisted Selection in Backcross Breeding, trong PROCEEDINGS OF THE SYMPOSIUM "ANALYSIS OF MOLECULAR MARKER DATA," pp. 41-43 (1994). Phép lai ban đầu tạo nên sự tăng của thế hệ F1. Thuật ngữ "BC1" là để chỉ lần sử dụng thứ hai của cha mẹ lặp lại, "BC2" là để chỉ lần sử dụng thứ ba của cha mẹ lặp lại, và tiếp tục như vậy. Theo các phương án, ít nhất một hoặc nhiều thế hệ con được xác định và/hoặc được lựa chọn cho sự có mặt của gen hoặc locus mong muốn (ví dụ, trong mẫu axit nucleic từ cây thế hệ con hoặc bộ phận của cây). Theo các phương án, hai hoặc nhiều thế hệ (hoặc thậm chí tất cả các thế hệ) của thế hệ con được xác định và/hoặc được lựa chọn cho sự có mặt của gen hoặc cấu trúc ADN hoặc locus mong muốn.

Thuật ngữ “cấu trúc khảm” hoặc “gen khảm” hoặc “polynucleotit khảm” hoặc “axit nucleic khảm” hoặc “protein khảm” (hoặc các thuật ngữ tương tự) như được sử dụng ở đây là để chỉ cấu trúc hoặc phân tử axit nucleic hoặc protein bao gồm hai hoặc nhiều polynucleotit hoặc môtip axit amin hoặc vùng axit amin, một cách tương ứng, từ các nguồn gốc khác nhau được tập hợp thành một phân tử axit nucleic hoặc protein. Thuật ngữ “cấu trúc khảm”, “gen khảm”, “polynucleotit khảm” hoặc “axit nucleic khảm” là để chỉ cấu trúc hoặc phân tử bất kỳ mà chứa, không giới hạn, (1) các polynucleotit (ví dụ, ADN), bao gồm các polynucleotit điều hòa và mã hóa mà không được tìm thấy cùng nhau trong tự nhiên (nghĩa là, ít nhất một trong các polynucleotit trong cấu trúc là dị loại đối với ít nhất một trong số các polynucleotit khác của nó), hoặc (2) các polynucleotit mã hóa các phần của protein không liên kề trong tự nhiên, hoặc (3) các phần của các vùng khởi động mà không liên kề trong tự nhiên. Ngoài ra, cấu trúc khảm, gen khảm, polynucleotit khảm hoặc axit nucleic khảm có thể bao gồm các polynucleotit điều hòa và các polynucleotit mã hóa mà có nguồn gốc từ các nguồn khác nhau, hoặc bao gồm các polynucleotit điều hòa và các polynucleotit mã hóa có nguồn gốc từ cùng nguồn, nhưng được bố trí theo cách khác với cách được tìm thấy trong tự nhiên. Theo một số phương án theo sáng chế, cấu trúc khảm, gen khảm, polynucleotit khảm hoặc axit nucleic khảm bao gồm catxet biểu hiện bao gồm polynucleotit theo sáng chế dưới sự kiểm soát của các polynucleotit điều hòa, cụ thể là dưới sự kiểm soát của các polynucleotit điều hòa hoạt động chức năng ở các thực vật hoặc vi khuẩn.

"Trình tự mã hóa" là trình tự axit nucleic mà được phiên mã thành ARN chẳng hạn như mARN, rARN, tARN, snARN, ARN có nghĩa hoặc ARN đối nghĩa. Trong một số khía cạnh, ARN sau đó được dịch mã trong sinh vật, chẳng hạn như cây ngô, để sản xuất protein, ví dụ protein diệt côn trùng theo sáng chế. Trong các khía cạnh khác, ARN không được dịch mã để sản xuất protein nhưng đóng vai trò là phân tử ARN để điều hòa việc biểu hiện của protein diệt côn trùng tái tổ hợp theo sáng chế.

Như được sử dụng ở đây, trình tự "được tối ưu hóa codon" có nghĩa là trình tự nucleotit trong đó các codon được chọn để phản ánh sự lệch codon cụ thể mà sinh vật hoặc tế bào chủ có thể có. Điều này được thực hiện thường theo cách mà bảo tồn trình tự axit amin của polypeptit được mã hóa bởi trình tự nucleotit cần được tối ưu hóa. Theo các phương án nhất định, trình tự ADN của kết cấu ADN tái tổ hợp theo sáng chế bao gồm các codon được tối ưu hóa cho tế bào (ví dụ, tế bào động vật, thực vật, hoặc nấm) trong đó cấu trúc cần được biểu hiện. Ví dụ, cấu trúc cần được biểu hiện trong tế bào thực vật có thể có tất cả hoặc các phần của codon trình tự của nó (ví dụ, thành phần ức chế gen thứ nhất hoặc thành phần biểu hiện gen) được tối ưu hóa cho việc biểu hiện ở thực vật. Xem, ví dụ, Bảng sáng chế Mỹ Số 6,121,014, được kết hợp ở đây bằng việc viện dẫn.

Các thuật ngữ "bao gồm" hoặc "gồm," khi được sử dụng trong bản mô tả này, thể hiện sự có mặt của các dấu hiệu, các phần nguyên, các bước, các hoạt động, các phần tử, hoặc các thành phần được nêu, nhưng không loại trừ sự có mặt hoặc bổ sung của một hoặc nhiều các dấu hiệu, các phần nguyên, các bước, các hoạt động, các phần tử, hoặc các thành phần hoặc các nhóm của chúng.

Như được sử dụng ở đây, cụm từ chuyển tiếp "về cơ bản bao gồm" (và các biến thể ngữ pháp) có nghĩa là phạm vi của yêu cầu bảo hộ được hiểu là để bao trùm các nguyên liệu hoặc các bước cụ thể được nêu trong yêu cầu bảo hộ" và các mục không làm thay đổi (các) đặc tính cơ bản và mới" của sáng chế được yêu cầu bảo hộ. Do đó, thuật ngữ "về cơ bản bao gồm" khi được sử dụng trong yêu cầu bảo hộ của sáng chế không nhằm mục đích được hiểu là tương đương với "bao gồm."

"Kiểm soát" các côn trùng có nghĩa là ức chế, thông qua tác dụng gây độc, khả năng của các côn trùng gây hại trong việc sống sót, sinh trưởng, ăn, hoặc sinh sản, hoặc để hạn chế hư hại hoặc mất mát của các cây trồng liên quan đến côn trùng hoặc để bảo

vệ tiềm năng năng suất của cây trồng khi sinh trưởng trong sự có mặt của các côn trùng gây hại. "Kiểm soát" các côn trùng có thể có hoặc có thể không có nghĩa là tiêu diệt các côn trùng, mặc dù nó ưu tiên có nghĩa là tiêu diệt các côn trùng.

Như được sử dụng ở đây, thuật ngữ "ngô" là đồng nghĩa với thuật ngữ "bắp" hoặc "*Zea mays*."

Như được sử dụng ở đây, các thuật ngữ "lai chéo" hoặc "được lai chéo" là để chỉ sự hợp nhất của các giao tử thông qua thụ phấn để tạo ra thế hệ con (ví dụ, các tế bào, các hạt hoặc các thực vật). Thuật ngữ bao gồm cả lai chéo hữu tính (sự thụ phấn của cây này với cây khác) và tự thụ phấn (tự thụ phấn, ví dụ, khi hạt phấn và noãn từ cùng một cây). Thuật ngữ "lai chéo" là để chỉ hoạt động hợp nhất các giao tử thông qua thụ phấn để sản xuất thế hệ con.

"Exon" là để chỉ phần của ADN mà mang trình tự mã hóa cho protein hoặc một phần của nó. Các exon được phân tách bởi các trình tự không mã hóa, can thiệp (các intron). Cho các mục đích của sáng chế, định nghĩa của thuật ngữ "exon" bao gồm các cải biến với trình tự nucleotit của exon có nguồn gốc từ gen mục tiêu, với điều kiện exon được cải biến không làm giảm đáng kể hoạt động của trình tự điều hòa 5' có liên quan.

"Catxet biểu hiện" như được sử dụng ở đây có nghĩa là trình tự axit nucleic có khả năng định hướng việc biểu hiện của trình tự nucleotit cụ thể trong tế bào chủ thích hợp, bao gồm vùng khởi động được liên kết hoạt động với trình tự nucleotit được quan tâm mà được liên kết hoạt động với các tín hiệu kết thúc. Cũng thường bao gồm các trình tự cần thiết cho việc dịch mã phù hợp của trình tự nucleotit. Catxet biểu hiện bao gồm trình tự nucleotit được quan tâm có thể có ít nhất một trong các thành phần của nó dị loại đối với ít nhất một trong các thành phần khác của nó. Catxet biểu hiện cũng có thể xuất hiện trong tự nhiên nhưng đã được thu trong dạng tái tổ hợp hữu dụng cho việc biểu hiện dị loại. Việc sử dụng như vậy của catxet biểu hiện khiến nó không xuất hiện một cách tự nhiên trong tế bào mà nó đã được đưa vào. Tuy nhiên, thông thường, catxet biểu hiện là dị loại đối với vật chủ, nghĩa là, trình tự axit nucleic cụ thể của catxet biểu hiện không xuất hiện một cách tự nhiên trong tế bào vật chủ và phải được đưa vào tế bào vật chủ hoặc tổ tiên của tế bào vật chủ bởi quá trình biến nạp. Việc biểu hiện của trình tự nucleotit trong catxet biểu hiện có thể nằm dưới sự kiểm soát của vùng khởi động cấu thành hoặc của vùng khởi động cảm ứng mà bắt đầu phiên mã chỉ khi tế bào

vật chủ được tiếp xúc với một số kích thích bên ngoài cụ thể. Trong trường hợp sinh vật đa bào, chẳng hạn như thực vật, vùng khởi động có thể cũng là đặc hiệu cho mô, hoặc cơ quan, hoặc giai đoạn phát triển cụ thể.

Catxet biểu hiện có thể cũng tùy chọn bao gồm vùng kết thúc phiên mã và/hoặc dịch mã (nghĩa là, vùng kết thúc) mà hoạt động chức năng ở các thực vật. Nhiều các vùng kết thúc phiên mã khác nhau là sẵn có cho việc sử dụng trong các catxet biểu hiện và chịu trách nhiệm cho việc kết thúc phiên mã ngoài trình tự nucleotit dị loại được quan tâm và hiệu chỉnh việc polyadenyl hóa mARN. Vùng kết thúc có thể có nguồn gốc từ vùng khởi động phiên mã, có thể có nguồn gốc từ trình tự nucleotit được liên kết hoạt động được quan tâm, có thể có nguồn gốc từ vật chủ thực vật, hoặc có thể có nguồn gốc từ nguồn khác (nghĩa là, ngoại lai hoặc dị loại với vùng khởi động, trình tự nucleotit được quan tâm, vật chủ thực vật, hoặc kết hợp bất kỳ của chúng). Các vùng kết thúc phiên mã thích hợp bao gồm, nhưng không bị giới hạn ở, vùng kết thúc CAMV 35S, vùng kết thúc tml, vùng kết thúc synthaza nopalin và/hoặc vùng kết thúc rbcs E9 ở đậu. Chúng có thể được sử dụng trong cả các cây một lá mầm và các cây hai lá mầm. Ngoài ra, vùng kết thúc phiên mã nguyên bản của trình tự mã hóa có thể được sử dụng. Vùng kết thúc sẵn có bất kỳ đã biết là hoạt động chức năng ở các thực vật có thể được sử dụng trong phạm vi của sáng chế.

Thuật ngữ "việc biểu hiện" khi được sử dụng với tham chiếu tới polynucleotit, chẳng hạn như gen, khung đọc mở (ORF) hoặc một phần của nó, hoặc gen chuyển ở các thực vật, là để chỉ quá trình chuyển đổi thông tin di truyền được mã hóa trong gen vào ARN (ví dụ, mARN, rARN, tARN, hoặc snARN) thông qua "việc phiên mã" của gen (nghĩa là, thông qua hoạt động enzym của ARN polymeraza), và vào protein khi có thể áp dụng (ví dụ nếu gen mã hóa protein), thông qua "việc dịch mã" của mARN. Sự biểu hiện gen có thể được điều hòa ở nhiều giai đoạn trong quá trình. Ví dụ, trong trường hợp các cấu trúc dsARN hoặc đối nghĩa, một cách tương ứng, việc biểu hiện có thể có nghĩa là việc phiên mã của chỉ ARN đối nghĩa hoặc chỉ dsARN. Theo một số phương án theo sáng chế, "việc biểu hiện" là để chỉ việc dịch mã và sự tích lũy ổn định của (mARN) có nghĩa hoặc ARN chức năng. Theo một số phương án theo sáng chế, "việc biểu hiện" là để chỉ việc sản xuất của protein.

Như được sử dụng ở đây, thuật ngữ "được phiên mã chức năng" có nghĩa là việc biểu hiện của polynucleotit điều hòa theo sáng chế sản xuất protein được quan tâm với

lượng trao cho mô trong đó protein được tạo ra kiểu hình cho mục đích dự định của nó. Ví dụ, không giới hạn, polynucleotit mã hóa protein diệt côn trùng, mà “được phiên mã chức năng” trong mô lá của thực vật chuyển gen, ví dụ cây ngô chuyển gen, tạo ra số lượng của protein diệt côn trùng mà gây độc với sinh vật có hại là côn trùng mà ăn mô lá trong đó protein diệt côn trùng được tạo ra. Mặt khác, thuật ngữ “được phiên mã chức năng” có nghĩa là số lượng của protein được tạo ra trong mô bất kỳ của thực vật chuyển gen là ít nhất 50 ng/mg tổng protein tan được (TSP), hoặc ít nhất 100 ng/mg TSP, hoặc ít nhất 500 ng/mg TSP, hoặc ít nhất 800 ng/mg TSP, hoặc ít nhất 1000 ng/mg TSP, hoặc ít nhất 2000 ng/mg TSP hoặc ít nhất 3000 ng/mg TSP. Theo một số phương án theo sáng chế, polynucleotit được liên kết hoạt động với vùng khởi động theo sáng chế có thể “được phiên mã chức năng” trong các mô không phải hạt phấn và được phiên mã ở các mức rất thấp trong mô hạt phấn của thực vật chuyển gen và do đó tạo ra mức rất thấp của protein trong mô hạt phấn. Lượng rất thấp như vậy của protein có thể nằm trong khoảng từ 1 ng/mg TSP đến khoảng 15 ng/mg TSP. Các mức độ biểu hiện thấp như vậy không có nghĩa là polynucleotit “được phiên mã chức năng” trong hạt phấn, cụ thể là so sánh với các mô không phải hạt phấn khác trong cùng thực vật chuyển gen. Theo các phương án khác, mức của protein được tạo ra trong mô không phải hạt phấn của thực vật chuyển gen, ví dụ cây ngô chuyển gen, là ít nhất cao hơn 100 lần so với trong mô hạt phấn trong cùng thực vật chuyển gen, hoặc ít nhất cao hơn 200 lần, hoặc ít nhất cao hơn 300 lần, hoặc ít nhất cao hơn 400 lần, hoặc ít nhất cao hơn 500 lần, hoặc ít nhất cao hơn 600 lần, hoặc ít nhất cao hơn 700 lần, hoặc ít nhất cao hơn 800 lần, hoặc ít nhất cao hơn 900 lần hoặc ít nhất cao hơn 1000 lần so với trong mô hạt phấn của cùng thực vật chuyển gen.

"Gen" là vùng được xác định mà nằm trong hệ gen và bao gồm trình tự axit nucleic mã hóa và thường cũng bao gồm các axit nucleic khác, chủ yếu điều hòa, chịu trách nhiệm cho việc kiểm soát sự biểu hiện, mà có nghĩa là việc phiên mã và dịch mã, của phần mã hóa. Gen có thể cũng bao gồm các trình tự kết thúc và các trình tự không được dịch mã 5' đến 3' khác. Các thành phần khác mà có thể có mặt là, ví dụ, các intron. Trình tự axit nucleic điều hòa của gen thường có thể không được liên kết hoạt động với trình tự axit nucleic được kết hợp như được tìm thấy trong tự nhiên và do đó sẽ là gen khảm.

"Gen được quan tâm" là để chỉ phân tử axit nucleic bất kỳ mà, khi được biến nạp vào thực vật, trao cho thực vật tính trạng được mong muốn chẳng hạn như kháng thuốc kháng sinh, kháng virus, kháng côn trùng, kháng bệnh, hoặc kháng các sinh vật gây hại khác, dung hòa thuốc diệt cỏ, dung hòa căng thẳng phi sinh học, vô sinh ở giống đực, chuyển hóa axit béo được cải biến, chuyển hóa cacbohydrat được cải biến, cải thiện giá trị dinh dưỡng, cải thiện hiệu quả trong quá trình xử lý công nghiệp hoặc khả năng sinh sản được thay đổi. "Gen được quan tâm" cũng có thể là gen mà được biến nạp vào các thực vật để tạo ra các chất chuyển hóa hoặc các enzym có giá trị thương mại trong thực vật.

Trình tự axit nucleic hoặc phân tử axit nucleic "dị loại" là trình tự axit nucleic hoặc phân tử axit nucleic không được kết hợp trong tự nhiên với tế bào vật chủ mà vào đó nó được thêm, bao gồm nhiều bản sao không xuất hiện tự nhiên của trình tự axit nucleic xuất hiện trong tự nhiên. Trình tự axit nucleic hoặc phân tử axit nucleic dị loại có thể bao gồm trình tự khảm chẳng hạn như catxet biểu hiện khảm, trong đó vùng khởi động và vùng mã hóa có nguồn gốc từ nhiều nguồn sinh vật. Trình tự vùng khởi động có thể là trình tự vùng khởi động cấu thành, trình tự vùng khởi động đặc hiệu mô, trình tự vùng khởi động một cảm ứng hóa học, trình tự vùng khởi động cảm ứng vết thương, trình tự vùng khởi động cảm ứng căng thẳng, hoặc trình tự vùng khởi động đặc hiệu giai đoạn phát triển.

Trình tự axit nucleic "tương đồng" là trình tự axit nucleic được kết hợp tự nhiên với tế bào vật chủ mà vào đó nó được thêm.

Thuật ngữ "tương đồng" hoặc "giống hệt" hoặc "gần như giống hệt," trong bối cảnh của hai trình tự axit amin hoặc axit nucleic, là để chỉ hai hoặc nhiều trình tự hoặc trình tự con mà có ít nhất 60%, tốt hơn là ít nhất 80%, tốt hơn nữa là 90%, còn tốt hơn nữa là 95%, và tốt nhất là ít nhất 99% độ tương đồng gốc axit amin hoặc nucleotit, khi được so sánh và được căn cho sự tương đương tối đa, như được đo bằng cách sử dụng một trong các thuật toán so sánh trình tự sau đây hoặc bằng việc kiểm tra trực quan. Tốt hơn là, sự xác thực đáng kể tồn tại trên vùng của các trình tự mà ít nhất khoảng 50 gốc hoặc bazơ về độ dài, tốt hơn nữa là trên vùng của ít nhất khoảng 100 gốc hoặc bazơ, và tốt nhất là các trình tự là về cơ bản tương đồng trên ít nhất khoảng 150 gốc hoặc bazơ. Trong phương án đặc biệt ưu tiên, các trình tự là về cơ bản tương đồng trên toàn bộ

chiều dài của các vùng mã hóa. Hơn thế nữa, các trình tự axit amin hoặc axit nucleic gần như giống hệt thực hiện gần như cùng chức năng.

Để so sánh trình tự, một trình tự thường đóng vai trò là trình tự tham chiếu để so sánh các trình tự thử nghiệm. Khi sử dụng thuật toán so sánh trình tự, các trình tự tham chiếu và thử nghiệm được nhập vào máy tính, các tọa độ trình tự con được chỉ định nếu cần, và các thông số chương trình thuật toán trình tự được chỉ định. Thuật toán so sánh trình tự sau đó tính toán phần trăm độ tương đồng trình tự cho (các) trình tự thử nghiệm tương đối so với trình tự tham chiếu, dựa trên các thông số chương trình được chỉ định.

Việc căn chỉnh tối ưu của các trình tự để so sánh có thể được tiến hành, ví dụ, bởi thuật toán tương đồng cục bộ của Smith & Waterman, *Adv. Appl. Math.* 2: 482 (1981), bởi thuật toán căn chỉnh tương đồng của Needleman & Wunsch, *J. Mol. Biol.* 48: 443 (1970), bởi phương pháp tìm kiếm sự tương tự của Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85: 2444 (1988), bởi các cách thực hiện được máy tính hóa của các thuật toán này (GAP, BESTFIT, FASTA, và TFASTA trong gói phần mềm di truyền Wisconsin, nhóm máy tính di truyền (Wisconsin Genetics Software Package, Genetics Computer Group), 575 Science Dr., Madison, Wis.), hoặc bằng việc kiểm tra trực quan (xem thông thường, Ausubel et al., *infra*).

Một ví dụ của thuật toán mà thích hợp cho việc xác định phần trăm độ tương đồng trình tự và độ tương tự trình tự là thuật toán BLAST, mà được mô tả trong Altschul et al., *J. Mol. Biol.* 215: 403-410 (1990). Phần mềm để thực hiện phân tích BLAST có sẵn công khai thông qua Trung Tâm Thông Tin Công Nghệ Sinh Học Quốc Gia (National Center for Biotechnology Information, U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894 USA). Thuật toán này gồm xác định lần thứ nhất các cặp trình tự có điểm cao (HSP) bằng cách xác định các từ ngắn có chiều dài W trong trình tự truy vấn, mà khớp hoặc thỏa mãn một số điểm ngưỡng giá trị dương T khi được sắp xếp thẳng hàng với từ có cùng chiều dài trong trình tự cơ sở dữ liệu. T được đề cập đến dưới dạng ngưỡng điểm từ lân cận (Altschul et al., 1990). Các kết quả so khớp từ lân cận ban đầu này đóng vai trò làm hạt giống để bắt đầu tra cứu tìm kiếm các HSP dài hơn chứa chúng. Các kết quả so khớp từ sau đó được kéo dài theo cả hai hướng dọc theo mỗi trình tự để miễn là điểm sắp xếp thẳng hàng tích lũy có thể được tăng lên. Điểm tích lũy được tính bằng cách sử dụng, đối với trình tự nucleotit, các tham số M (điểm thưởng đối với cặp gốc ăn khớp; luôn > 0) và N (điểm phạt đối với các gốc bất

cặp không khớp; luôn < 0). Đối với trình tự axit amin, ma trận điểm được sử dụng để tính điểm tích lũy. Sự kéo dài của các kết quả so khớp theo mỗi hướng được dừng lại khi điểm sắp xếp thẳng hàng tích lũy giảm đi một lượng X so với giá trị đạt được lớn nhất của nó, điểm tích lũy tiến đến không hoặc thấp hơn, do sự tích lũy của một hoặc nhiều sự sắp xếp thẳng hàng gốc điểm âm, hoặc đạt đến đầu của một trong hai trình tự. Các tham số thuật toán BLAST W , T , và X xác định độ nhạy và tốc độ của sự sắp xếp thẳng hàng. Chương trình BLASTN (dùng cho các trình tự nucleotit) sử dụng như các mặc định độ dài từ (W) là 11, kỳ vọng (E) là 10, ngưỡng là 100, $M=5$, $N=-4$ và so sánh của cả hai chuỗi. Đối với các trình tự axit amin, chương trình BLASTP sử dụng như các mặc định độ dài từ (W) là 3, và kỳ vọng (E) là 10, và ma trận điểm BLOSUM62 (xem Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89: 10915 (1989)).

Ngoài ra để tính toán phần trăm độ tương đồng trình tự, thuật toán BLAST cũng thực hiện việc phân tích thống kê sự tương tự giữa hai trình tự (xem, ví dụ, Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90: 5873-5787 (1993)). Một số đo của độ tương tự cung cấp bởi thuật toán BLAST là khả năng xảy ra tổng nhỏ nhất ($P(N)$), mà cung cấp biểu thị của khả năng xảy ra mà tại đó sự ăn khớp giữa hai trình tự nucleotit hoặc axit amin sẽ xảy ra ngẫu nhiên. Ví dụ, trình tự axit nucleic thử nghiệm được coi là tương tự với trình tự tham chiếu nếu khả năng xảy ra tổng nhỏ nhất khi so sánh trình tự axit nucleic thử nghiệm với trình tự axit nucleic tham chiếu nhỏ hơn khoảng 0,1, tốt hơn là nhỏ hơn khoảng 0,01, và tốt nhất là nhỏ hơn khoảng 0,001.

Một dấu hiệu khác của việc hai trình tự axit nucleic về cơ bản tương đồng là việc hai phân tử lai với nhau trong các điều kiện nghiêm ngặt. Thuật ngữ "lai đặc hiệu với" là để chỉ việc liên kết, tạo bộ đôi, hoặc lai của phân tử chỉ cho trình tự nucleotit cụ thể trong các điều kiện nghiêm ngặt khi trình tự này có mặt trong hỗn hợp phức (ví dụ, toàn bộ tế bào) ADN hoặc ARN. "Về cơ bản liên kết" là để chỉ việc lai bổ sung giữa axit nucleic mẫu dò và axit nucleic mục tiêu và bao gồm những sai lệch nhỏ mà có thể được điều chỉnh bằng cách giảm sự nghiêm ngặt của phương tiện lai để đạt được sự phát hiện trình tự axit nucleic mục tiêu như mong muốn.

"Các điều kiện lai nghiêm ngặt" và "các điều kiện rửa lai nghiêm ngặt" trong bối cảnh các thử nghiệm lai axit nucleic chẳng hạn như việc lai Southern và Northern là phụ thuộc vào trình tự, và khác nhau trong các thông số môi trường khác nhau. Các trình tự dài hơn lai đặc hiệu ở các nhiệt độ cao hơn. Hướng dẫn mở rộng cho việc lai của các

axit nucleic được tìm thấy trong Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology-Hybridization with Nucleic Acid Probes part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays" Elsevier, New York. Thông thường, các điều kiện rửa và lai mức nghiêm ngặt cao được lựa chọn là khoảng 5°C thấp hơn so với điểm nóng chảy nhiệt (T_m) cho trình tự đặc hiệu ở cường độ ion và pH xác định. Thông thường, trong "các điều kiện nghiêm ngặt" mẫu dò sẽ lai với trình tự con mục tiêu của nó, nhưng không với các trình tự khác.

T_m là nhiệt độ (trong cường độ ion và độ pH được xác định) tại đó 50% của trình tự đích lai với đầu dò khớp hoàn hảo. Các điều kiện rất nghiêm ngặt được lựa chọn là bằng với T_m đối với mẫu dò cụ thể. Ví dụ của các điều kiện lai nghiêm ngặt cho việc lai của các axit nucleic bổ sung mà có nhiều hơn 100 gốc bổ sung trên bộ lọc trong Southern hoặc northern blot là 50% formamit với 1 mg của heparin ở 42°C, với việc lai được tiến hành qua đêm. Ví dụ của các điều kiện rửa ở mức nghiêm ngặt cao là NaCl 0,15M ở 72°C trong khoảng 15 phút. Ví dụ của các điều kiện rửa nghiêm ngặt là rửa 0,2× SSC ở 65°C trong 15 phút (xem, Sambrook, *infra*, cho phần mô tả của đệm SSC). Thông thường, việc rửa với mức nghiêm ngặt cao được thực hiện trước việc rửa với mức nghiêm ngặt thấp để loại bỏ tín hiệu đoạn dò nền. Việc rửa với mức nghiêm ngặt trung bình ví dụ là cho việc nhân đôi của, ví dụ, nhiều hơn 100 nucleotit, là 1× SSC ở 45°C trong 15 phút. Việc rửa với mức nghiêm ngặt thấp ví dụ cho việc nhân đôi của, ví dụ, nhiều hơn 100 nucleotit, là 4-6× SSC ở 40°C trong 15 phút. Đối với các đoạn dò ngắn (ví dụ, khoảng 10 đến 50 nucleotit), các điều kiện nghiêm ngặt thường bao gồm các nồng độ muối là nhỏ hơn khoảng 1,0 M ion Na, thường ion Na nồng độ từ khoảng 0,01 đến 1,0 M (hoặc các muối khác) ở mức pH 7,0 đến 8,3, và nhiệt độ thường là ít nhất khoảng 30°C. Các điều kiện nghiêm ngặt cũng có thể đạt được với việc bổ sung của các chất làm mất ổn định chẳng hạn như formamit. Nói chung, tỷ lệ tín hiệu với nhiễu là 2× (hoặc cao hơn) so với được quan sát cho đoạn dò không liên quan trong thử nghiệm lai cụ thể chỉ báo sự phát hiện của việc lai đặc hiệu. Các axit nucleic mà không lai với nhau trong các điều kiện nghiêm ngặt vẫn gần như giống hệt nhau nếu các protein mà chúng mã hóa là gần như giống hệt nhau. Điều này xảy ra, ví dụ, khi bản sao của axit nucleic được tạo ra sử dụng sự thoái hóa codon tối đa được cho phép bởi mã di truyền.

Dưới đây là các ví dụ của các tập điều kiện rửa/lai mà có thể được sử dụng để nhân dòng các trình tự nucleotit tương đồng mà gần như giống hệt với các trình tự

nucleotit tham chiếu của sáng chế: trình tự nucleotit tham chiếu tốt hơn là lai với trình tự nucleotit tham chiếu trong natri dodecyl sulfat (SDS) 7%, NaPO_4 0,5 M, EDTA 1 mM ở 50°C với việc rửa trong $2\times$ SSC, SDS 0,1% ở 50°C, tốt hơn là trong natri dodecyl sulfat (SDS) 7%, NaPO_4 0,5 M, EDTA 1 mM ở 50°C với việc rửa trong $1\times$ SSC, SDS 0,1% ở 50°C, tốt hơn nữa là trong natri dodecyl sulfat (SDS) 7%, NaPO_4 0,5 M, EDTA 1 mM ở 50°C với việc rửa trong $0,5\times$ SSC, 0,1% SDS ở 50°C, tốt hơn là trong natri dodecyl sulfat (SDS) 7%, NaPO_4 0,5 M, EDTA 1 mM ở 50°C với việc rửa trong $0,1\times$ SSC, SDS 0,1% ở 50°C, tốt hơn nữa là trong natri dodecyl sulfat (SDS) 7%, NaPO_4 0,5 M, EDTA 1 mM ở 50°C với việc rửa trong $0,1\times$ SSC, SDS 0,1% ở 65°C.

Một dấu hiệu khác cho thấy hai trình tự axit nucleic hoặc protein là gần như giống hệt nhau là ở chỗ protein được mã hóa bởi axit nucleic thứ nhất phản ứng chéo về mặt miễn dịch với, hoặc liên kết đặc hiệu với, protein được mã hóa bởi axit nucleic thứ hai. Do đó, protein thường là về cơ bản giống với protein thứ hai, ví dụ, trong đó hai protein chỉ khác nhau bởi việc thể bảo toàn.

"Intron" là để chỉ phần can thiệp của ADN mà hầu như chỉ xảy ra trong gen sinh vật nhân chuẩn, nhưng không được dịch mã thành các trình tự axit amin trong sản phẩm gen. Các intron được loại bỏ từ mARN tiền trưởng thành thông qua quá trình được gọi là nối, mà khiến các exon không bị ảnh hưởng, để tạo thành mARN. Cho các mục đích của sáng chế, định nghĩa của thuật ngữ "intron" bao gồm các cải biến với trình tự nucleotit của intron có nguồn gốc từ gen mục tiêu, với điều kiện intron được cải biến không làm giảm đáng kể hoạt động của trình tự điều hòa 5' có liên quan.

Thuật ngữ phân tử axit nucleic, polynucleotit hoặc protein "được phân lập" là phân tử axit nucleic, polynucleotit hoặc protein mà không còn tồn tại trong môi trường tự nhiên của nó. Phân tử axit nucleic được phân lập, polynucleotit hoặc protein theo sáng chế có thể tồn tại ở dạng được tinh sạch hoặc có thể tồn tại trong vật chủ tái tổ hợp chẳng hạn như trong vi khuẩn chuyển gen hoặc thực vật chuyển gen. Do đó, yêu cầu bảo hộ với phân tử axit nucleic "được phân lập", như được liệt kê ở đây, bao gồm phân tử axit nucleic mà được bao gồm trong hệ gen thực vật chuyển gen.

"Phân tử axit nucleic" hoặc "trình tự axit nucleic" là đoạn của ADN hoặc ARN mạch đơn hoặc đôi mà có thể được phân lập từ nguồn bất kỳ. Trong phạm vi theo sáng

chế, phân tử axit nucleic thường là đoạn ADN. Theo một số phương án, các phân tử axit nucleic theo sáng chế là các phân tử axit nucleic được phân lập.

"Được liên kết hoạt động" là để chỉ việc kết hợp của các polynucleotit trên đoạn axit nucleic đơn sao cho chức năng của một đoạn này ảnh hưởng đến chức năng của đoạn khác. Ví dụ, vùng khởi động được liên kết hoạt động với polynucleotit mã hóa hoặc ARN chức năng khi nó có khả năng ảnh hưởng đến việc biểu hiện của polynucleotit mã hóa hoặc ARN chức năng này (nghĩa là, polynucleotit mã hóa hoặc ARN chức năng là dưới sự kiểm soát phiên mã của vùng khởi động). Việc mã hóa các polynucleotit trong định hướng có nghĩa hoặc đối nghĩa có thể được liên kết hoạt động với các polynucleotit điều hòa.

Các thuật ngữ "protein," "peptit" và "polypeptit" được sử dụng thay thế cho nhau ở đây.

"Thực vật" là thực vật bất kỳ ở giai đoạn phát triển bất kỳ, đặc biệt là thực vật cho hạt.

"Tế bào thực vật" là đơn vị cấu trúc và sinh lý của thực vật, bao gồm chất nguyên sinh và thành tế bào. Tế bào thực vật có thể có dạng tế bào đơn được phân lập hoặc tế bào được nuôi cấy, hoặc dưới dạng một phần của đơn vị có tổ chức cao hơn chẳng hạn như, ví dụ, mô thực vật, cơ quan thực vật, hoặc toàn bộ thực vật.

"Môi trường tế bào thực vật" có nghĩa là các môi trường của các đơn vị thực vật chẳng hạn như, ví dụ, chất nguyên sinh, các tế bào môi trường tế bào, các tế bào trong các mô thực vật, hạt phấn, các ống hạt phấn, noãn, túi phôi, hợp tử và phôi ở các giai đoạn phát triển khác nhau.

"Nguyên liệu thực vật" là để chỉ lá, thân, rễ, hoa hoặc các phần hoa, quả, hạt phấn, các tế bào trứng, các hợp tử, các hạt, các phần cắt, các môi trường mô và tế bào, hoặc phần bất kỳ khác hoặc sản phẩm của thực vật.

"Cơ quan thực vật" một phần được biệt hóa và có cấu trúc thấy được và tách biệt của thực vật chẳng hạn như rễ, thân, lá, nụ hoa, hoặc phôi.

"Mô thực vật" như được sử dụng ở đây có nghĩa là nhóm các tế bào thực vật được tổ chức thành đơn vị chức năng và cấu trúc. Mô bất kỳ của thực vật *in planta* hoặc trong môi trường đều được bao gồm. Thuật ngữ này bao gồm, nhưng không bị giới hạn ở, toàn

bộ thực vật, các cơ quan thực vật, các hạt thực vật, môi trường mô và các nhóm bất kỳ của các tế bào thực vật được tổ chức thành các đơn vị chức năng và/hoặc cấu trúc. Việc sử dụng của thuật ngữ này kết hợp với, hoặc trong trường hợp không có, bất kỳ loại mô thực vật cụ thể nào như được liệt kê ở trên hoặc mặt khác được chấp nhận bởi định nghĩa này không nhằm để loại trừ loại bất kỳ khác của mô thực vật. Ví dụ, “tapetum (tế bào dinh dưỡng trong bao phấn)” là mô trong túi bào tử, đặc biệt là bao phấn, của các cây ngô mà cung cấp dinh dưỡng cho sự phát triển bào tử.

“Polynucleotit” là để chỉ polyme được cấu thành từ nhiều nucleotit monome được liên kết cộng hóa trị theo chuỗi. “Các polynucleotit” như vậy bao gồm ADN, ARN, các oligo nucleotit được cải biến (ví dụ, các oligonucleotit bao gồm các bazơ mà không thường cho ARN hoặc ADN sinh học, chẳng hạn như các oligonucleotit được 2'-O-metyl hóa), và tương tự. Theo một số phương án, axit nucleic hoặc polynucleotit có thể là mạch đơn, mạch đôi, mạch đa, hoặc các kết hợp của chúng. Trừ khi được nêu rõ khác đi, axit nucleic hoặc polynucleotit cụ thể theo sáng chế tùy ý bao gồm hoặc mã hóa các polynucleotit bổ sung, bên cạnh polynucleotit bất kỳ được chỉ định rõ ràng.

“Polynucleotit được quan tâm” là để chỉ bất kỳ polynucleotit nào mà, khi được biến nạp vào sinh vật, ví dụ, thực vật, đem lại cho sinh vật đó đặc tính mong muốn chẳng hạn như kháng côn trùng, kháng bệnh, dung hòa thuốc diệt cỏ, kháng thuốc kháng sinh, cải thiện giá trị dinh dưỡng, cải thiện hiệu quả trong quá trình xử lý công nghiệp, việc sản xuất các chất chuyển hóa hoặc các enzym có giá trị thương mại hoặc khả năng sinh sản được thay đổi.

“Vùng khởi đầu” là trình tự ADN không được dịch mã nằm trước của vùng mã hóa mà chứa vị trí gắn cho ARN polymeraza và khởi đầu việc dịch mã của ADN. Vùng khởi động có thể cũng bao gồm các thành phần khác mà đóng vai trò như các vùng điều hòa của sự biểu hiện gen.

“Sự biểu hiện mong muốn” là việc biểu hiện của các sản phẩm gen mà tốt hơn là được biểu hiện ở mức cao hơn trong một hoặc một số các mô thực vật (giới hạn theo không gian) và/hoặc tới một hoặc một số các giai đoạn phát triển thực vật (giới hạn theo thời gian) trong khi trong các giai đoạn phát triển/các mô khác có mức biểu hiện tương đối thấp.

Như được sử dụng ở đây, thuật ngữ "tái tổ hợp" là để chỉ dạng của axit nucleic (ví dụ, ADN hoặc ARN) hoặc protein hoặc cơ quan mà thường sẽ không được tìm thấy trong tự nhiên và như được tạo ra bởi sự can thiệp của con người. Như được sử dụng ở đây, "phân tử axit nucleic tái tổ hợp" là phân tử axit nucleic bao gồm sự tái tổ hợp của các polynucleotit mà sẽ không xảy ra trong tự nhiên cùng với nhau và là kết quả của sự can thiệp của con người, ví dụ, phân tử axit nucleic mà bao gồm sự kết hợp của ít nhất hai polynucleotit dị loại với nhau, hoặc phân tử axit nucleic mà được tổng hợp nhân tạo, ví dụ, tổng hợp polynucleotit sử dụng trình tự nucleotit được lắp ráp, và bao gồm polynucleotit mà lệch với polynucleotit mà thường sẽ tồn tại trong tự nhiên, hoặc phân tử axit nucleic mà bao gồm gen chuyển được kết hợp nhân tạo vào ADN hệ gen của tế bào vật chủ và ADN sườn được kết hợp của hệ gen tế bào vật chủ. Ví dụ khác của phân tử axit nucleic tái tổ hợp là phân tử ADN là kết quả từ việc chèn của gen chuyển vào ADN hệ gen ở thực vật, mà cuối cùng có thể dẫn đến việc biểu hiện của phân tử protein hoặc ARN tái tổ hợp trong cơ quan đó. Như được sử dụng ở đây, "thực vật tái tổ hợp" là thực vật mà thường không xuất hiện trong tự nhiên, là kết quả của sự can thiệp của con người, và chứa gen chuyển hoặc phân tử axit nucleic dị loại được kết hợp vào hệ gen của nó. Như kết quả của sự thay đổi hệ gen này, thực vật tái tổ hợp là khác hẳn với thực vật kiểu dại có liên quan.

"Trình tự điều hòa" hoặc "phần tử điều hòa," và tương tự, được hiểu ở đây là để chỉ trình tự nucleotit mà kiểm soát việc biểu hiện của trình tự mã hóa được kết hợp hoạt động bằng việc cung cấp sự nhận biết ARN polymeraza và các tác nhân khác cần thiết cho việc phiên mã chính xác và thường được đặt phía trước (5') trình tự mã hóa của nó. "Các trình tự điều hòa" bao gồm các trình tự điều hòa 5' nằm gần và các phần tử xa hơn nằm phía trước vùng mã hóa được kết hợp, mà ảnh hưởng đến việc phiên mã, quá trình xử lý ARN hoặc sự ổn định, hoặc việc dịch mã của trình tự mã hóa được kết hợp. "Các trình tự điều hòa" có thể còn bao gồm các trình tự 3', bao gồm các trình tự không phiên mã 3' và/hoặc không dịch mã 3', nằm xuôi chiều vùng mã hóa được kết hợp, và có thể bao gồm vị trí kết thúc phiên mã, nghĩa là các vùng kết thúc. "Các trình tự điều hòa" có thể bao gồm các vùng tăng cường, các vùng khởi động, các trình tự dẫn đầu không được dịch mã, các intron, và các trình tự tín hiệu polyadenyl hóa. Chúng bao gồm các trình tự tự nhiên và tổng hợp cũng như các trình tự mà có thể là sự kết hợp của các trình tự tự nhiên và tổng hợp. "Vùng tăng cường" là trình tự ADN mà có thể kích thích hoạt động

vùng khởi đầu và có thể là phần tử bẩm sinh của vùng khởi động hoặc phần tử dị loại được chèn để tăng cường mức hoặc độ đặc hiệu mô của vùng khởi động. Nó có khả năng hoạt động theo cả hai định hướng (thông thường hoặc đảo ngược), và có khả năng hoạt động chức năng ngay cả khi được di chuyển ngược chiều hoặc xuôi chiều từ vùng khởi động. Các ví dụ của các vùng tăng cường bao gồm, trong số các ví dụ khác, vùng tăng cường virus khảm cây huyền sâm (cF₁ MV) và vùng tăng cường virus khảm cây súp lơ (e35S). Ý nghĩa của thuật ngữ "các trình tự điều hòa" bao gồm các trình tự "khởi đầu phiên mã" hoặc "vùng khởi động" và "các trình tự điều hòa vùng khởi động." Các thuật ngữ này có thể được sử dụng thay thế cho nhau ở đây.

"Các thành phần điều hòa" là để chỉ các trình tự tham gia vào việc kiểm soát việc biểu hiện của trình tự nucleotit. Các thành phần điều hòa bao gồm vùng khởi động được liên kết hoạt động với trình tự nucleotit được quan tâm và các tín hiệu kết thúc. Chúng cũng thường bao gồm các trình tự cần thiết cho việc dịch mã chính xác trình tự nucleotit.

"Biểu hiện đặc hiệu" là việc biểu hiện của các sản phẩm gen mà giới hạn ở một hoặc một số các mô thực vật (giới hạn theo không gian) và/hoặc ở một hoặc một số các giai đoạn phát triển thực vật (giới hạn theo thời gian).

"Vùng khởi động đặc hiệu theo mô" là để chỉ các vùng khởi động được điều hòa mà không được biểu hiện trong tất cả các tế bào thực vật nhưng chỉ trong một hoặc nhiều loại tế bào trong các cơ quan cụ thể (chẳng hạn như lá, rễ hoặc hạt), các mô cụ thể (chẳng hạn như phôi hoặc lá mầm), hoặc các loại tế bào cụ thể (chẳng hạn như nhu mô lá hoặc tế bào lưu trữ hạt giống). Cũng bao gồm các vùng khởi động mà được điều hòa tạm thời, chẳng hạn như trong quá trình tạo phôi sớm hoặc muộn, trong quá trình chín của quả khi hạt hoặc quả đang phát triển, ở lá đã được biệt hóa hoàn toàn, hoặc khi bắt đầu lão hóa.

"Catxet phiên mã" sẽ bao gồm trong hướng 5'-3' của việc dịch mã, vùng khởi động phiên mã và dịch mã, trình tự ADN được quan tâm, và vùng kết thúc phiên mã và dịch mã hoạt động chức năng ở các thực vật. Vùng kết thúc có thể có nguồn gốc từ vùng khởi động phiên mã, có thể có nguồn gốc từ trình tự ADN được quan tâm, hoặc có thể có nguồn gốc từ nguồn khác.

"Vị trí khởi đầu phiên mã" là vị trí xung quanh nucleotit thứ nhất mà là một phần của trình tự được phiên mã, mà cũng được xác định như vị trí +1. Đối với vị trí này tất

cả các trình tự khác của gen và các vùng kiểm soát của nó được đánh số. Các trình tự xuôi dòng (nghĩa là các trình tự protein mã hóa khác theo chiều 3') được coi là dương, trong khi các trình tự ngược dòng (hầu hết các vùng kiểm soát theo chiều 5') được coi là âm.

"Việc biến nạp tạm thời" trong bối cảnh polynucleotit có nghĩa là polynucleotit được đưa vào tế bào và không tích hợp vào hệ gen của tế bào. Việc biến nạp tạm thời có thể được phát hiện bởi, ví dụ, thử nghiệm miễn dịch hấp thụ liên kết enzym (ELISA) hoặc Western blot, mà có thể phát hiện sự có mặt của peptit hoặc polypeptit được mã hóa bởi một hoặc nhiều các phân tử axit nucleic được đưa vào cơ quan sinh vật. Việc biến nạp ổn định của tế bào có thể được phát hiện bởi, ví dụ, thử nghiệm lai Southern blot của ADN hệ gen của tế bào với các trình tự axit nucleic mà lai đặc hiệu với trình tự nucleotit của phân tử axit nucleic được đưa vào cơ quan sinh vật (ví dụ, thực vật). Việc biến nạp ổn định của tế bào có thể được phát hiện bởi, ví dụ, thử nghiệm lai Northern blot của ARN của tế bào với các trình tự axit nucleic mà lai đặc hiệu với trình tự nucleotit của phân tử axit nucleic được đưa vào thực vật hoặc cơ quan sinh vật khác. Việc biến nạp ổn định của tế bào có thể cũng được phát hiện bởi, ví dụ, phản ứng chuỗi polymeraza (PCR) hoặc phản ứng khuếch đại khác như đã biết trong lĩnh vực, sử dụng các trình tự đoạn mồi đặc hiệu mà lai với (các) trình tự mục tiêu của phân tử axit nucleic, dẫn đến việc khuếch đại của (các) trình tự mục tiêu, mà có thể được phát hiện theo các phương pháp tiêu chuẩn. Việc biến nạp có thể cũng được phát hiện bởi các quá trình lai và/hoặc giải trình tự trực tiếp đã biết trong lĩnh vực.

"Việc biến nạp" là quá trình đưa axit nucleic dị loại vào cơ quan sinh vật hoặc tế bào vật chủ. Theo các phương án cụ thể, "việc biến nạp" có nghĩa là sự tích hợp ổn định của phân tử ADN vào hệ gen (nhân hoặc plastit) của cơ quan sinh vật được quan tâm.

"Được biến nạp / chuyển gen / tái tổ hợp" là để chỉ sinh vật chủ chẳng hạn như vi khuẩn hoặc thực vật mà vào đó phân tử axit nucleic dị loại đã được thêm vào. Phân tử axit nucleic có thể được tích hợp một cách ổn định vào hệ gen của vật chủ hoặc phân tử axit nucleic có thể cũng có mặt dưới dạng phân tử ngoài nhiễm sắc thể. Phân tử ngoài nhiễm sắc thể như vậy có thể tự động sao chép. Các tế bào, các mô, hoặc các thực vật được biến nạp được hiểu là bao gồm không chỉ sản phẩm cuối của quá trình biến nạp, mà cả thế hệ con chuyển gen của chúng. Vật chủ "không được biến nạp", "không chuyển

gen", hoặc "không tái tổ hợp" là để chỉ sinh vật kiểu đại, ví dụ, vi khuẩn hoặc thực vật, mà không chứa phân tử axit nucleic dị loại.

“Vectơ” được xác định là bao gồm, không kể những cái khác, plasmit, cosmit, phago hoặc vectơ nhị phân *Agrobacterium* bất kỳ ở dạng vòng hoặc mạch thẳng đơn hoặc đôi mà có thể có hoặc có thể không tự chuyển giao hoặc vận động, và có thể biến nạp vật chủ nhân sơ hoặc nhân thực hoặc bằng việc tích hợp vào hệ gen tế bào hoặc tồn tại ngoài nhiễm sắc thể (ví dụ plasmit tự sao chép với tâm sao chép). Được bao gồm cụ thể là các vectơ con thoi mà có nghĩa là phương tiện ADN có khả năng, tự nhiên hoặc theo thiết kế, sao chép ở hai sinh vật chủ khác nhau, mà có thể được lựa chọn từ các xạ khuẩn và các loài liên quan, vi khuẩn và sinh vật nhân thực (ví dụ các tế bào nấm, nấm men, động vật có vú, thực vật cấp cao hơn).

Sáng chế đề cập chung đến các polynucleotit điều hòa ít nhất một phần của nó có chức năng khởi đầu phiên mã định hướng việc biểu hiện của polynucleotit mã hóa protein được liên kết hoạt động tới hầu hết các mô thực vật, nhưng về cơ bản loại bỏ việc biểu hiện trong các mô sinh sản đực chẳng hạn như hạt phấn và/hoặc mô cụm hoa. Các polynucleotit điều hòa có thể được sử dụng trong các cấu trúc khảm, các catxet biểu hiện, các vectơ tái tổ hợp và tương tự để thúc đẩy việc biểu hiện của protein được quan tâm, ví dụ protein diệt côn trùng, trong các mô thực vật mà các sinh vật gây hại mục tiêu thường ăn, và hạn chế hoặc giảm sự tích lũy của protein diệt côn trùng trong các mô thường không bị ăn bởi các côn trùng là sinh vật gây hại hoặc trong mô mà có thể bị tổn hại bởi protein diệt côn trùng, hạt phấn hoặc cụm hoa này.

Các trình tự vùng khởi động được thu bằng việc nhân dòng các trình tự hệ gen mà tương đồng với các trình tự cADN được biểu hiện trong các mô ngoài hạt phấn, nghĩa là các trình tự cADN “không phải hạt phấn”. Các trình tự hệ gen có thể được thu bằng các phương pháp lai hoặc bằng cách sử dụng các phương pháp PCR để kéo dài trình tự theo hoặc chiều 5’ hoặc 3’ từ trình tự đã biết (đôi khi được gọi là “di chuyển hệ gen”). Ví dụ, để thu được các trình tự hệ gen 5’ tới trình tự đã biết của cADN, các đoạn môi được làm cho trình tự gần đầu 5’ của cADN. Thư viện hệ gen được xây dựng với đầu 5’ của mỗi trình tự ADN hệ gen được ghép nối với vùng tiếp hợp oligonucleotit ngắn. PCR với đoạn môi lai với trình tự tiếp hợp và đoạn môi 5’ của trình tự cADN không phải hạt phấn cho phép việc khuếch đại của trình tự hệ gen ở 5’ với trình tự tương đồng của trình tự không phải hạt phấn. Các trình tự ADN thu được từ hệ gen đi bộ được

xác định trình tự và nếu các vùng 5' bổ sung được mong muốn, quá trình được lặp lại với các đoạn môi hiện tại ở đầu 5' của dòng thu được dài nhất. Các trình tự hệ gen tương đồng với các trình tự cADN không phải hạt phấn cũng được thu bằng việc lai trong các điều kiện mức nghiêm ngặt cao. Các điều kiện mức nghiêm ngặt cao lựa chọn cho việc lai của mẫu dò được làm từ trình tự cADN không phải hạt phấn để lai với trình tự tương đồng của nó trong ADN hệ gen. ADN hệ gen được bao gồm trong thư viện ADN hệ gen của các trình tự ADN hệ gen ở ngò 5-20 kb trong vectơ phagơ lambda. Các dòng hệ gen mà lai với cADN không phải hạt phấn được phân lập và được giải trình tự.

Các vùng khởi động theo sáng chế được nằm trong vùng 5' của trình tự cADN trừ hạt phấn ngay trước trình tự mã hóa. Kích thước của vùng điều hòa tốt hơn là trong khoảng giữa khoảng 2 kb đến 8 kb và bao gồm trình tự không được phiên mã 5', cụ thể là trình tự không được phiên mã 5' và 5'-UTR và tất cả hoặc một phần của trình tự nucleotit thể hiện intron thứ nhất. Các vùng khởi động được lấy ví dụ ở đây như được quy định trong các SEQ ID NO:1-11. Các trình tự điều hòa theo sáng chế có thể còn bao gồm một phần của trình tự 3' mà bắt đầu ngay sau trình tự mã hóa của codon dừng dịch mã cADN trừ hạt phấn bao gồm trình tự được phiên mã nhưng không được dịch mã (UTR) và trình tự không được phiên mã mà đóng vai trò là vùng kết thúc phiên mã và tín hiệu poly-adenyl hóa. Cụ thể là, trình tự 3' là trong khoảng giữa khoảng 1,0 kb và khoảng 2,5 kb. Các vùng kết thúc được lấy ví dụ ở đây như được quy định trong các SEQ ID NO:12-20.

Các dòng hệ gen có thể bao gồm các trình tự intron, không được tìm thấy trong các dòng mARN hoặc cADN. Các trình tự hệ gen có thể còn bao gồm các trình tự không được dịch mã 5', các trình tự không được dịch mã 3', và các trình tự điều hòa 5' và 3'. Các trình tự vùng khởi động được tìm thấy trong trình tự hệ gen 5' tới trình tự cADN. Các trình tự hệ gen được nhân dòng mà tương đồng với các trình tự cADN trừ hạt phấn. Các trình tự mà là 5' với trình tự tương đồng với trình tự cADN mà được đề cập tới ở đây như là vùng sườn 5' mà bao gồm vùng khởi động.

Theo một số phương án, sáng chế đề xuất cấu trúc ADN khả mã bao gồm trình tự mã hóa protein diệt côn trùng được liên kết hoạt động với trình tự điều hòa, ít nhất một phần của nó có chức năng khởi đầu phiên mã định hướng việc biểu hiện của protein diệt côn trùng được mã hóa tới gần như tất cả các mô của thực vật với ngoại trừ các mô sinh sản dục, chẳng hạn như hạt phấn và/hoặc cụm hoa, sao cho có ít hoặc không có sản

phẩm biểu hiện có mặt là hạt phấn và/hoặc các mô cụm hoa tới mức đáng kể bất kỳ. Các trình tự axit nucleic theo sáng chế có thể được cung cấp dưới dạng ADN hoặc dưới dạng ARN, như được cụ thể hóa; việc bộc lộ của một dạng nhất thiết xác định dạng khác, là được biết đến bởi người có trình độ trung bình trong lĩnh vực. Hơn thế nữa, việc bộc lộ của trình tự axit nucleic đã cho nhất thiết xác định sự bổ sung chính xác của trình tự đó, là được biết đến bởi người có trình độ trung bình trong lĩnh vực.

Theo một số phương án, sáng chế đề xuất catxet biểu hiện bao gồm vùng khởi động bao gồm trình tự nucleotit a) có ít nhất 95% độ tương đồng tới ít nhất 99% độ tương đồng với bất kỳ trong số các SEQ ID NO:1-11; hoặc b) của bất kỳ trong số các SEQ ID NO:1-11, được liên kết hoạt động với polynucleotit dị loại được quan tâm, mà được liên kết hoạt động với vùng không được dịch mã 3' bao gồm tín hiệu polyadenyl hóa, trong đó polynucleotit dị loại được dịch mã chức năng trong mô của thực vật chuyển gen mà không phải là hạt phấn. Theo các phương án khác, polynucleotit dị loại mã hóa protein diệt côn trùng hoặc ARN chuỗi kép (dsARN). Theo các phương án khác, protein diệt côn trùng là protein Cry hoặc protein Vip3.

Theo các phương án khác nữa, protein Cry diệt côn trùng được lựa chọn từ nhóm bao gồm Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, Cry1Af, Cry1Ag, Cry1Ah, Cry1Ai, Cry1Aj, Cry1Ba, Cry1Bb, Cry1Bc, Cry1Bd, Cry1Be, Cry1Bf, Cry1Bg, Cry1Bh, Cry1Bi, Cry1Ca, Cry1Cb, Cry1Da, Cry1Db, Cry1Dc, Cry1Dd, Cry1Ea, Cry1Eb, Cry1Fa, Cry1Fb, Cry1Ga, Cry1Gb, Cry1Gc, Cry1Ha, Cry1Hb, Cry1Hc, Cry1Ia, Cry1Ib, Cry1Ic, Cry1Id, Cry1Ie, Cry1If, Cry1Ig, Cry1Ja, Cry1Jb, Cry1Jc, Cry1Jd, Cry1Ka, Cry1La, Cry1Ma, Cry1Na, Cry1Nb, Cry2Aa, Cry2Ab, Cry2Ac, Cry2Ad, Cry2Ae, Cry2Af, Cry2Ag, Cry2Ah, Cry2Ai, Cry2Aj, Cry2Ak, Cry2Al, Cry2Ba, Cry3Aa, Cry3Ba, Cry3Bb, Cry3Ca, Cry4Aa, Cry4Ba, Cry4Ca, Cry4Cb, Cry4Cc, Cry5Aa, Cry5Ab, Cry5Ac, Cry5Ad, Cry5Ba, Cry5Ca, Cry5Da, Cry5Ea, Cry6Aa, Cry6Ba, Cry7Aa, Cry7Ab, Cry7Ac, Cry7Ba, Cry7Bb, Cry7Ca, Cry7Cb, Cry7Da, Cry7Ea, Cry7Fa, Cry7Fb, Cry7Ga, Cry7Gb, Cry7Gc, Cry7Gd, Cry7Ha, Cry7Ia, Cry7Ja, Cry7Ka, Cry7Kb, Cry7La, Cry8Aa, Cry8Ab, Cry8Ac, Cry8Ad, Cry8Ba, Cry8Bb, Cry8Bc, Cry8Ca, Cry8Da, Cry8Db, Cry8Ea, Cry8Fa, Cry8Ga, Cry8Ha, Cry8Ia, Cry8Ib, Cry8Ja, Cry8Ka, Cry8Kb, Cry8La, Cry8Ma, Cry8Na, Cry8Pa, Cry8Qa, Cry8Ra, Cry8Sa, Cry8Ta, Cry9Aa, Cry9Ba, Cry9Bb, Cry9Ca, Cry9Da, Cry9Db, Cry9Dc, Cry9Ea, Cry9Eb, Cry9Ec, Cry9Ed, Cry9Ee, Cry9Fa, Cry9Ga,

Cry10Aa, Cry11Aa, Cry11Ba, Cry11Bb, Cry12Aa, Cry13Aa, Cry14Aa, Cry14Ab, Cry15Aa, Cry16Aa, Cry17Aa, Cry18Aa, Cry18Ba, Cry18Ca, Cry19Aa, Cry19Ba, Cry19Ca, Cry20Aa, Cry20Ba, Cry21Aa, Cry21Ba, Cry21Ca, Cry21Da, Cry21Ea, Cry21Fa, Cry21Ga, Cry21Ha, Cry22Aa, Cry22Ab, Cry22Ba, Cry22Bb, Cry23Aa, Cry24Aa, Cry24Ba, Cry24Ca, Cry25Aa, Cry26Aa, Cry27Aa, Cry28Aa, Cry29Aa, Cry29Ba, Cry30Aa, Cry30Ba, Cry30Ca, Cry30Da, Cry30Db, Cry30Ea, Cry30Fa, Cry30Ga, Cry31Aa, Cry31Ab, Cry31Ac, Cry31Ad, Cry32Aa, Cry32Ab, Cry32Ba, Cry32Ca, Cry32Cb, Cry32Da, Cry32Ea, Cry32Eb, Cry32Fa, Cry32Ga, Cry32Ha, Cry32Hb, Cry32Ia, Cry32Ja, Cry32Ka, Cry32La, Cry32Ma, Cry32Mb, Cry32Na, Cry32Oa, Cry32Pa, Cry32Qa, Cry32Ra, Cry32Sa, Cry32Ta, Cry32Ua, Cry33Aa, Cry34Aa, Cry34Ab, Cry34Ac, Cry34Ba, Cry35Aa, Cry35Ab, Cry35Ac, Cry35Ba, Cry36Aa, Cry37Aa, Cry38Aa, Cry39Aa, Cry40Aa, Cry40Ba, Cry40Ca, Cry40Da, Cry41Aa, Cry41Ab, Cry41Ba, Cry42Aa, Cry43Aa, Cry43Ba, Cry43Ca, Cry43Cb, Cry43Cc, Cry44Aa, Cry45Aa, Cry46Aa, Cry46Ab, Cry47Aa, Cry48Aa, Cry48Ab, Cry49Aa, Cry49Ab, Cry50Aa, Cry50Ba, Cry51Aa, Cry52Aa, Cry52Ba, Cry53Aa, Cry53Ab, Cry54Aa, Cry54Ab, Cry54Ba, Cry55Aa, Cry56Aa, Cry57Aa, Cry57Ab, Cry58Aa, Cry59Aa, Cry59Ba, Cry60Aa, Cry60Ba, Cry61Aa, Cry62Aa, Cry63Aa, Cry64Aa, Cry65Aa, Cry66Aa, Cry67Aa, Cry68Aa, Cry69Aa, Cry69Ab, Cry70Aa, Cry70Ba, Cry70Bb, Cry71Aa, Cry72Aa và Cry73Aa. Theo các phương án khác nữa, protein Cry là protein Cry1. Theo các phương án khác, protein Cry1 là protein Cry1A. Theo các phương án khác nữa, protein Cry1A là protein Cry1Ab hoặc Cry1Ai.

Theo các phương án khác, protein diệt côn trùng Vip3 được mã hóa bởi di loại được liên kết hoạt động với vùng khởi động theo sáng chế được lựa chọn từ nhóm bao gồm Vip3Aa1, Vip3Aa2, Vip3Aa3, Vip3Aa4, Vip3Aa5, Vip3Aa6, Vip3Aa7, Vip3Aa8, Vip3Aa9, Vip3Aa10, Vip3Aa11, Vip3Aa12, Vip3Aa13, Vip3Aa14, Vip3Aa15, Vip3Aa16, Vip3Aa17, Vip3Aa18, Vip3Aa19, Vip3Aa20, Vip3Aa21, Vip3Aa22, Vip3Aa23, Vip3Aa24, Vip3Aa25, Vip3Aa26, Vip3Aa27, Vip3Aa28, Vip3Aa29, Vip3Aa30, Vip3Aa31, Vip3Aa32, Vip3Aa33, Vip3Aa34, Vip3Aa35, Vip3Aa36, Vip3Aa37, Vip3Aa38, Vip3Aa39, Vip3Aa40, Vip3Aa41, Vip3Aa42, Vip3Aa43, Vip3Aa44, Vip3Ab1, Vip3Ab2, Vip3Ac1, Vip3Ad1, Vip3Ad2, Vip3Ae1, Vip3Af1, Vip3Af2, Vip3Af3, Vip3Ag1, Vip3Ag2, Vip3Ag3, HM117633, Vip3Ag4, Vip3Ag5, Vip3Ah1, Vip3Ba1, Vip3Ba2, Vip3Bb1, Vip3Bb2 và Vip3Bb3. Theo các phương án

khác nữa, protein Vip3 là protein Vip3A. Theo các phương án khác, protein Vip3A là protein Vip3Aa. Theo các phương án khác, protein Vip3Aa là protein Vip3Aa19 hoặc Vip3Aa20.

Theo một số phương án, thực vật chuyển gen trong đó polynucleotit dị loại theo sáng chế được dịch mã chức năng là thực vật một lá mầm. Theo các phương án khác, thực vật một lá mầm là cây ngô. Theo các phương án khác, cây ngô là cây ngô lai cùng dòng hoặc cây ngô lai. Theo một số phương án, mô của cây ngô trong đó polynucleotit dị loại được biểu hiện chức năng là lá, râu hoặc bẹ.

Theo một số phương án, vùng không được dịch mã 3' của catxet biểu hiện theo sáng chế bao gồm trình tự vùng kết thúc được lựa chọn từ bất kỳ trong số các SEQ ID NO:12-20. Theo các phương án khác nữa, catxet biểu hiện bao gồm SEQ ID NO:29.

Theo một số phương án, sáng chế đề xuất vectơ tái tổ hợp bao gồm catxet biểu hiện theo sáng chế. Theo các phương án khác, vectơ tái tổ hợp là vectơ nhị phân có khả năng hoạt động chức năng trong nhiều sinh vật. Theo các phương án khác, sinh vật là vi khuẩn hoặc thực vật. Theo các phương án khác, vectơ nhị phân bao gồm bất kỳ trong số các SEQ ID NO:21-28.

Theo một số phương án, sáng chế đề xuất tế bào thực vật được biến nạp với vectơ tái tổ hợp hoặc catxet biểu hiện theo sáng chế. Theo các phương án khác, vectơ tái tổ hợp hoặc catxet biểu hiện được biểu hiện nhất thời trong tế bào thực vật. Theo các phương án khác, vectơ tái tổ hợp hoặc catxet biểu hiện theo sáng chế được tích hợp ổn định vào hệ gen của tế bào thực vật. Theo các phương án khác, tế bào thực vật là tế bào thực vật một lá mầm. Theo các phương án khác nữa, tế bào thực vật một lá mầm là tế bào cây ngô. Theo các phương án khác, catxet biểu hiện bao gồm trình tự của catxet biểu hiện trong bất kỳ trong số các SEQ ID NO:21-28. Theo các phương án khác nữa, catxet biểu hiện bao gồm SEQ ID NO:29.

Theo một số phương án, sáng chế đề xuất thực vật chuyển gen bao gồm catxet biểu hiện hoặc vectơ theo sáng chế. Theo các phương án khác, catxet biểu hiện hoặc vectơ hoặc một phần của nó được tích hợp ổn định vào hệ gen của thực vật chuyển gen. Theo các phương án khác, thực vật chuyển gen là một lá mầm. Theo các phương án khác nữa, thực vật chuyển gen một lá mầm là cây ngô chuyển gen. Theo các phương án khác, cây ngô chuyển gen là cây ngô lai cùng dòng hoặc cây ngô lai. Theo các phương

án khác, cây ngô chuyển gen biểu hiện protein được quan tâm ở các mức chức năng trong mô bất kỳ mà không phải là hạt phấn. Theo các phương án khác nữa, mô trong đó protein được quan tâm được biểu hiện ở các mức chức năng là lá, râu hoặc bẹ.

Theo các phương án khác, protein được quan tâm mà được biểu hiện trong thực vật chuyển gen ở các mức chức năng trong mô bất kỳ mà không phải là hạt phấn là protein Cry hoặc protein Vip3. Theo các phương án khác, protein Cry hoặc protein Vip3 được lựa chọn từ danh sách của các protein Cry và Vip3 được mô tả ở trên. Theo các phương án khác, protein Cry là protein Cry1. Theo các phương án khác, protein Cry1 là protein Cry1A. Theo các phương án khác nữa, protein Cry1A là protein Cry1Ab hoặc protein Cry1Ai. Theo các phương án khác nữa, protein Vip3 là protein Vip3A. Theo các phương án khác, protein Vip3A là protein Vip3Aa. Theo các phương án khác nữa, protein Vip3Aa là protein Vip3Aa19 hoặc Vip3Aa20.

Theo một số phương án, sáng chế đề xuất hạt chuyển gen của thực vật chuyển gen bất kỳ theo sáng chế. Theo các phương án khác, hạt chuyển gen bao gồm catxet biểu hiện hoặc vector, hoặc một phần của nó, theo sáng chế. Theo các phương án khác, hạt chuyển gen là hạt ngô. Theo các phương án khác, hạt ngô đóng vai trò là phương tiện nhân giống ngô. Theo các phương án khác, hạt ngô chuyển gen đóng vai trò là hạt thu hoạch.

Theo một số phương án, sáng chế đề xuất phân tử axit nucleic được phân lập bao gồm trình tự nucleotit được lựa chọn từ nhóm bao gồm: a) trình tự được quy định trong bất kỳ trong số các SEQ ID NO:1-20; b) trình tự nucleotit mà lai trong các điều kiện mức nghiêm ngặt cao với trình tự nucleotit của a); và c) trình tự nucleotit bao gồm đoạn trình tự nucleotit của a), trong đó đoạn duy trì chức năng của trình tự nucleotit của a); trong đó phân tử axit nucleic khởi đầu việc phiên mã chức năng hoặc kết thúc việc phiên mã chức năng của polynucleotit được liên kết hoạt động được quan tâm trong mô của thực vật chuyển gen mà không phải là hạt phấn.

Theo một số phương án, polynucleotit được quan tâm mà được liên kết hoạt động với phân tử axit nucleic được phân lập theo sáng chế mã hóa protein diệt côn trùng hoặc ARN chuỗi kép (dsARN). Theo các phương án khác, protein diệt côn trùng là protein Cry hoặc protein Vip3. Theo các phương án khác, protein Cry hoặc protein Vip3 được lựa chọn từ danh sách của các protein Cry và Vip3 được mô tả ở trên. Theo các phương

án khác, protein Cry là protein Cry1. Theo các phương án khác, protein Cry1 là protein Cry1A. Theo các phương án khác, protein Cry1A là protein Cry1Ab hoặc Cry1Ai. Theo các phương án khác nữa, protein Vip3 là protein Vip3A. Theo một phương án khác, protein Vip3A là protein Vip3Aa. Theo các phương án khác, protein Vip3Aa là protein Vip3Aa19 hoặc Vip3Aa20.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:1. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:1 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:12.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:2. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:2 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO: 13 hoặc SEQ ID NO:14.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:3. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:3 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO: 13 hoặc SEQ ID NO:14.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:4. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:4 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO: 13 hoặc SEQ ID NO:14.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:5 Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:5 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO: 13 hoặc SEQ ID NO:14.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:6. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:6 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:15.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:7. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:7 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:16.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:8. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:8 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:17.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:9. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:9 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:18.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:10. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:10 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:19.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:11. Theo các phương án khác, phân tử axit nucleic được phân lập theo sáng chế bao gồm trình tự nucleotit vùng khởi động mà được quy định trong SEQ ID NO:11 và trình tự nucleotit vùng kết thúc mà được quy định trong SEQ ID NO:20.

Theo một số phương án, phân tử axit nucleic được phân lập theo sáng chế khởi đầu việc phiên mã chức năng hoặc kết thúc việc phiên mã chức năng của polynucleotit được liên kết hoạt động được quan tâm trong mô bất kỳ không phải hạt phần của cây ngô chuyển gen.

Theo một số phương án, sáng chế đề xuất phương pháp biểu hiện protein hoặc polynucleotit được quan tâm trong thực vật chuyển gen hoặc tế bào thực vật chuyển gen bao gồm đưa vào thực vật hoặc tế bào thực vật catxet biểu hiện hoặc vector theo sáng chế bao gồm vùng khởi động được liên kết hoạt động với polynucleotit dị loại mà mã hóa protein hoặc polynucleotit được quan tâm, trong đó vùng khởi động bao gồm bất kỳ trong số các SEQ ID NO:1-11, và trong đó protein hoặc polynucleotit được quan tâm được biểu hiện chức năng trong mô của thực vật chuyển gen mà không phải là hạt phần. Theo các phương án khác, catxet biểu hiện hoặc vector cũng bao gồm trình tự vùng kết thúc được liên kết hoạt động với polynucleotit dị loại mà bao gồm bất kỳ trong số các SEQ ID NO:12-20.

Theo các phương án khác của phương pháp biểu hiện protein hoặc polynucleotit được quan tâm trong thực vật chuyển gen hoặc tế bào thực vật chuyển gen, protein được quan tâm là protein diệt côn trùng hoặc polynucleotit được quan tâm là dsARN mà diệt côn trùng. Theo các phương án khác, protein diệt côn trùng là protein Cry hoặc protein Vip3. Theo các phương án khác, protein Cry hoặc protein Vip3 được lựa chọn từ danh sách của các protein Cry và Vip3 được mô tả ở trên. Theo các phương án khác, protein Cry là protein Cry1. Theo các phương án khác nữa, protein Cry1 là protein Cry1A. Theo các phương án khác, protein Cry1A là protein Cry1Ab hoặc Cry1Ai. Theo các phương án khác nữa, protein Vip3 là protein Vip3A. Theo một phương án khác, protein Vip3A là protein Vip3Aa. Theo các phương án khác, protein Vip3Aa là protein Vip3Aa19 hoặc Vip3Aa20.

Theo các phương án khác của phương pháp biểu hiện protein hoặc polynucleotit được quan tâm trong thực vật chuyển gen hoặc tế bào thực vật chuyển gen, thực vật chuyển gen là cây ngô chuyển gen. Theo các phương án khác, cây ngô chuyển gen là cây ngô lai cùng dòng hoặc cây ngô lai. Theo các phương án khác nữa, mô trong đó protein được quan tâm được biểu hiện ở các mức chức năng trong cây ngô chuyển gen là lá, râu hoặc bẹ.

Theo một số phương án theo sáng chế, để mô tả thêm nữa các trình tự được yêu cầu cho việc biểu hiện trong mô bất kỳ mà không phải là hạt phấn cũng như các trình tự điều hòa mà ảnh hưởng đến mức độ biểu hiện tổng thể, việc xóa vùng khởi động trừ hạt phấn có thể được thực hiện. Việc xóa được thực hiện trong vùng sườn 5' của mỗi dòng hệ gen trừ hạt phấn. Trong hầu hết các vùng khởi động 500-1000 cặp bazơ (base pairs - bp) của trình tự sườn 5' là đủ cho hoạt động khởi đầu, bao gồm hoạt động đặc hiệu của mô. Việc xóa của vùng sườn 5' có thể tạo thành các vùng khởi động xấp xỉ 50 cặp bazơ, 100 cặp bazơ, 250 cặp bazơ, 500 cặp bazơ, 750 cặp bazơ và 1000 cặp bazơ hoặc nhiều hơn. Các trình tự xóa vùng khởi động này thực hiện mục đích gấp hai lần. Việc xóa cho phép việc lập bản đồ thêm nữa các trình tự điều hòa trong trình tự sườn 5' của mỗi dòng hệ gen trừ hạt phấn. Ngoài ra, việc xóa còn cung cấp hộp công cụ của vùng khởi động và các trình tự điều hòa mà khác nhau về các mức độ biểu hiện và các mẫu hình biểu hiện của chúng do đó cung cấp sự linh hoạt bổ sung trong việc chọn các trình tự vùng khởi động cho việc điều hòa gen thích hợp.

Cũng rõ ràng với người có trình độ trung bình trong lĩnh vực rằng các đột biến, việc chèn, việc xóa và/hoặc việc thay thế một hoặc nhiều nucleotit có thể được thêm vào các trình tự nucleotit của các SEQ ID NO:1-11 bằng cách sử dụng các phương pháp đã biết trong lĩnh vực kỹ thuật. Ngoài ra, việc xáo trộn các trình tự theo sáng chế có thể cung cấp các trình tự nucleotit mới và đa dạng. Ví dụ, các SEQ ID NO:3-5 là các biến thể của SEQ ID NO:1 mà có một hoặc nhiều hơn việc thế, việc xóa hoặc việc bổ sung được so với SEQ ID NO:2.

Để kiểm tra chức năng của các trình tự ADN cải biến theo sáng chế, chẳng hạn như các đoạn xóa của các SEQ ID NO:1-11, trình tự được quan tâm được liên kết hoạt động với gen đánh dấu nhìn thấy hoặc có thể chọn lọc và việc biểu hiện của gen đánh dấu được kiểm tra trong các thử nghiệm biểu hiện nhất thời với các mô được phân lập, chẳng hạn như mô lá, hoặc các tế bào hoặc bằng việc biến nạp ổn định vào các thực vật. Người có trình độ trung bình trong lĩnh vực sẽ biết rằng các trình tự ADN có khả năng điều khiển việc biểu hiện của trình tự mã hóa liên quan được xây dựng theo cách mô đun. Theo đó, các mức độ biểu hiện từ các đoạn ADN ngắn nhất có thể là khác với các mức độ biểu hiện từ đoạn ADN dài nhất và có thể là khác nhau. Ví dụ, việc xóa của thành phần ngược dòng điều hòa xuôi sẽ dẫn đến việc tăng về các mức độ biểu hiện của trình tự mã hóa liên quan trong khi việc xóa của thành phần điều hòa ngược sẽ làm giảm các

mức độ biểu hiện của trình tự mã hóa liên quan. Cũng được biết tới bởi người có trình độ trung bình trong lĩnh vực rằng việc xóa của các thành phần đặc hiệu cho sự phát triển hoặc mô cụ thể sẽ dẫn đến cấu hình biểu hiện bị thay đổi một phần hoặc tạm thời của trình tự mã hóa liên quan.

Theo một phương án khác theo sáng chế, ADN và các trình tự ADN hệ gen tương đồng với các SEQ ID NO:1-11 có thể được phân lập từ mầm ngô khác bằng cách sử dụng kỹ thuật lai hoặc các kỹ thuật PCR đã biết trong lĩnh vực. Các trình tự được phân lập có thể là giống hệt với các SEQ ID NO:1-11 hoặc chúng có thể là gần như giống hệt với các SEQ ID NO:1-11. Các trình tự thu được từ mầm ngô khác không nhất thiết phải chứa các trình tự nucleotit giống hệt là giống hệt về chức năng với các trình tự được bộc lộ ở đây. Một số việc xóa, việc bổ sung, và việc thay thế nucleotit có thể không có tác động hoặc chỉ tác động nhỏ lên việc biểu hiện gen. Phân tử axit nucleic được phân lập được ưu tiên, theo sáng chế, bao gồm trình tự nucleotit mà có ít nhất 90%, hoặc ít nhất 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% hoặc ít nhất 99% độ tương đồng với bất kỳ trong số các trình tự nucleotit được quy định trong các SEQ ID NO:1-11. Phân tử axit nucleic được phân lập được ưu tiên hơn bao gồm trình tự nucleotit mà có ít nhất 90% độ tương đồng với bất kỳ trong số các trình tự nucleotit được quy định trong các SEQ ID NO:1-11. Phân tử axit nucleic được phân lập được ưu tiên hơn nữa bao gồm trình tự nucleotit mà có ít nhất 95% độ tương đồng với bất kỳ trong số các trình tự nucleotit được quy định trong các SEQ ID NO:1-11. Phân tử axit nucleic được phân lập được ưu tiên hơn nữa bao gồm trình tự nucleotit mà có ít nhất 99% độ tương đồng với bất kỳ trong số các trình tự nucleotit được quy định trong các SEQ ID NO:1-11. Phân tử axit nucleic được phân lập được ưu tiên nhất bao gồm bất kỳ trong số các trình tự nucleotit được quy định là các SEQ ID NO:1-11.

Theo các phương án khác, cADN và các trình tự ADN hệ gen có thể được nhân dòng từ các thực vật khác mà thể hiện sự tương đồng của các vùng khởi động và các gen ngô trừ hạt phấn. Những tương đồng này cho phép thu được các vùng khởi động trừ hạt phấn bổ sung hữu dụng cho việc điều hòa của nhiều gen trong các mô thực vật ngoài hạt phấn. Việc lai sử dụng cADN ngô và các trình tự hệ gen hoặc các phần của nó được sử dụng để sàng lọc các trình tự gần như giống hệt hoặc tương đồng trong các hệ gen thực vật khác. Các trình tự này có thể chỉ bao gồm tập con của các nucleotit của các SEQ ID NO:1-11. Độ dài ưu tiên của sự tương đồng là 20 cặp bazơ (bp) về độ dài, tốt hơn là, 50

cặp bazơ về độ dài, và tốt nhất là ít nhất 100 cặp bazơ về độ dài. Theo một phương án của sáng chế, mẫu dò lai được chuẩn bị từ bất kỳ trong số các SEQ ID NO:1-11 hoặc các phần. Việc lai của các trình tự này có thể được tiến hành trong các điều kiện mức nghiêm ngặt cao. Mặt khác, các điều kiện mức nghiêm ngặt thấp hoặc trung bình có thể được sử dụng để cho phép một số sự không khớp trong trình tự sao cho các mức độ tương đồng thấp hơn được phát hiện (dò dị loại). Thông thường, mẫu dò là nhỏ hơn khoảng 1000 nucleotit về độ dài, tốt hơn là nhỏ hơn 500 nucleotit về độ dài.

Theo các phương án khác theo sáng chế, cADN và các trình tự hệ gen được phân lập bằng việc chuẩn bị các đoạn môi bao gồm các trình tự trong bất kỳ trong số các SEQ ID NO:1-11. Các đoạn môi có thể được sử dụng trong phản ứng PCR với cADN hoặc ADN hệ gen từ thực vật để thu được các trình tự tương đồng hoặc các trình tự gần như giống hệt với bất kỳ trong số các SEQ ID NO:1-11.

Cấu tạo của catxet biểu hiện

Các catxet biểu hiện được xây dựng bao gồm các trình tự sườn 5' của các dòng hệ gen trừ hạt phấn. Theo các phương án theo sáng chế, vùng khởi động được sử dụng trong mỗi catxet biểu hiện bao gồm vùng sườn 5' cho tới và bao gồm phần bắt đầu của việc dịch mã. Phần bắt đầu của việc dịch mã được biểu thị bởi ATG thứ nhất của khung đọc mở (ORF) được tìm thấy trong cADN và trình tự hệ gen tương đồng. Do đó, vùng khởi động có thể bao gồm trình tự dẫn đầu không được dịch mã 5' cũng như vị trí bắt đầu phiên mã, vùng khởi động lõi và các thành phần điều hòa bổ sung. Theo các phương án khác theo sáng chế, các catxet biểu hiện được xây dựng bao gồm trình tự sườn 5' của các dòng hệ gen trừ hạt phấn cho tới và bao gồm vị trí khởi đầu phiên mã. Vị trí khởi đầu phiên mã có thể được xác định bởi nucleotit thứ nhất của dòng cADN dài nhất thu được. Ngoài ra, vị trí khởi đầu phiên mã có thể còn được xác định bằng việc sử dụng các kỹ thuật đã biết trong lĩnh vực bao gồm RACE PCR, ánh xạ bảo vệ RNaza và phân tích mở rộng đoạn môi.

Các catxet biểu hiện có thể còn bao gồm vùng kết thúc phiên mã, xuôi chiều (3') tới vùng khởi động. Nhiều vùng kết thúc phiên mã khác nhau là sẵn có cho việc sử dụng trong các catxet biểu hiện. Vùng kết thúc phiên mã chịu trách nhiệm cho việc kết thúc phiên mã ngoài gen chuyển và hiệu chỉnh việc polyadenyl hóa mARN của phiên mã mARN. Các vùng kết thúc phiên mã thích hợp các vùng đã biết là hoạt động chức năng

ở các thực vật và bao gồm vùng kết thúc CaMV 35S, vùng kết thúc *tml*, vùng kết thúc *synthaza nopalín* và vùng kết thúc *rbcs E9* ở đậu. Chúng có thể được sử dụng trong cả các cây một lá mầm và các cây hai lá mầm. Ngoài ra, vùng kết thúc phiên mã nguyên bản của gen có thể được sử dụng. Ví dụ, trình tự sườn 3' bao gồm trình tự hệ gen 3' tới vùng tương đồng với dòng cADN trừ hạt phần có thể được sử dụng. Các vùng kết thúc như vậy được lấy ví dụ ở đây bao gồm bất kỳ trong số các SEQ ID NO:12-20.

Theo một số phương án theo sáng chế trình tự mã hóa dị loại, ví dụ, trình tự mã hóa diệt côn trùng, trình tự mã hóa đánh dấu nhìn thấy, hoặc trình tự mã hóa đánh dấu có thể lựa chọn, được nhân dòng giữa vùng khởi động theo sáng chế và vùng kết thúc phiên mã bởi đó trình tự mã hóa dị loại được liên kết hoạt động với vùng khởi động và vùng kết thúc phiên mã được liên kết hoạt động với trình tự mã hóa dị loại. Các ví dụ của các đánh dấu nhìn thấy hữu dụng cho sáng chế bao gồm, nhưng không bị giới hạn ở, β -glucuronidaza (GUS), Cloamphenicol Acetyl Transferaza (CAT), Luxiferaza (LUC) và các protein với các đặc tính huỳnh quang, chẳng hạn như Protein huỳnh quang xanh lá cây (GFP) từ *Aequora victoria*. Về nguyên tắc, nhiều protein là phù hợp cho mục đích này, với điều kiện protein không làm ảnh hưởng đến các chức năng thiết yếu của thực vật. Các ví dụ thêm nữa về các trình tự mã hóa dị loại hữu dụng cho sáng chế bao gồm, nhưng không bị giới hạn ở, kháng thuốc kháng sinh, kháng virus, kháng côn trùng, kháng bệnh, hoặc kháng các sinh vật gây hại khác, dung hòa thuốc diệt cỏ, cải thiện giá trị dinh dưỡng, cải thiện hiệu quả trong quá trình xử lý công nghiệp hoặc khả năng sinh sản được thay đổi. Theo các phương án khác theo sáng chế, gen mã hóa để kháng các côn trùng mà ăn các mô của các thực vật mà không phải là hạt phần được nhân dòng giữa vùng khởi động theo sáng chế và vùng kết thúc đã biết trong lĩnh vực hoặc của sáng chế. Theo một phương án khác của sáng chế trình tự mã hóa ARN chức năng chẳng hạn như ARN đối nghĩa, ARN có nghĩa để ức chế có nghĩa, hoặc ARN chuỗi kép cũng có thể được nhân dòng giữa vùng khởi động và vùng kết thúc phiên mã.

Nhiều trình tự đã được phát hiện là nâng cao sự biểu hiện gen từ trong đơn vị phiên mã và các trình tự này có thể được sử dụng kết hợp với các vùng khởi động của sáng chế để làm tăng việc biểu hiện của chúng trong các thực vật chuyển gen. Các trình tự intron khác nhau đã thể hiện là tăng cường sự biểu hiện, cụ thể là trong các tế bào thực vật một lá mầm. Ví dụ, một số các trình tự dẫn đầu không được dịch mã có nguồn gốc từ các virus cũng được biết là tăng cường sự biểu hiện, và chúng đặc biệt hữu ích

trong các tế bào thực vật hai lá mầm. Cụ thể là, các trình tự dẫn đầu từ virus khảm thuốc lá (TMV, "trình tự W"), virus đốm vàng ở ngô (MCMV), và virus khảm cỏ linh lăng (AMV) đã thể hiện là hiệu quả trong việc tăng cường biểu hiện (ví dụ Gallie *et al.* *Nucl. Acids Res.* 15: 8693-8711 (1987); Skuzeski *et al.* *Plant Molec. Biol.* 15: 65-79 (1990)). Vùng khởi động trừ hạt phần theo sáng chế có thể bao gồm, hoặc được cải biến để bao gồm, một hoặc nhiều phần tử vùng tăng cường. Theo một số phương án, vùng khởi động có thể bao gồm nhiều phần tử vùng tăng cường. Các vùng khởi động chứa các phần tử vùng tăng cường cung cấp các mức cao hơn của sự phiên mã trong các mô không phải hạt phần như so sánh với các vùng khởi động mà không bao gồm chúng. Các phần tử vùng tăng cường thích hợp cho việc sử dụng trong các thực vật bao gồm phần tử vùng tăng cường PCISV (Patent Mỹ Số 5,850,019), phần tử vùng tăng cường CaMV 35S (Patent Mỹ Số 5,106,739 và 5,164,316) và phần tử vùng tăng cường FMV (Maiti *et al.* (1997) *Transgenic Res.* 6:143-156); vùng kích hoạt dịch mã của virus khảm thuốc lá (TMV) được mô tả trong đơn WO87/07644, hoặc của virus khảm thuốc lá (TEV) được mô tả bởi Carrington & Freed 1990, *J. Virol.* 64: 1590-1597, ví dụ, hoặc các intron chẳng hạn như intron adh1 của ngô hoặc intron 1 của lúa actin. Tương tự xem thêm trong PCT WO96/23898, WO2012/021794, WO2012/021797, WO2011/084370, và WO2011/028914.

Sự biến nạp thực vật

Các quá trình biến nạp các thực vật là đã biết trong lĩnh vực và được mô tả trong suốt tài liệu. Các ví dụ không giới hạn của các phương pháp biến nạp các thực vật bao gồm sự biến nạp thông qua sự vận chuyển axit nucleic được trung gian bởi vi khuẩn (ví dụ, thông qua *Agrobacterium*), sự vận chuyển axit nucleic được trung gian bởi virus, sự vận chuyển axit nucleic được trung gian bởi silicon cacbua hoặc axit nucleic, sự vận chuyển axit nucleic được trung gian bởi liposom, vi tiêm, sự bắn phá vi hạt, sự biến nạp trung gian bởi canxi-phosphat, sự biến nạp trung gian bởi cyclodextrin, điện di, sự biến nạp trung gian bởi nano, siêu âm, thâm nhập, hấp thu axit nucleic trung gian bởi PEG, cũng như cơ chế sinh học hoặc vật lý (cơ học), hóa học, điện tử bất kỳ khác mà dẫn đến việc đưa axit nucleic vào tế bào thực vật, bao gồm bất kỳ kết hợp nào của chúng. Các hướng dẫn chung cho các phương pháp biến nạp thực vật khác nhau đã biết trong lĩnh vực bao gồm Miki *et al.* ("Procedures for Introducing Foreign ADN into Plants" in *Methods in Plant Molecular Biology và Biotechnology*, Glick, B. R. and Thompson, J.

E., Eds. (CRC Press, Inc., Boca Raton, 1993), pages 67-88) và Rakowoczy-Trojanowska (*Cell. Mol. Biol. Lett.* 7:849-858 (2002)).

Đối với sự biến nạp trung gian bởi *Agrobacterium*, các vector nhị phân hoặc các vector mang ít nhất một trình tự biên T-ADN là thích hợp, trong đó cho việc biến nạp gen trực tiếp (ví dụ, bắn phá hạt và tương tự) vector bất kỳ là thích hợp và ADN mạch thẳng chứa chỉ cấu trúc được quan tâm có thể được sử dụng. Trong trường hợp biến nạp gen trực tiếp, việc biến nạp với loài ADN đơn hoặc đồng biến nạp có thể được sử dụng (Schocher et al., *Biotechnology* 4:1093- 1096 (1986)). Đối với cả biến nạp gen trực tiếp và biến nạp được trung gian bởi *Agrobacterium*, việc biến nạp thường (nhưng không nhất thiết) được thực hiện với đánh dấu chọn lọc mà có thể là sự lựa chọn dương tính (Phosphomannoza Isomeraza; PMI), cung cấp khả năng kháng thuốc kháng sinh (kanamycin, hygromycin hoặc methotrexat) hoặc chất diệt cỏ (glyphosat hoặc glufosinat). Tuy nhiên, sự lựa chọn của đánh dấu chọn lọc là không nhất thiết đối với sàng chế.

Sự biến nạp trung gian bởi *Agrobacterium* là phương pháp thường được sử dụng để biến nạp các thực vật bởi mức hiệu quả cao của nó trong việc biến nạp và bởi tính ứng dụng rộng với nhiều loài khác nhau. Sự biến nạp trung gian bởi *Agrobacterium* thường bao gồm việc biến nạp của vector nhị phân mang ADN ngoại lai được quan tâm tới chủng *Agrobacterium* thích hợp mà có thể phụ thuộc vào sự bổ sung của các gen *vir* được mang bởi chủng *Agrobacterium* vật chủ hoặc trên Ti plasmit đồng cư trú hoặc nhiễm sắc thể (Uknes et al. (1993) *Plant Cell* 5:159-169). Việc biến nạp của vector nhị phân tái tổ hợp vào *Agrobacterium* có thể được hoàn tất bởi quá trình giao hợp ba bố mẹ sử dụng *Escherichia coli* mang vector nhị phân tái tổ hợp, chủng *E. coli* hỗ trợ mang plasmit mà có khả năng đưa vector nhị phân tái tổ hợp tới chủng *Agrobacterium* mục tiêu. Mặt khác, vector nhị phân tái tổ hợp có thể được biến nạp vào *Agrobacterium* bằng việc biến nạp axit nucleic (Höfgen & Willmitzer (1988) *Nucleic Acids Res.* 16:9877).

Các cây hai lá mầm cũng như một lá mầm có thể được biến nạp sử dụng *Agrobacterium*. Các phương pháp để biến nạp trung gian bởi *Agrobacterium* của lúa bao gồm các phương pháp đã biết rõ để biến nạp lúa, chẳng hạn như phương pháp được mô tả trong bất kỳ trong số các tài liệu sau đây: Đơn sáng chế châu Âu EP 1198985 A1, Aldemita và Hodges (*Planta* 199: 612-617, 1996); Chan et al. (*Plant Mol Biol* 22 (3): 491-506, 1993), Hiei et al. (*Plant J* 6 (2): 271-282, 1994), mà các phần bộc lộ được kết

hợp bằng việc viện dẫn ở đây như được quy định đầy đủ. Trong trường hợp biến nạp ngô, phương pháp ưu tiên là như được mô tả trong hoặc Ishida et al. (Nat. Biotechnol 14(6): 745-50, 1996) hoặc Frame et al. (Plant Physiol 129(1): 13-22, 2002), mà các phần bộc lộ được kết hợp bằng việc viện dẫn ở đây như được quy định đầy đủ. Các phương pháp đã nêu được mô tả thêm như ví dụ trong B. Jenes et al., Techniques for Gene Transfer, in: Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S. D. Kung and R. Wu, Academic Press (1993) 128-143 và trong Potrykus Annu. Rev. Plant Physiol. Plant Molec. Biol. 42 (1991) 205-225). Các axit nucleic hoặc cấu trúc cần được biểu hiện tốt hơn là được nhân dòng vào vectơ, mà vectơ này thích hợp để biến nạp *Agrobacterium tumefaciens*, ví dụ pBin19 (Bevan et al., Nucl. Acids Res. 12 (1984) 8711). *Agrobacteria* được biến nạp bởi vectơ này sau đó có thể được sử dụng theo cách đã biết cho việc biến nạp thực vật, chẳng hạn như các thực vật được sử dụng làm mô hình, như *Arabidopsis* hoặc các cây trồng chẳng hạn như, làm ví dụ, các cây thuốc lá, ví dụ bằng việc ngâm các lá bị giập hoặc các lá được cắt nhỏ trong dung dịch *agrobacterial* và sau đó nuôi cấy chúng trong môi trường thích hợp. Việc biến nạp của thực vật bởi *Agrobacterium tumefaciens* được mô tả, ví dụ, bởi Hagen and Willmitzer in Nucl. Acid Res. (1988) 16, 9877 hoặc được biết không kể những cái khác từ F. F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, eds. S. D. Kung và R. Wu, Academic Press, 1993, pp. 15-38.

Việc biến nạp của thực vật bởi *Agrobacterium* tái tổ hợp thường bao gồm việc đồng nuôi cấy của *Agrobacterium* với các mẫu từ thực vật và các phương pháp tiếp theo đã biết trong lĩnh vực. Mô được biến nạp được tái sinh trên môi trường chọn lọc mang chỉ thị chống kháng sinh hoặc diệt cỏ giữa các biên T-ADN plasmit nhị phân.

Như được thảo luận trước đó, phương pháp khác để biến nạp các thực vật, các phần thực vật và các tế bào thực vật bao gồm thúc đẩy các hạt trơ hoặc hoạt tính sinh học tại các tế bào và các mô thực vật. Xem, ví dụ, patent Mỹ số 4,945,050; 5,036,006 và 5,100,792. Thông thường, phương pháp này bao gồm thúc đẩy các hạt trơ hoặc hoạt tính sinh học ở các tế bào thực vật trong các điều kiện hiệu quả để thâm nhập vào bề mặt bên ngoài của tế bào và tạo sự kết hợp bên trong của chúng. Khi các hạt trơ được sử dụng, vectơ có thể được thêm vào tế bào bằng việc phủ các hạt với vectơ chứa axit nucleic được quan tâm. Mặt khác, một hoặc các tế bào có thể được bao quanh bởi vectơ sao cho vectơ được đưa vào tế bào bởi việc kích hoạt các hạt. Các hạt hoạt tính sinh học (ví dụ,

tế bào nấm men sấy khô, vi khuẩn khô hoặc thực khuẩn thể, mỗi trong số chúng chứa một hoặc nhiều axit nucleic tìm cách được thêm) cũng có thể được đẩy vào mô thực vật.

Theo các phương án khác, polynucleotit theo sáng chế có thể được biến nạp trực tiếp vào hệ gen plastit. Ưu điểm chính của việc biến nạp plastit là ở chỗ các plastit thường có khả năng biểu hiện các gen vi khuẩn mà không có sự cải biến đáng kể, và các plastit có khả năng biểu hiện nhiều khung đọc mở trong sự kiểm soát của vùng khởi động đơn. Công nghệ biến nạp plastit được mô tả sâu trong các Patent Mỹ số 5,451,513, 5,545,817, và 5,545,818, trong đơn PCT số WO 95/16783, và trong McBride et al. (1994) *Proc. Natl. Acad. Sci. USA* 91, 7301-7305. Kỹ thuật cơ bản cho việc biến nạp lục lạp bao gồm bổ sung các vùng của sườn ADN plastit được nhân dòng đánh dấu chọn lọc cùng với gen được quan tâm vào mô mục tiêu thích hợp, ví dụ, sử dụng bắn hạt sinh học hoặc biến nạp thể nguyên sinh (ví dụ, sự biến nạp trung gian bởi PEG hoặc canxi clorua). Các vùng sườn 1 đến 1,5 kb, gọi là các trình tự mục tiêu, tạo điều kiện cho việc tái tổ hợp đồng nhất với hệ gen plastit và do đó cho phép việc thay thế hoặc cải biến của các vùng đặc hiệu của plastom. Ban đầu, các đột biến điểm trong các gen 16S rARN và rps12 lục lạp mang lại khả năng chống spectinomycin hoặc streptomycin có thể được sử dụng làm các đánh dấu chọn lọc cho việc biến nạp (Svab, Z., Hajdukiewicz, P., và Maliga, P. (1990) *Proc. Natl. Acad. Sci. USA* 87, 8526-8530; Staub, J. M., và Maliga, P. (1992) *Plant Cell* 4, 39-45). Sự có mặt của các vị trí nhân bản giữa các đánh dấu này cho phép việc tạo ra của vectơ hướng đích plastit cho việc giới thiệu của các gen ngoại lai (Staub, J.M., and Maliga, P. (1993) *EMBO J.* 12, 601-606). Các sự tăng đáng kể về tần suất biến nạp có thể được thu bằng việc thay thế rARN lặn hoặc các gen kháng thuốc kháng sinh r-protein với đánh dấu chọn lọc trội, gen aadA vi khuẩn mã hóa enzym spectinomycin-cletoxyf aminoglycosit- 3'- adenylyltransf tẩy xóa (Svab, Z., and Maliga, P. (1993) *Proc. Natl. Acad. Sci. USA* 90, 913-917). Trước đó, đánh dấu này đã được sử dụng thành công cho việc biến nạp tần suất cao của hệ gen plastit của tảo xanh lá cây *Chlamydomonas reinhardtii* (Goldschmidt-Clermont, M. (1991) *Nucl. Acids Res.* 19:4083-4089). Các đánh dấu chọn lọc khác hữu dụng cho việc biến nạp plastit là đã biết trong lĩnh vực và được bao gồm trong phạm vi theo sáng chế. Thông thường, cần xấp xỉ 15-20 chu kỳ phân chia tế bào tiếp theo sau việc biến nạp để đạt được trạng thái đồng loại. Việc biểu hiện plastit, trong đó các gen được chèn bởi việc tái tổ hợp tương đồng vào tất cả trong số vài nghìn bản sao của hệ gen plastit tròn trong mỗi tế bào thực

vật, tận dụng ưu điểm số lượng bản sao khổng lồ so với các gen được biểu hiện trong nhân để cho phép các mức độ biểu hiện mà có thể vượt quá 10% của tổng số protein thực vật hòa tan. Theo một phương án, polynucleotit theo sáng chế có thể được chèn vào vectơ mục tiêu plastit và được biến nạp vào hệ gen plastit của vật chủ thực vật mong muốn. Do đó, có thể thu được các thực vật tương đồng cho hệ gen plastit chứa trình tự nucleotit theo sáng chế, mà có khả năng biểu hiện ở mức cao polynucleotit.

Các phương pháp chọn lọc các môi trường thực vật, các tế bào thực vật hoặc mô thực vật chuyển gen, được biến nạp là thông thường trong lĩnh vực kỹ thuật và có thể được thực hiện trong các phương pháp theo sáng chế được đề xuất ở đây. Ví dụ, vector tái tổ hợp theo sáng chế cũng có thể bao gồm catxet biểu hiện bao gồm trình tự nucleotit của đánh dấu chọn lọc, mà có thể được sử dụng để chọn lọc thực vật, phần thực vật hoặc tế bào thực vật được biến nạp. Như được sử dụng ở đây, “đánh dấu chọn lọc” có nghĩa là trình tự nucleotit mà khi được biểu hiện truyền kiểu hình riêng biệt với thực vật, phần thực vật hoặc tế bào thực vật biểu hiện đánh dấu và do đó cho phép các thực vật, các phần thực vật hoặc các tế bào thực vật được biến nạp để phân biệt được với các thực vật, các phần thực vật hoặc các tế bào thực vật không có đánh dấu này. Trình tự nucleotit này có thể mã hóa hoặc đánh dấu chọn lọc hoặc sàng lọc được, phụ thuộc vào liệu đánh dấu có mang đặc điểm mà có thể được chọn lọc bằng cách hóa học, chẳng hạn như bằng cách sử dụng chất chọn lọc (ví dụ, kháng sinh, diệt cỏ, hoặc tương tự), hoặc liệu chọn lọc đơn giản là tính trạng mà có thể được nhận biết thông qua việc quan sát hoặc thử nghiệm, chẳng hạn như bằng việc sàng lọc (ví dụ, tính trạng locus R). Đương nhiên, nhiều ví dụ của các đánh dấu chọn lọc phù hợp là đã biết trong lĩnh vực và có thể được sử dụng trong các catxet biểu hiện được mô tả ở đây.

Các ví dụ của các đánh dấu chọn lọc bao gồm, nhưng không bị giới hạn ở, trình tự nucleotit mã hóa *neo* hoặc *nptII*, mà tạo nên sự kháng kanamycin, G418, và tương tự (Potrykus *et al.* (1985) *Mol. Gen. Genet.* 199:183-188); trình tự nucleotit mã hóa *bar*, mà tạo nên sự kháng phosphinothricin; trình tự nucleotit mã hóa 5-enolpyruvylshikimat-3-phosphat (EPSP) synthaza được thay đổi, mà tạo nên việc kháng glyphosat (Hinchee *et al.* (1988) *Biotech.* 6:915-922); trình tự nucleotit mã hóa nitrilaza chẳng hạn như *bxn* từ *Klebsiella ozaenae* mà tạo nên việc kháng bromoxynil (Stalker *et al.* (1988) *Science* 242:419-423); trình tự nucleotit mã hóa axetolactat synthaza được thay đổi (ALS) mà tạo nên việc kháng imidazolinon, sulfonylurea hoặc các hóa chất ức chế ALS khác (đơn

sáng chế châu Âu số 154204); trình tự nucleotit mã hóa dihydrofolat reductaza kháng methotrexat (DHFR) (Thillet *et al.* (1988) *J. Biol. Chem.* 263:12500-12508); trình tự nucleotit mã hóa dalapon dehalogenaza mà tạo nên việc kháng dalapon; trình tự nucleotit mã hóa mannoza-6-phosphat isomeraza (cũng được gọi là phosphomannoza isomeraza (PMI)) mà tạo nên khả năng chuyển hóa mannoza (patent Mỹ số 5,767,378 và 5,994,629); trình tự nucleotit mã hóa anthranilat synthaza được thay đổi mà tạo nên việc kháng 5-metyl tryptophan; hoặc trình tự nucleotit mã hóa *hph* mà tạo nên việc kháng hygromyxin. Người có trình độ trung bình trong lĩnh vực có khả năng lựa chọn đánh dấu chọn lọc thích hợp cho việc sử dụng trong catxet biểu hiện của sáng chế.

Các đánh dấu chọn lọc bổ sung bao gồm, nhưng không bị giới hạn ở, trình tự nucleotit mã hóa β -glucuronidaza hoặc *uidA* (GUS) mà mã hóa enzym cho nó các chất tạo màu khác nhau đã được biết; trình tự nucleotit locus R mà mã hóa sản phẩm mà điều hòa việc tạo ra của các sắc tố anthocyanin (màu đỏ) trong các mô thực vật (Dellaporta *et al.*, "Molecular cloning of the maize R-nj allele by transposon-tagging with Ac" 263-282 In: *Chromosome Structure and Function: Impact of New Concepts*, 18th Stadler Genetics Symposium (Gustafson & Appels eds., Plenum Press 1988)); trình tự nucleotit mã hóa β -lactamaza, enzym mà cho nó các chất tạo màu khác nhau đã được biết (ví dụ, PADAC, cephalosporin tạo màu) (Sutcliffe (1978) *Proc. Natl. Acad. Sci. USA* 75:3737-3741); trình tự nucleotit mã hóa *xylE* mà mã hóa catechol dioxygenaza (Zukowsky *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1101-1105); trình tự nucleotit mã hóa tyrosinaza, enzym có khả năng oxy hóa tyrosin thành DOPA và dopaquinon, mà đến lượt nó cô đặc để tạo thành melanin (Katz *et al.* (1983) *J. Gen. Microbiol.* 129:2703-2714); trình tự nucleotit mã hóa β -galactosidaza, enzym mà cho nó có các chất tạo màu; trình tự nucleotit mã hóa luxiferaza (*lux*) mà cho phép việc phát hiện phát quang sinh học (Ow *et al.* (1986) *Science* 234:856-859); trình tự nucleotit mã hóa aequorin mà có thể được sử dụng trong việc phát hiện phát quang sinh học nhạy canxi (Prasher *et al.* (1985) *Biochem. Biophys. Res. Comm.* 126:1259-1268); hoặc trình tự nucleotit mã hóa protein huỳnh quang xanh lá cây (Niedz *et al.* (1995) *Plant Cell Reports* 14:403-406). Người có trình độ trung bình trong lĩnh vực có khả năng lựa chọn đánh dấu chọn lọc thích hợp cho việc sử dụng trong catxet biểu hiện của sáng chế.

Ngoài ra, như đã biết trong lĩnh vực, các thực vật chuyển gen nguyên vẹn có thể được tái sinh từ môi trường nuôi cấy các tế bào thực vật, mô thực vật được biến nạp

hoặc các chất nguyên sinh được nuôi cấy sử dụng bất kỳ trong số nhiều kỹ thuật đã biết khác nhau. Việc tái sinh thực vật từ các tế bào thực vật, môi trường mô thực vật hoặc các chất nguyên sinh được nuôi cấy được mô tả, ví dụ, trong Evans *et al.* (*Handbook of Plant Cell Cultures*, Vol. 1, MacMilan Publishing Co. New York (1983)); và Vasil I. R. (ed.) (*Cell Culture and Somatic Cell Genetics of Plants*, Acad. Press, Orlando, Vol. I (1984), và Vol. II (1986)).

Ngoài ra, các đặc tính di truyền được thiết kế trong các thực vật, các phần thực vật, các tế bào thực vật hoặc các hạt chuyển gen theo sáng chế được mô tả ở trên có thể được truyền lại qua sinh sản hữu tính hoặc sinh dưỡng và do đó có thể được duy trì và nhân giống ở các cây thế hệ con. Thông thường, việc duy trì và nhân giống sử dụng các phương pháp nông nghiệp đã biết được phát triển để phù hợp với các mục đích cụ thể chẳng hạn như thu hoạch, gieo hạt hoặc xới đất.

Polynucleotit do đó có thể được đưa vào thực vật, phần thực vật hoặc tế bào thực vật theo số lượng cách nhất định mà đã biết trong lĩnh vực, như được mô tả ở trên. Do đó, không có phương pháp cụ thể nào để đưa một hoặc nhiều polynucleotit vào cây trồng, thay vì sử dụng bất kỳ phương pháp nào cho phép một hoặc nhiều polynucleotit được biểu hiện nhất thời trong tế bào thực vật hoặc tích hợp ổn định vào bộ gen của cây. Khi có nhiều hơn một polynucleotit được đưa vào, các polynucleotit tương ứng có thể được lắp ráp như một phần của một phân tử axit nucleic đơn lẻ, hoặc dưới dạng các phân tử axit nucleic riêng biệt và có thể nằm trên cùng hoặc các phân tử axit nucleic khác nhau. Do đó, các polynucleotit có thể được đưa vào tế bào quan tâm trong một sự kiện biến đổi đơn lẻ, trong các sự kiện biến đổi riêng biệt, hoặc ví dụ, trong thực vật, như một phần của quá trình nhân giống.

Các phương án bổ sung theo sáng chế bao gồm các sản phẩm được thu hoạch được tạo ra từ các thực vật chuyển gen hoặc các bộ phận của chúng theo sáng chế, cũng như sản phẩm được xử lý được tạo ra từ các sản phẩm được thu hoạch. Sản phẩm được thu hoạch có thể là toàn bộ thực vật hoặc bộ phận thực vật bất kỳ, như được mô tả ở đây. Do đó, theo một số phương án, các ví dụ không nhằm giới hạn của sản phẩm được thu hoạch bao gồm hạt, quả, hoa hoặc phần của chúng (ví dụ, bao phấn, đầu nhụy, và tương tự), lá, thân, và tương tự. Theo các phương án khác, sản phẩm được xử lý bao gồm, nhưng không bị giới hạn ở, bột mì, bột thô, dầu, tinh bột, ngũ cốc, và tương tự được tạo ra từ hạt đã thu hoạch hoặc phần thực vật khác theo sáng chế, trong đó hạt hoặc

phần thực vật khác đã nêu bao gồm trình tự phân tử axit nucleic/polynucleotit/nucleotit của sáng chế.

Theo các phương án khác, sáng chế đề xuất phần chiết từ hạt chuyển gen hoặc thực vật chuyển gen theo sáng chế, trong đó phần chiết bao gồm phân tử axit nucleic, polynucleotit, trình tự nucleotit hoặc protein diệt côn trùng theo sáng chế. Các phần chiết từ các thực vật hoặc các phần thực vật có thể được tạo ra theo các quá trình đã biết trong lĩnh vực (*See, de la Torre et al., Food, Agric. Environ.* 2(1):84-89 (2004); Guidet, *Nucleic Acids Res.* 22(9): 1772-1773 (1994); Lipton et al., *Food Agric. Immun.* 12:153-164 (2000)).

Đánh giá hoạt động vùng khởi động

Một số phương pháp là sẵn có để đánh giá hoạt động vùng khởi động. Các catxet biểu hiện được xây dựng với đánh dấu nhìn thấy, như được mô tả ở trên. Các phương pháp biến nạp tạm thời có thể được sử dụng để đánh giá hoạt động vùng khởi động. Sử dụng các phương pháp biến nạp chẳng hạn như bắn phá vi đạn, biến nạp *Agrobacterium* hoặc biến nạp thể nguyên sinh, các catxet biểu hiện được chuyển tới các mô hoặc các tế bào thực vật. Hoạt động của gen báo cáo, chẳng hạn như hoạt tính β -glucuronidaza, hoạt tính luxiferaza hoặc huỳnh quang GFP được giám sát sau khi biến nạp theo thời gian, ví dụ 2 giờ, 5 giờ, 8 giờ, 16 giờ, 24 giờ, 36 giờ, 48 giờ và 72 giờ sau khi vận chuyển ADN sử dụng các phương pháp đã biết trong lĩnh vực kỹ thuật. Hoạt động gen báo cáo có thể được giám sát bởi hoạt động enzym, bằng cách nhuộm tế bào hoặc mô với chất nền cho enzym được mã hóa bởi gen báo cáo hoặc bằng cách trực quan hóa dưới bước sóng ánh sáng thích hợp. Protein diệt côn trùng, ví dụ Cry1Ab hoặc Vip3 theo sáng chế, có thể hoạt động như đánh dấu nhìn thấy được bởi đó các tế bào thực vật được biến nạp được thử nghiệm hoạt tính diệt côn trùng. Các trình tự vùng khởi động độ dài đầy đủ, việc xóa và các đột biến của trình tự vùng khởi động có thể được thử nghiệm và các mức độ biểu hiện của chúng được so sánh. Ngoài ra, các mức ARN có thể được đo sử dụng các phương pháp đã biết trong lĩnh vực chẳng hạn như thăm Northern, các thử nghiệm bảo vệ ARNse và PCR phiên mã ngược cạnh tranh. Các thử nghiệm này đo mức của việc biểu hiện của vùng khởi động bằng việc đo nồng độ 'trạng thái dừng' của mARN báo cáo được phiên mã tiêu chuẩn. Phép đo này là gián tiếp vì nồng độ của mARN báo cáo không chỉ phụ thuộc vào tốc độ tổng hợp của nó mà còn phụ thuộc vào tốc độ phân giải mARN. Do đó mức trạng thái dừng là tích số của tốc độ tổng hợp và tốc độ phân giải.

Tuy nhiên, tốc độ suy giảm có thể được coi là tiến hành với tốc độ cố định khi các trình tự được phiên mã giống hệt nhau, và do đó giá trị này có thể dùng làm thước đo tốc độ tổng hợp.

Việc xác nhận thêm hoạt động của vùng khởi động thu được bằng cách chuyển đổi ổn định vùng khởi động trong catxet biểu hiện bao gồm đánh dấu có thể nhìn thấy được hoặc gen quan tâm vào thực vật như được mô tả ở trên. Sử dụng các phương pháp khác nhau được mô tả ở trên, chẳng hạn như các thử nghiệm hoạt tính enzym, phân tích ARN và thử nghiệm protein như đã mô tả ở trên, hoạt động của vùng khởi động được giám sát quá trình phát triển và thêm vào đó bằng cách giám sát sự biểu hiện trong các mô khác nhau trong các thể biến nạp sơ cấp và qua các thế hệ thực vật chuyển gen tiếp theo.

Ví dụ thực hiện sáng chế

Sáng chế sẽ được mô tả thêm bằng việc tham chiếu tới các ví dụ cụ thể sau đây. Các ví dụ này được cung cấp chỉ nhằm các mục đích minh họa, và không nhằm giới hạn trừ khi được nêu rõ khác đi. Các kỹ thuật nhân dòng phân tử và ADN tái tổ hợp tiêu chuẩn được sử dụng ở đây là đã biết trong lĩnh vực và được mô tả bởi Ausubel (ed.), *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. (1994); J. Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual, 3d Ed.*, Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press (2001); và bởi T.J. Silhavy, M.L. Berman, và L.W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984).

Ví dụ 1. Việc xác định các thành phần điều hòa ở ngô.

Các thành phần điều hòa được xác định từ các gen ở ngô trên chip Affymetrix (Zm80K) độc quyền Syngenta mà được biểu hiện ở mức cao trong nhiều loại mô nhưng không trong hạt phấn, nghĩa là việc biểu hiện hạt phấn là rất thấp hoặc không thể được phát hiện. Các thành phần điều hòa từ tám gen khác nhau ở ngô được sử dụng để tạo ra các cấu trúc để thúc đẩy việc biểu hiện của các gen kiểm soát côn trùng trong các cây và các mô ở ngô chuyển gen. Phần mô tả của các thành phần điều hòa được sử dụng được thể hiện trong Bảng 1.

Bảng 1. Mô tả các thành phần điều hòa được sử dụng để tạo ra các cấu trúc biểu hiện.

Thành phần điều hòa	SEQ ID NO:	Mô tả
PMP370-3	1	Vùng khởi động từ gen Zm032370 (Syngenta chip ID), được định danh như giống profilin. Bao gồm 5' UTR, intron và exon thứ nhất, và một phần của exon thứ hai.
PMP393-1	2	Vùng khởi động từ gen Zm061393 (Syngenta chip ID), được định danh như sucroza synthaza. Bao gồm 5' UTR, các exon thứ nhất và thứ hai, và các intron thứ nhất và thứ hai. Cũng bao gồm đột biến T thành A để loại bỏ codon khởi đầu.
PMP393-2	3	prZm061393 với vùng tăng cường TMV được bổ sung ở đầu 3'.
PMP393-3	4	prZm061393-02 với đột biến T thành A ở cặp bazơ 8 để loại bỏ khung đọc mở (ORF)
PMP393-4	5	prZm041393-01 với đột biến T thành A ở cặp bazơ 29 để loại bỏ thành phần chéo ngoài ý muốn ORF trên mạch đối nghĩa.
PMP855-1	6	Vùng khởi động từ gen ZmU45855-3 (Syngenta chip ID) từ ngô Zea. Bao gồm 5' UTR, exon thứ nhất, intron thứ nhất, và một phần của exon thứ hai.
PMP747-1	7	Vùng khởi động từ gen Zm001747 (Syngenta chip ID) từ ngô Zea, được định danh như mã hóa protein sốc nhiệt (HSP). Bao gồm 5' UTR.
PMP004-1	8	Vùng khởi động từ gen Zm021004 (Syngenta chip ID) từ ngô Zea, có sự tương đồng với các gen mã hóa các protein gắn ran, và bao gồm vùng không được dịch mã 5', exon thứ nhất, intron thứ nhất, một phần của exon thứ hai, và nó kết thúc trong trình tự Kozak được tối ưu hóa ở ngô.
PMP335-1	9	Vùng khởi động từ gen Zm015335 (Syngenta chip ID) từ ngô Zea tương tự với các gen mã hóa các protein ribosom 40S, và bao gồm vùng không được dịch mã 5', exon thứ nhất, intron thứ nhất, exon thứ hai một phần, và nó kết thúc trong trình tự Kozak được tối ưu hóa ở ngô.
PMP722-1	10	Vùng khởi động từ gen Zm009722 từ ngô Zea, và bao gồm vùng không được dịch mã 5', exon thứ

		nhất, intron thứ nhất và một phần của exon thứ hai, và nó kết thúc trong trình tự Kozak được tối ưu hóa ở ngô.
PMP948-2	11	Vùng khởi động từ gen Zm058948 (Syngenta chip ID) từ ngô Zea. Bao gồm vùng không được dịch mã 5', exon thứ nhất, intron thứ nhất, exon thứ hai một phần, và kết thúc trong trình tự Kozak được tối ưu hóa ở ngô.

Các thành phần kết thúc cũng được xác định từ cùng các gen từ đó các vùng khởi động được xác định ở trên được nhân dòng. Phần mô tả của các vùng kết thúc là trong Bảng 2.

Bảng 2. Mô tả các thành phần vùng kết thúc theo sáng chế.

Thành phần vùng kết thúc	SEQ ID NO:	Mô tả
t370-2	12	trình tự điều hòa 3' từ gen AF032370 (Syngenta chip ID) từ ngô Zea. Bao gồm 3' UTR và vùng không được phiên mã xuôi chiều.
t393-1	13	Trình tự điều hòa 3' từ gen Zm061393 (Syngenta chip ID) từ ngô Zea. Bao gồm 3' UTR và vùng không được phiên mã xuôi chiều.
t393-2	14	tZm061393-01 với hai đột biến điểm để loại bỏ ORFs không mong muốn.
t855-1	15	Trình tự điều hòa 3' từ gen ZmU45855-3 (Syngenta chip ID) từ ngô Zea. Bao gồm 3' UTR và vùng không được phiên mã xuôi chiều.
t747-1	16	Trình tự điều hòa 3' từ gen Zm001747 (Syngenta chip ID) từ ngô Zea. Bao gồm 3' UTR và vùng không được phiên mã xuôi chiều.
t004-1	17	Trình tự điều hòa 3' từ gen Zm021004 (Syngenta chip ID) từ ngô Zea. Bao gồm 3'-UTR và vùng không được phiên mã xuôi chiều.
t335-1	18	Trình tự điều hòa 3' từ gen Zm015335 (Syngenta chip ID) từ ngô Zea. Bao gồm 3' UTR và vùng không được phiên mã xuôi chiều.

t722-1	19	Trình tự điều hòa 3' từ gen Zm009722 (Syngenta chip ID) từ ngô Zea. Bao gồm 3'-UTR và vùng không được phiên mã xuôi chiều.
t948-2	20	Trình tự điều hòa 3' từ gen Zm058948 (Syngenta chip ID) từ ngô Zea. Bao gồm intron thứ hai, exon thứ ba, 3'-UTR, và vùng không được phiên mã xuôi chiều.

Ví dụ 2: Xây dựng các vectơ biểu hiện trừ hạt phấn

Để kiểm tra các thành phần điều hòa được mô tả ở trên (các vùng khởi động và các vùng kết thúc) trong các cây ngô chuyển gen, tám vectơ nhị phân được tạo ra. Mỗi vectơ nhị phân chứa hai catxet biểu hiện.

Catxet biểu hiện thứ nhất bao gồm vùng tăng cường phiên mã eFMV/e35S, mà được liên kết hoạt động vùng khởi động theo sáng chế cần được kiểm tra, mà được liên kết hoạt động với trình tự mã hóa dị loại (*cry1Ab*) mà mã hóa protein Cry1Ab diệt côn trùng (Patent Mỹ số 5,625,136), mà được liên kết hoạt động với trình tự vùng kết thúc so khớp của vùng khởi động. Ví dụ, catxet biểu hiện được tạo ra để kiểm tra vùng khởi động prZm061393-01 (SEQ ID NO:2) bao gồm vùng kết thúc tZm061393-01 (SEQ ID NO:13).

Catxet biểu hiện thứ hai, mà là giống nhau cho mỗi trong số các vectơ nhị phân, bao gồm vùng khởi động ubiquitin ở ngô (prUbi1-10) (Christensen et al, 1992 PMB 18: 675) được liên kết hoạt động với trình tự mã hóa phosphomannoza isomeraza (PMI; Patent Mỹ số 5,767,378) mà được liên kết hoạt động với vùng kết thúc Ubi1 (tUbi1-01). Việc biểu hiện của PMI cho phép sự lựa chọn dương tính của các thực vật chuyển gen trên mannoza.

Cả hai catxet biểu hiện được nhân dòng vào vectơ thích hợp cho việc biến nạp ở ngô được trung gian bởi *Agrobacterium*. Các vectơ nhị phân mà được xây dựng được liệt kê và được mô tả trong Bảng 3.

Bảng 3. Các vectơ nhị phân được xây dựng để kiểm tra các vùng khởi động trừ hạt phấn ở ngô.

Vectơ nhị phân	SEQ ID NO:	Mô tả
pSYN18499	21	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP370-3 (PMP370-3:cCry1Ab:t370-2)
pSYN18500	22	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP393-1 và t393-1.
pSYN18501	23	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP855-1 và t855-1.
pSYN18498	24	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát PMP747-1. (PMP747:cCry1Ab:t747)
pSYN18617	25	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP004-1. (PMP004-1:cCry1Ab: t004-1).
pSYN18618	26	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP335-1. (prZm015335-01:cCry1Ab:t335-1)
pSYN18619	27	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP722-1 (PMP722-1:cCry1Ab:t722-1).
pSYN18705	28	Vectơ nhị phân với việc chọn lọc PMI chứa cCry1Ab dưới sự kiểm soát của PMP948-2 (PMP948-2:cCry1Ab:t948-2).

Ví dụ 3: Sự biểu hiện nhất thời của Cry1Ab ở ngô được thúc đẩy bởi các vùng khởi động trừ hạt phấn

Các vectơ được mô tả ở trên được biến nạp vào chủng *Agrobacterium tumefaciens* LBA4404 chứa plasmit hỗ trợ (pSBI) sử dụng phương pháp lạnh đông-rã đông (An et al., Binary vector. In: Gelvin S B, Schilproot R A (eds), Plant molecular biology manual. Kluwer Academic Publishers, Dordrecht, pp A3 1-19 (1988)). Việc điều chế các môi trường *Agrobacterium* được tiến hành như được mô tả bởi Azhakanandam et al., Plant Mol. Biol. 63: 393-404 (2007). Tóm lại, *agrobacteria* được cải biến di truyền được sinh trưởng qua đêm trong 50 mL môi trường YP chứa axetosyringon 100 μ M và MES 10 μ M (độ pH 5,6), và sau đó được tạo hạt bằng việc ly tâm ở 4000 x g trong 10 phút. Các hạt được hòa lại trong môi trường chuyển nhiễm (các

muối Murashige và Skoog với các vitamin, 2% sucroza, MES 500 μ M (độ pH 5,6), MgSO_4 10 μ M, và axetosyringon 100 μ M) tới OD600=0,5 và sau đó được giữ ở 28°C trong 2-3 giờ.

Thử nghiệm biểu hiện tạm thời *in planta* việc gieo hạt ngô được thực hiện về cơ bản như được mô tả trong patent Mỹ số 8,642,839, được kết hợp ở đây bằng việc viện dẫn toàn bộ nó. Tóm lại, các hạt ngô được nảy mầm trong các điều kiện nhà kính trong các chậu 2,5 inch được điền hỗn hợp nảy mầm Fafard. Cây con được nuôi theo chu kỳ ngày/đêm 14/10 với cường độ ánh sáng ban ngày là 2000 μ -mol-m-2s-1 được duy trì với nguồn sáng bổ sung. Nhiệt độ được duy trì giữa 23°C-26°C. Thí nghiệm nông chọn lọc được thực hiện hầu như sử dụng các lá sơ cấp và thứ cấp của giai đoạn V2 (Ritchie S. W., Hanway J. J. Benson G. O. (edts): How a Corn Plant Develops: Iowa State Univ Special Report No. 48, July 2005). Để khiến việc thâm nhập dễ dàng hơn, các cây con được tưới nước 1-2 giờ trước việc nông chọn lọc, mà giữ cho lá ẩm ướt và khí khổng mở. Việc thâm nhập các lá riêng lẻ được tiến hành trên các cây non ngô sử dụng thân xi lanh 5 mL (xi lanh 5 ml BD với Luer-Lok.TM. Tip, BD.TM. Franklin Lakes, N.J. 07427, USA), bằng việc ấn đầu của xi lanh lên bề mặt dưới của lá. Các lá nhìn thấy được thứ nhất và thứ hai của giai đoạn V2 được thâm nhập với: 1 ml huyền phù agrobacterium/28 giây/lá. Các cây được thâm nhập được biến nạp và được duy trì trong các điều kiện buồng sinh trưởng được thiết đặt ở 25°C với chu kỳ ngày/đêm 16/8 với cường độ ánh sáng là 1900 μ -mol-m-2s-1. Mô thực vật được thu hoạch sau 4 ngày sau việc thâm nhập cho việc phân tích sau đó của việc phát hiện protein Cry1Ab và protein PMI sử dụng ELISA. Các kết quả chứng minh rằng cả hai protein được phát hiện trong cả tám nhóm vector, biểu thị rằng các vùng khởi động theo sáng chế có khả năng điều khiển việc biểu hiện gen chuyển ở các tế bào ngô.

Ví dụ 4: Sự biểu hiện của Cry1Ab trong ngô được biến nạp ổn định sử dụng các vùng khởi động trừ hạt phấn

Việc biến nạp *Agrobacterium* của các phôi ngô chưa trưởng thành được thực hiện về cơ bản như được mô tả trong Negrotto *et al.*, 2000, Plant Cell Reports 19: 798-803, được kết hợp ở đây bằng việc viện dẫn toàn bộ nó. Tuy nhiên, các thành phần phương tiện khác nhau đã biết trong lĩnh vực có thể được thay thế. Tóm lại, chủng *Agrobacterium* LBA4404 (pSB1) chứa vector biến nạp thực vật được mô tả ở trên được sinh trưởng trên môi trường rắn YEP (dịch chiết nấm men (5 g/L), pepton (10g/L), NaCl

(5g/L), 15g/l agar, độ pH 6,8) trong khoảng 2 – 4 ngày ở khoảng 28°C. Xấp xỉ $0,8 \times 10^9$ *Agrobacterium* được phân tán trong môi trường LS-inf được bổ sung với As 100 μ M (Negrotto *et al.*, *supra*). Vi khuẩn được tạo ra trước trong môi trường này trong khoảng 30-60 phút.

Các phôi chưa trưởng thành từ kiểu gen thích hợp được loại bỏ khỏi các tai từ 8 – 12 ngày tuổi vào chất lỏng LS-inf + As 100 μ M. Các phôi được rửa một lần với môi trường chuyển nhiễm tươi. Dung dịch *Agrobacterium* sau đó được bổ sung và phôi được lắc trong khoảng 30 giây và được cho xử lý với vi khuẩn trong khoảng 5 phút. Các phôi sau đó được chuyển phía lớp màng cứng hướng lên sang môi trường LSAs và được nuôi cấy trong bóng tối trong từ hai đến ba ngày. Sau đó, giữa 20 và 25 phôi mỗi đĩa petri được chuyển sang môi trường LSDc được bổ sung với xefotaxim (250 mg/l) và nitrat bạc (1,6 mg/l) và được nuôi cấy trong bóng tối trong 28°C trong 10 ngày.

Các phôi chưa trưởng thành, tạo ra mô sẹo tạo phôi được chuyển sang môi trường LSD1M0.5S. Các môi trường nuôi cấy được chọn lọc trong môi trường này trong khoảng 6 tuần với bước nuôi cấy phụ ở khoảng 3 tuần. Các mô sẹo sống sót được chuyển sang môi trường Reg1 được bổ sung với mannoza. Tiếp theo sau việc nuôi cấy trong ánh sáng (phác đồ 16 giờ sáng/ 8 giờ tối), các mô xanh sau đó được chuyển sang môi trường Reg2 mà không có các chất điều hòa sinh trưởng và được ủ trong khoảng 1-2 tuần. Các cây con được chuyển sang các hộp Magenta GA-7 (Magenta Corp, Chicago Ill.) chứa môi trường Reg3 và được sinh trưởng trong ánh sáng. Sau khoảng 2-3 tuần, các cây được kiểm tra sự có mặt của gen *pmi* và trình tự mã hóa *cry1Ab-09*. Các cây dương tính từ thử nghiệm PCR được chuyển sang nhà kính và được kiểm tra mức biểu hiện của protein Cry1Ab-09 ở lá và hạt phấn và được kiểm tra hoạt tính chống lại sâu đục thân ngô châu Âu (*Ostrinia nubilalis*).

Các nồng độ của protein Cry1Ab được xác định ở các giai đoạn sinh dưỡng (V) và sinh sản (R) của sự phát triển của các cây ngô chuyển gen. Các giai đoạn “V” được chỉ định các số là V1, V2, V3, v.v. bằng số lượng các lá thông qua giai đoạn VT trong đó nhánh cuối cùng của cụm hoa là nhìn thấy. Các giai đoạn “R” được chỉ định từ R1 đến R6 từ râu đến sự trưởng thành sinh lý của tai, một cách tương ứng. Cry1Ab được định lượng bởi ELISA trong các lá giai đoạn V3-V4 và VT cũng như râu, bẹ và hạt phấn từ 8-12 sự kiện cho mỗi vùng khởi động được kiểm tra sử dụng các quá trình ELISA đã biết trong lĩnh vực. Tóm lại, mô lá được đông khô và sau đó được giảm thành bột mịn

bằng quá trình xử lý sử dụng hoặc máy xay cà phê, máy trộn, máy nghiền Grindomix™ (Brinkmann Instruments; Westbury, N.Y., USA), cối với chày hoặc cối xay, hoặc sự kết hợp của các thiết bị này. Tất cả quá trình xử lý được thực hiện trong sự có mặt của đá khô hoặc nitơ lỏng. Các mẫu được trộn đều để đảm bảo độ đồng nhất. Phần trăm khối lượng khô của mỗi mẫu được xác định và các mẫu được xử lý được lưu trữ ở khoảng -80°C cho đến khi đông khô.

Đối với mỗi mẫu được phân tích, một phần 1,0 g của nguyên liệu thực vật dạng bột (ngoài hạt phần) được cân vào ống 15-ml polypropylen, được phân tán trong 3 ml đệm chiết [CAPS 50 mM, NaCl 0,1 M, EDTA 2 mM, dithiothreitol 1 mM, 4-(1-aminoethyl) benzensulfonyl florua Hcl 1 mM, leupeptin 1 mM, độ pH 10], và được chiết sử dụng bộ làm đồng nhất Autogizer® (Tomtek; Hamden, Conn., USA). Sau việc ly tâm trong 15 phút ở 10,000 x g ở 4°C, dịch nổi được sử dụng cho việc phân tích Cry1Ab và PMI bởi ELISA. Sau việc thử nghiệm với iodoacetamid như được mô tả bởi Hill và Straka (1988), tổng protein trong các phần chiết được định lượng sử dụng chất phản ứng thử nghiệm protein BCA™ (Pierce; Rockford, Ill., USA).

Các phần chiết hạt phần ở ngô được chuẩn bị bằng việc phân tán hạt phần 1:30 (w/v) trong đệm chiết. Sau khoảng 30 phút trên đá, các huyền phù hạt phần được làm gián đoạn bởi ba con đường thông qua tế bào áp suất French ở khoảng 15000 psi, được tiếp theo sau bởi việc ly tâm ở 14,000 x g trong khoảng 5 phút ở 4°C. Các phân tích Cry1Ab và PMI bởi ELISA được thực hiện trên các dịch nổi như được mô tả dưới đây. Tổng protein được định lượng như được mô tả ở trên.

Các dịch chiết được chuẩn bị như được mô tả ở trên được phân tích định lượng cho Cry1Ab và PMI bởi ELISA (Tijssen, 1985). Cry1Ab được định lượng sử dụng dòng đơn được tinh sạch ái lực miễn dịch, kháng thể kháng Cry1Ab và kháng thể kháng Cry1Ab đa dòng được tinh sạch ái lực miễn dịch. PMI được định lượng sử dụng các kháng thể thô đa dòng được tinh sạch Protein A và kháng thể đa dòng được tinh sạch ái lực miễn dịch đặc hiệu cho PMI. Giới hạn định lượng dưới của ELISA kẹp đôi cho Cry1Ab và PMI được ước tính dựa trên nồng độ thấp nhất của protein tham chiếu tinh khiết nằm trên phần tuyến tính của đường chuẩn, thể tích tối đa của dịch chiết đôi chứng mà có thể được phân tích mà không có nhiễu nền, và khối lượng tương ứng của mẫu mà phần mẫu đại diện.

Các kết quả Cry1Ab ELISA được thể hiện trong Bảng 4. Mức độ phát hiện (LOD) đối với thử nghiệm Cry1Ab ELISA được xác định là khoảng 0,37 ng/mg tổng số protein hòa tan (TSP). Các mức định lượng của protein Cry1Ab được phát hiện trong lá V3-V4, râu, bẹ và lá VT từ tất cả các sự kiện đối với mỗi vùng khởi động được thử nghiệm, cao hơn đáng kể so với trong hạt phấn. Các nồng độ của Cry1Ab trong hạt phấn đối với mỗi vùng khởi động chỉ cao hơn LOD đối với ELISA một chút. Năm trong số tám vùng khởi động được thử nghiệm các mức độ biểu hiện của Cry1Ab trong các mô không phải phấn hoa cao hơn LOD gấp 1000 lần. Ba trong số tám vùng khởi động thể hiện Cry1Ab ở mức cao hơn LOD khoảng 500 đến khoảng 950 lần. Bằng việc so sánh, bảy trong số tám vùng khởi động được thử nghiệm biểu hiện Cry1Ab trong hạt phấn ở các mức cao hơn LOD ít hơn 10 lần và một vùng khởi động, prZmU45855, biểu hiện Cry1Ab trong hạt phấn ở nồng độ cao hơn LOD 15 lần. Tuy nhiên, một số thực vật từ mỗi sự kiện của mỗi nhóm vector bao gồm vùng khởi động 001747-01, 061393-01, AF032370-02, 009722-01, 015335-01 hoặc 058948-02 không có mức Cry1Ab có thể phát hiện được trong hạt phấn. Cry1Ab được thể hiện ở mức độ diệt côn trùng trong tất cả các mô không phải hạt phấn được thử nghiệm, trong khi mức độ Cry1Ab trong hạt phấn đối với tất cả các vùng khởi động được thử nghiệm thấp hơn nồng độ cần thiết cho hoạt động diệt côn trùng.

Bảng 4. Biểu hiện Cry1Ab trong mô của cây ngô chuyển gen.

Vector	Vùng khởi động	Mức Cry1Ab trung bình (ng/mg TSP)				
		Lá V3-V4	Tơ	Vỏ	Lá VT	Hạt phấn
pSYN18499	PMP370-2	389	485	290	229	1,6
pSYN18500	PMP393-1	927	359	437	402	2,2
pSYN18501	PMP855-1	910	461	238	693	5,7
pSYN18498	PMP747-1	324	491	505	341	1,1
pSYN18617	PMP004-1	1037	194	960	462	2,7
pSYN18618	PMP335-1	254	109	243	256	0,3
pSYN18619	PMP722-1	1132	283	390	476	1,3
pSYN18705	PMP948-2	292	215	409	318	1,7

Ví dụ 5. Sự biểu hiện của Cry1Ai trong ngô được biến nạp ổn định

Cây ngô được biến đổi như mô tả ở trên với catxet biểu hiện bao gồm trình tự tăng cường phiên mã cFMV/e35S được liên kết hoạt động với vùng khởi động PMP393-4 theo sáng chế (SEQ ID NO: 5), được liên kết hoạt động với trình tự mã hóa dị loại mà mã hóa protein diệt côn trùng Cry1Ai chiều dài đầy đủ (công bố đơn Mỹ số 20190177377), mà được liên kết hoạt động với vùng kết thúc t393-2 theo sáng chế (SEQ ID NO: 14). Mô lá và hạt phần V3-V4 từ 40 sự kiện được thử nghiệm như được mô tả ở trên. Mức độ phát hiện (LOD) đối với thử nghiệm Cry1Ai ELISA được xác định là khoảng 0,75 ng/mg tổng số protein hòa tan (TSP). Các kết quả của ELISA trên các mô từ 40 sự kiện ngô chuyển gen đã chứng minh rằng mức protein Cry1Ai trong lá V3-V4 nằm trong khoảng từ khoảng 10 ng/mg TSP đến khoảng 80 ng/mg TSP, trong khi không có protein Cr1Ai không được phát hiện trong hạt phần, chỉ ra rằng mức độ biểu hiện trong hạt phần thấp hơn LOD của ELISA, tức là <0,75 ng/mg TSP, trong tất cả 40 sự kiện.

Tất cả các công bố và các đơn sáng chế được đề cập trong bản mô tả này là biểu thị ở mức của người có trình độ trung bình trong lĩnh vực mà sáng chế thuộc về. Tất cả các công bố và các đơn sáng chế được kết hợp ở đây bằng việc viện dẫn tới cùng mức như mỗi công bố riêng lẻ hoặc đơn sáng chế được chỉ báo cụ thể và riêng lẻ được kết hợp bằng việc viện dẫn.

Mặc dù sáng chế nêu trên đã được mô tả một số chi tiết bằng cách minh họa và ví dụ cho các mục đích hiểu rõ ràng, sẽ là rõ ràng với người có trình độ trung bình trong lĩnh vực rằng các sự thay đổi và các cải biến nhất định có thể được thực hiện trong phạm vi của sáng chế.

Danh mục trình tự

Danh mục trình tự theo định dạng văn bản ASCII, được nộp theo 37 C.F.R. § 1.821, với tên gọi "81937-REG-ORG-NAT-1_ST25.txt", có kích thước 239 kilobyte, được tạo ra vào ngày 04/09/2019 và được nộp thông qua EFS-Web được cung cấp thay cho bản sao giấy. Danh mục trình tự này được kết hợp ở đây bằng việc viện dẫn vào bản mô tả cho phần bộc lộ của nó.

YÊU CẦU BẢO HỘ

1. Catxet biểu hiện bao gồm vùng khởi động được quy định trong SEQ ID NO:5 được liên kết hoạt động với polynucleotit dị loại được quan tâm, mà được liên kết hoạt động với vùng không được dịch mã 3' bao gồm tín hiệu polyadenyl hóa, trong đó polynucleotit dị loại được dịch mã chức năng trong mô của thực vật chuyển gen mà không phải là hạt phấn.
2. Catxet biểu hiện theo điểm 1, trong đó polynucleotit dị loại mã hóa protein diệt côn trùng hoặc ARN chuỗi kép (dsARN).
3. Catxet biểu hiện theo điểm 2, trong đó protein diệt côn trùng là protein Cry hoặc protein Vip3.
4. Catxet biểu hiện theo điểm 3, trong đó protein Cry là protein Cry1.
5. Catxet biểu hiện theo điểm 4, trong đó protein Cry1 là protein Cry1A.
6. Catxet biểu hiện theo điểm 5, trong đó protein Cry1A là protein Cry1Ab hoặc Cry1Ai.
7. Catxet biểu hiện theo điểm 3, trong đó protein Vip3 là protein Vip3A.
8. Catxet biểu hiện theo điểm 7, trong đó protein Vip3A là protein Vip3Aa.
9. Catxet biểu hiện theo điểm 8, trong đó protein Vip3Aa là protein Vip3Aa20.
10. Catxet biểu hiện theo điểm 1, trong đó thực vật chuyển gen là thực vật một lá mầm.
11. Catxet biểu hiện theo điểm 10, trong đó thực vật một lá mầm là cây ngô.
12. Catxet biểu hiện theo điểm 11, trong đó cây ngô là cây ngô lai cùng dòng hoặc cây ngô lai.
13. Catxet biểu hiện theo điểm 11, trong đó mô là lá, râu hoặc bẹ.
14. Vector tái tổ hợp bao gồm catxet biểu hiện theo điểm 1.
15. Phương pháp biểu hiện protein hoặc polynucleotit được quan tâm ở thực vật hoặc tế bào thực vật bao gồm đưa vào thực vật hoặc tế bào thực vật catxet biểu hiện bao gồm vùng khởi động được liên kết hoạt động với polynucleotit dị loại mà mã hóa

protein hoặc polynucleotit được quan tâm, trong đó vùng khởi động bao gồm SEQ ID NO:5, và trong đó protein hoặc polynucleotit được quan tâm đã nêu được biểu hiện chức năng trong mô của thực vật chuyển gen mà không phải là hạt phấn.

16. Phương pháp theo điểm 15, trong đó protein được quan tâm là protein diệt côn trùng hoặc polynucleotit được quan tâm là dsARN mà diệt côn trùng.

17. Phương pháp theo điểm 16, trong đó protein diệt côn trùng là protein Cry hoặc protein Vip3.

18. Phương pháp theo điểm 17, trong đó protein Cry là protein Cry1.

19. Phương pháp theo điểm 18, trong đó protein Cry1 là protein Cry1A.

20. Phương pháp theo điểm 19, trong đó protein Cry1A là protein Cry1Ab hoặc Cry1Ai.

21. Phương pháp theo điểm 17, trong đó protein Vip3 là protein Vip3A.

22. Phương pháp theo điểm 21, trong đó protein Vip3A là protein Vip3Aa.

23. Phương pháp theo điểm 22, trong đó protein Vip3Aa là protein Vip3Aa20.

24. Phương pháp theo điểm 15, trong đó thực vật chuyển gen là cây ngô.

25. Phương pháp theo điểm 24, trong đó mô là lá, bẹ hoặc râu.

26. Phân tử axit nucleic được phân lập bao gồm trình tự nucleotit được quy định trong SEQ ID NO:5, trong đó phân tử axit nucleic đã nêu khởi đầu việc phiên mã chức năng của polynucleotit được liên kết hoạt động được quan tâm trong mô của thực vật chuyển gen mà không phải là hạt phấn.

27. Phân tử axit nucleic được phân lập theo điểm 26, trong đó polynucleotit được liên kết hoạt động được quan tâm mã hóa protein diệt côn trùng hoặc dsARN.

28. Phân tử axit nucleic được phân lập theo điểm 27, trong đó protein diệt côn trùng là protein Cry hoặc protein Vip3.

29. Phân tử axit nucleic được phân lập theo điểm 28, trong đó protein Cry là protein Cry1.

30. Phân tử axit nucleic được phân lập theo điểm 29, trong đó protein Cry1 là protein Cry1A.

31. Phân tử axit nucleic được phân lập theo điểm 30, trong đó protein Cry1A là protein Cry1Ab hoặc Cry1Ai.
32. Phân tử axit nucleic được phân lập theo điểm 28, trong đó protein Vip3 là protein Vip3A.
33. Phân tử axit nucleic được phân lập theo điểm 32, trong đó protein Vip3A là protein Vip3Aa.
34. Phân tử axit nucleic được phân lập theo điểm 33, trong đó protein Vip3Aa là protein Vip3Aa20.
35. Phân tử axit nucleic được phân lập theo điểm 26, trong đó thực vật chuyển gen đã nêu là cây ngô chuyển gen.
36. Catxet biểu hiện theo điểm 1, trong đó vùng không được dịch mã 3' bao gồm trình tự vùng kết thúc được lựa chọn từ bất kỳ trong số các SEQ ID NO:12-20.
37. Vector tái tổ hợp theo điểm 14, trong đó vector đã nêu là vector nhị phân bao gồm bất kỳ trong số các SEQ ID NO:21-28.
38. Phương pháp biểu hiện đặc hiệu trình tự mã hóa dị loại trong mô thực vật chuyển gen mà không phải là hạt phấn, bao gồm: a) biến nạp các tế bào thực vật với vector, trong đó vector đã nêu bao gồm catxet biểu hiện theo điểm 1; b) phát triển các tế bào thực vật chuyển gen bao gồm catxet biểu hiện đã nêu; và c) tạo ra các thực vật chuyển gen từ các tế bào thực vật được biến nạp đã nêu trong đó trình tự mã hóa dị loại được biểu hiện đặc hiệu trong mô mà không phải là hạt phấn trong sự kiểm soát của phân tử axit nucleic đã nêu.

DANH MỤC TRÌNH TỰ

<110> Syngenta Crop Protection AG

Clarke, Joseph Azhakanandam, Kasi Zhou, Ailing Conville, Jard

<120> CATXET BIỂU HIỆN BAO GỒM VÙNG KHỞI ĐỘNG, VECTO TÁI TỔ HỢP VÀ PHƯƠNG PHÁP BIỂU HIỆN PROTEIN HOẶC POLYNUCLEOTIT ĐƯỢC QUAN TÂM

<130> 81937-US-L-ORG-NAT-1

<160> 29

<170> PatentIn phiên bản 3.5

<210> 1

<211> 2738

<212> ADN

<213> Zea mays

<400> 1

gagatatgtg atgtatatgt gatatatgtg atgattatgt gatataTTTT ttgtttgttt	60
ggatggaata ataaaaacaa ataaaaaggg tatgtttggcc acttttggcg agtgtaacac	120
tcagcaaaga ggtactttgc cgagtgtcac agtcatatca ttoggcaaag aaggcacaca	180
tgggaaccga taaagcttct ttgccgagtg ttgtggcctt agcactcggc aaagaagcaa	240
cttttgcoga gagcctccta gtgtactcgg caaaggaact gacaaaggag cccactgatg	300
attcctttga tgagtggtag tccggcagac acacggcaaa gataaagtct ttgccgagtg	360
ccaccttaata cggccggcaa aggaactggc aaaggacaca cgatgagttt atttgtcgag	420
tgctagtaca atagacactc gacaaagagt gagcctttgc agagtgtcac cgtggcactc	480
ggcaaagtcc actctttgtc gagtgtcacc gtgacactca gcgaagtctc cgtcgtgtgc	540
atctgtcgcc gtgacgtcga cttttttttg ccgagtaccg agtggttgcc gagtgtccga	600
caaaaaatac tcgggttaaag accgttgccg atgtcagttc accaagacct tttttatcaa	660
gtgtcatact tgataaagtc ttccccaagt attttcaaaa ctttgtcgat tacctcacac	720
accaaggggt aggatcgtgt tgaggcgggt tgctgcttgt ttcttgcttt cggcgaacca	780
acggaccatg agcgcataaa ccaatgggcg agcggccgca tatacaggcc tgcgcggcga	840
gcctcacctt gctggatttg aatgccccct cggcatggtc aatgaccatg catctttgtg	900
cagaacatat aaaatgcaca aattaattaa ttaataaaaa tacatgaaag agtacataat	960
ttgtttgtgg attattaatt atatttataa atatattata gtatattata ttttatcttt	1020
atctaactac aaaaataatt atagattgtt aacttttttt tgcattactg gtaagatttt	1080
gtcgatttla taattttgtc taacatgttt tattatcaat ctaaaattgc aaatttatatg	1140
ttaacgaact aaatttgtaa taaaatatat gtaataaagg tcgtctaaaa gtaagctaaa	1200
attcacgcac ctaaaatttla tttagtaaag tgttttttct tgacatatga cactagtgtt	1260
ttcaaagctt gtctcttaat ttaggattaa attaataggc ctctctaaat tgctttttat	1320
aagataaata aatcctattc cgtctgctca tcttcaaaga acatttcaat gaaacaaata	1380


```

tttgggaaaa ccaatggaaa aagaataaaa ataaaaattg aaaaagagaa gtggaggaaa      1440
agtaggggttc cattccacgg cctgtcgagt gtcgaccac aaaccacgca accgatatat      1500
tccccagcac cccagtcctc cagccgtcca accgagacac cgcgtgcgaa ccaagcagac      1560
cacaacaaga agaagcgtag tcgtcgccgg aaggaaaggc gcggagcaag atctcgtggc      1620
aggcgtacgt cgacgagcac ctgaagtgcg agatcgaagg acagcatctc agcgccggcg      1680
ccatcgtcgg tcacgacggc agcgtttggg cgcagtcgga gagcttcccc gaggtcattc      1740
actcccacct tatctcttcc cgctctactt gcttagttgc tttatccatg ttcagtgctc      1800
ctgcgctgat ttagaccgcg gcgactctgg atctgcccc gccgtgcggc ccggccggat      1860
cggagcggag ggatgtccgg atctcggctg ctgaggcgag atgcggatct aaggccccct      1920
ccctcgggtga cgtgggttgc gctgctgctg ttctccgggc gtttgatgcc aatatagctg      1980
agatcaagct tgatgatctg cgtactgtgg atttgctagt gagggatggc cggatcgggc      2040
tgttggcgtc cgacctgatt tgggtgttcg cttgaataga cttacttacc agtggacagt      2100
ggtagtgatc gttcagcagc agtagagcaa tttgttttac atgtaaaatt tgagtggctg      2160
ctagagatgc accaatggcc gagctgcatg tgttttattcc tatatagcga cagttgtgtg      2220
attgatccaa aacgcaaaat tttagttttc atcaaatgat gatttcctga tgccactacc      2280
tgtgtgtgag gcacttatta aatgaaaagt aactttcggg tgggatctgg gaatctgtat      2340
tgtgttccgt ggttggtcac agcttgtggc ctgcattggt ccataggcca aatggctagg      2400
caaggaaaat aaatcatgta atttggaaaa aaataactgc catagtcagt agtgtacagc      2460
agtacatgct tgtccggggc ttttatttca ccatcgtggt gtggtatatg gagtacttgg      2520
attgtgcgat gctttataca tttgtttggt ctagcagagt ttgagtttat atttctatcc      2580
ttatgtaggg cgtcacgagt gtcgtcaatt gttacatcct tacatccttg tgttcacaat      2640
atgatccatc aactctttct tctgcgggtct cttatccctg tgttctcatg tccatgtttc      2700
ctgcttcttg ttgcagttaa agccttagga ggtaaacc      2738

```

<210> 2

<211> 2627

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Trình tự điều hòa 5' từ gen Zm061393 với đột biến T thành A để loại bỏ codon bắt đầu.

<400> 2

```

tctgcctttc tgttcttcaa acgatgtctc atgtctgcgc tggacaactt tcttgttgcc      60
gcctgtcgct tgcgctgtgc tgactggacg cagctccgga gggttggttg tgcttggttt      120
tcgtagagaa ctgcgcactt gccgccgca cgttcttggt gtttctcctt ccccgctgtg      180
ttctgcgcac gggctttttc tgagagaccc atgtttccct tttactttta taaacagtat      240

```

acatgctatg tttctagaag gaggggaaac ctaatcccc taatccaatg gcggggagga 300
 aataggggtg ggtgggggtg ggggagggaa atatctcgct actttttaat ccggacaagc 360
 tcatttgctg ttgctgtctga atgatgatga ctgcaatgct gatcgcacct cgggtgtcgg 420
 atcaccagct tttggctgct ctacacaaat cagctgcaag aagattagag cacaaaagaa 480
 ttacagaaaag agagcctttt tcttttcttc cttgtggggg tcttttcatt tcgtgctctc 540
 ctttctctgc cagccagtcc gtccttgctg ccaactgcacc tgcacacagg tcaccccgac 600
 ccgcactgtt ctagactcca ttagaaaaaa aaaggctctga acctttccga aaccagccag 660
 ccattgggtct ggcaggccag catatgctaa ttggattttt ttgccgcac attgagtgcg 720
 ccatcaggat ttggaaatcc tggttttgag taatacagta atttggcatt atccattgcc 780
 gaattcccaa gctccgtcag cttgaacgtg gaccctacc atctgcacca gctcggcacc 840
 tcacgtctgc agcgttagga gcctaggagc agctgcccg tcttttattg gtccctctcc 900
 cgtcccagag aaaccctccc tccctcctcc attggactgc ttgctccctg ttgaccattg 960
 gggatatgctt gctgccttgc tctcctgttc atctccgtgc taaacctctg tctctgggt 1020
 gggtttttgc tgggattttg agctaactct ctgggtcccg tagaaaaaga tcatgtcccc 1080
 tgacgtgctc aagcgtctgc cttagccgag tcttgcccc ccgccatttt ttgcggtttc 1140
 ggtgtgttcc cgtgactcgc cgggtgcgtc atcgctgaa tcttgtctgg gctctgctga 1200
 catgttcttg gctagtggg tttatagatt cctctgatct aaaccgtgcc tgtgctgcgc 1260
 acagaactct cccctgtcct ttcctggggg tttggttacg tgggtgtagt aagcttgat 1320
 ttgcacatgg ataaagtgt tctaagctcc gtgggttgc tgcagatctt ctgttattgc 1380
 gtgccgtgct cacttttttt gcaatccgag gaatgaattt gtcgtttact cgttttgggtg 1440
 gattattagc gcgaaaaaaa aactcttttt ttttgttct tttactacga aaagcatctt 1500
 cttggatttt gctatcttct tttactacga aaaactcttg agtctaggaa tttgaatttg 1560
 tgatgtccat tcttgcaagt cgctgtgctt tattgggaag ccaaactcta ttattttctg 1620
 cctctagggt ctgaatggaa tcagtactct tgagacagaa aatcaatcca atcaagtga 1680
 tttctttctt taaaaatatt atcacagaac taagtgcctg tgcggaatca gtactggctt 1740
 ttgtttgggt gaggatcaat acttgctttt gtttgggggt ggcaactgtt ttgctataag 1800
 attccatgtg ttcctgttga gatgaatcat atatagtata gctgcatact acaaatctgt 1860
 ttttcaaatt taggttgctt tggcatgac tatttttttg tcagacagac tttctaagt 1920
 gtagctcttg atttcttgtt cttgtacaac tgggtgctgct gaatcttgac cgtatagctc 1980
 gaattgcagt attctgaacc atcgagccaa ggctgccaa ctgactcgcc tccacagtct 2040
 tcgcgaacgc cttggtgcca ccttctcttc ccatcccaat gaactgatag cactcttttc 2100
 cagggtgggt taccaaaatc atataacttg catttcattc ggtactgaaa gttgttaatt 2160
 tgttattctc ttcatgcctg tcttaatagc acaccagat gtaaacacga gattatgcaa 2220
 cttcttactt ggtttctttt gttggcacca tcatgcatgc taattgctaa ggatgttacc 2280
 tattcatcct tgactcatat tatcatatgt aatgatttta tgatcacgag actattgatt 2340
 gtgaagcata gtatagctgt tcttcagttt ttgtaccctt ttgtttttt ccttaagcta 2400

gaactggtac aatttagttg ataagacagt gtagtttgta gtacgtcatt tgacagattg 2460
 tttgtcttta gctggtaaag tgccatttaa tatctgtatc cttcagatct aataaaaagg 2520
 atatgagatg tccatcacaa gaggggaaaa attacatgat ctgagatgta acatccgttt 2580
 ttatttgtga aataccactt ctacaggtat cttcactagg gtaaacc 2627

<210> 3

<211> 2695

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vùng tăng cường TMV được bổ sung vào đầu 3'.

<400> 3

tctgcctttc tgttcttcaa acgatgtctc atgtctgcgc tggacaactt tcttggtgcc 60
 gcctgtcgct tgcgctglgc tgactggacg cagctccgga gggttggttg tgcttggttt 120
 tcgtagagaa ctgcgcactt gccgccgca cgttcttggg gtttcctcct ccccgctgtg 180
 ttctgcgcac gggctttttc tgagagacc cagtttcctt tttactttta taaacagtat 240
 acatgctatg tttctagaag gaggggaaac ctaatcccc taatccaatg gcggggagga 300
 aataggggtg ggtgggggtg ggggaggga atatctcgct actttttaat ccggacaagc 360
 tcatttgctt ttgcgtctga atgatgatga ctgcaatgct gatcgcacct cgggtgtcgg 420
 atcaccagct tttggctgct ctacacaaat cagctgcaag aagattagag cacaaaagaa 480
 ttacagaaaag agagcctttt tcttttcttc cttgtggggg tcctttcatt tcgtgctctc 540
 ctttctctgc cagccagtcc gtccttgctt cactgcacc tgcacacagg tcaccccgac 600
 ccgcactgtt ctagactcca ttagaaaaaa aaagggtctga acctttccga aaccagccag 660
 ccattgggtt ggcaggccag catatgctaa ttggattttt ttgccgcac attgagtgcg 720
 ccatcaggat ttggaaatcc tggttttgag taatacagta atttggcatt atccattgcc 780
 gaaltcccaa gctccgtcag cttgaacgtg gaccctacc atctgcacca gctcggcacc 840
 tcacgctcgc agcgctagga gcctaggagc agctgccgt ctatttattg gtccctctcc 900
 cgtcccagag aaacctccc tccctccctc attggactgc ttgctccctg ttgaccattg 960
 gggatatgctt gctgccttgc tctcctgttc atctccgtgc taaacctctg tcctctgggt 1020
 gggtttttgc tgggattttg agctaactct ctgggtcccg tagaaaaaga tcatgtcccc 1080
 tgacgtgctc aagcgctcgc cttagccgct tcttgcccc ccgccatttt ttgcggtttc 1140
 ggtgtgttcc cgtgactcgc cgggtgcgtc atcgccgtgaa tcttgtctgg gctctgctga 1200
 catgttcttg gctagtggg tttatagatt cctctgatct aaaccgtgcc tgtgctgcgc 1260
 acagaactct cccctgtcct ttccctgggg tttggttacg tgggtgtagt aagcttggat 1320
 ttgcacatgg ataaagtgt tctaagctcc gtgggttgct tgagatcttg ctgttattgc 1380
 gtgccgtgct cacttttttt gcaatccgag gaatgaattt gtcgtttact cgttttggtg 1440

gattattagc gcgaaaaaaa aactcttttt tttttgttct tttactacga aaagcatctt 1500
 cttggatttt gctatcttct tttactacga aaaactcttg agtctaggaa tttgaatttg 1560
 tgatgtccat tcttgcagtg cgctgtgctt tattgggaag ccaaatccta ttattttctg 1620
 cctctagggc ctgaatggaa tcagtactct tgagacagaa aatcaatcca atcaagttga 1680
 tttctttctt taaaaatatt atcacagaac taagtgtctg tgcggaatca gtactggctt 1740
 ttgtttggtg gaggatcaat acttgccttt gtttgggggt ggcaactgtt ttgctataag 1800
 attccatgtg ttcctgttga gatgaatcat atatagtata gctgcatact acaaactctg 1860
 ttttcaaatt taggttgctt tggcatgac ttttttttg tcagacagac tttctaagt 1920
 gtagctcttg atttcttgtt cttgtacaac tgggtgtgct gaatcttgac cgtatagctc 1980
 gaattgcagt attctgaacc atcgagccaa ggctgccaaag ctgactcgcc tccacagtct 2040
 tcgcgaacgc cttggtgcc ccttctctc ccatcccaat gaactgatag cactcttttc 2100
 caggtgggct taccaaaatc atataacttg catttcattc ggtactgaaa gttgttaatt 2160
 tgttattctc ttcattgctg tcttaatagc acaccagat gtaaacacga gattatgcaa 2220
 cttcttactt ggtttctttt gttggcacca tcatgcatgc taattgctaa ggatgttacc 2280
 tattcatcct tgactcatat tatcatatgt aatgatttta tgatcacgag actattgatt 2340
 gtgaagcata gtatagctgt tcttcagttt ttgtaccctt ttgttttttt ccttaagcta 2400
 gaactggtac aatttagttg ataagacagt gtagtttgta gtacgtcatt tgacagattg 2460
 tttgtcttta gctggtaaag tgccatttaa tatctgtatc cttcagatct aataaaaagg 2520
 atatgagatg tccatcacia gaggggaaaa attacatgat ctgagatgta acatccgttt 2580
 ttattttgtg aataccactt ctacaggtat cttcactagg gtttgtcgaa caacaacaaa 2640
 caacaacaaa caaagtcgaa caacaacaaa caacaacaaa caaagtcgac caaaa 2695

<210> 4

<211> 2694

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Bao gồm đột biến điểm để ức chế ORFS.

<400> 4

tctgcctatc tgttcttcaa acgatgtctc atgtctgcgc tggacaactt tcttgttgcc 60
 gcctgtcgct tgcgctgtgc tgactggacg cagctccgga ggtttggttg tgcttggttt 120
 tcgtagagaa ctgcgcactt gccgcccgca cgttcttggg gtttctctct ccccgctgtg 180
 ttctgcgcac gggctttttc tgagagaccc aagtttccct tttactttta taaacagtat 240
 acatgctatg tttctagaag gaggggaaac ctaatcccc taatccaatg gcggggagga 300
 aataggggtg ggtgggggtg ggggagggaa atatctcgct actttttaat ccggacaagc 360
 tcatttgcgt ttgcgtctga atgatgatga ctgcaatgct gatcgcacct cgggtgtcgg 420

atcaccagct tttggctgct ctcaccaa	at	cagctgcaag aagattagag cacaaaagaa	480
ttacagaaaag agagcctttt tcttttcttc	ct	cttgtgggggt tcctttcatt tcgtgctctc	540
ctttctctgc cagccagtcc gtcccttgct	cc	actgcacc tgcacacagg tcaccccgac	600
ccgcactggt ctagactcca ttagaaaaaa	aa	aggtctga acctttccga aaccagccag	660
ccattggtct ggcaggccag catatgctaa	tt	ggattttt ttgccgcac attgagtgcg	720
ccatcaggat ttggaaatcc tggttttgag	ta	atacagta atttggcatt atccattgcc	780
gaattcccaa gctccgtcag cttgaacgtg	g	acccctacc atctgcacca gctcggcacc	840
tcacgctcgc agcgctagga gcctaggagc	ag	ctgcccgt ctattttattg gtccctctcc	900
cgtccagag aaacctccc tccctcctcc	att	ggactgc ttgctccctg ttgaccattg	960
gggtatgctt gctgccttgc tctcctgttc	at	ctccgtgc taaacctctg tcctctgggt	1020
gggtttttgc tgggattttt agctaactctg	ct	ggccccg tagaaaaaga tcatgtcccc	1080
tgacgtgctc aagcgctcgc cttagccgcg	tc	cttgcccc ccgccatttt ttgcggtttc	1140
ggtgtgttcc cgtgactcgc cgggtgcgtc	at	cgccctgaa tcttgtctgg gctctgtga	1200
catgttcttg gctagtggg tttatagatt	cc	ctctgatct aaacctgtcc tgtgctgcgc	1260
acagaactct cccctgtcct ttccctgggg	tt	tgggttacg tggtggtagt aagcttggat	1320
ttgcacatgg ataaagtgt tctaagctcc	gt	gggttgct tgagatcttg ctgttattgc	1380
gtgccgtgct cacttttttt gcaatccgag	ga	atgaattt gtcgtttact cgttttggtg	1440
gattattagc gcgaaaaaaa aactcttttt	tt	ttttgttct tttactacga aaagcatctt	1500
cttggatttt gctatcttct tttactacga	aaa	actcttg agtctaggaa tttgaatttg	1560
tgatgtccat tcttgcagtgc cgtgtgctt	tatt	gggaag ccaaatccta ttattttctg	1620
cctctagggt ctgaatggaa tcagtactct	tg	agacagaa aatcaatcca atcaagtga	1680
tttctttctt taaaaatatt atcacagaac	ta	agtgtcttg tgcggaatca gtactggctt	1740
ttgtttggtg gaggatcaat acttgctttt	gttt	gggggt ggcaactgtt ttgctataag	1800
attccatgtg ttccctgttg gatgaatcat	ata	tagtata gctgcatact acaaactctgt	1860
ttttcaaatt taggttgctt tggcatgac	tatt	tttttttg tcagacagac tttctaagtg	1920
gtagctcttg atttcttgtt cttgtacaac	tgg	tgtgctgt gaatcttgac cgtatagctc	1980
gaattgcagt attctgaacc atcgagccaa	gg	ctgccaaag ctgactcgcc tccacagtct	2040
tcgcgaacgc cttggtgcc ccttctcctc	cc	atcccaat gaactgatag cactcttttc	2100
cagggtgggct taccaaaatc atataacttg	catt	ttcattc ggtactgaaa gttgttaatt	2160
tgttattctc ttcatgcctg tcttaatagc	ac	accagat gtaaacacga gattatgcaa	2220
cttcttactt ggtttctttt gttggcacca	tc	atgcatgc taattgctaa ggatgttacc	2280
tattcatcct tgactcatat tatcatatgt	aat	gattttta tgatcacgag actattgatt	2340
gtgaagcata gtatagctgt tcttcagttt	tt	gtaccctt ttgttttttt ccttaagcta	2400
gaactggtac aatttagttg ataagacagt	g	tagtttgta gtacgtcatt tgacagattg	2460
tttgtcttta gctggtaaag tgccatttaa	tat	ctgtatc cttcagatct aataaaaagg	2520
atatgagatg tccatcacaa gaggggaaaa	att	acatgat ctgagatgta acatccgttt	2580

ttattttgtga aataccactt ctacaggtat cttcactagg gtttgtcgaa caacaacaaa 2640
 caacaacaaa caaagtcgaa caacaacaaa caacaacaaa caaagtcgac caaa 2694

<210> 5

<211> 2626

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Bao gồm đột biến điểm để ức chế thành phần chéo ORF trên mạch đối nghĩa.

<400> 5

tctgcctttc tgttcttcaa acgatgtcac atgtctgcgc tggacaactt tcttgttgcc 60
 gcctgtcgct tgcgctgtgc tgactggacg cagctccgga ggtttggttg tgcttggttt 120
 tcgtagagaa ctcgccactt gccgcccga cgttcttggt gtttcctcct ccccgctgtg 180
 ttctgcgcac gggctttttc tgagagaccc atgtttccct tttactttta taaacagtat 240
 acatgctatg tttctagaag gaggggaaac ctaatcccc taatccaatg gcggggagga 300
 aataggggtg ggtgggggtg ggggaggga atatctcgct actttttaat ccggacaagc 360
 tcatttgcgt ttgcgtctga atgatgatga ctgcaatgct gatcgcacct cgggtgtcgg 420
 atcaccagct tttggctgct ctacacaaat cagctgcaag aagattagag caaaaagaa 480
 ttacagaaag agagcctttt tcttttcttc cttgtgggt tcctttcatt tcgtgctctc 540
 ctttctctgc cagccagtcc gtccttgct ccactgcacc tgcacacagg tcaccccgac 600
 ccgcactgtt ctagactcca ttagaaaaaa aaaggtctga acctttccga aaccagccag 660
 ccattggtct ggcaggccag catatgctaa ttggattttt ttgccgcac attgagtgcg 720
 ccatcaggat ttggaaatcc tggttttgag taatacagta atttggcatt atccattgcc 780
 gaattcccaa gtcctgtcag cttgaacgtg gaccctacc atctgcacca gtcggcacc 840
 tcacgtctgc agcgttagga gcctaggagc agctgccgt ctatttattg gtccctctcc 900
 cgtcccagag aaaccctccc tccctcctcc attggactgc ttgctccctg ttgaccattg 960
 gggatatgctt gctgccttgc tctcctgttc atctccgtgc taaacctctg tctctgggt 1020
 gggtttttgc tgggattttg agctaactct ctgggtcccg tagaaaaaga tcatgtcccc 1080
 tgacgtgctc aagcgtcgc cttagccgcg tccctgcccc ccgccatttt ttgcggtttc 1140
 ggtgtgttcc cgtgactcgc cgggtgcgtc atcgccgtgaa tcttgtctgg gctctgctga 1200
 catgttcttg gctagttagg tttatagatt cctctgatct aaaccgtgcc tgtgctgcgc 1260
 acagaactct cccctgtcct ttcctgggg tttgggttac tgggtggtagt aagcttggat 1320
 ttgcacatgg ataaagtgt tctaagctcc gtgggttgct tgagatcttg ctgttattgc 1380
 gtgccgtgct cacttttttt gcaatccgag gaatgaattt gtcgtttact cgttttggtg 1440
 gattattagc gcgaaaaaaa aactcttttt ttttgttct tttactacga aaagcatctt 1500

```

cttggatttt gctatcttct ttactacga aaaactcttg agtctaggaa tttgaatttg 1560
tgatgtccat tcttgacgtg cgctgtgctt tattgggaag ccaaatccta ttattttctg 1620
cctctagggg ctgaatggaa tcagtactct tgagacagaa aatcaatcca atcaagttga 1680
tttctttctt taaaaatatt atcacagaac taagtgcctg tgcggaatca gtactggctt 1740
ttgtttggtg gaggatcaat acttgctttt gtttgggggt ggcaactgtt ttgctataag 1800
attccatgtg ttcctgttga gatgaatcat ataatgtata gctgcatact acaaatctgt 1860
ttttcaaatt taggttgctt tggcatgac tatttttttg tcagacagac tttctaagt 1920
gtagctcttg atttcttggt cttgtacaac tgggtgctgt gaatcttgac cgtatagctc 1980
gaattgcagt attctgaacc atcgagccaa ggctgccaa ctgactcgcc tccacagtct 2040
tcgcgaacgc cttggtgcc ccttctctc ccatcccaat gaactgatag cactcttttc 2100
caggtggggt taccaaaatc atataacttg catttcattc ggtactgaaa gttgttaatt 2160
tgttattctc ttcattgctg tcttaatagc acaccagat gtaaacacga gattatgcaa 2220
cttcttactt ggtttctttt gttggcacca tcatgcatgc taattgctaa ggatgttacc 2280
tattcatcct tgactcatat tatcatatgt aatgatttta tgatcacgag actattgatt 2340
gtgaagcata gtatagctgt tcttcagttt ttgtaccctt ttgttttttt ccttaagcta 2400
gaactggtac aatttagttg ataagacagt gtagtttgta gtacgtcatt tgacagattg 2460
tttgtcttta gctggtaaag tgccatttaa tatctgtatc cttcagatct aataaaaagg 2520
atatgagatg tccatcacia gaggggaaaa attacatgat ctgagatgta acatccgttt 2580
ttatttgtga aataccactt ctacaggtat cttcactagg gtaaac 2626

```

<210> 6

<211> 2510

<212> ADN

<213> Zea mays

<400> 6

```

atgccttacg atcatctgac atcttaaata tttcaaactg cagtatttca aaaactgtgg 60
ttttgtcaaa aactttgttc ccaaactgag actaaaatgt agtgctagat aataaggcct 120
taccggttg aatgacaaaa aagaaacaga aaacacctga aaaatagtat tgccttggct 180
ctaaattata aatcgttttg gttttattta atatctattt tttaagcata ctaaaatcta 240
tgtacctatg aaaaaataaa tatgaactac aatttaagac tgagtacagt gtacagcgat 300
aagttatatg aaccacaatt gtttttgaaa ggaaaaaac taaattgaca agagaaacaa 360
tcaactcttt tagaaacttt gtataccaaa aaacgaaata aagaagaaga aaaaaatata 420
cacgatgagt ttattagccc acaaaacatg aaaaaaatg agagttaaag ctaactcag 480
gccacaaaag aatcacaagc gcccatattc acagccgccg aaaggtttcg gccccagtt 540
ggatctcaga gccagtggtg aaataggact gacttgtctc tgatggttac gttacgtagg 600
gattgaaaca aacaggcagt gagccgatgt ggcagaccgg agcgcagaag cgacgacgac 660

```

catgaactga aggcgctggc agtggggccat gcgcccagaa cgctgacgaa ccacgacgac 720
gcggacgttc tacgaggtcg gtggggccgt cggcagcccg ccggagcata ttcgccgtcg 780
acagaggcag agaaggggaa tgtgtgggag catattccac ccgtgggaag cctcggagtc 840
ggagatgtgc agatctgccg gacgcacatg cgcacagctg ccgcaataga aaccagtgtc 900
tttttatttt attttataag aaaaaactag aaaagaaaaat ccagtcttct atacaggcct 960
gcacagagag agaggaggga aggtcaaate gtagaaaaaa aatgcttcct tttattgtta 1020
cgacattggc ctgaatatat actcattgtt tctttcctta acttcaacaa ttattgattt 1080
tgacaatcca ttttttattt tagttcgcca aggttttggg tggagaagaa tttaaaacct 1140
agccaagcag tcgagcacga gagccgccgc gcgccgccgc gccgttggca ggacctgctg 1200
acctctgacg accggacacg gacgttttcc aaagtccaaa ctcggggtag gtaggtcgtg 1260
aggctgctgc gttcgccaa ggaagaagca tctagttgaa gctagaaacg ggcagggggg 1320
agggggccgac gctatcggaa atctgaagcc cagcacggac ggacggacgg gcagggggcg 1380
cggcacatcc tcgtctgacc gccgcgccgc gtccgacatg cgcaagctgc tgctcttccc 1440
gcgccccccac ggcgggccggg ccggcagagc acgggacgcg cggggcgggc ctgacgtcgt 1500
tttcacgtgg tccgtcaccg ttgcttgcac agcgtacagg agtatagtcc catttccgaa 1560
acaaaaaaa gtaaacaaaa aaatgccgtc gcattagtgt gctggcaagg gaagaccaag 1620
atggttttca aatcaaacat gtaaaaaaat gtttctttct tttttgaacg aacttaccag 1680
ttcgaattct ttaccggcat tattggitta aattgttccc caagaaagaa agaaaagaaa 1740
agcgggggga aatatgcctt ccttcttctt tgcttcacag accacgcgcg gataccgcg 1800
acaacagcca ggcagcaagc agagcaccca ccgccggaag cggccgtcgg acagaacgac 1860
acgtggggca cggcacagtc cggggccgca cgtcatcgaa gacacctgac ctgcccattg 1920
gtcttctaga gaaaggcgga cggacaggtc acccccgcac ctcatccag gagttgcatt 1980
tcgctttctc ctttatttat ttataccaat aaaaaataaa tccgcctttt ctctcccta 2040
tcgtgtgtct tctctcgcg ggcttttaaaa acgcacacaa gcgctaaaac cctctccacc 2100
gtccacctca gctcccatat ccgtccctt acctctccag catctcccg tctcgtcgt 2160
ctcgtctccg ctctcacct cgcgcgtagc caagggtaac tctcgtctcc cggccggccc 2220
ccgctcggg gttttccatt tactactact cgtgctccc tctgctccg tgctcagact 2280
cagatccgac caaagcgggt tcgctgacta aattctcctt cgtttttttt tctttttctg 2340
gacggattcc tggacgcagg caagatcaag atcggaatca acggtgagtc tgtaccccca 2400
acacaaactc gttcttctct ctcggttcgt tgggtctgga ctctggagtg atctgagtgg 2460
ggtctctgcg cgctgcgtg ctgaggttct tgaaggatcg gcagtaaacc 2510

<210> 7

<211> 2046

<212> ADN

<213> Zea mays

<400> 7

aaatgaatta atatattatc taaattcaga cgggaaaaag agagaaggaa tcaggaaggc	60
tgagtgcaat attagttttt cattgagcac aatottatta tgtttgatga acatottgaa	120
caaagtatga gtggaaaaca acacttatgt tatatctgca gcacagattg ttagaagtga	180
aggctacctt gagatgggtga taggtctctg aacctatctg cattgcttct tcgaagttca	240
tagcaccaac aagaagaatc ataatttctt acaaataaga aatgtttata gttactccaa	300
aactacagtt aaacatattg atatgggtga tatacatgaa caactaactg tcttaatatg	360
tttctgttta ggcaaaagat gacacaagtg gacaaagtaa gggaataatg cattcaactt	420
ctcagactct aaaggtttgg atgaatactt gaatgggaag accatttcca gcatgctttc	480
caccattaat gactgtattt gcagggacag gaagagttgt agcgcttttg ccaacaagat	540
ccgctatatg cttgtagagt ggaacctgca cagcaacaag catccaatga tctgagaata	600
agaattatgt gcggtggcat gacacattta ccaaattatg atgattttag cagaagagca	660
gcacctcttt ttcagcagca ccagctttgc aagctgcaat tgacactgcc agcatagtat	720
ttgctccaag ctcagcctgc atttcgcaca atatcatatc tatttccatt tggataaaat	780
ctattttgcca gaaatatggg aacggaatca agatacagaa aaaagttgag gctgtagtgg	840
taagtgggaa cacaagtaca gtgaaaagat agcattattt tccgaacagt caaattaacc	900
gaaataagta gtatcacata ctgaccctaa attgactaca cccaatgtaa agaattgctt	960
gtgcatattg agccatcagt tgtttgtttg tgtaatggat ttgacaatgc agaattcagg	1020
agcaaaaaaa accgcaaagg agagagaggg aaaggatctc cgagtatccc acctgtggcg	1080
cgccttgtea agtccatgat ggcttggtcg atctgcgcct gctgctgcgg atccatccca	1140
acgagtgcct ccgacacctt gtcgttgatc acccgaccg cgtaggcgac cccctggcg	1200
aggaacttcc gcctctcgga ggccccggcc gcgtcggcgg cggcgccac gagcgcgccc	1260
acgctagccg tagatgctcg gtgcacggcc ttgttggtgt gcaactcgac ctcaacggcc	1320
ggtgcgctac gcccgtaag gatctggcgc gcccgaccc tctgtatcac ggtgggcgcc	1380
tccgcctca tgtggccgac gatgaaaagc gccgggtcag gggccttggc gcggagcgcc	1440
gtattcaggg cctcctcgct cttgcgcgag agcaggtgct tcccaggta ctctgcact	1500
gacatggctg ccacagcggg ggaattttct tgctctcacg acttcgccac gtcgtcctca	1560
cctccctggc ttcacagac tccgccttat cgaggtgctc ggtgggtcgc cttaatccga	1620
agtccgaact aggaagacga gacagagagg tctgaggaat gggcctcgtt tgatgttgag	1680
ctgaattatt tttccagcat aagcccaggc cttggtccat gaacaaaatt actagaaacc	1740
cagcccagta ctacgatcta aaagagggac tgccactgga tagctctctc tagcattctc	1800
cacgctccaa tacagcggcg tagggctctat ccgggtctat ccgcgaacac gtgagaactc	1860
tccagaaact gttttctcct ccacttcctc tctctcgctt tccctctata aaaagacccc	1920
ttctaggaat tgagggagac agcaagcagc gatccgaagc tcaatcaatt cactcaaacc	1980
tcttccccaa atcttcgatt agattctcgt tgacaagaag actataaccg aacctgaccg	2040
taaacc	2046

<210> 8

<211> 2489

<212> ADN

<213> Zea mays

<400> 8

tagtcaccac accatgatca tcttgtttgt tgtagaatac ccacttggtta cctacaacaa	60
tttggtttgg acgtggaact aaatgcaaga ccttattcct cgtgaagttg ttgagcacct	120
cttgcatgtgc caccacccag tctggatctc taagtgcacg ctctatcatg tatggctcaa	180
tagaagacac acaagagtaa tggtcacaaa aatgagcaac ttgagagcga gtgattaacc	240
cctttttgaat gtcaccaagg atggagttga cggggtgata tgcgtgaatt gcttggtgaa	300
ctcttgagtg aggtgggtctt tgatcttgaa tctcttggtc atcctccttt tcttgaacaa	360
cttcatctcc cccttgattg attcgttctt cttgaggtgg ctcatgtct tgaacttgat	420
cttcttcttc ttcttgagct gggtcctcat ctttaagttgg tggagatggt tgtatggaat	480
atgatagttg atcatgtgct tgtgaaggct cttcgggttt ttgtggacac atgtctccaa	540
tggacatggt tcttagcgcg atgcatggag cctctttatc atctaaatca tcaagatcaa	600
cttgctctag ttgagagcct ctagtggaca cacatcttcg ggtaaagtgg ttttatggaa	660
tattttttct tcacagatac aaatagtcac gaatatttca gatatttcat ggttttgtcg	720
aatacaaaata taaaatcggg tagagaaaac taaatttaat tatatccatt tccatccata	780
ttaaaattga atacggatat agatatccat attagcattt tatttgaata cgaatatata	840
taatttggat gtctagacat tcgaatccat ctctaattag tgtgggaatg agggacactg	900
aaaaacaatg acgtgcatgg tgacatcata caatagtaca attctgacga cgatgaagga	960
atttacgtgc ggatcagcga caccacctgg tttggtggtt cctgtcgccg gaggcgcaag	1020
aaataaagag ggcacataac tacgtgaaat caagcccaat tagtctgcct tggctccctt	1080
atgctgtttt aaaaagttaa gggttaagtc attagtagac tgttgcggtt ttagaaattt	1140
ttagagaatt ttgtttacaa cagcccctaa actaaagttt ttgggaacaa atttttagtag	1200
tcttttaagt tgctctaaga ctatattttt ttagttgagg aggacagtga caattttgga	1260
gttgctctaa aaccatgttt tttagtcgag tgggacagtg gcaattactt aactacaatg	1320
cacaacacca ggaatccagg atgaaaaatt actacaccga gggctagttt gggaacctca	1380
ttttcccaag agattttcat tttcccaaag aaaattagtt tatttttcct tgggaaaata	1440
gaaatccttt ggaaaattgg agtttccaaa ctagccttga ttttttttcc taagatatgt	1500
gcagatcttt ctttgagagg acacaaaaaa aatggattgg gattgggctc atcgaaggcc	1560
gaatattcct atccatcggt cgtgccggat taggcccaga accagaaaag agctaggccg	1620
ggctgcagta gcactatcct ctgtttcaca atattatgca cttttgatca ctttatttat	1680
gtcaaaaaata cttgatacat cacgttttat tttcactttc gcctctataa aagtattaag	1740
ggattttctag acaattcaaa atgtaatttt ataagactat gtttgtcatt ttataaaaaa	1800

aaatagtttg attattttgg tgaacgtgcc ttggtcaaaa tttgtgggaa cggaggatgc	1860
tatcaaattc gtctgcgcag atgtacgccc agtaacgaag tatcgtcaat cgactgatga	1920
ccccgtcacc gtcagcaaga cagcaactca acattcaaatt tcgaccgtaa taacatccac	1980
atacatacac ggagtatcaa tctagactag aggagacggg tgaacgtggg gagcctccgc	2040
ccataatgca accctactag tgctagcttt cgccgcgcaa aaaacgtccc cccccccacg	2100
totcaaacttt atagccgccc cctccccacc gccgcggccg ccacgcgcag cagcaacccg	2160
gtagcaggag cgcagccagc aagctcaggc cccagccct actgccaccg cgccgaacga	2220
caaggccgag ccggcggagc accgtccagc tgaggaggag gaggccgcg cgccggcgga	2280
ggatgaggac accggcgccc aggtcgcgcc catcgtgaag ctggaggagg tcgccgttac	2340
cactggagag gaggacgagg acgcgctcct ggacttgtga ggcatccgcg gccgcttcga	2400
ttcccccccc tccccgatcc gatttgccca tgtcttgttg atctgatgtg cggcggctgt	2460
gcaggaaggc gaagctctac cggtaaacc	2489

<210> 9

<211> 2320

<212> ADN

<213> Zea mays

<400> 9

tttgtattgt tgtatgtaat gaataatctt tatatatata tatatatata tatatatata	60
tatatatata tatatatata tatatatgtg tgtgtgtggg ggggtgtggg ggtgtgtgta	120
ttgatgtctc ttggacaaca agattacaca aaacacaatt agaataaaaa aatatcctcg	180
tgtatataaa cttgtccgta taccatatta gaacacataa attttaggca acattttttc	240
catcaacatt cttcaatcat caacaaaat ttacggatac acaagataag agggggtatg	300
taaggttgta cgtaatgggc tacatgataa catcaaagat tatgcaagca aatctcaatc	360
caccaggcga tcataaacat catagttcac atgcttcagt tttaaagcca tcgtagtacc	420
cgtgcaaaga caaaaacatt agaattattc aatacaagtt gcacaagata gttcaataaa	480
atttaaacca caatagtatt atccaacaaa gctagtccat accattatag tttctagtaa	540
acaagaatag agaacatata ttaagcaaac agaccacata ggataaggct aaggatgtat	600
ttttgtctat tgttttctgt gacatcgatc tcgtttgcac gagtaaaact aaaacatgag	660
aaaattccga taaaaaatag gaatctagaa atacgaacgg aaaacactaa accattttta	720
ttcctgtttc taaaatatat catctcgttt ctgtttttct ctgtaaatat gaaaacgac	780
agatcatagt taaaacggaa tacagtaccg gacgaaacaa tttttctctc tcgtttcccc	840
cttatagcat ctccaaaagc tccccagaag tctcccctaa atctattttt ttggaaaaac	900
acaaaaacat gtctccaaca gttcctctaa agcgccccca actttttcat agcccttaaa	960
actccctcat ttgtagctac aaatgagggg ttttttgggc tccccagaaa caaactgctg	1020
ctttaagatg tttggttgag gagccaagta gaatggagtc gtttcatccc tgattctagg	1080

aacggagccg ttctgttctg tgttttggttaa tctggaacgg agcggctctg ttttttgttt 1140
 ggttgacagag tgaacggaac ggagcgtgac tgtgagagcg ggatgagaac ggagcggctc 1200
 cgttcggttg attttttggg gcggaatggt tccggatttg aggagaatat tccataattg 1260
 gagtcattcc gttctagttc ctttataacc aaacagcaac aaaactggga tagaatggtt 1320
 ccgttctact tggctcttca accaaacact aactaaggga cctgttggag aaatgattaa 1380
 aatttaccct cacttattat ttagatatcc cttaaaaacta attttgagaa gtcgttttat 1440
 ggagtgcctc tggagatgct cttagtctgt agatctgatt gtgtgtatgt atgtaagcat 1500
 atgcgtctgt actacgatcc acggtaaaaa agtcacaaac atataataat agcgtcattc 1560
 aataaattga taacgtaaag tataaatgaa cttttaactc tgtttgagta aacaactcgt 1620
 gcagagactg caatgaaatc tattatTTTT tctaaaaataa ttatacaagt tgaggaaatg 1680
 ctatttggtc cactaagcga cgatgtattt tgttttttaa aaaaatcgac gaggtactgc 1740
 tcggttatta ttttcacatg caccgcgcgt tgtttttggg ccggcccatt tgtattgcga 1800
 atttgcgag acgaatatga ccgaatggag tttagaaagc ccagctcact taggattgtc 1860
 tattttctca agaaaagaga gaatggggccc aaaggcctaa acacaaaaac ccgatccgtc 1920
 tatgggatgt catggacatg gagtctggga ccgtccggca gatgagacga cagccgtcgg 1980
 atcagaaacc ctagcgcggg aggtctctcc tattaatacc caccctgcac cccgcgggag 2040
 gagtctttcc taggggtttc tagcttctag ccgccgccgc gtccgcctcg ccaagcgcag 2100
 cagccgccgc agcacatcgc tctctcgatc tcagccatag cggaggtgga gcagcaacag 2160
 gacacgccga agctcttcaa ccgttggaac ttcgatgatg tccaggtacg cgaacgagtc 2220
 ttgcgatctc tgcattgttc gattcttttag ccttgccgtc actagcagtg gatggaaccg 2280
 acgatgaaat ctgcaggtga acgacatctc gctgtaaacc 2320

<210> 10

<211> 2686

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Bao gồm trình tự Kozak được tối ưu hóa ở ngô.

<400> 10

ggatcctgag gtgtggcttg tatgtttcct gacccctgg tgggtgtccc agtttgcaat 60
 acatcatttt ctgtagccct tgtacttgct gtgatccctg tttattgttc agggaaaact 120
 tagtgcagtg tattagaaat atagaaacct cacatttcga agagcagaat agtgtttgat 180
 aacagttttt ttaacgcaac aacgtttgtt agactgtgtc cagcggccag cgggtagtgt 240
 aaaataggtg acgcgaaact atatgtgata ttgtttgaca ctatttgacg agtatagttt 300
 aaaatagggg gtgtgagtaa tctgctgaag atagccttaa gagttaagat accaaggtat 360
 agtttatcat atgcaaaaaga aaagaaaaaa aaggaaatac aaaccattgt ttttatgttg 420

gttagagcta ggtaattatc tcttgaccag tgtattttcca aacctgtttct tctcagtttc 480
 tgtgtccgt cgatactgaa cattgttggt cattttctcat ttgttgaacg atggtatttc 540
 aggagctaga gaagaggaga gaggagcttg agtctctgtt aactgctgac cggatccgct 600
 gcgtttgggg gcggcaggat ctccgaccgg atccgcccgc cgtgcccgcg accggatccg 660
 ctgcgtttgg gggggggcag gatctccgac cggatccgcc cgccgtggcc gcaaccggat 720
 ccgctgcgtt tggggcgcca ggatctccga ccggatccgc ccaccgtggc ggcgaccgga 780
 tccgctgcgt ttggggggcg caaggaagcg ggtggggctt ctagggttca ggtggcgggc 840
 gtcggggtag agagtgcctg cggcgattct ggccggcatg cgcaggggtg ggcaggggtg 900
 tggtcgagcg ggccatgcac aaccttccat ccgcgggatt ttgcgggctg tgctggcgcg 960
 acggatttgc aggagtgtgt ggccgaggc tttgaggcg cgccgattt ggggactgcg 1020
 ggcgttggcg attttgcggg cgttgctggc gcggcagatt ctgcgggctg tgctggcgcg 1080
 gggactgcgg acgcgacagc tgtgctcgcc tgcggccgcg aggaaggcga cggggactgc 1140
 gggcgcgggg agggaaatcg cgggcggagc gcgcgatggc agaacacaca gaactatgga 1200
 cgctacact aagggggtgt ttggtttcta gggactaatg tttagtcctt tcatTTTTatt 1260
 cctTTTTagt atataaattg ctaaatatag aaactaaaat aaagttttag tttctatatt 1320
 tagcaatttt agaacaaaaa tggaataaaa tgtagggact aaacattagt ccctaaaaac 1380
 caaacacccc ttaaatacat aagaagtagt agagattatt attatttgtg tctattgata 1440
 tgatcattat gatattatat tttactgttt taatctaaat atagttatta tattcagata 1500
 attattttat taaatccaaa aatatttaac ttctaagtta ataaggagag aatactctaa 1560
 tatcataata atttcttcta aagtgaccct cggatttggga gatgactgac agggagggct 1620
 gtgcaccctt ctttctttcc ttttcaattg aagaacttgg gttgtgcttg ctacacaaac 1680
 cgaatgaccg atgcacaccg acgcgcacgg agcggaccag cacagcccgg aaactgcctt 1740
 acgccgacac gttttccggt gcggcccagc agcagcaggg aggaagggtt gctcggctcg 1800
 tgggcctggc tccgcgtgcc gtccctccta taaatgcgag tcctcgccga gtccactggg 1860
 caccagaagc tcaactactg ctcgttgccg gctgcccccg cgccccggc cagtccatcc 1920
 cctcgctcgc tccccaaact cagcaggcag atcagataca tccatccatt cgcgcaccgg 1980
 aaggtgagcg ccgtgaacga accatccgcc ctgctagctg cgatctgtag ccttgcgctg 2040
 ctttcgcgcc tagatcgtca cgtcacctat cagcatccgt gcggttctag atctgtggtt 2100
 tttccttccc ctggtggtcg aatccttcca tccaccagac caccacggga cctcgtggat 2160
 tccttttggg tttcctgtgc cgagagccaa aatcgagggg gggggcttgt tttttattgg 2220
 ctcggtctcc cgctgtctcg tgatctgatt tgctgtagta atcagcagga aaggaagggt 2280
 tgaactaaga gcgcctggc ggtttcgctg tcgctgaacc cggacgcgcc gctcttcate 2340
 ccggcgggcg tgctgcaggt ggaggacttc tcgccgcagt ggtgggacct catcaccacc 2400
 actgcctggt tccgcgacca ctggtcccgc gagegcgccc acctggacga gatcgccgag 2460
 cagatcgacg cggccggcct cctccccgac gacgaggacc tcttctacga cgaccaggtc 2520
 gagcagggcc ccgtcgccgc cgcccttaag ataggtactg atgtctctct ctctctctct 2580

ctcttactct cccctcgatt ttagatctgc ctgaaggacg aatcatagtg acctcacgtt 2640
 ggtgcgtttt tctccaccag attcgggtgct caaggcgctg taaacc 2686

<210> 11

<211> 7672

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Bao gồm trình tự Kozak được tối ưu hóa ở ngô.

<400> 11

tatagaatag ctactatcc tatttattat agtttaagta tatagccaat attttaaatt 60
 tactatttat taaattctag ggaagatagt ctcaattcat aactttatta taatacgttt 120
 gaaattttta atcttttagga aattttctta attcacctag atacgattct ggagtgttac 180
 aagctgcgaa tatactgggtg ccattgagta tacataaatg gatttaggtg gtgctcaata 240
 ggtgaaaatg agatactaatt cacttaaatt tcaaaatttc tatgggtcca ctgtactcgg 300
 ataggtctat ctagggtcgg acaaaatgct cgtggctcgc tggctcgcgc gtttcgtggg 360
 cagctcggct cggctcggat cggctcattt gaattttgtc acgagctgag ctgacattct 420
 agctcgggtc gttaacgagc cagctcgcga gctaaacgag ctaccatatt ctagtaaaac 480
 gaaattatat tcatatcatt tatagaataa ttgatgaaca tgttatatat atgtgagatg 540
 tctatggcct atgaattaaa ctaatgatta atgaactatg cctatgtgtt aatttggtct 600
 atgcaaatat aattatgggt taaactgatg aacatgcatg tgaattgtga attaatgagt 660
 gatgaattgt gctaatttgg tgttatattg acatggtttg tgaaactatg agtataatta 720
 ctattttcta ttgttaaatt agtttgaaat taactaaaaa ataattatta tatacatttt 780
 attttttttc tgctctgggt cgcgagctaa acgagccagc tcgacctcgt aaacgagccg 840
 agccgagctg actctgtggc tcgttacctt aacgagccga gccgagctgg ctcgttagct 900
 taacgagcca gctcgaactc ggacgagccg agccgagctg gctcgttatc caccctagg 960
 tctatctagc ttctgatgtt tgcaaacctt agagttggag tgttcagcca gctactcctt 1020
 tgctttgctg aataaccata ccaaacacgc ccatattaat acccgctcgg cgggtggtct 1080
 gcaatcaaac gcaggccgca gtcgcgtgcg gaactagagg tccttcagag aagtgccgtg 1140
 ccagtgccac cgccggccgc atcatcgttc cgccccctg gtacgagcac ttgcgagagc 1200
 tgcaacctac atccctttta cataaatcta ttgtctcgtg ttgccgttga cgccggaata 1260
 gtcttcgcat cccttttaca taaatccgat gttttcttcc tccgattcct ttgaggaata 1320
 atcacgggtc agggcaggtg ttctgccgtt tgcccttttc tttatattct ccttagaaga 1380
 aatatttagt tggaggctgg acatagccgg aggagctaac taatcgagcg gtgtactggc 1440
 aaaacaaaag gagcggagca agaaagggga gaaaaaacta gccactgccg gagcgctatt 1500
 ggccgtgttg ggccgtgaag cttgcatcaa tacttccctc gccccgattt ggttccaaaa 1560

tcatacaagt cccaaagttg tcaagatatt ggaggtatgc aagcgacttg gatctcaaaa	1620
tagaagaaat ttcggatctg agcacaaatc tgagttgaaa aaactgcaac tcaaaatcat	1680
caaaaaaaga agaagaaaaga aacgaatata ttcgctcctc ttctcagccg aacccaaagg	1740
aattgaatcc aaaccctggg taggcagaca gtgagatatg gaggagagca ggaggcgaac	1800
aagagaggct gcggccacga atatctcacg aacaagcaca tcatgggtcc acggagcggg	1860
cagggtgacg ggctcccgac ggcgagctac atctcggaag agcaccaggg cagcatgtcg	1920
tgttgggcag gttggccgtc tggcgacgg cggacggtga ctctgtgtca gggcgcacct	1980
gctcgattaa ggcgccctgac tactcatgtc ttggctctctt tgcttggtgt tgctatatgc	2040
tgctcgtaac tcatgagcat actaagttga ctgctcagtc tgctgagtct gtttttctag	2100
ggtatagtgc tgagcacaag ggatatcatt gttgggatat gattgctcgt tggatgaggg	2160
tctcttggga tgttgtcttt gatgaggctc attcttttta ttcttgcct tctttcgatg	2220
ctttgtcaac atccttgggt gatcccatct cttttctata ttttctagat gcccggttta	2280
ctattggacc tgcctcacgc ttggtgcgc cagcatagta gccttagctc cttctgacat	2340
gttcactctc ctttcgggtgc cttcctttgt ggtgccttct atagtgtttt ctttggagcc	2400
tgctgcttta gccctgact acgctatgaa cacttgccta caccgcccgg gtcacatcaatt	2460
cttttggtag accatcatcc tctcatgcgt tgccctctta tgatgtgcgc tcttctgcaa	2520
ctcattcatt ttcttgcat ttacctttga ctgatgctcc ctattcatct ctggatccag	2580
cttcctcagt tgactctttg ctggagccac ctcttagacg gagtcacgt tttcgtcagc	2640
cacctaattg gtactctcct tcaggtttag tgcctaccgt tctttctgag ctgacttctt	2700
atcatgatgc tattcttcat ctgtaacgac aacatgcgat ttctgaggag attgctactc	2760
ttgagcgcac tagcacgttg gaacttgctt cttgtccatc acgtgtttgt cctatcacca	2820
gtatgtgggt ctataaggtc aagaccggtt ctgatggttc tcttgatcgc tataaatctc	2880
gtctagttgc ccaaggcttc cagtaggaac atggttgtgg ctatgatgag atttttgcac	2940
ctgttgctca tatgaccact gttcgcactc ttcttgctat ggccctctgt cgtgcgtggt	3000
ccatctctca tcttgatgtc aagaatacct ttcttgatgg taagctactt gagttctata	3060
tgtagccatc gcctaggat tctatttctg cttgtatggt ttgttgctct cgccgttccc	3120
cttatggcct caagcaggct ccacattctt ggtttcagct ctttgcttct atgataactg	3180
ttgttggttt ttctaccagt aatcatggtc ctgcactctt tgtgtactac ctctctcgg	3240
ggcggactc ttctttatgt tgatgatata attatcactg gagataacct tgagtatggt	3300
gactttgtta aggcacgtct tagttatcat ttctcatgt ctgatcttgg tctctgtgt	3360
tactttcttg ggacaaagggt ttcttctttg tctcagggcc tttatctatc tcaagaggag	3420
tacattcaag attttcttca tcgggcttct cttaccgatc actagattgt tgagactccc	3480
aagcagctca atcttcacct tagtgccgat gatggcgagt cttttcccga ccatactcgt	3540
tatcgtcaac atactgtagg aagttttgtt tatctctgtg tcaactcgtct tgacatttca	3600
tatgttgtgt gtatcctgag ttagtttgct tcagatccca tccaggta ca ctatagtcac	3660
ttgctttgtg tctacaata tctttgtgga accatatcta gatgtatggt ctttccacat	3720

tctagctcgt	tgcaactgca	atcttggtct	gatgctactt	gggctagtga	ttttttcgat	3780
agttgggtctc	tttctcaata	ttgtgttttt	cttgggtggtt	ctctcattgc	tcggaagact	3840
aagtagcagg	tagcagtttc	tcgtttgagt	accgaggctg	agttgcgtgc	tatggccctt	3900
gtgactgcag	aggttacttg	gttacgatag	ttgcttgagg	attttcatgt	ttctgtttcc	3960
atgaacgactc	cttttgctctg	acagtacagg	tgttatcagt	attgctcgtg	atgcggtgaa	4020
gcatgaggtc	accaagcata	ttggagttga	tgtttcgtat	acacgagctg	aagtctagga	4080
tgatgttatc	ttgatttggg	atgtgccttt	agagcttcag	tlggctaatt	tcttcacgag	4140
ggcacaggct	cgcgctgagc	ataaattttt	cctctcaaaa	ctcagtgtta	tagatccacc	4200
ttgagtttga	gggagtatta	gatagatatg	ggtttatttg	tatttttcca	ttttataagg	4260
gtattagata	gataggcaac	gactgctatg	caagtagtca	ttctgtgcaa	gcgtgcaagc	4320
aaaccatctg	atccattata	tcgtgatcca	accgtgggtc	acatttaaca	cttaaaccct	4380
tccaccacca	actcaataat	ctttataaaa	aaacccttaa	caaacaatgg	ttatatctgt	4440
ggttggatcg	taatctaata	gatcagatgg	tttgcttgta	cgcttgca	gaatgactgc	4500
ttgcatagca	gttggttgct	agatagatat	gggtttattt	gtatttttct	cttaagggtt	4560
tttgtgtata	tttgtactca	tgtacctata	tatttgtgct	agttgacccc	ataatgaata	4620
gacctgctat	tcataataat	tgcaaaccat	gaaaatttga	ttattacgaa	ctatccaaat	4680
actcgaacac	atggggcatta	tagctcacia	aatggaagg	ttgagctgct	gcttgaagaa	4740
cctcaacatc	tttgaacaac	aacctcaacg	aaacttgat	atgaaccaac	ttccaaacaa	4800
tcccttggtg	aaggatagta	atgacttcag	ggcattgatc	acacatatcc	gacggtggaa	4860
ctactgtaac	aaccctcttt	tctgtggaat	atagttgaaa	ctctacaact	tgacaaaaac	4920
caagatgacg	acatatgggtg	gaactaacia	aacaagagga	ctacactacc	tcattagctt	4980
attaagcaca	atctcttggtc	accacaacia	cgaacaacia	aaccatcatt	tggatgctct	5040
gtgggcgact	aatgcaaat	tctttgcatg	gttgatcatc	ccaaattggg	ggcacttagc	5100
tataggctag	cagtgaagg	atggccgaac	aacatgcatt	gtccactatg	ttgggtgtagc	5160
catgagacca	accaccacat	aatgccaaa	cgttcattca	ccaaaaaat	ctaggcaaca	5220
atggcttgga	tttcttacct	gcagctccac	caagctaact	ggagttcaat	taggtcaacg	5280
tatgggtggt	ggtcgagtat	agcagtcaca	aatgatgttc	taaagatggg	gttgtgttaa	5340
cacatcttgc	ttgtagcacg	agaacactgg	aaggagtga	accaaagaat	ctttcaacac	5400
aaggacctat	caacgctatc	catgattggg	aaattcaagg	acgaaactag	aatttgggtg	5460
aacacatgca	caaggcacct	aggagagcct	ttcttttgta	ctgttaatcc	ctttttaaac	5520
tctctctgtc	cttaggagtt	cgtttcttcc	gctctattca	atgaagttag	gcacaatctt	5580
gtgtgatttc	attagaaaaa	cacaagtaaa	ttgcatggtc	agtacttgaa	gtattacagg	5640
aatctcgtct	gcccccaaac	tattaaacct	tatatttggc	tcctaattgt	acttaactga	5700
tctcattctg	gtcaaaactaa	acatggtgat	ggcaaggagc	cgatatggtc	goccatgtgg	5760
atgtgattta	agcaaaaaat	ctcatgggtcc	atagctgtgt	caacaagcca	acatgccatc	5820
gcttccctat	gccgagactg	cccatgtcgc	tcgcttttac	tgtcatcatc	atcaaaactgc	5880

ctgtcatgtc tacggatgcc atgaccgctg tcacacatga tgtggagatg aacctgtcca 5940
 tcaacttcca cgtgctgcc ctatcgctag ctgacaccgt cttggtcatt gctgtgtagg 6000
 gctaggctaa gagtcgctga atgatccttt cgctctcctt tacaggaaca tgcgtgtttac 6060
 tttgtgtcgc caaggcgtgc tagagtacct cttctacacc tccagcacca gtagccttat 6120
 tgttagcttg cacatcccac ataagcaggc cgatgtgaat gataacttca gggacgtcga 6180
 cggcatgtca ctgccaagag tcatttgggtg ggaagcgttg tcatgccatc tgtcgtgcc 6240
 ttttgtcctc agttcgaccg ccattaccgt gagcacaacc tttgcgcatg gttggcgcgt 6300
 tccatcaccc ttattccggt tctcgtgtt ggtcttgccc caaggctatg gttagcagac 6360
 cgtgcatatg gccggcaaaa gactattttg cactgtagat tgcactcttt atatagtga 6420
 gtttaaaata ggagatgaga tgaataaggc tgctggagat agcctaaacc cttgcagctc 6480
 gtgcttgcat cgggggagcc aaaaggcgtc cactccacc atcgccgaag cactgagcac 6540
 tactctggct tgtgtttcag caccacaccg cagagtgcct agggccacca acctctctt 6600
 gcctctgtgc ccagagcacc atcagctctg ctgcctccct ctgttccttg tgcttgctag 6660
 gcaggcaatt ccgagctggg gcccaacttg taacgctgat ttcaccatct tgccactgcc 6720
 gggcaccaag tggacacatt tgacttggcc tagtgggttt tctgcataaa tcacatacat 6780
 gtggatgcc tatcaggctc tttgggtgtg tcgtgtctac tttcgacaag gatgagatca 6840
 cttaaacata ttagggagcc aagtatgtaa tttcatagtt tagggacct caaaaaatc 6900
 gtataatact ttagaacagc cgtgcagttt actcaatcaa cacatacaaa gtcagatctt 6960
 aagctctgat acttcaaagg aatggttgag ccagttgac aaacaatctt gcttcattca 7020
 ttgaattgtt tataggagtg gctatgtaac tactgggtgg ttttgtttga cctgtcatcc 7080
 aaattgtgta gtcaaccata aacatacacg tcacacaata ctttttggat gtgacagata 7140
 ggatttaggc gagagaatgt acaatgtcac tgaaaaatta ccactgtatg gaaaggacaa 7200
 tctaagtga aagagaacca gggcctaatt gtttcaggac ttcaaactcc ggccaaatga 7260
 atttacagt cttaaattaa ctcatgttaa tcatgatagc caaagcatgg gcaaaagaga 7320
 aactatgaat aaatcgacaa tgtattctat atagcagtaa tataccatgt cagcagcttt 7380
 tacactaatg ggctgtatct tctgcagtt attttaactg gcaatattct atgtcacagt 7440
 aatatttgtt aaattttttc cagaatagca actgaactag aagtctagta tttcttaatt 7500
 ggataacaaa aggaattagt gtgcatttgg cttacgaaca atcagtcacc caacattgaa 7560
 tttgaagttc tgtttcctct ttgttcagac gacactctcc aaatgaatgc cttatatattt 7620
 gtgttgctcc tcttttctgc agagtgttca gtaacttctt ccgatgtaaa cc 7672

<210> 12

<211> 1014

<212> ADN

<213> Zea mays

<400> 12

aagttcgtca tgttctgttt tggtcatttg ggcaccaaag tttgcgcctc atttggttct 60
gtaatccgtg aggtcgtgca tgtacttggc gtattgcatg cagtgaataa tttagcttgg 120
gtttgtttgt tgggggcagt gttggggacg gatttggatt ggggtttatg cttggcatcg 180
cgtcgtatcg aaactcagct gctgtttcgc tgagtaatgt acatttcctt ggtaatggta 240
cttgtggact ctgatgcttt tatgggaacg agtgcatttt actgccgtgc ttggccgtgc 300
tgttgaatta ttcccgtagt gtattaaata ttggatgcac cagattgttt gtcccattta 360
cgcataattt tcccgtcata tgttgaatta ttacgaatt ataaaatatg taatttatat 420
ttatttttat gaagtatata aaattttact tttttgttat ataattaaat attttttgtt 480
ttgataaaat ggtatgtttt attttaaata aaaatctata aaaaagatg atatggtggc 540
tggaaggtaa cgataaatta tgtatactaa agtattaacg atagtaaata tagttggacg 600
atagatgatg tgacgggggtt tgttcttttg attttaactt ataatgcttg tctatagatt 660
acagccgcat caattttaa atctgctttt aatctgaagc gaacaaccga taaaaataga 720
cattaggcca tctccaacag ttccgaccgtg caaagtatta ttttgctatt gcactatttg 780
aataatggat tttagaatag ggagtgataa aaagtgatg aaaaagtta ataaaaatag 840
tcttagatcc aacaaatata cgtgccctc ccatcagttg cctggaccca tgccaagaag 900
tgccagccat agtcgtatga aaaatgattc taaggcggtt cgcaatgggt gtcttaaata 960
gttagctatt gtgaagctga aactaagaga agaataaaaa tataatttat ttca 1014

<210> 13

<211> 1216

<212> ADN

<213> Zea mays

<400> 13

tacggggaaa gaaggagaag aagaagaaga agcccaggcc ggagaaccat cgcctgcatt 60
tcgatctgtt tcaccgcaat tcgcattgtt agtcgtgtat tggagttatg tgtacttgg 120
ttccaagaac tttggttcct tctcgttttt ttttccttgc ttgtttgagc gtttttgggc 180
agcgtgccc tggttcctag tatggtggga attggctgca ccttttgctt cgaataaaaa 240
tgctgctcg ttacctgtc ttccagagt caatgcgatg ttctgttgcc caggctgtgt 300
ggttctgact gatggcgatg ttgtgttctt ctgttaatcg cctgttttaa cgtggtaggc 360
tgatgcttgt tcttgttgag aaagcttgct gtgccagaca tggctgcttg aatacaagt 420
aaggaaaaaa aaagccatgc caagtaaagt tgcacaaaat ttcaactgct cagtggacca 480
ctggaccatg ttcttgggta ttgcagttgc agggcttcac atggcgtttg gacagcagtc 540
ttggattgat gcataaagag gtggtgggta atgaggacgc aaggccgttc cctcagagtc 600
agtcacaagg ttgcagaggt cacggttctc ttccctttcc gcttctgtc acatcggaat 660
tgttgtttac gccatctgcc catcaccac caagtctatg tttctgtact ggatctttca 720
atggcggaac gcgcttagtt cttcgtcaca gtccaatcac atgatcta atcgatgtctt 780

aatctcgcctg taaaaagggt gggacggtgg gtgcagggta gggaccaggg aaggcctgcc	840
taaacgtatc cataaacatg cacagcaacc ctaagatatt atactgccta ctccctaaga	900
tatagttatt tctagtctat tttttttccg tccacatcca aataagtgat aatatataga	960
catacatata tatactatat tcatcataga ttaatgaacg aatgtatact tagttttaa	1020
ctaattatat tttaggaagg atggagtatg aaacatgaca atacaacaaa aaaaaatcat	1080
gtaattgcat atcgtcaaag ttatctgaag taaccaatcc agggggaaat cccgttagca	1140
aacatacaag agcaccgccc cactacatcc cagaaaataa aacaaaacca gaactcagat	1200
ggataaataa tactac	1216

<210> 14

<211> 1216

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Bao gồm hai đột biến để loại bỏ ORFS không mong muốn.

<400> 14

tacggggaaa gaaggagaag aagaagaaga agcccaggcc ggagaaccat cgctgcatt	60
tcgatctgtt tcaccgcaat tcgcattgtt agtcgtgtat tggagttagt tgtacttgg	120
ttccaagaac tttggttcc tctcgttttt ttttccttgc ttgtttgagc gtttttgggc	180
agcgtctggc tggttcctag tatggtggga attggctgca ccttttgctt cgaataaaaa	240
tgctgtctcg ttcacctgtc ttccagagtg caatgcgatg ttctgttgcc caggtcgtgt	300
ggttctgact gatggcgatg ttgtgttctt ctgttaatcg cctgttttaa cgtggtaggc	360
tgttgcttgt tcttggtgag aaagcttgct gtgccagaca tggctgcttg aatacaagt	420
aaggaaaaaa aaagccatgc caagtaaagt tgcacaaaat ttcaactgct cagtggacca	480
ctggaccatg ttcttggtta ttgcagttgc agggcttcac atggcgtttg gacagcagtc	540
ttggattgat ccataaagag gtggtgggta atgaggacgc aaggccgttc cctcagagtc	600
agtcacaagg ttgcagaggt cacggttctc ttccctttcc gcttcctgtc acatcggaat	660
tgttggttac gccatctgac catcaccac caagtctatg tttctgtact ggatctttca	720
atggcggaac gcgcttagtt cttcgtcaca gtccaatcac atgatcta atcgatgtctt	780
aatctcgcctg taaaaagggt gggacggtgg gtgcagggta gggaccaggg aaggcctgcc	840
taaacgtatc cataaacatg cacagcaacc ctaagatatt atactgccta ctccctaaga	900
tatagttatt tctagtctat tttttttccg tccacatcca aataagtgat aatatataga	960
catacatata tatactatat tcatcataga ttaatgaacg aatgtatact tagttttaa	1020
ctaattatat tttaggaagg atggagtatg aaacatgaca atacaacaaa aaaaaatcat	1080
gtaattgcat atcgtcaaag ttatctgaag taaccaatcc agggggaaat cccgttagca	1140
aacatacaag agcaccgccc cactacatcc cagaaaataa aacaaaacca gaactcagat	1200

ggataaataa tactac

1216

<210> 15

<211> 1004

<212> ADN

<213> Zea mays

<400> 15

agagagatat ttctgcctcc ctatcgtggg tcgtcccca tggcctttgg tcgcagacca 60
 tctttgctgc ttgtctatgc tgagaataaa tgtgaacggg gccctggac gctggatcca 120
 tgctggtttt ggacacgggt gtctttttgt gtttaactta tctgctaccg tcctgtaacg 180
 aattcactaa gttctagttc ttttgtgctt tgtaagaat ataaacaatg aggtttcaat 240
 tttgggtggc ccatccgggc tgattgcgta ctagtatcat gcatactga gctggctttt 300
 ccggagtaac agtttttttt tgtttcgtgt caatgattct cagcctgttc atttggcttt 360
 aatccatata gggtactact ataatgttgt cttgtctcta tagattgcag cagtctaaca 420
 catgttcttg gtatcctttt ggggtgtgtt aatgctatct gctctgtgtt gtctttttcg 480
 ttaatcggga cagagttttc tctctgtttt gacagtttcg gtggttcttt tttgtgtgtg 540
 gtcagcaggt cacgaactgc tgggtggcca gcgaaatagc gcagactatt atgggtcttc 600
 tgctggtgaa tatggaagga cgagggtaaa ttttctgctt tgatgtgtct gccctctgcc 660
 gctgtgctgc atgcatgtgt ttgggagcga agcagcgtgc tgtttgggcg tatggttggc 720
 atggtgttaa attatttcgc tgattcagct ttgggttata actggttttg atggtcacgc 780
 tgtgtcgtaa gagagatcgt ggcagcaaaa tggagaaaaa ggcagtgaaa caaactttca 840
 ggggtgtggaa tgtatgggat tcgtatggtt gaggccagga tttgtcaatt ggtcaagttg 900
 aaaagtaaag gcgagacaga ttggcgagga ccgaggagag attgggggttc agaaacttcg 960
 gttacacttt aatccctcc atctccctcc actcctccct tcta 1004

<210> 16

<211> 1023

<212> ADN

<213> Zea mays

<400> 16

acctcggctc gcggtcgtcg tacctgcgtg gtttgaggaa cggcagttcg cctcggctcg 60
 tctgtgaaat aaaattgggt tacaagaatt atggcgtttg tcaatatggg cgtaatgtcg 120
 taggatgggt gaatgtgggc acaaactttg cgtatgttgg gtctactggg ggtgtctgaa 180
 tctatgtatg gatgtcatga gtttgtctac tagagttcta gttcctgtgg tgttcgtatg 240
 atgtatatca tgagtttaact ctatctaaaa tttcttcatt atggtatctt taagaaataa 300
 attctttatt atgggttttt taaagggact atgtccatcc taacaagagc aacgtccaat 360

aattctctaa aacttaaatt taagaaccga ataaaaaatg agtagttttt aaatactttc 420
tatctaacct tattttactct tctccatatt ttagtagttt tttaaataga cttactaaat 480
ttagttactc tatattttgg taactcgaca gaatacaatc tgtgattaat ttttctcaca 540
tgtggtagct agatacgaca cttttttttt tacttttttag cacatgcaca atggagtagt 600
tagattcaac acttatgaca caagtttttt cttcacggtt ggactattat tgtcgcatat 660
gttgccaccg cgaaccaccg ctccctctata ggttatgtca atcgaccttt cctcttttgt 720
cttcgacatc atccaagact aacaaagtta gattcaaadc aaactagatg tccagactgg 780
caaagttaga tttaaatca aactactatc tacatataag caactattag ggactaagtt 840
atttttctac ttcaatagtt gttgctagca acttgctaaa cataatttta gaaaactttt 900
ttagagaact attagagttg ctctaaaagg tttttagtgc ccatttactg ttttgtgact 960
attattaggt ggagtttggg cagtaaaaag gccaaataga aaatgggggt gagattggga 1020
aaa 1023

<210> 17

<211> 1028

<212> ADN

<213> Zea mays

<400> 17

aggttgaagg caaacaaggg tcaaattgat gccattccat tcatttcggt tccaagggtc 60
agcttccccc caaattttca ttgtgttttc tccgagatga atgtttgtgt tcgggtgaaat 120
cagagtcgtc agtcatctac atagcttttc ttggttgata gactgttatt ttaagtcgca 180
tgtttatctg ggatggctgg ggtcagcatg tttgtacaat tatttggagt tgcttttgga 240
atggtcgcgg tttgatgagt tgccacagc catgagatgt cttttcgctc cacttttatg 300
gttcattcgt tctcaataat atgggatgct atacttgtgt tgatccatta tcttgatgca 360
tcgttgtctg tgcactgcaa caacaatacc catctgaaca cccctatcaa taaaatacca 420
ttttttttct ttccatccca ctaatctagg ccactttct cactcttttc ttatccacta 480
tattgtcaat atagaatctg gggagagaga gagagagaga gagattgaga gagagagaga 540
gagagagatt taggctccct ccttccctat attcaaaaata ggtatcgcc tggggtcacc 600
tgttggaatg atttatttta gtatcgcca tatttaattt aaacaaatgt tgaattttac 660
atgcatattg atttaaagtt tgtagtttta tgagcatcac caactaagat ctctaagcc 720
aaagagcatc tccactagtt ccaaaaaact ctctaaattt aatttaggat gttagtaacc 780
agaaattatg ctccaacagt ttccctaaatg agttccctaa atataccaac tttttaaata 840
tctctattta gtcaaacttg agaaattgtt tcacactccc aatagttgtc atcaatcaca 900
tcgtaaaaatc atttgagttc gcacccgtgt acaagtgggt cacttttaaat aaattaaatt 960
acgaaaatga taaatttact atcgagttag cataatttaa aaatatataa caacataaga 1020
actgggaa 1028

<210> 18

<211> 1002

<212> ADN

<213> Zea mays

<400> 18

```

actgagcttg tatectggtg cactctgcgc tggaaacttt tatgtcgtg gcagtcgtat      60
cggttcttgt tttaccaatg tttagagttt tttgagacct atatgcggtt ttggttttca    120
gtgcacaatt aaaattactg agtaatgtag ttgattggga acagaaatgt ttgggtgctg      180
gtttaccgaa ctccagttct cttgtcattt ttctttattc tatagtctgt attatgtatg      240
cgtatgagta ttgagatgat tctgcatttg aattgtctgc tttgttgctg tcgctgtatg      300
cgtaccaatg gtaacagggg agttgtggga agtagacacg gccggttcta tgttttcgtg      360
cttccgcggt tcaggctggt taagcctatg gagcgtacgc acgctcctcc cgtctctccg      420
tgtcctgcat gctggcaaca ggagtgcggc ccagcgcacg ccctaatacga cggcgggtata      480
tttgtctgtc cctccatttt gtggtgaggg tattcgcaac cgttatcctt aaattttttc      540
tcctatatca ctattcccct atttttcctt atattttttc atcttcagca gcggttctcc      600
taaatactcc ctctataccc actacaacta taaatattat tttccatata tattcatcat      660
ttattaccac tttttttcaa ctaaaaaata ctcgcatgca tggattttac ggaagggggg      720
ctgtcacagt atccccctga tctgctgtga gagaaaaggg ggacactagg taggggtgcaa      780
ggtaggggag agcgggtgcgg gtggttagcgt ggttactgca gccgctacga cgtgagcagt      840
gttaggggag aggatggaag ggtggtgcgc tgctgcagat aacccgagcg ccaaactctc      900
atgggtgata attaggtata agaaaagata ttttatgggt aggagagtat agagagaatt      960
tagtggtaac ttctatggaa gatggaaaaa taggggtgaa at                          1002

```

<210> 19

<211> 1001

<212> ADN

<213> Zea mays

<400> 19

```

gttcgctggg ggaactcatc aggaaggctg ctgcccctct tgcagccttg ctctggctg      60
ccgccgctgt cgtggtctgc tctttcaagt cgaagtaacg gtggttcgag ctagtggata      120
gtgtggctca actgtagaag ttccttttgt atagcaagca agtaaaaaaa aaaaatgacc      180
aaaaaatata acaaaatgca gctgtaagtt tactgtctgt ctctaagtcg tgttcagtca      240
tccagtgtgt ctagtctagg gaaaccccat aaaaatggtg aagggtggaat cccatcccag      300
tgtcataatt aaggatgcac ttcttctgta agcaaagtga tgatgtacaa tggccggccg      360
gcagtctaaa tgttacaact agctcttctt ggtgaattca ccggtccaca ctgatgtgct      420

```

gctatgtatc attactatcc agttagggct tgttcggtta ttccctacgcc atatggattg 480
 gacgggattg gaaaatttta gtagacattt tgactttctat ggattttaaac ccaccaatc 540
 ccctccaatc cacatggatt gagatgaaac cgaacaagcc cttagttgga tggatggatg 600
 atctcacgcg cttgagttta tctagttgct gatatgggag agccctcaa cacctcaaata 660
 attgtggata agtttaatcc tacactgtca gtcttcagtt ataaggcacg cactcttcga 720
 cgttggggcg tgtcttttgt ttcaaagggt gaggcaacct caacctcgtt taaaacagag 780
 agcaaagcta atttccaaac tgatgtaagt catgtgtcct taattaccaa agtagcaatg 840
 atgacattgt tatgttgtgt gctaattgagc cattacgtga acatcagttc ctgccttgcc 900
 gcttcgttca gccggggcgg ggtatatattt gagtactaa aaggacacgt cggatgaatg 960
 gaaaattgcc tcttgtcttt accaagagtg tcgcccgcga g 1001

<210> 20

<211> 2037

<212> ADN

<213> Zea mays

<400> 20

cgacgtacaa atctcatctg tgccttgctc tagtttccca aatggaatta actatgcatg 60
 atttgitttg aaactcttat tgcattccatc cagataatgc atccaccata aggtaatatc 120
 ttgatgacat ctgtgcctga tgggtgtacca aatgtctcta tctctgcatt gagccacgag 180
 taggaggata gcctaggggt gccttgactc caaagttgta ttgaaaaaga tggatgaagc 240
 aggcaaatgc tgcctgaatc catgactcag ggcacagatt ttccactcaa aggaagataa 300
 gattgcatta cttcatgac ttttgaactg cctctgcaag acgggactcg gatagtggat 360
 gcaaagatct aatactggcc tcaggcaacg agttgtttca ctcgaaagtc tagaaatgac 420
 cgggctcaaa ttttgcaccc caaggaaagt gagtttgcat tacttcatga ctttttgaac 480
 tgcctctgca agactggact cagattacgc ttgattgggt gccggcctca ctttcgctg 540
 gcttgcgca gccctgcgtct atagaaatgc gccggactca cgtctccgct gatgcaggca 600
 ttcgactgaa aaaacattta aactgcaccc atgcgtgcgg gctgagctta tgcatacaa 660
 gtaaccaatc acaggcttaa gttcagtcaa cgcattgcgt aagcttggat gtggctgacc 720
 gggcaaccaa tcacacagat agtggatgca cggatctaata attggctaata ttgggttaaac 780
 ttgtctaacc ttagacgtgg caagtgagtc agcggatcaa atctgctcta aaattgtctg 840
 cctcctagat gtcccttggtg ttccaagatt taatcatcac tgcactatct ctttgcgttg 900
 cttcgtgca gcttcgctt acttgcatc gcttaatcag gattactttg atcaactagg 960
 tttctaactt ctactacctt cacttgcaca ggggtgccgt cctgctagcc ggtgtgcttg 1020
 ctgtgcgac gtttggcatg tgcttgttga ggggttgcta ggggattgga gaggattgaa 1080
 gggattaaat ctctctctat tcaattttga ataggagggg atttaataccc cttcaatccc 1140
 cctcaaacca ctagtaaccg aacgtggcct gagggggcgg gcgagtcctt atattgaatg 1200

aaactacata aaatagcatg ccgtctctgt cactggcaat ggacgggtggg gcctagcgca 1260
 actcagcgca caactgtgtg tcttgatttt tcttctgttt atcacggcat tagtgccatg 1320
 ccgttttatg ttacagtgtt gtgtgctcgc aagcatccga aaatatgcgt ctgagtttag 1380
 ggttgggtca aacttgctga atttgggggt ctgttataat atgttgagca tgaataaaga 1440
 tggatgctgg tgactctgtc gccatcgccg tccatcatga gtgtcctgta attcaactta 1500
 tatctatcat gtatgtatgt atgtatgtat gtatgtatgt atatgctgtc tactatgctt 1560
 ctttgtttta actgaaatgt gtgttacagt gttacttctc tgggggtccat ttaaaacggc 1620
 atttcgttta cgataggaac cagccattat aatctttaac caataatttc gctaaccaat 1680
 ttcaactatt gcaatgcgaa cttaatatta tcagatttat aaccgaatgc gctatcaaat 1740
 aatcataagg ttgtaatcat aataatataa tataaaataa atgagtgtc gaagtgaat 1800
 tttagagagc gttataagaa aaattgatgt gatctccaag aataatagcc cctcccggt 1860
 cccggtacaa acatagggtt tctttagaat gcaggattgt gagaacatag gaataggaaa 1920
 aatataggaa ttctatagga atgtatatgg aaaacagagg attgaaaaac acagaaaaaa 1980
 tgtgaaagca agtctttgga tgaagcgtag gaaacttata ggaataggaa ttcataa 2037

<210> 21

<211> 16119

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vectơ tái tổ hợp.

<400> 21

attcctgtgg ttggcatgca catacaaag gacgaacgga taaacctttt cacgcccttt 60
 taaatatccg attattctaa taaacgctct tttctcttag gtttaccgc caatatatcc 120
 tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga 180
 attaaggag tgacgttatg acccccgccg atgacgcggg acaagccgtt ttacgtttgg 240
 aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccatttaaa tcaattgggc 300
 gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc 360
 taccaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg 420
 gaacaaaata acgtggaaaa gagctgtcct gacagcccac tactaatgc gtatgacgaa 480
 cgcagtgcg accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct 540
 cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaagggtg 600
 ctctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga 660
 cagtggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc 720
 aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga 780
 acaatcccac tatccttcgg taccggaccc gagatatgtg atgtatatgt gatatatgtg 840

atgattatgt gatataat	ttgtttgttt ggatggaata ataaaaacaa ataaaaagg	900	
tatgttggcc acttttggcg	agtgtaacac tcagcaaaga ggtactttgc cgagtgtcac	960	
agtcatatca ttcggcaaag	aaggcacaca tgggaaccga taaagcttct ttgccgagt	1020	
ttgtggcctt agcactcggc	aaagaagcaa cttttgccga gagcctccta gtgtactcgg	1080	
caaaggaact gacaaaggag	cccactgatg attcctttga tgagtggtag tccggcgagac	1140	
acacggcaaa gataaagtct	ttgccgagt	ccacctaata cggccggcaa aggaactggc	1200
aaaggacaca cgatgagttt	atgtgtcgag tgctagtaca atagacactc gacaaagagt	1260	
gagcctttgc agagtgtcac	cgtggcactc ggcaaagtcc actctttgtc gagtgtcacc	1320	
gtgacactca gcgaagtctc	cgtcgtgtc atctgtcgcc gtgacgtcga cttttttttg	1380	
ccgagtaccg agtggttgcc	gagtgtccga caaaaaatac tcggttaaag accgttgccg	1440	
atgtcagttc accaagacct	tttttatcaa gtgtcatact tgataaagtc ttccccaagt	1500	
attttcaaaa ctttgtcgat	tacctcacac accaaggggt aggatcgtgt tgaggcgggt	1560	
tgctgcttgt ttcttgcttt	cggcgaacca acggaccatg agcgcataaa ccaatgggcg	1620	
agcggccgca tatacaggcc	tgcgcggcga gcctcacctt gctggatttg aatgccccct	1680	
cggcatggtc aatgaccatg	catctttgtg cagaacatat aaaatgcaca aattaattaa	1740	
ttaataaaaa tacatgaaag	agtacataat ttgtttgtgg attattaatt atatttataa	1800	
atatattata gtatattata	ttttatcttt atctaactac aaaaataatt atagattgtt	1860	
aacttttttt tgcatctactg	gtaagatttt gtcgatttta taattttgtc taacatgttt	1920	
tattatcaat ctaaaattgc	aaattatatg ttaacgaact aaatttgtaa taaaatatat	1980	
gtaataaagg tcgtctaaaa	gtaagctaaa attcacgcac ctaaatttta ttagtaaaag	2040	
tgttttttct tgacatatga	cactagtgtt ttcaaagctt gtctcttaat ttaggattaa	2100	
attaataggc ctctctaaat	tgctttttat aagataaata aatcctatlc cgtctgtctca	2160	
tcttcaaaga acattttcaat	gaaacaaata tttgggaaaa ccaatggaaa aagaataaaa	2220	
ataaaaaattg aaaaagagaa	gtggaggaaa agtagggttc cattccacgg cctgtcgagt	2280	
gtcgaccac aaaccacgca	accgatatat tccccagcac cccagtcctc cagccgtcca	2340	
accgagacac cgcgtgcgaa	ccaagcagac cacaacaaga agaagcgtag tcgtcgccgg	2400	
aaggaaaggc gcggagcaag	atctcgtggc aggcgtacgt cgacgagcac ctgaagtgcg	2460	
agatcgaagg acagcatctc	agcgccgccc ccatcgtcgg tcacgacggc agcgtttggg	2520	
cgcagtccga gagcttcccc	gaggtcattc actcccacct tatctcttcc cgctctactt	2580	
gcttagttgc tttatccatg	ttcagtgtct ctgcgctgat ttagaccgcg gcgactctgg	2640	
atctgccccca gccgtgcggc	ccggccggat cggagcggag ggatgtccgg atctcggtcg	2700	
ctgaggcgag atgcggatct	aaggccccct cctcgggtga cgtggttgct gctgctgctg	2760	
ttctccgggc gtttgatgcc	aatatagctg agatcaagct tgatgatctg cgtactgtgg	2820	
atgtgctagt gagggatggc	cggatcgggc tgttggcgct cgacctgatt tggttgttcg	2880	
cttgaataga cttacttacc	agtggacagt ggtagtgatc gttcagcagc agtagagcaa	2940	
tttgttttac atgtaaaatt	tgagtggctg ctagagatgc accaatggcc gagctgcatg	3000	

tgtttattcc	tatatagcga	cagttgtgtg	attgatccaa	aacgcaaaat	tttagttttc	3060
atcaaagtat	gatttcctga	tgccactacc	tgtgtgtgag	gcacttatta	aatgaaaagt	3120
aactttcggg	tgggatctgg	gaatctgtat	tgtgttccgt	ggttggtcac	agcttgtggc	3180
ctgcattgtt	ccataggcca	aatggctagg	caaggaaaat	aaatcatgta	atgttgaaaa	3240
aaataactgc	catagtcagt	agtgtagcgc	agtacatgct	tgtccggggc	ttttatttca	3300
ccatcgtgtt	gtggtatatg	gagtacttgg	attgtgcgat	gctttataca	tttgtttgtt	3360
ctagcagagt	ttgagtttat	atttctattc	ttatgtaggg	cgtcacgagt	gtcgtcaatt	3420
gttacatcct	tacatccttg	tggtcacaat	atgatccatc	aactctttct	tctgcggtct	3480
cttatccctg	tggttctcatg	tccatgtttc	ctgcttcttg	ttgcagttaa	agccttagga	3540
ggtaaaccat	ggacaacaac	ccaacatca	acgagtgcac	cccctacaac	tgccctgagca	3600
accccgaggt	ggaggtgctg	ggcggcgagc	gcacgcagac	cggctacacc	cccctgcaca	3660
tcagccctgag	cctgaccag	ttcctgctga	gcgagttcgt	gcccggcgcc	ggcttcgtgc	3720
tgggcctggg	ggacatcatc	tggggcatct	tcggccccag	ccagtgggac	gccttcctgg	3780
tgcagatcga	gcagctgac	aaccagcgca	tcgaggagtt	cggccgcaac	caggccatca	3840
gccgcctgga	gggcctgagc	aacctgtacc	aaatctacgc	cgagagcttc	cgcgagtggg	3900
aggccgaccc	caccaacccc	gccctgcgcg	aggagatgcg	catccagttc	aacgacatga	3960
acagcgccct	gaccaccgcc	atccccctgt	tcgcctgtga	gaactaccag	gtgcccctgc	4020
tgagcgtgta	cgtgcaggcc	gccaacctgc	acctgagcgt	gctgcgcgac	gtcagcgtgt	4080
tcggccagcg	ctggggcttc	gacgccgcca	ccatcaacag	ccgctacaac	gacctgaccc	4140
gcctgatcgg	caactacacc	gaccacgcgc	tgcgctggtg	caacaccggc	ctggagcgcg	4200
tgtgggggtc	agacagccgc	gactggatca	ggtacaacca	gttccgcccgc	gagctgaccc	4260
tgaccgtgct	ggacatcgtg	agcctgttcc	ccaactacga	cagccgcacc	taccccatcc	4320
gcaccgtgag	ccagctgacc	cgcgagatct	acaccaaccc	cgtgctggag	aacttcgacg	4380
gcagcttcgc	cggcagcgcc	cagggcatcg	agggcagcat	ccgcagcccc	cacctgatgg	4440
acatcctgaa	cagcatcacc	atctacaccg	acgcccaccg	cggcgagtac	tactggagcg	4500
gccaccagat	catggccagc	cccgtcggct	tcagcggccc	cgagttcacc	ttccccctgt	4560
acggcactat	gggcaacgct	gcacctcagc	agcgcatcgt	ggcacagctg	ggccagggag	4620
tgtaccgcac	cctgagcagc	acctgttacc	gtcgaccttt	caacatcggc	atcaacaacc	4680
agcagctgag	cgtgctggac	ggcaccgagt	tcgcctacgg	caccagcagc	aacctgccca	4740
gcgccgtgta	cgcgaagagc	ggcaccgtgg	acagcctgga	cgagatcccc	cctcagaaca	4800
acaacgtgcc	acctcgacag	ggcttcagcc	accgtctgag	ccacgtgagc	atgttccgca	4860
gtggcttcag	caacagcagc	gtgagcatca	tccgtgcacc	tatgttcagc	tggattcacc	4920
gcagtgccga	gttcaacaac	atcatcccca	gcagccagat	caccagatc	cccctgacca	4980
agagcaccaa	cctgggcagc	ggcaccagcg	tggtgaaggg	ccccggcttc	accggcgggc	5040
acatcctgcg	ccgcaccagc	cccggccaga	tcagcaccct	gcgcgtgaac	atcaccgccc	5100
ccctgagcca	gcgctaccgc	gtccgcaccc	gctacgccag	caccaccaac	ctgcagttcc	5160

acaccagcat	cgacggccgc	cccatcaacc	agggcaactt	cagcgccacc	atgagcagcg	5220
gcagcaacct	gcagagcggc	agcttccgca	ccgtgggctt	caccaccccc	ttcaacttca	5280
gcaacggcag	cagcgtgttc	accctgagcg	cccacgtgtt	caacagcggc	aacgaggtgt	5340
acatcgaccg	catcgagttc	gtgcccgcg	aggtgacctt	cgaggccgag	tacgacctgg	5400
agaggggtca	gaaggccgtg	aacgagctgt	tcaccagcag	caaccagatc	ggcctgaaga	5460
ccgacgtgac	cgactaccac	atcgatcagg	tgtaggagct	caagttcgtc	atgttctgtt	5520
ttggtcattt	gggcaccaa	gtttggcct	catttggttc	tgtaatccgt	gaggtcgtgc	5580
atgtacttgg	cgtattgcat	gcagtgaata	atttagcttg	ggtttggttg	ttgggggcag	5640
tgttggggac	ggatttggat	tggggtttat	gcttggcatc	gcgtcgtatc	gaaactcagc	5700
tgctgtttcg	ctgagtaatg	tacatttccc	tggtaatgg	acttggtggac	tctgatgctt	5760
ttatgggaac	gagtgcattt	tactgccgtg	cttggccgtg	ctgttgaatt	attcccgtag	5820
tgtattaaat	attggatgca	ccagattgtt	tgtcccattt	acgcataatt	ttcccgtcat	5880
atgttgaatt	atttacgaat	tataaaatat	gtaatttata	tttattttta	tgaagtatat	5940
aaaattttac	ttttttgtta	tataattaaa	tattttttgt	tttgataaaa	tggtatgttt	6000
tattttaaat	aaaaatctat	aaaaaaagat	gatatgggtg	ctggaaggta	acgataaatt	6060
atgtatacta	aagtattaac	gatagtaa	atagttggac	gatagatgat	gtgacggggt	6120
ttgttctttt	gatttttaac	tataatgctt	gtctatagat	tacagccgca	tcaatttaaa	6180
tagctgcttt	taatctgaag	cgaacaaccg	ataaaaatag	acattaggcc	atctccaaca	6240
gttcgaccgt	gcaaagtatt	attttgctat	tgactattt	gaataatgga	ttttagaata	6300
gggagtgata	aaaagtgaga	taaaaagttt	aataaaaaata	gtcttagatc	caacaaatat	6360
acgctgccct	cccatcagtt	gcctggaccc	atgccaagaa	gtgccagcca	tagtcgtatg	6420
aaaaatgatt	ctaagggcgt	tcgcaatgg	tgtcttaaat	agttagctat	tgtgaagctg	6480
aaactaagag	aagaataaaa	atataattta	tttcacggac	cgcgatcgct	taattaagct	6540
tgcatgcctg	cagtgcagcg	tgaccgggtc	gtgcccctct	ctagagataa	tgagcattgc	6600
atgtctaagt	tataaaaaat	taccacatat	tttttttgtc	acacttgttt	gaagtgcagt	6660
ttatctatct	ttatacatat	atttaaactt	tactctacga	ataatataat	ctatagtact	6720
acaataatat	cagtgtttta	gagaatcata	taaatgaaca	gttagacatg	gtctaaagga	6780
caattgagta	ttttgacaac	aggactctac	agttttatct	ttttagtgtg	catgtgttct	6840
cctttttttt	tgcaaatagc	ttcacctata	taatacttca	tccattttat	tagtacatcc	6900
atttaggggt	taggggtta	ggtttttata	gactaatttt	tttagtacat	ctattttatt	6960
ctatttttagc	ctctaaatta	agaaaactaa	aactctat	tagttttttt	atttaataat	7020
ttagatataa	aatagaataa	aataaagtga	ctaaaaatta	aacaaatacc	ctttaagaaa	7080
ttaaaaaaac	taaggaaaca	tttttcttgt	ttcgagtaga	taatgccagc	ctgttaaacg	7140
ccgtcgacga	gtctaacgga	caccaaccag	cgaaccagca	gcgtcgcgtc	gggccaagcg	7200
aagcagacgg	cacggcatct	ctgtcgtctg	ctctggaccc	ctctcgagag	ttccgctcca	7260
ccgttggaact	tgctccgctg	tcggcatcca	gaaattgcgt	ggcggagcgg	cagacgtgag	7320

ccggcacggc aggcggcctc ctccctctct caccggcaccg gcagctacgg gggattccctt 7380
 tcccaccgct ccttcgcttt ccttctctcg cccgccgtaa taaatagaca cccctccac 7440
 accctctttc cccaacctcg tgttggttcgg agcgcacaca cacacaacca gatctcccc 7500
 aaatccaccc gtcggcacct ccgcttcaag gtacgccgct cgtcctcccc cccccccct 7560
 ctctaccttc tctagatcgg cgttccggtc catgggttagg gcccggtagt tctacttctg 7620
 ttcattgtttg tgtagatcc gtgtttgtgt tagatccgtg ctgctagcgt tcgtacacgg 7680
 atgcgacctg tacgtcagac acgttctgat tgctaacttg ccagtgtttc tctttgggga 7740
 atcctgggat ggctctagcc gttccgcaga cgggatcgat ttcattgattt tttttgttc 7800
 gttgcatagg gtttggtttg ccttttctt tttttcaat atatgccgtg cacttgtttg 7860
 tcgggtcatc ttttcattgt tttttttgtc ttggttggtga tgatgtggtc tggttgggag 7920
 gtcgttctag atcggagtag aattctgttt caaactacct ggtggattta ttaattttgg 7980
 atctgtatgt gtgtgccata catattcata gttacgaatt gaagatgatg gatggaaata 8040
 tcgacttagg ataggtatac atgttgatgc gggttttact gatgcatata cagagatgct 8100
 ttttgttcgc ttggttggtga tgatgtggtg tggttgggag gtcgttcatt cgttctagat 8160
 cggagtagaa tactgtttca aactacctgg tgtattttatt aattttggaa ctgtatgtgt 8220
 gtgtcataca tcttcattgt tacgagttta agatggatgg aaatategat ctaggatagg 8280
 tatacatgtt gatgtgggtt ttactgatgc atatacatga tggcatatgc agcatctatt 8340
 catatgctct aaccttgagt acctatctat tataataaac aagtatgttt tataattatt 8400
 ttgactctga tataacttga tgatggcata tgcagcagct atatgtggat ttttttagcc 8460
 ctgccttcat acgctattta tttgcttggg actgtttctt ttgtcgatgc tcacctgtt 8520
 gtttggtgtt acttctgcag ggatccccga tcatgcaaaa actcattaac tcagtcaaaa 8580
 actatgcctg gggcagcaaa acggcggttg ctgaacttta tggatggaa aatccgtcca 8640
 gccagccgat ggccgagctg tggatgggag cacatccgaa aagcagttca cgagtgcaga 8700
 atgccgccgg agatategtt tcaactgcgt atgtgattga gagtataaaa tcgactctgc 8760
 tcggagaggc cgttgccaaa cgctttggcg aactgccttt cctgttcaaa gtattatgcg 8820
 cagcacagcc actctccatt caggttcac caaacaaaca caattctgaa atcggttttg 8880
 ccaaagaaaa tgccgcaggt atcccgatgg atgccgccga gcgtaactat aaagatccta 8940
 accacaagcc ggagctggtt tttgcgctga cgcttttctt tgcgatgaac gcgtttcgtg 9000
 aattttccga gattgtctcc ctactccagc cggtcgcagg tgcacatccg gcgattgctc 9060
 actttttaca acagcctgat gccgaacgtt taagcgaact gttcgccagc ctgttgata 9120
 tgcaggggtga agaaaaatcc cgcgcgctgg cgattttaaa atcggccctc gatagccagc 9180
 aggggtgaacc gtggcaaacc attcgtttaa tttctgaatt ttaccggaa gacagcggtc 9240
 tgttctcccc gctattgctg aatgtggtga aattgaaccc tggcgaagcg atgttctctg 9300
 tcgctgaaac accgcacgct tacctgcaag gcgtggcgct ggaagtgatg gcaaaactccg 9360
 ataacgtgct gcgtgcgggt ctgacgccta aatacattga tattccggaa ctggttgcca 9420
 atgtgaaatt cgaagccaaa cgggctaacc agttgttgac ccagccggtg aaacaagggtg 9480

cagaactgga	cttccccgatt	ccagtggatg	attttgcctt	ctcgcctgcat	gaccttagtg	9540
ataaagaaac	caccattagc	cagcagagt	ccgccatttt	gttctgcgtc	gaaggcgatg	9600
caacgttgtg	gaaagggttct	cagcagttac	agcttaaacc	gggtgaatca	gcgtttattg	9660
ccgccaacga	atcaccgggtg	actgtcaaag	gccacggccg	tttagcgcgt	gtttacaaca	9720
agctgtaaga	gcttactgaa	aaaattaaca	tctcttgcta	agctgggagc	tcgtcatggg	9780
tcgtttaagc	tgccgatgtg	cctgcgtcgt	ctggtgccct	ctctccatat	ggaggttgtc	9840
aaagtatctg	ctgttcgigt	catgagtcgt	gtcagtggtg	gtttaataat	ggaccgggtg	9900
tgttgtgtgt	gcgtactacc	cagaactatg	acaaatcatg	aataagtttg	atgtttgaaa	9960
ttaaagcctg	tgctcattat	gttctgtctt	tcagttgtct	cctaataattt	gcctgcaggt	10020
actggctatc	taccgtttct	tacttaggag	gtgtttgaat	gcactaaaac	taatagttag	10080
tggctaaaat	tagttaaaac	atccaaacac	catagctaata	agttgaacta	ttagctattt	10140
ttggaaaatt	agttaatagt	gaggtagtta	tttgtttagct	agctaattca	actaacaatt	10200
tttagccaac	taacaattag	tttcagtga	ttcaaacacc	cccttaatgt	taacgtgggt	10260
ctatctaccg	tctcctaata	tatggttgat	tgttcgggtt	gttgctatgc	tattgggttc	10320
tgattgctgc	tagttcttgc	tgaatccaga	agttctcgta	gtatagctca	gattcatatt	10380
atatttttga	gtgataagt	atccaggtta	ttactatgtt	agctagggtt	tttttacaag	10440
gataaattat	ctgtgatcat	aattcttatg	aaagctttat	gtttcctgga	ggcagtgga	10500
tgcaatgcat	gacagcaact	tgatcacacc	agctgaggta	gatacggtaa	caaggttctt	10560
aaatctgttc	accaaactat	tggagaacac	acatacacat	tcttgccagt	cttggttaga	10620
gaaatttcat	gacaaaatgc	caaagctgtc	ttgactcttc	acttttggcc	atgagtcgtg	10680
acttagtttg	gtttaatgga	ccggttctcc	tagcttggtc	tactcaaaac	tggtgttgat	10740
gcgaataagt	tgtgatgggt	gatctctgga	ttttgttttg	ctctcaatag	tggacgagat	10800
tagatagccc	ggaaatttac	cggtgcccgg	gcggccagca	tggccgtatc	cgcaatgtgt	10860
tattaagttg	tctaagcgtc	aatttggtta	caccacaata	tatcctgcca	ccagccagcc	10920
aacagctccc	cgaccggcag	ctcggcacia	aatcaccact	cgatacaggc	agcccatcag	10980
aattaattct	catgtttgac	agcttatcat	cgactgcacg	gtgcaccaat	gcttctggcg	11040
tcaggcagcc	atcggaagct	gtggtatggc	tgtgcaggtc	gtaaatcact	gcataattcg	11100
tgtcgctcaa	ggcgcactcc	cgttctggat	aatgtttttt	gcgccgacat	cataacgggt	11160
ctggcaaata	ttctgaaatg	agctgttgac	aattaatcat	ccggctcgta	taatgtgtgg	11220
aattgtgagc	ggataacaat	ttcacacagg	aaacagacca	tgagggaagc	gttgatcgcc	11280
gaagtatcga	ctcaactatc	agaggtagtt	ggcgtcatcg	agcgccatct	cgaaccgacg	11340
ttgctggccg	tacatttgta	cggctccgca	gtggatggcg	gcctgaagcc	acacagtgat	11400
attgatttgc	tggttacgggt	gaccgtaagg	cttgatgaaa	caacgcggcg	agctttgatc	11460
aacgaccttt	tggaaacttc	ggcttcccct	ggagagagcg	agattctccg	cgtgttagaa	11520
gtcaccattg	ttgtgcacga	cgacatcatt	ccgtggcggt	atccagctaa	gcgcgaactg	11580
caatttgagg	aatggcagcg	caatgacatt	cttgcaggta	tcttcgagcc	agccaogate	11640

gacattgatc tggctatctt gctgacaaaa gcaagagaac atagcgttgc cttggtaggt 11700
ccagcggcgg aggaactctt tgatccgggt cctgaacagg atctatttga ggcgctaaat 11760
gaaaccttaa cgctatggaa ctgcgccccc gactgggctg gcgatgagcg aaatgtagtg 11820
cttacgttgt cccgcatttg gtacagcgca gtaaccggca aaatcgcgcc gaaggatgtc 11880
gctgccgact gggcaatgga ggcctgccc gcccagtatc agcccgcatc acttgaagct 11940
aggcaggctt atcttggaca agaagatcgc ttggcctcgc gcgcagatca gttggaagaa 12000
tttgttcact acgtgaaagg cgagatcacc aaagtagtcg gcaaataaag ctctagtgga 12060
tctccgtacc cggggatctg gctcgcggcg gacgcacgac gccggggcga gaccataggc 12120
gatctcctaa atcaatagta gctgtaacct cgaagcgttt cacttgtaac aacgattgag 12180
aatttttgtc ataaaattga aatacttggg tcgcattttt gtcacccgcg gtcagccgca 12240
attctgacga actgcccatt tagctggaga tgattgtaca tccttcacgt gaaaatttct 12300
caagcgcgtg gaacaagggt tcagatttta gattgaaagg tgagccgttg aaacacgttc 12360
ttcttgtcga tgacgacgtc gctatgcggc atcttattat tgaatacctt acgatccacg 12420
ccttcaaagt gaccgcggta gccgacagca ccagttcac aagagtactc tcttccgcga 12480
cggtcgatgt cgtggttgtt gatctagatt taggtcgtga agatgggctc gagatcgttc 12540
gtaatctggc ggcaaagtct gatattccaa tcataattat cagtggcgac cgccttgagg 12600
agacggataa agttgttgca ctcgagctag gagcaagtga ttttatcgct aagccgttca 12660
gtatcagaga gtttctagca cgcattcggg ttgccttgcg cgtgcgcccc aacgttgtcc 12720
gctccaaaga ccgacggtct ttttgtttta ctgactggac acttaatctc aggcaacgtc 12780
gcttgatgtc cgaagctggc ggtgaggtga aacttacggc aggtgagttc aatcttctcc 12840
tcgcgttttt agagaaaccc cgcgacgttc tatcgcgcg gcaacttctc attgccagtc 12900
gagtacgcca cgaggaggtt tatgacagga gtatagatgt tctcattttg aggctgcgcc 12960
gcaaacttga ggcagatccg tcaagccctc aactgataaa aacagcaaga ggtgccgggt 13020
atttctttga cgcggacgtg caggtttcgc acggggggac gatggcagcc tgagccaatt 13080
cccagatccc cgaggaatcg gcgtgagcgg tcgcaaacca tccggcccgg taaaaatcgg 13140
cgcggcgctg ggtgatgacc tgggtggagaa gttgaaggcc gcgcaggccg ccagcggca 13200
acgcacgag gcagaagcac gccccggtga atcgtggcaa gcggcgcgtg atcgaatccg 13260
caaagaatcc cggcaaccgc cggcagccgg tcgcgcgtcg attaggaagc cgcccaaggg 13320
cgacgagcaa ccagattttt tcgttccgat gctctatgac gtgggcaccc gcgatagtcg 13380
cagcatcatg gacgtggccg ttttccgtct gtcgaagcgt gaccgacgag ctggcgaggt 13440
gatccgctac gagcttccag acgggcacgt agaggtttcc gcagggccgg ccggcatggc 13500
cagtgtgtgg gattacgacc tgggtactgat ggcggtttcc catctaaccg aatccatgaa 13560
ccgataccgg gaaggggaagg gagacaagcc cggccgcgtg ttccgtccac acgttgcgga 13620
cgtactcaag ttctgccggc gagccgatgg cggaaagcag aaagacgacc tggtagaaac 13680
ctgcattcgg ttaaacacca cgcacgttgc catgcagcgt acgaagaagg ccaagaacgg 13740
ccgcctggtg acggtatccg aggggtgaagc cttgattagc cgctacaaga tcgtaaagag 13800

cgaaaccggg	cggccggagt	acatcgagat	cgagctagct	gattggatgt	accgcgagat	13860
cacagaaggc	aagaaccggg	acgtgctgac	ggttcacccc	gattactttt	tgatcgatcc	13920
cggcacggc	cgttttctct	accgcctggc	acgccgcgcc	gcaggcaagg	cagaagccag	13980
atggttggtc	aagacgatct	acgaacgcag	tggcagcgcc	ggagagttca	agaagttctg	14040
tttcaccgtg	cgcaagctga	tcgggtcaaa	tgacctgccg	gagtacgatt	tgaaggagga	14100
ggcggggcag	gctggcccga	tcctagtcac	gcgctaccgc	aacctgatcg	agggcgaagc	14160
atccgccggg	tcctaagtga	cggagcagat	gctagggcaa	attgccctag	caggggaaaa	14220
aggtcgaaaa	ggtctctttc	ctgtggatag	cacgtacatt	gggaacccaa	agccgtacat	14280
tgggaaccgg	aaccggtaca	ttgggaaccc	aaagccgtac	attgggaacc	ggtcacacat	14340
gtaagtgact	gatataaaag	agaaaaaagg	cgattttttc	gcctaaaact	ctttaaact	14400
tatttaaact	cttaaaaccc	gcctggcctg	tgcataactg	tctggccagc	gcacagccga	14460
agagctgcaa	aaagcgccca	cccttcggtc	gctgcgctcc	ctacgccccg	ccgcttcgcg	14520
tcggcctatc	gcggccgctg	gccgctcaaa	aatggctggc	ctacggccag	gcaatctacc	14580
agggcgcgga	caagccgcgc	cgtcgccact	cgaccgccgg	cgctgaggtc	tgccctcgta	14640
agaaggtgtt	gctgactcat	accaggcctg	aatcgcccca	tcattccagcc	agaaagtgag	14700
ggagccacgg	ttgatgagag	ctttgttgta	ggtggaccag	ttggtgattt	tgaacttttg	14760
ctttgccacg	gaacgggtctg	cgttgtcggg	aagatgcgtg	atctgatcct	tcaactcagc	14820
aaaagttcga	tttattcaac	aaagccgcgc	tcccgctcaag	tcagcgtaat	gctctgccag	14880
tgttacaacc	aattaaccaa	ttctgattag	aaaaactcat	cgagcatcaa	atgaaactgc	14940
aatttattca	tatcaggatt	atcaatacca	tatttttgaa	aaagccgttt	ctgtaatgaa	15000
ggagaaaact	caccgaggca	gttccatagg	atggcaagat	cctgggtatcg	gtctgcgatt	15060
ccgactcgtc	caacatcaat	acaacctatt	aatttccctt	cgtcaaaaat	aaggttatca	15120
agtgagaaat	caccatgagt	gacgactgaa	tcgggtgaga	atggcaaaaag	ctctgcatta	15180
atgaatcggc	caacgcgcgc	ggagaggcgc	tttgcgattt	gggcgctctt	ccgcttctct	15240
gctcactgac	tcgctgcgct	cggtcgttcg	gctgcggcga	gcggtatcag	ctcactcaaa	15300
ggcggtaata	cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	15360
aggccagcaa	aaggccagga	accgtaaaaa	ggccgcgttg	ctggcggtttt	tcctataggct	15420
ccgccccctt	gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	gaaaccgcac	15480
aggactataa	agataccagg	cgtttccccc	tggaagctcc	ctcgtgcgct	ctcctgttcc	15540
gaccctgccg	cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	tggcgctttc	15600
tcatagctca	cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	agctgggctg	15660
tgtgcacgaa	ccccccgttc	agcccgaccg	ctgcgcctta	tccggtaact	atcgtcttga	15720
gtccaaccgc	gtaagacacg	acttatcgcc	actgycagca	gccactggta	acaggattag	15780
cagagcgagg	tatgtaggcg	gtgctacaga	gttcttgaag	tgggtggcta	actacggcta	15840
cactagaaga	acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	tcggaaaaag	15900
agttggtagc	tcttgatccg	gcaaacaaac	caccgctggt	agcgggtggtt	tttttgtttg	15960

caagcagcag attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac 16020
 ggggtctgac gctcagtga acgaaaactc acgttaaggg attttgggtca tgagattatc 16080
 aaaaaggatc ttcacctaga tctttttgat ccggaatta 16119

<210> 22

<211> 16210

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vectơ tái tổ hợp

<400> 22

attcctgtgg ttggcatgca catacaaag gacgaacgga taaacctttt cagccctttt 60
 taaatatccg attattctaa taaacgctct tttctcttag gtttaccgc caatatatcc 120
 tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga 180
 attaaggagg tcacgttatg acccccgccg atgacgcggg acaagccgtt ttacgtttgg 240
 aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccatttaaa tcaattgggc 300
 gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc 360
 taccaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg 420
 gaacaaaata acgtggaaaa gagctgtcct gacagccac tactaatgc gtatgacgaa 480
 cgcagtgcag accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct 540
 cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaagggtg 600
 ctctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga 660
 cagtgggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc 720
 aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga 780
 acaatccac tatccttcgg taccggacc tctgcctttc tgttcttcaa acgatgtctc 840
 atgtctgcgc tggacaactt tcttggtgcc gctgtcgtc tgcgctgtgc tgactggacg 900
 cagctccgga gggttggttg tgcttggttt tcgtagagaa ctgcgcactt gccgccgca 960
 cgttcttggg gtttctcct ccccgctgtg ttctgcgcac gggctttttc tgagagaccc 1020
 atgtttccct tttactttta taaacagtat acatgctatg tttctagaag gaggggaaac 1080
 ctaatcccc taatccaatg gcggggagga aatagggtgg ggtgggggtg ggggagggaa 1140
 atatctcgct actttttaat ccggacaagc tcatttgctt ttgcgtctga atgatgatga 1200
 ctgcaatgct gatcgcacct cgggtgtcgg atcaccagct tttggctgct ctaccaaatt 1260
 cagctgcaag aagattagag cacaaaagaa ttacagaaag agagcctttt tcttttcttc 1320
 cttgtggggg tcttttcatt tcgtgctctc ctttctctgc cagccagtcc gtccttgctt 1380
 ccactgcacc tgcacacagg tcaccccgac ccgcactgtt ctagactcca ttagaaaaaa 1440
 aaagggtctga acctttccga aaccagccag ccattgggtc ggccaggccag catatgctaa 1500

ttggattttt	ttgccgcata	attgagtgcg	ccatcaggat	ttggaaatcc	tggttttgag	1560
taatacagta	atttggcatt	atccattgcc	gaattcccaa	gctccgtcag	cttgaacgtg	1620
gacccctacc	atctgcacca	gctcggcacc	tcacgctcgc	agcgctagga	gcctaggagc	1680
agctgcccgt	ctattttattg	gtccctctcc	cgtcccagag	aaaccctccc	tcctcctccc	1740
attggactgc	ttgctccctg	ttgaccattg	gggtatgctt	gctgccttgc	tctcctgttc	1800
atctccgtgc	taaacctctg	tcctctgggt	gggtttttgc	tgggattttg	agctaactctg	1860
ctggtcccg	tagaaaaaga	tcatgtcccc	tgacgtgctc	aagcgctcgc	cttagccgcg	1920
tccttgcccc	ccgccatttt	ttgcgggtttc	ggtgtgttcc	cgtgactcgc	cggtgtgcgc	1980
atcgccctgaa	tcttgtctgg	gctctgctga	catgttcttg	gctagtgtgg	tttatagatt	2040
cctctgatct	aaaccgtgcc	tgtgctgcgc	acagaactct	cccctgtcct	ttcctggggt	2100
tttggttacg	tggtggtagt	aagcttggat	ttgcacatgg	ataaagttgt	tctaagctcc	2160
gtgggttgct	tgagatcttg	ctgttattgc	gtgccgtgct	cacttttttt	gcaatccgag	2220
gaatgaattt	gtcgtttact	cgttttgggtg	gattattagc	gcgaaaaaaaa	aactcttttt	2280
tttttgttct	tttactacga	aaagcatctt	cttggatttt	gctatcttct	tttactacga	2340
aaaactcttg	agtctaggaa	tttgaatttg	tgatgtccat	tcttgacgtg	cgctgtgctt	2400
tattgggaag	ccaaatccta	ttattttctg	cctctagggt	ctgaatggaa	tcagtactct	2460
tgagacagaa	aatcaatcca	atcaagttga	tttcttttct	taaaaatatt	atcacagaac	2520
taagtgcctg	tgcggaatca	gtactggcct	ttgtttgggtg	gaggatcaat	acttgctttt	2580
gttttgggggt	ggcaactggt	ttgctataag	attccatgtg	ttcctgttga	gatgaatcat	2640
atatagtata	gctgcatact	acaaatctgt	ttttcaaatt	taggttgctt	tggcatgato	2700
tatttttttg	tcagacagac	tttctaagtg	gtagctcttg	atttcttggt	cttgtacaac	2760
tgggtgctgct	gaatcttgac	cgtatagctc	gaattgcagt	attctgaacc	atcgagccaa	2820
ggctgccaa	ctgactcgcc	tcacacgtct	tcgcgaacgc	cttgggtgcca	ccttctcctc	2880
ccatcccaat	gaactgatag	cactcttttc	caggtgggct	tacccaaatc	atataacttg	2940
catttcatct	ggtactgaaa	gttgtttaatt	tgttattctc	ttcatgcctg	tcttaatagc	3000
acaccagat	gtaaacacga	gattatgcaa	cttcttactt	ggtttctttt	gttggcacca	3060
tcatgcatgc	taattgctaa	ggatgttacc	tattcatcct	tgactcatat	tatcatatgt	3120
aatgatttta	tgatcacgag	actattgatt	gtgaagcata	gtatagctgt	tcttcagttt	3180
ttgtaccctt	ttgttttttt	ccttaagcta	gaactgggtac	aatttagttg	ataagacagt	3240
gtagtttgta	gtacgtcatt	tgacagattg	tttgtcttta	gctggtaaa	tgccatttta	3300
tatctgtatc	cttcagatct	aataaaaagg	atatgagatg	tccatcacia	gaggggaaaa	3360
attacatgat	ctgagatgta	acatccgttt	ttatttgtga	aataccactt	ctacagggtat	3420
cttcactagg	gtaaaccatg	gacaacaacc	ccaacatcaa	cgagtgcata	ccctacaact	3480
gcctgagcaa	ccccgagggtg	gaggtgctgg	ggggcgagcg	catcgagacc	ggctacaccc	3540
ccatcgacat	cagcctgagc	ctgaccacgt	tcctgctgag	cgagttcgtg	cccggcgccg	3600
gcttcgtgct	gggcctgggtg	gacatcatct	ggggcatctt	cggccccagc	cagtgggacg	3660

ccttcctggt	gcagatcgag	cagctgatca	accagcgcat	cgaggagttc	gcccgcgaacc	3720
aggccatcag	ccgcctggag	ggcctgagca	acctgtacca	aatctacgcc	gagagcttcc	3780
gcgagtggga	ggccgaqccc	accaaccccg	ccctgcgcga	ggagatgcgc	atccagttca	3840
acgacatgaa	cagcgccctg	accaccgcca	tccccctggt	cgccgtgcag	aactaccagg	3900
tgcccttget	gagcgtgtac	gtgcaggccg	ccaacctgca	cctgagcgtg	ctgcgcgacg	3960
tcagcgtggt	cggccagcgc	tggggcttcg	acgccgccac	catcaacagc	cgctacaacg	4020
acctgacccg	cctgatcggc	aactacaccg	accacgccgt	gcgctgggtac	aacaccggcc	4080
tggagcgcgt	gtgggggtcca	gacagccgcg	actggatcag	gtacaaccag	ttccgccgcg	4140
agctgaccct	gaccgtgctg	gacatcgtga	gcctgttccc	caactacgac	agccgcacct	4200
accccatccg	caccgtgagc	cagctgaccc	gcgagattta	caccaacccc	gtgctggaga	4260
acttcgacgg	cagcttccgc	ggcagcgccc	agggcatcga	gggcagcatc	cgcagccccc	4320
acctgatgga	catcctgaac	agcatcacca	tctacaccga	cgcaccgcgc	ggcgagtact	4380
actggagcgg	ccaccagatc	atggccagcc	ccgtcggctt	cagcggcccc	gagttcacct	4440
tccccctgta	cggcactatg	ggcaacgctg	cacctcagca	gcgcacgtg	gcacagctgg	4500
gccagggagt	gtaccgcacc	ctgagcagca	ccctgtaccg	tcgacctttc	aacatcggca	4560
tcaacaacca	gcagctgagc	gtgctggacg	gcaccgagtt	cgcctacggc	accagcagca	4620
acctgcccag	cgccgtgtac	cgcaagagcg	gcaccgtgga	cagcctggac	gagatccccc	4680
ctcagaacaa	caacgtgccca	cctcgacagg	gcttcagcca	ccgtctgagc	cacgtgagca	4740
tgttccgcag	tggcttcagc	aacagcagcg	tgagcatcat	ccgtgcacct	atgttcagct	4800
ggattcacccg	cagtgccgag	ttcaacaaca	tcatccccag	cagccagatc	accagatcc	4860
ccctgaccaa	gagcaccaac	ctgggcagcg	gcaccagcgt	ggtgaagggc	cccggcttca	4920
ccggcgggca	catcctgcgc	cgcaccagcc	ccggccagat	cagcacccctg	cgcgtgaaca	4980
tcaccgcccc	cctgagccag	cgctaccgcg	tcgcacccg	ctacgccagc	accaccaacc	5040
tgcagttcca	caccagcatc	gacggccgcg	ccatcaacca	gggcaacttc	agcgccacca	5100
tgagcagcgg	cagcaacctg	cagagcggca	gcttccgcac	cgtgggcttc	accacccct	5160
tcaacttcag	caacggcagc	agcgtgttca	ccctgagcgc	ccacgtgttc	aacagcggca	5220
acgaggtgta	catcgaccgc	atcgagttcg	tgcccgcgga	ggtgaccttc	gaggccgagt	5280
acgacctgga	gagggctcag	aaggccgtga	acgagctggt	caccagcagc	aaccagatcg	5340
gcctgaagac	cgacgtgacc	gactaccaca	tcgatcaggt	gtaggagctc	tacggggaaa	5400
gaaggagaag	aagaagaaga	agcccaggcc	ggagaacat	cgcctgcatt	tcgatctggt	5460
tcaccgcaat	tcgcattggt	agtcgtgtat	tggagttatg	tgtacttggt	ttccaagaac	5520
tttggttcc	tctcgttttt	ttttccttgc	ttgtttgagc	gtttttgggc	agcgtggcc	5580
tggttccctag	tatggtggga	attggctgca	ccttttgctt	cgaataaaaa	tgccctgctcg	5640
ttcacctgtc	ttccagagtg	caatgcgatg	ttctgttgcc	caggtcgtgt	ggtttctgact	5700
gatggcgatg	ttgtgttctt	ctgttaatcg	cctgttttaa	cgtggtaggc	tgatgcttgt	5760
tcttgttgag	aaagcttgct	gtgccagaca	tggctgcttg	aatacaagtg	aaggaaaaaa	5820

aaagccatgc	caagtaaagt	tgcacaaaat	ttcaactgct	cagtggacca	ctggaccatg	5880
ttcttgggta	ttgcagttgc	agggcttcac	atggcggttg	gacagcagtc	ttggattgat	5940
gcataaagag	gtgggtggta	atgaggacgc	aaggccgttc	cctcagagtc	agtcacaagg	6000
ttgcagaggt	cacggttctc	ttccctttcc	gcttcctgtc	acatcggaat	tggtgtttac	6060
gccatctgcc	catcacccac	caagtctatg	tttctgtact	ggatctttca	atggcggaac	6120
gcgcttagtt	cttcgtcaca	gtcgaatcac	atgatctaata	cgatgtcttt	aatctcgctg	6180
taaaaagggt	gggacggtag	gtgcagggtg	gggaccagg	aaggcctgcc	taaacgtatc	6240
cataaacatg	cacagcaacc	ctaagatatt	atactgccta	cttcctaaga	tatagttatt	6300
tctagtctat	tttttttccg	tccacatcca	aataagtgat	aatatataga	catacatata	6360
tatactatat	tcatcataga	ttaatgaacg	aatgtatact	tagtttaaac	ctaattatat	6420
tttaggaagg	atggagtatg	aaacatgaca	atacaacaaa	aaaaaatcat	gtaattgcat	6480
atcgtcaaag	ttatctgaag	taaccaatcc	agggggaaat	cccgttagca	aacatacaag	6540
agcaccgccc	cactacatcc	cagaaaataa	aacaaaacca	gaactcagat	ggataaataa	6600
tactaccgga	ccgcgatcgc	ttaattaagc	ttgcatgcct	gcagtgcagc	gtgaccgggt	6660
cgtgcccctc	tctagagata	atgagcattg	catgtctaag	ttataaaaaa	ttaccacata	6720
ttttttttgt	cacacttggt	tgaagtgcag	tttatctatc	tttatacata	tattttaaaact	6780
ttactctacg	aataatataa	tctatagtac	tacaataata	tcagtgtttt	agagaatcat	6840
ataaatgaac	agttagacat	ggtctaaagg	acaattgagt	attttgacaa	caggactcta	6900
cagttttatc	tttttagtgt	gcatgtgttc	tccttttttt	ttgcaaatag	cttcacctat	6960
ataatacttc	atccatttta	ttagtacatc	catttagggg	ttaggggtta	tggtttttat	7020
agactaattt	tttttagtaca	tctattttat	tctatttttag	cctctaaatt	aagaaaacta	7080
aaactctatt	ttagtttttt	tattttaataa	tttagatata	aaatagaata	aaataaagtg	7140
actaaaaatt	aaacaaatac	cctttaagaa	attaaaaaaa	ctaaggaaac	atttttcttg	7200
tttcgagtag	ataatgccag	cctgttaaac	gccgtcgacg	agtctaacgg	acaccaacca	7260
gcgaaccagc	agcgtcgcg	cgggccaaag	gaagcagacg	gcacggcatc	tctgtcgctg	7320
cctctggacc	cctctcgaga	gttccgctcc	accgttgga	ttgctccgct	gtcggcatcc	7380
agaaattg	tgccggagcg	gcagacgtga	gccggcacgg	caggcgccct	cctcctcctc	7440
tcacggcacc	ggcagctacg	ggggattcct	ttcccaccgc	tccttcgctt	tccttcctc	7500
gcccgcgta	ataaatagac	acccctcca	caccctcttt	cccacacctc	gtgttggtcg	7560
gagcgcacac	acacacaacc	agatctcccc	caaataccacc	cgtcggcacc	tcgcttcaa	7620
ggtacgccgc	tcgtcctccc	cccccccccc	tctctacctt	ctctagatcg	gcgttcgggt	7680
ccatgggttag	ggcccggtag	ttctacttct	gttcatgttt	gtgttagatc	cgtgtttgtg	7740
ttagatccgt	gctgctagcg	ttcgtacacg	gatgcgacct	gtacgtcaga	caggttctga	7800
ttgctaactt	gccagtgttt	ctctttgggg	aatcctggga	tggctctagc	cgttccgcag	7860
acgggatcga	tttcatgatt	ttttttgttt	cgttgcatag	ggtttggttt	gcccttttcc	7920
tttatttcaa	tatatgccgt	gcacttggtt	gtcgggtcat	cttttcatgc	ttttttttgt	7980

cttggtttgtg atgatgtggt ctggttgggc ggtcgttcta gatcggagta gaattctgtt 8040
 tcaaactacc tgggtggaattt attaatTTTT gatctgtatg tgtgtgccat acatattcat 8100
 agttacgaat tgaagatgat ggatggaaat atcgatctag gataggtata catgttgatg 8160
 cgggtttttac tgatgcatat acagagatgc tttttgttcg cttggtttgtg atgatgtggt 8220
 gtggttgggc ggtcgttcat tcgttctaga tcggagtaga atactgtttc aaactacctg 8280
 gtgtatttat taattttgga actgtatgtg tgtgtcatac atcttcatag ttacgagttt 8340
 aagatggatg gaaatatcga tctaggatag gtatacatgt tgalgtgggt tttactgatg 8400
 catatacatg atggcatatg cagcatctat tcatatgctc taaccttgag tacctatcta 8460
 ttataataaaa caagtatgtt ttataattat tttgatcttg atatacttgg atgatggcat 8520
 atgcagcagc tatatgtgga ttttttttagc cctgccttca tacgctatTT atttgcttgg 8580
 tactgtttct tttgtcgatg ctccacctgt tgtttggtgt tacttctgca gggatccccg 8640
 atcatgcaaa aactcattaa ctcagtgcaa aactatgcct ggggcagcaa aacggcgttg 8700
 actgaacttt atggtatgga aaatccgtcc agccagccga tggccgagct gtggatgggc 8760
 gcacatccga aaagcagttc acgagtgcag aatgccgccg gagatatcgt ttcactgcgt 8820
 gatgtgattg agagtgataa atcgactctg ctcgagagagg ccgttgccaa acgctttggc 8880
 gaactgcctt tcctgttcaa agtattatgc gcagcacagc cactctccat tcaggttcat 8940
 ccaaacaaac acaatttctga aatcggtttt gccaaagaaa atgccgcagg tatcccgatg 9000
 gatgccgcgc agcgtaacta taaagatcct aaccacaagc cggagctgggt ttttgcgctg 9060
 acgcctttcc ttgcgatgaa cgcgtttcgt gaattttccg agattgtctc cctactccag 9120
 ccggtcgcag gtgcacatcc ggcgattgct cactttttac aacagcctga tgccgaacgt 9180
 ttaagcgaac tgttcgccag cctgttgaat atgcaggggtg aagaaaaatc ccgcgcgctg 9240
 gcgattttta aatcggccct cgatagccag caggggtgaac cgtggcaaac gattcgttta 9300
 atttctgaat tttaccgga agacagcgggt ctgttctccc cgctattgct gaatgtgggtg 9360
 aaattgaacc ctggcgaagc gatgttctctg ttcgctgaaa caccgcacgc ttacctgcaa 9420
 ggcgtggcgc tggaagtgat ggcaaaactcc gataacgtgc tgcgtgcggg tctgacgcct 9480
 aaatacattg atattccgga actggttgcc aatgtgaaat tcgaagccaa accggctaac 9540
 cagttgttga ccagccgggt gaaacaaggt gcagaactgg acttcccgat tccagtggat 9600
 gattttgcct tctcgctgca tgaccttagt gataaagaaa ccaccattag ccagcagagt 9660
 gccgccattt tgttctgcgt cgaaggcgat gcaacgttgt ggaaagggtc tcagcagtta 9720
 cagcttaaac cgggtgaatc agcgtttatt gccgccaacg aatcaccgggt gactgtcaaa 9780
 ggccacggcc gtttagcgcg tgtttacaac aagctgtaag agcttactga aaaaattaac 9840
 atctcttgct aagctgggag ctcgatcatg gtcgtttaag ctgccgatgt gcctgcgtcg 9900
 tctggtgccc tctctccata tggaggttgt caaagtatct gctgttcgtg tcatgagtcg 9960
 tgtcagtgtt ggtttaataa tggaccgggt gtgttggtgtg tgcgtactac ccagaactat 10020
 gacaaatcat gaataagttt gatgtttgaa attaaagcct gtgctcatta tgttctgtct 10080
 ttcagttgtc tctaataatt tgccctgcagg tactggctat ctaccgtttc ttacttagga 10140

ggtgtttgaa tgcactaaaa ctaatagtta gtggctaaaa ttagttaaaa catccaaaca 10200
 ccatagctaa tagttgaact attagctatt tttggaaaat tagttaatag tgaggtagtt 10260
 atttgttagc tagctaattc aactaacaat ttttagccaa ctaacaatta gtttcagtgc 10320
 attcaaacac ccccttaatg ttaacgtggg tctatctacc gtctcctaata atatggttga 10380
 ttgttcgggt tgttgctatg ctattgggtt ctgattgctg ctagtctctg ctgaatccag 10440
 aagttctcgt agtatagctc agattcatat tatttatttg agtgataagt gatccagggt 10500
 attactatgt tagctagggt ttttttaciaa ggataaatta tctgtgatca taattcttat 10560
 gaaagcttta tgtttcctgg aggcagtggc atgcaatgca tgacagcaac ttgatcacac 10620
 cagctgaggt agatacggta acaagggttct taaatctgtt caccaaataca ttggagaaca 10680
 cacatacaca ttcttgccag tcttggttag agaaatttca tgacaaaatg ccaaagctgt 10740
 cttgactctt cacttttggc catgagtcgt gacttagttt ggtttaatgg accggttctc 10800
 ctagcttggt ctactcaaaa ctgttggtga tgcgaataag ttgtgatggg tgatctctgg 10860
 attttgtttt gctctcaata gtggacgaga ttagatagcc cggaaattta ccggtgccc 10920
 ggcggccagc atggccgtat ccgcaatgtg ttattaagtt gtctaagcgt caatttgttt 10980
 acaccacaat atatcctgcc accagccagc caacagctcc ccgaccggca gctcggcaca 11040
 aaatcaccac tcgatacagg cagcccatca gaattaattc tcatgtttga cagcttatca 11100
 tcgactgcac ggtgcaccaa tgcttctggc gtcaggcagc catcggaagc tgtggtatgg 11160
 ctgtgcaggt cgtaaatcac tgcataattc gtgtcgctca aggcgcactc ccgttctgga 11220
 taatgttttt tgcgcccaga tcataacggg tctggcaaat attctgaaat gagctgttga 11280
 caattaatca tccggctcgt ataattgtgtg gaattgtgag cggataacaa tttcacacag 11340
 gaaacagacc atgagggagc cgttgatcgc cgaagtatcg actcaactat cagaggtagt 11400
 tggcgtcatc gagcgccatc tcgaaccgac gttgctggcc gtacatttgt acggctccgc 11460
 agtggatggc ggcctgaagc cacacagtga tattgatttg ctgggttacgg tgaccgtaag 11520
 gcttgatgaa acaacgcggc gagctttgat caacgacctt ttggaaactt cggcttcccc 11580
 tggagagagc gagattctcc gcgctgtaga agtcaccatt gttgtgcagc acgacatcat 11640
 tccgtggcgt tatccagcta agcgcgaact gcaatttgga gaatggcagc gcaatgacat 11700
 tcttgacagg atcttcgagc cagccacgat cgacattgat ctggctatct tgctgacaaa 11760
 agcaagagaa catagcgttg ccttggttagg tccagcggcg gaggaactct ttgatccggg 11820
 tcctgaacag gatctatttg aggcgctaaa tgaaacctta acgctatgga actcgccgcc 11880
 cgactgggct ggcgatgagc gaaatgtagt gcttacgttg tcccgcattt ggtacagcgc 11940
 agtaaccggc aaaatcgccg cgaaggatgt cgctgcgcac tgggcaatgg agcgccctgc 12000
 ggcccagtat cagcccgta tacttgaagc taggcaggct tatcttggac aagaagatcg 12060
 cttggcctcg cgcgcagatc agttggaaga atttgttcac tacgtgaaag gcgagatcac 12120
 caaagtagtc ggcaaataaa gctctagtgg atctccgtac ccggggatct ggctcgcggc 12180
 ggacgcacga cgcggggcg agaccatagg cgatctccta aatcaatagt agctgtaacc 12240
 tcgaagcgtt tcacttgtaa caacgattga gaattttgt cataaaattg aaatacttgg 12300

ttcgcatttt	tgtcatccgc	ggtcagccgc	aattctgacg	aactgcccac	ttagctggag	12360
atgattgtac	atccttcacg	tgaaaatttc	tcaagcgctg	tgaacaaggg	ttcagatttt	12420
agattgaaag	gtgagccggt	gaaacacggt	cttcttgtcg	atgacgacgt	cgctatgcgg	12480
catcttatta	ttgaatacct	tacgatccac	gccttcaaag	tgaccgcggt	agccgacagc	12540
accagttca	caagagtact	ctcttccgcg	acggtcgatg	tcgtaggttg	tgatctagat	12600
ttaggtcgtg	aagatgggct	cgagatcggt	cgtaatctgg	cggcaaagtc	tgatattcca	12660
atcataatta	tcagtggcga	ccgccttgag	gagacggata	aagttgttgc	actcgagcta	12720
ggagcaagtg	atcttatcgc	taagccgttc	agtatcagag	agtttctagc	acgcattcgg	12780
gttgccctgc	gcgtgcgccc	caacgttgtc	cgctccaaag	accgacggtc	tttttgtttt	12840
actgactgga	cacttaatct	caggcaacgt	cgcttgatgt	ccgaagctgg	cggtgaggtg	12900
aaacttacgg	caggtgagtt	caatcttctc	ctcgcgtttt	tagagaaacc	ccgcgacggt	12960
ctatcgcgcg	agcaacttct	cattgccagt	cgagtacgcg	acgaggaggt	ttatgacagg	13020
agtatagatg	ttctcatttt	gaggctgcgc	cgcaaacttg	aggcagatcc	gtcaagccct	13080
caactgataa	aaacagcaag	aggtgccggt	tatttctttg	acgcggacgt	gcaggtttcg	13140
cacgggggga	cgatggcagc	ctgagccaat	tcccagatcc	ccgaggaatc	ggcgtgagcg	13200
gtcgcaaacc	atccggcccc	gtacaaatcg	gcgcggcgct	gggtgatgac	ctggtggaga	13260
agttgaaggc	cgcgcaggcc	gccagcggc	aacgcacga	ggcagaagca	cgccccgggtg	13320
aatcgtggca	agcggccgct	gatcgaatcc	gcaaagaatc	ccggcaaccg	ccggcagccg	13380
gtgcgcgctc	gattaggaag	ccgcccgaag	gcgacgagca	accagatttt	ttcgttccga	13440
tgctctatga	cgtgggcacc	cgcgatagtc	gcagcatcat	ggacgtggcc	gttttccgctc	13500
tgctgaagcg	tgaccgacga	gctggcgagg	tgatccgcta	cgagcttcca	gacgggcacg	13560
tagaggtttc	cgcagggccg	gccggcatgg	ccagtgtgtg	ggattacgac	ctggtactga	13620
tggcggtttc	ccatctaacc	gaatccatga	accgataccg	ggaagggaag	ggagacaagc	13680
ccggcccgct	gttccgtcca	cacgttgccg	acgtactcaa	gttctgccgg	cgagccgatg	13740
gcggaaagca	gaaagacgac	ctggtagaaa	cctgcattcg	gttaaacc	acgcacgttg	13800
ccatgcagcg	tacgaagaag	gccaagaacg	gccgcctggt	gacggtatcc	gaggggtgaag	13860
ccttgattag	ccgctacaag	atcgtaaaga	gcgaaaccgg	gcggccggag	tacatcgaga	13920
tcgagctagc	tgattggatg	taccgcgaga	tcacagaagg	caagaaccgg	gacgtgctga	13980
cggttcaccc	cgattacttt	ttgatcgatc	ccggcatcgg	ccgttttctc	taccgcctgg	14040
cacgccgcgc	cgcaggcaag	gcagaagcca	gatggttggt	caagacgatc	tacgaacgca	14100
gtggcagcgc	cggagagttc	aagaagttct	gtttcacctg	gcgcaagctg	atcgggtcaa	14160
atgacctgcc	ggagtacgat	ttgaaggagg	aggcggggca	ggctggcccc	atcctagtca	14220
tgcgtaccg	caacctgatc	gagggcggaag	catccgccgg	ttcctaattg	acggagcaga	14280
tgctagggca	aattgcccta	gcaggggaaa	aaggtcgaaa	aggtctcttt	cctgtggata	14340
gcacgtacat	tgggaaccca	aagccgtaca	ttgggaaccg	gaacccttac	attgggaacc	14400
caaagccgta	cattgggaac	cggtcacaca	tgtaagtgc	tgatataaaa	gagaaaaaag	14460

gcgatttttc cgcctaaaaac tctttaaaaac ttattaaaaac tcttaaaaacc cgcctggcct 14520
 gtgcataact gtctggccag cgcacagccg aagagctgca aaaagcgcct acccttcggt 14580
 cgtcgcgtc cctacgcccc gccgcttcgc gtcggcctat cgcggccgct ggccgctcaa 14640
 aaatggctgg cctacggcca ggcaatctac cagggcgcgg acaagccgcg ccgtcggcac 14700
 tcgaccgccc gcgctgaggt ctgcctcgtg aagaaggtgt tgctgactca taccaggcct 14760
 gaatcgcccc atcatccagc cagaaagtga gggagccacg gttgatgaga gctttgttgt 14820
 aggtggacca gttggtgatt ttgaactttt gctttgccac ggaacggtct gcgttgctcg 14880
 gaagatgctg gatctgatcc ttcaactcag caaaagttcg atttattcaa caaagccgcc 14940
 gtcccgctcaa gtcagcgtaa tgctctgcca gtgttacaac caattaacca attctgatta 15000
 gaaaaactca tcgagcatca aatgaaactg caattttattc atatcaggat tatcaatacc 15060
 atatttttga aaaagccggt tctgtaatga aggagaaaac tcaccgaggc agttccatag 15120
 gatggcaaga tcctgggtatc ggtctgcgat tccgactcgt ccaacatcaa tacaacctat 15180
 taatttcccc tcgtcaaaaa taaggttatc aagtgagaaa tcaccatgag tgacgactga 15240
 atccggtgag aatggcaaaa gctctgcatt aatgaatcgg ccaacgcgcg gggagaggcg 15300
 gtttgcgat tgggcgctct tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc 15360
 ggctgcggcg agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag 15420
 gggataacgc aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa 15480
 aggcgcggtt gctggcggtt ttccatagc tccgcccccc tgacgagcat cacaaaaatc 15540
 gacgctcaag tcagagggtg cgaaacccga caggactata aagataccag gcgtttcccc 15600
 ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg 15660
 cttttctccc ttcggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt 15720
 cgggtgtaggt cgttcgctcc aagctgggct gtgtgcacga acccccgtt cagcccgacc 15780
 gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc 15840
 cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 15900
 agttcttgaa gtggtggcct aactacggt acactagaag aacagtattt ggtatctgcg 15960
 ctctgctgaa gccagttacc ttcgaaaaa gagttggtag ctcttgatcc ggcaaacaaa 16020
 ccaccgctgg tagcgggtgt ttttttgttt gcaagcagca gattacgcgc agaaaaaag 16080
 gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact 16140
 cacgttaagg gatcttggtc atgagattat caaaaaggat cttcacctag atccttttga 16200
 tccggaatta 16210

<210> 23

<211> 15881

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vectơ tái tổ hợp

<400> 23

```

attcctgtgg ttggcatgca catacaaattg gacgaacgga taaacctttt cacgcccttt      60
taaatatccg attattctaa taaacgctct tttctcttag gtttaccgcg caatatatcc      120
tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga      180
attaagggag tcacgttatg acccccgcgc atgacgcggg acaagccgtt ttacgttttg      240
aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccatttaaa tcaattgggc      300
gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc      360
taccaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg      420
gaacaaaata acgtggaaaa gagctgtcct gacagcccac tcactaatgc gtatgacgaa      480
cgcagtgcag accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct      540
cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaaggtgg      600
ctcctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga      660
cagtgggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc      720
aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga      780
acaatcccac tatccttcgg taccggaccc atgccttacg atcatctgac atcttaaata      840
tttcaaactg cagtatttca aaaactgtgg ttttgtcaaa aactttgttc ccaaactgag      900
actaaaatgt agtgctagat aataaggcct taccggttg gaaatgacaaa aagaaacaga      960
aaacacctga aaaatagtat tgcccttggt ctaaattata aatcgttttg gttttattta     1020
atatctattt tttaagcata ctaaaatcta tgtacctatg aaaaataaaa tatgaactac     1080
aatttaagac tgagtacagt gtacagcgat aagttatatg aaccacaatt gtttttgaaa     1140
ggaaaaaaac taaattgaca agagaaacaa tcaactcttt tagaaacttt gtataccaaa     1200
aaacgaaata aagaagaaga aaaaaatata cacgatgagt ttattagccc acaaaacatg     1260
aaaaaaaaatg agagttaaag cctaactcag gccacaaaag aatcacaagc gcccatattc     1320
acagccgcgc aaaggtttcg gccccagtt ggatctcaga gccagtggtg aaataggact     1380
gacttgtctc tgatggttac gttacgtagg gattgaaaca aacaggcagt gagccgatgt     1440
ggcagaccgg agcgcagaag cgacgacgac catgaactga aggcgctggc agtgggccat     1500
gcgcccagaa cgctgacgaa ccacgacgac gcggacgttc tacgaggtcg gtggggccgt     1560
cggcagcccg ccggagcata ttcgccgtcg acagaggcag agaaggggaa tgtgtgggag     1620
cataatccac ccgtgggaag cctcggagtc ggagatgtgc agatctgccg gacgcacatg     1680
cgcacagctg ccgcaataga aaccagtgtc tttttatttt attttataag aaaaaactag     1740
aaaagaaaat ccagtcttct atacaggcct gcacagagag agaggaggga aggtcaaatc     1800
gtagaaaaaa aatgcttctt tttattgtta cgacattggc ctgaatatat actcattggt     1860
tctttcctta acttcaacaa ttattgattt tgacaatcca ttttttattt tagttcgcca     1920
aggttttggt tggagaagaa tttaaaacct agccaagcag tcgagcacga gagccgccgc     1980
cgccgcgcgc gccgttgga ggacctgtg acctctgacg accggacacg gacgttttcc     2040

```


aaagtcctaaa	ctcggggtag	gtaggtcgtg	aggctgctgc	gttcgccaag	ggaagaagca	2100
tctagttgaa	gctagaaacg	ggcagggggg	agggggccgac	gctatcggaa	atctgaagcc	2160
cagcacggac	ggacggacgg	gcagggggcg	cggcacatcc	tcgctcgacc	gccgcgccgc	2220
gtccgacatg	cgcaagctgc	tgctcttccc	gcgccccccac	ggcggccggg	ccggcagagc	2280
acgggacgcg	cgggggcggg	ctgacgtcgt	tttcacgtgg	tccgtcaccg	ttgcttgcat	2340
agcgtacagg	agtatagtcc	cattttccgaa	accaaaaaaa	gtaaacaaaa	aaatgccgtc	2400
gcattagttg	gctggcaagg	gaagaccaag	atggttttca	aatcaaaccat	gtaaaaaaat	2460
gtttcttttct	tttttgaacg	aactttaccag	ttcgaattct	ttaccggcat	tatttggttta	2520
aattgtttccc	caagaaagaa	agaaaagaaa	agcgggggga	aatatgcctt	ccttccttct	2580
tgcgtcacag	accacgcgcg	gatacacccg	acaacagcca	ggcagcaagc	agagcaccca	2640
ccgccggaag	cggccgtcgg	acagaacgac	acgtggggca	cggcacagtc	cgggccccga	2700
cgtcatcgaa	gacacctgac	ctgcccctgc	gtcttctaga	gaaaggcgga	cggacaggtc	2760
acccccgcac	ctcattccag	gagttgcatt	tcgctttctc	ctttatttat	ttataccaat	2820
aaaaaataaa	tccgcctttt	ctcctcccta	tcgtgtgtct	tcctctcgcc	ggctttaaaa	2880
acgcacacaa	gcgctaaaaa	cctctccacc	gtccacctca	gctcccatat	ccgctccctt	2940
acctctccag	catactcccg	tctcgtcgt	ctcgtctccg	ctcctcacct	cgccgctagc	3000
caagggtaac	tcctcgctcc	cgcccgggcc	ccgcgtcggg	gttttccatt	tactactact	3060
cgctgctccc	tcctgctccg	tgctcagact	cagatccgac	caaagcgggt	tcgctgacta	3120
aattctcctt	cgtttttttt	tctttttctg	gacggattcc	tggacgcagg	caagatcaag	3180
atcggaatca	acggtgagtc	tgtaccccca	acacaaactc	gttcttctctg	ctcgggttcgt	3240
tgggtctgga	ctctggagtg	atctgagtg	ggctctctgcg	cgctgctg	ctcaggtttc	3300
tgaaggatcg	gcagtaaacc	atggacaaca	accccaacat	caacgagtgc	atccctaca	3360
actgcctgag	caaccccgag	gtggaggtgc	tgggcggcga	gcgcacgag	accggctaca	3420
cccccatcga	catcagcctg	agcctgaccc	agttcctgct	gagcgagttc	gtgcccggcg	3480
ccggcttcgt	gctgggcctg	gtggacatca	tctggggcat	cttcggcccc	agccagtggg	3540
acgccttcct	ggtgcagatc	gagcagctga	tcaaccagcg	catcgaggag	ttcgcccga	3600
accaggccat	cagccgcctg	gagggcctga	gcaacctgta	ccaaatctac	gccgagagct	3660
tccgcgagtg	ggaggccgac	cccaccaacc	ccgcctgctg	cgaggagatg	cgcattccagt	3720
tcaacgacat	gaacagcgcc	ctgaccaccg	ccatccccct	gttcgcctg	cagaactacc	3780
aggtgcccc	gctgagcgtg	tacgtgcagg	ccgccaaact	gcacctgagc	gtgctgcgcg	3840
acgtcagcgt	gttcggccag	cgctggggct	tcgacgcgc	caccatcaac	agccgctaca	3900
acgacctgac	ccgcctgac	ggcaactaca	ccgaccacgc	cgtgcgctgg	tacaacaccg	3960
gcctggagcg	cgtgtgggg	ccagacagcc	gcgactggat	caggtacaac	cagttccgcc	4020
gcgagctgac	cctgaccgtg	ctggacatcg	tgagcctgtt	ccccaaactac	gacagccgca	4080
cctaccccat	ccgcaccgtg	agccagctga	cccgcgagat	ttacaccaac	cccgctgctg	4140
agaacttcga	cggcagcttc	cgcggcagcg	cccagggcat	cgagggcagc	atccgcagcc	4200

cccacctgat	ggacatcctg	aacagcatca	ccatctacac	cgacgccac	cgcggcgagt	4260
actactggag	cgccaccag	atcatggcca	gccccgtcgg	cttcagcggc	cccagattca	4320
ccttccccct	gtacggcact	atgggcaacg	ctgcacctca	gcagcgcata	gtggcacagc	4380
tgggccaggg	agtgtaccgc	accctgagca	gcaccctgta	ccgtcgacct	ttcaacatcg	4440
gcatcaacaa	ccagcagctg	agcgtgctgg	acggcaccga	gttcgcctac	ggcaccagca	4500
gcaacctgcc	cagcgccgtg	taccgcaaga	gcggcaccgt	ggacagcctg	gacgagatcc	4560
cccctcagaa	caacaacgtg	ccacctcgac	agggtttcag	ccaccgtctg	agccacgtga	4620
gcatgtttcg	cagtggcttc	agcaacagca	gcgtgagcat	catccgtgca	cctatgtttca	4680
gctggattca	ccgcagtgcc	gagttcaaca	acatcatccc	cagcagccag	atcaccacaga	4740
tccccctgac	caagagcacc	aacctgggca	gcggcaccag	cgtggtgaag	ggccccggct	4800
tcaccggcgg	cgacatcctg	cgccgcacca	gccccggcca	gatacagcacc	ctgcgcgtga	4860
acatcaccgc	ccccctgagc	cagcgctacc	gcgtccgcat	ccgctacgcc	agcaccacca	4920
acctgcagtt	ccacaccagc	atcgacggcc	gccccatcaa	ccagggcaac	ttcagcgcca	4980
ccatgagcag	cggcagcaac	ctgcagagcg	gcagctttccg	caccgtgggc	ttcaccaccc	5040
ccttcaactt	cagcaacggc	agcagcgtgt	tcacctgag	cgccacgtg	ttcaacagcg	5100
gcaacgaggt	gtacatcgac	cgcatacgagt	tcgtgcccg	cgaggtgacc	ttcgaggccg	5160
agtacgacct	ggagagggct	cagaaggccg	tgaacgagct	gttcaccagc	agcaaccaga	5220
tcggcctgaa	gaccgacgtg	accgactacc	acatcgatca	ggtgtaggag	ctcagagaga	5280
tattttctgcc	tccctatcgt	gggtcgtccc	caatggcctt	tggtcgcaga	ccatctttgc	5340
tgcttgtcta	tgctgagaat	aaatgtgaac	ggtgccctg	gacgctggat	ccatgctggt	5400
tttgacacg	gttgtctttt	tgtgtttaac	ttatctgcta	ccgtcctgta	acgaattcac	5460
taagttctag	ttcttttctg	ctttgttaag	aatataaaca	atgaggtttc	aattttggtg	5520
gcgccatccg	gtctgattgc	gtactagtat	catgcataac	tgagctggtc	tttcgggagt	5580
aacagttttt	ttttgtttcg	tgtcaatgat	tctcagcctg	ttcattttggt	cttaattccat	5640
atcggttact	actataatgt	tgtcttctct	ctatagattg	cagcagttca	acacatgttc	5700
ttggtatcct	tttggggtgtg	tttaatgcta	tttgctctgt	gttgtctttt	tcgttaatcg	5760
ggacagagtt	ttctctctgt	tttgacagtt	tcgggtggttc	ttttttgtgt	gtggtcagcg	5820
agtcacgaac	tgctggttgg	ccagcgaaat	agcgcagact	attatgggtc	ttctgctggt	5880
gaatatggaa	ggacgaggtt	aaattttctg	ctttgatgtg	tctgccctct	gccgctgtgc	5940
tgcatgcatg	tgtttgggag	cgaagcagcg	tgctgtttgg	gcgtatggtc	ggcatggtgt	6000
taaattattt	cgctgattca	gctttgggtt	atcactgggt	ttgatgggtca	cgctgtgtcg	6060
taagagagat	cgtggcagca	aaatggagaa	aaaggcagtg	aaacaaactt	tcagggtgtg	6120
gaatgtatgg	gattcgtatg	gttgaggcca	ggatttgtca	attggtcaag	ttgaaaagta	6180
aaggcgagac	agattggcga	ggaccgagga	gagattgggg	ttcagaaaact	tcggttacac	6240
tttaatcccc	tccatctccc	tccactcctc	ccttctacgg	accgcgacgc	cttaattaag	6300
cttgcatgcc	tgacgtgcag	cgtgaccggg	tcgtgccctt	ctctagagat	aatgagcatt	6360

gcatgtctaa gttataaaaa attaccacat attttttttg tcacacttgt ttgaagtgca	6420
gtttatctat ctttatacat atattttaaac tttactctac gaataatata atctatagta	6480
ctacaataat atcagtgttt tagagaatca tataaatgaa cagttagaca tgggtctaaag	6540
gacaattgag tatttttgaca acaggactct acagttttat ctttttagtg tgcattgtgtt	6600
ctcctttttt ttgcaaata gcttcaccta tataatactt catccatttt attagtagat	6660
ccatttaggg tttaggggta atgggttttta tagactaatt ttttttagtac atctatttta	6720
ttctatttta gcctctaaat taagaaaact aaaactctat ttttagtttt ttattttaata	6780
atthagatat aaaatagaat aaaataaagt gactaaaaat taaacaaata ccctttaaga	6840
aattaaaaaa actaaggaaa catttttctt gtttcgagta gataatgcca gcctgttaaa	6900
cgccgtcgac gagtctaacg gacaccaacc agcgaaccag cagcgtcgcg tcgggccaag	6960
cgaagcagac ggcacggcat ctctgtcgct gcctctggac ccctctcgag agttccgctc	7020
caccgttggg cttgctccgc tgtcggcatc cagaaattgc gtggcgagac ggcagacgtg	7080
agccggcacg gcaggcggcc tctcctcct ctacggcac cggcagctac gggggattcc	7140
tttcccaccg ctcttcgct ttccttctc cgcccgccgt aataaataga caccctctcc	7200
acaccctctt tcccacacct cgtgttggtc ggagcgcaca cacacacaac cagatctccc	7260
ccaaatccac ccgtcggcac ctccgcttca aggtacgccg ctcgctctcc ccccccccc	7320
ctctctacct tctctagatc ggcgttcggg tccatgggta gggcccggta gttctacttc	7380
tgttcatgtt tgtgttagat ccgtgtttgt gttagatccg tgcgtctagc gttcgtacac	7440
ggatgcgacc tgtacgtcag acacgttctg attgctaact tgccagtgtt tctctttggg	7500
gaatcctggg atggctctag ccgttcgcga gacgggatcg atttcatgat tttttttgtt	7560
tcgttgcata gggtttggtt tgcccttttc ctttatttca atatatgccg tgcacttgtt	7620
tgtcgggtca tcttttcatg cttttttttg tcttggttgt gatgatgtgg tctgggtggg	7680
cggtcgttct agatcggagt agaattctgt ttcaaactac ctggtggatt tattaatttt	7740
ggatctgtat gtgtgtgcca tacatatcca tagttacgaa ttgaagatga tggatggaaa	7800
tatcgatcta ggataggtat acatgttgat gcgggtttta ctgatgcata tacagagatg	7860
ctttttgttc gcttggttgt gatgatgtgg tgtgggtggg cggtcgttca ttcggtctag	7920
atcggagtag aatactgttt caaactacct ggtgtattta ttaatttttg aactgtatgt	7980
gtgtgtcata catcttcata gttacgagtt taagatggat ggaaatatcg atctaggata	8040
ggtatacatg ttgatgtggg ttttactgat gcatatacat gatggcatat gcagcatcta	8100
ttcatatgct ctaaccttga gtacctatct attataataa acaagtatgt ttataatta	8160
ttttgatctt gatatacttg gatgatggca tatgcagcag ctatatgtgg attttttttag	8220
ccctgccttc atacgtatct tatttgcttg gtactgtttc ttttgctgat gctcaccctg	8280
ttgtttggtg ttacttctgc agggatcccc gatcatgcaa aaactcatta actcagtgca	8340
aaactatgcc tggggcgagc aaacggcggt gactgaactt tatgggatgg aaaatccgtc	8400
cagccagccg atggccgagc tgtggatggg cgcacatccg aaaagcagtt cacgagtgca	8460
gaatgccgcc ggagatatcg tttcactgcg tgatgtgatt gagagtgata aatcgactct	8520

gctcggagag	gccgttgcca	aacgcttttg	cgaactgcct	ttcctgttca	aagtattatg	8580
cgcagcacag	ccactctcca	ttcaggttca	tccaaacaaa	cacaattctg	aaatcggttt	8640
tgccaaagaa	aatgccgcag	gtatcccgat	ggatgccgcc	gagcgtaact	ataaagatcc	8700
taaccacaag	ccggagctgg	tttttgcgct	gacgcctttc	cttgcgatga	acgcgtttcg	8760
tgaattttcc	gagattgtct	ccctactcca	gccggtcgca	ggtgcacatc	cggcgattgc	8820
tcacttttta	caacagcctg	atgccgaacg	tttaagcgaa	ctgttcgcca	gcctgttgaa	8880
tatgcagggg	gaagaaaaat	cccgcgcgct	ggcgatttta	aaatcggccc	tcgatagcca	8940
gcaggggtgaa	ccgtggcaaa	cgattcgttt	aattttctgaa	ttttaccggg	aagacagcgg	9000
tctgtttctcc	ccgctattgc	tgaatgtggg	gaaattgaac	cctggcgaag	cgatgttctt	9060
gttcgctgaa	acaccgcacg	cttacctgca	aggcgtggcg	ctggaagtga	tggcaaactc	9120
cgataacgtg	ctgcgtgcgg	gtctgacgcc	taaatacatt	gatattccgg	aactggttgc	9180
caatgtgaaa	ttcgaagcca	aaccggctaa	ccagttgttg	accagccggg	tgaaacaagg	9240
tgcagaactg	gacttcccca	ttccagtgga	tgattttgcc	ttctcgcctg	atgaccttag	9300
tgataaagaa	accaccattt	gccagcagag	tgccgccatt	ttgtttctgcg	tcgaaggcga	9360
tgcaacgttg	tggaaagggt	ctcagcagtt	acagcttaaa	ccgggtgaat	cagcgtttat	9420
tgccgccaac	gaatcacccg	tgactgtcaa	aggccacggc	cgttttagcg	gtgtttacaa	9480
caagctgtaa	gagcttactg	aaaaaattaa	catctcttgc	taagctggga	gctcgtcatg	9540
ggtcgtttta	gctgccgatg	tgccctgcgc	gtctgggtgc	ctctctccat	atggagggtg	9600
tcaaagtatc	tgctgttcgt	gtcatgagtc	gtgtcagtgt	tggtttaata	atggaccggt	9660
tgtgtttgtg	gtgcgtacta	cccagaacta	tgacaaatca	tgaataagtt	tgatgtttga	9720
aattaaagcc	tgtgctcatt	atgttctgtc	tttcagttgt	ctcctaatat	ttgcctgcag	9780
gtactggcta	tctaccgttt	cttacttagg	aggtgtttga	atgcactaaa	actaatagtt	9840
agtggctaaa	attagttaaa	acatccaaac	accatagcta	atagttgaac	tattagctat	9900
ttttggaaaa	ttagttaata	gtgaggtagt	tatttggttag	ctagctaatt	caactaacia	9960
tttttagcca	actaacaatt	agtttcagtg	cattcaaaca	cccccttaat	gttaacgtgg	10020
ttctatctac	cgtctcctaa	tatatgggtg	attgttcggg	ttgttgctat	gctattgggt	10080
tctgattgct	gctagttctt	gctgaatcca	gaagttctcg	tagtatagct	cagattcata	10140
ttatttat	gagtgataag	tgatccaggt	tattactatg	ttagctaggt	tttttttaca	10200
aggataaatt	atctgtgatc	ataattctta	tgaaagcttt	atgtttcctg	gaggcagtg	10260
catgcaatgc	atgacagcaa	cttgatcaca	ccagctgagg	tagatacggg	aacaagggtc	10320
ttaaatctgt	tcaccaaata	attggagaac	acacatacac	attcttgcca	gtcttggtta	10380
gagaaatttc	atgacaaaat	gccaaagctg	tcttgactct	tcacttttgg	ccatgagtcg	10440
tgacttagtt	tggtttaatg	gaccggttct	cctagcttgt	tctactcaaa	actgttggtg	10500
atgcgaataa	gttggtgatg	ttgatctctg	gattttgttt	tgctctcaat	agtggacgag	10560
attagatagc	ccggaaattt	accgggtgcc	gggcggccag	catggccgta	tcgcgaatgt	10620
gttattaagt	tgtctaagcg	tcaatttggt	tacaccacaa	tatatcctgc	caccagccag	10680

ccaacagctc	cccgaccggc	agctcggcac	aaaatcacca	ctcgatacag	gcagcccato	10740
agaattaatt	ctcatgtttg	acagctttatc	atcgactgca	cggtgcacca	atgctttctgg	10800
cgtcaggcag	ccatcggaag	ctgtgggtatg	gctgtgcagg	tcgtaaatca	ctgcataatt	10860
cgtgtcgcctc	aaggcgcact	cccgtttctgg	ataatgtttt	ttgcgccgac	atcataacgg	10920
ttctggcaaa	tattctgaaa	tgagctgttg	acaattaatc	atccggctcg	tataatgtgt	10980
ggaattgtga	gcggataaca	atttcacaca	ggaaacagac	catgagggaa	gcgttgatcg	11040
ccgaagtatc	gactcaacta	tcagaggtag	ttggcgctcat	cgagcgccat	ctcgaaccga	11100
cgttgctggc	cgtacatttg	tacggctccg	cagtggatgg	cggcctgaag	ccacacagtg	11160
atattgattt	gctggttacg	gtgaccgtaa	ggcttgatga	aacaacgcgg	cgagctttga	11220
tcaacgacct	tttgaaaact	tcggcttccc	ctggagagag	cgagattctc	cgcgctgtag	11280
aagtcaccat	tgttgtgcac	gacgacatca	ttccgtggcg	ttatccagct	aagcgcgaac	11340
tgcaatttgg	agaatggcag	cgcaatgaca	ttcttgccagg	tatcttcgag	ccagccacga	11400
tcgacattga	tctggctatc	ttgctgacaa	aagcaagaga	acatagcggt	gccttggtag	11460
gtccagcggc	ggaggaactc	tttgatccgg	ttctgaaca	ggatctatct	gaggcgctaa	11520
atgaaacctt	aacgctatgg	aactcgccgc	ccgactgggc	tggcgatgag	cgaaatgtag	11580
tgcttacggt	gtcccgcat	tggtacagcg	cagtaaccgg	caaaatcgcg	ccgaaggatg	11640
tcgctgccga	ctggggcaatg	gagcgccctgc	cggcccagta	tcagcccgtc	atacttgaag	11700
ctaggcaggc	ttatcttggg	caagaagatc	gcttggcctc	gcgcgcagat	cagttggaag	11760
aatttgttca	ctacgtgaaa	ggcgagatca	ccaaagtagt	cggcaaataa	agctctagt	11820
gatctccgta	cccggggatc	tggctcgcgg	cggacgcacg	acgccggggc	gagaccatag	11880
gcgatctcct	aaatcaatag	tagctgtaac	ctcgaagcgt	ttcacttgta	acaacgattg	11940
agaatttttg	tcataaaaatt	gaaatacttg	gttcgcattt	ttgtcatccg	cggtcagccg	12000
caattctgac	gaactgcccc	tttagctgga	gatgattgta	catccttcac	gtgaaaattt	12060
ctcaagcgct	gtgaacaagg	gttcagattt	tagattgaaa	ggtgagccgt	tgaaacacgt	12120
tcttcttgtc	gatgacgacg	tcgctatgcg	gcattcttatt	attgaatacc	ttacgatcca	12180
cgccttcaaa	gtgaccgcgg	tagccgacag	caccagttc	acaagagtac	tctcttccgc	12240
gacggctgat	gtcgtggttg	ttgatctaga	tttaggtcgt	gaagatgggc	tcgagatcgt	12300
tcgtaatctg	gcggcaaagt	ctgatattcc	aatcataatt	atcagtggcg	accgccttga	12360
ggagacggat	aaagttgttg	cactcgagct	aggagcaagt	gattttatcg	ctaagccgtt	12420
cagtatcaga	gagtttctag	cacgcattcg	ggltgccttg	cgcgtgcgcc	ccaacgttgt	12480
ccgctccaaa	gaccgacggt	ctttttgttt	tactgactgg	acacttaatc	tcaggcaacg	12540
tcgcttgatg	tccgaagctg	gcggtgaggt	gaaacttacg	gcaggtgagt	tcaatcttct	12600
cctcgcgttt	ttagagaaac	cccgcgacgt	tctatcgcgc	gagcaacttc	tcatttgccag	12660
tcgagtacgc	gacgaggagg	tttatgacag	gagtatagat	gttctcattt	tgaggctgcg	12720
ccgcaaactt	gaggcagatc	cgtcaagccc	tcaactgata	aaaacagcaa	gaggtgccgg	12780
ttatttcttt	gacgcggacg	tgcaggtttc	gcacgggggg	acgatggcag	cctgagccaa	12840

ttcccagatc	cccaggaat	cggcgtgagc	ggtcgcaaac	catccggccc	ggtacaaatc	12900
ggcgcggcgc	tgggtgatga	cctggtggag	aagttgaagg	ccgcgcaggc	cgcccagcgg	12960
caacgcacgc	aggcagaagc	acgccccggc	gaatcgtggc	aagcggccgc	tgatcgaatc	13020
cgcaaagaat	cccggcaacc	gccggcagcc	ggcgcgccgt	cgattaggaa	gccgcccag	13080
ggcgacgagc	aaccagattt	tttcgttccg	atgctctatg	acgtgggcac	ccgcgatagt	13140
cgcagcatca	tggacgtggc	cgttttccgt	ctgtcgaagc	gtgaccgacg	agctggcgag	13200
gtgatccgct	acgagcttcc	agacgggcac	gtagaggttt	ccgcagggcc	ggccggcatg	13260
gccagtgtgt	gggattacga	cctggtactg	atggcggttt	cccatctaac	cgaatccatg	13320
aaccgatacc	gggaagggaa	gggagacaag	cccggccgcg	tgttccgtcc	acacgttgcg	13380
gacgtactca	agttctgccg	gcgagccgat	ggcggaaaagc	agaaagacga	cctggtagaa	13440
acctgcattc	ggttaaacac	cacgcacgtt	gccatgcagc	gtacgaagaa	ggccaagaac	13500
ggccgcctgg	tgacgggtatc	cgaggggtgaa	gccttgatta	gccgctacaa	gatcgtaaag	13560
agcgaaccgc	ggcggccgga	gtacatcgag	atcgagctag	ctgattggat	gtaccgcgag	13620
atcacagaag	gcaagaaccc	ggacgtgctg	acggttcacc	ccgattactt	tttgatcgat	13680
cccggcatcg	gccgttttct	ctaccgcctg	gcacgccgcg	ccgcaggcaa	ggcagaagcc	13740
agatggttgt	tcaagacgat	ctacgaacgc	agtggcagcg	ccggagagtt	caagaagtcc	13800
tgtttcaccg	tgcgcaagct	gatcgggtca	aatgacctgc	cggagtlacga	tttgaaggag	13860
gaggcggggc	aggctggccc	gacccatgct	atgcgctacc	gcaacctgat	cgagggcgaa	13920
gcatccgccg	gttcctaattg	tacggagcag	atgctagggc	aaattgccct	agcaggggaa	13980
aaaggtcgaa	aaggtctctt	tcctgtggat	agcacgtaca	ttgggaaccc	aaagccgtac	14040
attgggaacc	ggaacccgta	cattgggaac	ccaaagccgt	acattgggaa	ccggtcacac	14100
atgtaagtga	ctgatataaa	agagaaaaaa	ggcgattttt	ccgcctaaaa	ctctttaaaa	14160
cttattaaaa	ctcttaaaac	ccgcctggcc	tgtgcataac	tgtctggcca	gcgcacagcc	14220
gaagagctgc	aaaaagcgcc	tacccttcgg	tcgctgcgct	ccctacgccc	cgccgcttcg	14280
cgtcggccta	tcgcggccgc	tggccgctca	aaaatggctg	gcctacggcc	aggcaatcta	14340
ccagggcgcg	gacaagccgc	gccgtcgcca	ctcgaccgcc	ggcgtgagg	tctgcctcgt	14400
gaagaaggctg	ttgctgactc	ataccaggcc	tgaatcgccc	catcatccag	ccagaaagtg	14460
aggagagccac	ggttgatgag	agctttgttg	taggtggacc	agttggtgat	tttgaacttt	14520
tgctttgcca	cggaacggtc	tcggttgctg	ggaagatgcg	tgatctgato	cttcaactca	14580
gcaaaagtcc	gatttattca	acaaagccgc	cgtcccgtca	agtcagcgta	atgctctgcc	14640
agtgttacia	ccaattaacc	aattctgatt	agaaaaactc	atcgagcatc	aaatgaaact	14700
gcaatttatt	catatcagga	ttatcaatac	catatttttg	aaaaagccgt	ttctgtaatg	14760
aaggagaaaa	ctcaccgagg	cagttccata	ggatggcaag	atcctgggat	cggtctgcga	14820
ttccgactcg	tccaacatca	atacaacctc	ttaattttcc	ctcgtcaaaa	ataaggttat	14880
caagtgagaa	atcaccatga	gtgacgactg	aatccggtga	gaatggcaaa	agctctgcat	14940
taatgaatcg	gccaacgcgc	ggggagaggc	ggtttgcgta	ttggggcgctc	ttccgcttcc	15000

tcgctcactg actcgctgcg ctcggtcggtt cggctgcggc gagcgggtatc agctcactca 15060
 aaggcggtaa tacgggttatc cacagaatca ggggataacg caggaaagaa catgtgagca 15120
 aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcggtt tttccatagg 15180
 ctccgcccc ctgacgagca tcacaaaaat cgacgctcaa gtcagaggtg gcgaaacccg 15240
 acaggactat aaagatacca ggcgtttccc cctggaagct ccctcgtagc ctctcctggt 15300
 ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttcggaag cgtggcgctt 15360
 tctcatagct cagcgtgtag gtatctcagt tcgggtgtagg tcgttcgctc caagctgggc 15420
 tgtgtgcacg aacccccgt tcagcccgac cgctgcgcct tatccggtaa ctatcgtctt 15480
 gagtccaacc cggtaagaca cgacttatcg ccaactggcag cagccactgg taacaggatt 15540
 agcagagcga ggtatgtagg cgggtgtaca gagttcttga agtgggtggc taactacggc 15600
 tacactagaa gaacagtatt tgggtatctgc gctctgtga agccagttac cttcggaaaa 15660
 agagttggta gctcttgatc cggcaaacaa accaccgctg gtagcgggtg tttttttgtt 15720
 tgcaagcagc agattacgcg cagaaaaaaaa ggatctcaag aagatccttt gatcttttct 15780
 acgggggtctg acgctcagtg gaacgaaaac tcacgttaag ggatttttgt catgagatta 15840
 tcaaaaagga tcttcaccta gatccttttg atccggaatt a 15881

<210> 24

<211> 15436

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vector tái tổ hợp

<400> 24

attcctgtgg ttggcatgca catacaaatg gacgaacgga taaacctttt cagccctttt 60
 taaatatccg attattctaa taaacgctct tttctcttag gtttaccgc caatatatcc 120
 tgtcaaacac tgatagttaa aactgaaggc gggaaacgac aatctgatca tgagcggaga 180
 attaagggag tcacgttatg accccgcgcg atgacgcggg acaagccgtt ttacgtttgg 240
 aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccattttaa tcaattgggc 300
 gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc 360
 taccaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg 420
 gaacaaaata acgtggaaaa gagctgtcct gacagccac tcactaatgc gtatgacgaa 480
 cgcagtgacg accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct 540
 cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaaggtgg 600
 ctctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga 660
 cagtggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc 720
 aaccacgtct tcaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga 780

acaatcccac	tatccttcgg	taccggaccc	aatgaatta	atatattatc	taaattcaga	840
cgggaaaaag	agagaaggaa	tcaggaaggc	tgagtgcaat	attagttttt	cattgagcac	900
aatcttatta	tgtttgatga	acatcttgaa	caaagtatga	gtggaaaaca	acacttatgt	960
tatatctgca	gcacagattg	ttagaagtga	aggctacctt	gagatgggtga	taggtctctg	1020
aacccatctg	cattgcttct	tcgaagttca	tagcaccaac	aagaagaatc	ataatttctt	1080
acaaataaga	aatgttttata	gttactccaa	aactacagtt	aaacatattg	atatggttga	1140
tatacatgaa	caactaactg	tcttaatatg	tttctgttta	ggcaaaagat	gacacaagtg	1200
gacaaagtaa	gggaataatg	cattcaactt	ctcagactct	aaaggtttgg	atgaataactt	1260
gaatgggaag	accattttcca	gcatgctttc	caccattaat	gactgtattt	gcagggacag	1320
gaagagttgt	agcgcttttg	ccaacaagat	ccgctatatg	cttgtagagt	ggaacctgca	1380
cagcaacaag	catccaatga	tctgagaata	agaattatgt	gcggtggcat	gacacattta	1440
ccaaattatg	atgatttttag	cagaagagca	gcacctcttt	ttcagcagca	ccagctttgc	1500
aagctgcaat	tgacactgcc	agcatagtat	ttgctccaag	ctcagcctgc	atttcgcaca	1560
atatcatatc	tattttccatt	tggataaaaat	ctatttgcca	gaaatatggg	aacggaatca	1620
agatacagaa	aaaagttgag	gctgtagtgg	taagtgggaa	cacaagtaca	gtgaaaagat	1680
agcattattt	tccgaacagt	caaattaacc	gaaataagta	gtatcacata	ctgaccctaa	1740
attgactaca	cccaatgtaa	agaattgctt	gtgcatattg	agccatcagt	tgtttgtttg	1800
tgtaatggat	ttgacaatgc	agaattcagg	agcaaaaaaa	accgcaaagg	agagagaggg	1860
aaaggatctc	cgagtatccc	acctgtggcg	cgccttgcca	agtcctatgat	ggcctggctg	1920
atctgcgcct	gctgctgcgg	atccatccca	acgagtgcct	ccgacacctt	gtcgttgatc	1980
acccgcaccg	cgtaggcgac	ccccctggcg	aggaacttcc	gcctctcgga	ggccccggcc	2040
gcgtcggcgg	cggcgcccac	gagcgcgccc	acgctagccg	tagatgctcg	gtgcacggcc	2100
ttgttggtgt	gcaactcgac	ctcaacggcc	ggtgcgctac	gcccgtcaag	gatctggcgc	2160
gcccgcaccc	tcgtgatcac	ggtgggcgcc	tccgcctca	tgtggccgac	gatgaaaagc	2220
gccgggtcag	gggccttggc	gcggagcgcc	gtattcaggg	cctcctcgct	cttgccgcgag	2280
agcaggtgct	tccccaggta	ctcctgcact	gacatggctg	ccacagcggg	ggaattttct	2340
tgtcctcacg	acttcgccac	gtcgtcctca	cctccctggc	ttcccacgac	tccgcgctat	2400
cgaggtgctc	ggtgggtcgc	cttaatccga	agtccgaact	aggaagacga	gacagagagg	2460
tctgaggaat	gggcctcggt	tgatgttgag	ctgaattatt	tttcagcat	aagcccaggt	2520
cttgggtccat	gaacaaaatt	actagaaacc	cagcccagta	ctacgatcta	aaagagggag	2580
tgccactgga	tagctctctc	tagcattctc	cacgctccaa	tacagcggcg	tagggctctat	2640
ccgggtctat	ccgcgaacac	gtgagaactc	tccagaaact	gctttctcct	ccacttcctc	2700
tctctcgctt	tccctctata	aaaagacccc	ttctaggaat	tgagggagac	agcaagcagc	2760
gatccgaagc	tcaatcaatt	cactcaaacc	tcttccccaa	atcttcgatt	agattctcgt	2820
tgacaagaag	actataaccg	aacctgaccg	taaaccatgg	acaacaaccc	caacatcaac	2880
gagtgcaccc	cctacaactg	cctgagcaac	cccaggtggg	aggtgctggg	cggcgagcgc	2940

atcgagaccg gctacacccc catcgacatc agcctgagcc tgacccagtt cctgctgagc	3000
gagttcgtgc ccggcgccgg cttcgtgctg ggccctggtg acatcatctg gggcatcttc	3060
ggccccagcc agtgggacgc cttcctggtg cagatcgagc agctgatcaa ccagcgcac	3120
gaggagttcg cccgcaacca ggccatcagc cgcctggagg gcctgagcaa cctgtaccaa	3180
atctacgccg agagcttccg cgagtgggag gccgacccca ccaaccccg cctgcgcgag	3240
gagatgcgca tccagttcaa cgacatgaac agcgccctga ccaccgccat ccccctgttc	3300
gccgtgcaga actaccaggt gccctgctg agcgtgtacg tgcaggccgc caacctgcac	3360
ctgagcgtgc tgcgcgacgt cagcgtgttc ggccagcgt ggggcttcga cgccgccacc	3420
atcaacagcc gctacaacga cctgacccgc ctgatcggca actacaccga ccacgccgtg	3480
cgctggtaca acaccggcct ggagcgcgtg tggggtccag acagccgcga ctggatcagg	3540
tacaaccagt tccgcccga gctgacctg accgtgctgg acatcgtgag cctgttcccc	3600
aactacgaca gccgcaccta ccccatccgc accgtgagcc agctgacccg cgagatttac	3660
accaaccccg tgctggagaa cttcgacggc agcttccgcg gcagcgccca gggcatcgag	3720
ggcagcatcc gcagccccc cctgatggac atcctgaaca gcatcaccat ctacaccgac	3780
gccaccgcg gcgagtacta ctggagcggc caccagatca tggccagccc cgtcggcttc	3840
agcggccccc agttcacctt ccccctgtac ggcactatgg gcaacgctgc acctcagcag	3900
cgcacgtggtg cacagctggg ccagggagtg taccgcaccc tgagcagcac cctgtaccgt	3960
cgacctttca acatcggcat caacaaccag cagctgagcg tgctggacgg caccgagttc	4020
gcctacggca ccagcagcaa cctgcccagc gccgtgtacc gcaagagcgg caccgtggac	4080
agcctggacg agatccccc tcagaacaac aacgtgccac ctcgacaggg cttcagccac	4140
cgtctgagcc acgtgagcat gttccgcagt ggcttcagca acagcagcgt gagcatcacc	4200
cgtgcacctt tgttcagctg gattcacccg agtgccgagt tcaacaacat catccccagc	4260
agccagatca cccagatccc cctgaccaag agcaccaacc tgggcagcgg caccagcgtg	4320
gtgaagggcc ccggcttcac cggcggcgac atcctgcgcc gcaccagccc cggccagatc	4380
agcacccctg gcgtgaacat caccgcccc ctgagccagc gctaccgct cgcacccgc	4440
tacgccagca ccaccaacct gcagttccac accagcatcg acggccgccc catcaaccag	4500
ggcaacttca gcgccaccat gagcagcggc agcaacctgc agagcggcag cttccgcacc	4560
gtgggcttca ccacccctt caacttcagc aaoggcagca gcgtgttcac cctgagcgcc	4620
cacgtgttca acagcggcaa cgaggtgtac atcgaccgca tcgagttcgt gcccgccgag	4680
gtgaccttcg aggcgcagta cgacctggag agggctcaga aggcctgaa cgagctgttc	4740
accagcagca accagatcgg cctgaagacc gacgtgaccg actaccacat cgatcagggtg	4800
taggagctca cctcggctctg cggctcgtcgt acctgcgtgg tttgaggaac ggcagttcgc	4860
ctcggctcgtt ctgtgaaata aaattgggtt acaagaatta tggcgtttgt caatatggtc	4920
gtaatgtcgt aggatgggtg aatgtggtca caaactttgc gtatgttggg tctactgggtg	4980
gtgtctgaat ctatgtatgg atgtcatgag tttgtctact agagttctag ttccctgtggt	5040
gttcgtatga tgtatatcat gagttaactc tatctaaaat ttcttcatta tggtatTTTT	5100

aagaaataaa	ttctttatta	tggttttttt	aaagggacta	tgtccatcct	aacaagagca	5160
acgtccaata	attctctaaa	acttaaattt	aagaaccgaa	taaaaaatga	gtagttttta	5220
aatactttct	atctaacctt	atctactctt	ctccatattt	tagtagtttt	ttaaatagac	5280
ttactaaatt	tagttactct	atatttttgg	aactcgacag	aatacaatct	gtgattaatt	5340
tttctcacat	gtggtagcta	gatacgacac	tttttttttt	acttttttag	acatgcacaa	5400
tggagtagtt	agattcaaca	cttatgacac	aagttttttc	ttcaccgttg	gactattatt	5460
gtcgcatatg	ttgccaccgc	gaaccaccgc	tcctctatag	gttatgtcaa	tcgaccttcc	5520
ctcttttgtc	ttcgacatca	tccaagacta	acaaagttag	attcaaatac	aactagatgt	5580
ccagactggc	aaagtttagat	ttaaaatcaa	actactatct	acataataag	aactattagg	5640
gactaagtta	tttttctact	tcaatagttg	ttgctagcaa	cttgctaaac	ataatttttag	5700
aaaacttttt	tagagaacta	ttagagttgc	tctaaaagggt	tttgtagtcc	cattttactgt	5760
tttgtgacta	ttattaggtg	gagtttgggc	agtaaaaagg	ccaaatagaa	aatgggggtg	5820
agattgggaa	aacggaccgc	gatcgcttaa	ttaagcttgc	atgcctgcag	tcgacgctga	5880
cccggtcgtg	cccctctcta	gagataatga	gcattgcatg	tctaagttat	aaaaaattac	5940
cacatatttt	ttttgtcaca	cttgtttgaa	gtgcagttta	tctatcttta	tacatatatt	6000
taaacttttac	tctacgaata	atataatcta	tagtactaca	ataatatcag	tgttttagag	6060
aatcatataa	atgaacagtt	agacatggtc	taaaggacaa	ttgagtattt	tgacaacagg	6120
actctacagt	tttatctttt	tagtgtgcat	gtgttctcct	ttttttttgc	aaatagcttc	6180
acctatataa	tacttcatcc	attttattag	tacatccatt	tagggtttag	ggttaatggt	6240
ttttatagac	taattttttt	agtacatcta	ttttattcta	tttttagcctc	taaattaaga	6300
aaactaaaac	tctatttttag	tttttttatt	taataattta	gatataaaat	agaataaaat	6360
aaagtgacta	aaaattaaac	aaataccctt	taagaaatta	aaaaaactaa	ggaaacattt	6420
ttcttgtttc	gagtagataa	tgccagcctg	ttaaacgccg	tcgacgagtc	taacggacac	6480
caaccagcga	accagcagcg	tcgcgtcggg	ccaagcgaag	cagacggcac	ggcatctctg	6540
tcgctgcctc	tggacccttc	tcgagagttc	cgctccaccg	ttggacttgc	tcogctgtcg	6600
gcatccagaa	attgcgtggc	ggagcggcag	acgtgagccg	gcacggcagg	cggcctcctc	6660
ctcctctcac	ggcaccggca	gctacggggg	attccttttc	caccgctcct	tcgctttccc	6720
ttcctcgccc	gccgtaataa	atagacaccc	cctccacacc	ctctttcccc	aacctcggtg	6780
tgttcggagc	gcacacacac	acaaccagat	ctccccaaa	tcaccccgtc	ggcacctcgg	6840
cttcaaggta	cgcgcgtcgt	cctccccccc	ccccctctc	taccttctct	agatcggcgt	6900
tccggtccat	ggtttagggc	cggtagttct	acttctgttc	atgtttgtgt	tagatccgtg	6960
tttgtgttag	atccgtgctg	ctagcgttcg	tacacggatg	cgacctgtac	gtcagacacg	7020
ttctgattgc	taacttgcca	gtgtttctct	ttgggggaatc	ctgggatggc	tctagccgtt	7080
cgcgagacgg	gatcgatttc	atgatttttt	ttgtttcggt	gcataggggt	tggtttgccc	7140
ttttccttta	tttcaatata	tgccgtgcac	ttgtttgtcg	ggtcatcttt	tcattgctttt	7200
ttttgtcttg	gttgtgatga	tgtggtctgg	ttgggcgggtc	gttctagatc	ggagtagaat	7260

tctgtttcaa	actacctggt	ggatttatta	atthttggatc	tgtatgtgtg	tgccatacat	7320
attcatagtt	acgaattgaa	gatgatggat	ggaaatatcg	atctaggata	ggtatacatg	7380
ttgatgcggg	ttttactgat	gcatatacag	agatgctttt	tgttcgcttg	gttgtgatga	7440
tgtggtgtgg	ttgggcggtc	gttcattcgt	tctagatcgg	agtagaatac	tgthttcaaac	7500
tacctggtgt	atthtattaat	tttggaactg	tatgtgtgtg	tcatacatct	tcatagttac	7560
gagthttaaga	tggaatggaaa	tatcgatcta	ggataggat	acatgttgat	gtgggtthtta	7620
ctgatgcata	tacatgatgg	catatgcagc	atctattcat	atgctctaac	cttgagtacc	7680
tatctattat	aataaacaag	tatgtthttat	aattatthttg	atcttgatat	acttggaatga	7740
tgccatatgc	agcagctata	tgtggatthtt	tttagccctg	ccttcatacg	ctatthtattt	7800
gcttggtact	gtthtctthtg	tcgatgctca	ccctgttggt	tgggtgttact	tctgcaggga	7860
tccccgatca	tgcaaaaact	cattaactca	gtgcaaaaact	atgcctgggg	cagcaaaaacg	7920
gcgttgactg	aactthtatgg	tatggaaaat	ccgtccagcc	agccgatggc	cgagctgtgg	7980
atgggcgcac	atccgaaaag	cagthtcacga	gtgcagaatg	ccgccggaga	tatcgthttca	8040
ctgcgtgatg	tgattgagag	tgataaatcg	actctgctcg	gagaggccgt	tgccaaaacgc	8100
thttggcgaac	tgccthttcct	gttcaaagta	ttatgcgcag	cacagccact	ctccattcag	8160
gttcatccaa	acaaacacaa	ttctgaaatc	ggtthttgcc	aagaaaatgc	cgcaggatct	8220
ccgatggatg	ccgccgagcg	taactataaa	gatcctaacc	acaagccgga	gctgggtthtt	8280
gcgctgacgc	ctthtcttgc	gatgaacgcg	thtctgtgaat	thtccgagat	tgtctcccta	8340
ctccagccgg	tcgcagggtgc	acatccggcg	attgctcact	thttacaaca	gcctgatgcc	8400
gaacgtthta	gcgaactgtt	cgccagcctg	ttgaatatgc	aggggtgaaga	aaaatcccg	8460
gcgctggcga	thttaaaatc	ggccctcgat	agccagcagg	gtgaaccgtg	gcaaacgatt	8520
cgtthtaattt	ctgaattthta	cccggaagac	agcggtctgt	tctccccgct	attgctgaat	8580
gtgggtgaaat	tgaaccttg	cgaagcgatg	thtctgttcg	ctgaaacacc	gcacgcttac	8640
ctgcaaggcg	tggcgctgga	agtgatggca	aactccgata	acgtgctgcg	tgcgggtctg	8700
acgcctaaat	acattgatat	tcgggaactg	gttgccaatg	tgaaattcga	agccaaaccg	8760
gctaaccagt	tgthtgacca	gccgggtgaa	caagggtgcag	aactggactt	cccgattcca	8820
gtggatgatt	thgcctthctc	gctgcatgac	cttagtgata	aagaaaccac	cattagccag	8880
cagagtgcgg	ccatthttgtt	ctgogtcgaa	ggcgatgcaa	cgttgtggaa	aggtthctcag	8940
cagttacagc	thaaacccgg	tgaatcagcg	thtattgcgg	ccaacgaatc	accgggtgact	9000
gtcaaaggcc	acggccgttht	agcgctgtt	tacaacaagc	tgtaagagct	tactgaaaaa	9060
attaacatct	cttgctaagc	tgggagctcg	tcattgggtcg	thtaagctgc	cgatgtgcct	9120
gcgtcgtctg	gtgcctctctc	tcataatgga	ggttgtcaaa	gtatctgctg	thcgtgtcat	9180
gagtcgtgtc	agtgttggtt	taataatgga	ccggttgtgt	tgtgtgtgcg	tactaccag	9240
aactatgaca	aatcatgaat	aagthttgatg	thtgaaatta	aagcctgtgc	tcattatgtt	9300
ctgtctthta	gttgtctctc	aatathttgcc	tgccaggtact	ggctatctac	cgtthtcttac	9360
thaggagggtg	thtgaaatgca	ctaaaactaa	tagthtagtgg	ctaaaattag	thaaaacatc	9420

caaacacccat agctaatagt tgaactatta gctatTTTTg gaaaattagt taatagttag	9480
gtagttatTTt gtagctagc taattcaact aacaatTTTT agccaactaa caattagttt	9540
cagtgcattc aaacaccccc ttaatgttaa cgtggttcta tctaccgtct cctaatatat	9600
ggttgattgt tCGgttTgtt gctatgctat tgggttctga ttgctgctag ttcttgctga	9660
atccagaagt tctcgtagta tagctcagat tcatattatt tatttgagtg ataagtgatc	9720
caggttatta ctatgttagc taggtTTTTt ttacaaggat aaattatctg tgatcataat	9780
tcttatgaaa gctttatgtt tcctggaggc agtggcatgc aatgcatgac agcaacttga	9840
tcacaccagc tgaggtagat acggtaacaa ggttcttaaa tctgttcacc aaatcattgg	9900
agaacacaca tacacattct tgccagtctt ggtagagaa atttcatgac aaaatgccaa	9960
agctgtcttg actcttcact tttggccatg agtcgtgact tagtttggtt taatggaccg	10020
gttctcctag ctgtttctac tcaaaactgt tgttgatgcg aataagttgt gatggttgat	10080
ctctggattt tgttttgctc tcaatagtgg acgagattag atagcccgga aatttacogg	10140
tgcccgggcg gccagcatgg cegtatccgc aatgtgttat taagttgtct aagcgtcaat	10200
ttgtttacac cacaatatat cctgccacca gccagccaac agctccccga ccggcagctc	10260
ggcacaaaat caccactcga tacaggcagc ccatcagaat taattctcat gtttgacagc	10320
ttatcatcga ctgcacggtg caccaatgct tctggcgtca ggcagccatc ggaagctgtg	10380
gtatggctgt gcaggtcgta aatcaactga taattcgtgt cgtcaaggc gcaactccgt	10440
tctggataat gttttttgcg ccgacatcat aacggttctg gcaaatatc tgaaatgagc	10500
tgttgacaat taatcatccg gctcgtataa tgtgtggaat tgtgagcggg taacaatttc	10560
acacaggaaa cagaccatga ggggaagcgtt gatcgccgaa gtatcgactc aactatcaga	10620
ggtagttggc gtcacgagc gccatctcga accgacgttg ctggccgtac atttgtagcg	10680
ctcgcagtg gatggcggcc tgaagccaca cagtgatatt gatttgctgg ttacggtgac	10740
cgtaaggctt gatgaaacaa cgcggcgagc tttgatcaac gaccttttg aaacttcggc	10800
ttcccttga gagagcgaga ttctccgcgc tgtagaagtc accattgttg tgcacgacga	10860
catcatccg tggcgttatc cagctaagcg cgaactgcaa tttggagaat ggcagcgcaa	10920
tgacattctt gcaggatatc tcgagccagc cagcatcgac attgatctgg ctatcttgct	10980
gacaaaagca agagaacata gcgttgctt ggtaggtcca gcggcgagg aactctttga	11040
tccggttctt gaacaggatc tatttgaggc gctaaatgaa accttaacgc tatggaactc	11100
gccgcccagc tgggctggcg atgagcgaaa tgtagtgtt acgttgctcc gcatttggtg	11160
cagcgcagta accggcaaaa tcgcgccgaa ggatgtcgtt gccgactggg caatggagcg	11220
cctgccggcc cagtatcagc ccgtcatact tgaagctagg caggcttata ttggacaaga	11280
agatcgcttg gcctcgcgcg cagatcagtt ggaagaattt gttcactacg tgaaaggcga	11340
gatcaccaaa gtagtcggca aataaagctc tagtggatct ccgtaccg ggtatctggct	11400
cgcggcggac gcacgacgc ggggcgagac cataggcgat ctctaaatc aatagtagct	11460
gtaacctcga agcgtttcac ttgtaacaac gattgagaat ttttgtcata aaattgaaat	11520
aottggttcg catttttgct atccgcggtc agccgcgaatt ctgacgaact gccatttag	11580

ctggagatga ttgtacatcc ttcacgtgaa aattttctcaa gcgctgtgaa caagggttca	11640
gatttttagat tgaaaggtga gccgttgaaa caggttcttc ttgtcgatga cgacgtcgct	11700
atgcggcatc ttattattga ataccttacg atccacgcct tcaaagtga cgcggtagcc	11760
gacagcacc agttcacaag agtactctct tccgcgacgg tcgatgtcgt ggttggtgat	11820
ctagatttag gtctgaaga tgggctcgag atcgttcgta atctggcggc aaagtctgat	11880
attccaatca taattatcag tggcgaccgc cttgaggaga cggataaagt tgttgcactc	11940
gagctaggag caagtgattt tatcgctaag ccgttcagta tcagagagtt tctagcacgc	12000
attcggggtt ccttgccggt gcccccaac gttgtccgct ccaaagaccg acggtctttt	12060
tgttttactg actggacact taatctcagg caacgtcgct tgatgtccga agctggcgg	12120
gaggtgaaac ttacggcagg tgagttcaat cttctcctcg cgtttttaga gaaacccgc	12180
gacgttctat cgcgcgagca acttctcatt gccagtcgag tacgcgacga ggaggtttat	12240
gacaggagta tagatgttct ctttttgagg ctgcgccga aacttgaggc agatccgtca	12300
agccctcaac tgataaaaac agcaagaggt gccggttatt tctttgacgc ggacgtgcag	12360
gtttcgcacg gggggacgat ggcagcctga gccaatcccg agatccccga ggaatcggcg	12420
tgagcggctg caaaccatcc ggcccggtac aaatcggcgc ggcgctgggt gatgacctgg	12480
tggagaagtt gaaggccgcg caggccgcc agcggcaacg catcgaggca gaagcacgcc	12540
ccggtgaatc gtggcaagcg gccgctgac gaatccgcaa agaatcccgg caaccgcgg	12600
cagccggtgc gccgtcgatt aggaagccgc ccaagggcga cgagcaacca gattttttcg	12660
ttccgatgct ctatgacgtg ggcacccgcg atagtcgcag catcatggac gtggccggtt	12720
tccgtctgtc gaagcgtgac cgacgagctg gcgaggtgat ccgctacgag cttccagacg	12780
ggcacgtaga ggtttccgca gggccggccg gcatggccag tgtgtgggat tacgacctgg	12840
tactgatggc ggtttcccat ctaaccgaat ccatgaaccg ataccgggaa gggaaggag	12900
acaagcccgg ccgctgttcc cgtccacacg ttgcggacgt actcaagtcc tgccggcgag	12960
ccgatggcgg aaagcagaaa gacgacctgg tagaaacctg cattcggtta aacaccacgc	13020
acgttgccat gcagcgtacg aagaaggcca agaacggccg cctggtgacg gtatccgagg	13080
gtgaagcctt gattagccgc tacaagatcg taaagagcga aaccgggccc ccggagtaca	13140
tcgagatcga gctagctgat tggatgtacc gcgagatcac agaaggcaag aaccgggacg	13200
tgctgacggg tcaccccgat tactttttga tcgatcccg catcggcgct tttctctacc	13260
gcctggcacg ccgcgccgca ggcaaggcag aagccagatg gttgttcaag acgatctacg	13320
aacgcagtgg cagcgccgga gagttcaaga agttctgttt caccgtgcgc aagctgatcg	13380
ggtcaaata cctgccggag tacgatttga aggaggaggc ggggcaggct ggcccgatcc	13440
tagtcatgcg ctaccgcaac ctgatcgagg gcgaagcatc cgcgggttcc taatgtacgg	13500
agcagatgct agggcaaatt gccctagcag gggaaaaagg tcgaaaagggt ctctttctcg	13560
tggatagcac gtacattggg aacccaaagc cgtacattgg gaaccggaac ccgtacattg	13620
ggaacccaaa gccgtacatt gggaaccggg cacacatgta agtgactgat ataaaagaga	13680
aaaaaggcga tttttccgcc taaaactctt taaaacttat taaaactctt aaaacccgcc	13740

tggcctgtgc ataactgtct ggccagcgca cagccgaaga gctgcaaaaa gcgcctaccc 13800
 ttcggctcgct gcgctcccta cgccccgcgc ctctcgctcg gcctatcgcg gccgctggcc 13860
 gctcaaaaat ggctggccta cggccaggca atctaccagg gcgcggacaa gccgcgccgt 13920
 cgccactcga ccgcccggcg tgaggtctgc ctctgaaga aggtgttgct gactcatacc 13980
 aggctgaat cgccccatca tccagccaga aagtgaggga gccacggttg atgagagctt 14040
 tgttgtaggt ggaccagttg gtgattttga acttttgctt tgccacggaa cggctctcgct 14100
 tgtcgggaag atgcgtgatc tgatccttca actcagcaaa agttcgattt attcaacaaa 14160
 gccgcgcgtc cgtcaagtca gcgtaatgct ctgccagtgt tacaaccaat taaccaattc 14220
 tgattagaaa aactcatcga gcatcaaatg aaactgcaat ttattcatat caggattatc 14280
 aataccatat ttttgaaaaa gccgtttctg taatgaagga gaaaactcac cgaggcagtt 14340
 ccataggatg gcaagatcct ggtatcggtc tgcgattccg actcgtccaa catcaataca 14400
 acctattaat ttcccctcgt caaaaataag gttatcaagt gagaaatcac catgagtgac 14460
 gactgaatcc ggtgagaatg gcaaaagctc tgcattaatg aatcggccaa cgcgcgggga 14520
 gaggcgggtt gcgtattggg cgtctcttcg ctctctcgt cactgactcg ctgcgctcgg 14580
 tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag 14640
 aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc 14700
 gtaaaaaggc cgcgttgctg gcgtttttcc ataggctcgg ccccctgac gagcatcaca 14760
 aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt 14820
 ttcccctcgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc 14880
 tgtccgcctt tctcccttcg ggaagcgtgg cgttttctca tagctcacgc tgtaggtatc 14940
 tcagttcggg taggtcgtt cgtcccaagc tgggctgtgt gcacgaacct cccgttcagc 15000
 ccgaccgctg cgccttatec ggtaactatc gtcttgagtc caaccggta agacacgact 15060
 tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggatg gtaggcgggtg 15120
 ctacagagtt cttgaagtgg tggcctaact acggctacac tagaagaaca gtatttggtg 15180
 totgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca 15240
 aacaaaccac cgtcggtagc ggtgggtttt ttgtttgcaa gcagcagatt acgcgcagaa 15300
 aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaaacg 15360
 aaaactcacg ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc 15420
 ttttgatcgg gaatta 15436

<210> 25

<211> 15884

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vector tái tổ hợp

<400> 25

attcctgtgg ttggcatgca catacaaattg gacgaacgga taaacctttt cacgcccttt	60
taaatatccg attattctaa taaacgctct tttctcttag gtttaccgc caatatatcc	120
tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga	180
attaagggag tcacgttatg accccgcgcg atgacgcggg acaagccgtt ttacgtttgg	240
aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccattttaa tcaattgggc	300
gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc	360
tacaaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg	420
gaacaaaata acgtggaaaa gagctgtcct gacagccac tcactaatgc gtatgacgaa	480
cgcagtgcag accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct	540
cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaaggtgg	600
ctcctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga	660
cagtgggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc	720
aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga	780
acaatcccac tatccttcgg taccggaccc tagtcaccac accatgatca tcttgtttgt	840
tgtagaatac ccacttgcta cctacaacaa tttggtttgg acgtggaact aaatgcaaga	900
ccttattcct cgtgaagttg ttgagcacct cttgcattgc caccaccag tctggatctc	960
taagtgcac ctcctatcatg tatggctcaa tagaagacac acaagagtaa tgttcacaaa	1020
aatgagcaac ttgagagcga gtgattaacc ccttttgaat gtcaccaagg atggagttga	1080
cggggtgata tcgctgaatt gcttggtgaa ctcttgagtg aggtggctct tgatcttgaa	1140
tctcttggtc atcctccttt tcttgaacaa cttcatctcc cccttgattg attogttcct	1200
cttgaggtgg ctcatgtct tgaacttgat cttctctctc ttcttgagct ggttctcat	1260
cttaagttgg tggagatggt tgtatggaat atgatagttg atcatgtgct tgtgaaggct	1320
cttcgggttt ttgtggacac atgtctccaa tggacatggt tcttagcgcg atgcatggag	1380
cctctttatc atctaaatca tcaagatcaa cttgctctag ttgagagcct ctagtggaca	1440
cacatcttcg ggttaaatgg ttttatggaa tattttttct tcacagatac aaatagtatc	1500
gaatatttca gatatttcat ggttttgcg aatacaata taaaatcgga tagagaaaac	1560
taaatttaat tatatccatt tccatccata ttaaaattga atacggatat agatatccat	1620
attagcattt tatttgaata cgaatatata taatttggat gtctagacat tcgaatccat	1680
ctctaattag tgtgggaatg agggacactg aaaaacaatg acgtgcatgg tgacatcata	1740
caatagtaca attctgacga cgatgaagga atttacgtgc ggatcagoga caccacctgg	1800
tttgggtggt cctgtcgccg gaggcgcaag aaataaagag ggcacataac tacgtgaaat	1860
caagcccaat tagtctgcct tggctccctt atgctgtttt aaaaagttaa gggttaagtc	1920
attagtagac tgttgcgggt ttagaaattt ttagagaatt ttgtttacaa cagcccctaa	1980
actaaagttt ttgggaacaa attttagtag tcttttaagt tgctctaaga ctatattttt	2040
ttagttgagg aggacagtga caattttgga gttgctctaa aaccatgttt ttagtcgag	2100

tgggacagtg gcaattactt aactacaatg cacaacacca ggaatccagg atgaaaaatt	2160
actacaccga gggctagttt gggaacctca ttttcccaag agattttcat tttcccaaag	2220
aaaattagtt tattttttcct tgggaaaata gaaatccctt ggaaaattgg agttttccaaa	2280
ctagccttga ttttttttcc taagatatgt gcagatcttt ctttgagagg acacaaaaaa	2340
aatggattgg gattgggctc atcgaaggcc gaatatctct atccatcgtt cgtgccggat	2400
taggcccaga accagaaaag agctaggccg ggctgcagta gcactatcct ctgtttcaca	2460
atattatgca cttttgatca ctttatttat gtcaaaaata cttgatacat caggttttat	2520
tttcactttc gcctctataa aagtattaag ggattttctag acaattcaaa atgtaatttt	2580
ataagactat gtttgtcatt ttataaaaaa aaatagtttg attatlttgg tgaacgtgcc	2640
ttggtcaaaa tttgtgggaa cggaggatgc tatcaaattc gtctgcgcag atgtacgcc	2700
agtaacgaag tatcgtcaat cgaactgatga ccccgtcacc gtcagcaaga cagcaactca	2760
acattcaaat tcgaccgtaa taacatccac atacatacac ggagtatcaa tctagactag	2820
aggagacggg tgaacgtggt gagcctccgc ccataatgca accctactag tgetagcttt	2880
cggccgcgaa aaaacgtccc ccccccaag tctcaacttt atagccgcc cctccccacc	2940
gccgcggccg ccacgcgcag cagcaaccgc gtagcaggag cgcagccagc aagctcaggc	3000
ccccagccct actgccaccg cgcggaacga caaggccgag ccggcggagc accgtccagc	3060
tgaggaggag gaggccgcgg cggccggcga ggatgaggac accggcgccc aggtcgcgcc	3120
catcgtgaag ctggaggagg tgcgcgttac cactggagag gaggacgagg acgcgcctct	3180
ggacttgtga ggcateccgc gcgcgttcga tttccccccc tccccgatec gatttgccca	3240
tgtcttgttg atctgatgtg cggcggctgt gcaggaaggc gaagctctac cggtaaacca	3300
tggacaacaa cccaacatc aacgagtgc tccccacaa ctgcctgagc aaccocgagg	3360
tggaggtgct gggcggcgag cgcategaga ccggctacac ccccatcgac atcagcctga	3420
gcctgaccca gtctcctgctg agcaggttcg tgcccggcgc cggcttcgtg ctgggcctgg	3480
tggacatcat ctggggcctc ttccggccca gccagtggga cgccttcctg gtgcagatcg	3540
agcagctgat caaccagcgc atcgaggagt tcgccgcaa ccaggccatc agccgcctgg	3600
agggcctgag caacctgtac caaatctacg ccgagagctt ccgcgagtgg gaggccgacc	3660
ccaccaaccc cgcctgcgc gaggagatgc gcatccagtt caacgacatg aacagcgccc	3720
tgaccacccg catccccctg ttccgcgtgc agaactacca ggtgcccctg ctgagcgtgt	3780
acgtgcaggc cgcacaacctg cacctgagcg tgetgcgcga cgtcagcgtg ttccggccagc	3840
gctggggctt cgacgcgcgc accatcaaca gccgtacaa cgacctgacc cgctgatcg	3900
gcaactacac cgaccacgcc gtgcgctggt acaacaccgg cctggagcgc gtgtggggtc	3960
cagacagccg cgactggatc aggtacaacc agttccgcgc cgagctgacc ctgaccgtgc	4020
tggacatcgt gagcctgttc cccaactacg acagccgcac ctaccccatc cgcaccgtga	4080
gccagctgac ccgcgagatt tacaccaacc ccgtgctgga gaacttcgac ggcagcttcc	4140
gcggcagcgc ccagggcctc gagggcagca tccgcagccc ccacctgatg gacatcctga	4200
acagcatcac catctacacc gacgcccacc gcggcgagta ctactggagc ggccaccaga	4260

tcatggccag	ccccgtcggc	ttcagcggcc	ccgagttcac	cttccccctg	tacggcacta	4320
tgggcaacgc	tgcacctcag	cagcgcacgc	tggcacagct	gggccaggga	gtgtaccgca	4380
ccctgagcag	cacctgttac	cgtcgacctt	tcaacatcgg	catcaacaac	cagcagctga	4440
gcgtgctgga	cggcaccgag	ttcgcctacg	gcaccagcag	caacctgccc	agcgccgtgt	4500
accgcaagag	cggcaccgtg	gacagcctgg	acgagatccc	ccctcagaac	aacaacgtgc	4560
cacctcgaca	gggcttcagc	caccgtctga	gccacgtgag	catgttccgc	agtggcttca	4620
gcaacagcag	cgtgagcadc	atccgtgcac	ctatgttcag	ctggattcac	cgcagtgccg	4680
agttcaacaa	catcatcccc	agcagccaga	tcaccagat	ccccctgacc	aagagcacca	4740
acctgggcag	cggcaccagc	gtggtgaagg	gccccggctt	caccggcggc	gacatcctgc	4800
gccgcaccag	ccccggccag	atcagcacc	tgcgcgtgaa	catcaccgcc	cccctgagcc	4860
agcgctaccg	cgtccgcadc	cgtacgcca	gcaccacca	cctgcagttc	cacaccagca	4920
tcgacggccg	ccccatcaac	cagggcaact	tcagcgccac	catgagcagc	ggcagcaacc	4980
tgcagagcgg	cagcttccgc	accgtgggct	tcaccacccc	cttcaacttc	agcaacggca	5040
gcagcgtggt	cacctgagc	gcccacgtgt	tcaacagcgg	caacgaggtg	tacatcgacc	5100
gcacgcagtt	cgtgcccgcc	gaggtgacct	tcgaggccga	gtacgacctg	gagagggctc	5160
agaaggccgt	gaacgagctg	ttcaccagca	gcaaccagat	cggcctgaag	accgacgtga	5220
ccgactacca	catcgatcag	gtgtaggagc	tcaggttgaa	ggcaaacaag	ggtcaaatgg	5280
atgccattcc	attcatttgc	tttccaaggt	tcagcttccc	cgcgaatttt	cattgtgttt	5340
tctccgagat	gaatgtttgt	gttcggtgaa	atcagagtcg	tcagtcactc	acatagcttt	5400
tcttggttga	tagactgtta	ttttaagtcg	catgtttatc	tgggatggct	ggggtcagca	5460
tgtttgtaca	attatttgga	gttgcttttg	gaatggctgc	ggtttgatga	gttgccctaca	5520
gccatgagat	gtcttttcgc	tcacttttta	tggttcattc	gttctcaata	atatgggatg	5580
ctatacttgt	gttgatccat	tatcttgatg	catcgttgtc	tgtgcactgc	aacaacaata	5640
cccatctgaa	cacctctatc	aataaaatac	catttttttt	ctttccatcc	cactaatcta	5700
ggcccacttt	ctcactcttt	tcttatccac	tatattgtca	atatagaatc	tggggagaga	5760
gagagagaga	gagagattga	gagagagaga	gagagagaga	tttaggctcc	ctcccttcc	5820
atattcaaaa	taggtatcgc	ctttgggtca	cctgttgga	tgatttattt	tagtatcgtc	5880
catatttaat	ttaaacaat	gttgaatttt	acatgcatat	tgatttaaag	tttggttagtt	5940
tatgagcadc	accaactaag	atctctaata	ocaaagagca	tctccactag	ttccaaaaaa	6000
ctctctaaat	ttaatttagg	atgttagtaa	ccagaaatta	tgtccaaca	gtttccctaaa	6060
tgagttcctt	aaatatacca	actttttaaa	tatctctatt	tagtcaaaact	tgagaaattg	6120
tttcacactc	ccaatagttg	tcataaatca	catcgtaaaa	tcatttgagt	tcgcacccgt	6180
gtacaagtgg	tgcactttta	ataaattaaa	ttacgaaaat	gataaattta	ctatcgagtt	6240
agcataattt	aaaaatatat	aacaacataa	gaactgggaa	cggaccgcga	tcgcttaatt	6300
aagcttgcat	gcctgcagtg	cagcgtgacc	cggtcgtgcc	cctctctaga	gataatgagc	6360
atlgcatgtc	taagttataa	aaaattacca	catatttttt	ttgtcacact	tgtttgaagt	6420

gcagtttatac	tatctttata	catatatatta	aactttactc	tacgaataat	ataatctata	6480
gtactacaat	aatatcagtg	ttttagagaa	tcatataaat	gaacagttag	acatgggtcta	6540
aaggacaatt	gagtattttg	acaacaggac	tctacagttt	tatcttttta	gtgtgcatgt	6600
gttctccttt	ttttttgcaa	atagcttcac	ctatataata	cttcatccat	tttattagta	6660
catccattta	gggttttaggg	ttaatggttt	ttatagacta	atTTTTTTtag	tacatctatt	6720
ttattctatt	ttagcctcta	aattaagaaa	actaaaactc	tatttttagtt	tttttattta	6780
ataatttaga	tataaaatag	aataaaataa	agtgactaaa	aattaaacaa	atacccttta	6840
agaaattaaa	aaaactaagg	aaacattttt	cttgtttcga	gtagataatg	ccagcctggt	6900
aaacgccgtc	gacgagtcta	acggacacca	accagcgaac	cagcagcgtc	gcgtcggggc	6960
aagcgaagca	gacggcacgg	catctctgtc	gctgcctctg	gacccctctc	gagagttccg	7020
ctccaccgtt	ggacttgctc	cgctgtcggc	atccagaaat	tgctgtggcg	agcggcagac	7080
gtgagccggc	acggcaggcg	gcctcctcct	cctctcacgg	caccggcagc	tacgggggat	7140
tcctttccca	ccgtccttc	gctttccctt	cctcgcccg	cgtaataaat	agacaccccc	7200
tccacaccct	ctttcccaa	cctcgtgttg	ttcggagcgc	acacacacac	aaccagatct	7260
cccccaaatac	caccgcgcg	cacctccgct	tcaagggtacg	ccgctcgtcc	cccccccccc	7320
cccctctcta	ccttctctag	atcggcggtc	cgggccatgg	ttagggcccg	gtagttctac	7380
ttctgttcat	gtttgtgtta	gatccgtgtt	tgtgttagat	ccgtgctgct	agcgttcgta	7440
cacggatgcg	acctgtacgt	cagacacggt	ctgattgcta	acttgccagt	gtttctcttt	7500
ggggaatcct	gggatggctc	tagccgttcc	gcagacggga	tcgatttcat	gatttttttt	7560
gtttcgttgc	atagggtttg	gtttgccctt	ttcctttatt	tcaatatatg	ccgtgcactt	7620
gtttgtcggg	tcatcttttc	atgctttttt	ttgtcttggg	tgtgatgatg	tggtctgggt	7680
gggcggtcgt	tctagatcgg	agtagaatcc	tgtttcaaac	tacctgggtg	atttattaat	7740
tttgatctg	tatgtgtgtg	ccatacatat	tcatagttac	gaattgaaga	tgatggatgg	7800
aaatatcgat	ctaggatagg	tatacatgtt	gatgcggggt	ttactgatgc	atatacagag	7860
atgctttttg	ttcgcttggg	tgtgatgatg	tggtgtgggt	gggcggtcgt	tcattcgttc	7920
tagatcggag	tagaatactg	tttcaaacta	cctggtgtat	ttattaattt	tggaactgta	7980
tgtgtgtgtc	atacatcttc	atagttacga	gtttaagatg	gatggaaata	tcgatctagg	8040
ataggtatac	atgttgatgt	gggttttact	gatgcatata	catgatggca	tatgcagcat	8100
ctattcatat	gctctaacct	tgagtaccta	tctattataa	taaacaagta	tgttttataa	8160
ttattttgat	cttgatatac	ttggatgatg	gcatatgcag	cagctatatg	tggttttttt	8220
tagccctgcc	ttcatacgct	atttattttg	ttggtactgt	ttcttttgtc	gatgctcacc	8280
ctgttgtttg	gtgttacttc	tgcaaggatc	cccgatcatg	caaaaactca	ttactcagt	8340
gcaaaaactat	gcctggggca	gcaaaacggc	gttgactgaa	ctttatggta	tggaaaatcc	8400
gtccagccag	ccgatggccg	agctgtggat	gggcgcacat	ccgaaaagca	gttcacaggt	8460
gcagaatgcc	gccggagata	tcgttttact	gcgtgatgtg	attgagagtg	ataaatcgac	8520
tctgctcgga	gaggccgttg	ccaaacgctt	tggcgaactg	cctttcctgt	tcaaagtatt	8580

atgcgcagca cagccactct ccattcaggt tcatccaaac aaacacaatt ctgaaatcgg	8640
ttttgccaaa gaaaatgccg caggtatccc gatggatgcc gccgagcgta actataaaga	8700
tcctaaccac aagccggagc tggtttttgc gctgacgcct ttccttgcca tgaacgcgtt	8760
tcgtgaatth tccgagattg tctccctact ccagccggtc gcaggtgcac atccggcgat	8820
tgctcactth ttacaacagc ctgatgccga acgtttaagc gaactgttcg ccagcctgth	8880
gaatatgcag ggtgaagaaa aatcccgcgc gctggcgatt ttaaaatcgg ccctcgatag	8940
ccagcagggg gaaccgtggc aaacgattcg tttaatthct gaatthtacc cggaagacag	9000
cggctctgth tccccgctat tgctgaatgt ggtgaaattg aacctggcg aagcgatgth	9060
cctgttcgct gaaacaccgc acgcttacct gcaaggcgth gcgctggaag tgatggcaaa	9120
ctccgataac gtgctgcgth cgggtctgac gcctaaatac attgatath cggaactggt	9180
tgccaatgth aaattcgaag ccaaaccggc taaccagtht ttgaccagc cggtgaaaca	9240
aggtgcagaa ctggactthc cgattccagt ggatgathth gccttctcgc tgcathacct	9300
tagtgataaa gaaaccacca ttagccagca gagtgcgcgc atthtgtthc gcgtcgaagg	9360
cgatgcaacg ttgtggaaag gttctcagca gttacagctt aaaccgggtg aatcagcgth	9420
tattgcgcgc aacgaatcac cgggtgactgt caaaggccac ggccgtthtag cgcgtgttht	9480
caacaagctg taagagctta ctgaaaaaat taacatctct tgctaagctg ggagctcgtc	9540
atgggtcgtt taagctgcgc atgtgcctgc gtcgtctggt gccctctctc catatggagg	9600
ttgtcaaagt atctgctgth cgtgtcatga gtcgtgtcag tgttggttht ataatggacc	9660
ggttggtgtg tgtgtgcgta ctaccagaa ctatgacaaa tcatgaataa gthtcatgth	9720
tgaaattaaa gccgtgtctc attatgtthc gtctthcagt tgtctcctaa tathtgcctg	9780
caggtactgg ctatctaccg thtcttactt aggaggtgth tgaatgcact aaaactaata	9840
gthtagtggc aaaattagth aaacatcca aacaccatag ctaatagtht aactattagc	9900
taththtggg aaattagtht atagttaggt agthaththt tagctagcta attcaactaa	9960
caaththtag ccaactaaca attagththc gtgcattcaa acacccctt aatgttaacg	10020
tggthctatc taccgtctcc taatatatgg ttgattgttc ggtthgttgc tatgctattg	10080
ggtthctgatt gctgctagth ctgtctgaat ccagaagthc tcgtagtata gctcagathc	10140
atathththt thttagtgat aagtgatcca ggtthattact atgttagcta ggtthththt	10200
acaaggataa attatctgtg atcataathc ttatgaaagc thtatgtthc ctggaggcag	10260
tggcatgcaa tgcathgacg caacttgatc acaccagctg aggtagatac ggtaacaagg	10320
thcttaaatc tgttcaccaa atcattggag aacacacata cacathcttg ccagthcttg	10380
ttagagaaat thcatgacaa aatgccaaag ctgtcttgac ththcactth tggccatgag	10440
tcgtgactth gthtggthth atggaccggc thctctagct tgtthctactc aaaactgttg	10500
ttgatgcgaa taagthtgta tggthgatct ctggaththt ththgtctc aatagtggac	10560
gagattagat agcccgaaa thtaccggth cccgggcggc cagcatggcc gtatccgcaa	10620
tgtgttatta agthgtctaa gcgtcaatht gthtacacca caatathctc tgccaccagc	10680
cagccaacag ctccccgacc ggcagctcgg caaaaaatca cactcgata caggcagccc	10740

atcagaatta attctcatgt ttgacagctt atcatcgact gcacgggtgca ccaatgcttc 10800
 tggcgctcagg cagccatcgg aagctgtggt atggctgtgc aggtcgtaaa tcaactgcata 10860
 attcgtgtcg ctcaaggcgc actcccgttc tggataatgt tttttgcgcc gacatcataa 10920
 cggttcttggc aaatattctg aaatgagctg ttgacaatta atcatccggc tcgtataatg 10980
 tgtggaattg tgagcggata acaattttcac acaggaaaca gaccatgagg gaagcgttga 11040
 tcgccgaagt atcgactcaa ctatcagagg tagttggcgt catcgagcgc catctcgaac 11100
 cgacgttgct ggccgtacat ttgtacggct ccgcagtgga tggcggcctg aagccacaca 11160
 gtgatattga tttgctgggtt acggtgaccg taaggcttga tgaaacaacg cggcgagctt 11220
 tgatcaacga ctttttgaa acttcggctt cccctggaga gagcgagatt ctccgcgctg 11280
 tagaagtcac cattgttgtg cacgacgaca tcattccgtg gcgttatcca gctaagcgcg 11340
 aactgcaatt tggagaatgg cagcgcaatg acattcttgc aggtatcttc gagccagcca 11400
 cgatcgacat tgatctggct atcttgctga caaaagcaag agaacatagc gttgccttgg 11460
 taggtccagc ggccgaggaa ctctttgatc cggttcctga acaggatcta tttgaggcgc 11520
 taaatgaaac cttaacgcta tggaactcgc cgcccgactg ggctggcgat gagcgaaatg 11580
 tagtgcttac gttgtccgc atttggtaca gcgcagtaac cggcaaaatc gcgccgaagg 11640
 atgtcgctgc cgactgggca atggagcgc tgccggccca gtatcagccc gtcatacttg 11700
 aagctaggca ggcttatctt ggacaagaag atcgcttggc ctgcgcgcga gatcagttgg 11760
 aagaatttgt tcaactacgtg aaaggcgaga tcaccaaaagt agtcggcaaa taaagctcta 11820
 gtggatctcc gtaccgggg atctggctcg cggcggacgc acgacgccgg ggcgagacca 11880
 taggcgatct cctaaatcaa tagtagctgt aacctcgaag cgtttcactt gtaacaaoga 11940
 ttgagaattt ttgtcataaa attgaaatac ttggttcgca tttttgtcat ccgcggctag 12000
 ccgcaattct gacgaactgc ccatttagct ggagatgatt gtacatcctt cacgtgaaaa 12060
 tttctcaagc gctgtgaaca agggttcaga ttttagattg aaagggtgagc cgttgaaaca 12120
 cgttcttctt gtcgatgacg acgtcgctat gcggcatctt attattgaat accttacgat 12180
 ccacgccttc aaagtgaccg cggtagccga cagcaccag ttcacaagag tactctcttc 12240
 cgcgacggtc gatgtcgtgg ttgttgatct agatttaggt cgtgaagatg ggctogagat 12300
 cgttcgtaat ctggcggcaa agtctgatat tccaatcata attatcagtg gcgaccgcct 12360
 tgaggagacg gataaagttg ttgcaactga gctaggagca agtgatttta tcgctaagcc 12420
 gttcagtatc agagagtttc tagcacgcat tcgggttgcc ttgcgcgtgc gccccaacgt 12480
 tgtccgctcc aaagaccgac ggtctttttg ttttactgac tggacactta atctcaggca 12540
 acgtcgcttg atgtccgaag ctggcgggtga ggtgaaactt acggcagggtg agttcaatct 12600
 tctcctcgcg tttttagaga aaccccgcca cgttctatcg cgcgagcaac ttctcattgc 12660
 cagtcgagta cgcgacgagg aggtttatga caggagtata gatgttctca ttttgaggct 12720
 gcgccgcaaa cttgaggcag atccgtcaag ccctcaactg ataaaaacag caagagggtgc 12780
 cggttatttc tttgacgcgg acgtgcaggt ttgcacggg gggacgatgg cagcctgagc 12840
 caattcccag atccccgagg aatcggcgtg agcggtcgca aaccatccgg cccggtacaa 12900

atcggcgcg	cgctgggtga	tgacctggtg	gagaagttga	aggccgcgca	ggccgcccag	12960
cggcaacgca	tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	13020
atccgcaaag	aatcccggca	accgcccggca	gccggtgcgc	cgtcgattag	gaagccgccc	13080
aagggcgacg	agcaaccaga	ttttttcgtt	ccgatgctct	atgacgtggg	caccgcgcat	13140
agtcgcagca	tcatggacgt	ggccgttttc	cgtctgtcga	agcgtgaccg	acgagctggc	13200
gaggtgatcc	gctacgagct	tccagacggg	cacgtagagg	tttcgcgagg	gccggccggc	13260
atggccagtg	tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	13320
atgaaccgat	accgggaagg	gaagggagac	aagcccggcc	gcgtgttccg	tccacacgtt	13380
gcggacgtac	tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaaga	cgacctggta	13440
gaaacctgca	ttcggttaaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	13500
aacggccgcc	tggtagcgtt	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	13560
aagagcgaaa	ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	13620
gagatcacag	aaggcaagaa	cccggacgtg	ctgacggttc	accccgatta	ctttttgatc	13680
gatcccggca	tcggccgttt	tctctaccgc	ctggcacgcc	gcgccgcagg	caaggcagaa	13740
gccagatggg	tgttcaagac	gatctacgaa	cgcagtggca	gcgccggaga	gttcaagaag	13800
ttctgtttca	ccgtgcgcaa	gctgatcggg	tcaaatgacc	tgccggagta	cgatttgaag	13860
gaggaggcgg	ggcaggctgg	cccgatccta	gtcatgcgct	accgcaacct	gatcgagggc	13920
gaagcatccg	ccggttccta	atgtacggag	cagatgctag	ggcaaattgc	cctagcaggg	13980
gaaaaaggtc	gaaaaggctc	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	14040
tacattggga	accggaacct	gtacattggg	aaccctaaagc	cgtacattgg	gaaccggtca	14100
cacatgtaag	tgactgatat	aaaagagaaa	aaaggcgatt	tttcgccta	aaactcttta	14160
aaacttatta	aaactcttaa	aaccgcctg	gcctgtgcat	aactgtctgg	ccagcgcaca	14220
gccgaagagc	tgcaaaaagc	gcctacctt	cggtcgctgc	gctccctacg	ccccgccgct	14280
tcgcgtcggc	ctatcgcggc	cgctggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	14340
ctaccagggc	gcggacaagc	cgcgcgctcg	ccactcgacc	gccggcgctg	aggtctgcct	14400
cgtgaagaag	gtgttgctga	ctcataccag	gcctgaatcg	ccccatcctc	cagccagaaa	14460
gtgagggagc	cacggttgat	gagagctttg	ttgtagggtg	accagttggg	gattttgaac	14520
ttttgctttg	ccacggaacg	gtctgcgttg	tcgggaagat	gcgtgatctg	atccttcaac	14580
tcagcaaaaag	ttcgatttat	tcaacaaagc	cgcctccccg	tcaagtcagc	gtaatgctct	14640
gccagtgtta	caaccaatta	accaattctg	attagaaaaa	ctcatcgagc	atcaaatgaa	14700
actgcaatth	attcatatca	ggattatcaa	taccatattt	ttgaaaaagc	cgtttctgta	14760
atgaaggaga	aaactcaccg	aggcagttcc	ataggatggc	aagatcctgg	tatcggctcg	14820
cgattccgac	tcgtccaaca	tcaatacaac	ctattaattt	cccctcgta	aaaataaggt	14880
tatcaagtga	gaaatcacca	tgagtgcgca	ctgaatccgg	tgagaatggc	aaaagctctg	14940
cattaatgaa	tcggccaacg	cgcggggaga	ggcggtttgc	gtattgggag	ctcttccgct	15000
tcctcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggg	atcagctcac	15060

tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga 15120
 gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat 15180
 aggctccgcc cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac 15240
 ccgacaggac tataaagata ccaggcggtt cccctggaa gctccctcgt gcgctctcct 15300
 gttccgaccc tgccgcttac cggatacctg tccgccttct tcccttcggg aagcgtggcg 15360
 ctttctcata gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg 15420
 ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt 15480
 cttgagtcca acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg 15540
 attagcagag cgaggatatgt aggcggtgct acagagttct tgaagtgggt gcctaactac 15600
 ggctacacta gaagaacagt atttgggtatc tgcgctctgc tgaagccagt taccttcgga 15660
 aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 15720
 gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 15780
 tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga 15840
 ttatcaaaaa ggatcttcac ctagatcctt ttgatccgga attt 15884

<210> 26

<211> 15689

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vector tái tổ hợp

<400> 26

attcctgtgg ttggcatgca catacaaatg gacgaacgga taaacctttt cagccctttt 60
 taaatatccg attattctaa taaacgctct tttctcttag gtttaccgc caatatatcc 120
 tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga 180
 attaaggagag tcacgttatg acccccgccg atgacgcggg acaagccgtt ttacgtttgg 240
 aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccattttaa tcaattgggc 300
 gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc 360
 taccaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg 420
 gaacaaaata acgtggaaaa gagctgtcct gacagccac tcaactaatgc gtatgacgaa 480
 cgcagtgcag accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct 540
 cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaagggtg 600
 ctctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga 660
 cagtgggtccc aaagatggac cccacccac gagggagcatc gtggaaaaag aagacgttcc 720
 aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga 780
 acaatccac tatccttcgg taccggaccc ttgtattgt tglatgtaat gaataatctt 840

tatatatata tatatatata tatatatata tatatatata tatatatata tatatatgtg	900
tgtgtgtggg ggggtgtggg ggtgtgtgta ttgatgtctc ttggacaaca agattacaca	960
aaacacaatt agaataaaaa aatatcctcg tgtatataaa cttgtccgta taccatatta	1020
gaacacataa atttttaggca acattttttc catcaacatt cttcaatcat caacccaaat	1080
ttacggatac acaagataag aggggggtatg taagggttgta cgtaatgggc tacatgataa	1140
catcaaagat tatgcaagca aatctcaatc caccaggcga tcataaacat catagtccac	1200
atgcttcagt tttaaagcca tcgtagttcc cgtgcaaaga caaaaacatt agaattattc	1260
aatacaagtt gcacaagata gttcaataaa atttaaacca caatagtatt atccaacaaa	1320
gctagtccat accattatag tttctagtaa acaagaatag agaacatata ttaagcaaac	1380
agaccacata ggataaggct aaggatgtat ttttgtctat tgttttctgt gacatcgatc	1440
tcgtttgcac gagtaaaact aaaacatgag aaaattccga taaaaaatag gaatctagaa	1500
atacgaacgg aaaacactaa accatttttta ttcctgtttc taaaatatat catctcgttt	1560
ctgtttttct ctgtaaatat gaaaacgatc agatcatagt taaaacggaa tacagtaccg	1620
gacgaaacaa tattttctctc tcgtttcccc cttatagcat ctccaaaagc tccccagaag	1680
tctcccctaa atctatTTTT ttggaaaaac acaaaaacat gtctccaaca gttcctctaa	1740
agcgccccca actttttcat agcccttaaa actccctcat ttgtagctac aaatgagggg	1800
ttttttgggc tccccagaaa caaactgctg ctttaagatg tttgggttgag gagccaagta	1860
gaatggagtc gtttcatccc tgattctagg aacggagccg ttctgttctg tgtttggtaa	1920
tctggaacgg agcggctctg ttttttgttt ggttgcagag tgaacggaac ggagcgtgac	1980
tgtgagagcg ggatgagaac ggagcggctc cgttcggttg attttttggg gcggaatggt	2040
tccggatttg aggagaatat tccataattg gagtcattcc gttctagttc ctttataacc	2100
aaacagcaac aaaactggga tagaatggtt ccgttctact tggctcttca accaaacact	2160
aactaaggga cctgttgag aaatgattaa aatttaccct cacttattat ttagatatcc	2220
cttaaaacta attttgagaa gtcgttttat ggagtgtctt tggagatgct cttagtctgt	2280
agatctgatt gtgtgtatgt atgtaagcat atgcgtctgt actacgatcc acggtaaaaa	2340
agtcacaaac atataataat agcgtcattc aataaattga taacgtaaag tataaatgaa	2400
cttttaactc tgtttgagta aacaactcgt gcagagactg caatgaaatc tattatTTTT	2460
tctaaaataa ttatacaagt tgaggaaatg ctatttgttc cactaagcga cgatgtatTT	2520
tgttttttaa aaaaatcgac gaggtactgc tcggttatta ttttcacatg caccgcgcgt	2580
tgtttttggg ccggcccatt tglattgcca atttgcgag acgaatatga ccgaatggag	2640
tttagaaagc ccagctcact taggattgtc tattttctca agaaaagaga gaatgggccc	2700
aaaggcctaa acacccaaac ccgatccgct tatgggatgt catggacatg gagtctggga	2760
ccgtccggca gatgagacga cagccgtcgg atcagaaacc ctagcgcggg aggctctccc	2820
tattaatacc caccctgcac cccgcgggag gagtctttcc tagggtttcg tagcttctag	2880
ccgccgccgc gtccgcctcg ccaagcgcag cagccgccgc agcacatcgc tctctcgatc	2940
tcagccatag cggaggtgga gcagcaacag gacacgccga agctcttcaa ccgctggacc	3000

ttcgatgatg tccaggtacg cgaacgagtc ttgcgatctc tgcatgcttc gattcttttag	3060
ccttgccgct actagcagtg gatggaaccg acgatgaaat ctgcaggtga acgacatctc	3120
gctgtaaacc atggacaaca accccaacat caacgagtgc atccccata actgcctgag	3180
caaccccgag gtggaggtgc tgggcggcga gcgcacgag accggctaca ccccatcga	3240
catcagcctg agcctgaccc agttcctgct gagcgagttc gtgcccggcg ccggcttcgt	3300
gctgggcctg gtggacatca tctggggcat cttcggtccc agccagtggg acgccttct	3360
ggtgcagatc gagcagctga tcaaccagcg catcgaggag ttgcgccgca accaggccat	3420
cagccgctg gagggcctga gcaacctgta ccaaatctac gccgagagct tccgcgagtg	3480
ggaggccgac cccaccaacc ccgcctcgcg cgaggagatg cgcattccagt tcaacgacat	3540
gaacagcgcc ctgaccaccg ccatccccct gttcgccgtg cagaactacc aggtgccct	3600
gctgagcgtg tacgtgcagg ccgccaacct gcacctgagc gtgctgcgcg acgtcagcgt	3660
gttcggccag cgctggggct tcgacgcgcg caccatcaac agccgctaca acgacctgac	3720
ccgcctgac ggcaactaca ccgaccacgc cgtgcgctgg tacaacaccg gcctggagcg	3780
cgtgtggggc ccagacagcc gcgactggat caggtacaac cagttccgcc gcgagctgac	3840
cctgaccgtg ctggacatcg tgagcctggt cccaactac gacagccgca cctaccccat	3900
ccgcaccgtg agccagctga ccgcgcgagat ttacaccaac ccctgtctgg agaacttcga	3960
cggcagcttc cgcggcagcg ccaggggcat cgagggcagc atccgcagcc ccacctgat	4020
ggacatcctg aacagcatca ccatctacac cgacgcccac cgcggcgagt actactggag	4080
cggccaccag atcatggcca gccccgtcgg cttcagcggc ccgcagttca cttccccct	4140
gtacggcact atgggcaacg ctgcacctca gcagcgcac gtggcacagc tgggccaggg	4200
agtgtaccgc accctgagca gcacctgta ccgtcgacct ttcaacatcg gcatcaacaa	4260
ccagcagctg agcgtgctgg acggcaccga gttcgctac ggcaccagca gcaacctgcc	4320
cagcgcctg taccgcaaga gcggcaccgt ggacagcctg gacgagatcc cccctcagaa	4380
caacaacgtg ccacctcgac agggcttcag ccaccgtctg agccacgtga gcatgttcg	4440
cagtggcttc agcaacagca gcgtgagcat catccgtgca cctatgttca gctggattca	4500
ccgcagtgc gagttcaaca acatcatccc cagcagccag atcaccagaa tccccctgac	4560
caagagcacc aacctgggca gcggcaccag cgtggtgaag ggccccggct tcaccggcg	4620
cgacatcctg cgcgcacca gccccggcca gatcagcacc ctgcgcgtga acatcaccgc	4680
ccccctgagc cagcgtacc gcgtccgcac ccgtacgcc agcaccacca acctgcagtt	4740
ccacaccagc atcgacggcc gccccatcaa ccagggcaac ttcagcgcca ccatgagcag	4800
cggcagcaac ctgcagagcg gcagcttcg caccgtgggc ttcaccaccc ctttcaactt	4860
cagcaacggc agcagcgtgt tcacctgag cgcacacgtg ttcaacagcg gcaacgaggt	4920
gtacatcgac cgcacagagt tcgtgccgcg cgaggtgacc ttcgaggccg agtacgacct	4980
ggagagggct cagaaggccg tgaacgagct gttcaccagc agcaaccaga tcggcctgaa	5040
gaccgacgtg accgactacc acatcgatca ggtgtaggag ctactgagc ttgtatcctg	5100
gtgcactctg cgctggaaac ttttatgtcg ctggcagtcg tatcggttct tgttttacca	5160

atgttttagag	ttttttgaga	cctatatgcg	gttttggttt	tcagtgcaca	attaaaatta	5220
ctgagtaatg	tagttgattg	ggaacagaaa	tgtttggtgc	ctggtttacc	gaactccagt	5280
tctcttgtea	tttttcttta	ttctatagtc	tgtattatgt	atgcgtatga	gtattgagat	5340
gattctgcat	ttgaattgtc	tgctttggtg	ctgtcgctgt	atgcgtacca	atggtaacag	5400
ggtagttgtg	ggaagtagac	acggccggtt	ctatgttttc	gtgcttcgcg	ggttcaggct	5460
ggttaagcct	atggagcgta	cgcacgctcc	tcccgctctc	ccgtgtcctg	catgctggca	5520
acaggagtgc	ggcccagcgc	acgccctaata	cgacggcggt	atatttgtct	gtccctccat	5580
tttgtgggtga	ggctattcgc	aaccgttatc	cttaaatttt	ttctcctata	tcactattcc	5640
cctatttttc	cttatatttt	ttcatcttca	gcagcggttc	tcctaaatac	tcctctata	5700
cccactacaa	ctataaatat	tattttccat	atctattcat	catttattac	cacttttttt	5760
caactaaaaa	atactcgcat	gcatggattt	tacggaagg	gggctgtcac	agtatccct	5820
tgatctgctg	tgagagaaaa	gggggacact	aggtagggtg	caaggtaggg	agcagcgggtg	5880
cgggtggtag	cgtgggttact	gcagccgcta	cgacgtgagc	agtgttaggg	gagaggatgg	5940
aagggtgggtg	cgtgctgca	gataaccgga	gcgccaaaca	ctcatgggtg	ataattaggt	6000
ataagaaaag	atattttatg	gttaggagag	tatagagaga	atttagtggt	aacttctatg	6060
gaagatggaa	aaataggggt	gaaatcggac	cgcgatcgct	taattaagct	tgcatgcctg	6120
cagtgcagcg	tgaccgggtc	gtgcccctct	ctagagataa	tgagcattgc	atgtctaagt	6180
tataaaaaat	taccacatat	tttttttgtc	acacttggtt	gaagtgcagt	ttatctatct	6240
ttatacatat	atttaaacct	tactctacga	ataatataat	ctatagtact	acaataatat	6300
cagtgtttta	gagaatcata	taaatgaaca	gttagacatg	gtctaaagga	caattgagta	6360
ttttgacaac	aggactctac	agttttatct	ttttagtgtg	catgtgttct	cotttttttt	6420
tgcaaatagc	ttcacctata	taataacttca	tccattttat	tagtacatcc	atttaggggt	6480
tagggttaat	ggttttttata	gactaatttt	tttagtacat	ctattttatt	ctatttttagc	6540
ctctaaatta	agaaaactaa	aactctatct	tagttttttt	atttaataat	ttagatataa	6600
aatagaataa	aataaagtga	ctaaaaatta	aacaaatacc	ctttaagaaa	ttaaaaaac	6660
taaggaaaca	tttttcttgt	ttcgagtaga	taatgccagc	ctgttaaacg	ccgtcgacga	6720
gtctaacgga	caccaaccag	cgaaccagca	gcgtcgcgct	gggccaagcg	aagcagacgg	6780
caaggcatct	ctgtcgctgc	ctctggaccc	ctctcgagag	ttccgctcca	ccgttggaact	6840
tgctccgctg	tcggcatcca	gaaattgcgt	ggcggagcgg	cagacgtgag	ccggcacggc	6900
aggcggcctc	ctcctcctct	cacggcaccg	gcagctacgg	gggattcctt	tcccaccgct	6960
ccttcgcttt	cccttcctcg	cccgccgtaa	taaatagaca	ccccctccac	accctctttc	7020
cccaacctcg	tgttgttcgg	agcgcacaca	cacacaacca	gatctcccc	aaatccaccc	7080
gtcggcacct	ccgcttcaag	gtacgcgct	cgctctcccc	ccccccccct	ctctaccttc	7140
tctagatcgg	cgttcgggtc	catggttagg	gcccggtagt	totacttctg	ttcatgtttg	7200
tgttagatcc	gtgtttgtgt	tagatccgtg	ctgctagcgt	togtacacgg	atgcgacctg	7260
tacgtcagac	acgttctgat	tgctaacttg	ccagtgtttc	tctttgggga	atcctgggat	7320

ggctctagcc gttccgcaga cgggatcgat ttcattgattt tttttgtttc gttgcatagg 7380
 gtttggtttg ccccttttctt ttattttcaat atatgccgtg cacttggtttg tcgggtcacc 7440
 ttttcatgct tttttttgtc ttggttgtga tgatgtggtc tgggtgggcg gtcgttctag 7500
 atcggagtag aattctgttt caaactacct ggtggattta ttaatttttg atctgtatgt 7560
 gtgtgccata catattcata gttacgaatt gaagatgatg gatggaaata tcgatctagg 7620
 ataggtatac atgttgatgc gggttttact gatgcatata cagagatgct ttttgttcgc 7680
 ttggttgtga tgatgtggtg tgggtgggcg gtcgttcatt cgttctagat cggagtagaa 7740
 tactgtttca aactacctgg tgtatttatt aattttggaa ctgtatgtgt gtgtcataca 7800
 tcttcatagt tacgagttta agatggatgg aaatatcgat ctaggatagg tatacatggt 7860
 gatgtgggtt ttactgatgc atatacatga tggcatatgc agcatctatt catatgctct 7920
 aaccttgagt acctatctat tataataaac aagtatgttt tataattatt ttgatcttga 7980
 tatacttgga tgatggcata tgcagcagct atatgtggat ttttttagcc ctgccttcat 8040
 acgtatttta tttgcttggg actgtttctt ttgtcgatgc tcacctgtt gtttggtgtt 8100
 acttctgcag ggatccccga tcatgcaaaa actcattaac tcagtgc aaa actatgcctg 8160
 gggcagcaaa acggcggtga ctgaacttta tggatggaa aatccgtcca gccagccgat 8220
 ggccgagctg tggatgggcg cacatccgaa aagcagttca cgagtgcaga atgccgccgg 8280
 agatatcggt tcaactgcgtg atgtgattga gagtgataaa tcgactctgc tcggagagggc 8340
 cgttgcaaaa cgctttggcg aactgccttt cctgttcaaaa gtattatgcg cagcacagcc 8400
 actctccatt caggttcacc caaacaacaa caattctgaa atcgggttttg ccaaagaaaa 8460
 tgcgcaggt atcccgatgg atgccgccga gcgttaactat aaagatccta accacaagcc 8520
 ggagctggtt tttgcgctga cgcttttctt tgcgatgaac gcgtttcgtg aattttccga 8580
 gattgtctcc ctactccagc cggtcgcagg tgcacatccg gcgattgctc actttttaca 8640
 acagcctgat gccgaacgtt taagcgaact gttcgccagc ctgttgaata tgcaggggtga 8700
 agaaaaatcc cgcgcgctgg cgatttttaa atcggccctc gatagccagc aggggtgaacc 8760
 gtggcaaacg attcgtttaa tttctgaatt ttacccggaa gacagcggtc tgttctcccc 8820
 gctattgctg aatgtggtga aattgaaccc tggcgaagcg atgttcctgt tcgctgaaac 8880
 accgcacgct tacctgcaag gcgtggcgct ggaagtgatg gcaaactccg ataactgtct 8940
 gcgtgcgggt ctgacgccta aatacattga tattccggaa ctggttgcca atgtgaaatt 9000
 cgaagccaaa ccggctaacc agttgttgac ccagccggtg aaacaagggtg cagaactgga 9060
 cttcccgatt ccagtggatg attttgccct ctgcgtgcat gaccttagtg ataaagaaac 9120
 caccattagc cagcagagtg ccgccatttt gttctgcgtc gaaggcgatg caacgttgtg 9180
 gaaaggttct cagcagttac agcttaaacc gggatgaatca gcgtttattg ccgccaacga 9240
 atcaccggtg actgtcaaaag gccacggccg tttagcgcgt gtttacaaca agctgtaaga 9300
 gcttactgaa aaaattaaca tctcttgcta agctgggagc tcgtcatggg tcgtttaagc 9360
 tgccgatgtg cctgcgtcgt ctggtgccct ctctccatat ggaggttgtc aaagtatctg 9420
 ctgttcgtgt catgagtcgt gtcagtgttg gtttaataat ggaccggttg tgttgtgtgt 9480

gcgtactacc	cagaactatg	acaaatcatg	aataagtttg	atgtttgaaa	ttaaagcctg	9540
tgctcattat	gttctgtctt	tcagttgtct	cctaataattt	gcctgcaggt	actggctatc	9600
taccgtttct	tacttaggag	gtgtttgaat	gcactaaaac	taatagttag	tggtctaaaat	9660
tagttaaaac	atccaaacac	catagctaata	agttgaacta	ttagctattt	ttggaaaatt	9720
agttaatagt	gaggtagtta	tttgtttagct	agctaattca	actaacaatt	tttagccaac	9780
taacaattag	tttcagtgc	ttcaaacacc	cccttaatgt	taacgtgggt	ctatctaccg	9840
tctcctaata	tatggttgat	tgttcgggtt	gttgctatgc	tattgggttc	tgattgctgc	9900
tagttcttgc	tgaatccaga	agttctcgta	gtatagctca	gattcatatt	atattattga	9960
gtgataagtg	atccagggtta	ttactatggt	agctagggtt	tttttacaag	gataaattat	10020
ctgtgatcat	aattcttatg	aaagctttat	gtttcctgga	ggcagtgcca	tgcaatgcat	10080
gacagcaact	tgatcacacc	agctgaggta	gatacggtaa	caaggttctt	aaatctgttc	10140
accaaatacat	tgagagaacac	acatacacat	tcttgccagt	cttggttaga	gaaatttcat	10200
gacaaaatgc	caaagctgtc	ttgactcttc	acttttggcc	atgagtcgtg	acttagtttg	10260
gtttaatgga	cgggttctcc	tagcttggtc	tactcaaaac	tgttgttgat	gcgaataagt	10320
tgtgatgggt	gatctctgga	ttttgttttg	ctctcaatag	tggacgagat	tagatagccc	10380
ggaaatttac	cgggtgcccgg	gcggccagca	tggccgtatc	cgcaatgtgt	tattaagttg	10440
tctaagcgtc	aatttgttta	caccacaata	tatcctgcca	ccagccagcc	aacagctccc	10500
cgaccggcag	ctcggcacia	aatcaccact	cgatacaggc	agcccatcag	aattaattct	10560
catgtttgac	agcttatcat	cgactgcacg	gtgcaccaat	gcttctggcg	tcaggcagcc	10620
atcggaagct	gtggtatggc	tgtgcaggtc	gtaaatcaat	gcataattcg	tgtcgtctca	10680
ggcgcaactcc	cgttctggat	aatgtttttt	gogcogacat	cataacgggt	ctggcaataa	10740
ttctgaaatg	agctgttgac	aattaatcat	ccggctcgta	taatgtgtgg	aattgtgagc	10800
ggataacaat	ttcacacagg	aaacagacca	tgagggaagc	gttgatcgcc	gaagtatcga	10860
ctcaactatc	agaggtagtt	ggcgtcatcg	agcgccatct	cgaaccgacg	ttgctggccg	10920
tacatttgta	cggctccgca	gtggatggcg	gcctgaagcc	acacagtgat	attgatttgc	10980
tgggtacggg	gacogtaagg	cttgatgaaa	caacgcggcg	agctttgatc	aacgaccttt	11040
tggaaacttc	ggcttcccct	ggagagagcg	agattctccg	cgctgtagaa	gtcaccattg	11100
ttgtgcacga	cgacatcatt	cogtggcggt	atccagctaa	gcgcgaactg	caatttggag	11160
aatggcagcg	caatgacatt	cttgacaggta	tcttcagagc	agccacgacg	gacattgacg	11220
tggctatctt	gctgacaaaa	gcaagagaac	atagcgttgc	cttggttaggt	ccagcggcgg	11280
aggaactctt	tgatccgggt	cctgaacagg	atctatttga	ggcgctaaat	gaaaccttaa	11340
cgtatggaa	ctcgccgccc	gactgggctg	gcgatgagcg	aaatgtagtg	cttacgttgt	11400
cccgcatttg	gtacagcgca	gtaaccggca	aaatcgcgcc	gaaggatgtc	gctgccgact	11460
gggcaatgga	gcgcctgccc	gcccagtatc	agcccgatcat	acttgaagct	aggcaggctt	11520
atcttggaca	agaagatcgc	ttggcctcgc	gcgcagatca	gttggaagaa	tttgttcact	11580
acgtgaaagg	cgagatcacc	aaagtagtcg	gcaataaaag	ctctagtggg	tctccgtacc	11640

cggggatctg	gctcgcggcg	gacgcacgac	gccggggcga	gaccataggc	gatctcctaa	11700
atcaatagta	gctgtaacct	cgaagcggtt	cacttgtaac	aacgattgag	aatttttgtc	11760
ataaaattga	aatacttggg	tcgcattttt	gtcatccgcg	gtcagccgca	attctgacga	11820
actgcccatt	tagctggaga	tgattgtaca	tccttcacgt	gaaaatttct	caagcgctgt	11880
gaacaagggt	tcagatttta	gattgaaagg	tgagccgttg	aaacacgttc	ttcttgtcga	11940
tgacgacgtc	gctatgcggc	atcttattat	tgaatacctt	acgatccacg	ccttcaaagt	12000
gaccgcggta	gccgacagca	cccagttcac	aagagtactc	tcttccgcga	cggtcgatgt	12060
cgtggttgtt	gatctagatt	taggtcgtga	agatgggctc	gagatcgttc	gtaatctggc	12120
ggcaaagtct	gatattccaa	tcataattat	cagtggcgac	cgccttgagg	agacggataa	12180
agttgttgca	ctcgagctag	gagcaagtga	ttttatcgct	aagccgttca	gtatcagaga	12240
gtttctagca	cgcattcggg	ttgccttgcg	cgtgcgcccc	aacgttgtcc	gtcccaaaga	12300
ccgacgggtc	ttttgtttta	ctgactggac	acttaatctc	aggcaacgtc	gcttgatgtc	12360
cgaagctggc	ggtgaggtga	aacttacggc	aggtgagttc	aatcttctcc	tcgcgttttt	12420
agagaaaacc	cgcgacgttc	tatcgcgcg	gcaacttctc	attgccagtc	gagtacgcga	12480
cgaggaggtt	tatgacagga	gtatagatgt	tctcattttg	aggctgcgcc	gcaaacttga	12540
ggcagatccg	tcaagccctc	aactgataaa	aacagcaaga	ggtgccgggt	atttctttga	12600
cgcggacgtg	caggtttcgc	acggggggac	gatggcagcc	tgagccaatt	cccagatccc	12660
cgaggaatcg	gcgtgagcgg	tcgcaaacca	tccggcccgg	tacaaatcgg	cgcggcgctg	12720
ggtgatgacc	tggtggagaa	gttgaaggcc	gcgcaggccg	cccagcggca	acgcacgcag	12780
gcagaagcac	gccccggtga	atcgtggcaa	gcggccgctg	atcgaatccg	caaagaatcc	12840
cggcaaccgc	cggcagccgg	tcgcgcgtcg	attaggaagc	cgcceaaggg	cgacgagcaa	12900
ccagattttt	tcgttccgat	gctctatgac	gtgggcaccc	gcgatagtcg	cagcatcatg	12960
gacgtggccg	ttttccgtct	gtcgaagcgt	gaccgacgag	ctggcgaggt	gatccgctac	13020
gagcttccag	acgggcacgt	agaggtttcc	gcagggccgg	ccggcatggc	cagtgtgtgg	13080
gattacgacc	tggtactgat	ggcggtttcc	catctaaccg	aatccatgaa	ccgataccgg	13140
gaagggaagg	gagacaagcc	cggccgcgtg	ttccgtccac	acgttgcgga	cgtactcaag	13200
ttctgcccgc	gagccgatgg	cggaaagcag	aaagacgacc	tggtagaaac	ctgcattcgg	13260
ttaaacacca	cgcacgttgc	catgcagcgt	acgaagaagg	ccaagaacgg	ccgcctgggtg	13320
acggtatccg	aggggtgaagc	cttgattagc	cgctacaaga	tcgtaaagag	cgaaaccggg	13380
cggccggagt	acatcgagat	cgagctagct	gattggatgt	accgcgagat	cacagaaggc	13440
aagaaccggg	acgtgctgac	ggttcacccc	gattactttt	tgatcgatcc	cggcacgggc	13500
cgttttctct	accgcctggc	acgccgcgcc	gcaggcaagg	cagaagccag	atggttgttc	13560
aagacgatct	acgaacgcag	tggcagcgcc	ggagagttca	agaagttctg	tttcaccgtg	13620
cgcaagctga	tcgggtcaaa	tgacctgcgc	gagtacgatt	tgaaggagga	ggcggggcag	13680
gctggcccga	tcctagtcat	gcgctaccgc	aacctgatcg	agggcggaagc	atccgcgggt	13740
tcctaattga	cggagcagat	gctagggcaa	attgccttag	caggggaaaa	aggtcgaaaa	13800

ggtctctttc ctgtggatag cacgtacatt gggaacccaa agccgtacat tgggaaccgg 13860
 aaccggtaca ttgggaaccc aaagccgtac attgggaacc ggtcacacat gtaagtgact 13920
 gatataaaag agaaaaaagg cgattttttcc gcctaaaact ctttaaaact tattaanaact 13980
 cttaaaaccc gcctggcctg tgcataactg tctggccagc gcacagccga agagctgcaa 14040
 aaagcgcta ccttcgggtc gctgcgctcc ctacgccccg ccgcttcgag tgggcctatc 14100
 ggggcccgtg gccgctcaaa aatggctggc ctacggccag gcaatctacc agggcgcgga 14160
 caagccgcgc cgtcgccact cgaccgcccg cgctgaggtc tgcctcgtga agaaggtgtt 14220
 gctgactcat accaggcctg aatcgccccca tcatccagcc agaaagtgag ggagccacgg 14280
 ttgatgagag ctttggttga ggtggaccag ttggtgattt tgaacttttg ctttgccacg 14340
 gaacggtctg cgttgctggg aagatgcgtg atctgactct tcaactcagc aaaagttcga 14400
 tttattcaac aaagccgcgc tcccgtaag tcagcgtaat gctctgccag tgttacaacc 14460
 aattaaccaa ttctgattag aaaaactcat cgagcatcaa atgaaactgc aattttattca 14520
 tatcaggatt atcaatacca tttttttgaa aaagccgttt ctgtaatgaa ggagaaaact 14580
 caccgaggca gttccatagg atggcaagat cctggatatg gctcgcgatt ccgactcgtc 14640
 caacatcaat acaacctatt aatttccctt cgtcaaaaat aagggttatca agtgagaaat 14700
 caccatgagt gacgactgaa tccgggtgaga atggcaaaaag ctctgcatta atgaatcggc 14760
 caacgcgcgc ggagaggcgc tttgcgtatt gggcgctctt ccgcttcttc gctcactgac 14820
 tgcgtgcgct cggctcgttc gctgcggcga gcggtatcag ctcaactcaa ggcggttaata 14880
 cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa 14940
 aaggccagga accgtaaaaa ggccgcgttg ctggcggttt tccataggct ccgccccctt 15000
 gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa 15060
 agataaccagg cgtttccccc tggaaagctcc ctgctgcgct ctctgttcc gaccctgccg 15120
 cttaaccgat acctgtcgc ctttctccct tcgggaagcg tggcgctttc tcatagctca 15180
 cgctgtaggt atctcagttc ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa 15240
 cccccgttc agcccagccg ctgcgcctta tccgtaact atcgtcttga gtccaacccg 15300
 gtaagacacg acttatcgcc actggcagca gccactggta acaggattag cagagcgagg 15360
 tatgtaggcg gtgctacaga gttcttgaag tggtagccta actacggcta cactagaaga 15420
 acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc 15480
 tcttgatccg gcaaacaaac caccgctggg agcggtggtt tttttgtttg caagcagcag 15540
 attacgcgca gaaaaaaagg atctcaagaa gatcctttga tctttctac ggggtctgac 15600
 gctcagtggg acgaaaactc acgttaaggg attttggtca tgagattatc aaaaaggatc 15660
 ttcacctaga tctttttgat ccggaatta 15689

<210> 27

<211> 16054

<212> ADN

<213> Trình tự nhân tạo

<220>

<223> Vector tái tổ hợp

<400> 27

attcctgtgg ttggcatgca catacaaatg gacgaacgga taaacctttt cagccctttt	60
taaatatccg attattctaa taaacgctct tttctcttag gtttaccgc caatatatcc	120
tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga	180
attaagggag tcacgttatg acccccgccg atgacgcggg acaagccgtt ttacgtttgg	240
aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccatttaaa tcaattgggc	300
gcgccagctg cttgtgggga ccagacaaaa aaggaatggt gcagaattgt taggcgcacc	360
tacaaaaagc atctttgctt ttattgcaaa gataaagcag attcctctag tacaagtggg	420
gaacaaaata acgtggaaaa gagctgtcct gacagccac tcactaatgc gtatgacgaa	480
cgcagtgcag accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct	540
cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaaggtgg	600
ctcctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga	660
cagtgggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc	720
aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga	780
acaatcccac tatccttcgg taccggaccc ggatcctgag gtgtggcttg tatgtttcct	840
gacccctgg tgggtgttccc agtttgcaat acatcatttt ctgtagccct tgtacttgct	900
gtgatccctg tttattgttc agggaaaact tagtgacgtg tattagaaat atagaaacct	960
cacatttcga agagcagaat agtgtttgat aacagttttt ttaacgcaac aacgtttgtt	1020
agactgtgtc cagcggccag cgggtagtgt aaaataggtg acgcgaaact atatgtgata	1080
ttgtttgaca ctatttgag agtatagttt aaaatagggg gtgtgagtaa tctgctgaag	1140
atagccttaa gagttaagat accaaggtat agtttatcat atgcaaaaga aaagaaaaaa	1200
aaggaaatac aaaccattgt ttttatgttg gttagagcta ggtaattatc tcttgaccag	1260
tgtatttcca aacctgttct tctcagtttc tgtgctccgt cgatactgaa catgtgtgtt	1320
cattttctcat ttgttgaacg atggtatttc aggagctaga gaagaggaga gaggagcttg	1380
agtctctgtt aactgctgac cggatccgct gcgtttgggg gcggcaggat ctccgaccgg	1440
atccgcccgc cgtgcccgcg accggatccg ctgcgtttgg gggggggcag gatctccgac	1500
cggatccgcc cgccgtggcc gcaaccggat ccgctgcgtt tggggcggca ggatctccga	1560
ccggatccgc ccaccgtggc ggcgaccgga tccgctgcgt ttggggggcg caaggaagcg	1620
ggtgggggctt ctagggttca ggtggcgggc gtcggggtag agagtgcctg cggcgattct	1680
ggcggggcatg cgcaggggtg ggcaggggtg tggtcgagcg ggccatgcac aaccttccat	1740
ccgcgggatt ttgcgggcgt tgctggcgcg acggatttgc aggagttgct ggcgcgggagc	1800
tttgagggcg cggccgattt ggggactgcg ggcgttggcg attttgcggg cgttgctggc	1860
gcggcagatt ctgcgggcgt tgctggcgcg gggactgcgg acgcgacagc tgtgctgcgc	1920

tgcggccgcg aggaaggcga cggggactgc gggcgcgggg agggaaatcg cgggcggagc	1980
gcgcgatggc agaacacaca gaactatgga cgcctacact aaggggggtgt ttggtttcta	2040
gggactaatg tttagtcctt tcatttttatt ctttttttagt atataaattg ctaaatatag	2100
aaactaaaat aaagtttttag ttictatatatt tagcaattttt agaacaaaaa tggaaataaaa	2160
tgtagggact aaacattagt ccctaaaaaac caaacacccc ttaaatacat aagaagtagt	2220
agagattatt attattttgtg tctattgata tgatcattat gatattatat tttactgttt	2280
taatctaaat atagttatta tattcagata attattttat taaatccaaa aatatttaac	2340
ttctaatagta ataaggagag aatactctaa tatcataata atttcttcta aagtgacctt	2400
cggattttgga gatgactgac agggaggggt gtgcaccctt ctttctttcc ttttcaattg	2460
aagaacttgg gttgtgcttg ctacacacaac cgaatgaccg atgcacaccg acgcgcacgg	2520
agcggaccag cacagcccg aaactgcctt acgcgcacac gttttccgggt gcggcccagc	2580
agcagcaggg aggaaggttc gctcggtcgc tgggcctggc tccgcgtgcc gtccctccta	2640
taaatgcgag tctcgcgga gtccactggg caccagaagc tcaactactg ctcggtgcg	2700
gctgcccccg cggccccggc cagtccatcc cctcgctcgc tccccaaactc cagcaggcag	2760
atcagataca tccatccatt cgcgcacccg aaggtgagcg ccgtgaacga accatccgcc	2820
ctgctagctg cgatctgtag ccttgcgctg ctttcgcgcc tagatcgta cgtcacctat	2880
cacgatccgt gcggttctag atctgtggtt tttccttccc ctggtggctg aatccttcca	2940
tccaccagac caccacggga cctcggtgat tcccttttgtt tttcctgtgc cgagagccaa	3000
aatcgagggg gggggcttgt tttttalttg ctcggtctcc cgtgtctctg tgatctgatt	3060
tgctgtagta atcagcagga aaggaagggt tgaactaaga gcgcctggc ggtttctgctg	3120
tgcgtgaacc cggacgcgc gctcttccatc ccggcggcgc tgctgcaggt ggaggacttc	3180
tgcgcgcagt ggtgggaact catcaccacc actgcctggt tccgcgacca ctggtcccgc	3240
gagcgcgcgc acctggacga gatgcgcgag cagatogaag cggccggcct cctccccgac	3300
gacgaggacc tcttctacga cgaccaggtc gagcagggcc ccgtcgcgc cgcccttaag	3360
ataggtactg atgtctctct ctctctctct ctcttactct cccctcgatt ttagatctgc	3420
ctgaaggagc aatcatagt acctcacgtt ggtgcgtttt tctccaccag attcggtgct	3480
caaggcgctg taaaccatgg acaacaaccc caacatcaac gagtgcaccc cctacaactg	3540
cctgagcaac cccgaggtgg aggtgctggg cggcgagcgc atcgagaccg gctacacccc	3600
catcgacatc agcctgaycc tgaccacagt cctgctgagc gagttcgtgc ccggcgccgg	3660
cttcgtgctg ggccctggtg acatcatctg gggcatcttc ggccccagcc agtgggacgc	3720
cttcctggtg cagatcgagc agctgatcaa ccagcgcac caggagttcg cccgcaacca	3780
ggccatcagc ccctggagg gcctgagcaa cctgtaccaa atctacgcg agagcttccg	3840
cgagtgggag gccgacccca ccaaccccg cctgcgcgag gagatgcgca tccagttcaa	3900
cgacatgaac agegcctga ccaccgccat cccctgttc gccgtgcaga actaccagg	3960
gcccctgctg agcgtgtacg tgcaggccgc caacctgcac ctgagcgtgc tgcgcgacgt	4020
cagcgtgttc ggccagcgct ggggcttcga cgcgcaccc atcaacagcc gctacaacga	4080

cctgacccgc	ctgatcggca	actacaccga	ccacgccgtg	cgctggtaca	acaccggcct	4140
ggagcgcgtg	tggggtccag	acagccgcga	ctggatcagg	tacaaccagt	tccgccgcga	4200
gctgaccctg	accgtgctgg	acatcgtgag	cctgttcccc	aactacgaca	gccgcaccta	4260
ccccatccgc	accgtgagcc	agctgacccg	cgagatttac	accaaccccc	tgctggagaa	4320
cttcgacggc	agcttccgcg	gcagcgccca	gggcatcgag	ggcagcatcc	gcagccccca	4380
cctgatggac	atcctgaaca	gcatacccat	ctacaccgac	gccacccgcg	gcgagtacta	4440
ctggagcggc	caccagatca	tggccagccc	cgtcggcttc	agcggccccg	agttcacctt	4500
ccccctgtac	ggcactatgg	gcaacgctgc	acctcagcag	cgcatacgtg	cacagctggg	4560
ccagggagtg	taccgcaccc	tgagcagcac	cctgtaccgt	cgacctttca	acatcggcat	4620
caacaaccag	cagctgagcg	tgctggacgg	caccgagttc	gcctacggca	ccagcagcaa	4680
cctgcccagc	gcctgttacc	gcaagagcgg	caccgtggac	agcctggacg	agatcccccc	4740
tcagaacaac	aacgtgccac	ctcgacaggg	cttcagccac	cgtctgagcc	acgtgagcat	4800
gttccgcagt	ggcttcagca	acagcagcgt	gagcatcatt	cgtgcaccta	tgttcagctg	4860
gattcacccg	agtgcagagt	tcaacaacat	catccccage	agccagatca	cccagatccc	4920
cctgaccaag	agcaccaacc	tgggcagcgg	caccagcgtg	gtgaagggcc	ccggcttcac	4980
cggcggcgac	atcctgcgcc	gcaccagccc	cggccagatc	agcaccctgc	gcgtgaacat	5040
caccgcccc	ctgagccage	gctaccgcgt	ccgcataccg	tacgccagca	ccaccaacct	5100
gcagttccac	accagcatcg	acggccgcgc	catcaaccag	ggcaacttca	gcgccaccat	5160
gagcagcggc	agcaacctgc	agagcggcag	cttccgcacc	gtgggcttca	ccacccccct	5220
caacttcage	aacggcagca	gcgtgttcac	cctgagcgcc	cacgtgttca	acagcggcaa	5280
cgaggtgtac	atcgaccgca	togagttcgt	gcccgcgcag	gtgaccttcg	aggccagagta	5340
cgacctggag	aggggtcaga	aggccgtgaa	cgagctgttc	accagcagca	accagatcgg	5400
cctgaagacc	gacgtgaccg	actaccacat	cgatacaggtg	taggagctcg	ttcgctgggg	5460
gaactcatca	ggaaggetgc	tgcctctctt	gcagccttgc	tcctggctgc	cgcgcgtgtc	5520
gtggtctgct	ctttcaagtc	gaagtaacgg	tggttcgagc	tagtggtatg	tgtggtctaa	5580
ctgtagaagt	tccttttgta	tagcaagcaa	gtaaaaaaaa	aaaatgacca	aaaaatataa	5640
caaaatgcag	ctgtaagttt	actgctgctc	tctaagtcgt	gttcagtcac	ccagtggtgc	5700
tagtctaggg	aaaccccata	aaaatgggtga	aggtggaatc	ccatcccagt	gtcataatta	5760
aggatgcact	tcttctgtaa	gcaaagtgtat	gatgtacaat	ggccggcccg	cagtctaaat	5820
gttacaacta	gtctctcttg	gtgaattcac	cggtcacac	tgatgtgctg	ctatgtatca	5880
ttactatcca	gttagggctt	gttcgggttat	tcctacgcca	tatggattgg	acgggattgg	5940
aaaatttttag	tagacatttt	gacttctatg	gatttaaacc	cacccaatcc	cctccaatcc	6000
acatggaltg	agatgaaacc	gaacaagccc	ttagttggat	ggatggatga	tctcacgcgc	6060
ttgagtttat	ctagttgctg	atatgggaga	gcccctcaac	acctcaaata	ttgtggataa	6120
gtttaatcct	acactgtcag	tcttcagtta	taaggcacgc	actcttcgac	gttgggcgct	6180
gtcttttggt	tcaaagggtg	aggcaacctc	aacctcgttt	aaaacagaga	gcaaagctaa	6240

tttccaaact gatgtaagtc atgtgtcctt aattaccaa gtagcaatga tgacattggt	6300
atgttgtgtg ctaatgagcc attacgtgaa catcagttcc tgccttgccg cttcgttcag	6360
ccggggccggg gtatatTTTT agtcactaaa aggacacgtc ggatgaatgg aaaattgcct	6420
cttgtcttta ccaagagtgt cggccgagcg cggaccgcga tcgcttaatt aagcttgcac	6480
gcctgcagtg cagcgtgacc cggtcgtgcc cctctctaga gataatgagc attgcatgtc	6540
taagttataa aaaattacca catatTTTT ttgtcacact tgtttgaagt gcagtttata	6600
tatctttata catatatTTT aactttactc tacgaataat ataactata gtactacaat	6660
aatatcagtg ttttagagaa tcatataaat gaacagttag acatgggtcta aaggacaatt	6720
gagtattttg acaacaggac tctacagttt tatcttttta gtgtgcatgt gttctccttt	6780
ttttttgcaa atagcttcac ctatataata cttcatccat tttattagta catccattta	6840
gggttttaggg ttaatggttt ttatagacta attttttttag tacatctatt ttattctatt	6900
ttagcctcta aattaagaaa actaaaactc tatttttagtt tttttattta ataatttaga	6960
tataaaatag aataaaataa agtgactaaa aattaaacaa atacccttta agaaattaaa	7020
aaaactaagg aaacatTTTT cttgtttcga gtagataatg ccagcctgtt aaacgcgcgc	7080
gacgagtcta accgacacca accagcgaac cagcagcgtc gcgtcggggc aagcgaagca	7140
gacggcacgg catctctgtc gctgcctctg gacccctctc gagagtcccg ctccaccgtt	7200
ggacttgctc cgctgtcggc atccagaaat tgcgtggcgg agcggcagac gtgagccggc	7260
acggcaggcg gcctcctcct cctctcacgg caccggcagc tacgggggat tcccttccca	7320
ccgctccttc gctttccctt cctcgccgcg cgtaataaat agacaccccc tccacacct	7380
ctttcccaa cctcgtgttg ttcgagcgc acacacacac aaccagatct ccccaaatc	7440
caccgcgcgg cacctcgct tcaaggtacg ccgctcgtcc tcccccccc cccctctcta	7500
ccttctctag atcggcgctt cgggtccatg ttagggcccg gtagttctac ttctgttcat	7560
gtttgtgtta gatccgtgtt tgtgttagat ccgtgctgct agcgttcgta cacggatgcg	7620
acctgtacgt cagacacgtt ctgattgcta acttgccagt gtttctcttt ggggaatcct	7680
gggatggctc tagccgttcc gcagacggga tcgatttcat gatttttttt gtttcggttc	7740
atagggtttg gtttgccctt ttcttttatt tcaatatatg ccgtgcactt gtttgcggg	7800
tcactttttc atgctttttt ttgtcttggg tgtgatgatg tgggtctggt gggcggtcgt	7860
tctagatcgg agtagaatc tgtttcaaac tacctggtgg atttattaat ttggatctg	7920
tatgtgtgtg ccatacatat tcatagttac gaattgaaga tgatggatgg aaatatcgat	7980
ctaggatagg tatacatgtt gatgcggggt ttactgatgc atatacagag atgctttttg	8040
ttcgcttggg tgtgatgatg tgggtgtggt gggcggtcgt tcattcgctc tagatcggag	8100
tagaatactg tttcaaacta cctggtgtat ttattaattt tggaactgta tgtgtgtgtc	8160
atacatcttc atagttacga gtttaagatg gatggaaata tcgatctagg ataggtatac	8220
atgttgatgt ggggttttact gatgcatata catgatggca tatgcagcat ctattcatat	8280
gctctaacct tgagtaccta tctattataa taaacaagta tgttttataa ttattttgat	8340
cttgatatac ttggatgatg gcatatgcag cagctatatg tggatttttt tagccctgcc	8400

ttcatacgcct	atattatttgc	ttgggtactgt	ttctttttgtc	gatgctcacc	ctgttggtttg	8460
gtgttacttcc	tgcagggatc	cccgatcatg	caaaaactca	ttaactcagt	gcaaaaactat	8520
gcctgggggca	gcaaaaacggc	gttgactgaa	ctttatggta	tggaaaatcc	gtccagccag	8580
ccgatggccg	agctgtggat	gggcgcacat	ccgaaaagca	gttcacgagt	gcagaatgcc	8640
gccggagata	tcgttttact	gcgtagatgtg	attgagagtg	ataaatcgac	tctgctcgga	8700
gaggccggtt	ccaaacgctt	tggcgaactg	ccttttctgt	tcaaagtatt	atgcgcagca	8760
cagccactct	ccattcaggt	tcattccaaac	aaacacaatt	ctgaaatcgg	ttttgccaaa	8820
gaaaatgccg	caggtatccc	gatggatgcc	gccgagcgta	actataaaga	tcctaaccac	8880
aagccggagc	tgggtttttgc	gctgacgcct	ttccttgcca	tgaacgcgtt	tcgtgaattt	8940
tccgagattg	tctccctact	ccagccggtc	gcaggtgcac	atccggcgat	tgctcacttt	9000
ttacaacagc	ctgatgccga	acgttttaagc	gaactgttcg	ccagcctgtt	gaatatgcag	9060
ggtgaagaaa	aatcccgcgc	gctggcgatt	ttaaaatcgg	ccctcgatag	ccagcagggg	9120
gaaccgtggc	aaacgattcg	tttaattttct	gaattttacc	cggaagacag	cggtctgttc	9180
tccccgctat	tgctgaatgt	ggtgaaattg	aaccctggcg	aagcgatgtt	cctgttcgct	9240
gaaacaccgc	acgcttacct	gcaaggcggtg	gcgctggaag	tgatggcaaa	ctccgataac	9300
gtgctgcgtg	cgggtctgac	gcctaaatac	attgatattc	cggaactggg	tgccaatgtg	9360
aaattcgaag	ccaaaccggc	taaccagttg	ttgaccacgc	cggtgaaaca	aggtgcagaa	9420
ctggacttcc	cgattccagt	ggatgatattt	gccttctcgc	tgcatgacct	tagtgataaa	9480
gaaaccacca	ttagccagca	gagtgcgcgc	attttgttct	gcgtcgaagg	cgatgcaacg	9540
ttgtggaaaag	gttctcagca	gttacagctt	aaaccgggtg	aatcagcggt	tattgccgcc	9600
aacgaatcac	cggtgactgt	caaaggccac	ggccgttttag	cgcggtgttta	caacaagctg	9660
taagagctta	ctgaaaaaat	taacatctct	tgctaagctg	ggagctcgtc	atgggtcgtt	9720
taagctgccg	atgtgcctgc	gtcgtctggg	gccctctctc	catatggagg	ttgtcaaagt	9780
atctgctgtt	cgtgtcatga	gtcgtgtcag	tggttggttta	ataatggacc	ggttgtgttg	9840
tgtgtgcgtg	ctaccagaa	ctatgacaaa	tcataaataa	gtttgatgtt	tgaaattaaa	9900
gcctgtgctc	attatgttct	gtctttcagt	tgtctcctaa	tatttgccctg	caggtactgg	9960
ctatctaccg	tttcttactt	aggaggtgtt	tgaatgcact	aaaactaata	gttagtggtc	10020
aaaattagtt	aaaacatcca	aacaccatag	ctaatagttg	aactattagc	tattttttgga	10080
aaattagtta	atagtggagg	agttatttgt	tagctagcta	attcaactaa	caattttttag	10140
ccaactaaca	attagtttca	gtgcattcaa	acacccctt	aatgttaacg	tggttctatc	10200
taccgtctcc	taatatatgg	ttgattgttc	ggtttggttg	tatgctattg	ggttctgatt	10260
gctgctagtt	cttgcgtgaat	ccagaagttc	tcgtagtata	gctcagattc	atattattta	10320
tttgagtgat	aagtgatcca	ggttattact	atgttagcta	ggtttttttt	acaaggataa	10380
attatctgtg	atcataatcc	ttatgaaagc	tttatgtttc	ctggaggcag	tggcatgcaa	10440
tgcatgacag	caacttgatc	acaccagctg	aggtagatac	ggtaacaagg	ttcttaaatac	10500
tgttcaccaa	atcattggag	aacacacata	cacattcttg	ccagtcttgg	ttagagaaat	10560

ttcatgacaa aatgccaaag ctgtcttgac tcttcacttt tggccatgag tegtgactta	10620
gttttggttta atggaccggt tctcctagct tgttctactc aaaactgttg ttgatgcgaa	10680
taagttgtga tggttgatct ctggatTTTT ttttgctctc aatagtggac gagattagat	10740
agccccgaaa tttaccggtg cccgggcggc cagcatggcc gtatccgcaa tgtgttatta	10800
agttgtctaa gcgtcaatTT gtttacacca caatatatcc tgccaccagc cagccaacag	10860
ctccccgacc ggcagctcgg cacaaaatca ccaactcgata caggcagccc atcagaatta	10920
attctcatgt ttgacagctt atcatcgact gcacggtgca ccaatgcttc tggcgtcagg	10980
cagccatcgg aagctgtggt atggctgtgc aggtcgtaaa tcaactgcata attcgtgtcg	11040
ctcaaggcgc actcccgctc tggataatgt tttttgcgcc gacatcataa cggttctggc	11100
aaatattctg aaatgagctg ttgacaatta atcatccggc tcgtataatg tgtggaattg	11160
tgagcggata acaatttcac acaggaaaca gaccatgagg gaagcgttga tcgccgaagt	11220
atcgactcaa ctatcagagg tagttggcgt catcgagcgc catctcgaac cgacgttgct	11280
ggcgtacat ttgtacggct ccgcagtgga tggcggcctg aagccacaca gtgatattga	11340
tttgctggtt acggtgaccg taaggcttga tgaaacaacg cggcgagctt tgatcaacga	11400
ccttttgga aacttcggctt cccctggaga gagcgagatt ctccgcgctg tagaagtcac	11460
cattgttgtg cacgacgaca tcattccgtg gcgttatcca gctaagcgcg aactgcaatt	11520
tggagaatgg cagcgcaatg acattcttgc aggtatcttc gagccagcca cgatcgacat	11580
tgatctggct atcttgctga caaaagcaag agaacatagc gttgccttgg taggtccagc	11640
ggcggaggaa ctctttgatc cggttcctga acaggatcta tttgaggcgc taaatgaaac	11700
cttaacgcta tggaactcgc cgcgcgactg ggctggcgat gagcgaaatg tagtgcttac	11760
gttgctccgc attttggtaca gcgcagtaac cggcaaaatc gcgcgaagg atgtcgctgc	11820
cgactgggca atggagcgcc tgccggccca gtatcagccc gtcatacttg aagctaggca	11880
ggcttatctt ggacaagaag atcgcttggc ctccgcgcga gatcagttgg aagaatttgt	11940
tcaactacgtg aaaggcgaga tcaccaaagt agtcggcaaa taaagctcta gtggatctcc	12000
gtaccggggg atctggctcg cggcggacgc acgacgcggg ggcgagacca taggcgatct	12060
cctaaatcaa tagtagctgt aacctogaag cgtttcactt gtaacaacga ttgagaattt	12120
ttgtcataaa attgaaatac ttggttcgca tttttgtcat ccgcggtcag ccgcaattct	12180
gacgaactgc ccatttagct ggagatgatt gtacatcctt cacgtgaaaa tttctcaagc	12240
gctgtgaaca agggttcaga ttttagattg aaaggtgagc cgttgaaaca cgttcttctt	12300
gtcgatgacg acgtcgctat ggggcattctt attattgaat accttacgat ccacgccttc	12360
aaagtgaccg cggtagccga cagcacccag ttcacaagag tactctcttc cgcgacggtc	12420
gatgtcgtgg ttgttgatct agatttaggt cgtgaagatg ggctcgagat cgttcgtaat	12480
ctggcggcaa agtctgatat tccaatcata attatcagtg gcgaccgcct tgaggagacg	12540
gataaagttg ttgcaactga gctaggagca agtgatttta tcgctaagcc gttcagtatc	12600
agagagtttc tagcacgcac tcgggttgcc ttgcgcgtgc gccccaacgt tgtccgctcc	12660
aaagaccgac ggtctttttg ttttactgac tggacactta alctcaggca acgtcgcttg	12720

atgtccgaag	ctggcggtga	ggtgaaactt	acggcaggtg	agttcaatct	tctcctcgcg	12780
tttttagaga	aaccccgga	cgttctatcg	cgcgagcaac	ttctcattgc	cagtcgagta	12840
cgcgacgagg	aggtttatga	caggagtata	gatgtttctca	ttttgaggct	gcgccgcaaa	12900
cttgaggcag	atccgtcaag	ccctcaactg	ataaaaaacag	caagaggtgc	cggttatcttc	12960
tttgacgcgg	acgtgcaggt	ttcgcacggg	gggacgatgg	cagcctgagc	caattcccag	13020
atccccgagg	aatcggcggtg	agcggtcgca	aaccatccgg	cccgggtacaa	atcggcgcg	13080
cgctgggtga	tgacctggtg	gagaagttga	aggccgcgca	ggccgcccag	cggcaacgca	13140
tcgaggcaga	agcacgcccc	ggtgaatcgt	ggcaagcggc	cgctgatcga	atccgcaaa	13200
aatccccgga	accgcgggca	gccgggtgcg	cgtcgattag	gaagccgccc	aagggcgacg	13260
agcaaccaga	ttttttcgtt	ccgatgctct	atgacgtggg	caccgcgat	agtcgcagca	13320
tcattggaagt	ggccgttttc	cgtctgtcga	agcgtgaccg	acgagctggc	gaggtgatcc	13380
gctacgagct	tccagacggg	cacgtagagg	tttcgcgagg	gccggccggc	atggccagtg	13440
tgtgggatta	cgacctggta	ctgatggcgg	tttcccatct	aaccgaatcc	atgaaccgat	13500
accgggaagg	gaagggagac	aagccccggc	gcgtgttccg	tccacacgtt	gcggacgtac	13560
tcaagttctg	ccggcgagcc	gatggcggaa	agcagaaa	cgacctggta	gaaacctgca	13620
ttcgggttaa	caccacgcac	gttgccatgc	agcgtacgaa	gaaggccaag	aacggccgcc	13680
tgggtgacgg	atccgagggg	gaagccttga	ttagccgcta	caagatcgta	aagagcgaaa	13740
ccgggcggcc	ggagtacatc	gagatcgagc	tagctgattg	gatgtaccgc	gagatcacag	13800
aaggcaagaa	cccgacgtg	ctgacggttc	accccgatta	ctttttgatc	gatccccgca	13860
tcggcggttt	tctctaacgc	ctggcacgcc	gcgcgcgagg	caaggcagaa	gccagatggt	13920
tgttcaagac	gatctacgaa	cgcagtggca	gcgcgcgaga	gttcaagaag	ttctgtttca	13980
ccgtgcgcaa	gctgatcggg	tcaaattgacc	tgccggagta	cgatttgaag	gaggaggcgg	14040
ggcaggctgg	ccgatcccta	gtcatgcgct	accgcaacct	gatcgagggc	gaagcatccg	14100
ccggttcccta	atgtaaggag	cagatgctag	ggcaaattgc	cctagcaggg	gaaaaagggtc	14160
gaaaaagggt	ctttcctgtg	gatagcacgt	acattgggaa	cccaaagccg	tacattggga	14220
accggaaccc	gtacattggg	aacccaaagc	cgtacattgg	gaaccgggtca	cacatgtaag	14280
tgactgatata	aaaagagaaa	aaaggcgatt	tttcgcgcta	aaactcttta	aaacttatta	14340
aaactcttaa	aaccgcgctg	gcctgtgcat	aactgtctgg	ccagcgcaca	gccgaagagc	14400
tgcaaaaagc	gcctaccctt	cggtcgctgc	gctccctacg	ccccgcgct	tcgcgtcggc	14460
ctatcgcggc	cgttggccgc	tcaaaaatgg	ctggcctacg	gccaggcaat	ctaccagggc	14520
goggacaagc	cgcgcgcgtg	ccactcgacc	gccggcgctg	aggtctgcct	cgtgaagaag	14580
gtgttgctga	ctcataccag	gcctgaatcg	ccccatcctc	cagccagaaa	gtgagggagc	14640
caagggtgat	gagagctttg	ttgtagggtg	accagttggg	gattttgaac	ttttgctttg	14700
ccacggaacg	gtctgogttg	tcgggaagat	gcgtgatctg	atccttcaac	tcagcaaaa	14760
ttcgatttat	tcaacaaagc	cgcgctcccg	tcaagtcagc	gtaatgctct	gccagtggtta	14820
caaccaatta	accaattctg	attagaaaaa	ctcatcgagc	atcaaatgaa	actgcaattt	14880

attcatatca ggattatcaa taccatattt ttgaaaaagc cgtttctgta atgaaggaga 14940
 aaactcaccg aggcagttcc ataggatggc aagatcctgg tatcgggtctg cgattccgac 15000
 tcgtccaaca tcaatacaac ctattaattt cccctcgtca aaaataaggt tatcaagtga 15060
 gaaatcacca tgagtgacga ctgaatccgg tgagaatggc aaaagctctg cattaatgaa 15120
 tcggccaacg cgcggggaga ggcggtttgc gtattgggag ctcttcgct tctcgtctca 15180
 ctgactcgtt gcgctcggtc gttcgggtgc ggcgagcggc atcagctcac tcaaaggcgg 15240
 taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc 15300
 agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc 15360
 cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac 15420
 tataaagata ccaggcgttt cccctggaa gctccctcgt gcgctctcct gttccgacct 15480
 tgccgcttac cggatacctg tccgcctttc tcccttcggg aagcgtggcg ctttctcata 15540
 gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg ggctgtgtgc 15600
 acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca 15660
 acccggttaag acacgaacta tcgccactgg cagcagccac tggtaacagg attagcagag 15720
 cgaggatatgt aggcgggtgt acagagttct tgaagtgggt gcctaactac ggctacacta 15780
 gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg 15840
 gtagctcttg atccggcaaa caaacccacg ctggtagcgg tggttttttt gtttgcaagc 15900
 agcagattac gcgcagaaaa aaaggatctc aagaagatcc ttgatcttt tctacgggg 15960
 ctgacgtctc gtggaacgaa aactcacgtt aagggatatt ggtcatgaga ttatcaaaaa 16020
 ggatcttcac ctagatcctt ttgatccgga atta 16054

<210> 28

<211> 22075

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Vectơ tái tổ hợp

<400> 28

attcctgtgg ttggcatgca catacaaatg gacgaacgga taaacctttt cagccctttt 60
 taaatatccg attattctaa taaacgtctt tttctcttag gtttaccgc caatatatcc 120
 tgtcaaacac tgatagttta aactgaaggc gggaaacgac aatctgatca tgagcggaga 180
 attaagggag tcacgttatg acccccgcgc atgacgcggg acaagccggt ttacgtttgg 240
 aactgacaga accgcaacgc tgcaggaatt ggccgcagcg gccatttaaa tcaattgggc 300
 gcgcacgctg cttgtgggga ccagacaaaa aagggaatgt gcagaattgt taggcgcacc 360
 taccaaaagc atctttgcct ttattgcaaa gataaagcag attcctctag tacaagtggg 420
 gaacaaaata acgtggaaaa gagctgtcct gacagcccac tcaactaatgc gtatgacgaa 480

cgcagtgacg accacaaaac tcgagacttt tcaacaaagg gtaatatccg gaaacctcct	540
cggattccat tgcccagcta tctgtcactt tattgtgaag atagtggaaa aggaaggtgg	600
ctcctacaaa tgccatcatt gcgataaagg aaaggctatc gttgaagatg cctctgccga	660
cagtgggtccc aaagatggac cccacccac gaggagcatc gtggaaaaag aagacgttcc	720
aaccacgtct tcaaagcaag tggattgatg tgatatctcc actgacgtaa gggatgacga	780
acaatcccac tatccttcgg taccggaccc tatagaatag ctactatcc tatttattat	840
agtttaagta tatagccaat attttaaatt tactatttat taaattctag ggaagatagt	900
ctcaattcat aactttatta taatacgttt gaaattttta atcttttagga aattttctta	960
attcacctag atacgattct ggagtgttac aagctgcgaa tatactgggtg ccattgagta	1020
tacataaatg gatttaggtg gtgctcaata ggtgaaaatg agatactaact cacttaaatt	1080
tcaaaatttc tatggtgccca ctgtactcgg atagggtctat ctagggtcgg acaaaatgct	1140
cgtggctcgc tggctcgcgc gtttcgtggg cagctcggct cggctcggat cggctcattt	1200
gaattttgtc acgagctgag ctgacattct agctcggttc gttaacgagc cagctcgcga	1260
gctaaacgag ctaccatatt ctagtaaaac gaaattatat tcataatcatt tatagaataa	1320
ttgatgaaca tgttatatat atgtgagatg tctatggcct atgaattaaa ctaatgatta	1380
atgaactatg cctatgtgtt aatttgggtc atgcaaatat aattatgggt taaactgatg	1440
aacatgcatg tgaatttgtga attaatgagt gatgaattgt gctaatttgg tgttatatgt	1500
acatggtttg tgaaactatg agtataatta ctattttcta ttgttaaatt agtttgaaat	1560
taactaaaaa ataattatta tatacatttt attttttttc tgcctcggct cgcgagctaa	1620
acgagccagc tcgacctcgt aaacgagccg agccgagctg actctgtggc tcgttacctt	1680
aacgagccga gccgagctgg ctcgtagct taacgagcca gctcgaactc ggacgagccg	1740
agccgagctg gctcgttata caccctagg tctatctage ttctgatgtt tgcaaacctt	1800
agagttggag tgttcagcca gctactcctt tgctttgctg aataaccata ccaaacacgc	1860
ccatattaat acccgctcgg cgggtggttct gcaatcaaac gcaggccgca gtcgcgtgcg	1920
gaactagagg tccttcagag aagtgccgtg ccagtgccac cgccggccgc atcatcgttc	1980
cgccccctg gtacgagcac ttgcgagagc tgcaacctac atccctttta cataaatcta	2040
ttgtctcgta ttgccgttga cgccggaata gtcttcgcat cctttttaca taaatccgat	2100
gttttctttc tccgattcct ttgaggaatc atcacgggtc agggcaggtg ttctgccgtt	2160
tgcccttttc tttatattct ccttagaaga aatatttagt tggaggctgg acatagccgg	2220
aggagctaac taatcgagcg gtgtactggc aaaacaaaag gagcggagca agaaagggga	2280
gaaaaaacta gccactgccg gagcgctatt ggccgtgttg ggccgtggaag cttgcatcaa	2340
tacttcctc gccccgattt ggttcacaaa tcatacaagt cccaaagttg tcaagatatt	2400
ggaggatatc aagcgacttg gatctcaaaa tagaagaaat ttoggatctg agcacaaatc	2460
tgagttgaaa aaactgcaac tcaaaatcat caaaaaaaga agaagaaaga aacgaatata	2520
ttcgctcctc ttctcagccg aacccaaagg aattgaatcc aaacctggg taggcagaca	2580
gtgagatatg gaggagagca ggaggcgaac aagagaggct gcggccacga atatctcacg	2640

aacaagcaca tcatgggtcc acggagcggg caggggtgacg ggctcccgcac ggcgagctac 2700
atctcgggaag agcaccaggg cagcatgtcg tgttgggacg gttggccgctc tggcggacgg 2760
cggacgggtga ctcgtgggtca ggggtgcacct gctcgattaa ggcgcctgac tactcatgtc 2820
ttgggtctctt tgcttgtgtt tgctatatgc tgctcgtacc tcatgagcat actaagttga 2880
ctgctcagtc tgctgagtct gtttttctag ggtatagtgc tgagcacaag ggatatcatt 2940
gttgggatat gattgctcgt tggatgaggg tctcttggga tgttgtcttt gatgaggctc 3000
attcttttta ttcttgtcct tctttcgatg ctttgtcaac atccttgggt gatcccatct 3060
cttttctata ttttctagat gcccggtgtta ctattggacc tgcttcacgc ttgggtgcgc 3120
cacgatagta gccttagctc cttctgacat gttcatctct ctttcgggtgc cttcctttgt 3180
gggtgccttct atagtgtttt ctttggagcc tgctgcttta gccctgact acgctatgaa 3240
cacttgtcta caccgcgcgg gtcataaatt cttttggtac accatcatcc tctcatgcgt 3300
tgccctctta tgatgtgcgc tcttctgcaa ctattcatt ttcttgcgat ttacctttga 3360
ctgatgctcc ctattcatct ctggatccag cttcctcagt tgactctttg ctggagccac 3420
ctcttagacg gagtcacgt tttcgtcagc cacctaattg gtactctcct tcaggtttag 3480
tcgctaccgt tctttctgag ctgacttctt atcatgatgc tattcttcat ctgtaacgac 3540
aacatgcgat ttctgaggag attgctactc ttgagcgcac tagcacgttg gaacttgttc 3600
cttgtccatc acgtgtttgt cctatcacca gtatgtgggt ctataaggtc aagaccggtt 3660
ctgatgggtc tcttgatcgc tataaatctc gtctagttgc ccaaggcttc cagtaggaac 3720
atggttgttg ctatgatgag atttttgcac ctgttgcctc tatgaccact gttcgcactc 3780
ttcttgcctat ggcctctgtt cgtgcgtggg ccatctctca tcttgatgtc aagaatacct 3840
ttcttgatgg taagctactt gagttctata tgtagccatc gcctaggtat tctatttctg 3900
cttgtatggg ttgttgtctt cgcggttccc cttatggcct caagcaggct ccacattctt 3960
ggtttcagct ctttgcctct atgataactg ttgttgggtt ttctaccagt aatcatggtc 4020
ctgcactctt tgtgtaactc ctcctctcgg ggtcggactc ttctttatgt tgatgatata 4080
attatcactg gagataacct tgagtatgtt gactttgtta aggcacgtct tagttatcat 4140
tttctcatgt ctgatcttgg tctctgtgt tactttcttg ggacaaagggt ttcttctttg 4200
tctcagggcc tttatctatc tcaagaggag tacattcaag attttcttca tcgggcttct 4260
cttaccgac actagattgt tgagactccc aagcagctca atcttcacct tagtgccgat 4320
gatggcgagt cttttcccga ccatactcgt tatcgtcaac atactgtagg aagtittgtt 4380
tatctctgtg tcaactcgtc tgacatttca tatgttgtgt gtatcctgag ttagtttgc 4440
tcagatccca tccagggtaca ctatagtcac ttgctttgtg tctacaata tctttgtgga 4500
accatatcta gatgtatgtt ctttccacat tctagctcgt tgcaactgca atcttgttct 4560
gatgctactt gggtatgtga tttttctgat agttgggtctc tttctcaata ttgtgttttt 4620
cttgggtggt ctctcattgc tcggaagact aagtagcagg tagcagtttc tcgtttgagt 4680
accgaggctg agttgcgtgc tatggccctt gtgactgcag aggttacttg gttacgatag 4740
ttgcttgagg attttcatgt ttctgtttcc atgacgactc ctttltgtctg acagtacagg 4800

tgttatcagt attgctcgtg atgcggtgaa gcatgaggtc accaagcata ttggagttga	4860
tgtttcgtat acacgagctg aagtctagga tgatgttata ttgatttggt atgtgccttt	4920
agagcttcag ttggctaatt tcttcacgag ggcacaggct cgcgctgagc ataaatTTTT	4980
cctctcaaaa ctcaagtgtta tagatccacc ttgagtttga gggagtatta gatagatatg	5040
ggtttatttg tattttttcca ttttataagg gtattagata gataggcaac gactgctatg	5100
caagtagtca ttctgtgcaa gcgtgcaagc aaaccatctg atccattata tcgtgatcca	5160
accgtgggtc acatttaaca cttaaaccct tccaccacca actcaataat ctttataaaa	5220
aaacccttaa caaacaatgg ttatatctgt ggttggatcg taatctaata gatcagatgg	5280
tttgcttgta cgcttgca caaatgactgc ttgcatagca gttgttgctt agatagatat	5340
gggtttatTT gtatttttct cttaaggggt tttgtgtata tttgtactca tgtacctata	5400
tatttTgtgt agttgacccc ataatgaata gacctgctat tcataatatt tgcaaaccat	5460
gaaaatttga ttattacgaa ctatccaaat actcgaacac atgggcatta tagctcacia	5520
aaatggaagg ttgagctgct gcttgaagaa cctcaacatc tttgaacaac aacctcaacg	5580
aaacttTtat atgaaccaac ttccaaacaa tcccttTgtg aaggatagta atgacttcag	5640
ggcattgata acacatatcc gacgggtggaa ctactgtaac aacctctttt tctgtggaat	5700
atagttgaaa ctctacaact tgaccaaaac caagatgacg acatatgggtg gaactaacia	5760
aacaagagga ctacactacc tcatttagctt attaaacaca atctcttTgg accacaacia	5820
cgaacaacia aacctatcatt tggatgctct gtgggagact aaatgcaaat tctttgcatg	5880
gttgatcatc ccaaattTgt ggcacttagc tataggctag cagtgagagg atggccgaac	5940
aacatgcatt gtccactatg ttggtgtagc catgagacca accaccacat aaatgccaaa	6000
cgttcattca ccaaaaaaat ctaggcaaca atggcttTga tttcttacct gcagctccac	6060
caagctaact ggagttcaat taggtcaacg tatgggtTgt ggtcgagtat agcagtcaca	6120
aatgatgttc taaagatggg gttgtgttaa cacatcttgc ttgtagcacg agaactTgg	6180
aaggagtga accaaagaat ctttcaacac aaggacctat caacgctatc catgattTgg	6240
aaattcaagg acgaaactag aatttTgggtg aacacatgca caaggcacct aggagagcct	6300
ttcttttTga ctgttaatcc ctttttaaac tctctctgtc cttaggagtt cgtttcttcc	6360
gctctattca atgaagttag gcacaatctt gtgtgatttc attagaaaaa cacaagtaaa	6420
ttgcatgggc agtacttgaa gtattacagg aatctcgtct gccccaaaac tattaacct	6480
tatatTTggc tccctaattgt acttaactga tctcattctg gtcaaaactaa acatgggtgat	6540
ggcaaggagc cgatatgggc gcccatgtgg atgtgattta agcaaaaaat ctcatggTcc	6600
atagctgtgt caacaagcca acatgcoatc gcttccctat gccgagactg cccatgtcgc	6660
tcgcttttac tgtcatcatc atcaaaactgc ctgtcatgtc tacggatgcc atgaccgctg	6720
tcacacatga tgtggagatg aacctgtcca tcaacttcca cgtgctgcc caatcgttag	6780
ctgacaccgt cttggTcatt gctgtgtagg gctaggctaa gagtcgctga atgatccttt	6840
cgtctctctt tacaggaaca tgcTgtttac tttgtgtcgc caaggcgtgc tagagtacct	6900
cttctacacc tccagcacca gtagccttat tgttagcttg cacatcccac ataagcaggc	6960

cgatgtgaat gataacttca gggacgtcga cggcatgtca ctgccaagag tcatttggtg	7020
ggaagcggtt tcatgccatc tgtcgtgcc ttttgtcctc agttcgaccg ccattaccgt	7080
gagcacaacc tttgcgcatg gttggccgct tccatcaccc ttattccgtt tcctcgtgtt	7140
ggtcttgccc caaggctatg gtttagcagac cgtgcatatg gccggcaaaa gactattttg	7200
caactgtagat tgcactcttt atatagttaa gtttaaaata ggagatgaga tgaataaggc	7260
tgctggagat agcctaaacc cttgcagctc gtgcttgcat cgggggagcc aaaaggcgctc	7320
cacctccacc atcgccgaag cactgagcac tactctggct tgtgtttcag caccacaccg	7380
cagagtgtt agggccacca acctcctctt gcctctgtgc ccagagcacc atcagctctg	7440
ctgcctccct ctgttccttg tgcttgctag gcaggcaatt ccgagctggg gcccaacttg	7500
taacgctgat ttaccatct tgcactgcc gggcaccaag tggacacatt tgacttgccc	7560
tagtgggttt tctgcataaa tcacatacat gtggatgcc tatlaggctc tttggtgttg	7620
tcgtgtctac tttcgacaag gatgagatca cttaaacata ttagggagcc aagtatgtaa	7680
tttcatagtt tagggaccta cacaaaaatc gtataatact ttagaacagc cgtgcagttt	7740
actcaatcaa cacatacaaa gtcagatctt aagctctgat acttcaaagg aatgggtgag	7800
cccagttgac aaacaatctt gcttcattca ttgaattgtt tataggagtg gctatgtaac	7860
tactgggttg ttttgtttga cctgtcatcc aaattgtgta gtcaaccata aacatacacg	7920
tcacacaata catltttggat gtgacagata ggatttaggc gagagaatgt acaatgtcac	7980
tgaaaaatta ccaactgtatg gaaaggacaa tctaagttaa aagagaacca gggcctaattg	8040
gtttcaggac ttcaaactcc ggccaaatga atttacagtg cttaaattaa ctcattgtta	8100
tcattgatagc caaagcatgg gcaaaagaga aactatgaat aaatcgacaa tgtattctat	8160
atagcagtaa tataccatgt cagagcttt tacactaatg ggctgtattt ttctgcagtt	8220
attttaactg gcaatattct atgtcacagt aatatttgtt aaattttttc cagaatagca	8280
actgaactag aagtctagta tttcttaatt ggataacaaa aggaattagt gtgcatttgg	8340
cttacgaaca atcagtcacc caacattgaa tttgaagtcc tgtttcctct ttgttcagac	8400
gacactctcc aaatgaatgc cttatatatt gtgttgctcc tcttttctgc agagtgttca	8460
gtaactttct ccgatgtaaa ccattggaca caaccccaac atcaacagat gcatccctca	8520
caactgcctg agcaaccccg aggtggaggt gctggggcgc gagcgcatcg agaccggcta	8580
cacccccatc gacatcagcc tgagcctgac ccagttcctg ctgagcgagt tcgtgcccg	8640
cgcgggcttc gtgctgggccc tgggtggacat catctggggc atcttcggcc ccagccagtg	8700
ggacgccttc ctgggtgcaga tcgagcagct gatcaaccag cgcacgagag agttcgcccg	8760
caaccaggcc atcagccgcc tggagggcct gagcaacctg taccaaatct acgcccagag	8820
cttcgcgag tgggaggccg accccaccaa cccgcctctg cgcgaggaga tgcgcatcca	8880
gttcaacgac atgaacagcg cctgaccac cgcctatccc ctgttcgcgc tgcagaacta	8940
ccaggtgccc ctgctgagcg tgtacgtgca ggccgccaac ctgcacctga gcgtgctgcg	9000
cgacgtcagc gtgttcggcc agcgtgggg cttcgacgcc gccaccatca acagccgcta	9060
caacgacctg accgcctga tcggcaacta caccgaccac gcctgctgct ggtacaacac	9120

cggcctggag	cgcggtgtggg	gtccagacag	ccgcgactgg	atcaggtaca	accagttccg	9180
ccgcgagctg	accctgaccg	tgctggacat	cgtgagcctg	ttccccaact	acgacagccg	9240
cacctacccc	atccgcaccg	tgagccagct	gacccgcgag	atttacacca	accccggtgt	9300
ggagaacttc	gacggcagct	tccgcggcag	cgcccagggc	atcgagggca	gcattccgag	9360
ccccacctg	atggacatcc	tgaacagcat	caccatctac	accgacgccc	accgcggcga	9420
gtactactgg	agcgggccacc	agatcatggc	cagccccgtc	ggcttcagcg	gccccgagtt	9480
caccttcccc	ctgtacggca	ctatgggcaa	cgctgcacct	cagcagcgca	tcgtggcaca	9540
gctggggccag	ggagtgtacc	gcacctgag	cagcaccctg	taccgtcgac	ctttcaacat	9600
cggcatcaac	aaccagcagc	tgagcgtgct	ggacggcacc	gagttcgctt	acggcaccag	9660
cagcaacctg	cccagcgccg	tgtaccgcaa	gagcggcacc	gtggacagcc	tggacgagat	9720
ccccctcag	aacaacaacg	tgccacctcg	acagggcttc	agccaccgtc	tgagccacgt	9780
gagcatgttc	cgcagtggct	tcagcaacag	cagcgtgagc	atcatccgtg	cacctatgtt	9840
cagctggatt	caccgcagtg	ccgagttcaa	caacatcatt	cccagcagcc	agatcaccca	9900
gatccccctg	accaagagca	ccaacctggg	cagcggcacc	agcgtggtga	agggccccgg	9960
cttcaccggc	ggcgacatcc	tgcgcgcgac	cagccccggc	cagatcagca	ccctgcgcgt	10020
gaacatcacc	gccccctga	gccagcgcta	ccgcgtccgc	atccgctacg	ccagcaccac	10080
caacctgcag	ttccacacca	gcategacgg	ccgccccatc	aaccagggca	acttcagcgc	10140
cacctgagc	agcggcagca	acctgcagag	cggcagcttc	cgcaccgtgg	gcttcaccac	10200
ccccttcaac	ttcagcaacg	gcagcagcgt	gttcacctg	agcggccaacg	tgttcaacag	10260
cggcaacgag	gtgtacatcg	accgcacoga	gttcgtgccc	gccgaggtga	ccttcgaggg	10320
cgagtacgac	ctggagaggg	ctcagaaygc	cgtgaacgag	ctgttcacca	gcagcaacca	10380
gatcggcctg	aagaccgacg	tgaccgacta	ccacatcgat	caggtgtagg	agctcgacgt	10440
acaaatctca	tctgtgcctt	gctctagttt	cccaaattga	attaactatg	catgattttg	10500
ttggaaactc	ttattgcatt	catocagata	atgcattcac	cataaggtaa	tattttgatg	10560
acattctgtc	ctgatggtgt	accaaattgc	tctattctctg	cattgagcca	cagtagggag	10620
gatagccctg	gggtgccttg	actccaaagt	tgtattgaaa	aagatggatg	aagcaggcaa	10680
atgctgcctg	aatccatgac	tcagggcaca	gattttccac	tcaaagggaag	ataagattgc	10740
attacttcat	gatcttttga	actgcctctg	caagacggga	ctcggatagt	ggatgcaaag	10800
atctaatact	ggcctcaggg	aacgagttgt	ttcactcgaa	agtctagaaa	tgaccgggct	10860
caaattttgc	accccaaggga	aagtgagttt	gcattacttc	atgacctttt	gaactgcctc	10920
tgaagactg	gactcagatt	acgcttgatt	ggttgcgggc	ctcaccttcg	cctggcttgc	10980
gogagccctg	gtctatagaa	atgcgcggga	ctcacgtctc	cgtcgatgca	ggcattcgac	11040
tgaaaaaaca	tttaaactgc	acctatgcgt	gcgggctgag	cttatgtcat	acaagtaacc	11100
aatcacaggc	ttaagttcag	tcaacgcatt	cgttaagctt	ggatgtggct	gaccgggcaa	11160
ccaatcacac	agatagtggg	tgcacggatc	taatattggc	taatttggtt	aaacttgtct	11220
aaccttagac	gtggcaagtg	agtcagcgga	tcaaattctgc	tctaaaattg	tctgcctcct	11280

agatgtcctt ggtgttccaa gattttaalca tcaactgcact atttctttgc gttgcttcgc 11340
 tgcagcttcg cgttacttgc attcgcttaa tcaggattac tttgatcaac taggtttcta 11400
 acttctacta ccttcacttg cacaggggtgc ccgtcctgct agccgggtgtg cttgctgtgc 11460
 gatcgtttgg catgtgcttg ttgaggggtt gctaggggat tggagaggat tgaagggatt 11520
 aaatctcctc ctattcaatt ttgaatagga ggggatttaa tccccctcaa tccccctcaa 11580
 accactagta accgaacgtg gcctgagggg gcgggcgagt ctttatattg aatgaaacta 11640
 cataaaatag catgccgtct ctgtcactgg caatggacgg tgggtgcctag cgcaactcag 11700
 cgcacaactg tgtgtcttga tttttcttct gtttatcacg gcattagtgc catgccgttt 11760
 tatgttacag tgttgtgtgc tcgcaagcat ccgaaaatat gcgtctgagt ttagggttgg 11820
 gtcaaaacttg tcgaatttgg ggttctgtta taatatgttg agcatgaata aagatggatg 11880
 ctggtgactc tgtcgccatc gccgtccatc atgagtgtcc tgtaattcaa cttatatcta 11940
 tcatgtatgt atgtatgtat gtatgtatgt atgtatatgc tgtctactat gcttctttgt 12000
 ttttaactgaa atgtgtgtta cagtgttact tctctggggt ccatttaaaa cggcatttcg 12060
 tttacgatag gaaccagcca ttataatctt taaccaataa tttcgctaac caatttcaac 12120
 tattgcaatg cgaacttaat attatcagat ttataaccga atgcgctatc aaataatcat 12180
 aaggttgtaa tcataataat ataataaaa ataaatgagt gctcgaagtg aaattttaga 12240
 gagcgttata agaaaaattg atgtgatctc caagaataat agcccctccc ggctcccgggt 12300
 acaaacatag ggcttcttta gaatgcagga ttgtgagaac ataggaatag gaaaaatata 12360
 ggaattctat aggaatgtat atggaaaaca gaggattgaa aaacacagaa aaaatgtgaa 12420
 agcaagtctt tggatgaagc gtaggaaact tataggaata ggaattcata acggaccgcg 12480
 atcgcttaat taagcttgca tgcctgcagt gcagcgtgac ccggtcgtgc ccctctctag 12540
 agataatgag cattgcatgt ctaagttata aaaaattacc acatattttt tttgtcacac 12600
 ttgtttgaag tgcagtttat ctatctttat acatatattt aaactttact ctacgaataa 12660
 tataatctat agtactacaa taatatcagt gttttagaga atcatataaa tgaacagtta 12720
 gacatggctt aaaggacaat tgagtatttt gacaacagga ctctacagtt ttatcttttt 12780
 agtgtgcatg tgttctcctt tttttttgca aatagcttca cctatataat acttcatcca 12840
 ttttattagt acatccattt agggtttagg gttaatgggt tttatagact aattttttta 12900
 gtacatctat tttattctat tttagcctct aaattaagaa aactaaaact ctattttagt 12960
 ttttttattt aataatttag atataaaata gaataaaata aagtgactaa aaattaaaca 13020
 aatacccttt aagaaattaa aaaaactaag gaaacatttt tcttgtttcg agtagataat 13080
 gccagcctgt taaacgccgt cgacgagtct aacggacacc aaccagcgaa ccagcagcgt 13140
 cgcgtcgggc caagcgaagc agacggcacg gcatctctgt cgctgcctct ggacccctct 13200
 cgagagttcc gctccaccgt tggacttgct ccgctgtcgg catccagaaa ttgcgtggcg 13260
 gagcggcaga cgtgagccgg cacggcaggc ggctcctcc tctctcacg gcaccggcag 13320
 ctacggggga ttcctttccc accgctcctt cgctttccct tctcgcccg ccgtaataaa 13380
 tagacacccc ctccacaccc tctttcccca acctcgtgtt gttcggagcg cacacacaca 13440

caaccagatc	tcccccaaat	ccaccgcgtc	gcacctccgc	ttcaagggtac	gccgctcgtc	13500
ctcccccccc	ccccctctct	accttctcta	gatcggcgtt	ccggtccatg	gttagggccc	13560
ggtagttcta	cttctgttca	tgtttgtgtt	agatccgtgt	ttgtgttaga	tccgtgctgc	13620
tagcgttcgt	acacgggatgc	gacctgtacg	tcagacacgt	tctgattgct	aacttgccag	13680
tgtttctctt	tggggaatcc	tgggatggct	ctagccgttc	cgcagacggg	atcgatttca	13740
tgattttttt	tgtttcgttg	cataggggtt	ggtttgccct	tttcctttat	ttcaatatat	13800
gccgtgcact	tgtttgtcgg	gtcatctttt	catgcttttt	tttgtcttgg	ttgtgatgat	13860
gtgggtctgg	tgggcggtcg	ttctagatcg	gagtagaatt	ctgtttcaaa	ctacctgggt	13920
gattttattaa	ttttggatct	gtatgtgtgt	gccatacata	ttcatagtta	cgaattgaag	13980
atgatggatg	gaaatatcga	tctaggatag	gtatacatgt	tgatgcgggt	tttactgatg	14040
catatacaga	gatgcttttt	gttcgcttgg	ttgtgatgat	gtgggtgtgg	tgggcggtcg	14100
ttcattcgtt	ctagatcgga	gtagaatact	gtttcaaact	acctgggtga	tttattaatt	14160
ttggaactgt	atgtgtgtgt	catacatctt	catagttacg	agtttaagat	ggatggaaat	14220
atcgatctag	gataggtata	catgttgatg	tgggttttac	tgatgcatat	acatgatggc	14280
atatgcagca	tctattcata	tgctctaacc	ttgagtacct	atctattata	ataaacaagt	14340
atgttttata	attattttga	tcttgatata	cttggtgatg	ggcatatgca	gcagctatat	14400
gtggattttt	ttagccctgc	cttcatacgc	tattttattg	cttggtactg	tttcttttgt	14460
cgatgctcac	cctgtttgtt	ggtgttactt	ctgcagggat	ccccgatcat	gcaaaaactc	14520
attaactcag	tgcaaaaacta	tgccctggggc	agcaaaacgg	cgttgactga	actttatggt	14580
atggaaaatc	cgtccagcca	gccgatggcc	gagctgtgga	tgggcgcaca	tccgaaaagc	14640
agttcacgag	tgacagaatgc	cgcgggagat	atcgtttcac	tgcgatgatg	gattgagagt	14700
gataaatcga	ctctgctcgg	agaggccgtt	gccaaacgct	ttggcgaact	gcctttcctg	14760
ttcaaagtat	tatgcgcgag	acagccactc	tccattcagg	ttcatccaaa	caaacacaat	14820
tctgaaatcg	gttttgccaa	agaaaatgcc	gcaggatatcc	cgatggatgc	cgcggagcgt	14880
aactataaag	atcctaacca	caagccggag	ctgggtttttg	cgctgacgcc	tttccttgcg	14940
atgaacgcgt	ttcgtgaatt	ttccgagatt	gtctccctac	tccagccggg	cgcagggtgca	15000
catccggcga	ttgctcactt	tttacaacag	cctgatgccg	aacgtttaag	cgaactgttc	15060
gccagcctgt	tgaatatgca	gggtgaagaa	aaatcccgcg	cgctggcgat	tttaaaatcg	15120
gcctcgcata	gccagcaggy	tgaaccgtgg	caaacgattc	gtttaatttc	tgaattttac	15180
ccggaagaca	gcggctctgt	ctccccgcta	ttgctgaatg	tggtgaaatt	gaaccctggc	15240
gaagcgatgt	tectgttcgc	tgaaacaccg	cacgcttacc	tgcaaggcgt	ggcgctggaa	15300
gtgatggcaa	actccgataa	cgtgctgcgt	gcgggtctga	cgcctaaata	cattgatatt	15360
ccggaactgg	ttgccaatgt	gaaattcgaa	gccaaaccgg	ctaaccagtt	gttgaccacg	15420
ccggtgaaac	aaggtgcaga	actggacttc	ccgattccag	tggatgattt	tgccttctcg	15480
ctgcatgacc	ttagtataa	agaaaccacc	attagccagc	agagtgccgc	cattttgttc	15540
tgcgtcgaag	gcgatgcaac	gttgtggaaa	ggttctcagc	agttacagct	taaaccgggt	15600

gaatcagcgt	ttattgccgc	caacgaatca	ccggtgactg	tcaaaggcca	cggccgttta	15660
gcgcgtgttt	acaacaagct	gtaagagctt	actgaaaaaa	ttaacatctc	ttgctaagct	15720
gggagctcgt	catgggtcgt	ttaagctgcc	gatgtgcctg	cgtcgtctgg	tgccctctct	15780
ccatatggag	gttgtcaaa	tatctgctgt	tctgtcatg	agtcgtgtca	gtgttggttt	15840
aataatggac	cggttgtgtt	gtgtgtgcgt	actaccaga	actatgacaa	atcatgaata	15900
agtttgatgt	ttgaaattaa	agcctgtgct	cattatgttc	tgtctttcag	ttgtctccta	15960
atatttgcct	gcaggtactg	gctatctacc	gtttcttact	taggaggtgt	ttgaatgcac	16020
taaaactaat	agttagtggc	taaaattagt	taaaacatcc	aaacaccata	gctaatagtt	16080
gaactattag	ctatTTTTTg	aaaattagtt	aatagtgagg	tagttatttg	ttagctagct	16140
aattcaacta	acaatTTTTa	gccaactaac	aattagtttc	agtgcaattca	aacacccctt	16200
taatgttaac	gtggttctat	ctaccgtctc	ctaataatg	gttgattgtt	cggtttggtg	16260
ctatgctatt	gggttctgat	tgctgctagt	tcttgctgaa	tccagaagtt	ctcgtagtat	16320
agctcagatt	catattatTT	atttgagtga	taagtgatcc	aggttattac	tatgttagct	16380
aggTTTTTTT	tacaaggata	aattatctgt	gatcataatt	cttatgaaag	ctttatgttt	16440
cctggaggca	gtggcatgca	atgcatgaca	gcaacttgat	cacaccagct	gaggtagata	16500
cggtaacaag	gttctttaaT	ctgttcacca	aatcattgga	gaacacacat	acacattctt	16560
gccagtcttg	gttagagaaa	tttcatgaca	aaatgccaaa	gctgtcttga	ctcttcactt	16620
ttggccatga	gtcgtgactt	agtttggttt	aatggaccgg	ttctcctagc	ttgttctact	16680
caaaactgtt	gttgatgcga	ataagttgtg	atggttgatc	tctggatttt	gttttgctct	16740
caatagtgga	cgagattaga	tagcccgga	atttaccggt	gcccgggcgg	ccagcatggc	16800
cgtatccgca	atgtgttatt	aagttgtcta	agcgtcaatt	tgtttacacc	acaatatatc	16860
ctgccaccag	ccagccaaca	gctccccgac	cggcagctcg	gcacaaaatc	accactcgat	16920
acaggcagcc	catcagaatt	aattctcatg	tttgacagct	tatcatcgac	tgacgggtgc	16980
accaatgctt	ctggcgtcag	gcagccatcg	gaagctgtgg	tatggctgtg	caggctgtaa	17040
atcactgcat	aattcgtgtc	gctcaaggcg	cactcccgtt	ctggataatg	ttttttgcgc	17100
cgacatcata	acggttctgg	caaataattct	gaaatgagct	gttgacaatt	aatcatccgg	17160
ctcgtataat	gtgtggaatt	gtgagcggat	aacaatttca	cacaggaaac	agaccatgag	17220
ggaagcgttg	atcgccgaag	tatcgactca	actatcagag	gtagttggcg	tcacgcagcg	17280
ccatctcgaa	ccgacgttgc	tggccgtaca	tttgtacggc	tccgcagtgg	atggcggcct	17340
gaagccacac	agtgatattg	atttgctgg	tacggtgacc	gtaaggcttg	atgaaacaac	17400
gcggcgagct	ttgatcaacg	accttttgga	aacttcggct	tcccctggag	agagcgagat	17460
tctccgcgct	gtagaagtca	ccattgttgt	gcacgacgac	atcattccgt	ggcgttatcc	17520
agctaagcgc	gaactgcaat	ttggagaatg	gcagcgcaat	gacattcttg	caggtatctt	17580
cgagccagcc	acgatcgaca	ttgatctggc	tatcttgctg	acaaaagcaa	gagaacatag	17640
cgttgccctg	gtaggtccag	cggcggagga	actctttgat	cgggttctctg	aacaggatct	17700
atttgaggcg	ctaaatgaaa	ccttaacgct	atggaactcg	ccgcccgaact	gggctggcga	17760

tgagcgaaat gtagtgctta cgttgtcccg catttggtac agcgcagtaa ccggcaaaat	17820
cgcgccgaag gatgtcgctg ccgactgggc aatggagcgc ctgccggccc agtatcagcc	17880
cgtcatactt gaagctaggg aggccttatct tggacaagaa gatcgcttgg cctcgcgcgc	17940
agatcagttg gaagaatttg ttcactacgt gaaaggcgag atcaccaaag tagtcggcaa	18000
ataaagctct agtggatctc cgtaccggg gatctggctc gcggcgagc cagcagccg	18060
gggcgagacc ataggcgatc tctaaatca atagtagctg taacctcgaa gcgtttcact	18120
tgtacaacg attgagaatt ttgtcataa aattgaaata cttggttcgc atttttgtca	18180
tccgcggtca gccgaattc tgacgaactg ccattttagc tggagatgat tgtacatcct	18240
tcacgtgaaa atttctcaag cgctgtgaac aagggttcag attttagatt gaaaggtgag	18300
ccgttgaaac acgttcttct tgtcgatgac gacgtcgcta tgcggcatct tattattgaa	18360
taccttacga tccacgcctt caaagtgacc gcggtagccg acagcaccca gttcacaca	18420
gtactctctt ccgcgacggg cgatgtcgtg gttgttgatc tagatttagg tcgtgaagat	18480
gggctcgaga tcgttcgtaa tctggcgcca aagtctgata tccaatcat aattatcagt	18540
ggcgaccgcc ttgaggagac ggataaagtt gttgcactcg agctaggagc aagtgatttt	18600
atcgctaagc cgttcagtat cagagagttt ctagcacgca ttcgggttgc cttgcgcgtg	18660
cgccccaaac ttgtccgctc caaagacoga cggctttttt gttttactga ctggacactt	18720
aatctcaggg aacgtcgctt gatgtccgaa gctggcggtg aggtgaaact tacggcaggt	18780
gagttcaatc ttctctcgc gtttttagag aaaccccgcg acgttctatc gcgcgagcaa	18840
cttctcattg ccagtcgagt acgcgacgag gaggtttatg acaggagtat agatgttctc	18900
attttgaggg tcgcgcgcaa acttgaggca gatccgtcaa gccctcaact gataaaaaca	18960
gcaagaggtg ccggttatct ctttgacgcg gacgtgcagg ttccgcacgg ggggacgatg	19020
gcagcctgag ccaattccca gatcccgag gaatcggcgt gagcggtcgc aaaccatccg	19080
gcccggtaca aatcggcgcg gcgctgggtg atgacctggt ggagaagttg aaggccgcgc	19140
aggccgccc ggcgcaacgc atcgaggcag aagcacgccc cgggtgaatc tggcaagcgg	19200
ccgtgatcg aatccgcaaa gaatccggc aacgcgcggc agccggtgcg ccgtcgatta	19260
ggaagccgcc caaggcgac gagcaaccag attttttcgt tccgatgctc tatgacgtgg	19320
gcaccgcgca tagtcgcagc atcatggacg tggccgtttt ccgtctgtcg aagcgtgacc	19380
gacgagctgg cgaggtgatc cgctacgagc ttccagacgg gcacgtagag gtttcgcag	19440
ggccggccgg catggccagt gtgtgggatt acgacctggt actgatggcg gtttcccatc	19500
taaccgaatc catgaacga taccgggaag ggaagggaga caagcccgcc cgcgtgttcc	19560
gtccacacgt tcgggacgta ctcaagttct gccggcgagc cgatggcgga aagcagaaag	19620
acgacctggt agaaacctgc attcggttaa acaccacgca cgttgccatg cagcgtacga	19680
agaaggccaa gaacggccgc ctggtgacgg tatccgaggg tgaagccttg attagccgct	19740
acaagatcgt aaagagcgaa accgggcggc cggagtacat cgagatcgag ctagctgatt	19800
ggatgtaccg cgagatcaca gaaggcaaga acccgacgt gctgacgggt caccocgatt	19860
actttttgat cgatcccgcc atcgccggtt ttctctaccg cctggcacgc cgcgccgag	19920

gcaaggcaga	agccagatgg	ttgttcaaga	cgatctacga	acgcagtggc	agcgccggag	19980
agttcaagaa	gttctgtttc	accgtgcgca	agctgatcgg	gtcaaataac	ctgccggagt	20040
acgatttgaa	ggaggaggcg	gggcaggctg	gcccgatcct	agtcatgcgc	taccgcaacc	20100
tgatcgaggg	cgaagcatcc	gccggttcct	aatgtacgga	gcagatgcta	gggcaaattg	20160
ccctagcagg	ggaaaaaggt	cgaaaaggtc	tctttcctgt	ggatagcacg	tacattggga	20220
acccaaagcc	gtacattggg	aaccggaacc	cgtacattgg	gaacccaaag	ccgtacattg	20280
ggaaccggtc	acacatgtaa	gtgactgata	taaaagagaa	aaaaggcgat	ttttccgcct	20340
aaaactcttt	aaaacttatt	aaaactctta	aaaccgcctt	ggcctgtgca	taactgtctg	20400
gccagcgcac	agccgaagag	ctgcaaaaag	cgctaccctt	tccgtcgtcg	cgctccctac	20460
gccccgccgc	ttcgcgctcg	cctatcgcg	ccgctggccg	ctcaaaaatg	gctggcctac	20520
ggccaggcaa	tctaccaggg	cgcggaaca	ccgcgccgtc	gccactcgac	cgccggcgct	20580
gaggtctgcc	tctgtaagaa	ggtgttgctg	actcatacca	ggcctgaatc	gccccatcat	20640
ccagccagaa	agtgagggag	ccacggttga	tgagagcttt	gttgtaggtg	gaccagttgg	20700
tgattttgaa	cttttgcttt	gccacggaac	ggtctgcgtt	gtcgggaaga	tgcgtgatct	20760
gatccttcaa	ctcagcaaaa	gttcgattta	ttcaacaaag	ccgccgtccc	gtcaagtcag	20820
cgtaatgctc	tgccagtgtt	acaaccaatt	aaccaattct	gattagaaaa	actcatcgag	20880
catcaaata	aactgcaatt	tattcatatc	aggattatca	ataccatatt	tttgaaaaag	20940
ccgtttctgt	aatgaaggag	aaaactcacc	gaggcagttc	cataggatgg	caagatcctg	21000
gtatcgggtc	gcgattccga	ctcgtccaac	atcaatacaa	cctattaatt	tcccctcgtc	21060
aaaaataagg	ttatcaagtg	agaaatcacc	atgagtgcgc	actgaatccg	gtgagaatgg	21120
caaaagctct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	cgtattgggc	21180
gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	cggcgagcgg	21240
tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	21300
agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	21360
cgtttttcca	taggctccgc	ccccctgacg	agcatcaca	aaatcgacgc	tcaagtcaga	21420
ggtggcgaaa	cccgacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	21480
tgcgctctcc	tgttccgacc	ctgcgcctta	ccggatacct	gtccgccttt	ctcccttcgg	21540
gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcgggtg	taggtcgttc	21600
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	21660
gtaactatcg	tcttgagtcc	aaccgggtaa	gacacgactt	atcgccactg	gcagcagcca	21720
ctggtaacag	gattagcaga	gcgaggatat	taggcgggtg	tacagagttc	ttgaagtggg	21780
ggcctaacta	cggctacact	agaagaacag	tatttggtat	ctgcgctctg	ctgaagccag	21840
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	21900
gtgggttttt	tgtttgcaag	cagcagatta	cgcgagaaaa	aaaaggatct	caagaagatc	21960
ctttgatctt	ttctacgggg	tctgacgctc	agtggaacga	aaactcacgt	taagggattt	22020
tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttgatccgg	aatta	22075

<210> 29

<211> 7910.

<212> ADN

<213> Trình tự Nhân tạo

<220>

<223> Catxet biểu hiện Cry1Ai

<400> 29

```

agctgcttgt ggggaccaga caaaaaagga atggtgcaga attgttaggc gcacctacca      60
aaagcaactt tgcctttatt gcaaagataa agcagattcc tctagtacaa gtggggaaca      120
aaataacgtg gaaaagagct gtccctgacag cccactcact attgcgtttg acgaacgcag      180
tgacgaccac aaaactcgag acttttcaac aaaggggtatt atccggaaac ctccctcgat      240
tccattgccc agctatctgt cacttttattg tgaagatagt ggaaaaggaa ggtggctcct      300
acaaatgcca tcattgcat aaaggaaagg ctatcgttga agatgcctct gccgacagtg      360
gtcccaaaga tggaccccca cccacgagga gcacgtgga aaaagaagac gttccaacca      420
cgtcttcaaa gcaagtggat tgatgtgata tctccactga cgtaaggggt gacgaacaat      480
cccactatcc ttctgcctaa ttagctaacg gaccctctgc ctttctgttc tcaaacgat      540
gtcacatgtc tgcgctggac aactttcttg ttgccgctg tcgcttgccg tgtgctgact      600
ggacgcagct ccggagggtt ggttgtgctt ggttttcgta gagaactcgc cacttgccgc      660
ccgcacgttc ttggtgttcc ctccctccccg ctgtgttctg cgcacgggct ttttctgaga      720
gacccatgtt tcccttttac ttttataaac agtatacatg ctatgtttct agaaggaggg      780
gaaacctaatt cccctaata caatggcggg gaggaatat ggtgggggtg ggtgggggga      840
gggaaatatc tcgctacttt ttaatccgga caagctcatt tgcgtttgcg tctgaatgat      900
gatgactgca atgctgatcg caccctgggt gtccgatcac cagcttttgg ctgctctcac      960
caaatcagct gcaagaagat tagagcaca aagaattaca gaaagagagc ctttttcttt      1020
tcttcttctg tgggttctct tcatttctgt ctctctcttc tctgccagcc agtccgtcct      1080
tgcttccact gcacctgcac acaggtcacc ccgacccgca ctgttctaga ctccattaga      1140
aaaaaaaaagg tctgaacctt tccgaaacca gccagccatt ggtctggcag gccagcatat      1200
gctaattgga tttttttgcc gcacatttga gtgcgccatc aggatttggg aatcctgggt      1260
ttgagtaata cagtaatttg gcattatcca ttgccgaatt cccaagctcc gtcagcttga      1320
acgtggaccc ctaccatctg caccagctcg gcacctcacg ctgcgagcgc taggagccta      1380
ggagcagctg cccgtctatt tatttggtccc tctcccgccc cagagaaacc ctccctccct      1440
cctccattgg actgcttgct ccctgttgac cattggggta tgcttgctgc attgctctcc      1500
tgttcatctc cgtgctaaac ctctgtcttc tgggtgggtt tttgctggga ttttgagcta      1560
atctgctggg cccggtagaa aaagatcatg tccctgacg tgcacaagcg ctgcgcttag      1620
ccgcgtcctt gcccccgcc attttttgcg gtttcgggtg gttcccgctg ctgcgcgggt      1680

```


gcgtcatcgc	ctgaatcttg	tctgggctct	gctgacatgt	tcttggctag	ttgggtttat	1740
agattcctct	gatctaaacc	gtgcctgtgc	tgcgcacaga	actctcccct	gtcctttcct	1800
ggggtttttg	ttacgtggtg	gtagtaagct	tggatttgca	catggataaa	gttgttctaa	1860
gctccgtggg	ttgcttgaga	tcttgctgtt	attgctgtcc	gtgctcactt	tttttgcaat	1920
ccgaggaatg	aatttgctgt	ttactcgttt	tggtaggatta	ttagcgcgaa	aaaaaaactc	1980
tttttttttt	gttctttttac	tacgaaaagc	atcttctttg	attttgctat	cttctttttac	2040
tacgaaaaaac	tcttgagtct	aggaatttga	atgtgtgatg	tccatttctg	cagtgcgctg	2100
tgctttattg	ggaagccaaa	tcctattatt	ttctgcctct	agggctctgaa	tggaaatcagt	2160
actcttgaga	cagaaaatca	atccaatcaa	gttgattttct	ttcttttaaaa	atattatcac	2220
agaactaagt	gcttgtgcgg	aatcagtaact	ggctttttgtt	tggtaggagga	tcaataacttg	2280
cttttgtttg	gggggtggcaa	ctgttttgct	ataagattcc	atgtgttcct	gttgagatga	2340
atcatatata	gtatagctgc	atactacaaa	tctgtttttc	aaatttaggt	tgctttggca	2400
tgatctatlt	ttttgtcaga	cagactttct	aagtggtagc	tcttgatttc	ttgttcttgt	2460
acaactggtg	ctgctgaatc	ttgaccgtat	agctcgaatt	gcagtattct	gaaccatcga	2520
gccaaggctg	ccaagctgac	tgcctccac	agtcttcgcg	aacgccttgg	tgccaccttc	2580
tcctcccatc	ccaatgaact	gatagcactc	ttttccaggt	gggcttacca	aaatcatata	2640
acttgcatlt	cattcggtac	tgaaagttgt	taatttgcta	ttctcttcat	gcctgtctta	2700
atagcacacc	cagatgtaaa	cacgagatta	tgcaacttct	tacttggttt	cttttgttgg	2760
caccatcatg	catgctaatt	gctaaggatg	ttacctattc	atccttgact	catattatca	2820
tatgtaatga	ttttatgata	acgagactat	tgattgtgaa	gcatagtata	gctgttcttc	2880
agtttttgta	cccttttggt	tttttcccta	agctagaact	ggtacaattt	agttgataag	2940
acagtgtagt	ttgtagtacg	tcatttgaca	gattgtttgt	cttttagctgg	taaagtgcca	3000
tttaatatct	gtatccttca	gatctaataa	aaaggatatg	agatgtccat	cacaagaggg	3060
gaaaaattac	atgatctgag	atgtaacatc	cgttttttatt	tgtgaaatac	cacttctaca	3120
ggtatcttca	ctagggtaaa	ccatggacaa	taatcctaatt	attaacgagt	gcattccata	3180
caactgcctg	agcaatcctg	aggtggaggt	tctcgggcgg	gagaggattg	agacgggcta	3240
cacgccgatc	gacatctccc	tgagcctcac	ccagttcctc	ctgtccgagt	togtgcgggg	3300
cgccggcttc	gtgctggggc	tcgctgacat	catctggggc	atcttcgggc	ccagccagtg	3360
ggacgccttc	ctgggtccaga	tgcagcagct	catcaaccag	aggatcgagg	agttcgcgcg	3420
caaccaggcc	atctccaggc	tggagggcct	cagcaacctg	taccagatct	acgccagatc	3480
cttccgcgag	tgggaggcgg	acccgaccaa	ccgggtctct	agggaggaga	tgcgcatacca	3540
gttcaacgac	atgaactcgg	cctgaccac	cgctatcccc	ctcttcgcgc	tgcagaacta	3600
ccaggtgcgc	ctcctgtcgg	tctacgtgca	ggctgccaac	ctgcacctct	cggtgctgag	3660
ggacgtgagc	gtcttcgggc	agcgtgggg	cttcgacgcg	gccacgatca	actcccgcga	3720
caacgacctc	acgaggctga	tgggcaacta	caccgactac	gccgtccgct	ggtacaacac	3780
cggcctcgag	aggggtgtgg	gcccggacag	cagggactgg	gtcaggtaca	accagttccg	3840

cagggagctg	accctcacgg	tgctggacat	cgtcgccctc	ttctccaact	acgactcgag	3900
gaggtacccc	atcaggaccg	tgtcccagct	gacgagggag	atctacacca	accccgctct	3960
cgagaacttc	gacggcagct	tccgcggcat	ggcccagagg	atcgagcaga	acatccgcc	4020
gccgcacctg	atggacatcc	tcaactccat	caccatctac	acggacgtgc	acaggggctt	4080
caactactgg	tcggggccacc	agatcacggc	ttccccagtg	ggcttcagcg	gcccggagtt	4140
cgccttccca	ctgttcggca	acgcgggcaa	cgcgggccccg	ccagtgtctcg	tctccctcac	4200
cggcctcggc	atcttccgca	ccctgtccag	ccgctctac	aggaggatca	tcctcggcag	4260
cggccccaat	aatcaggagc	tgttcgtgct	cgacggcacg	gagttctcct	tcgcgagcct	4320
gaccacgaac	ctcccgctca	cgatctacag	gcagaggggc	acggtggact	ccctggacgt	4380
catcccgccc	caggacaact	cggctccgcc	cgcgccgggc	ttctcccaca	ggctgagcca	4440
cgtgaccatg	ctcagccagg	cggccggcgc	cgtctacacc	ctcagggccc	cgaccttctc	4500
ctggcagcac	aggagcgcgg	agttcaacaa	catcatcgcc	tccgacagca	tcacgcagat	4560
cccggcggtg	aagggcaact	tcctcttcaa	cggctccgtc	atcagcggcc	ccggcttcac	4620
cggcggcgac	ctggtgcgcc	tcaactccag	cggcaacaac	atccagaaca	ggggctacat	4680
cgaggtcccc	atccacttcc	cgtccaccag	caccaggtac	aggggtgcgcg	tcaggtacgc	4740
ttcggtgacc	ccgatccacc	tgaacgtcaa	ctggggcaac	tccagcatct	tctccaacac	4800
ggtgccggct	acggccacca	gcctggacaa	cctccagtcc	agcgacttcg	gctacttcga	4860
gtccgccaac	gcgttcacgt	ccagcctcgg	caacatcgctg	ggcgtccgca	acttctcggg	4920
gacggctggc	gtgatcatcg	acaggttcga	gttcatcccc	gtcacggcca	ccctggaggc	4980
tgagtacaac	ctcgagcgcg	cccagaaggc	cgtgaacgcc	ctgttcacct	ccacgaacca	5040
gctgggcctc	aagaccaacg	tcacggacta	ccacatcgac	caggtgtcca	acctggtcac	5100
ctgcctcagc	gacgagttct	gcctggacga	gaagagggag	ctgagcgaga	aggtgaagca	5160
cgccaagcgc	ctctccgacg	agaggaacct	cctgcaggac	agcaacttca	aggacattaa	5220
taggcagccc	gagaggggct	ggggcggctc	cacgggcatac	accatccagg	gcggcgacga	5280
cgtcttcaag	gagaactacg	tcaccctgag	cggcacgttc	gacgagtgct	accccacgta	5340
cctctaccag	aagatcgacg	agtccaagct	gaaggcgttc	acccgctacc	agctcagggg	5400
ctacatcgag	gacagccagg	acctggaggt	gtacctcatc	cgctacaacg	ccaagcacga	5460
gacgctgaac	gtgcccggca	cgggctccct	gtggccccctc	gcgggtcaaga	gccccatcgg	5520
caggtgcggc	gagcccaaca	ggtgcgcccc	caggatcgag	tgggaagccgg	acgtggactg	5580
cagctgcagg	gacggcgaga	agtgcgctca	ccactcccac	cacttcagcc	tcgacatcga	5640
cgtcggctgc	acggacctga	acgaggacct	cggcgtgtgg	gtcatcttca	agattaaaac	5700
ccaggacggc	cacgcgaaga	tcggcaacct	ggagttcctc	gaggagaagc	tgctgctggg	5760
cgaggctctg	gccaggggtga	agaaggccga	gaagaagtgg	cgcgacaaga	gggagaagct	5820
cgagtgggag	acgaacatcg	tctacaagga	ggcgaaggag	tccgtggacg	ccctgttcgt	5880
cgacagccag	tacaaccgcc	tccagaccga	cacgaacatc	gcgatgatcc	acgccgcgga	5940
caagagggtg	caccgcatca	gggaggccta	cctgcccagag	ctgagcgtga	ttccgggcgt	6000

caacgctgcg atcttcgagg agctggaggg cctcatcttc accgcttct cctgtacga	6060
cgcccgcaac gtgatcaaga acggcgactt caactacggc ctcagctgct ggaacgtcaa	6120
gggccacgtg gacgtcgagg agcagaacaa ccacaggtcg gtgctggtea tccccgagt	6180
ggaggccgag gtgagccagg aggtgcgcgt ctgcccgggc aggggctaca tcctccgcgt	6240
caccgcctac aaggagggct acggcgaggg ctgcgtcacg atccacgaga tcgaggacaa	6300
caccgacgag ctgaagttct ccaactgcgt ggaggaggag gtctaccga ataataccgt	6360
gacgtgcaac gactacaccg ccacgcagga ggagtacgag ggcacctaca cctcgaggaa	6420
caggggctac gacggcgctt acgagtcctaa cagctcggtc cccgcggact acgctagcgc	6480
gtacgaggag aaggcctaca ccgacggcag gagggacaa ccgtgcgagt ccaaccgcgg	6540
ctaccgggac tacacgcgcg tgcggcgagg ctacgtgacc aaggagctgg agtacttccc	6600
ggagacggac aagggtctga tcgagctggg ggagaccgag gggactttcc tggttgatcc	6660
agtggagctg ctgctgatgg aggagtgaga gctctacggg gaaagaagga gaagaagaag	6720
aagaagccca ggccggagaa ccatgcctg catttcgact tgtttcaccg caattcgcat	6780
tgttagtcgt gtattggagt tatgtgtact tggtttccaa gaactttggg tccttctcgt	6840
ttttttttcc ttgcttggtt gagcggtttt gggcagcgct ggccctgggtc ctagtatggt	6900
gggaattggc tgcacctttt gcttcgaata aaaatgcctg ctcgttcacc tgtcttccag	6960
agtgcaatgc gatgttctgt tgcacaggtc gtgtgggtct gactgatggc gatgttgtgt	7020
tcttctgtta atcgctgtt ttaacgtggg aggcgtgtgc ttgttcttgt tgagaaagct	7080
tgtgtgtcca gacatggctg cttgaataca agtgaaggaa aaaaaaagcc atgccaaagta	7140
aagttgcaca aaatttcaac tgcctcagtg accactggac catgttcttg gttattgcag	7200
ttgcagggct tcacatggcg tttggacagc agtcttggat tgatecataa agaggtgggtg	7260
gttaatgagg acgcaaggcc gttccctcag agtcagtcac aaggttgcag aggtcacggg	7320
tctcttccct ttccgcttcc tgtcacatcg gaattgttgt ttacgccatc tgcacatcac	7380
ccaccaagtc tatgtttctg tactggatct ttcaatggcg gaacgcgctt agttcttcgt	7440
cacagtcgaa tcacatgata taatcgatgt ctttaatctc gctgtaaaaa ggggtgggacg	7500
gtgggtgcag ggtagggacc agggaaggcc tgcctaaacg tatccataaa catgcacagc	7560
aaocctaaga tattatactg cctacttccct aagatatagt tttttctagt ctattttttt	7620
tccgtccaca tccaaataag tgataatata tagacataca tatatatact atattcatca	7680
tagattaatg aacgaatgta tacttagttt aaacctaatt atattttagg aaggatggag	7740
tatgaaacat gacaatacaa caaaaaaaaaa tcatgtaatt gcataatcgtc aaagttatct	7800
gaagtaacca atccaggggg aaatcccggt agcaaacata caagagcacc gcccactac	7860
atcccagaaa ataaaacaaa accagaactc agatggataa ataatactac	7910