(12) **BẢN MÔ TẢ SÁNG CHẾ THUỘC BẰNG ĐỘC QUYỀN SÁNG CHẾ**

(19) **Cộng hòa xã hội chủ nghĩa Việt Nam (VN)**
CỤC SỞ HỮU TRÍ TUỆ

(11) **1-0022195**

(51)[7] **C12N 15/85**

(13) **B**

(21) 1-2008-02238
(22) 12.02.2007
(86) PCT/GB2007/000488 12.02.2007
(87) WO2007/091099 16.08.2007
(30) 11/352,177 10.02.2006 US
0621234.4 25.10.2006 GB
(45) 25.11.2019 380
(43) 25.05.2009 254
(73) Oxitec Limited (GB)
71 Milton Park, Abingdon, Oxfordshire OX14 4RX, United Kingdom
(72) Alphey, Luke (GB)
(74) Công ty Luật TNHH T&G (TGVN)

(54) PHƯƠNG PHÁP BIỂU HIỆN PROTEIN CHỨC NĂNG Ở CÔN TRÙNG VÀ PHƯƠNG PHÁP PHÂN CHIA GIỚI TÍNH CÔN TRÙNG

(57) Sáng chế đề cập đến hệ thống biểu hiện polynucleotit có khả năng cắt nối thay thế trên các bản phiên mã ARN của trình tự polynucleotit được biểu hiện trong sinh vật.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến phương pháp biểu hiện protein chức năng bằng cách sử dụng hệ thống biểu hiện gen, kết hợp với các trình tự kiểm soát cắt nối, các trình tự kiểm soát này tạo ra cơ chế cắt nối thay thế đặc hiệu giới tính.

Tình trạng kỹ thuật của sáng chế

Sự cắt nối thay thế kéo theo sự loại bỏ của một hoặc nhiều intron và nối các exon lại với nhau. Phản ứng này được xúc tác bởi phức hệ, bộ máy đại phân tử bao gồm 5 loại ARN và hàng trăm loại protein (tham khảo tài liệu Jurica, M. S. & Moore, M. J. (2003) Mol. Cell 12, 5-14). Sự cắt nối thay thế tạo ra nhiều ARN từ một gen, do đó tăng tính đa dạng của proteome (tham khảo tài liệu Graveley, B. R. (2001) Trends Genet. 17, 100-107).

Sự cắt nối thay thế cũng đóng vai trò quan trọng trong sự điều hòa mức biểu hiện gen trong nhiều quá trình phát triển nằm trong khoảng từ xác định giới tính đến cơ chế gây chết tế bào theo chương trình (tham khảo tài liệu Black, D. L. (2003) Annu. Rev. Biochem. 72, 291-336), và các nhược điểm trong việc cắt nối thay thế có liên quan mật thiết tới các rối loạn chức năng ở người (tham khảo tài liệu Caceres, J. F. & Kornblihtt, A. R. (2002) Trends Genet. 18, 186-193). Nói chung, sự cắt nối thay thế được điều hòa bởi các protein kết hợp với tiền mARN và vừa tăng cường vừa ức chế khả năng nhận biết các vị trí cắt nối cạnh nhau của phức hệ để exon điều hòa (tham khảo tài liệu Smith, C. W. & Valcarcel, J. (2000) Trends Biochem. Sd. 25, 381-388).

Một exon luân phiên riêng biệt sẽ nằm trong hoặc cắt ra khỏi ARN trưởng thành trong mỗi tế bào, được xác định bằng sự tập hợp số yếu tố điều hòa cắt nối chủ động và bị động và sự tương tác của các yếu tố đó với tiền mARN và các thành phần của phức hệ (tham khảo tài liệu Smith, C. W. & Valcarcel, J. (2000) Trends Biochem. Sd. 25, 381-388).

Các điểm phân cắt (spliceosome) là phức hợp lớn của ARN phân tử nhỏ và các hạt protein (small nuclear RNA and protein particles: snRNP), các ARN này kết hợp với tiền mARN để đạt được sự cắt nối ARN, bằng cách loại bỏ các intron khỏi ARN của sinh vật nhân chuẩn, do đó tạo ra mARN mà sau đó mARN này được dịch mã thành protein trong ribosom.

Mặc dù ít nhất 74% bộ gen người mã hóa các mARN được cắt nối thay thế (tham khảo tài liệu Johnson, J. M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P. M., Armour C. D., Santos, R., Schadt, E. E., Stoughton, R. & Shoemaker, D. D. (2003) Science 302, 2141-2144), chỉ có một số ít yếu tố điều hòa cắt nối là được nhận biết.

Peng Gong và đồng tác giả trong tài liệu Nature Biotechnology 23(4), 453-456 (2005) mô tả sự tạo ra ruồi giấm Mediterranean biểu hiện chất hoạt chuyển có teher kiềm cheeds tetracycline (tTAV) mà gây ra sự chết sớm ở con của chúng.

Bản chất kỹ thuật của sáng chế

Do đó, trong khía cạnh thứ nhất, sáng chế đề cập đến phương pháp biểu hiện protein chức năng trong côn trùng thông qua sự cắt nối thay thế đặc hiệu giới tính sản phẩm phiên mã ARN, trong đó phương pháp này bao gồm các bước: a) phiên mã, trong côn trùng nêu trên, polynucleotit ngoại sinh từ hệ thống biểu hiện để tạo ra bản phiên mã ARN, trong đó hệ thống biểu hiện này bao gồm: trình tự khởi đầu được liên kết một cách có kiểm soát với trình tự polynucleotit ngoại sinh mã hóa cho protein chức năng, nằm giữa codon khởi đầu và codon kết thúc, và trình tự kiểm soát cắt nối intron bao gồm miền liên kết protein; trong đó miền liên kết protein bao gồm trình tự ADN liền kề được nêu trong SEQ ID NO:1 hoặc ARN tương ứng với nó, trong đó trình tự kiểm soát cắt nối intron bao gồm trình tự cho cắt nối GT ở đầu 5' của nó (5'-GT), và trong đó trình tự kiểm soát cắt nối intron được kẹp sườn bởi nucleotit guanin (G); và trong đó trình tự kiểm soát cắt nối intron là từ đầu 3' đến codon khởi đầu ATG của trình tự polynucleotit ngoại sinh; b) cắt nối thay thế bản phiên mã ARN của

trình tự polynucleoti ngoại sinh, cùng hoạt động với vị trí cắt, để tạo ra sản phẩm ARN thông tin (mARN) được cắt nối thứ nhất, mà không bao gồm khung đọc mở liên tục kéo dài từ codon khởi đầu đến codon kết thúc, và một sản phẩm ARN thông tin được cắt nối thay thế khác mà bao gồm khung đọc mở liên tục kéo dài từ codon khởi đầu đến codon kết thúc, xác định protein chức năng nêu trên đối với việc biểu hiện.

Hệ thống biểu hiện có thể là ADN hoặc ARN hoặc thể lai hoặc tổ hợp của cả hai. Các tác giả sáng chế đã dự tính rằng hệ thống này bao gồm cả các ribo- và deoxy-ribonucleotit, nghĩa là các phần của ADN và các phần của ARN. Chúng có thể tương ứng với các yếu tố di truyền khác nhau, chẳng hạn hệ thống này là thể lai ADN/ARN, với một số yếu tố chức năng được tạo ra bởi ADN và các yếu tố chức năng khác được tạo ra bởi ARN.

Sự điều hòa là điều hòa đặc hiệu giới tính, và tốt hơn là đặc hiệu giai đoạn, đặc hiệu phôi hoặc đặc hiệu mô. Tuy nhiên, cũng được ưu tiên nếu sử dụng phối hợp cả 4 cách điều hòa trên. Được ưu tiên đặc biệt nếu, dạng phối hợp của các phương thức trên được sử dụng, mà bao gồm cả điều hòa đặc hiệu giới tính. Một ví dụ về dạng phối hợp được ưu tiên đặc biệt là phối hợp điều hòa đặc hiệu giới tính, đặc hiệu mô, đặc hiệu giai đoạn của sự cắt nối thay thế.

Hệ thống này được điều chỉnh để phù hợp với việc biểu hiện gen. Trình tự polynucleotit được biểu hiện bao gồm trình tự mã hóa cho một protein hoặc polypeptit, nghĩa là có tối thiểu một exon và tốt hơn là 2 exon hoặc nhiều hơn, có khả năng mã hóa một polypeptit, ví dụ một protein hoặc một đoạn protein của chúng.

Exon được hiểu là một vùng ADN bất kỳ trong gen, có mặt trong một phân tử ARN lấy ra từ gen đó, hơn là được cắt ra khỏi một phân tử ARN được phiên mã. Đối với gen mã hóa protein, các phân tử ARN trưởng thành tương ứng với các phân tử mARN trưởng thành, phân tử mARN trưởng thành đó có thể mã hóa cho một hoặc nhiều protein hoặc trình tự polypeptit. Các exon của nhiều gen trong tế bào nhân chuẩn được xen kẽ với các đoạn ADN không mã hóa.

Có ít nhất một trình tự polynucleotit khác loại có thể mã hóa một protein chức năng, nằm giữa codon khởi đầu và codon kết thúc được biểu hiện ở côn trùng. Ngoài ra, ít nhất một trình tự polynucleotit khác loại mã hóa hoặc bao gồm polynucleotit của ARN can thiệp (interference ARN: ARNi), được biểu hiện trong côn trùng.

Những trình tự này, được biểu hiện trong côn trùng, cũng có thể được đề cập đến như các trình tự mà mức biểu hiện của nó được điều chỉnh trong côn trùng nêu trên.

Tốt hơn là, trình tự polynucleotit được biểu hiện bao gồm hai hoặc nhiều exon mã hóa, là các đoạn hoặc trình tự polynucleotit mã hóa axit amin khi được dịch mã từ mARN. Tốt hơn là, các exon khác nhau được cắt một cách khu biệt cùng nhau để tạo ra các mARN thay thế. Tốt hơn, mARN được cắt nối thay thế nêu trên có tiềm năng mã hóa khác nhau, nghĩa là mã hóa cho các protein hoặc các trình tự polynucleotit khác nhau. Do đó, mức biểu hiện của trình tự mã hóa được điều chỉnh bởi sự cắt nối thay thế theo các cách điều hòa nêu trên.

Trình tự polynucleotit được biểu hiện có thể bao gồm các polynucleotit của ARN can thiệp (interference ARN: ARNi). Các trình tự này có khả năng cung cấp, chẳng hạn, một hoặc nhiều ARN sợi kép (dsARN) dạng thẳng, tốt hơn là ở dạng sản phẩm phiên mã sơ cấp lần lượt được xử lý bởi enzym tương tự ARN Pol III "Dicer". Các sợi này bao gồm, ví dụ, các sợi ARN đơn có thể tạo vòng, như các sợi chúng có thể được tìm thấy trong cấu trúc kẹp tóc-ngắn ARN (short hairpin ARN: shARN) hoặc các đoạn dài hơn mà về cơ bản tự bổ trợ.

Do đó, khi hệ thống này là ADN, trình tự polynuleotit của ARN can thiệp là deoxyribonucleotit mà khi được dịch mã thành ribonucleotit tiền ARN, cung cấp một mạch dsARN, như đề cập ở trên.

Các polynucleotit của ARN can thiệp được ưu tiên đặc biệt khi các polynucleotit này được bố trí để giảm đến mức tối thiểu sự can thiệp vào sự cắt nối thay thế. Điều này có thể đạt được bằng cách đặt các polynucleotit ở xa các trình tự kiểm soát cắt nối thay thế, tốt hơn là đầu 3' của các trình tự kiểm soát.

Trong một phương án được ưu tiên khác, các vùng về cơ bản tự bổ trợ có thể được tách biệt với nhau bởi một hoặc nhiều trình tự kiểm soát cắt nối, chẳng hạn một intron mà điều hòa sự cắt nối thay thế. Tốt hơn là, các vùng tự bổ trợ được xắp sếp thành một loạt gồm hai hoặc nhiều hơn hai đoạn lặp đã đảo chiều, mỗi đoạn lặp đảo chiều được tách biệt bởi trình tự điều khiển cắt nối, tốt hơn là intron, như đã đề cập ở trên.

Trong cấu hình này, các đoạn dịch mã được cắt nối thay thế khác nhau có thể có các vùng tự về cơ bản tự bổ trợ được tách biệt bởi trình tự không tự bổ trợ với độ dài khác nhau trong sản phẩm phiên mã hoàn chỉnh (sau khi cắt nối thay thế). Cần được hiểu rằng các vùng mà về cơ bản tự bổ trợ là các vùng có khả năng tạo thành dạng kẹp tóc, ví dụ, dưới dạng các phần của trình tự có khả năng ghép đôi với các phần khác của trình tự này. Hai phần đó không bổ trợ một cách chính xác cho nhau, do có thể có một số sự ghép đôi không đúng hoặc sự dung nạp mạch ở mỗi phần mà không ghép đôi với nhau. Các mạch này có thể không có sự tương đương ở các phần khác, đến mức mất tính đối xứng và tạo ra các phần nhô ra, như được biết đến với sự bổ trợ ghép đôi nói chung.

Theo một phương án được ưu tiên khác, một hoặc nhiều đoạn của trình tự về cơ bản bổ trợ cho một phần khác của sản phẩm phiên mã sơ cấp được xác định vị trí, tương ứng với ít nhất một trình tự kiểm soát cắt nối, để nó không được bao gồm trong tất cả các sản phẩm phiên mã tạo ra bằng cách cắt nối thay thế sản phẩm phiên mã ban đầu. Bằng cách đó, một số sản phẩm phiên mã được tạo ra có xu hướng tạo ra dsARN trong khi các sản phẩm phiên mã khác lại không tạo ra; bằng cách điều hòa sự cắt nối thay thế, ví dụ điều hòa đặc hiệu giới tính, điều hòa đặc hiệu giai đoạn, điều hòa đặc hiệu phôi, điều hòa đặc hiệu mô, và các phương pháp tổ hợp của chúng, dsARN có thể được tạo ra theo cách đặc hiệu giới tính, đặc hiệu giai đoạn, đặc hiệu phôi hoặc đặc hiệu mô, hoặc tổ hợp của chúng.

Hệ thống này có khả năng biểu hiện ít nhất một protein, nghĩa là protein chức năng được biểu hiện ở côn trùng. Ít nhất một protein quan tâm này có thể

có tác dụng trị liệu hoặc tốt hơn là, có thể là một chất chỉ thị, ví dụ DsRed, Green Fluorescent Protein (GFP) hoặc một hoặc nhiều thể đột biến hoặc các cải biến của chúng, hoặc các chất chỉ thị khác mà đã biết trong lĩnh vực kỹ thuật này.

Tốt nhất là, protein chức năng được biểu hiện ở côn trùng có hiệu quả gây chết, gây hại hoặc khử trùng. Trong bản mô tả này, khi đề cập đến hiệu quả gây chết có nghĩa là nó được mở rộng đến hiệu quả gây hại hoặc khử trùng, chẳng hạn có khả năng giết chính côn trùng đó hoặc giết con của nó hoặc có khả năng làm giảm hoặc phá hủy chức năng của một số mô của côn trùng, trong số đó các mô có khả năng sinh sản là được đặc biệt ưu tiên, tới mức côn trùng này hoặc con của chúng bị bất thụ. Do đó, một số hiệu quả gây chết, chẳng hạn thuốc độc, sẽ giết chết côn trùng hoặc phá hủy mô trong thời gian ngắn so với tuổi thọ của côn trùng, trong khi các chất khác có thể đơn giản làm giảm chức năng của côn trùng, chẳng hạn khả năng sinh sản.

Hiệu quả gây chết dẫn đến bất thụ được đặc biệt ưu tiên, vì điều này cho phép côn trùng này cạnh tranh với các côn trùng kiểu dại trong môi trường tự nhiên ("môi trường hoang dại"), nhưng côn trùng bất thụ sau đó không thể sinh ra con có khả năng sống sót. Theo cách này, sáng chế đạt được kết quả tương tự như công nghệ SIT (Sterile Insect Technique: SIT) ở côn trùng, mà không gặp phải các vấn đề của SIT, chẳng hạn chi phí, gây nguy hiểm cho người sử dụng và giảm tính cạnh tranh của côn trùng được chiếu xạ.

Tốt hơn là, hệ thống bao gồm ít nhất một cơ chế phản hồi chủ động, cụ thể là ít nhất protein chức năng được biểu hiện biệt hóa, thông qua sự cắt nối thay thế, và ít nhất một trình tự khởi đầu của nó, trong đó sản phẩm của gen được biểu hiện đóng vai trò làm yếu tố điều hòa phiên mã chủ động của ít nhất một trình tự khởi đầu, và nhờ đó sản phẩm, hoặc mức biểu hiện của sản phẩm, được kiểm soát. Tốt hơn là, trình tự tăng cường được kết hợp với trình tự khởi đầu, sản phẩm gen đóng vai trò tăng cường hoạt tính của trình tự khởi đầu thông qua trình tự tăng cường. Tốt nhất là yếu tố kiểm soát là sản phẩm gen tTA hoặc dạng

tương đồng của nó, và trong đó một hoặc nhiều đơn vị điều khiển tetO được liên kết một cách có kiểm soát với trình tự khởi đầu và là trình tự tăng cường, tTA hoặc dạng tương đồng của nó để tăng cường hoạt tính của trình tự tăng cường thông qua tetO. Tốt hơn là protein chức năng mã hóa sản phẩm tTAV hoặc tTAF và tốt hơn là, trình tự khởi đầu về căn bản là không có hoạt tính khi không có mặt yếu tố kiểm soát phiên mã chủ động. Phù hợp hơn nếu các trình tự khởi đầu của hệ thống được chọn từ: hsp70, trình tự khởi đầu tối thiểu P, trình tự khởi đầu tối thiểu CMV, trình tự khởi đầu tối thiểu trên cơ sở Act5C, đoạn khởi đầu BmA3, đoạn khởi đầu từ hunchback, trình tự khởi đầu lõi Adh, và trình tự khởi đầu tối thiểu Act5C, hoặc tổ hợp của chúng.

Theo một phương án, tốt hơn là protein chức năng là yếu tố gây chết tế bào theo chương trình, như protein AIF đã mô tả trong CDNAe et al (JouARNl of Cell Science 115, 4727-4734 (2002)) chẳng hạn hoặc các thể tương đồng của chúng. Các thể tương đồng của AIF đã được tìm thấy trong các loài động vật có vú và thậm chí cả trong các loài động vật không xương sống, bao gồm côn trùng, giun tròn, nấm và thực vật, nghĩa là gen AIF đã được bảo toàn trong toàn bộ giới sinh vật nhân chuẩn. Cũng được ưu tiên là Hid, sản phẩm protein của gen *head involution defective* của *Drosophila melanogaster*, hoặc *Reaper* (Rpr), sản phẩm của gen gây chết của *Drosophila*, hoặc thể đột biến của nó. Sử dụng Hid đã được mô tả bởi Heinrich và Scott (tham khảo Proc. Natl Acad. Sci USA 97, 8229-8232 (2000). Sử dụng dẫn xuất của thể đột biến, Hid^{Ala5} đã được mô tả bởi Horn và Wimmer (tham khảo tài liệu Nature Biotechnology 21, 64-70 (2003)). Sử dụng dẫn xuất đột biến của Rpr, Rpr^{KR}, đã được mô tả trong tài liệu này (cũng xem White et al 1996, Wing et al., 2001, và Olson et al., 2003). Cả Rpr và Hid đều là các protein tiền chết theo chương trình, được xem là liên kết với IAP1. IAP1 là protein kháng lại việc chết theo chương trình được bảo toàn tốt . Vì vậy, Hid và Rpr được mong chờ hoạt động với khoảng phát sinh loài rộng (tham khảo tài liệu Huang et al., 2002, Vernooy et al., 2000) ngay cả khi các trình tự của chúng không được bảo toàn tốt.

Cũng được ưu tiên là NipplDm, thể tương đương Drosophila của Nippl ở động vật có vú (Parker et al Biochemical Journal 368, 789-797 (2002); Bennett et al, Genetics 164, 235-245 (2003)). NipplDm là ví dụ khác của protein với hiệu ứng gây chết khi được biểu hiện ở mức độ phù hợp, được hiểu bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này. Thực vậy, nhiều ví dụ khác của protein với hiệu ứng gây chết sẽ được biết rõ bởi người có hiểu biết trung bình trong lĩnh vực kỹ thuật này.

Cũng được ưu tiên nếu bản thân protein chức năng là một chất hoạt hóa phiên mã, như hệ thống tTAV đã được mô tả ở trên.

Được ưu tiên nếu trình tự khởi đầu có thể được hoạt hóa nhờ các điều kiện môi trường, ví dụ có mặt hoặc vắng mặt yếu tố đặc biệt, chẳng hạn tetraxyclin trong hệ thống tet đã nêu ở trên, sao cho mức biểu hiện của gen quan tâm có thể được dễ dàng điều khiển bởi các chuyên gia trong lĩnh vực. Ngoài ra, một ví dụ được ưu tiên của trình tự khởi đầu phù hợp là trình tự khởi đầu sốc nhiệt hsp70 cho phép người dùng điều khiển mức biểu hiện nhờ sự thay đổi nhiệt độ môi trường mà ở đó có mặt vật chủ, ví dụ trong phòng thí nghiệm hoặc trên cánh đồng. Một ví dụ khác về sự kiểm soát nhiệt độ được ưu tiên đã được mô tả bởi Fryxell và Miller (JouRNAl of Economic Entomology 88, 1221-1232 (1995)).

Cũng được ưu tiên nếu trình tự khởi đầu là trình tự khởi đầu đặc hiệu phôi sryα (Horn & Wirnmer (2003) của *Drosophila melanogaster* hoặc các dạng tương tự của nó, hoặc các trình tự khởi đầu khác từ các gen đặc hiệu phôi hoặc các gen hoạt hóa phôi, chẳng hạn gen *Drosophila* chậm như *molasses* (slam), hoặc các dạng tương tự của nó từ các loài khác.

Cũng thích hợp nếu hệ thống bao gồm nhân tố ngược chiều 5' và/hoặc nhân tố xuôi chiều 3' để kiểm soát mức biểu hiện. Các ví dụ bao gồm các trình tự tăng cường như trình tự tăng cường thể mỡ từ các gen protein noãn hoàn *Drosophila* và các trình tự tăng cường vùng tương đồng (homology region – hr) từ baculovirut, ví dụ AcMNPV. Cũng được đánh giá cao nếu các sản phẩm ARN sẽ bao gồm các UTR 5' và 3' chẳng hạn.

Trình tự kiểm soát sự cắt nối thay thế cho phép tăng cường mức độ kiểm soát mức biểu hiện của protein, ngoài trình tự khởi đầu và/hoặc trình tự tăng cường. Ví dụ, mức biểu hiện đặc hiệu giới hoặc mô trong phôi côn trùng rất khó thực hiện bằng các phương pháp thông thường. Các trình tự khởi đầu với tính đặc hiệu này là chưa được biết đến, ngay cả trong *Drosophila*. Tuy nhiên, bằng cách sử dụng kiểm soát tổ hợp theo sáng chế, trình tự khởi đầu đặc hiệu phôi, ví dụ sryα, có thể được kết hợp với hệ thống cắt nối thay thế phù hợp.

Tốt hơn là tổ hợp bất kỳ gồm trình tự khởi đầu và tổ chức cắt nối thay thế được dự tính. Tốt hơn là gen khởi đầu đặc hiệu với một protein cụ thể có hiệu quả trong thời gian ngắn hạn chế về không gian, ví dụ hiệu ứng tự quản tế bào.

Thêm nữa, tốt hơn là trình tự khởi đầu có thể đặc hiệu đối với một nhóm protein lớn hơn hoặc một protein đặc hiệu có hiệu quả trong thời gian dài và/hoặc hiệu quả rộng khắp trong toàn hệ thống, chẳng hạn hormon, yếu tố phát triển tích cực hoặc tiêu cực, phân tử morphogen hoặc phân tử được tiết khác hoặc phân tử truyền tín hiệu bề mặt tế bào. Điều này cho phép, ví dụ, một mô hình biểu hiện rộng hơn để tổ hợp gồm các trình tự khởi đầu morphogen với cơ chế cắt nối thay thế đặc hiệu giai đoạn có thể dẫn đến morphogen được biểu hiện chỉ ở một giai đoạn nhất định trong chu trình sống, nhưng hiệu quả của morphogen sẽ vượt ra ngoài giai đoạn đó (nghĩa là morphogen có thể vẫn tác động và vẫn có hiệu quả). Các ví dụ được ưu tiên là các phân tử truyền tín hiệu/morphogen Hedgehog, Wingless/WNTs, TGFβ/BMPs, EGF và thể tương đồng của nó, mà là các phân tử truyền tín hiệu bảo toàn tiến hóa đã biết.

Cũng dự tính rằng trình tự khởi đầu đã hoạt hóa bởi một loạt các nhân tố protein, ví dụ chất hoạt hóa chuyển dịch, hoặc có tác dụng toàn thân, chẳng hạn hormon hoặc morphogen, có thể được sử dụng kết hợp với cơ chế cắt nối thay thế để đạt được quyền điều khiển đặc hiệu mô hoặc điều khiển đặc hiệu giới tính và giai đoạn hoặc tổ hợp của điều khiển đặc hiệu giai đoạn, mô, phôi và giới tính.

Cũng dự tính có nhiều hơn một trình tự khởi đầu và tùy ý trình tự tăng cường của nó, có thể được sử dụng trong hệ thống hiện tại, là các công cụ thay thế đối với việc khởi đầu phiên mã của các protein giống nhau hoặc bởi vì trên thực tế hệ thống di truyền có nhiều hơn một hệ thống biểu hiện gen (nghĩa là một gen và cùng với một trình tự khởi đầu của nó).

Theo một khía cạnh khác, sáng chế đề cập đến phương pháp chuyển nạp bao gồm bước biểu hiện hai hoặc nhiều phân tử ARN, bắt nguồn từ sản phẩm phiên mã sơ cấp sợi đơn hoặc về cơ bản các sản phẩm phiên mã sơ cấp tương tự, bằng cách cắt nối thay thế, hai hoặc nhiều hơn hai phân tử ARN tốt hơn là mã hóa các protein hoặc các trình tự polypeptit khác nhau trong côn trùng bằng cách cho côn trùng tiếp xúc với hệ thống biểu hiện gen và tốt hơn là gây ra sự biểu hiện của hệ biểu hiện này. Các phương pháp để đưa vào hoặc chuyển nạp hệ thống gen và gây ra sự biểu hiện đã được biết đến trong lĩnh vực kỹ thuật này đối với côn trùng tương ứng.

Sáng chế cũng đề cập đến côn trùng (nghĩa là thể chuyển nạp) được chuyển nạp bởi hệ thống theo sáng chế.

Mô tả vắn tắt các hình vẽ

Hình 1 minh họa các sản phẩm phiên mã khác nhau được sinh ra bởi việc cắt nối thay thế của gen *Aedes aegypti doublesex* (*Aadsx*). Cần phải hiểu rõ rằng *Aedes aegypti* cũng được biết đến như là *Stegomyia aegypti*. Hình này cho thấy gen *Aadsx* từ exon thứ tư, mà không được cắt nối thay thế, nghĩa là có mặt trong tất cả các sản phẩm phiên mã được mô tả ở đây. Việc đánh số là từ nucleotit thứ nhất của exon thứ tư (acgacgaact...). Chú ý rằng biểu đồ này không theo tỷ lệ - các intron dài hơn các exon. Tổng số các vùng được cắt nối thay thế gồm trên 43kb.

Hình 2 thể hiện RT-PCR của các con cái và các con đực từ dòng chuyển gen LA3491 *Aedes aegypti* sử dụng các đoạn mồi 688 - ie1-transcr (SEQ ID NO. 4) và 790 - Aedsx-m-r2 (SEQ ID NO. 5). Bằng cách sử dụng các primer này, việc cắt nối ở kiểu F2 sẽ tạo ra dải khoảng 985bp trong khi việc cắt nối ở

kiểu M1 sẽ tạo ra dải khoảng 516bp. Một dải khoảng 985bp (F2) chỉ được xuất hiện trong các đường biểu hiện các con cái và một dải sản phẩm phiên mã 1(M1) đặc hiệu con đực khoảng 516bp chỉ xuất hiện ở các con đực. Các dải này được xác định trình tự và cho thấy đã xảy ra việc cắt nối chính xác, nghĩa là lần lượt là loại F2 và M1 tương ứng. Sự có mặt của các dải trong các mẫu đối chứng không RT (-RT CON) cho thấy không có sự nhiễm tạp ADN bộ gen trong các mẫu. Các đường 1 và 11 là Marker (SmartLadder™ từ Eurogentec, các dải từ 1,5kb đến 0,2kb được chỉ ra). Các đường 2 và 3 là các đối chứng âm tính (không có enzyme phiên mã ngược) và các đường 2-9 là các sản phẩm phản ứng được tạo thành trên các phần chiết từ các con đực hoặc các con cái được biểu thị.

Hình 3 thể hiện RT-PCR của các con cái các con đực và các con cái từ các dòng chuyển gen LA3491 *Aedes aegypti* bằng cách sử dụng các đoạn mồi 688 - ie1-transcr (SEQ ID NO. 4) và 761 - Aedsx-fem-r (SEQ ID NO. 6). Bằng cách sử dụng các đoạn mồi này, việc cắt nối trong khuôn F2 sẽ tạo ra một dải khoảng 525bp. Một dải khoảng 525bp có mặt trong các sản phẩm phản ứng trên các phần chiết từ các con cái, nhưng không có trong các sản phẩm phản ứng tương ứng trên các phần chiết từ các con đực. Việc xác định trình tự của dải 525bp này khẳng định điều đó là đúng, nghĩa là đã xảy ra việc cắt nối loại F2. Marker (SmartLadder™ from Eurogentec, các dải từ 1,5kb đến 0,2kb được chỉ ra).

Hình 4 thể hiện RT-PCR của các con đực và các con cái từ các dòng chuyển gen LA3491 *Aedes aegypti* sử dụng các đoạn mồi 688 - ie1-transcr (SEQ ID NO. 4) và AedsxR1 (SEQ ID NO. 47). Sử dụng việc cắt nối các đoạn mồi này trong khuôn F1 sẽ tạo ra một dải bằng 283bp. Một dải khoảng 283bp phần lớn có mặt ở các con cái, mặc dù bằng chứng cho thấy một lượng nhỏ sản phẩm cắt nối này có mặt ở các con đực. Việc xác định trình tự xác nhận rằng rằng dải này quả thực tương ứng với việc cắt nối trong khuôn F1. Marker (SmartLadder™ từ Eurogentec, các dải từ 1, 5kb đến 0,2kb được chỉ ra).

Hình 5 là miêu tả bằng biểu đồ của các plasmit dựa theo sự cắt nối trong minigen *Aedes aegypti* dsx. Để rõ ràng, cần hiểu là intron con cái thứ nhất là của

việc cắt nối F1, F2 hoặc F3 bất kỳ, và tTAV trong biểu đồ liên quan đến tTAV2 (cần hiểu rõ là các protein khác hoặc các kiểu tTA hoặc tTAV khác có thể được lựa chọn sử dụng). Trong mỗi plasmit này, ngoài LA3491, trình tự khác loại được thêm vào exon F2. "ATG giả định" là trình tự codon ATG bất kỳ trong trình tự exon được định vị 5' tương ứng với ADN khác loại. Trong LA3646, các đơn vị mã hoá khởi đầu dịch mã giả định này ("ATG giả định") được loại bỏ hoặc được cải biến. Trong trường hợp cấu trúc LA3612, sự dịch mã từ một (5') ATG ngược chiều nằm trong khung với vùng mã hóa ubi-tTAV sẽ vẫn (giả sử không có codon kết thúc can thiệp) tạo ra tTAV chức năng, tiếp theo là sự tách các gốc ubiquitin và tTAV dưới sự tác động của proteaza. Các catxet (cassete) cắt nối thay thế được liên kết theo cách hoạt động được với gen khởi đầu thích hợp, gen kết thúc phiên mã và các trình tự điều hòa khác.

Hình 6 thể hiện sự sắp thẳng hàng của các exon đặc hiệu con cái thứ hai và các trình tự kẹp sườn của các gen *dsx* từ sâu hồng hại bông (Pectinophora gossypiella, PBW-dsx, SEQ ID NO. 146), con tằm (Bombyx mori, bombyx-dsx, SEQ ID NO. 147) và sâu bướm đục trái (Cydia pomonella, codling- dsx, SEQ ED NO. 148). Exon đặc hiệu con cái thứ hai được in đậm. Các tác giả đã xác định các sản phẩm phiên mã của một trình tự nulceotit ngắn, lặp lại, có trình tự bảo toàn và sự định vị gần giữa các loài bướm đêm tương đối xa này; các sản phẩm phiên mã này chỉ được đặt ở đầu 5' so với exon đặc hiệu con cái. Các đoạn lặp bảo toàn AGTGAC/T được gạch chân. Dấu hoa thị (*) là các nucleotit giống hệt nhau, dấu gạch ngang (-) là các khoảng trống để sắp thẳng hàng theo cách tốt nhất. Các exon được thể hiện trong SEQ ID NOS. bằng cách đánh số nucleotit như sau: SEQ ID NO. 146 289-439; SEQ ID NO. 147 339-492; và SEQ ID NO. 148 285-439.

Hình 7 là sơ đồ thể hiện các vị trí liên kết Tra/Tra2 giả định thuộc vùng mã hóa *dsx* của *plasmit* LA3491. Sơ đồ này có tỷ lệ gần đúng và thể hiện một trình tự khoảng 4kb.

Hình 8 cho thấy RT-PCR của các con cái và đực trưởng thành từ dòng chuyển gen LA3545AeC *Aedes aegypti* sử dụng các đoạn mồi AgexonlF (SEQ ID NO. 141) và TETRRl (SEQ ID NO. 142). Bằng cách sử dụng các đoạn mồi này, việc cắt nối trong khuôn mẫu tương đương với khuôn mẫu của gen AaAct4 nguyên thể sẽ tạo ra các dải khoảng 347bp đối với biến thể cắt nối kiểu cái và dải khoảng 595bp đối với biến thể cắt nối kiểu đực. Dải khoảng 347bp (F) chỉ được tìm thấy trong các phản ứng trên các phần chiết từ các con cái; dải khoảng 595bp (M) được tìm thấy ở cả các con đực và các con cái. Việc xác định trình tự xác nhận rằng việc cắt nối chính xác xảy ra ở cả con đực và con cái. Marker (SmartLadder™ từ Eurogentec, các dải từ 1, 5kb đến 0, 2kb được chỉ ra).

Hình 9 thể hiện RT-PCR của các con đực và các con cái từ dòng chuyển gen LA3604AeA *Aedes aegypti* sử dụng các đoạn mồi AgexonlF (SEQ ID NO. 141) và TETRRl (SEQ ID NO. 142). Bằng cách sử dụng các đoạn mồi này, việc cắt nối ở con cái sẽ tạo ra dải khoảng 575 bp, trong khi sự bao hàm exon cát-xét đặc hiệu con đực sẽ làm tăng dải đến khoảng 823bp. Dải khoảng 575bp được nhận thấy ở mỗi con cái được phân tích, trong khi một dải khoảng 823bp được nhận thấy ở mỗi con đực được phân tích. Các dải này dường như về cơ bản đặc hiệu với các giới tính tương ứng. Xác định trình tự của các dải này cho thấy việc cắt nối chính xác đã xảy ra ở cả con đực và con cái. Marker: SmartLadder™ từ Eurogentec, các dải từ 1,5kb đến 0,2kb được chỉ ra.

Hình 10 dưới đây là sơ đồ biểu thị các *plasmit* LA3545 và LA3604. S1: exon được phân chia 1; M1: trình tự thêm vào có trong exon đặc hiệu con đực 1 ; S2: exon được phân chia 2 (chỉ đầu 5'); ubi: trình tự mã hoá ubiquitin; tTAV2: trình tự mã hoá tTAV2.

Hình 11: Gel thể hiện sự cắt nối đặc hiệu giới tính của các intron thu được từ Cctra (dải 780bp ở các giống cái) loài *Ceratitis capitata* được biến nạp với LA3488. Việc cắt nối trong dạng F1 sẽ tạo ra một sản phẩm khoảng 780bp. Một băng có kích thước như vậy có thể xác định được rõ ràng từ các giống cái (dòng 4), nhưng không xác định được rõ ràng từ các giống đực, cũng không rõ ràng

trong các dòng với các phản ứng không có enzym xúc tác quá trình phiên mã ngược ("không RT"). Vì vậy, intron CcTRA thu được có khả năng cắt nối đặc hiệu giới tính thay thế trong phạm vi trình tự mới này. Dòng 1: Marker (SmartLadder™ từ Eurogentec, các băng gần đúng 0,8, 1,0 và 1,5kb được biểu thị); Các dòng 2 và 3: gen LA3488/+ của các giống đực loài *Ceratitis capitata* (lần lượt là, RT và đối chứng không RT); Các dòng 4 và 5: gen LA3488/+ của các giống cái loài *Ceratitis capitata* (lần lượt là, RT và đối chứng không RT).

Hình 12: Gel thể hiện sự cắt nối đặc hiệu giới tính của các intron thu được từ Cctra loài *Ceratitis capitata* được biến nạp với LA1038. Việc cắt nối trong dạng F1 sẽ tạo ra một sản phẩm khoảng 230bp. Một băng có kích thước như vậy có thể xác định được rõ ràng từ các giống cái (các dòng 1,2,7,8,9 và 10), nhưng không xác định được rõ ràng từ các giống đực. Vì vậy, intron CcTRA thu được có khả năng cắt nối đặc hiệu giới tính thay thế trong phạm vi trình tự mới này. Dòng 15: Marker (SmartLadder™ từ Eurogentec, các băng gần đúng của 0,2, 0,4 và 0,6kb được biểu thị); Các dòng 1, 2, 7, 8, 9 và 10: gen LA670; LA1038 của các giống cái loài *Ceratitis capitata*; Các dòng 3,4,5,6,11,12,13 và 14: gen LA670; LA1038 của các giống đực loài *Ceratitis capitata.*

Hình 13: Là ảnh chụp gel đề cập sự cắt nối đặc hiệu giới tính của các intron thu được từ Cctra loài *Ceratitis capitata* được biến nạp với LA3054. Việc cắt nối trong dạng F1 sẽ tạo ra một sản phẩm khoảng 340bp. Một băng có kích thước như vậy có thể xác định được rõ ràng trong dòng 7, nhưng không xác định được rõ ràng từ các giống đực.Vì vậy, intron CcTRA thu được có khả năng cắt nối đặc hiệu giới tính thay thế trong phạm vi trình tự mới này. Dòng 1 : Marker (SmartLadder™ từ Eurogentec, các băng gần đúng 0,4, 0,6 và 0,8kb được biểu thị); Các dòng từ 2 đến 5: gen LA3054 của các giống đực loài *Ceratitis capitata*; Dòng 7: gen LA3054 của giống cái loài *Ceratitis capitata.*

Hình 14: Gel thể hiện sự cắt nối đặc hiệu giới tính của các intron thu được từ Cctra loài *Ceratitis capitata* được biến nạp với LA3056. Việc cắt nối trong dạng F1 sẽ tạo ra một sản phẩm khoảng 200bp. Một băng có kích thước như vậy

có thể xác định được rõ ràng từ một giống cái (dòng 6), nhưng không xác định được rõ ràng từ các giống đực (các dòng từ 2 đến 4).Vì vậy, intron CcTRA thu được có khả năng cắt nối đặc hiệu giới tính thay thế trong phạm vi trình tự mới này. Dòng 1: Marker (SmartLadder™ từ Eurogentec, các băng gần đúng của 0,2, 0,4, 0,6 và 0,8kb được biểu thị); Các dòng từ 2 đến 5: gen LA3056/+ của các giống đực loài *Ceratitis capitata*; Các dòng từ 6 đến 7: gen LA3056/+ của các giống cái loài *Ceratitis capitata*.

Hình 15: Gel thể hiện sự cắt nối đặc hiệu giới tính của các intron thu được từ BzTra loài *Anastrepha ludens* được biến nạp với LA3376. Việc cắt nối trong dạng F1 sẽ tạo ra một sản phẩm khoảng 672bp. Một băng có kích thước như vậy có thể xác định được rõ ràng từ các giống cái (dòng 4), nhưng không xác định được rõ ràng từ các giống đực, cũng không rõ ràng trong các dòng với các phản ứng của enzym xúc tác quá trình phiên mã ngược đã được bỏ qua ("không RT"), các đoạn mồi đã sử dụng là SRY và AV3F. Vì vậy, Bztra intron thu được có khả năng cắt nối đặc hiệu giới tính thay thế trong phạm vi trình tự mới và loài này. Dòng 1: Marker (SmartLadder™ từ Eurogentec, các băng gần đúng 0,6, 0,8 và 1,0kb được biểu thị); Các dong 2 và 3: gen LA3376/+ của các giống đực loài *Anastrepha ludens* (lần lượt là, RT và đối chứng không RT); Các dòng 4 và 5: gen LA3376/+ của các giống cái loài *Anastrepha ludens* (lần lượt là, RT và đối chứng không RT).

Hình 16: biểu đồ plasmit LA3582 (SEQ ID NO. 144).

Hình 17: biểu đồ plasmit LA3576 (SEQ ID NO. 143).

Hình 18: các SEQ ID NO 149 và 150 đề cập các trình tự DSX minigen1, DSX minigen2 và biểu đồ plasmit LA3619.

Hình 19: Sử dụng yếu tố P để tạo ra sự biểu hiện đặc hiệu dòng mầm của một gen mong muốn (Gen E). Sự cài xen của yếu tố P IVS3 và các trình tự exon kẹp sườn ngược chiều dung hợp ubiquitin-Gen E cho phép sự biểu hiện đặc hiệu dòng mầm của gen E dưới dạng một trình tự khởi đầu hoạt hóa dòng mầm. A - Trình tự khởi đầu hoạt hóa dòng mầm - A; B-P yếu tố khung đọc mở ; C - P

intron TVS3'; D - Ubiquitin; E - vùng mã hóa cho protein mong muốn, ví dụ tTAV.

Hình 20: Sự biểu hiện đặc hiệu giới tính sử dụng gen dsx. A: Intron được sử dụng như intron CcTRA ở trên, nhưng mang lại sự biểu hiện đặc hiệu ở giống đực. Một đoạn của gen dsx (ở đây là phiên bản của *Anopheles*) được cài xen vào một vùng mã hóa khác loại (hộp được đánh bóng). Intron được loại bỏ hoàn toàn ở các con đực, nhưng ở các con cái vùng mã hóa bị kết thúc sớm. B: Một phương pháp biểu hiện đặc hiệu giống đực khác, mà ở đó một vùng mã hóa khác loại được dung hợp với một đoạn của gen dsx. C: Sự biểu hiện đặc hiệu giống cái: vùng mã hóa khác loại được cài xen vào exon đặc hiệu giống cái, hoặc dung hợp trong khung với một đoạn của gen Dsx, hoặc với các codon khởi đầu và kết thúc của chính nó. D: Sự biểu hiện khác nhau: Các biểu hiện B và C có thể được tổ hợp để mang lại sự biểu hiện gen a ở các giống cái và gen b ở các giống đực.

Hình 21: Sự cắt nối thay thế đặc hiệu giới tính của Cctra. A: Cctra được cắt nối ở các giống cái để thu được ba sản phẩm phiên mã: F1, mà mã hóa protein Tra chức năng, và M1 và M2, mà không mã hóa cho protein này, bởi vì chúng bao gồm các exon bổ sung với các codon kết thúc (được vẽ lại bởi Pane et al. 2002). Các con đực chỉ tạo ra các sản phẩm phiên mã M1 và M2 và vì vậy không hề tạo ra protein Tra chức năng. B: Nếu intron này có chức năng tương tự trong vùng mã hóa khác loại, điều này tương tự sẽ cho phép giống cái, nhưng không cho phép giống đực, sản sinh ra protein X chức năng.

Hình 22: Biểu đồ mô tả cấu trúc của plasmit pLA3435 (SEQ ID NO.46).

Hình 23: Bản đồ plasmit của gen dsx pLA3359 ở loài *Anopheles gambiae* được đặt dưới sự kiểm soát của một trình tự khởi đầu Hr5-IE1 để đánh giá sự cắt nối thông qua sự biểu hiện tạm thời.

Hình 24: gen dsx pLA3433 ở loài *Anopheles gambiae* được đặt dưới sự kiểm soát của một trình tự khởi đầu Hr5-IE1, với sự bổ sung của exon 2, để đánh giá sự cắt nối thông qua sự biểu hiện tạm thời.

Hình 25: Biểu đồ mô tả cấu trúc của pLA1188.

Hình 26: Biểu đồ mô tả cấu trúc của pLA3077.

Hình 27: Biểu đồ mô tả cấu trúc của pLA3097.

Hình 28: Biểu đồ mô tả cấu trúc của pLA3233.

Hình 29: Biểu đồ mô tả cấu trúc của pLA3014.

Hình 30: Biểu đồ mô tả cấu trúc của pLA3166.

Hình 31: Biểu đồ mô tả cấu trúc của pLA3376.

Hình 32: Biểu đồ mô tả cấu trúc của pLA3242.

Hình 33: Trình tự kẹp sườn của gen Cctra. Sự cắt nối của intron CcTRA trong LA3077 và LA3097 chính là trình tự mà bạn sẽ nhìn thấy trong intron CcTRA tự nhiên. Sự cắt nối trong LA1188 dẫn đến sự loại bỏ của 4 nucleotit bổ sung. Trong mọi trường hợp các intron đều được kẹp sườn bởi 5' exon TG và 3' GT.

Hình 34: Gel thể hiện sự cắt nối đặc hiệu giới tính chính xác của các intron thu được từ CcTra (dải 776bp ở giống cái) loài *Ceratitis capitata* được biến nạp với LA3077. Đường 1: Dấu chuẩn (SmartLadder™ từ Eurogentec, các dải khoảng 0,8, 1,0 và 1,5kb được chỉ ra); Các đường 2 và 3: gen LA3077/+ của giống đực loài *Ceratitis capitata*; Các đường 4 và 5: gen LA3077/+ của giống cái loài *Ceratitis capitata*.

Hình 35: Dữ liệu kiểu hình đối với các cấu trúc đặc hiệu giống cái đã được biến nạp ở loài *Ceratitis capitata.* Cột 1: Cấu trúc có tên gọi LA#, ví dụ, LA3077, LA3097, LA3233, v.v., được biểu thị bằng con số, với các dòng cài xen độc lập được đề cập bằng các chữ cái; Cột 2 và 3: Kết quả không tetraxyclin (Non-tetraxyclin - NT) đối với mỗi dòng được biến nạp trong tổng số các con đực (4) và tổng số các con cái (5).

Hình 36: Các sản phẩm phiên mã của các cấu trúc intron CcTRA ở loài *Drosophila* và *Ceratitis capitata.* Đường trên cùng mô tả cấu trúc ADN chứa intron gen chuyển được kẹp sườn bởi gen mong muốn (hộp mở). Hộp màu đỏ là các exon đặc hiệu giống cái. Các intron được biểu thị bởi các đường nét liền.

Mũi tên ở trên đường đầu tiên là các vị trí của các oligonucleotit được sử dụng trong các thí nghiệm RT-PCR. Thang đo biểu thị tỷ lệ của hình.

Hình 37: Gel thể hiện sự cắt nối đặc hiệu giống cái chính xác của trình tự có nguồn gốc CcTra (dải 508bp) ở giống cái loài *Ceratitis capitata* được biến nạp với LA3014. Đường 1: Dấu chuẩn (SmartLadder™ từ Eurogentec, các dải 0,4 và 1,0kb được chỉ ra); Đường 2: gen LA3014/+ của giống đực loài *Ceratitis capitata*; Đường 4: gen LA3014/+ của giống cái loài *Ceratitis capitata*; Các đường 3 và 5: không mẫu đối chứng âm tính enzym phiên mã ngược (các băng cơ sở, chắc chắn từ ADN bộ gen, có thể nhìn thấy được trong các đường 2 và 4).

Hình 38: Dữ liệu kiểu hình cho gen chuyển ở loài *Anastrepha ludens* được biến nạp với LA3097 hoặc LA3233. Cột 1: Cấu trúc LA# (LA3097 hoặc LA3233) được chỉ ra, với các dòng cài xen độc lập được đề cập bằng các chữ cái; Cột 2 và 3: Kết quả không tetraxyclin (Non-tetraxyclin - NT) đối với mỗi dòng biến nạp đã nêu trong tổng số con đực (2) và tổng số con cái (3). Cột 4 và 5: Kết quả tetraxyclin (Tetraxyclin - TET) đối với mỗi dòng biến nạp đã nêu trong tổng số con đực (4) và tổng số con cái (5).

Hình 39: Gel thể hiện sự cắt nối đặc hiệu giới tính chính xác của sự cắt nối CcTRA (dải 348bp ở các giống cái) loài *Anastrepha ludens* được biến nạp với LA3097. Đường 1: Dấu chuẩn (SmartLadder™ từ Eurogentec, các dải khoảng 0,4 và 1,0kb được chỉ ra); Các đường 2, 3 và 4: gen LA3097/+ ở các giống đực loài *A. ludens*; Các đường 5, 6 và 7: gen LA3097/+ ở các giống cái loài *A. ludens*;

Hình 40: Gel thể hiện sự cắt nối đặc hiệu giới tính chính xác của BzTRA trong các vùng reaperKR (dải 200bp ở các giống cái) và tTAV3 (dải 670bp ở các giống cái loài) của LA3376, ở loài *Ceratitis capitata* được biến nạp với LA3376. Đường 1: Dấu chuẩn (SmartLadder™ từ Eurogentec, các dải khoảng 0,2, 0,6 và 1,0kb được chỉ ra); Các đường 2 và 3: gen LA3376/+ của các giống đực loài *C. capitata* được thử nghiệm cắt nối trong reaperKR; Các đường 4 và 5: gen LA3376/+ của các giống cái loài *C. capitata* được thử nghiệm cắt nối trong

reaperKR; Đường 6: SmartLadder™; Các đường 7 và 8: gen LA3376/+ của các giống đực loài *C. capitata* được thử nghiệm cắt nối trong tTAV; Các đường 9 và 10: gen LA3376/+ của các giống cái loài *C. capitata* được thử nghiệm cắt nối trong tTAV; Đường 11: SmartLadder™.

Hình 41: Gel thể hiện sự cắt nối CrTRA đặc hiệu giới tính chính xác trong CrTRA-reaperKR (dải 200bp ở các giống cái) loài *Ceratitis capitata* được biến nạp với LA3242. Đường 1: Dấu chuẩn (SmartLadder™ từ Eurogentec, các dải khoảng 0,2, 0,6 và 1,0kb được chỉ ra); Các đường từ 2 đến 7: Các giống đực loài *C. capitata* kiểu dại được tiêm với LA3242; Đường 8: SmartLadder™; Các đường từ 9 đến 14: Các giống cái loài *C. capitata* kiểu dại được tiêm với LA3242; Đường 15: SmartLadder™.

Hình 42: Biểu đồ mô tả các cấu trúc minigen Bmdsx. Hai cấu trúc minigen thu được từ *Bombyx mori* gen *dsx* được minh họa bằng đồ thị, cũng như sự cắt nối thay thế được dự báo trước của các cấu trúc này (cấu trúc mẫu giống cái được chỉ ra ở trên, mẫu giống đực ở dưới). (A) là cấu trúc minigen *dsx* ở loài *Bombyx mori* đã sử dụng trong Funaguma et al., 2005) (B) là pLA3435. A và B khác nhau ở vài đoạn: (i) Exon 1 được loại bỏ khỏi pLA3435 (ii) intron ở giữa các exon 3 và 4 đặc hiệu giống cái đã được loại bỏ và một trình tự ngắn khác loại đã được biến nạp vào trong pLA3435 (iii) Funaguma et al., sử dụng trình tự khởi đầu ie1 từ loài baculovirut *Bm*NPV và một *Bm* A3 3'UTR trong khi pLA3435 sử dụng trình tự khởi đầu tăng cường hr5-IE1 từ *Ac*NPV của baculovirut và một 3'SV40 3'UTR. (iv) pLA3435 sử dụng các trình tự intron dài hơn một chút so với (A) (xem trình tự ở Hình 15). Hai cấu trúc minigen thu được từ *Bombyx mori* gen *dsx* được minh họa bằng đồ thị, cũng như sự cắt nối thay thế được dự báo trước của các cấu trúc này (cấu trúc mẫu giống cái ở trên, mẫu giống đực ở dưới).

Hình 43: Sự cắt nối đặc hiệu giới tính của cấu trúc minigen BMdsx ở PBW. Phân tích sự biểu hiện tạm thời từ pLA3435 bằng cách sử dụng kỹ thuật RT-PCR chỉ ra sự có mặt của một đoạn 442bp (Các đường1,2,3 và 4) ở các giống

đực và một đoạn 612bp ở các giống cái (Đường 5), chỉ ra rằng minigen BMdsx với một đoạn khác loại đã được cài xen vào giữa exon 3 và 4 có khả năng cắt nối chính xác ở các loài bướm đêm khác nhau, PBW. Các dấu chuẩn là SmartLadder™ từ Eurogentec; các dải khoảng 0,2, 0,4 và 0,6kb được chỉ ra.

Hình 44: Sự cắt nối đặc hiệu giới tính của *Anopheles gambiae dsx.* Anopheles (A) đề cập sự cắt nối mà đã được báo cáo bởi Scali và cộng sự 2005. Tuy nhiên, khi kỹ thuật RT-PCR được thực hiện bằng cách sử dụng các đoạn mồi của sáng chế (spl-agdsx-e3 (SEQ ID NO. 60) và spl-agdsx-m (SEQ ID NO. 61)) một mẫu cắt nối khác đối với giống cái đã được bộc bộ, được thể hiện bởi Anopheles (B).

Hình 45: Nhận biết *Anopheles gambiae* giống đực và giống cái bằng cách sử dụng các mồi dsx.

ARN được tách chiết từ giống đực và giống cái loài *Anopheles gambiae* và các sản phẩm phiên mã dsx đã được khuếch đại bằng kỹ thuật RT-PCR sử dụng các đoạn mồi spl-agdsx-e3 (SEQ ID NO. 62) và spl-agdsx-m (SEQ ID NO. 63); mẫu dải thu được được chỉ ra trong gel nêu trên. Các dải được mong đợi cho các sản phẩm phiên mã của giống đực và giống cái được biểu thị bởi các mũi tên màu trắng, các dải này được tách dòng và giải trình tự và giống hệt trình tự được dự báo trước của các sản phẩm phiên mã dsx theo sáng chế (xem SEQ ID NO. 47 (LA3359) và SEQ ID NO. 48 (LA3433)). Trọng lượng phân tử của các dấu chuẩn được thể hiện bằng kb (SmartLadder™ từ Eurogentec; các kích thước là gần đúng).

Hình 46: Nhận biết *Stegomyia aegypti* giống đực và giống cái bằng cách sử dụng các đoạn mồi dsx. Các đoạn mồi dùng cho qúa trình RT-PCR đối với loài *Stegomyia aegypti* đối với A và B là aedesxF1 (SEQ ID NO. 64) và aedesxR5 (SEQ ID NO. 65) đã được thử nghiệm ban đầu trên các nhộng, một trạng thái sống của loài *Stegomyia aegypti* mà có thể xác định giới tính thuận tiện và chính xác; quá trình khuếch đại RT-PCR được chỉ ra trên hình ảnh gel (A). Nhộng đực và cái thể hiện dải đặc hiệu giới tính khác biệt. Sau đó các đoạn mồi đã được thử

nghiệm trên dịch chiết ARN từ các ấu trùng, mà không thể xác định được giới tính dễ dàng bằng hình thái học của chúng và quá trình khuếch đại RT-PCR được chỉ ra trên hình ảnh gel (B). Ấu trùng thể hiện một mẫu dải rõ rệt mà phân biệt một cách rõ ràng giống đực và giống cái. Hình ảnh gel (C) cho thấy một dải chiều dài khoảng 600bp từ RT-PCR sử dụng các đoạn mồi aedessxF1 và aedesxR2 (SEQ ID NO. 66) từ mỗi giống nhộng đực và cái riêng rẽ. Phân tích trình tự của dải này cho thấy biến thể cắt nối đặc hiệu giống cái mà không hề xuất hiện ở giống đực có chung phần exon mà aedesxR5 được dự báo gắn vào (exon 7, xem Hình 56). Trọng lượng phân tử của các dấu chuẩn được thể hiện bằng kb (SmartLadder™ từ Eurogentec; các kích thước là gần đúng).

Hình 47: Đồ thị mô tả một phần gen *dsx* ở loài *Stegomyia aegypti* (không theo tỷ lệ). Một đoạn của gen *dsx* ở loài *Stegomyia aegypti* được mô tả ở trên. Các exon 5a và 5b là đặc hiệu cho giống cái và exon 6 là exon đặc hiệu cho giống đực. Hai biến thể cắt nối đặc hiệu giống cái đã được phát hiện (F1 và F2) mà bao gồm các exon từ 1 đến 4, 5b, 6 và 7 (F1) hoặc 1-4,5a (F2); các sản phẩm phiên mã ở các giống đực (M1) bao gồm các exon từ 1 đến 4, 6 và 7 nhưng không bao gồm các exon 5a hoặc 5b và một sản phẩm phiên mã (C1) bao gồm các exon từ 1 đến 4 và 7 nhưng không bao gồm các exon 5a, 5b hoặc 6 được chỉ ra trong các giống đực và giống cái. Các con số sau ký tự # đối với mỗi exon có liên hệ với contig 1.370 (http://www.broad.mit.edu/annotation/disease vector/aedes aeqypti/), mà đọc theo chiều ngược lại, và sau * có liên hệ với trình tự nucleotit được chỉ ra trong SEQ ED NO. 43.

Hình 48: Đồ thị mô tả gen dsx ở loài *Stegomyia aegypti*. Toàn bộ gen dxs ở loài *Stegomyia aegypti* được mô tả ở trên Exon 5 là exon đặc hiệu cho giống cái và exon 6 là exon đặc hiệu cho giống đực được giả định. Nói chung, các sản phẩm phiên mã ở các giống cái bao gồm các exon 1, 2, 3, 4, 5 và 7, và các giống đực bao gồm các exon 1, 2, 3, 4, 6 và 7. Các con số sau ký tự # đối với mỗi exon có liên hệ với contig 1.370 (http://www.broad.mit.edu/annotation/disease vector/aedes aeqypti/) dịch mã đọc theo chiều ngược lại, và sau * có liên hệ với hình 12

Hình 49: Biểu đồ plasmit của pLA 1172. Vùng mã hóa cho tTAV đã được đặt dưới sự kiểm soát của một đoạn từ gen actin -4 của loài *Stegomyia aegypti* (Munoz et al 2005) mà bao gồm 5' UTR, intron thứ nhất, và các trình tự ngược chiều (trình tự khởi đầu giả định). Cấu trúc này cũng bao gồm một trình tự $tetO_7$ Nipper. Cấu trúc này có các đầu *piggyBac* và một dấu chuẩn DsRed2 để hợp nhất ổn định vào bộ gen.

Hình 50: Sự cắt nối đặc hiệu giới tính của tTAV ở các thể biến nạp LA1172.

Hinh ảnh gel của quá trình RT-PCR đối với ARN được tách chiết từ LA1172 đường 2 của các giống nhộng đực và cái. Các đoạn mồi được sử dụng là Agexonl (SEQ ID NO. 67) và Tra (tTAV) seq+ (SEQ ID NO. 68). Giải trình tự các dải RT-PCR cho thấy sự cắt nối mong đợi xuất hiện ở các giống đực và cái. Dữ liệu chỉ ra trong đồ thị nêu trên là đối với LA1172 đường 2, đường 8 đã chỉ ra một cách chính xác các kết quả tương tự (dữ liệu không thể hiện). Các dấu chuẩn là SmartLadder™ từ Eurogentec; các kích thước được biểu thị là gần đúng, tính bằng kb).

Hình 51: RT-PCR đối với các mẫu kiểu dại cho thấy các biến thể cắt nối đặc hiệu giới tính của gen Actin-4 ở loài *Stegomyia aegypti*.

Hình ảnh gel của quá trình RT-PCR đối với ARN được tách chiết từ các giai đoạn phát triển khác nhau, và giải phẫu các cá thể trưởng thành, của LA1172 đường 8. Các đoạn mồi đã sử dụng là Agexon1 (SEQ ID NO. 69) và Exon 3 (SEQ ID NO. 70). Hình ảnh gel cho thấy sự biểu hiện rõ rệt từ gen Actin - 4 chỉ xuất hiện ở giai đoạn nhộng, và sự biểu hiện của các cá thể trưởng thành thường bị hạn chế ở phần giữa của giống cái ở đó có các cơ hoạt động. Bảng 17 dưới đây đề cập trạng thái của mỗi đường.

Bảng 1

E = tổ hợp gồm ~ 100 phôi	MH = phần đầu từ giống đực trưởng thành
L4 = ấu trùng tuổi thứ 4	MT = phần giữa từ giống đực trưởng thành
ME = nhộng đực non (<4 giờ tuổi)	MA = phần bụng từ giống đực trưởng thành
FE = nhộng cái non (<4 giờ tuổi)	FH = phần đầu từ giống cái trưởng thành
MP = nhộng đực	FT = phần giữa từ giống cái trưởng thành
FP = nhộng cái	FA = phần bụng từ giống cái trưởng thành
	-ve = đối chứng nước

Các hình 52-58, 68 và 69 thể hiện các biểu đồ và trình tự plasmit khác nhau.

Hình 59 cho thấy RT-PCR của các con đực và các con cái từ dòng chuyển gen LA3097 *Aedes aegypti* sử dụng các đoạn mồi HSP (SEQ ID NO. 139) và VP16 (SEQ ID NO. 140). Bằng cách sử dụng các đoạn mồi này, việc cắt nối theo kiểu CcF1 (tương ứng với biến thể F1 của Ceratitis capitatd) sẽ tạo ra một dải xấp xỉ 765bp và việc cắt nối trong CcM1 1005bp và CcM2 1094bp. Trong cả các con đực và các con cái, dải mạnh khoảng 950bp (1) được quan sát với một dải yếu hơn khoảng 800bp (2). Marker (SmartLadder™ from Eurogentec, các dải từ 1,5kb đến 0,4kb được biểu thị). Phân tích trình tự của các dòng từ dải 2 (các biết thể cắt nối AeM1/AeFl chẳng hạn) từ các con đực và các con cái cho thấy rằng một trong số năm dòng từ các con cái thể hiện việc cắt nối AeM2 (20%), trong khi trong các con đực, ba trong số bốn dòng đã thể hiện việc cắt nối AeM2 (75%); tất cả các dòng khác thể hiện việc cắt nối AeF1. Điều này chỉ ra rằng có nhiều sản phẩm phiên mã AeF1 có trong các con cái hơn có trong các con đực và điều này sẽ giải thích tác động gây chết khác nhau giữa chúng.

Hình 60 minh họa các sản phẩm phiên mã khác nhau được sinh ra bởi việc cắt nối thay thế của Cctra từ dòng chuyển gen LA3Q91A.*Aedes aegypti*. 3097 là trình tự ADN của Cctra và các số chỉ dẫn hình vẽ được mô tả ở đây. Phần bóng và đóng khung cũng liên quan đến Hình 33. Chú ý rằng biểu đồ này không theo tỷ lệ.

Các hình từ 61 đến 66 đề cập rõ hơn đến các biểu đồ và trình tự plasmit khác nhau.

Hình 61 là biểu đồ plasmit pLA3054.

Hình 62 là biểu đồ plasmit pLA3056.

Hình 63 là biểu đồ plasmit pLA3488.

Hình 64 là biểu đồ plasmit pLA3641.

Hình 65 là biểu đồ plasmit pLA3570.

Hình 66 là biểu đồ plasmit #1188.

Hình 67 là pLA3596, như đã mô tả ở dưới đây.

Mô tả chi tiết sáng chế

Khi tham chiếu đến trình tự nucleotit hoặc protein, cần hiểu rằng bao gồm việc tham chiếu đến thể đột biến hoặc biến thể bất kỳ của nó, có hoạt tính sinh học về cơ bản tương đương. Tốt nhất là, thể đột biến hoặc biến thể có mức độ tương đồng về trình tự ít nhất 85%, tốt hơn là ít nhất 90%, tốt hơn nữa là ít nhất 95%, tốt nhất là ít nhất 99% và tốt hơn nữa là ít nhất 99,99% so với với trình tự tham chiếu.

Các trình tự được cung cấp có thể có một số cải biến về trình tự, nhưng vẫn được cắt chính xác. Có một số nucleotit được biết là quan trọng. Các nucleotit này là các nucleotit cần thiết đối với việc cắt nối, ví dụ như được chỉ ra trên hình 34 dưới đây. GU ban đầu và AG cuối cùng của intron có vai trò quan trọng đặc biệt, và vì vậy tốt hơn là như đã đề cập, chỉ có khoảng 5% intron GC ban đầu được thay thế. Trình tự liên ứng này được ưu tiên, mặc dù nó áp dụng cho tất cả việc cắt intron, nhưng không đặc hiệu để cắt nối thay thế. Trên hình 34, Pu = A hoặc G; Py = C hoặc U.

Tốt hơn là, hệ thống này là hoặc chứa một plasmit. Như đã nêu ở trên, hệ thống này có thể là ADN, ARN hoặc là hỗn hợp của chúng. Nếu hệ thống này chứa ARN thì nó có thể thích hợp hơn để phiên mã ngược ARN này thành and

bằng enzyme phiên mã ngược. Nếu cần đến sự phiên mã ngược, hệ thống này có thể còn chứa một trình tự mã hóa protein RT và một trình tự khởi đầu phù hợp của nó. Ngoài ra, RTaza và trình tự khởi đầu của nó có thể được cung cấp bởi một hệ thống riêng, chẳng hạn một virut. Trong trường hợp này, hệ thống này sẽ chỉ được hoạt hóa sau khi nhiễm virut. Người có hiểu biết trung bình trong lĩnh vực kỹ thuật này hiểu rằng enzyme phiên mã ngược hoặc ARN polymeraza phụ thuộc ARN này cần phải chứa trình tự hoạt hóa cis phù hợp.

Tuy nhiên, đặc biệt tốt nếu hệ thống này chứa chủ yếu là ADN và tốt hơn nữa nếu hệ thống này chỉ chứa ADN, ít nhất đối với trình tự được biểu hiện ở côn trùng.

Đã bộc lộ rằng ít nhất một trình tự polynucleotit khác loại được biểu hiện ở côn trùng là trình tự polynucleotit của ARN can thiệp (RNAi), được ưu tiên đặc biệt nếu trình tự polynucleotit này là trình tự polynucleotit có khả năng mã hóa cho một protein chức năng. Bản mô tả sẽ tập trung chủ yếu vào các trình tự polynucleotit mã hóa protein chức năng, nhưng cần hiểu rằng bản mô tả này cũng đề cập cả các polynucleotit của ARN can thiệp (ARNi), trừ tường hợp có quy định khác.

Cần hiểu rằng tham chiếu được tạo ra đối với codon khởi đầu và codon kết thúc mà trình tự polynucleotit giữa chúng được biểu hiện ở côn trùng được xác định, nhưng điều này không loại trừ sự định vị của ít nhất một trình tự kiểm soát sự cắt nối, các yếu tố của nó, hoặc các trình tự khác trong vùng này, chẳng hạn intron. Trên thực tế, cần hiểu rõ từ bản mô tả này rằng trong một số trường hợp, trình tự kiểm soát sự cắt nối có thể được đặt trong vùng này.

Hơn nữa, chẳng hạn, trình tự kiểm soát sự cắt nối có thể chồng lấn lên codon khởi đầu, trong trường hợp này, G của ATG có thể là G 5' ban đầu của trình tự kiểm soát sự cắt nối. Do đó, thuật ngữ "giữa" có thể được xem như là đề cập đến phần đầu của codon khởi đầu (nucleotit thứ nhất của đầu 3', là A), tốt nhất là nucleotit thứ hai của codon khởi đầu của đầu 3' (là T), đến nucleotit đầu tiên của codon kết thúc của đầu 5'. Ngoài ra, rõ ràng là bằng cách đọc một cách

đơn giản một trình tự polynucleotit, trình tự này có thể bao gồm cả codon kết thúc.

Ít nhất một trình tự polynucleotit khác loại được biểu hiện ở côn trùng là trình tự khác loại. "Khác loại" ở đây được hiểu là trình tự mà ở kiểu dại không được tìm thấy trong tổ hợp, hoặc liên kết với ít nhất một yếu tố hoặc một thành phần của ít nhất một trình tự điều khiển cắt intron. Ví dụ, khi trình tự kiểm soát sự cắt nối bắt nguồn từ một sinh vật cụ thể, và polynucleotit khác loại có trình tự mã hóa cho một protein hoặc polypeptit, nghĩa là nó là một trình tự polynucleotit mã hóa cho một protein chức năng, thì trình tự mã hóa này có thể có nguồn gốc, một phần hoặc toàn bộ, từ một gen của cùng sinh vật này, với điều kiện là nguồn gốc của ít nhất một phần trình tự polynucleotit đã phiên mã không giống với nguồn gốc của ít nhất một trình tự kiểm soát sự cắt nối. Ngoài ra, trình tự mã hóa có thể là từ một sinh vật khác và trong bản mô tả này có thể là trình tự ngoại sinh. Polynucleotit khác loại cũng có thể là "tái tổ hợp" trong đó trình tự mã hóa cho một protein hoặc polynucleotit được bắt nguồn từ các vị trí khác nhau, hoặc trong cùng hệ gen (có nghĩa là hệ gen của một loài hoặc loài phụ) hoặc từ các hệ gen khác (có nghĩa là các hệ gen từ các loài hoặc loài phụ khác).

Trình tự khác loại có thể liên quan đến một trình tự khác trình tự kiểm soát sự cắt nối và vì vậy có thể liên quan đến trình tự khởi đầu và các trình tự khác chẳng hạn 5' UTR và/hoặc 3'UTR có thể là khác loại với trình tự polynucleotit được biểu hiện ở côn trùng, với điều kiện là trình tự polynucleotit không kết hợp hoặc có thể liên kết theo cách hoạt động được với trình tự khởi đầu, 5' UTR và/hoặc 3 'UTR ở kiểu dại, nghĩa là các kiểu trình tự polynucleotit có trong tự nhiên, nếu có.

Cũng cần hiểu rằng, dạng khác loại cũng áp dụng cho các trình tự "thiết kế" hoặc các trình tự lai mà không bắt nguồn từ sinh vật cụ thể, nhưng dựa trên một số thành phần từ các sinh vật khác nhau, điều này cũng sẽ đáp ứng yêu cầu về việc trình tự và ít nhất một thành phần của trình tự kiểm soát sự cắt nối không liên kết hoặc kết hợp trong kiểu dại, ngay cả khi một phần hoặc yếu tố của trình

tự lai được tìm thấy cho đến khi không có ít nhất một phần hoặc yếu tố. Tốt hơn là, một phần gồm ít nhất 50 nucleotit của trình tự lai không được kết hợp với ít nhất một thành phần của trình tự kiểm soát sự cắt nối, tốt hơn nữa là 200 nucleotit, và tốt nhất là 500 nucleotit.

Cũng cần hiểu rằng các phiên bản tổng hợp của các trình tự xuất hiện trong tự nhiên được dự tính. Các trình tự tổng hợp như vậy cũng được coi như khác loại, trừ khi chúng là trình tự giống hệt với trình tự mà ở dạng tự nhiên hay kiểu dại thì thường được tìm thấy ở dạng kết hợp hoặc liên kết với ít nhất một yếu tố hoặc một thành phần của ít nhất một trình tự kiểm soát sự cắt nối.

Điều này áp dụng tương tự đối với polynucleotit khác loại là một polynucleotit của ARN can thiệp.

Theo một khía cạnh của sáng chế, khi trình tự polynucleotit được biểu hiện bao gồm một trình tự mã hóa cho một protein hoặc polynucleotit, cần hiểu rằng tham chiếu đến sự biểu hiện ở côn trùng dùng để chỉ việc tạo ra một hay nhiều trình tự ARN phiên mã, tốt nhất là mARN trưởng thành, nhưng có thể, tốt hơn là, cũng dùng để chỉ việc dịch mã polynucleotit trong côn trùng đó.

RT-PCR mà chứng tỏ sự có mặt của sản phẩm phiên mã không phải protein có thể được sử dụng để nhận biết các trình tự ARN phiên mã. Nó cũng đặc biệt hữu hiệu khi bản thân protein này không được dịch mã hoặc không có chức năng hoặc không được nhận biết bằng các kháng thể được tạo ra để kháng lại protein xuất hiện trong tự nhiên hoặc kiểu dại, do ARNi, sự cải biến sau dịch mã hoặc sự gấp nếp gây biến dạng.

Khi trình tự polynucleotit được biểu hiện bao gồm polynucleotit của ARN can thiệp, cần hiểu rằng tham chiếu đến sự biểu hiện ở côn trùng dùng để chỉ sự tương tác của các polynucleotit của ARN can thiệp, hoặc sản phẩm phiên mã của nó, theo con đường ARNi, chẳng hạn bằng cách liên kết Dicer hoặc nhờ sự hình thành của ARN can thiệp nhỏ (siARN). Thực vậy, đặc biệt tốt hơn là các polynucleotit của ARN can thiệp bao gồm các trình tự siARN và vì thế, tốt hơn là có chiều dài từ 20 đến 25 nucleotit, đặc biệt là khi sinh vật là động vật có vú.

Ở các loài côn trùng và đặc biệt là các loài giun tròn, tốt hơn là tạo ra phần dsARN, ví dụ tạo thành cấu trúc kẹp tóc, cấu trúc kẹp tóc này sau đó có thể được xử lý bằng hệ thống Dicer. Các tế bào của động vật có vú thường tạo ra một đáp ứng can thiệp kháng lại các trình tự dsARN dài, vì vậy đối với tế bào động vật có vú thường tạo ra các trình tự ngắn hơn như siARN. Các trình tự đối mã hoặc các trình tự tương đồng với các ARN nhỏ là các phân tử mARN xuất hiện trong tự nhiên hướng đích protein 3' UTR cũng được dự định làm các trình tự ARNi theo một phương án của sáng chế.

Mỗi trình tự kiểm soát sự cắt nối trong hệ thống có chứa ít nhất một vị trí nhận intron và có chứa ít nhất một vị trí cho intron. Số lượng các vị trí cho và nhận có thể thay đổi, tùy vào số lượng đoạn trình tự được cắt intron cùng nhau. Tốt hơn là, vị trí phân nhánh được bao hàm trong mỗi trình tự kiểm soát sự cắt nối. Một vị trí phân nhánh là một trình tự mà thể cho cắt nối được kết hợp vào ban đầu, xem hình 32, thể hiện việc cắt intron xuất hiện ở hai giai đoạn, trong đó exon 5’ được cắt ra và sau đó được nối vào exon 3’.

Trong hình này, A là nucleotit thiết yếu duy nhất, và vì vậy tốt hơn là được bao hàm. Không bị giới hạn bởi lý thuyết, người ta tin rằng việc cắt intron ra khỏi tiền mARN xảy ra thông qua một bước trung gian, như là cắt nối nhóm II. Đầu tiên, xuất hiện phân cắt tại một số mối nối 5’ đôi khi được gọi là vị trí thể cho. Sau đó, photphat ở đầu 5’của intron được nối với OH 2’ của adenin khoảng nucleotit 25 ngược chiều đầu 3’ của intron, đôi khi được gọi là vị trí nhận. Phần còn lại của phần A được gọi là điểm phân nhánh. Bước tiếp theo là phân cắt xuất hiện ở chỗ nối phân cắt intron 3’ và photphat 5’ xuôi chiều exon được nối với OH 3’ ngược chiều exon.

Được ưu tiên đặc biệt nếu cách thức hay cơ chế cắt nối thay thế là đặc hiệu giới tính. Tốt nhất là, trình tự kiểm soát sự cắt nối được bắt nguồn từ gen chuyển intron. Tuy nhiên, tốt hơn nếu cơ chế cắt nối thay thế được bắt nguồn từ gen chuyển Cctra của Medfiy, hoặc từ một gen khác cùng loài hoặc khác loài

Drosophila transformer, tốt nhất là từ *C. rosa*, hoặc *B. zotana,* đặc biệt là thu được từ ruồi giấm tephritid.

Cũng được ưu tiên nếu trình tự kiểm soát sự cắt nối được bắt nguồn từ cơ chế cắt nối thay thế của gen Actin-4, cụ thể là từ *Aedes* spp. và tốt nhất là từ AaActin-4, mà là gen từ *Aedes/Stegomyia aegypti* biểu hiện sự cắt nối trong mô, theo giai đoạn, đặc hiệu giới tính.

Tốt hơn là, sự cắt nối thay thế, đặc biệt là sự cắt nối thay thế qua trung gian Actin-4, có thể bổ sung các trình tự ảnh hưởng đến sự dịch mã ARN hoặc độ ổn định chẳng hạn.

Được ưu tiên nếu cơ chế cắt nối bao gồm ít nhất một đoạn gen lưỡng tính doublexex (dsx), tốt hơn nếu có nguồn gốc từ *Drosophila, B. mori, Pink Boll Worm, Codling Moth*, hoặc muỗi, *A. gambiae* hoặc đặc biệt từ *A. aegypti.*

Tốt hơn nếu trình tự điều khiển sự cắt nối thay thế và trình tự polynucleotit khác loại mã hóa cho một protein chức năng, nằm giữa codon khởi đầu và codon kết thúc. Các polynucleotit của ARN can thiệp (ARNi), được biểu hiện ở côn trùng mà được tạo ra ở dạng cấu trúc minigen hoặc catxet exon cũng được bộc lộ.

Được ưu tiên nếu trình tự kiểm soát sự cắt nối được bắt nguồn từ dsx (tốt nhất là minigen 1 như được mô tả trong phần ví dụ và được biểu thị trong SEQ ID NO. 149 (các exon có mặt ở các vị trí 1-135, 1311-2446 và 3900-4389 của SEQ ID NO. 149) mà được bao hàm trong cấu trúc LA3491) hoặc Actin-4.

Các ví dụ được ưu tiên đặc biệt theo sáng chế được nêu trong phần các ví dụ thực hiện sáng chế, và có thể được lựa chọn từ nhóm gồm các plasmit hoặc cấu trúc, điển hình là các ví dụ theo một trong số các Hình từ 19 đến 31, cụ thể là plasmit bất kỳ được thể hiện trên các Hình 16-18, 22- 24, 26-32, 49, 52-55, và 61-69, và/hoặc các SEQ ID NO. 46-48, 50-56, 143-145 và 151-162.

Tốt hơn là, protein chức năng được biểu hiện ở côn trùng là tTAV, tTAV2 hoặc tTAV3.

Hơn nữa, các protein biểu hiện ở côn trùng là ở dạng kết hợp với protein chức năng đã nêu, tốt hơn là một gen gây chết như được đề cập ở trên.

ORF liên tục cũng có thể được đề cập đến như là một ORF không gián đoạn, nghĩa là trình tự polynucleotit trong mARN trưởng thành, không chứa các nucleotit không mã hóa, chẳng hạn các nucleotit có khả năng dịch mã thành các axit amin. Trong định nghĩa này, tốt hơn là condon kết thúc được bao hàm

Theo một số phương án, ít nhất một trình tự kiểm soát cắt nối sẽ điều chỉnh sự cắt nối thay thế bằng cả nucleotit intron và exon. Tuy nhiên, theo một phương án, được ưu tiên đặc biệt nếu ít nhất một trình tự kiểm soát sự cắt nối là trình tự kiểm soát cắt nối intron. Nói cách khác, tốt hơn là ít nhất một trình tự kiểm soát cắt nối về cơ bản có nguồn gốc từ ccas polynucleotit, mà tạo thành phần intron và vì vậy được tách ra khỏi sản phẩm phiên mã sơ cấp bằng cách cắt nối, sao cho các nucleotit này không còn lại trong trình tự mARN trưởng thành.

Do đó, các trình tự intron khác biệt với các trình tự exon, mà vẫn còn lại trong phân tử mARN đã xử lý (sau cắt nối). Khi phân tử mARN đã xử lý hoàn thiện mã hóa cho một phân tử protein hoặc một trình tự polynucleotit, và có khả năng bị dịch mã, nghĩa là, có cấu trúc chính xác và các cải biến như gắn mũ và polyadenyl hóa chẳng hạn, nó được biết đến như là mARN trưởng thành hay mARN hoàn thiện và một số trình tự exon sau đó mã hóa cho các axit amin khi dịch mã.

Trong việc cắt intron, nối exon trên mARN, các trình tự có thể là intron trong một số trường hợp (chẳng hạn trong một số biến thể cắt nối), nhưng có thể là exon trong một số trường hợp khác (trong các biến thể khác). Do đó, ít nhất một trình tự điều khiển việc cắt nối theo sáng chế tốt nhất là về cơ bản được bắt nguồn từ polynucleotit mà tạo thành một phần của intron trong ít nhất một biến thể cắt nối thay thế, nghĩa là trong sản phẩm cắt nối đầu tiên của mARN hoặc ít nhất một sản phẩm cắt nối thay thế của mARN. Do đó các intron hoặc các trình tự intron có thể được xem xét như phần bị cắt ra trong ít nhất một sản phẩm phiên mã hoặc kiểu sản phẩm phiên mã.

Ví dụ, intron tra của *C. capitata* (intron Cctra) là ví dụ được ưu tiên đặc biệt của ít nhất một trình tự điều khiển việc cắt nối theo sáng chế. Theo Hình 2A của Pane và cộng sự, được tạo ra như trong Hình 33, cả 8 vị trí liên kết Tra/Tra2 giả định được đánh dấu là thuộc trình tự intron, các trình tự này là các phần của trình tự được cắt ra trong sản phẩm phiên mã F1, nhưng mặt khác 6 trong số 8 vị trí này là exon trong trường hợp chúng là các exon được bao hàm hoặc vẫn còn trong các sản phẩm phiên mã hoặc Ml hoặc M2 hoặc cả hai. Do đó, các vị trí kết nối gen Tra/Tra 2 là intron trong trường hợp chúng có khả năng điều khiển việc cắt nối thay thế, nhưng lại bị cắt khỏi, nghĩa là, không có mặt trong ít nhất một biến thể của việc cắt nối thay thế, nghĩa là ít nhất một mARN được cắt nối theo cách thay thế từ tiền mARN.

Trong quá trình cắt nối bình thường (không thay thế) và trong quá trình cắt nối thay thế, các intron thường bị tách khỏi tiền mARN để tạo thành mARN đã tách intron, mà sau đó có thể được dịch mã thành polypeptit, chẳng hạn là protein hoặc đoạn protein, mang trình tự axit amin. Do đó, người có hiểu biết trung bình trong lĩnh vực kỹ thuật này hiểu rõ cách xác định những trình tự của hệ thống này mà được xem là intron mà không phải exon.

Dĩ nhiên, cần hiểu rõ rằng chỉ một phần của mARN được dịch mã thực sự, có nghĩa là phần điển hình nằm giữa codon khởi đầu và codon kết thúc, mặc dù đôi khi có nhiều codon khởi đầu và codon kết thúc có mặt. Do đó, khi tham chiếu đến sự dịch mã của một trình tự mARN, cần hiểu rằng điều này đề cập đến sự dịch mã của phần bắt đầu từ nucleotit đầu tiên của codon khởi đầu và kết thúc sau nucleotit cuối cùng trước khi bắt đầu codon kết thúc mà được coi là phần mã hóa.

Như đã đề cập ở trên, các trình tự exon có thể liên quan tới sự điều hòa kiểm soát việc cắt nối thay thế, nhưng tốt hơn nếu ít nhất một số trình tự kiểm soát intron tham gia vào việc cắt nối thay thế. Nói cách khác, hệ thống biểu hiện gen theo sáng chế cũng có thể bao gồm trình tự kiểm soát cắt nối có mặt trong exon, miễn là có sự tham gia điều khiển của một số intron. Các ví dụ được ưu

tiên đặc biệt về các trình tự exon là các trình tự kiểm soát cắt nối bắt nguồn từ hoặc chứa các yếu tố gen dsx, mà không bị giới hạn bởi lý thuyết bất kỳ là trình tự exon trợ giúp cho cơ chế cắt nối thay thế.

Do đó, theo một số phương án, ít nhất một trình tự kiểm soát cắt nối bao gồm trình tự exon và cần hiểu rằng điều này được dự tính bởi các định nghĩa được sử dụng để mô tả trong sáng chế. Do đó, cần hiểu rõ rằng, một số nucleotit có thể được bao hàm trong định nghĩa của ít nhất một trình tự kiểm soát cắt nối và cũng thuộc định nghĩa của trình tự polynucleotit mã hóa cho một protein chức năng. Nói cách khác, định nghĩa của các yếu tố có thể chồng lên nhau, chẳng hạn một số nucleotit có thể được bao hàm bởi định nghĩa của nhiều hơn một yếu tố.

Tuy nhiên, người có hiểu biết trung bình trong lĩnh vực kỹ thuật này sẽ nhận ra đó không phải là điều bất thường trong sinh học phân tử, như là các nucleotit có thể đảm nhiệm nhiều hơn một vai trò. Chẳng hạn, theo sáng chế, một nucleotit có thể tạo thành một phần của trình tự mã hóa cho một protein chức năng, nhưng cũng có thể tạo thành một phần của trình tự phát hiện và giới hạn bởi một yếu tố cắt nối, ví dụ là protein TRA hoặc phức hợp TRA/TRA như đã được đề cập ở trên. Chẳng hạn, điều này không phải là bất thường do một số virut có hệ gen nồng độ cao, tại đó cacs polynucleotit mạch thẳng có thể mã hóa cho hai hoặc thậm chí ba protein khác nhau, mà mỗi protein được đọc trong một khung khác nhau.

Dĩ nhiên, trình tự hoặc các trình tự kiểm soát cắt nối cũng có thể là intron đơn, nghĩa là không bị ảnh hưởng bởi exon. Thực vậy, trình tự này được ưu tiên đặc biệt.

Theo một số phương án, được ưu tiên nếu ít nhất một trình tự kiểm soát cắt nối có khả năng bị loại bỏ từ tiền ARN, bằng cách cắt nối. Tốt hơn là, trình tự kiểm soát cắt nối không dẫn đến sự chuyển dịch khung ở ít nhất một biến thể cắt nối. Tốt hơn là đây là một biến thể cắt nối mã hóa cho một protein chức năng với chiều dài đầy đủ. Nói cách khác, tốt hơn là ít nhất một trình tự kiểm soát cắt

nối không gây ra việc loại bỏ nucleotit mà tạo thành một phần, hoặc được dự định để tạo ra một phần, trình tự polynucleotit mã hóa cho một protein chức năng, nằm giữa codon khởi đầu và codon kết thúc polynucleotit của ARN can thiệp (ARNi), được biểu hiện ở côn trùng cũng được bộc lộ. Bằng cách này, các nucleotit bị loại bỏ bởi cắt nối, ở ít nhất một biến thể cắt nối, không phải là nucleotit mã hóa cho các axit amin trong kiểu protein hoặc gen kiểu dại. Một hoặc nhiều biến thể cắt nối có thể có các nucleotit đã bị loại bỏ, nhưng ít nhất một biến thể phải giữ lại các nucleotit đó, để khung dịch chuyển không được bao hàm trong ít nhất một biến thể. Các nucleotit đã được loại bỏ đó là các nucleotit mà bị loại bỏ ngoài các trình tự mà bình thường bị cắt khỏi đi ví dụ như là intron.

Tuy nhiên, theo quan điểm đã nêu ở trên, các biến thể cắt nối khác có thể tạo ra cùng trình tự được dịch mã trong các khung khác nhau.

Sự tương tác của ít nhất một trình tự kiểm soát cắt nối với cơ chế cắt nối tế bào, ví dụ phức hệ spliceosome, dẫn đến hoặc gây ra sự loại bỏ một loạt, tốt hơn là 50 nucleotit liên tiếp từ sản phẩm phiên mã sơ cấp và sự ghép nối của các trình tự nucleotit không liên tiếp trong sản phẩm phiên mã sơ cấp với nhau (bởi vì chúng, hoặc các trình tự bổ trợ của chúng nếu trình tự đối mã được xem xét là không liên tục trên trình tự khuôn ban đầu mà từ đó sản phẩm phiên mã sơ cấp được phiên mã). Một loạt gồm ít nhất 50 nucleotit liên tiếp đó chứa một intron. Tốt hơn là hoạt động qua trung gian này là theo cách đặc hiệu giới tính, đặc hiệu giai đoạn, đặc hiệu phôi hoặc đặc hiệu mô, hoặc tổ hợp của chúng, sao cho sản phẩm phiên mã sơ cấp tương đương khác nhau ở giới tính, giai đoạn, kiểu mô, v.v., có xu hướng loại bỏ intron ở vị trí hoặc trình tự khác, hoặc trong một số trường hợp có thể loại bỏ một intron trong trường hợp này nhưng không loại bỏ trong trường hợp khác. Hiện tượng này, loại bỏ intron với các kích thước hay trình tự khác nhau, trong các trường hợp khác, hoặc sự loại bỏ sự biệt hóa của các intron có kích thước hoặc trình tự định trước, trong các trường hợp khác, được biết đến như việc cắt nối thay thế. Cắt nối thay thế là hiện tượng đã được biết rõ trong tự nhiên, và nhiều trường hợp là đã biết, như đã đề cập ở trên.

Theo một số phương án được ưu tiên, ít nhất một trình tự kiểm soát cắt nối được kết hợp với một khung đọc mở ngoại sinh, sao cho ở ít nhất một biến thể cắt nối, khung đọc mở ngoại sinh này bị phá vỡ, ví dụ, bằng codon kết thúc hoặc sự chuyển dịch khung trong khi ở ít nhất một biến thể cắt nối thay thế khung đọc mở ngoại sinh này không bị phá vỡ. Các sản phẩm phiên mã của kiểu thứ hai mã hóa hoặc có khả năng mã hóa cho một protein chức năng, trong khi kiểu thứ nhất mã hóa cho một protein bị biến đổi, phá vỡ hoặc thậm chí mất chức năng, hoạt tính hoặc độ ổn định so với kiểu thứ hai.

Nói chung, người có hiểu biết trung bình trong lĩnh vực kỹ thuật này hiểu rõ rằng khung đọc mở ngoại sinh có thể là liên hợp hoặc liên hợp của các trình tự từ các nguồn biến thể. Việc cắt nối để tạo ra protein chức năng có thể vẫn tạo ra một protein bị biến đổi so với khung đọc mở ngoại sinh ban đầu, ví dụ nếu cài xen intron bị cắt bỏ vào các trình tự exon trong tất cả các dạng cắt nối thay thế, và do đó giữ lại trong mARN trưởng thành của kiểu mã hóa thứ hai. Tuy nhiên, được ưu tiên đặc biệt nếu sản phẩm phiên mã loại bỏ tất cả hoặc về cơ bản là tất cả trình tự cắt nối thay thế thêm vào, khung đọc mở ngoại sinh được sửa chữa hoặc về cơ bản được sửa chữa thành dạng nguyên vẹn, với lượng nhỏ hoặc không có trình tự nội sinh kết hợp với intron còn lại trong mARN trưởng thành. Đoạn nội sinh được sử dụng ở đây trái ngược với đoạn ngoại sinh, được hiểu rằng có liên quan tới trình tự trong kiểu tự nhiên, được tìm thấy trong sự liên hợp, hoặc liên kết với yếu tố hoặc thành phần của trình tự kiểm soát cắt nối.

Ngoài ra, một hoặc nhiều sản phẩm phiên mã có thể loại bỏ các nucleotit thêm vào, khiến khung đọc mở ngoại sinh bị phá vỡ, không chỉ bởi các nucleotit thêm vào (chẳng hạn codon kết thúc hoặc chuyển dịch khung, mà còn bởi trình tự mã hóa làm mất chức năng) mà còn bởi việc loại bỏ các nucleotit từ khung đọc mở ngoại sinh, ví dụ theo cách gây ra chuyển dịch khung. Một hoặc nhiều biến thể cắt nối có thể có nucleotit bị cắt bỏ, nhưng ít nhất một một biến thể phải giữ lại các nucleotit đó, để không gây ra sự chuyển dịch khung trong ít nhất một biến thể. Các nucleotit đã loại bỏ đó là các nucleotit mà bị loại bỏ ngoài các trình tự mà thường được cắt đi như là intron, khi trình tự intron có thể được xem

xét như là trình tự tạo thành một phần của intron trong ít nhất một biến thể cắt nối thay thế của các dạng tương tự có trong tự nhiên.

Khi các nucleotit exon bị loại bỏ, sau đó chúng phải được cắt bỏ theo bội số của 3, nếu muốn ngăn chặn việc chuyển dịch khung, nhưng sẽ cắt một nucleotit hoặc bội số của 2 (mà không phải là bội số của 3) nếu muốn có sự chuyển dịch khung. Hiểu rõ rằng nếu chỉ một nucleotit hoặc một bội số của 2 nucleotit được loại bỏ thì sau đó có thể thu được trình tự protein hoàn toàn khác được mã hóa ở hoặc xung quanh vị trí nối trên mARN.

Trường hợp đặc biệt, theo một phương án của hệ thống, khi các catxet exon được sử dụng để phá vỡ khung đọc mở trong một số biến thể cắt nối, nhưng không phá vỡ trong các biến thể khác, ví dụ như gen chuyển, đặc biệt là Cctra.

Trong một khía cạnh được ưu tiên của sáng chế, tất cả hoặc một phần của khung đọc mở nằm trên catxet exon, chẳng hạn một số phương án Dsx bắt nguồn từ Aedes, được đề xuất với một vùng mã hóa tTAV trên catxet exon mà catxet exon chỉ có mặt trong các biến thể cắt nối đặc hiệu giống cái chẳng hạn.

Khi điều hòa cắt nối thay thế là đặc hiệu giới tính, được ưu tiên nếu biến thể cắt nối mã hóa cho một protein chức năng được biểu hiện ở côn trùng là biến thể cắt nối F1, có nghĩa là biến thể cắt nối chỉ được tìm thấy hoặc phần lớn được tìm thấy ở con cái và tốt hơn là biến thể phong phú nhất được tìm thấy ở con cái, mặc dù điều này là không bắt buộc. Do đó đối với các cấu hình mà tất cả hoặc một phần của khung đọc mở chức năng nằm trên một catxet exon, tốt hơn là catxet exon đó được chứa trong các sản phẩm phiên mã chỉ tìm thấy hoặc phần lớn được tìm thấy ở con đực, và tốt hơn là các sản phẩm phiên mã này, ở dạng riêng lẻ hoặc tổ hợp, là các biến thể phóng phú nhất tìm thấy ở con cái, mặc dù điều này là không cần thiết.

Trong một khía cạnh được ưu tiên, các trình tự được chứa trong trình tự lai hoặc tái tổ hợp hoặc cấu trúc mà được bắt nguồn từ các trình tự intron xuất hiện trongtự nhiên, mà bản thân chúng chịu sự cắt nối thay thế, trong tự nhiên hoặc nguồn gốc. Do đó, trình tự intron có thể được xem xét như là trình tự mà tạo

thành phần của intron trong ít nhất một biến thể cắt nối thay thế giống như trong tự nhiên. Như vậy, dự kiến bao gồm các trình tự duỗi thẳng tương ứng mạch đơn của trình tự intron xuất hiện trong tự nhiên, nhưng cũng bao gồm trình tự lai của trình tự trên, bao gồm trình tự lai từ hai trình tự intron khác nhau xuất hiện trong tự nhiên và cũng như các trình tự có sự xóa đoạn hoặc cài xen so với các mạch đơn duỗi thẳng liên tiếp của trình tự intron xuất hiện trong tự nhiên, và trình tự lai của nó. Các trình tự này xuất phát từ các trình tự intron xuất hiện trong tự nhiên có thể tự liên kết với các trình tự không phải là một phần của các intron bất kỳ xuất hiện trong tự nhiên. Nếu các trình tự này được phiên mã và tốt nhất là được giữ lại trong mARN trưởng thành trong ít nhất một biến thể cắt nối thì sau đó chúng có thể được coi là exon.

Cần hiểu rằng tham chiếu đến "sự dịch chuyển khung" cũng có thể được dùng để chỉ việc mã hóa trực tiếp của codon kết thúc, mà dường như dẫn đến việc tạo protein không có chức năng do sự phá vỡ trình tự mARN cắt nối gây ra bởi sự cài xen hoặc xóa bỏ các nucleotit. Sáng chế cũng bao gồm sự tổng hợp từ các biến thể cắt nối khác nhau của hai hay nhiều protein hoặc các trình tự polypeptit có chức năng khác nhau, ngoài ra cũng bao gồm sự tổng hợp hai hay nhiều protein hay trình tự polypeptit mà một hoặc nhiều trong số chúng có chức năng không biết trước hoặc chức năng chưa được thấy rõ. Sáng chế cũng dự kiến bao gồm việc tổng hợp từ các biến thể cắt nối khác nhau của hai hoặc nhiều protein khác nhau hoặc trình tự polypeptit có chức năng tương tự, nhưng ở vị trí khác nhau trong tế bào, ổn định hoặc có khả năng gắn kết hay liên kết với các protein hoặc các axit nucleic khác.

Tốt nhất là, trình tự kiểm soát cắt nối là intron và bao gồm nucleotit guanin (G) ở đầu 5' của nó. Nói cách khác, nucleotit 5' của trình tự trình tự kiểm soát cắt nối, 3' của vị trí cắt nối thể cho, và tốt hơn nữa là ở mặt phân cách hoặc mối nối của exon với trình tự kiểm soát cắt nối, là guanin (G), trong tiền mARN hoặc C trong trình tự ADN đối mã tương ứng của nó.

Hơn nữa, nucleotit liền kề (3’ của G) tốt nhất là cytozin (C) trong tiền mARN hoặc tương ứng với G trong trình tự ADN, nhưng tốt nhất là Uracil (U) trong tiền mARN, hoặc tương ứng với A trong trình tự đối mã ADN. Vì vậy, hai nucleotit 5’ của trình tự kiểm soát cắt nối tốt nhất là GT 5’ đối với sợi có nghĩa ADN, 5'-GU trên sản phẩm phiên mã sơ cấp.

Tốt hơn là, ít nhất một trình tự kiểm soát cắt nối intron cũng bao gồm nucleotit Guanin ở đầu 3' của nó và tốt hơn là AG-3' ở mối nối của vị trí nhận cắt nối với exon, như trong Hình 34.

Tốt hơn là, trình tự kẹp sườn 5’ của vị trí cắt nối thể cho trong hệ thống bao gồm 5'-TG, để trình tự này có thể là 5 '-TG-* - trình tự kiểm soát cắt nối -**-3', trong đó * là vị trí cắt nối thể cho ** là vị trí cắt nối thể nhận.

Tốt hơn là, trình tự kiểm soát cắt nối cũng được kẹp sườn ở phía 3’ của nó bằng nucleotit G, và tốt nhất là bằng nucleotit GT, chẳng hạn trình tự đó có thể là 5'-TG-*- trình tự kiểm soát cắt nối -**-GT-3'. Cần hiểu rõ rằng đây là trình tự ADN sợi có nghĩa (TG). Do đó, tiền ARN đã phiên mã sẽ được đọc là UG chẳng hạn, trong đó U thay thế cho T.

Các dẫn xuất của Guanin hoặc Thymin có cùng chức năng cũng được tính đến.

Được ưu tiên đặc biệt nếu sự cắt nối là đặc hiệu giới tính và hơn nữa được điều hòa hoặc kiểm soát bởi việc nối protein TRA hoặc phức hợp protein TRA/TRA2 hoặc thể tương đương của nó. Ở côn trùng chẳng hạn, protein TRA được biểu hiện khác nhau ở các giới khác nhau. Cụ thể, protein TRA có mặt phổ biến ở con cái và do đó gây ra sự cắt nối thay thế theo cách mà một trình tự mã hóa được biểu hiện theo cách đặc trưng giới tính, nghĩa là trong một số trường hợp một protein chỉ được biểu hiện ở con cái hoặc được biểu hiện ở mức độ lớn hơn nhiều so với con đực hoặc theo cách khác, trong các trường hợp khác, protein chỉ được biểu hiện ở con đực hoặc được biểu hiện ở con đực với mức độ lớn hơn nhiều so với con cái. Trong khi, tốt hơn nếu protein chỉ được biểu hiện ở con đực, nhưng tốt nhất là protein chỉ được biểu hiện ở con cái. Cơ chế để đạt

được cắt nối thay thế đặc hiệu giới tính qua trùng gian protein TRA hoặc phức hợp TRA/TRA-2 được biết đến và được mô tả bởi Pane và cộng sự chẳng hạn (tham khảo tài liệu Development 129, 3715-3725 (2002)).

Tốt hơn là, trình tự kiểm soát cắt nối bao gồm, và tốt hơn nữa là chứa, gen chuyển intron bắt nguồn từ gen chuyển của *Ceratitis capitata* (Cctrd), mà gen chuyển *Ceratitis capitata* này có một vùng cắt nối thay thế. Ở sản phẩm phiên mã F1, như minh họa trên Hình 33 (Hình 2 A của Pane và cộng sự (2002), nêu trên), đó là intron thứ nhất. Các gen tương đương gen chuyển ở các loài khác, ví dụ *Bactrocera oleae, Ceratitis rosa, Bactrocera zonata* và *Drosophila melanogaster* cũng có các vùng cắt nối thay thế ở vị trí tương tự trong đó trình tự mã hóa, gen chuyển intron bắt nguồn từ các côn trùng đó cũng được ưu tiên đặc biệt.

Kiểu cắt nối trong *Cctra* nói riêng được bảo toàn tốt, với các sản phẩm phiên mã của chúng được tìm thấy ở con đực bao gồm các nguyên liệu exon bổ sung so với sản phẩm phiên mã F1, đến mức các sản phẩm phiên mã này không mã hóa protein chức năng Tra với chiều dài đầy đủ. Ngược lại, sản phẩm phiên mã F1 mã hóa protein chức năng Tra với chiều dài đầy đủ, sản phẩm phiên mã này về căn bản đặc hiệu giống cái ở hầu hết các giai đoạn trong chu trình vòng đời, mặc dù đã nghiên cứu thấy rằng phôi sớm của cả hai giống có thể bao gồm một lượng nhỏ sản phẩm phiên mã này. Sáng chế mô tả trình tự được cắt khỏi sản phẩm phiên mã F1, nhưng không được cắt khỏi sản phẩm phiên mã đặc hiệu giống đực hoặc không đặc hiệu giới tính, như các intron gen chuyển, hoặc thậm chí là intron gen chuyển F1. Vì vậy phiên bản của trình tự đã được tìm thấy trong gen Cctra là intron CcTRA.

Vì vậy gen chuyển được điều hòa một phần bằng sự cắt nối thay thế đặc hiệu giới tính, trong khi sản phẩm chủ chốt của nó, protein Tra, bản thân nó tham gia vào việc cắt nối thay thế. Ở côn trùng, sự cắt nối thay thế đặc hiệu giới tính được điều hòa bởi protein TRA, hoặc một phức hợp bao gồm các protein TRA và TRA2, bao gồm các trình tự kiểm soát cắt nối ở loài hai cánh từ gen

lưỡng tính (dsx) và cũng như từ chính gen chuyển intron, mặc dù không bao gồm gen chuyển intron từ ruồi giấm (Dmtra), mà được điều hòa chủ yếu bởi sản phẩm gen *Sxl* ở ruồi giấm, hơn là TRA hoặc phức hợp TRA/TRA2.

Ngoài ruồi giấm, sản phẩm gen *Sxl* được biểu hiện biệt hóa ở các giới khác nhau. Sxl không được cho là đóng vai trò trong quá trình điều hòa cắt nối thay thế đặc hiệu giới tính ở các côn trùng không phải là ruồi giấm.

Ví dụ về protein TRA mà liên kết các vị trí protein liên kết (các trình tự nucleotit được nhận biết cụ thể bởi protein TRA) trong gen chuyển intron tốt hơn là từ bộ hai cánh, tốt hơn là từ họ Tephritidae, tốt hơn nữa nếu là từ giống *Ceratitis, Anastrepha* hoặc *Bactrocera*. Tuy nhiên, cũng được dự tính trước rằng các côn trùng hai cánh khác, như là các loại ruồi giấm hoặc các loại muỗi của các dạng khác nhau được đề cập dưới đây, cũng có khả năng tạo ra protein TRA hoặc các thể tương đồng của chúng mà là có khả năng liên kết với các vị trí thích hợp trên các trình tự kiểm soát cắt nối thu được từ gen dsx, gen chuyển hoặc gen chuyển intron, tức là gen chuyển intron đã được cắt nối thay thế được loại bỏ hàn toàn trong sản phẩm phiên mã F1, thậm chí trong các trường hợp, như là ruồi giấm, ở đó bản thân các gen chuyển tự nhiên (*Dmtra*) không được điều hòa bởi protein TRA. Theo một số phương án, "gen chuyển intron" có thể được định nghĩa như một trình tự kiểm soát cắt nối ở đó sự cắt nối thay thế của sản phẩm phiên mã ARN được điều hòa bởi TRA, chẳng hạn bởi sự liên kết của chúng, riêng lẻ hoặc trong tổ hợp (tức là khi phối hợp) với TRA2. Gen chuyển intron từ ruồi giấm được loại trừ.

Đặc biệt được ưu tiên là các trình tự kiểm soát cắt nối thu được từ gen chuyển intron. Theo đó gen chuyển có thể thu được, như đã được đề cập bất kỳ, từ *Ceratitis*, *Anastrepha* hoặc *Bactrocera*. Gen chuyển intron *Ceratitis capitata* từ gen chuyển được Pane và cộng sự (2002) mô tả ban đầu, nêu trên. Tuy nhiên, rõ ràng rằng các thể tương đồng tồn tại trong các loài khác, và có thể được nhận biết dễ dàng trong các loài đã biết và cũng như trong giống khác nhau của chúng. Vì vậy, khi tham chiếu đến gen chuyển, cần hiểu rằng cũng tham chiếu

đến gen chuyển tương đồng trong các loài khác, đặc biệt là trong các loài Ceratitis, Anastrapha hoặc Bactrocera.

Cần hiểu rằng thuật ngữ "thu được" khi được sử dụng cho gen chuyển dùng để chỉ các trình tự mà gần đúng với hoặc là bản sao chính xác của gen chuyển intron, như được mô tả trong nghiên cứu ở trên, trong trường hợp của Pane và cộng sự (2002). Tuy nhiên, cần hiểu rằng, khi đó các trình tự này là các trình tự intron, mà một vài nucleotit có thể được bổ sung hoặc xóa bỏ hoặc thay thế mà về cơ bản không làm mất chức năng.

Các ví dụ được ưu tiên của gen này bao gồm gen intron dsx, tốt hơn là được tạo ra dưới dạng một minigen. Trong trường hợp này, tốt hơn là có thể xóa bỏ, như sáng chế đã thực hiện trong các ví dụ, một số lượng khá lớn từ các intron cắt nối thay thế, ví dụ, 90% hoặc nhiều hơn intron trong một vài trường hợp, trong khi vẫn duy trì chức năng cắt nối thay thế. Vì vậy, trong khi dự tính xóa bỏ một lượng lớn, việc chèn, thay thế hoặc xóa bỏ nucleotit với lượng nhỏ hơn, thậm chí là một nucleotit cũng được ưu tiên.

Chiều dài chính xác của trình tự kiểm soát cắt nối thu được từ gen chuyển intron là yếu tố không cần thiết, miễn là nó có khả năng điều hòa sự cắt nối thay thế. Liên quan đến vấn đề này, người ta cho rằng khoảng 55 đến 60 nucleotit là chiều dài tối thiểu cho một gen chuyển intron đã cải biến, mặc dù chiều dài của kiểu gen chuyển intron kiểu dại (biến thể cắt nối F1) từ *C. capitata* là khoảng 1345 nucleotit.

Đặc biệt được ưu tiên là trình tự với chiều dài đầy đủ 1345 nucleotit của Cctra.

Đối với tất cả các trình tự nucleotit được đề cập ở đây, tốt hơn là có mức tương đồng về mặt trình tự nhất định, trừ khi có quy định khác. Vì vậy, tốt hơn là trình tự kiểm soát cắt nối có mức tương đồng về trình tự ít nhất 80% so với trình tự nêu trong SEQ ID NO, tốt hơn là mức độ tương đồng về trình tự ít nhất 80% so với trình tự nêu trong SEQ ID NO, tốt hơn nữa nếu mức độ tương đồng về trình tự ít nhất là 90% so với trình tự nêu trong SEQ ID NO, tốt hơn nữa nếu

mức độ tương đồng về trình tự ít nhất là 95% so với trình tự nêu trong SEQ ID NO, thậm chí tốt hơn nữa nếu mức độ tương đồng về trình tự ít nhất là 99% so với trình tự nêu trong SEQ ID NO, và tốt nhất nếu mức độ tương đồng về trình tự ít nhất là 99,9% so với trình tự nêu trong SEQ ID NO. Thuật toán phù hợp như BLAST có thể được sử dụng để xác định mức độ tương đồng về mặt trình tự. Nếu số lượng lớn trình tự được xóa bỏ trong kiểu dại, thì việc so sánh trình tự có thể trên toàn bộ chiều dài đầy đủ của kiểu dại hoặc trên các trình tự đã được sắp thẳng hàng với mức độ tương đồng tương tự.

Tuy nhiên, cần hiểu rằng mặc dù có tính tương đồng về mặt trình tự như trên, các yếu tố đã biết, đặc biệt là các nucleotit kẹp sườn và vị trí phân nhánh cắt nối phải được giữ lại, để thực hiện chức năng của hệ thống một cách có hiệu quả. Nói cách khác, trong khi các phần có thể được xóa bỏ hoặc thay đổi theo cách khác, chức năng hoặc hoạt tính cắt nối thay thế, có mức độ tương đồng ít nhất là 30%, tốt hơn là 50%, tốt hơn là 70%, tốt hơn nữa nếu là 90%, và tốt nhất là 95% so với kiểu dại cần phải được giữ lại. Tỷ lệ này cũng có thể được gia tăng tăng thêm đối với kiểu dại bằng cách tạo ra các vị trí mà liên kết các yếu tố cắt nối thay thế hoặc tương tác với phức hệ.

Đặc biệt, tốt hơn là tại vị trí trình tự kiểm soát cắt nối bao gồm một intron TRA được biến đổi, đoạn gen này bao gồm ít nhất 20 đến 40 cặp bazơ từ đầu 5' và, tốt hơn là, tính theo đầu 3' của intron đã biết. Hơn nữa, được ưu tiên hơn là ít nhất 3 hoặc 4 và tốt nhất nếu, ít nhất là 5, tốt hơn là 6, tốt hơn nữa nếu là 7 và tốt nhất là cả 8 trong số 8 miền liên kết TRA giả định từ gen chuyển intron của *C. capitata*, như được Pane và cộng sự, (2002) mô tả, hoặc thể tương đồng của chúng, được đề cập. Tất nhiên là, nếu như các vị trí khác được phát hiện đồng thời, thì dự tính rằng trình tự kiểm soát cắt nối có thể bao gồm nhiều hơn 8 vị trí. Thực tế là, dự tính rằng nhiều hơn 8 vị trí có thể được tạo ra trong trình tự kiểm soát cắt nối và sự cắt nối thay thế có thể được điều hòa theo cách này, đặc biệt là khi một vài vị trí được liên kết với các ái lực khác nhau dẫn đến kết quả cắt nối thay thế khác nhau.

Một trình tự liên ứng cho các miền liên kết TRA giả định của gen chuyển intron từ *C. capitata* được đề cập dưới đây là SEQ ID NO 1, một trình tự ADN, mặc dù ARN tương ứng cũng được ưu tiên.

Các trình tự liên ứng được ưu tiên là 1: TCWWCRATCAACA (SEQ ID NO. 1), trong đó W = A hoặc T và R = A hoặc G.

Tương tự áp dụng cho cá thể lưỡng tính, ở đó trình tự liên ứng cho protein TRA cũng là SEQ ID NO. 1 như một phức hợp protein bao gồm các protein Tra và TRA2 là các yếu tố điều hòa chủ chốt quá trình cắt nối thay thế của cá thể lưỡng tính, như là cho gen chuyển tương đồng (mặc dù không phải thể tương đồng gen chuyển tìm thấy trong ruồi giấm).

Như mô tả ở trên, trình tự kiểm soát cắt nối tốt hơn là thu được từ gen chuyển intron, tốt hơn nữa là từ họ *Tephritidae*. Đặc biệt được ưu tiên hơn là gen chuyển intron thu được từ *B. zonata* hoặc, tốt hơn là, từ các ruồi giấm không phải là Drosophilid khác. Tuy nhiên, đặc biệt được ưu tiên hơn là gen chuyển intron thu được từ giống *Ceratitis*, đặc biệt là *C. rosa* và, tốt nhất nếu là, *C. capitata*. Các loài này đã được biết đến, lần lượt là các loài ruồi giấm Natal và Mediterranean.

Liên quan đến gen chuyển intron thu được từ *B. zonata*, tác giả sáng chế đã chỉ ra rằng gen này có thể dẫn đến sự cắt nối thay thế đặc hiệu giới tính trong gen chuyển Mexfly (*Anastrapha ludens*) và trong gen chuyển Medfly (*C.capitata*). Tác giả sáng chế cũng chỉ ra rằng các protein khác nhau có thể được biểu hiện theo cách đặc hiệu giới tính thông qua sự cắt nối khác, bao gồm tTAV 3 và Rpr.

Liên quan đến gen chuyển intron thu được từ *C. rosa*, tác giả sáng chế đã đề xuất thành công sự cắt nối thay thế theo cách đặc hiệu giới tính của một gen chuyển trong Medfly.

Liên quan đến gen chuyển intron thu được từ *C. capitata* (Medfly), sáng chế đã chỉ ra rằng gen này có thể điều hòa sự cắt nối đặc hiệu giới tính trong gen chuyển Medfly, và các Tephritid khác, và các Tephritid khác như là *A. ludens*

(Mexfly). Không chỉ có vậy, sáng chế cũng chỉ ra rằng intron này có thể hoạt động một cách hiệu quả trong nhiều loài côn trùng, đặc biệt là bộ hai cánh Dipterans. Thực vậy, sáng chế đã chỉ ra rằng intron TRA từ *C. capitata* (được đề cập đến như là *Cctra*) có thể tạo ra việc cắt nối thay thế đặc hiệu giới tính trong gen chuyển từ ruồi giấm, mà không phải là Tephritid, và cũng như trong loài muỗi *Aedes aegypti*. Mặc dù các loài muỗi là bộ hai cánh Diptera, chúng chúng được phân kỳ từ ruồi giấm và Tephritid cách đây khoảng 250 triệu năm và, vì vậy, chúng có mối quan hệ gần hơn với Tephritid thay vì Drosophilid, mặc dù là khoảng thời gian phân kỳ đã được ước tính là khoảng từ 120 đến 150 triệu năm. Vì vậy, điều này cho thấy khả năng áp dụng rộng của sáng chế trên nhiều loài côn trùng.

Liên quan đến các trình tự kiểm soát cắt nối thu được từ intron dsx, sáng chế đã chỉ ra rằng trình tự này có thể được sử dụng cho việc cắt nối thay thế, theo kiểu đặc hiệu giới tính, ở nhiều loại côn trùng. Vì vậy, tốt hơn là gen dsx, thu được từ *Bombyx mori* (tằm), *Pectinophora gossypiella* (sâu hồng hại bông) *Pectinophora gossypiella*, *Cydia. Pomonella* (bướm đêm), ruồi giấm, và các loài muỗi như là *Anopheles* sp., ví dụ, *A. gambiae*. Các loài muỗi được ưu tiên đặc biệt bao gồm *Stegomyia* spp., đặc biệt là *S. aegypti* (cũng được biết đến như là *Aedes aegypti*).

Thực vậy, đối với *A. aegypti*, sáng chế đã chỉ ra một số lượng đáng kể các cấu trúc ADN, mà có khả năng tạo ra sự cắt nối thay thế đặc hiệu giới tính.

Cần hiểu rõ rằng hệ thống hoặc cấu trúc tốt hơn là được dùng như một plasmit, nhưng thường được thử nghiệm sau khi hợp nhất vào hệ gen. Việc sử dụng có thể được tiến hành theo các phương pháp đã biết trong lĩnh vực, như là ngoài đường tiêu hóa, trong tĩnh mạch, trong cơ, đường miệng, qua biểu bì, phân phối qua niêm mạc v.v. Việc tiêm vào phôi được ưu tiên đặc biệt. Plasmit có thể được mở vòng trước hoặc trong quá trình sử dụng, và không phải tất cả plasmit đều có thể được hợp nhất vào trong hệ gen. Ở đó chỉ một phần của plasmit được

hợp nhất vào trong hệ gen, được ưu tiên hơn nếu phần này bao gồm ít nhất một trình tự kiểm soát cắt nối có khả năng điều hòa sự cắt nối thay thế.

Tốt hơn, hệ thống biểu hiện polynucleotit là một hệ thống gen trội tái tổ hợp gây chết, hiệu quả gây chết là có điều kiện. Các điều kiện phù hợp bao gồm nhiệt độ, sao cho hệ thống được biểu hiện ở một nhiệt độ nhưng không được biểu hiện, hoặc được biểu hiện ở một mức độ thấp hơn, ở một nhiệt độ khác. Hệ thống gen gây chết có thể tác động đến các tế bào đặc hiệu hoặc các mô hoặc gây ảnh hưởng lên toàn bộ côn trùng. Các hệ thống mà không gây chết hoàn toàn, nhưng gây ra những ảnh hưởng đáng kể cũng được đề cập, ví dụ có thể dẫn tới mù, sức đề kháng kém (đối với côn trùng mà có thể bay bình thường), hoặc vô sinh. Các hệ thống mà gây cản trở cho việc xác định giới tính cũng được nhìn nhận, ví dụ làm biến đổi hoặc có xu hướng làm biến đổi toàn bộ hoặc một phần của côn trùng từ một kiểu giới tính này thành kiểu giới tính khác. cần hiểu rằng tất cả các hệ thống và trình tự như vậy được đề cập bằng thuật ngữ “gây chết” được sử dụng ở đây. Tương tự, thuật ngữ "làm chết", và các thuật ngữ tương tự dùng để chỉ sự biểu hiện hiệu quả của hệ thống gây chết và do đó có kiểu hình có hại hoặc đột biến giới tính, ví dụ bị chết.

Tốt hơn nữa nếu, hệ thống biểu hiện polynucleotit là một hệ thống gen trội tái tổ hợp gây chết, tác động gây chết là có điều kiện và không được biểu hiện trong điều kiện cho phép đòi hỏi có hợp chất mà không tồn tại trong môi trường tự nhiên của sinh vật, tới mức tác động gây chết của hệ thống gây chết xảy ra trong một trường tự nhiên của côn trùng.

Hay nói cách khác, các trình tự mã hóa mã hóa một yếu tố gây chết cho một hệ thống như là hệ thống tet được mô tả trong WO 01/39599 và/hoặc WO2005/012534.

Thực vậy, tốt hơn là sự biểu hiện của gen gây chết đã biết dưới sự kiểm soát của một protein hoạt hóa tương ứng. Cũng được ưu tiên là gen mà sự biểu hiện của nó được điều hòa bởi một protein hoạt hóa mã hóa sự cắt nối khác như là tTA. Điều này không tương thích với protein được điều hòa là một yếu tố gây

chết. Thực vậy, đặc biệt được ưu tiên cả hai. Liên quan đến vấn đề này, sáng chế đặc biệt ưu tiên hệ thống bao gồm một hệ thống phản hồi chủ động như mô tả trong WO2005/012534.

Tốt hơn là, hiệu quả gây chết của hệ thống trội gây chết là có tác dụng ngăn chặn có điều kiện.

Các sinh vật phù hợp mà có thể sử dụng hệ thống hiện thời bao gồm các động vật có vú như là chuột, các động vật gặm nhấm và gia súc. Cũng được ưu tiên là các loại cá, như là cá hồi almon và cá hồi trout. Các loại thực vật cũng được ưu tiên, nhưng đặc biệt được ưu tiên sinh vật chủ là côn trùng, tốt hơn là ruồi giấm hoặc tephritid. Tốt hơn là, sinh vật không phải là người, tốt hơn là động vật không có vú, tốt hơn là không phải là gia cầm, tốt hơn là một động vật không xương sống, tốt hơn là một động vật chân đốt.

Cụ thể, được ưu tiên là côn trùng từ bộ hai cánh, đặc biệt là bộ hai cánh bậc cao và đặc biệt là một ruồi giấm tephritid, tốt hơn là Medfly (*Ceratitis capitata*), tốt hơn là Mexfly (*Anastrepha ludens*), tốt hơn là ruồi giấm Oriental (*Bactrocera dorsalis*), ruồi giấm Olive (*Bactrocera oleae*), ruồi Melon (*Bactrocera cucurbitae*), ruồi giấm Natal (*Ceratitis rosa*), ruồi giấm Cherry (*Rhagoletis cerasi*), ruồi giấm Queensland (*Bactrocera tyroni*), ruồi giấm Peach (*Bactrocera zonata*), ruồi giấm Caribbean (*Anastrepha suspensa*) hoặc ruồi giấm West Indian (*Anastrepha obliqua*). Cũng được ưu tiên đặc biệt khi sinh vật chủ là một loài muỗi, tốt hơn là từ giống *Stegomyia*, *Aedes*, *Anopheles* hoặc *Culex*. Được ưu tiên đặc biệt là *Stegomyia aegyptae*, cũng được biết đến như là *Aedes aegypti, Stegomyia albopicta* (cũng được biết đến như là *Aedes albopictus*), *Anopheles stephensi*, *Anopheles albimanus* và *Anopheles gambiae*.

Cùng với bộ hai cánh, nhóm được ưu tiên khác là Calliphoridae, đặc biệt là screwworm thế hệ mới (*Cochliomyia hominivorax*), screwworm thế hệ cũ (*Chrysomya bezziana*) và ruồi xanh Australia (*Lucilia cuprina*). Các loài Lepidoptera và bộ cánh cứng cũng được ưu tiên, đặc biệt là các loài bướm, bao gồm sâu bướm (*Cydia pomonella*), và giun tơ (*Bombyx mori*), sâu hồng hại bông

(*Pectinophora gossypiella*), sâu tơ (*Plutella xylostella*), bướm Gypsy (*Lymantria dispar*), bọ cánh cam (*Amyelois transitella*), sâu ở cây đào (*Anarsia lineatelld*) và sâu đục thân lúa (*Tryporyza incertulas*), cũng được ưu tiên là các loài bướm cú, đặc biệt là Heliothinae. Trong số bộ cánh cứng, bọ cánh cứng Nhật Bản (*Popilla japonica*), bọ cánh cứng White-fringed (*Graphognatus spp.*), mọt ngũ cốc Boll (*Anthonomous grandis*), giun hại rễ ngô (*Diabrotica spp*) và bọ cánh cứng hại khoai tây Colorado (*Leptinotarsa decemlineata*) được ưu tiên đặc biệt.

Tốt hơn là, côn trùng không phải là Drosphilid, cụ thể là Dm. Vì vậy, trong một vài khía cạnh, sự biểu hiện trong các loài Drosophilids, đặc biệt là Dm được loại trừ. Trong các khía cạnh khác, trình tự kiểm soát cắt nối là không thu được từ gen chuyển intron của một Drosphilid, đặc biệt là Dm.

Được ưu tiên là sự biểu hiện của trình tự polynucleotit khác loại tạo ra một kiểu hình ở côn trùng. Đặc biệt tốt hơn là protein chức năng không phải là beta-galactozidaza, nhưng có thể được liên hợp với các chỉ thị có thể nhìn thấy được (bao gồm sự phát huỳnh quang), có khả năng tồn tại và phát triển, có khả năng sinh sản, thích hợp, có khả năng hoạt động, có khả năng quan sát được, và các thuộc tính khác. Tất nhiên, nó sẽ được đánh giá đúng, rằng, trong một vài khía cạnh, các hệ thống biểu hiện là điều kiện điển hình, với kiểu hình chỉ được biểu hiện trong một số điều kiện, chẳng hạn các điều kiện hạn chế.

Theo một khía cạnh khác, sáng chế cũng đề cập đến một phương pháp kiểm soát quần thể côn trùng trong môi trường tự nhiên của chúng, bao gồm:

i) nhân giống một lượng lớn côn trùng,

côn trùng này mang một hệ thống biểu hiện gen bao gồm một hệ thống theo sáng chế mà là một hệ thống gen trội gây chết,

ii) phân bổ lượng lớn động vật này vào môi trường ở vị trí để kiểm soát quần thể, và

iii) đạt được sự kiểm soát quần thể thông qua việc gây chết ở giai đoạn sớm bởi sự biểu hiện của hệ thống gây chết ở cá thể non mà tạo ra từ sự lai khác loài

của nguồn các cá thể đã nhân giống với các cá thể có giới tính đối lập của quần thể kiểu dại.

Tốt hơn là, gây chết ở giai đoạn sớm là trong thời kỳ phôi phát triển hoặc trước khi thuần thục giới tính, tốt hơn là trong sự phát triển sớm, tốt nhất là trong các giai đoạn ấu trùng hoặc giai đoạn phôi sớm của vòng đời.

Tốt hơn là, hiệu quả gây chết của hệ thống là có điều kiện và xuất hiện trong môi trường tự nhiên đã biết thông qua sự biểu hiện của một gen gây chết, sự biểu hiện của gen gây chết đã biết là dưới dạng kiểm soát của một protein hoạt chuyển có thể ức chế được, việc nhân giống đã biết là trong điều kiện tùy ý với sự có mặt của một cơ chất, mà cơ chất này là không có mặt trong môi trường tự nhiên đã biết và có thể ức chế sự hoạt chuyển đã nêu.

Tốt hơn là, hiệu quả gây chết được biểu hiện trong các phôi của con non này. Bản mô tả bộc lộ sinh vật là động vật đa bào không xương sống hoặc như được đề cập đến ở bất cứ đâu.

Sáng chế cũng đề cập đến phương pháp kiểm soát sinh học, trong đó phương pháp này bao gồm các bước:

i) nhân một lượng lớn sinh vật giống đực và giống cái được biến nạp với hệ thống biểu hiện theo sáng chế trong điều kiện tùy ý, cho phép các giống đực và giống cái sống sót, để mang lại một tác nhân kiểm soát giới tính sinh học kép;

ii) tùy ý trước bước tiếp theo, áp đặt hoặc cho phép tiếp xúc với các điều kiện giới hạn là nguyên nhân gây chết của các cá thể thuộc một giới và do đó tạo ra một tác nhân kiểm soát giới tính sinh học đơn lẻ bao gồm các cá thể giới tính còn lại mang hệ thống gen gây chết có điều kiện;

iii) giải phóng tác nhân kiểm soát giới tính kép hoặc tác nhân kiểm soát giới tính sinh học đơn lẻ vào môi trường ở nơi cần kiểm soát sinh học; và

iv) đạt được sự kiểm soát sinh học thông qua sự biểu hiện của hệ thống gen trong con non được tạo ra từ việc lai khác thứ cùng loài của các cá thể mang tác

nhân kiểm soát sinh học này với các cá thể giới tính đối lập của quần thể kiểu dại.

Tốt hơn là, có sự phân chia giới tính trước khi có sự phân phối sinh vật bởi sự biệu hiện của một hệ thống gen đặc hiệu giới tính gây chết.

Tốt hơn là, hiệu quả gây chết gây ra tỷ lệ chết lớn hơn 90% của nhóm đích gồm thế hệ con cháu tạo ra do sự giao phối giữa các sinh vật được giải phóng vào môi trường và quần thể kiểu dại.

Sáng chế cũng đề cập đến phương pháp phân chia giới tính, trong đó phương pháp này bao gồm:

i) nhân giống một lượng lớn sinh vật giống đực và giống cái được biến nạp với hệ thống gen biểu hiện dưới dạng các điều kiện tùy ý hoặc giới hạn, cho phép các giống đực và giống cái này sống sót; và

ii) loại bỏ các điều kiện cho phép hoặc hạn chế để gây ra tác động gây chết của gen gây chết ở một giới tính nhưng không gây chết đối với giới tính còn lại bởi sự cắt nối thay thế đặc hiệu giới tính của gen gây chết.

Tốt hơn là, hiệu quả gây chết dẫn đến tỷ lệ chết lớn hơn 90% của lớp đích của thế hệ con cháu được tạo ra do sự giao phối giữa côn trùng được giải phóng và quần thể kiểu dại.

Sáng chế cũng đề cập đến phương pháp kiểm soát sinh học hoặc quần thể, trong đó phương pháp này bao gồm các bước:

i) nhân giống một lượng lớn sinh vật giống đực và giống cái được biến nạp với hệ thống gen biểu hiện trong các điều kiện cho phép hoặc giới hạn, cho phép giống đực và giống cái này sống sót;

ii) loại bỏ điều kiện giới hạn hoặc cho phép này để gây ra tác động gây chết của gen gây chết ở một giới tính nhưng không gây chết đối với giới tính còn lại bởi sự cắt nối thay thế đặc hiệu giới tính của gen gây chết để thu được sự phân chia giới tính;

iii) làm mất khả năng sinh sản hoặc làm mất một phần khả năng sinh sản của các cá thể đã được phân chia; và

iv) thu được sự kiểm soát đã nêu thông qua sự giải phóng các cá thể đã được phân chia giới tính bị mất khả năng sinh sản hoặc bị mất một phần khả năng sinh sản vào môi trường tự nhiên của sinh vật.

Tốt hơn là, việc làm mất khả năng sinh sản đạt được thông qua việc sử dụng bức xạ ion hóa. Tuy nhiên, các phương pháp phổ biến ngăn ngừa sự chiếu xạ, như được sử dụng trong kỹ thuật làm mất khả năng sinh sản của côn trùng (Sterile Insect Technique - SIT) là đặc biệt được ưu tiên và có nhiều ưu điểm về chi phí và sức khỏe so với các phương pháp được thực hiện cùng hoặc sau khi sử dụng bức xạ.

Sáng chế cũng đề cập đến phương pháp hạn chế chọn lọc giống cái từ một quần thể. Phương pháp tương ứng đối với giống đực cũng được dự tính.

Các phương pháp phân chia giới tính là hết sức quan trọng về phương diện thương mại, ví dụ các loài tằm, ở đó các giống đực sản sinh tơ nhiều và tốt hơn giống cái. Vì vậy, các phương pháp sử dụng để phân chia giới tính mà các giống cái bị loại trừ và, cụ thể các loài tằm cái được ưu tiên đặc biệt.

Sáng chế cũng dự tính trước rằng protein chức năng có thể được biểu hiện biệt hóa, nhưng có thể được phát hiện trong nhiều hơn một biến thể cắt nối, vì vậy, và tốt hơn là, ví dụ trong cả hai giới. Các ví dụ về các protein như vậy bao gồm protein phát huỳnh quang, như eGFP, CopGFP và DsRed2. Protein này có thể được sử dụng trong phương pháp phân chia hoặc phân loại giới tính không gây chết, nhờ đó một protein có thể chia cắt hai kiểu giới tính nhưng không gây chết loại nào trong số chúng.

Sáng chế cũng bất ngờ phát hiện ra rằng vị trí của trình tự kiểm soát cắt nối có thể được thay đổi và thu được các kết quả tốt hơn. Tốt hơn là, trình tự kiểm soát cắt nối là trình tự kiểm soát cắt nối "thứ nhất", khi được đọc từ trình tự khởi đầu, theo hướng từ đầu 5' đến đầu 3'. Sáng chế đã phát hiện ra rằng có các cấu trúc cụ thể với một intron ở đầu 5' UTR của hệ thống mà điều này dẫn đến làm

giảm mức độ hoặc sự biểu hiện của protein cắt nối thay thế được điều hòa bởi trình tự kiểm soát cắt nối theo sáng chế.

Tốt hơn là, trình tự kiểm soát cắt nối là từ đầu 3' đến codon khởi đầu. Tốt hơn là, trình tự kiểm soát cắt nối được cài xen vào bên trong exon thứ nhất, tức là sự tháo xoắn ngay lập tức của trình tự từ đầu 3' đến vị trí khởi đầu phiên mã. Cần hiểu rằng các thuật ngữ mà có thể chỉ trình tự ADN mã hóa sản phẩm phiên mã, hoặc bản thân sản phẩm phiên mã ARN.

Nếu tự kiểm soát cắt nối bắt đầu từ đầu 3' đến codon khởi đầu, tốt hơn là nó cũng là từ đầu 5' đến codon kết thúc trong khung thứ nhất (nghĩa là từ đầu 3' và nằm trong khung với codon khởi đầu), nhờ đó việc cắt nối thay thế tạo ra các sản phẩm pheien mã mà mã hóa protein hoặc các trình tự polypeptit khác nhau. Vì vậy, theo một phương án được ưu tiên, cấu trúc hoặc trình tự polynucleotit này bao gồm các thành phần sau theo hướng từ đầu 5' đến đầu 3', đối với sợi mã hóa hoặc sản phẩm phiên mã sơ cấp, điểm khởi đầu phiên mã, điểm khởi đầu dịch mã, intron có khả năng cắt nối thay thế, trình tự mã hóa toàn bộ hoặc một phần protein, codon kết thúc.

Tốt hơn là trình tự kiểm soát cắt nối có thể được định nghĩa như ở trên bao gồm đầu 5' là G (GT/C) và đầu 3' tương đương của nó là G, cụ thể trong *tra*, nhưng như được đề cập ở trên, trình tự này có thể chứa một vài trình tự exon và vì vậy, có thể bao gồm đầu phần lớn nucleotit đầu 3' (đầu cuối) của exon (tức là G).

Được ưu tiên đặc biệt là trình tự kiểm soát cắt nối nằm ngay cạnh theo chiều 3', codon khởi đầu, nhờ đó G của ATG là đầu 5' của condon khởi đầu (đầu 5') của trình tự kiểm soát cắt nối. Trình tự này có ưu điểm đặc biệt do nó cho phép G của codon khởi đầu ATG trở thành trình tự kẹp sườn bên 5'G đối với trình tự kiểm soát cắt nối.

Theo cách khác, trình tự kiểm soát cắt nối là từ đầu 3' đến codon khởi đầu, nhưng trong giới hạn exon 1000 bp, tốt hơn là exon 500 pb, tốt hơn là exon 300 pb, tốt hơn là exon 200 pb, tốt hơn là exon 150 pb, tốt hơn là exon 100 pb, tốt

hơn nữa nếu là exon 75 bp, tốt hơn nữa nếu là exon 50 bp, tốt hơn nữa nếu là exon 30 bp, tốt hơn nữa nếu là exon 20 bp, và tốt nhất nếu là exon 10 bp hoặc thậm chí là exon 5, 4, 3, 2 hoặc 1 bp.

Sáng chế nhằm cải tiến hệ thống được định nghĩa là LAl 1188 trong WO 2005/012534. Plasmit này có một số nhược điểm, chủ yếu là các nucleotit exon được cắt bỏ với intron Cctra được sử dụng theo sáng chế, do đó tạo ra sự dịch chuyển khung trong sản phẩm phiên mã.

Cụ thể là, ngoài trình tự thu được từ Cctra (intron Cctra), 4 nucleotit của trình tự tTAV đã được loại bỏ trong sản phẩm phiên mã đặc hiệu giống cái, vì vậy, mặc dù các sản phẩm phiên mã cắt nối thay thế đã được tạo ra, bao gồm một sản phẩm phiên mã đặc hiệu giống cái, không có sản phẩm phiên mã nào có khả năng mã hóa protein tTAV chức năng. Vì vậy, cấu trúc này không có khả năng tạo ra sự biểu hiện đặc hiệu giới tính của protein chức năng tTAV.

Do việc cắt không nhằm vào trình tự cho cắt nối (5'-GT...) thường được sử dụng trong intron Cctra, cấu trúc này rõ ràng không bao gồm tất cả các trình tự điều hòa cần thiết để cắt nối ở dạng intron Cctra ở "trạng thái nguyên bản của nó". Tuy nhiên, đó là một vấn đề khác. Chắc chắn là chỉ có những đoạn bị mất mà gắn bên trình tự TG...GT, mà ở đó chỉ có thể là vật liệu 5' G.

Hiệu quả chính của sáng chế là, cụ thể liên quan đến gen *tra*, các yêu cầu đối với trình tự kiểm soát cắt nối exon là đến mức tối thiểu (ví dụ 2 nucleotit ở mỗi đầu) nhờ đó chúng có thể được thiết kế dễ dàng vào trong hầu hết các trình tự mã hóa, bằng cách sử dụng phần dư trong mã di truyền. Vì vậy các nucleotit exon "bổ sung" có thể là một phần của trình tự protein khác loại, và trình tự kẹp sườn của intron ở trạng thái tự nhiên của nó ở cùng thời điểm.

Hơn nữa, intron *Cctra* trong LA1188 là +132bp từ đầu 3' tới G của codon khởi đầu ATG (tới nucleotit exon cuối cùng). Thực vậy, mặc dù intron Cctra trong LA1188 là intron đầu tiên dịch mã từ đầu 5' đến đầu 3'; chiều từ codon khởi đầu ATG, nhưng nó không phải là intron "đầu tiên" khi đọc theo chiều từ đầu 5' đến đầu 3' từ trình tự khởi đầu. Thực tế, nó là intron thứ hai, nếu có nhiều

hơn một intron (thu được từ gen của *Drosophila melanogaster* Adh) ngược chiều với codon khởi đầu ATG). Thông tin này được nêu trong Bảng 3.

Cần hiểu rằng khi tham chiếu đến các condon khởi đầu ATG hoặc cá trình tự G kẹp sườn, hoặc các trình tự 5'- TG...GT-3', có nghĩa là liên quan đến trình tự ADN, nhưng điều đó cũng bao hàm trình tự ADN đối mã tương ứng và tương tự, trình tự ARN tương ứng.

Mô tả vắn tắt các trình tự được đề cập trong sáng chế

SEQ ID NO. 1 là trình tự liên ứng của gen chuyển.

SEQ ID NO. 2 là trình tự kẹp sườn LA3097 đầu 5'.

SEQ ED NO. 3 là trình tự kẹp sườn LA3097 đầu 3'.

SEQ ID NO. 4 là đoạn mồi 688 - ie1-transcr.

SEQ ID NO. 5 là đoạn mồi 790 - Aedsx-m-r2.

SEQ ID NO. 6 là đoạn mồi 761 - Aedsx-fem-r.

SEQ ID NO. 7 là đoạn mồi AedsxR1.

SEQ ID NO. 8 là trình tự liên ứng của Pane và cộng sự.

SEQ ID NO. 9 là trình tự liên ứng của Scali và cộng sự, 2005.

Các SEQ ID NO. 10-33 và 107-138 là trình tự liên ứng của các vị trí liên kết Tra/Tra2 giả định suy luận cho ruồi giấm (xem Bảng 2).

SEQ ID NO. 34: Khung đọc mở của tTAV.

SEQ ID NO. 35: Trình tự protein của tTAV.

SEQ ID NO. 36: Khung đọc mở của tTAV2.

SEQ ID NO. 37: Trình tự protein của tTAV2.

SEQ ID NO. 38: Khung đọc mở của tTAV3.

SEQ ID NO. 39: Trình tự protein của tTAV3.

SEQ ID NO. 40: Đoạn 1 trình tự *dsx* đặc hiệu của giống cái ởsâu hồng hại bông.

SEQ ID NO. 41: Đoạn 2 trình tự *dsx* đặc hiệu của giống cái loài sâu hồng hại bông (PBW, *Pectinophora gossypiella*).

SEQ ID NO. 42: Trình tự dsx đặc hiệu của giống đực loài sâu hồng hại bông (PBW, *Pectinophora gossypiella*)

SEQ ID NO. 43: Một phần trình tự gen *dsx* của *Aedes aegypti*. Tất cả trình tự exon bao gồm, nhưng chỉ giới hạn ở, một phần trình tự intron - xem giải thích trên hình 47 và 48.

SEQ ID NO. 44: Trình tự gen *dsx* của giống cái loài sâu bướm (*Cydia pomonella*): bao gồm một trình tự xoắn các nucleotit chưa biết, tốt hơn là 100, tốt hơn là nhỏ hơn 50, tốt hơn nữa nếu là nhỏ hơn 20, tốt hơn nữa nếu là nhỏ hơn 10, và tốt nhất nếu là nhỏ hơn 5.

SEQ ID NO. 45: Trình tự *dsx* của giống đực loài sâu bướm (*Cydia pomonella*).

SEQ ID NO. 46: Trình tự cấu trúc/plasmit dsx của pLA3435-Bombyx mori.

SEQ ID NO. 47: Trình tự cấu trúc dsx của pLA3359-*Anopheles gambiae*.

SEQ ID NO. 48: Trình tự cấu trúc của pLA3433-Agdsx (*Anopheles gambiae*) với bao gồm exon 2.SEQ ID NO. 49: Trình tự cấu trúc của pLA1188-cctra intron.

SEQ ID NO. 50: Trình tự cấu trúc của pLA3077-a intron CcTRA-tTAV.

SEQ ID NO. 51 : Trình tự cấu trúc của pLA3097-a intron CcTRA-tTAV.

SEQ ID NO. 52: Trình tự cấu trúc của pLA3233-Cctra-intron-tTAV2.

SEQ ID NO 53: Trình tự cấu trúc của pLA3014-Cctra-intron-Ubiquitin-reaperKR.

SEQ ID NO. 54: Trình tự cấu trúc của pLA3166-Cctra intron-Ubiquitin-reaperKR.

SEQ ID NO. 55: Trình tự của pLA3376-Bztra intron-reaperKR và Bztra-intron-tTAV3.

SEQ ID NO. 56: Trình tự cấu trúc của pLA3242-Crtra intron-reaperKR.

SEQ ID NO. 57: Một phần trình tự của một sản phẩm phiên mã giới tính đực được sinh ra trong ruồi giấm đen sẫm và da dày từ các LA3077 biến dạng mà khác với trình tự được sinh ra trong các loài Medfiy LA3077. Trình tự này phù hợp với sản phẩm phiên mã M3 được mô tả trong hình 36.

SEQ ID NO. 58: Một phần trình tự gen chuyển tương đồng của *Bactrocera zonata*. Trình tự của intron dự báo được cắt khỏi sản phẩm phiên mã đặc hiệu giống cái của gen chuyển ở *B. zonata* (trình tự từ +3 đến +970bp). Các nucleotit kẹp sườn exon ở các vị trí 1-2 và 971-972, tức là ở các đầu 5' và 3' của trình tự intron. Thực tế, đáng lưu ý rằng trình tự intron được kẹp sườn ở đầu 5' của nó bởi nucleotit Guanin, được xem là quyết định khả năng tách intron.

SEQ ID NO. 59: Một phần trình tự của gen chuyển tương đồng ở loài *Ceratitis rosa*. Trình tự của intron được dự báo được cắt khỏi sản phẩm phiên mã đặc hiệu giống cái của gen chuyển ở loài *C. rosa* tra (trình tự từ +3 đến 1311bp).

Các nucleotit kẹp sườn exon có mặt ở các vị trí 1-2 và 1312-3. Mặt khác, đáng lưu ý là trình tự intron được kẹp sườn ở đầu 5' của nó bởi một nucleotit Guanin, mà được xem là quyết định khả năng tách intron.

Các SEQ ID NO. 60-70: Các đoạn mồi như được đề cập đến trên các hình 44-46 và 50-51.

SEQ ID NO. 71 : Đoạn 3 ở gen dsx đặc hiệu cho giống cái của loài sâu hồng hại bông (PBW, *Pectinophora gossypiella*).

SEQ ID NO. 72: Khung đọc mở của *Drosophila melanogaster* ubiquitin.

SEQ ID NO. 73: Trình tự Protein của *Drosophila melanogaster* Ubiquitin.

Các SEQ ID NO. 74-105 là các đoạn mồi mà được đề cập trong các ví dụ ở trên.

SEQ ID NO. 106 là trình tự nucleotit LA1172, bao gồm bộ khung plasmit.

Các SEQ ID NO. 107-138 đã được mô tả ở trên.

SEQ ID NO. 139: Đoạn mồi HSP.

SEQ ID NO. 151: Trình tự plasmit nguyên vẹn LA3619.

SEQ ID NO. 140: Trình tự đoạn mồi VP16.

SEQ ID NO. 152: Trình tự plasmit nguyên vẹn LA3612.

SEQ ID NO. 141: Mồi Agexon1F.

SEQ ID NO. 142: Trình tự đoạn mồi TETRR1.

SEQ ID NO. 143: Trình tự plasmit LA3576.

SEQ ID NO. 153: Trình tự plasmit LA3491.

SEQ ID NO. 144 Trình tự plasmit LA3582.

SEQ ID NO. 154: Trình tự plasmit LA3515.

SEQ ID NO. 145: Trình tự plasmit LA3596.

SEQ ID NO. 155: Trình tự plasmit LA3545.

SEQ ID NO. 146: PBW-dsx (Hình 6).

SEQ ID NO. 156: Trình tự plasmit LA3604.

SEQ ID NO. 147: Bombyx-dsx (Hình 6).

SEQ ID NO. 157: Trình tự plasmit LA3646.

SEQ ID NO. 148: Codling-dsx (Hình 6).

SEQ ID NO. 158: Trình tự plasmit LA3054.

SEQ ID NO. 149: Minigen 1 DSX từ cấu trúc LA3491.

SEQ 3D NO. 159: Trình tự plasmit LA3056.

SEQ ID NO. 150: Minigen 2 DSX từ cấu trúc LA3534.

SEQ ID NO. 160: Trình tự plasmit LA3488.

SEQ ID NO. 161: Trình tự plasmit LA3641.

SEQ ID NO. 162: Trình tự plasmit LA3570.

Sáng chế sẽ được mô tả bởi các ví dụ không giới hạn sau đây.

Ví dụ thực hiện sáng chế

Biến nạp

Ví dụ 1- intron gen chuyển ở loài *Ceratitis capitata*

Tác giả sáng chế đã cài xencatxet intron Cctra vào khung đọc mở (open reading frame - ORF) tổng hợp. Hai phiên bản của sự cắt nối phù hợp này trong Medfly, nói cách khác, việc cắt nối của catxet intron Cctra nói chung là chính xác điều mà sẽ bình thường khi thực hiện trong phạm vi của gen Cctra nội sinh. Điều này là để tạo ra (chủ yếu hoặc chỉ) 3 biến thể cắt nối khác nhau ở giống cái, một trong số chúng là đặc hiệu giống cái (được gọi là F1) trong khi hai biến thể còn lại được tìm thấy ở cả giống đực và giống cái (được gọi là M1 và M2. Do mỗi sản phẩm phiên mã không đặc hiệu giới tính bao gồm vật liệu exon bổ sung với codon kết thúc, tác giả sáng chế cũng sắp xếp chúng sao cho chỉ có biến thể cắt nối của giống cái sẽ tạo ra protein chức năng.

Mỗi cấu trúc trong số các cấu trúc này (LA3077 và LA3097) có intron Cctra được kẹp sườn bởi TG và GT (để mang lại 5'...TG\intron\GT...3'. Một cấu trúc trước đây, mà hoạt động không hiệu quả, là LA1188. LA1188 đã được mô tả khá rõ – việc cắt nối chính xác như trên chỉ khác là 4 nucleotit bổ sung được loại bỏ. Intron này thuộc phạm vi 5'...TGGCAC|intron|GT...3'; việc cắt nối loại bỏ 4 bazơ bổ sung, tức là 5'...TG|GCACintron|GT...3' (Hình 33).

Trong tất cả các trường hợp intron là không thay đổi, và đơn giản là trình tự intron Cctra đầy đủ. Thông thường đối với các intron, nó bắt đầu bằng GT và kết thúc bằng AG. Hầu hết tất cả các intron đều bắt đầu với GT, cho nên sẽ bất ngờ khi sử dụng GC hiếm khác trong LA1188 [các intron GC-AG là các intron thay

thế đã biết theo một khảo sát ở quy mô lớn, 0,5% của tất cả các intron đã được báo cáo là có sử dụng GC-AG (Burset et al., 2001), mặc dù tỷ lệ này có thể nằm dưới mức ước tính, đặc biệt là cho các intron cắt nối thay thế, mà 5% trong số chúng có thể sử dụng GC-AG (Thanaraj và Clark, 2001).

Phân tích RT-PCR được tiến hành trên LA3077, (một cấu trúc phản hồi chủ động đối với intron CcTRA trong khung đọc mở tTAV). Ruồi trưởng thành đã được biến nạp thuộc cả hai giống đã được nuôi trên thức ăn về cơ bản chứa tetraxyclin ("nghèo tetraxyclin") trong 7 ngày. Sau đó các loài ruồi được thu nhận để tách chiết ARN và tiến hành RT-PCR sử dụng các đoạn mồi (HSP- SEQ ID NO. 104 và VP16 SEQ ID NO. 105) được sử dụng để phân tích mẫu cắt nối của intron CcTRA (Hình 34). Trong hai mẫu giống cái, tác giả sáng chế phát hiện ra rằng mẫu cắt nối chính xác của Cctra (776bp, tương ứng với việc loại bỏ chính xác của intron CcTRA) và nhận thấy không có dải như vậy ở các giống đực.

Sáng chế phát hiện ra rằng LA3077 và LA3097 tương ứng mang lại hiệu quả gây chết đặc hiệu giống đực có thể ức chế được. LA3077 đã được thử nghiệm kiểu hình thông qua việc lai chéo các loài ruồi dị hợp đối với LA3077 với ruồi kiểu dại, trên môi trường chứa và không chứa tetraxyclin. Hiệu quả gây chết đối với giống cái nằm trong khoảng từ 50 đến 70%. LA3097 (phiên bản cải biến của LA3077 ở đó intron CcTRA trực tiếp theo sau condon khởi đầu trong tTAV ORF), có mức độ gây chết đặc hiệu giống cái cao hơn nhiều, lên tới 100% (Hình 35). Intron CcTRA cũng đã được cài xen trong tTAV2 ở cùng vị trí như LA3097, trong cấu trúc LA3233, và điều này mang lại một kiểu hình tương tự như LA3097 (Hình 35).

Sáng chế cũng tạo ra các thể biến nạp của LA3077 trong ruồi giấm. Về mặt kiểu hình, cấu trúc này hoạt động hoàn hảo, điều đó cho thấy nó gây chết đặc hiệu cho giống cái một cách cực kỳ hiệu quả. Tuy nhiên, giải trình tự của các biến thể cắt nối của một trong số các đoạn cài xen đã chỉ ra rằng việc cắt nối của cấu trúc này trong ruồi giấm là không hề giống như việc cắt nối trong Medfly

(SEQ ID NO. 57). Sản phẩm phiên mã chủ chốt, sản phẩm đặc hiệu giống cái, là tương tự trong cả hai loại, nhưng ít nhất một phần sản phẩm phiên mã không đặc hiệu giới tính là khác nhau. Nó vẫn kết hợp trình tự exon bổ sung, với codon kết thúc, nhưng các mối nối khi cắt nối là không hề giống nhau (Hình 36). Nhận xét này là vô cùng quan trọng ở chỗ, nó chỉ ra rằng phương pháp này (sự điều hòa biểu hiện gen bằng cách sử dụng các intron cắt nối thay thế) có thể được sử dụng trong khoảng sinh loài khá rộng.

Một mẫu thử nghiệm để xác định liệu yếu tố điều hòa cắt nối exon cho đến nay đã được mô tả hay chưa (như là yếu tố tăng cường và ức chế) có thể làm biến đổi chức năng của intron cắt nối thay thế, có thể bao gồm bước thiết kế một cấu trúc và đưa nó vào một mô đích, sau đó nghiên cứu mẫu cắt nối của nó. Trong nhiều trường hợp, điều này sẽ không đòi hỏi sự biến nạp dòng phôi, vì thế thử nghiệm có thể khá nhanh, ví dụ bằng sự biểu hiện tạm thời trong các tế bào nuôi cấy mô phù hợp hoặc in vivo. Ví dụ, thử nghiệm intron vivo trong các côn trùng có thể đạt được bằng cách chuyển ADN bởi kỹ thuật vi tiêm. Tuy nhiên, người có hiểu biết trung bình sẽ hiểu rõ rằng, kỹ thuật vi tiêm kết hợp với đục lỗ bằng xung điện, hoặc đục lỗ bằng xung điện, biến nạp hóa học, các phương pháp sử dụng súng bắn gen, ví dụ, đều đã được sử dụng trong một số phạm vi khác nhau và các phương pháp đưa vào plasmit và biểu hiện protein từ đó đã được biết trong lĩnh vực này.

Gần đây sáng chế cũng đã thiết kế, và đã thu được gen chuyển với, intron CcTRA trong một gen khác (LA3014) (tất cả các ví dụ ở trên là tTAV). LA3014 bao gồm thể dung hợp ubiquitin-reaper^{KR} nằm xuôi chiều một intron CcTRA. Dữ liệu kiểu hình (Hình 35) đề cập rằng gen chuyển LA3014 Medfly đã mang lại sự gây chết đặc hiệu giống cái có thể ức chế được. Phép phân tích RT-PCR dùng để tách chiết ARN từ các giống đực và giống cái trưởng thành được nuôi cấy trên môi trường nghèo tetraxyclin, sử dụng các đoạn mồi (HSP, SEQ ID NO 74) và ReaperKR (SEQ ID NO. 75), chứng minh rằng sự cắt nối chính xác xuất hiện ở các giống cái (dải 508bp) và không có dải nào như vậy được phát hiện ở các giống đực (Hình 37). LA3166 là một cấu trúc khác với Cctra được đặt trong

vùng mã hóa ubiquitin được dung hợp với reaper^{KR}, nhưng được đặt ở một vị trí khác trong ubiquitin. LA3166 cũng có hiệu quả gây chết đặc hiệu trội cho giống cái mà có thể ức chế được trong Medfly (Hình 35).

Gần đây, tác giả sáng chế cũng đã thiết kế, và thu được gen chuyển với, các cấu trúc trên cơ sở Cctra "chỉ có intron" với intron trong một gen khác (tất cả các ví dụ ở trên là trong tTAV hoặc một trong số các biến thể của nó, tức là tTAV2 hoặc tTAV3). Các cấu trúc này hoạt động như được dự báo trước. Đây là một kết quả quan trọng, vì vậy chỉ ra rằng không có các trình tự exon thiết yếu trong Cctra mà tác giả sáng chế đã tiến hành sao chép tình cờ một cách dễ dàng (về chức năng, nếu không nhất thiết về trình tự) trong tTAV. Sáng chế cũng đề cập đến các cấu trúc ubi- rpr^{KR} thuộc loại này (LA3014 và LA3166), mà cũng phù hợp với phương pháp dung hợp ubiquitin được mô tả ở trên.

Để chứng minh khoảng phát sinh loài của intron CcTRA, tác giả sáng chế đã thiết kế gen chuyển LA3097 và LA3233 ở *Anastrepha ludens*. LA3097 và LA3233 được lựa chọn để tiêm vào *Anastrepha ludens* do chúng thể hiện hiệu quả gây chết đặc hiệu giống cái tốt nhất ở *Ceratitis capitata* (xem ví dụ 13). Dữ liệu kiểu hình được thiết kế cho 4 dòng LA3097 độc lập và 1 dòng LA3233 (xem firuge 38). Hiệu quả gây chết đặc hiệu giống cái ở *Anastrepha ludens* thường thấp hơn một chút so với *C. capitata* nhưng đạt được 100% ở một loài.

Anastrepha ludens được biến nạp với LA3097 và được nuôi trên tetraxyclin cho tới khi trứng đã nở được phân lập và duy trì trong môi trường nghèo tetraxyclin trong 7 ngày. Sau đó ARN được tách chiết và tiến hành phân tích RT-PCR sử dụng các đoạn mồi HSP (SEQ ID NO. 76) và TETRR1 (SEQ ID NO. 77). Kiểu cắt nối đặc hiệu giống cái chính xác (cá thể - F1) là ARN quan sát được được phân lập từ các giống cái (348bp) nhưng không từ các giống đực, cho thấy chức năng của intron CcTRA trong các loài khác nhau (Hình 39)

Dải rõ ràng nhất ở giống đực và dải đặc hiệu giống cái được tinh sạch và kết tủa để giải trình tự. Các sản phẩm phiên mã đặc hiệu giống cái đã được phát

hiện là được cắt nối chính xác ở các giống cái loài Mexfly như được mong đợi cho LA3097:

LA3097: AGCCACCATG|GT...intron...AG | GTCAGCCGCC

Hai trình tự kẹp sườn ở trên lần lượt là SEQ ID NOS. 2 và 3.

Ví dụ 2: Intron gen chuyển ở loài *Bactocera zonata*

Sáng chế đã phân lập intron gen chuyển từ loài *Bactocera zonata* (*B. zonata*) (SEQ ID NO. 58) bằng cách sử dụng các đoạn mồi ROSA1 (SEQ ID NO. 78), ROSA2 (SEQ ID NO. 79), và ROSA3 (SEQ ID NO. 80).

Trình tự của các đoạn mồi này đã được thiết kế căn cứ trên trình tự mã hóa được bảo toàn của gen chuyển tương đồng của các loài *Ceratitis capitata* và *Bactrocera oleae*. Bằng cách sử dụng ROSA2 và ROSA3 hoặc ROSA1 và ROSA3 làm các đoạn mồi, gen chuyển intron và vùng mã hóa kẹp sườn của nó đã được khuếch đại từ ADN bộ gen của loài *Bactrocera zonata*. Sau đó tác giả sáng chế đã sử dụng các sản phẩm PCR này làm khuôn và khuếch đại đoạn intron gen chuyển để thu được cấu trúc LA3376 (Hình 31 và SEQ ID NO. 55). Các đoạn mồi (BZNHE- SEQ ID NO. 81 và BZR-SEQ ID NO. 82) đã được sử dụng để thu được các cấu trúc; các mồi này chứa các trình tự bổ sung nhằm mục đích nhân dòng. Intron Bztra trong LA3376 được nhân dòng vào ORF của tTAV3 và ORF của reaper^{KR}. Các thể biến nạp Medfly đã được tạo ra và ARN được tách chiết từ các giống ruồi đực và cái.

Sau đó kỹ thuật RT-PCR được tiến hành trên cả hai đoạn cắt nối Reaper^{KR} (HB- SEQ ID NO. 83) và Reaper ^{KR}- SEQ ID NO. 84) và tTAV3 (SRY- SEQ ID NO. 85) và AV3F- SEQ ID NO. 86). Các đoạn mong đợi dài 200bp cho reaper^{KR} và 670bp cho tTAV3, tương ứng với việc cắt nối trong mẫu tương đương với sản phẩm phiên mã F1 của Cctra (Pane et al, 2002), đã được tạo ra ở các giống cái (Hình 40).

Ví dụ 3: Phân lập và cắt nối của intron gen chuyển ở loài *Ceratitis rosa* (*C. rosa*, ruồi giấm Natal)

Các đoạn mồi ROSA2 (SEQ ID NO. 87) và ROSA3 (SEQ ID NO. 88) được thiết kế dựa trên trình tự mã hóa bảo toàn của *Ceratitis capitata* và *Bactrocera oleae*. Bằng cách sử dụng ROSA2 và ROSA3 làm đoạn mồi, intron gen chuyển và vùng mã hóa kẹp sườn của nó đã được khuếch đại từ ADN bộ gen của loài *Ceratitis rosa* (SEQ ID NO. 59). Sau đó, tác giả sáng chế đã sử dụng các sản phẩm PCR này làm khuôn và khuếch đại đoạn intron gen chuyển để thu được các cấu trúc. Các đoạn mồi (CRNHE- SEQ ID NO 89 và CRR SEQ ID NO 90) đã được sử dụng trong quá trình xây dựng LA3242 (SEQ ID NO. 56 và hình 32). LA3242 bao gồm intron *C. rosa* ở đầu 5' của reaper^{KR}ORF. Phôi *Ceratitis capitata* được tiêm với ADN của LA3242, Phôi đã tiêm được nuôi cấy tới khi trưởng thành trên môi trường về cơ bản không chứa tetraxyclin. ARN được tách chiết từ các giống đực và cái trưởng thành; ARN này được sử dụng làm khuôn để thực hiện RT-PCR sử dụng các doạn mồi HB (SEQ ID NO. 91) và Reaper^{KR} (SEQ ID NO. 92). Băng cắt nối đặc hiệu cho giống cái được mong đợi, tương ứng với việc cắt nối mẫu tương đương với sản phẩm phiên mã F1của Cctra, được quan sát thấy ở các giống cái và không được quan sát thấy ở các giống đực (Hình 41).

Cá thể lưỡng tính

Ví dụ 4 Bombyx mori dsx trong PBW

Trình tự của loài *Bombyx mori* (bướm tằm) tương đồng của *Drosophila* Dsx (*Bmdsx*) đã được mô tả trước đó và sản phẩm cắt nối đặc hiệu giống đực và giống cái đã được nhận biết (Suzuki et al, 2001). Cả hai giống đực và giống cái sử dụng cùng một đoạn 3' polyA, và có hai exon đặc hiệu giống cái. Một nghiên cứu đã gợi ý rằng sự cắt nối đặc hiệu giới tính là không phụ thuộc vào tra/tra2, nói cách khác mặc dù mẫu có vẻ tương tự, cơ chế cắt nối cơ bản có thể khác (Suzuki et al., 2001), mặc dù dữ liệu của chúng, phần lớn thiếu các vị trí liên kết tra-tra2 có thể nhận biết được, tuy nhiên, điều đó là không thuyết phục. Ngoài ra, một cấu trúc minigen dsx của loài *B. mori* (bao gồm trình tự exon và trình tự

intron bị cắt ngắn) đã được biến nạp vào *B. mori* và các thể biến nạp dòng phôi cho thấy sự cắt nối đặc hiệu giới tính (Funaguma et al., 2005).

Sáng chế đã thiết kế một minigen Bmdsx dựa trên trình tự được sử dụng trong nghiên cứu của Funaguma và cộng sự, với một vài thay đổi đáng kể, và tiêm gen này vào trong loài bướm sâu hồng hại bông để xác định liệu có thể thu được sự cắt nối đặc hiệu giới tính ở loài phân lỳ hay không. Cấu trúc minigen mà tác giả sáng chế đã thiết kế không bao gồm exon 1, mà có mặt trong cả các giống đực và giống cái. Ngoài ra, sáng chế đã cắt bỏ đoạn intron ở giữa exon 3 và 4 (hai exon đặc hiệu cho giống cái), bao gồm một trình tự khác loại (chứa nhiều vị trí nhân dòng, multiple cloning sites -MCS), đã sử dụng trình tự khởi tăng cường Hr5-IE1 từ baculovirut *Ac*NPV và đã sử dụng một trình tự kết thúc phiên mã 3' thu được từ SV40 (xem hình 42 dưới dạng một biểu đồ). Các đoạn exon/intron kẹp sườn riêng biệt đã được sử dụng để khuếch đại và tái tổ hợp với nhau bằng kỹ thuật pCR và được nối vào trong một cấu trúc mang một đoạn trình tự khởi đầu tăng cường Hr5/IE1 và SV40 3'UTR (Hình 22 và SEQ ID NO. 22).

LA3435 đã được tiêm vào phôi sâu hồng hại bông (*Pectinophora gossypiella*). Các ấu trùng tuổi 5 đã được thu nhận sau 5-7 ngày và được phân tích riêng rẽ bằng kỹ thuật RT-PCR (sử dụng các đoạn mồi IE1 transcr- SEQ ID NO. 93 và SV40-RT-P2- SEQ ID NO. 94) để xác định xem liệu BMdsx có trải qua sự cắt nối đặc hiệu giống đực và giống cái hay không (Hình 43). Các phép phân tích của sáng chế phát hiện ra rằng băng đặc hiệu giống đực (dự đoán là 442bp) ở 4 mẫu (các dòng 1,2,3 và 4) và băng đặc hiệu giống cái (dự đoán là 612bp) ở 1 mẫu (dòng 5).

Sự cắt nối chính xác của *B. mori* dsx trong PBW chứng tỏ rằng tác giả sáng chế có thể đạt được (đã đạt được) sự biểu hiện đặc hiệu giới tính của một trình tự khác loại (ở đây là, MCS) ở bộ sâu bọ cánh vảy bằng cách sử dụng một hệ thống cắt nối thay thế. Hơn nữa, do hệ thống cắt nối này nhận được từ một loài khác, điều này gợi ý rằng các cấu trúc này có thể hoạt động trên một khoảng

phát sinh loài rộng lớn. Tuy nhiên, sự nhận biết các hệ thống cắt nối thay thế trong các loài mong muốn cũng đã được hình dung, và các phương pháp để nhận biết các hệ thống cắt nối thay thế đã được đề cập ở đây hoặc sẽ được chuyên gia trong lĩnh vực này biết rõ. Bằng việc đề cập đến một MCS trong ví dụ của sáng chế (xem hình 42), sự biểu hiện của một trình tự mong muốn, ví dụ một vùng mã hóa cho một protein mong muốn có thể thu được dễ dàng bằng cách cài xen trình tự đã biết. Nếu trình tự đã biết mã hóa một protein phù hợp, một kiểu hình đặc hiệu giới tính, ví dụ gây chết đặc hiệu giới tính có điều kiện, có thể được đưa, ví dụ vào loài sâu hồng hại bông.

Ví dụ 5: Phân lập dsx của bướm đêm (Codling moth)

Gen dsx từ loài Codling moth (*Cydia pomonella*) được phân lập bằng cách thực hiện 3' RACE sử dụng các đoạn mồi mà được dựa trên sự sắp thẳng hàng các trình tự liên kết từ *B. oleae*, *B. tyroni*, *C. capitata*, *D. melanogaster*, *B. mori*, và *A. gambiae*. ARN được phân lập từ một giống đực và giống cái của loài codling moth và kỹ thuật 3' RACE, để thu được cADN, đã được thực hiện bằng cách sử dụng mồi TT7T25 (SEQ ID NO. 95).

Kỹ thuật PCR được thực hiện bằng cách sử dụng các mồi ds1c (SEQ ID NO. 96) và TT7 (SEQ ID NO. 97). Sau đó hai chu kỳ của kỹ thuật PCR lồng đã được thực hiện trên sản phẩm của quá trình PCR lần thứ nhất bằng cách sử dụng các đoạn mồi codling2a (SEQ ID NO. 98) và TT7 (SEQ ID NO. 99) và sản phẩm của chu kỳ thứ hai của quá trình PCR sử dụng các đoạn mồi Codling2b (SEQ ID NO. 100) và TT7. Đã phân lập được các trình tự đặc hiệu giống đực và giống cái tương đương với các trình tự dsx tương đồng đã được phân lập trước đó (giống đực -SEQ ID NO. 43 và giống cái - SEQ ID NO. 42).

Ví dụ 6: Phân lập dsx của PBW

Gen dsx từ loài sâu hồng hại bông được phân lập bằng cách thực hiện 3' RACE sử dụng các đoạn mồi mà được căn cứ theo các trình tự liên kết từ *B. oleae*, *B. tyroni*, *C. capitata*, *D. melanogaster*, *B. mori* và *A. gambiae*.ARN được phân lập từ một giống đực và giống cái của loài bướm đêm và kỹ thuật 3'

RACE, để thu được cADN, đã được thực hiện bằng cách sử dụng mồi TT7T25 (trình tự được định nghĩa ở trên). Kỹ thuật PCR đã được thực hiện bằng cách sử dụng các mồi Pbwdsx2 (SEQ ID NO. 101) và TT7 (SEQ ID NO. 102). Sau đó kỹ thuật PCR lồng đã được thực hiện trên sản phẩm của quá trình PCR thứ nhất bằng cách sử dụng các mồi Pbwdsx3 (SEQ ID NO. 103) và TT7. Ba trình tự đặc hiệu cho giống cái đã được phân lập: PBWdsx-F1 (SEQ ID NO. 40), PBWdsx-F2 (Hình 10), và PBWdsx-F3 (SEQ ID NO. 71) và một trình tự đặc hiệu cho giống đực (SEQ ID NO. 42). Đã phân lập được các trình tự đặc hiệu cho giống đực và giống cái tương đương với các trình tự dsx tương đồng đã được phân lập trước đó.

Ví dụ 7: dsx trong *Anopheles gambiae*

Trình tự gen dsx của *Anopheles gambiae* đã được mô tả trước đó (Scali et al 2005). Tuy nhiên, khi tác giả cố gắng lặp lại trình tự được mô tả trong nghiên cứu đó, tác giả sáng chế phát hiện ra rằng có một vài điểm khác nhau trong sự cắt nối xuất hiện. Khi tác giả sáng chế đã cố gắng lặp lại sự khuếch đại của sản phẩm phiên mã đặc hiệu cho giống cái bằng cách sử dụng các đoạn mồi được thiết kế từ trình tự mARN (Số truy cập: AY903308 cho trình tự mã hóa cho giống cái và AY903307 cho trình tự mã hóa cho giống đực), sự khuếch đại đã bị lỗi. Tuy nhiên, khi Scali và cộng sự đã đề cập rằng đã có một exon tương đương, mà đã không được mô tả trước đó, tác giả sáng chế đã thiết kế các đoạn mồi để khuếch đại toàn bộ sản phẩm phiên mã dsx và gen. Bằng cách sử dụng các đoạn mồi này và các đoạn mồi đã được thiết kế từ trình tự bộ gen ADN (Số truy cập: GI: 19611767) tác giả sáng chế phát hiện ra rằng sự cắt nối của sản phẩm phiên mã ở giống cái là khác so với mô tả bởi Scali và cộng sự 2005 (Hình 44). Sản phẩm phiên mã đã chỉ ra rằng đoạn exon ở giống cái đã ở trong một vị trí khác. Có những giải thích khác nhau cho sựu khác nhau này, nhưng hầu hết giống nhau ở chỗ hoặc một vài phần khác nhau của trình tự trong *Anopheles* mà sáng chế đã sử dụng để từ đó thu được dữ liệu, hoặc trình tự được công bố không phải từ *Anopheles gambiae*, hoặc có nhiều hơn một đồng loài giống cái như mô tả cho *Stegomyia aegypti* trong ví dụ 20.

Sáng chế cũng đã sử dụng thành công các đoạn mồi, được thiết kế gần phiên bản cắt nối dsx ở loài *Anopheles gambiae* theo sáng chế, mà có thể phân biệt giữa giống đực và giống cái của *Anopheles gambiae* (Hình 45). Điều đó là bằng chứng tốt cho thấy hệ thống sẽ có chức năng như một cơ cấu cắt nối đặc hiệu giới tính khi dung hợp với một protein mong muốn, như là tTAV hoặc một tác nhân gây chết.

Gen dsx ở loài *Anopheles gambiae* mà sáng chế đã phân lập được từ ADN bộ gen, mà có một số thay đổi khác trong trình tự nucleotit so với trình tự bộ gen đã được báo cáo, đã được nhân dòng vào trong LA3359 (SEQ ID NO. 47) và LA3433 (SEQ ID NO. 48), các biểu đồ có thể lần lượt tìm thấy trong hình 23 và hình 24.

Ví dụ 8: dsx trong *Stegomyia aegypti*

Việc cắt nối của gen xảy ra tương tự như trong dsx của loài *Anopheles gambiae* (Scali et al 2005). Gen dsx ở loài *Stegomyia aegypti* được minh họa theo biểu đồ trên hình 47 hoặc 48. Một sản phẩm phiên mã đặc hiệu cho giống đực (M1) được tạo ra mà không bao gồm các exon 5a hoặc 5b. Hai sản phẩm phiên mã đặc hiệu cho giống cái khác nhau (F1 và F2) có cấu trúc như sau; F1 bao gồm các exon 1-4, 5a, 6 và 7 nhưng không bao gồm 5b, F2 bao gồm các exon 1-4 và 5b (hình 46). Ngoài ra, một sản phẩm phiên mã khác (C1) có mặt trong cả giống đực và giống cái, sản phẩm phiên mã này bao gồm các exon 1-4 và 7, nhưng không bao gồm các exon 5a, 5b hoặc 6.

Việc cắt nối của gen xảy ra tương tự như trong dsx của loài *Anopheles gambiae* (Scali et al 2005). gen dsx ở loài *Stegomyia aegypti* được minh họa theo biểu đồ trên hình 47 hoặc 48.

Actin 4

Ví dụ 9: Gen Actin - 4 ở loài *Stegomyia aegypti*

Theo một phương pháp để thu được sự biểu hiện đặc hiệu giới tính, mô và giai đoạn của một gen mong muốn là liên kết nó với gen Actin -4 ở loài

Stegomyia aegypti (AeAct-4). Gen này chỉ được biểu hiện trong các cơ phát triển vận động của giống cái loài *Stegomyia aegypti* (Munoz et al 2004). Đã sử dụng phép lai in-situ cho một ARN để phát hiện được dữ liệu biểu hiện của AeAct-4. Sáng chế đã thu được một đoạn của gen Actin - 4 của loài *Stegomyia aegypti*, bao gồm một vùng trình tự khởi đầu phiên mã được giả định, một đoạn intron cắt nối khác, và một phần của vùng không được phiên mã 5' (UTR) và đã đặt nó trước trình tự mã hóa cho tTAV (Hình 49) để thử nghiệm chức năng cắt nối đặc hiệu giới tính khi được dung hợp với tTAV.

Tác giả sáng chế đã hợp nhất LA1172 vào bộ gen của loài *Stegomyia aegypti* bằng cách sử dụng *piggyBac*. Hai dòng độc lập đã được tạo ra (các dòng 2 và 8). Cả hai dòng này đều thể hiện sự cắt nối chính xác của gen Actin-4-tTAV (Hình 50 và 51). Vì vậy, trình tự khởi đầu phiên mã Actin-4 và intron cắt nối thay thế có thể được sử dụng thành công để tạo ra sự cắt nối đặc hiệu giới tính, mô và giai đoạn của một gen mong muốn trong loài *Stegomyia aegypti.*

Ví dụ 10: Ngài

Gần đây, các tác giả sáng chế đã tạo ra các cấu trúc dựa trên dữ liệu biểu hiện tạm thời bằng cách sử dụng cấu trúc minigen tái tổ hợp bắt nguồn từ Bombyx mori. Cấu trúc này còn được mô tả dưới đây trong phần có được gọi "Moth *dsx* sequence alignment and conserved motifs"

Ví dụ 11: Sử dụng Bztra

Gần đây, các tác giả đã tạo ra hai cấu trúc dựa trên Bztra, được biểu hiện ở Mexfly (LA3376). LA3376 tạo ra khả năng gây chết đặc hiệu con cái có thể ức chế được. Trước đây, các tảc giả chỉ ra chức năng và cắt nối chính xác của LA3376 trong Medfly. Các thể biến nạp trong Mexfly {*Anastrepha ludens*) cũng được tạo ra với LA3376. Các cấu trúc này được phân tích việc cắt nối chính xác intron Bztra để chứng minh khoảng phát sinh loài của intron Bztra bởi RT-PCR sử dụng các đoạn mồi SRY và AV3F (Hình. 15 và đoạn nêu trên "các gel Medfly RT-PCR"). Cấu trúc này thể hiện việc cắt nối chính xác intron Bztra trong Mexfly.

Ví dụ 12: Dm *dsx* trong Medfly (DmDsx trong ví dụ Medfly chuyển gen: dung hợp nipper trong # 797)

Gần đây, các tác giả cũng tạo ra dữ liệu dựa trên cấu trúc Dmdsx trong Medfly. Cấu trúc này sử dụng một đoạn ở gen doublesex Drosophila melanogaster để tạo ra sự biểu hiện đặc hiệu giới tính của một đoạn Drosophila melanogaster gen NipplDm (các tác giả gọi đoạn này là "nipper"). Các tác giả không thấy rõ sự cắt nối đặc hiệu giới tính rõ rệt. Tuy nhiên, số liệu kiểu hình cho thấy mội số sự cắt nối đặc hiệu giới tính; các tác giả nhìn thấy khả năng gây chết gia tăng ở các con cái, đến khoảng 70% sự xâm nhập. Tất nhiên, sự xâm nhập không đầy đủ này có thể là do mức biểu hiện của nipper không có tính độc trong Medfly, v.v.. Các tác giả cũng nhận thấy có sự giảm đáng kể về số lượng con đực, nhưng bản thân nguồn tTA, LA670, được sử dụng trong thí nghiệm này có thể giết chết một vài con đực.

Các tác giả đã kiểm tra ba dòng gen chuyển Medfly độc lập mà có thể mang một thể dung hợp của nipper với trình tự DmDsx mà được dự định để được biểu hiện cụ thể ở các các con cái. Cấu trúc này có thể không hoạt động một cách hoàn hảo có thể do trình tự thiết yếu để cắt nối thay thế chính xác và/hoặc các vị trí liên kết SxI được quy định bởi DmDsx, và vì Medfly không sử dụng SxI trong con đường xác định giới tính, DmDsx có thể không có khả năng cắt nối hoàn toàn thể dung hợp này theo cách chính xác trong Medfly. Tuy nhiên, các tác giả đã thành công trong việc tạo ra khả năng gây chết gia tăng ở các con cái so với các con đực ở cả ba dòng với hiệu quả rất giống nhau (khả năng gây chết được quan sát ở các con cái lớn hơn xấp xỉ 75% so với con đực). Điều này chứng minh hệ *dsx* có thể hoạt động ở loài có liên quan gần gũi (sự tách tiến hóa có từ 120-150 triệu năm), và nếu trình tự Ccdsx được sử dụng nó có thể hoạt động tốt do yêu cầu SxI của Dmdsx .

Các kết quả 797 thể hiện dưới đây, sử dụng một hệ (Pub EGFP) nipper cắt nối TetO14 dsx. Chúng thể hiện rằng hệ này có khả năng gây chết tại giai đoạn phôi thai (-50%), và có khả năng tác động hiệu quả hơn ở các con cái (-75%).

797 được đánh dấu màu xanh lá cây (G), 670 với màu đỏ (R). 670 là nguồn tTAV, vì vậy mong muốn thấy một kiểu hình ở các con ruồi R+G; G (và R) chỉ là các con đối chứng. Trong đó có NF – không huỳnh quang (kiểu dại chẳng hạn) cũng là loài đối chứng. Tất cả các thế hệ con được nuôi trên môi trường không có tet.

Cả ba dòng độc lập dường như tác động theo cách giống nhau.

797A/797A M2 x 670/+:			
	Nhộng	Con trưởng thành	Con cái: con đực
G	184	176	85: 91
R+G	74	57	44:13
797C/797C M1 x 670/+:			
	Nhộng	Con trưởng thành	Con cái: con đực
G	169	157	89: 68
R+G	94	67	54:13
797C/797C M2 x 670/+:			
	Nhộng	Con trưởng thành	Con cái: con đực
G	406	377	197: 198
R+G	171	147	121:26
670A/797A C + x 797C/+:			
	Nhộng	Con trưởng thành	Con cái: con đực
NF	198	192	92: 100
G	162	147	67:80
R	149	72	43:29
R+G	45	22	20:2

Trung bình của ba dòng: số lượng các con cái R+G = 21% số lượng của các con đực R+G, vì vậy khả năng gây tử vong ở các con cái R+G về cơ bản vượt qua khả năng ở các con đực. Sự ảnh hưởng này không chỉ nhìn thấy ở các con cái đối chứng R hoặc G, mà còn không nhìn thấy ở kiểu dại.

Các ví dụ 13-15:

Gần đây, các tác giải đã chứng minh rằng:

(5) việc cắt nối đặc hiệu giới tính trong các cấu trúc minigen dựa trên *Aadsx* tái tổ hợp;

(6) kiểu hình đặc hiệu giới tính từ cấu trúc dựa theo *Cctra*; và

(7) việc cắt nối đặc hiệu giới tính ở các cấu trúc dựa trên Aedes-Actin4.

Ít nhất một trong số mỗi ví dụ này không chỉ thể hiện các minigen, mà thực tế còn thể hiện việc cắt nối để tạo ra tTAV/tTAV2 hoặc ubi-tTAV2

Ví dụ 13: Các minigen Aedes lưỡng tính (dsx)

Cũng xem phần được gọi là vị trí liên kết Aedes *dsx* Tra2. Các tác giả đã phân lập được gen *Aedes aegypti dsx* (Aadsx) và nhận biết 6 sản phẩm phiên mã từ vùng này (Hình. 1). Có: 2 sản phẩm phiên mã đặc hiệu con đực (M1 và M2), 3 sản phẩm phiên mã đặc hiệu con cái (Fl, F2 và F3) và một sản phẩm phiên mã được tìm thấy trong cả con đực và con cái (MF). Các tác giả tạo ra hai cấu trúc minigen. Trong các cấu trúc này, phần lớn trình tự intro được loại bỏ. Ví dụ, minigen *dsx* 1 độ dài xấp xỉ 4,4kb, trong khi các trình tự cuối cùng của nó được tách ra có độ dài xấp xỉ 26kb trong phạm vi tự nhiên của nó, trong ADN bộ gen của *Aedes aegypti*.

Việc cắt nối trong minigen2 của Hình 1 cho thấy việc cắt nối xuất hiện ở dạng "giống cái" trong cả các con cái và các con đực. Điều này có thể có nghĩa là hệ này phụ thuộc vào việc sử dụng chất nhận cắt nối thay thế. Trong kiểu này, có sự cạnh tranh giữa các chất nhận cắt nối khác, với một vài yếu tố đặc hiệu giới tính thiên về kiểu này, yếu tố đặc hiệu giới tính này chắc chắn là Tra. Nhưng việc loại bỏ các chất nhận cắt nối M1 và M23' thúc đẩy việc cắt nối ở các dạng F, bằng việc loại bỏ một chất nhận khác.

Vì vậy, được ưu tiên là một hoặc nhiều chất nhận cắt nối (Fl và/hoặc F2) 3' con cái đặc hiệu được đưa ra cùng với việc thêm chất nhận cắt nối đầu 3'. Tốt nhất, nếu chất nhận cắt nối thêm vào nêu trên là chất nhận đầu 3' của biến thể cắt nối M1 hoặc M2 (hoặc cả hai), mặc dù nó được dự tính là điều này không thực chất như các chất nhận cắt nối 3' được biết khác giống như chức năng.

Hình 1 minh họa các sản phẩm phiên mã khác nhau được sinh ra bởi việc cắt nối thay thế của gen *Aedes aegypti doublesex* (*Aadsx*). Cần phải hiểu rõ rằng

Aedes aegypti cũng được biết đến như là *Stegomyia aegypti*. Hình này cho thấy gen *Aadsx* từ exon thứ tư, mà không được cắt nối thay thế, nghĩa là có mặt trong tất cả các sản phẩm phiên mã được mô tả ở đây. Việc đánh số là từ nucleotit thứ nhất của exon thứ tư (acgacgaact...). Chú ý rằng biểu đồ này không theo tỷ lệ - các intron dài hơn các exon. Tổng số các vùng được cắt nối thay thế gồm trên 43kb.

Đoạn minigen có trong cấu trúc biểu hiện (LA3515). *Aedes aegypti* chuyển gen được sinh ra bởi sự tái tổ hợp vị trí đặc hiệu vào trong vị trí attP, sử dụng phương pháp của Nimmo và cộng sự (2006 : Nimmo, D.D., Alphey, L. Meredith, J.M. and Eggleston, P (2006). High efficiency site-specific genetic engineering of the mosquito genome. Insect Molecular Biology, 15: 129-136).

Thứ hai, minigen nhỏ hơn được xây dựng (DSX minigen2) và cấu trúc biểu hiện đối với minigen này được chèn vào vị trí atfP giống như minigen1 DSX, để cho phép so sánh trực tiếp (LA3534). *dsx* minigen2 không thể hiện việc cắt nối giới tính đặc hiệu. Điều này chỉ ra rằng các trình tự có mặt trong *dsx* minigenl nhưng không có trong *dsx* minigen2 (approx 2029bp, xem Hình 1 và SEQ ID NO. 150, trong đó các exon được tìm thấy ở các vị trí 29-163 và 1535-2572) là thiết yếu đối với việc cắt nối thay thế chính xác, thậm chí qua intro được cắt nối thay thế thứ nhất, và trình tự exon ngay bên cạnh nó, có mặt trong cả hai cấu trúc.

Các tác giả tạo ra hai gen chuyển (LA3491 và LA3534) sử dụng cấu trúc minigen của gen *Aedes aegypti dsx*. LA3491 là thể dung hợp của exon4 được chia ra, các exon catxet đặc hiệu con cái, và phần exon 3' được chia ra thứ nhất (exon 5 trong sản phẩm phiên mã M1).

Các sản phẩm phiên mã từ vùng minigen của LA3491 được phân tích bởi PCR transcriptaza ngược (RT-PCR) và sự sắp xếp trình tự. Các sản phẩm phiên mã tương ứng với việc cắt nối khác trong dạng F2 được tìm thấy trong các con cái nhưng không có trong các con đực (Hình 2 và 3) và trong dạng F1 có một vài sự biểu hiện con đực nhưng rất thấp (Hình 4). Trong khi các sản phẩm phiên mã tương ứng với dạng M1 được loại bỏ trong các con đực nhưng không được loại bỏ trong các con cái (Hình 2). Vì minigen không chứa chất nhận cắt nối 3'

của biến thể M2, sản phẩm phiên mã này không thể là từ cấu trúc này. Minigen này không chứa trình tự ngoại sinh bất kỳ, mặc dù đã chứng minh rõ rằng việc cắt nối đặc hiệu giới tính của đoạn *Aadsx*, có một đoạn "minigen" được loại bỏ cao.

Cần phải hiểu rằng các trình tự nhất định là quan trọng cho việc kiểm soát cắt nối và vì vậy cần được giữ lại, như được mô tả ở đây. Điều này có thể được thiết lập dễ dàng bởi sự loại bỏ các phần nhất định và kiểm tra việc cắt nối thay thế bởi RT-PCR chẳng hạn.

Hình. 2 thể hiện RT-PCR của các con cái và các con đực từ dòng chuyển gen LA3491 *Aedes aegypti* sử dụng các đoạn mồi 688 - ie1-transcr (SEQ ID NO. 4) và 790 - Aedsx-m-r2 (SEQ ID NO. 5). Bằng cách sử dụng các primer này, việc cắt nối ở kiểu F2 sẽ tạo ra dải khoảng 985bp trong khi việc cắt nối ở kiểu M1 sẽ tạo ra dải khoảng 516bp. Một dải khoảng 985bp (F2) chỉ được xuất hiện trong các đường biểu hiện các con cái và một dải sản phẩm phiên mã 1(M1) đặc hiệu con đực khoảng 516bp chỉ xuất hiện ở các con đực. Các dải này được xác định trình tự và cho thấy đã xảy ra việc cắt nối chính xác, nghĩa là lần lượt là loại F2 và M1 tương ứng. Sự có mặt của các dải trong các mẫu đối chứng không RT (-RT CON) cho thấy không có sự nhiễm tạp ADN bộ gen trong các mẫu. Các đường 1 và 11 là Marker (SmartLadder™ từ Eurogentec, các dải từ 1,5kb đến 0,2kb được chỉ ra). Các đường 2 và 3 là các đối chứng âm tính (không có enzyme phiên mã ngược) và các đường 2-9 là các sản phẩm phản ứng được tạo thành trên các phần chiết từ các con đực hoặc các con cái được biểu thị.

Hình. 3 thể hiện RT-PCR của các con cái các con đực và các con cái từ các dòng chuyển gen LA3491 *Aedes aegypti* bằng cách sử dụng các đoạn mồi 688 - ie1-transcr (SEQ ID NO. 4) và 761 - Aedsx-fem-r (SEQ ID NO. 6). Bằng cách sử dụng các đoạn mồi này, việc cắt nối trong khuôn F2 sẽ tạo ra một dải khoảng 525bp. Một dải khoảng 525bp có mặt trong các sản phẩm phản ứng trên các phần chiết từ các con cái, nhưng không có trong các sản phẩm phản ứng tương ứng trên các phần chiết từ các con đực. Việc xác định trình tự của dải 525bp này khẳng định điều đó là đúng, nghĩa là đã xảy ra việc cắt nối loại F2. Marker (SmartLadder™ from Eurogentec, các dải từ 1,5kb đến 0,2kb được chỉ ra).

Hình. 4 thể hiện RT-PCR của các con đực và các con cái từ các dòng chuyển gen LA3491 *Aedes aegypti* sử dụng các đoạn mồi 688 - ie1-transcr (SEQ ID NO. 4) và AedsxR1 (SEQ ID NO. 47). Sử dụng việc cắt nối các đoạn mồi này trong khuôn F1 sẽ tạo ra một dải bằng 283bp. Một dải khoảng 283bp phần lớn có mặt ở các con cái, mặc dù bằng chứng cho thấy một lượng nhỏ sản phẩm cắt nối này có mặt ở các con đực. Việc xác định trình tự xác nhận rằng rằng dải này quả thực tương ứng với việc cắt nối trong khuôn F1. Marker (SmartLadder™ từ Eurogentec, các dải từ 1, 5kb đến 0,2kb được chỉ ra).

LA3534 giống hệt với LA3491 chỉ khác là mất đoạn đầu 3' khoảng 2kb. Cấu trúc này cho thấy không có việc cắt nối khác nhau giữa con đực và con cái (Hình 1, minigen 2). Các gel RT-PCR không được thể hiện đối với trường hợp này. Dựa trên các các kết quả này, một số cấu trúc đã được thiết kế để kết hợp với việc cắt nối đặc hiệu giới tính của LA3491 (Hình 1, minigen 1) vào hệ phản hồi chủ động. LA3612 (Hình 5), mà hợp nhất thể dung hợp của ubiquitin và tTAV2 vào vùng mã hóa *dsx*, được thiết kế sao cho khi sản phẩm phiên mã con cái F2 được tạo ra, thì ubiquitin được tách ra và tTAV2 được giải phóng để bắt đầu và duy trì hệ phản hồi chủ động. LA3619 (Hình 5) có tTAV2 không chứa ubiquitin và sử dụng condon khởi đầu dịch mã của riêng nó. LA3646 (Hình 5) là giống hệt với LA3619 chỉ khác là các condon khởi đầu đối với gen *dsx* đã được đột biến; điều này cải thiện chất lượng của tTAV2 được sinh ra bởi việc loại bỏ sự dịch mã không đặc hiệu.

Hình. 5 là miêu tả bằng biểu đồ của các plasmit dựa theo sự cắt nối trong minigen *Aedes aegypti* dsx. Để rõ ràng, cần hiểu là intron con cái thứ nhất là của việc cắt nối F1, F2 hoặc F3 bất kỳ, và tTAV trong biểu đồ liên quan đến tTAV2 (cần hiểu rõ là các protein khác hoặc các kiểu tTA hoặc tTAV khác có thể được lựa chọn sử dụng). Trong mỗi plasmit này, ngoài LA3491, trình tự khác loại được thêm vào exon F2. "ATG giả định" là trình tự codon ATG bất kỳ trong trình tự exon được định vị 5' tương ứng với ADN khác loại. Trong LA3646, các đơn vị mã hoá khởi đầu dịch mã giả định này ("ATG giả định") được loại bỏ hoặc được cải biến. Trong trường hợp cấu trúc LA3612, sự dịch mã từ một (5') ATG ngược chiều nằm trong khung với vùng mã hóa ubi-tTAV sẽ vẫn (giả sử không có codon kết thúc can thiệp) tạo ra tTAV chức năng, tiếp theo là sự tách

các gốc ubiquitin và tTAV dưới sự tác động của proteaza. Các catxet (cassete) cắt nối thay thế được liên kết theo cách hoạt động được với gen khởi đầu thích hợp, gen kết thúc phiên mã và các trình tự điều hòa khác.

Ví dụ này cho thấy việc cắt nối đặc hiệu giới tính của đoạn "minigen" được nén cao trong phạm vi khác loại (trình tự khởi đầu khác loại chẳng hạn, 5' UTR và 3'UTR). Mặc dù nó không cho thấy sự biểu hiện khác nhau của một trình tự không phải từ *Aedes*, do các exon được cắt nối thay thế được bắt nguồn từ gen *Aadsx* và không chứa vật liệu bổ sung, điều này minh họa rõ ràng tính khả thi của phương pháp này. Trong trường hợp bất kỳ, trình tự khởi đầu 5' UTR và 3'UTR là khác loại. Các tác giả đã tạo ra các cấu trúc khác mà minh họa các phương pháp khác nhau để thu được sự biểu hiện biệt hóa (đặc hiệu giới tính) của một protein khác loại bởi *dsx* này .

So sánh trình tự TRA

Pane và đồng tác giả (2002) đề xuất rằng các trình tự nhất định liên quan đến các vị trí liên kết được biết của phức hợp Tra/Tra-2 trong Drosophila có thể là quan trọng trong việc điều chỉnh việc cắt nối Cctra, và điều này cũng được biết trong *Drosophila dsx* và cũng được đề xuất trong *Anopheles gambiae dsx* (Scali et al., 2005). Trình tự liên ứng là được mô tả khác nhau như:

UC(U/A)(U/A)C(A/G)AUCAACA (Pane et al.,), SEQ ID NO. 8, hoặc UC(U/A)(U/A)CAAUCAACA (Scali et al.,), SEQ ID NO. 9.

Điều đáng chú ý là các định nghĩa này là cực kỳ giống nhau. Pane và cộng sự nhận ra 8 cặp đôi không hoàn chỉnh với sự liên ứng này trong trình tự Cctra (7 hoặc nhiều hơn nucleotit cặp đôi với trình tự liên hiệp 13 nucleotit. Scali và cộng sự nhận biết 6 tiếp hợp trong *Agdsx* (9/13 hoặc tốt hơn). Các trình tự như vậy cũng được biết để điều chỉnh việc cắt nối thay thế của Drosophila gen fi-uitless; Scali và cộng sự xem xét 3 cặp đôi trong trình tự này (12/13 hoặc tốt hơn). Việc cắt nối chính xác của *dsx* cũng yêu cầu vùng có nhiều purin, như được mô tả bởi Scali và cộng sự.

Như có thể thấy từ Bảng 2 và Hình. 7, các tác giả đã nhận ra các cụm đáng kể gồm các vị trí liên kết đối với Tra/Tra2 trong minigen 1 *Aedes aegypti dsx*.

Sắp thẳng hàng trình tự Moth *dsx* và các motif được bảo toàn

Hình. 6 thể hiện sự sắp thẳng hàng của các exon đặc hiệu con cái thứ hai và các trình tự kẹp sườn của các gen *dsx* từ sâu hồng hại bông (Pectinophora gossypiella, PBW-dsx, SEQ ID NO. 146), con tằm (Bombyx mori, bombyx-dsx, SEQ ID NO. 147) và sâu bướm đục trái (Cydia pomonella, codling- dsx, SEQ ED NO. 148). Exon đặc hiệu con cái thứ hai được in đậm. Các tác giả đã xác định các sản phẩm phiên mã của một trình tự nulceotit ngắn, lặp lại, có trình tự bảo toàn và sự định vị gần giữa các loài bướm đêm tương đối xa này; các sản phẩm phiên mã này chỉ được đặt ở đầu 5' so với exon đặc hiệu con cái. Các đoạn lặp bảo toàn AGTGAC/T được gạch chân. Dấu hoa thị (*) là các nucleotit giống hệt nhau, dấu gạch ngang (-) là các khoảng trống để sắp thẳng hàng theo cách tốt nhất. Các exon được thể hiện trong SEQ ID NOS. bằng cách đánh số nucleotit như sau: SEQ ID NO. 146289-439; SEQ ID NO. 147 339-492; và SEQ ID NO. 148 285-439.

Các vị trí liên kết *Aedes dsx Tra2*.

Trong các con cái của Drosophila melanogaster, Tra và một sản phẩm từ gen tác động chủ yếu tra2, đóng vai trò làm các chất điều chỉnh việc cắt nối bằng cách liên kết với các vị trí tăng cường cắt nối trên pre-mARN của *dsx*, mà tác động vào vị trí chất nhận 3' yếu của exon đặc hiệu con cái (Scali *và cộng sự*). Trong các con đực, không có sự biểu hiện của TRA và vị trí chất nhận đầu 3 ' yếu và việc cắt nối xuất hiện ở exon con đực. Để tìm kiếm các vị trí liên kết Tra/Tra2 giả định, các tác giả sử dụng trình tự liên ứng của các vị trí liên kết được suy diễn cho Drosophila Tra/Tra2 và tìm thấy sự phân bố của các vị trí này trong trình tự gen *Aedes aegypti dsx*. Điều này được thể hiện trong Bảng 2, dưới đây.

Bảng 2: * = trong 3491, chỉ 9/13 nhưng 6/6 trong trung tâm. Bảng này không bao gồm 9/13 phần tử đồng nhất từ các phần này là trong 3491 với mức đồng nhất 6/6 với trình tự ở trung tâm của wwcrat. Trình tự trung tâm liên ứng (WWCRAT) đặc biệt được ưu tiên.

Tên	Trình tự w = T hoặc A r = A hoặc G	Có mặt trong minigen 1	Vị trí	Mức đồng nhất với trình tự liên ứng	Đồng nhất với wwcrat	SEQ ID NO.
Trình tự liên ứng	tc**wwcrat**caaca	/	/	/13	/6	138
1	tcaacaagcaaca	Y	14917	12	5	10
2	ttatcaaacaaca	Y	364	11	5	11
3	tcatcaattaaaa		1015	11	**6**	12
4	tcatcaatcaaac		6502	11	**6**	13
5	tcttcaaccaacc	Y	14958	11	5	14
6	cctacaatctaca	Y	14973	11	**6**	15
7	tcttagatcaaaa		16553	11	5	16
8	tcttcgatcatta		17386	11	**6**	17
9	ccaacaatctaca		28802	11	**6**	18
10	tcaaagatcacca		42096	11	5	19
11	tcttcggtcgacg	Y	256	11	5	20
12	tcgacaaacaaaa		1277	11	<5	21
13	tattcaaacaacg		4061	11	5	22
14	ttttcgataaaaa		4380	10	**6**	23
15	tcttcagtctgca		5399	10	**5**	24
16	gattcaatcatca		7723	10	**6**	25
17	ttatcgagcaaaa		8137	10	5	26
18	tcataactcaaga		9062	10	<5	27
19	tcagaaatcaaaa		9126	10	<5	28
20	tctttaatttaca		10639	10	5	29
21	tttacaatcctca		10646	10	**6**	30
22	tcatagatcagga		11214	10	5	31
23	acctcaaacaaca		11989	10	<5	32
24	tcatcgaacaccc		12020	10	5	33
25	tcaataatcgtca		12199	10	5	107
26	tcatcaaacgtca		13287	10	5	108
27	ttatcgttaaaca	Y	13439	10	5	109
28	taaacagtcaata	Y	13446	10	5	110
29	tacacgatcagca	Y	14096	10	5	111

Tên	Trình tự w = T hoặc A r = A hoặc G	Có mặt trong minigen 1	Vị trí	Mức đồng nhất với trình tự liên ứng	Đồng nhất với wwcrat	SEQ ID NO.
30	aatacaaacaaca	Y	14637	10	5	112
31	tcatcaacaagca	Y	14914	10	5	113
32	tctacaaaccaga	Y	14980	10	5	114
33	acatcgattcaca		16085	10	6	115
34	cgctcaatcaaca		16175	10	5	116
35	tctaccataaaaa		16511	10	5	117
36	aaatgaatcaaca		20044	10	5	118
37	acatcgttcaacg		21374	10	5	119
38	tcttgattcacca		21580	10	<5	120
39	tctgcagacaaca		22408	10	<5	121
40	tcttcggtaatca		23285	10	5	122
41	tctataaacaata	Y	25436	10	<5	123
42	taaacaataaata	Y	25440	10	**6**	124
43	taaacaagcaaaa		28242	10	5	125
44	tcaacgatcggcg		30309	10	**6**	126
45	tgatccatcatca		30910	10	5	127
46	tcaacatgcaaga		32295	10	<5	128
47	tcttaaataaaga		32862	10	5	129
48	tcaaagatctata		40551	10	5	130
49	taatgaattaaca		40847	10	5	131
50	tttaccatcaact		41712	10	5	132
51	taatgaaacaaca		43380	10	<5	133
52*	gtttcaattaaaa	Y	13500	9	**6**	134
53*	tattcaattataa	Y	13602	9	**6**	135
54*	tcttcaatcgttt	Y	15002	9	**6**	136
55*	tcaacgatccttt	Y	15533	9	**6**	137

Hình 7 là sơ đồ thể hiện các vị trí liên kết Tra/Tra2 giả định thuộc vùng mã hóa *dsx* của *plasmit* LA3491. Sơ đồ này có tỷ lệ gần đúng và thể hiện một trình tự khoảng 4kb. Có thể tính toán khả năng bắt cặp ngẫu nhiên với trình tự liên ứng Tra/Tra2. Giả sử tất cả 4 nucleotit xuất hiện ở tần suất như nhau, khả năng của một nucleotit bất kỳ trong trình tự ngẫu nhiên trở thành nucleotit đầu tiên là 10/13 hoặc bắt cặp tốt hơn với trình tự liên ứng là khoảng $7x10^{-4}$. Vì vậy, người ta mong đợi ít hơn một lần bắt cặp như vậy trên 1000 nucleotit thuộc trình tự ngẫu nhiên này. Việc tính toán là như dưới đây:

Việc cắt nối giới tính đặc hiệu: các khả năng có thể xẩy ra

Câu hỏi

Một trình tự liên ứng vị trí liên kết gồm có 13 baza. Mười vị trí này (được gắn) (gọi là bộ X) phải là một bazơ đặc hiệu. Ba bazơ khác (gọi là bộ Y) có thể mỗi bazơ là một trong số hai bazơ đặc hiệu.Việc đảm bảo là mỗi bazơ có thể là A, G_5 C và T là có khả năng giống nhau và bazơ này ở mỗi vị trí là độc lập trong số các bazơ tại các vị trí khác, có khả năng là một trình tự bazơ 13 được chọn cực kỳ ngẫu nhiên phù hợp với trình tự này? Các khả năng của một trình tự như vậy là một ghép đôi không xứng tương tự (cho phép tối là một, hai, ba hoặc bốn bazơ khác nhau)? Các câu trả lời được đưa ra trong Bảng 2 dưới đây và cách thực hiện được mô tả bên dưới.

Trả lời

Bảng 3

Số vị trí bắt cặp sai	Xác suất (phân đoạn)	Xác suất (đến 3 d.p.)
Không, nghĩa là bắt cặp chính xác	$\frac{1}{2^{23}}$	$1,192 \times 10^{-7}$
lên tới 1, nghĩa là ít nhất 12 vị trí bắt cặp	$\frac{17}{2^{22}}$	$4,053 \times 10^{-6}$
lên tới 2, nghĩa là ít nhất 11 vị trí bắt cặp	$\frac{133}{2^{21}}$	$6,342 \times 10^{-5}$
lên tới 3, nghĩa là ít nhất 10 vị trí bắt cặp	$\frac{23}{2^{15}}$	$7,019 \times 10^{-4}$
lên tới 4, nghĩa là ít nhất 9 vị trí bắt cặp	$\frac{33863}{2^{23}}$	$4,037 \times 10^{-3}$

Thực hiện

$$\text{P(bắt cặp chính xác)} = P_0 = \left(\frac{1}{4}\right)^{10}\left(\frac{1}{2}\right)^{3} = \frac{1}{4^{10}\times 2^{3}} = \frac{1}{2^{23}} = 1,192\times 10^{-7}$$ đến 3 d.p. (*3 d.p.* như dưới đây)

P(bắt cặp sai ở chính xác một vị trí) = P (bắt cặp sai ở một trong số 10 vị trí X hoặc bắt cặp sai tại một trong số 3 vị trí Y)

$$= P_1 = 10\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)\left(\frac{1}{2}\right)^3 + 3\left(\frac{1}{4}\right)^{10}\left(\frac{1}{2}\right)^3 = \frac{(10\times 3)+3}{4^{10}\times 2^3} = \frac{33}{2^{23}} = 3{,}934\times 10^{-6}$$

P(bắt cặp sai theo chính xác hai vị trí) = P(bắt cặp sai tại hai trong số 10 X vị trí hoặc bắt cặp sai tại một trong số 10 X một trong số 3Y hoặc bắt cặp sai tại hai trong số 3Y)

$$= P_2 = \frac{10!}{2!8!}\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2\left(\frac{1}{2}\right)^3 + 10\times 3\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)\left(\frac{1}{2}\right)^3 + 3\left(\frac{1}{4}\right)^{10}\left(\frac{1}{2}\right)^3$$
$$= \frac{\left((45\times 3^2)+(30\times 3)+3\right)}{2^{23}} = \frac{498}{2^{23}} = \frac{249}{2^{22}} = 5{,}937\times 10^{-5}$$

P(bắt cặp sai theo chính xác ba vị trí 3) = P(bắt cặp sai tại ba trong số 10 X vị trí hoặc bắt cặp sai tại hai trong số 10 X một trong số 3Y hoặc bắt cặp sai tại một trong số 3Y hoặc bắt cặp sai tại một trong số 10X và 2 trong số 3Y hoặc bắt cặp sai tại 3 trong số 3Y)

$$= P_3 = \frac{10!}{3!7!}\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3\left(\frac{1}{2}\right)^3 + \frac{10!}{2!8!}3\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2\left(\frac{1}{2}\right)^3 + 10\times 3\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)\left(\frac{1}{2}\right)^3 + \left(\frac{1}{4}\right)^{10}\left(\frac{1}{2}\right)^3$$
$$= \frac{\left((120\times 3^3)+(45\times 3^3)+(30\times 3)+1\right)}{2^{23}} = \frac{5356}{2^{23}} = \frac{1339}{2^{21}} = 6{,}385\times 10^{-4}$$

P(bắt cặp sai theo chính xác 4 vị trí) = P(bắt cặp sai tại 4 trong số 10 X hoặc bắt cặp sai tại 3 trong số 10 X và 1 trong số 3 Y hoặc bắt cặp sai tại 2 trong số 10 X và 2 trong số 3 Y hoặc bắt cặp sai tại 1 trong số 10 X và 3 trong số 3 Y)

$$= P_4 = \frac{10!}{4!6!}\left(\frac{1}{4}\right)^6\left(\frac{3}{4}\right)^4\left(\frac{1}{2}\right)^3 + \frac{10!}{3!7!}3\left(\frac{1}{4}\right)^7\left(\frac{3}{4}\right)^3\left(\frac{1}{2}\right)^3 + \frac{10!}{2!8!}3\left(\frac{1}{4}\right)^8\left(\frac{3}{4}\right)^2\left(\frac{1}{2}\right)^3 + 10\left(\frac{1}{4}\right)^9\left(\frac{3}{4}\right)\left(\frac{1}{2}\right)^3$$
$$= \frac{\left((210\times 3^4)+(120\times 3^4)+(45\times 3^3)+(10\times 3)\right)}{2^{23}} = \frac{27975}{2^{23}} = 3{,}335\times 10^{-3}$$

P(bắt cặp sai ở lên tới 1 vị trí)

$$= P_0 + P_1 = \frac{1+33}{2^{23}} = \frac{17}{2^{22}} = 4{,}053\times 10^{-6}$$

P(bắt cặp sai ở lên tới 2 vị trí) $= P_0 + P_1 + P_2 = \frac{1+33+498}{2^{23}} = \frac{532}{2^{23}} = \frac{133}{2^{21}} = 6{,}342 \times 10^{-5}$

P(bắt cặp sai ở lên tới 3 vị trí)

$$= P_0 + P_1 + P_2 + P_3 = \frac{1+33+498+5356}{2^{23}} = \frac{5888}{2^{23}}$$

$$= \frac{23}{2^{15}} = 7{,}019 \times 10^{-4}$$

P(bắt cặp sai ở lên tới 4 vị trí)

$$= P_0 + P_1 + P_2 + P_3 + P_4 = \frac{1+33+498+5356+27975}{2^{23}} = \frac{33863}{2^{23}}$$

$$= 4{,}037 \times 10^{-3}$$

Thí nghiệm 14: Cctra

Các tác giả có một dòng LA3097 (LA3097A) mà cho thấy sự biểu hiện rất tốt của các chỉ thị huỳnh quang; vẫn chưa biết liệu dòng này có phải là một trường hợp hợp nhất duy nhất hay không. Dòng này thể hiện dấu hiệu cắt nối đặc hiệu giới tính, nếu được nuôi không có tetraxyclin, tất cả các con cái chết dưới dạng các phôi, và nếu được nuôi trên tetraxyclin với lượng 30µg/M1 cả con đực và con cái đều sống sót.

Ví dụ này là quan trọng. Nó cho thấy Cctra tạo ra việc cắt nối giới tính đặc hiệu khác trong Aedes, và điều này có thể được sử dụng để tạo ra khả năng làm chết giới tính đặc hiệu. Vì vậy, điều này đưa ra bằng chứng về khoảng phylogene trong việc cắt nối Cctra. Vì vậy, ~~là~~ hoàn toàn hợp lý là sáng chế có thể ứng dụng với tất cả Diptera, như các tác giả cho thấy Cctra làm trong Drosophila, tephritids và các con muỗi, mà chủ yếu là quay quanh toàn bộ Dipteran Order.

Bất ngờ~~Ngạc nhiên~~ là Cctra làm trong Aedes, tạo thành sự phát triển trình tự nhanh chóng của tra.

Các tác giả thay đổi *Aedes aegypti* bằng cấu trúc LA3097. Các con đực khác loại từ dòng chuyển gen tạo thành được lai với loài tự nhiên và con lai được nuôi trong môi trường chứa nước được bổ sung với tetraxyclin đến nồng độ cuối cùng là 30 μg/M1. Các con trưởng thành được ghi lại như sau: 14 các con đực và một con cái, vì vậy cho thấy khả năng gây chết đặc hiệu con cái đáng kể.

Các loài này và giống thông thường có tỷ lệ giới tính xấp xỉ 1:1, vì vậy cấu trúc này tạo ra khả năng gây chết trong *Aedes aegypti*. Các cấu trúc tương đương mà không chứa trình tự intron CcTRA tạo ra khả năng gây chết không đặc hiệu giới tính. Vì vậy, intron CcTRA được sử dụng để tạo ra sự điều chỉnh khác nhau (đặc hiệu giới tính chẳng hạn) của sự biểu hiện gen trong các con muỗi, và điều có thể còn được sử dụng để tạo ra khả năng gây chết đặc hiệu giới tính và phương pháp loại bỏ chọn lọc con cái khỏi một quần thể.

Cụ thể hơn: trong 0μg/M1 tetraxyclin, các con đực sống sót chỉ với con nhộng, không làm cho nó thành con trưởng thành chẳng hạn. Các con cái chết quá sớm mà các tác giả không nhìn thấy chúng, chắc chắn dưới dạng phôi, vì vậy vẫn có một tác động khác giữa các giới tính. Tuy nhiên, khả năng gây chết nhộng trong các con đực đề xuất rằng hệ này không hoàn toàn biết mất trong các con đực. Dòng cài xen đơn mà các tác giả ghi nhận là không thông thường, bởi vì nó cho thấy chính xác sự biểu hiện mạnh của marker; các cài xen khác với nhiều mức biểu hiện điển hình có thể không thể hiện tốt khả năng gây chết con đực.

Cắt nối trong LA3097A

Phân tích việc cắt nối LA3097 từ các con muỗi chuyển gen LA3097A bởi RT-PCR cho thấy các con đực và các con cái được chia hai sản phẩm phiên mã, một dải xấp xỉ 950 bp và một dải yếu hơn xấp xỉ khoảng 800 bp (Hình. 59). Giải trình tự của các dải này cho thấy dải 900 bp tương ứng với biến thể cắt nối không đặc hiệu giới tính (AeM2, -920 bp), và dải yếu hơn là một hỗn hợp của một biến thể cắt nối không đặc hiệu giới tính (AeM1, ~804bp) và dạng con cái

(AeFl, ~765bp), xem Hình. 60. Việc cắt nối của sản phẩm phiên mã AeFl là giống hệt với việc cắt nối được nhận thấy đối với cấu trúc này trong Medfly (Hình. 33). Vệc cắt nối của các sản phẩm phiên mã M khác với việc cắt nối mà nhìn thấy trong tự nhiên (việc cắt nối Cctra trong Medfly, gen tự nhiên hoặc các tác giả quan sát từ LA3097 trong chuyển gen Medfly); trong AeM1m exon được cắt nối khác thứ hai (MEIb) không có trong sản phẩm phiên mã AeM1 thuần thục và trong AeM2, exon được tách thứ hai khác (ME2b) là tương tự không có trong sản phẩm phiên mã AeM2 thuần thục. Mặt khác, đối với mỗi trong số các sản phẩm phiên mã này, có mặt exon catxet thứ nhất mà không phải exon catxet thứ hai, liên quan đến mẫu đầu tiên Medfly. Chú ý rằng, do sự vắng mặt của exon catxet thứ hai trong AeM1, và khung đọc của tTAV2 liên quan với exon catxet thứ nhất trong cấu trúc này, việc cắt nối theo kiểu AeM1 không dẫn đến sự gián đoạn khung đọc mở tTAV2, mà còn thêm 39 nucleotit (tương ứng với 13 amino axit) giữa ATG và phần còn lại của khung đọc mở tTAV2. Dường như biến thể này của tTAV2 có thể duy trì một vài hoạt tính, liên quan tới tTAV2 kiểu nguyên thủy hoặc thông thường (như mã hóa bởi biến thể cắt nối F1). Sự có mặt của tetraxyclin, tác động kiểu hình được quan sát trong các con đực cũng như là trong các con cái, mặc dù trong các con đực yếu hơn so với các con cái. Sự sản sinh một biến thể hoạt hóa một phần của tTAV2 từ sản phẩm phiên mã AeM1 trong các con đực (và các con cái) có thể giải thích điều này.

Hình. 59 cho thấy RT-PCR của các con đực và các con cái từ dòng chuyển gen LA3097 *Aedes aegypti* sử dụng các đoạn mồi HSP (SEQ ID NO. 139) và VP16 (SEQ ID NO. 140). Bằng cách sử dụng các đoạn mồi này, việc cắt nối theo kiểu CcF1 (tương ứng với biến thể F1 của Ceratitis capitatd) sẽ tạo ra một dải xấp xỉ 765bp và việc cắt nối trong CcM1 1005bp và CcM2 1094bp. Trong cả các con đực và các con cái, dải mạnh khoảng 950bp (1) được quan sát với một dải yếu hơn khoảng 800bp (2). Marker (SmartLadder™ from Eurogentec, các dải từ 1,5kb đến 0,4kb được biểu thị).

Phân tích trình tự của các dòng từ dải 2 (các biết thể cắt nối AeM1/AeFl chẳng hạn) từ các con đực và các con cái cho thấy rằng một trong số năm dòng

từ các con cái thể hiện việc cắt nối AeM2 (20%), trong khi trong các con đực, ba trong số bốn dòng đã thể hiện việc cắt nối AeM2 (75%); tất cả các dòng khác thể hiện việc cắt nối AeF1. Điều này chỉ ra rằng có nhiều sản phẩm phiên mã AeF1 có trong các con cái hơn có trong các con đực và điều này sẽ giải thích tác động gây chết khác nhau giữa chúng.

Hình. 60 minh họa các sản phẩm phiên mã khác nhau được sinh ra bởi việc cắt nối thay thế của Cctra từ dòng chuyển gen LA3Q91A.*Aedes aegypti*. 3097 là trình tự ADN của Cctra và các số chỉ dẫn hình vẽ được mô tả ở đây. Phần bóng và đóng khung cũng liên quan đến Hình 33. Chú ý rằng biểu đồ này không theo tỷ lệ.

Ví dụ 15: Aedes Actin-4;

Các tác giả có 11 dòng LA3545, mà sử dụng gen Aedes actin-4 (AeAct-4 hoặc AaAct4) để điền khiển sự biểu hiện tTAV2. Trong cấu trúc LA3545, trình tự mã hóa tTAV2 được lồng vào exon thứ hai của AaAct4 (hình 10). Đối với các sản phẩm phiên mã được cắt nối theo kiểu đặc trưng của cắt nối AaAct4 trong các con cái, ATG của vùng mã hóa tTAV2 sẽ là (5 '-most) ATG thứ nhất của sản phẩm phiên mã. Việc cắt nối theo kiểu đặc trưng của việc cắt nối AaAct4 ở các con đực đưa một loạt các condon khởi đầu và kết thúc trước trình tự tTAV2 mà dự định để ức chế hoặc gây trở ngại cho quá trình dịch mã từ ATG của vùng mã hóa tTAV2. Các dòng này chỉ nên biểu hiện tTAV2 trong con nhộng cái. Việc cắt nối được thể hiện trong Hình. 8, dưới đây.

Hình 8 cho thấy RT-PCR của các con cái và đực trưởng thành từ dòng chuyển gen LA3545AeC *Aedes aegypti* sử dụng các đoạn mồi AgexonlF (SEQ ID NO. 141) và TETRRl (SEQ ID NO. 142). Bằng cách sử dụng các đoạn mồi này, việc cắt nối trong khuôn mẫu tương đương với khuôn mẫu của gen AaAct4 nguyên thể sẽ tạo ra các dải khoảng 347bp đối với biến thể cắt nối kiểu cái và dải khoảng 595bp đối với biến thể cắt nối kiểu đực. Dải khoảng 347bp (F) chỉ được tìm thấy trong các phản ứng trên các phần chiết từ các con cái; dải khoảng 595bp (M) được tìm thấy ở cả các con đực và các con cái. Việc xác định trình tự

xác nhận rằng việc cắt nối chính xác xảy ra ở cả con đực và con cái. Marker (SmartLadder™ từ Eurogentec, các dải từ 1, 5kb đến 0, 2kb được chỉ ra).

Các tác giả cũng có *Aedes aegypti* chuyển gen mang cấu trúc LA3604, mà tương tự với LA3545 chỉ khác là có một condon khởi đầu được thiết kế theo vị trí của exon 1 mà có trong cả sản phẩm phiên mã con đực và con cái (Hình 10). Nó được thiết kế với ATG thứ nhất trong kiểu sản phẩm phiên mã. LA3604 mã hoá tTAV2 được hợp nhất với ubiquitin (LA3545 mã hoá tTAV, trong khi LA3604 mã hoá ubi-tTAV2). Cấu trúc này chỉ tạo ra protein tTAV2 chức năng đầy đủ trong các con cái, thậm chí nếu dạng con đực được biểu hiện trong các con cái, exon con đực thêm vào chứa các condon khởi đầu và kết thúc mà sẽ ngăn ngừa sự dịch mã của protein dung hợp Ubi-tTAV2.

Việc cắt nối thay thế của AaAct4 xảy ra trong 5' UTR (của gen tự nhiên). Nó có thể hoặc không thể có vai trò điều chỉnh trong gen tự nhiên. Một khả năng là như sau: trong biến thể cắt nối đặc hiệu con cái, condon khởi đầu của vùng mã hoá AaAct4 là ATG thứ nhất của sản phẩm phiên mã. Tuy nhiên, trong biến thể cắt nối con đực đặc hiệu, có các trình tự ATG bổ sung khác nhau đầu 5' với đơn vị mã hoá bắt đầu của vùng mã hoá AaAct4; hầu hết trong số này có codon kết thúc trong khung, khoảng cách ngắn đầu 3'. Sắp xếp trình tự có thể gây trở ngại cho sự dịch mã có hiệu quả của protein AoAct4 và vì vậy làm giảm sự biểu hiện của protein ở các con đực so với các con cái. Đây là sự sắp xếp trong LA3545.

Tuy nhiên, tác động khác nhau lớn hơn giữa các con đực và các con cái srx được mong đợi nếu intron có trong vùng mã hoá (hơn 5' UTR), được cài xen vào giữa các condon khởi đầu và kết thúc của polynucleotit được biểu thị bằng biểu đồ chẳng hạn. Trong trường hợp này, exon catxet đặc hiệu con đực sẽ thay đổi khả năng mã hoá của sản phẩm phiên mã, hơn là chỉ can thiệp một cách đơn giản bằng sự dịch mã.

Điều này đạt được trong cấu trúc LA3604. Các tác giả đã điều biến exon thứ nhất được chia để có trình tự ATG trong phạm vi trình tự thích hợp đối với sự khởi đầu dịch mã. Trong trình tự được cải biến này, đây là ATG thứ nhất

trong các biến thể cắt nối loại con đực (M) hoặc con cái (F). Sau khi cắt nối ở dạng F, 5' ATG (được thiết kế) này là trong khung với vùng mã hoá ubi-tTAV. Vì vậy, các sản phẩm phiên mã loại F mã hoá protein dung hợp, bao gồm các phần đã được mã hoá bởi (i) phần mà thông thường là Act4 5' UTR (nhưng ở đây hiển nhiên được dịch mã, và vì vậy không hề có UTR), (ii) vùng mã hoá ubiquitin và (iii) vùng mã hoá tTAV2 .

Hoạt tính của các ubiquitin proteaza tế bào sẽ giải phóng protein tTAV2. Sự dịch mã từ 5' ATG đã được thiết kết sẽ được chấm dứt bởi codon kết thúc trong khung trong trình tự thêm vào (exon catxet) có trong các sản phẩm phiên mã được cắt nối ở dạng M. Vì vậy, điều này ngăn cản sự biểu hiện của tTAV2 chức năng trong các con đực, nhờ đó tạo ra sự biểu hiện đặc hiệu giới tính của tTAV2. Rõ ràng, điều này tạo ra một phương pháp chung để biểu hiện đặc hiệu giới tính một protein, bởi sự thay thế đoạn tTAV2 bằng protein hoặc trình tự quan tâm khác. Sử dụng chiến lược này, các tác giả đã tạo ra gen chuyển và việc cắt nối đặc hiệu giới tính được chỉ ra (Hình 9).

Hình. 9 thể hiện RT-PCR của các con đực và các con cái từ dòng chuyển gen LA3604AeA *Aedes aegypti* sử dụng các đoạn mồi AgexonlF (SEQ ID NO. 141) và TETRRl (SEQ ID NO. 142). Bằng cách sử dụng các đoạn mồi này, việc cắt nối ở con cái sẽ tạo ra dải khoảng 575 bp, trong khi sự bao hàm exon cát-xét đặc hiệu con đực sẽ làm tăng dải đến khoảng 823bp. Dải khoảng 575bp được nhận thấy ở mỗi con cái được phân tích, trong khi một dải khoảng 823bp được nhận thấy ở mỗi con đực được phân tích. Các dải này dường như về cơ bản đặc hiệu với các giới tính tương ứng. Xác định trình tự của các dải này cho thấy việc cắt nối chính xác đã xảy ra ở cả con đực và con cái. Marker: SmartLadder™ từ Eurogentec, các dải từ 1,5kb đến 0,2kb được chỉ ra.

Hình. 10, dưới đây, là sơ đồ biểu thị các *plasmit* LA3545 và LA3604. S1: exon được phân chia 1; M1: trình tự thêm vào có trong exon đặc hiệu con đực 1 ; S2: exon được phân chia 2 (chỉ đầu 5'); ubi: trình tự mã hoá ubiquitin; tTAV2: trình tự mã hoá tTAV2.

Nhiều dòng chuyển gen LA3545, tác động đặc hiệu giới tính- mô được quan sát: các con cái là không bay được. Hai dòng thể hiện 90-100% kiểu hình con cái không bay được, một dòng thể hiện 70% không bay được và 50% khác. Phenotip này là có lẽ đúng với sự biểu hiện con cái đặc hiệu của tTAV2 trong các cơ bay phát triển. Sự khác biệt về kiểu hình giữa các dòng này là do ảnh hưởng vị trí đối với sự biểu hiện của các trình tự khởi đầu AaAct4. Do vị trí các gen trong sự biểu hiện hệ gen có thể bị ảnh hưởng bởi số lượng các yếu tố (các vùng nhiễm sắc thể muộn hoặc nhiễm sắc thể điển hình, các yếu tố tăng cường và ức chế, gần với các gen khác) mà có thể nhìn thấy rõ ràng trong các marker huỳnh quang được sử dụng để nhận biết các gen chuyển. Tất cả mười một dòng LA3545 được nhận biết bởi vì chúng có các profin huỳnh quang khác nhau, thậm chí chúng có các trình tự khởi đầu và marker giống nhau. Sự biến đổi này giống với các tác động vị trí. Sau đó, điều này có nghĩa là các tác giả mong đợi một vài dòng LA3545 để biểu hiện tTAV2 nhiều hơn các dòng khác do các tác động vị trí, và các dòng nêu trên mà biểu hiện nhiều hơn sẽ tạo ra kiểu hình đặc hiệu con cái không bay được.

Để kiểm tra các giả thuyết này, các tác giả đã phát triển dòng *Aedes aegypti* tách với gen báo cáo tetO-DsRed2 (LA3576 xem Hình. 17 và SEQ ID NO. 143), khi được lai với các dòng LA3545 khác nhau, điều này sẽ cho phép thấy rõ Actin4-tTAV2 đang được biểu hiện ở đâu và khi nào. Vì 8 dòng LA3545 được với LA3576, tất cả biểu hiện sự phát huỳnh quang cơ bay gián tiếp con cái đặc hiệu trong ấu trùng muộn L4, nhộng và các con trưởng thành. Trong bốn dòng DsRed2 sự biểu hiện dường như đặc hiệu (nghĩa là độc quyền) với các cơ bay gián tiếp ở con cái; trong bốn mô khác biểu hiện DsRed2. Hiện tượng này, ở đó sự biểu hiện của gen chuyển phụ thuộc một phần vào vùng hoặc điểm trong bộ gen mà nó được cài xen vào, được gọi tác động vị trí, và sẽ được biết và được hiểu bởi người có trình độ trung bình trong lĩnh vực này.

Việc sử dụng LA3576 chứng tỏ rằng sự biểu hiện tTAV2 trong LA3604 là đặc hiệu con cái, xuất hiện chủ yếu trong các cơ bay gián tiếp và là đặc hiệu giai đoạn. Sau đó, nhiều cấu trúc yếu tố tác động tetO khác nhau được xây đựng để

phân tích các tác động của chúng. Gen chuyển tetO-MichelobX (LA3582, xem Hình. 15 và SEQ ID NO. 144) nếu được lai với LA3545, đều thể hiện các kiểu hình đặc hiệu con cái không bay được mà có thể được ức chế bởi tetraxyclin. Điều này chứng tỏ là Actin4 có thể được sử dụng để điều khiểu gen tác động theo cách đặc hiệu giai đoạn, mô và giới tính.

Do một vài dòng LA3545 có kiểu hình đặc hiệu con cái không bay được mà không có gen tác động, điều này chỉ ra là tTAV2 có thể đóng vai trò làm một phân tử tác động. tTAV2 gồm có tTA, miền liên kết tetO và VP 16, một protein virut phức hợp herpes. Sự phiên mã hoạt hóa VP 16 của các gen virut ngay từ sớm bằng cách sử dụng các trình tự đầu tận cùng amino của nó để gắn vào một hoặc nhiều protein được mã hoá thuộc vật chủ mà nhận các trình tự ADN trong các trình tự khởi đầu của chúng. Trong LA3604, gen tác động tetO-VP16 đã được thêm vào để tăng cường tác động của tTAV2. Trong ba dòng gen chuyển của LA3604, điều này dẫn đến 100% kiểu hình con cái đặc hiệu không bay được nếu được nuôi không có tetraxyclin, cho thấy rằng VP 16 là một phân tử tác động có hiệu quả. Chú ý rằng LA3604 condon khởi đầu tiềm năng (ATG) có đầu 5' nối với intron được cắt nối thay thế. Vì vậy, trong cấu trúc này, exon con đực đặc hiệu được mong đợi để gián gián đoạn việc mã hóa khung đọc mở tTAV (ubi-tTAV); vì trình tự đặc hiệu con đực chứa codon kết thúc khác, điều này sẽ làm giảm hoặc loại trừ sự sản sinh tTAV chức năng trong các con đực. Bằng cách so sánh, exon đặc hiệu con đực là từ 5' đến condon khởi đầu của tTAV trong LA3545. Tuy nhiên, bằng cách cài xen một số condon khởi đầu 5' vào condon khởi đầu của tTAV (mà là ATG thứ nhất của sản phẩm phiên mã con cái nhưng không phải của sản phẩm phiên mã con đực), không có condon nào trong số các condon khởi đầu thêm vào này là thích hợp cho việc sản xuất hữu hiệu tTAV chức năng do nó không thuộc khung hoặc cócodon kết thúc, bố trí như vậy cũng sẽ làm giảm hoặc hoặc loại trừ việc sản sinh tTAV chức năng trong các con đực, phù hợp với các dữ liệu kiểu hình nêu trên .

Ví dụ 16: Sử dụng vị trí ubiquitin và intron

Gần đây, các tác giả đã tạo ra các cấu trúc dựa theo *Cctra* với intron catxet CcTRA ở nhiều phạm vi khác nhau, được kẹp sườn bởi các trình tự khác nhau chẳng hạn. Các dòng khác nhau của Medfly chuyển gen mang các cấu trúc này được xây dựng. Điều này cho thấy hệ này là thông thường và không tinh vi, mà nó sẽ hoạt động với khoảng rộng của các trình tự khác loại quan tâm chẳng hạn.

Gần đây, các tác giả cũng đề xuất ít nhất một ví dụ về thể dung hợp Cctra-ubi-tTAV tạo ra việc việc cắt nối chính xác (DsRed-cctra-ubi-tTAV).

Các ví dụ được ưu tiên của protein chức năng, đặt trình tự mã hoá cho ubiquitin hoặc tTA, hoặc các thể đột biến chức năng của chúng và hoặc các biến thể như tTAV, tTAV2 hoặc tTAV3, 3' với intron. Các biến thể này được sắp xếp để các yếu tố này gần như liền kề với điểm cuối 3' của intron, tốt hơn là, nếu vùng mã hoá bắt đầu trong 20 nucleotit hoặc ít hơn ranh giới intron 3'), và tốt nhất là liền kề ngay lập tức điểm cuối 3' của intron, mặc dù điều này là ít liên quan nếu hệ Ubiquitin được sử dụng.

Các ví dụ được ưu tiên về các cấu trúc mã hoá theo sáng chế được ghi trong Bảng 4, dưới đây. Cần hiểu rõ rằng LAl 188 không thuộc phạm vi của sáng chế, do nó không mã hoá protein chức năng, nó không hoạt động một cách thích hợp chẳng hạn. Điều này cần được cho là do việc sử dụng không mong đợi thể cho cắt nối 4 bp 5' đối với mối nối với trình tự intron CcTRA, dẫn đến sự dịch chuyển khung được gây ra ở tất cả các trường hợp cắt nối. Vì vậy, các ví dụ này chỉ nhằm mục đích cung cấp thông tin.

Bảng 4

Cấu trúc No. (Hình #)	Intron tra từ các loài	Vị trí từ ATG (bp)	Intron tra được dung hợp với
LA1188 (80)	Medfly	+132	tTAV
LA3014 (29)	Medfly	+22	ubiquitin
LA3166 (30)	Medfly	+136	ubiquitin

Bảng 4 (tiếp ~~theo~~)

Cấu trúc No. (Hình #)	Intron tra từ các loài	Vị trí từ ATG (bp)	Intron tra được dung hợp với
LA3097 (27)	Medfly	+0	tTAV
LA3077 (26)	Medfly	+61	tTAV
LA3233 (28)	Medfly	+0	tTAV2
LA3376 (31)	Medfly	+0	tTAV2
LA3376 (31)	B. zonata	+3	reaper KR
LA3376 (31)	B.zonata	+0	tTAV3
LA3242 (32)	C. rosa	+3	reaperKR
LA1038 (14)	Medfly	+21	Nipp1 (nipper)
LA3054 (61)	Medfly	+811	DsRed-ubiquitin
LA3056 (62)	Medfly	+811	DsRed-ubiquitin
LA3488 (63)	Medfly	+949	Ubiquitin
LA3596 (67)	Medfly	+949	Ubiquitin

Bảng 4 thể hiện cấu trúc mà chứa trình tự kiểm soát cắt nối mà được bắt nguồn từ tra intron. Các intron được bắt nguồn từ C. capitata (Medfly), B. zonata hoặc C. rosa (see column 2). Các intron nêu trên được lồng vào trong vùng mã hoá sao cho khoảng cách giữa chất ức chế giả định ATG và nucleotit cuối cùng của exon ở ngay trước tra intron là như được chỉ ra trong cột 3. Intron được lồng vào trong hoặc liền kề với vùng mã hoá với ubiquitin, tTAV, reaper^{KR}, nipper hoặc ubiquitin-DsRed như được thể hiện trong cột 4. Các intron được sinh ra và được thể hiện là việc cắt nối thành công, bằng RT-PCR hoặc nhờ kiểu hình trong Medfly và, trong một vài trường hợp, cũng có trong Drosophila melanogaster (LA3077) hoặc Anastrepha ludens (LA3097, LA3233, LA3376). Ngoài ra, khoảng cách giữa ATG và điểm cuối của exon ở ngay trước tra intron (đảm bảo việc cắt nối trong dạng giống như F1) có thể nằm trong khoảng từ 0bp đến ít nhất +949bp mà không gây ra hậu quả bất lợi cho việc cắt nối (xem Bảng 4, cột 3). Vì vậy, hợp lý nếu đảm bảo rằng khoảng cách này có thể đạt đến tối đa ít nhất là 900 và tốt hơn là tối đa ít nhất 949 bp.

Thông tin khác về các ví dụ này được tóm tắt trong Bảng 5. Ưu tiên sử dụng trình tự không nội sinh để thu được sự kiểm soát việc cắt nối chính xác về sự biểu hiện (+0bp trong bảng 4). Các tác giả ưu tiên cài xen tra intron giữa các

dinucleotit liền kề TG... GT trong vùng mã hoá của protein lồng vào để được cắt nối thay thế nhằm đảm bảo việc cắt nối chính xác do việc này có thể là quan trọng, tuy nhiên, các tác giả sẽ không giới hạn ở phương án này nếu cần do các nucleotit kẹp sườn có thể cũng có chức năng chính xác. Các ví dụ LA1038, LA3054 và LA3056 bao gồm một vài trình tự exon kẹp sườn nội sinh từ gen Cctra tự nhiên. Trong bảng 5, nếu 6 nucleotit hoặc ít hơn (có đơn vị mã hoá bắt đầu ATG) có trong các thể dung hợp cụ thể với 3' hoặc 5' của mối nối cắt nối này, để ngắn gọn trongbảng này, các nucleotit này sẽ không được xem như một phần của thể dung hợp này. Bảng 4 có thể được liên hệ với bảng 3 để tìm intron tra nào (Cctra, Bztra hoặc Crtra) được sử dụng trong mỗi ví dụ. Hơn nữa, LAl 188 được bao hàm chỉ nhằm mục đích thông tin và nằm ngoài phạm vi của sáng chế.

Bảng 5

Cấu trúc No. (Hình #)	Intron tra được dung hợp với đầu 5'	Intron tra được dung hợp với đầu 3'	Trình tự exon tra được dung hợp với đầu 5' (bp)	Trình tự exon tra được dung hợp với đầu 3' (bp)
LA1188 (80)	Hsp70-tTAV	tTAV	+0bp	+0bp
LA3014 (29)	Hsp70-ubiqutin	ubiqutin-reaperKR-sv40	+0bp	+0bp
LA3166 (30)	Hsp70-ubiqutin	ubiqutin-reaperKR-sv40	+0bp	+0bp
LA3097 (27)	Hsp70	tTAV-K10	+0bp	+0bp

Bảng 5 (tiếp ~~theo~~)

Cấu trúc No. (Hình #)	Intron tra được dung hợp với đầu 5'	Intron tra được dung hợp với đầu 3'	Trình tự exon tra được dung hợp với đầu 5' (bp)	Trình tự exon tra được dung hợp với đầu 3' (bp)
LA3077 (26)	Hsp70-tTAV	tTAV-K10	+0bp	+0bp
LA3233 (28)	Hsp70	tTAV2-K10	+0bp	+0bp
LA3376 (31)	Hsp70	tTAV2-K10	+0bp	+0bp
LA3376 (31)	Sry-a	tTAV3-sv40	+0bp	+0bp
LA3376 (31)	HB	reaperKR-sv40	+0bp	+0bp
LA3242 (32)	HB	reaperKR-sv40	+0bp	+0bp
LA1038 (14)	Hsp70-tra	Tra-NippI (nipper)-sv40	+22bp	+20bp
LA3054 (61)	Opie2-nls-DsRed-tra	tra-ubiquitin-tTAV-sv40	+22bp	+20bp
LA3056 (62)	Opie2-nls-DsRed-tra	tra-ubiquitin-tTAV-sv40	+22bp	+242bp
LA3488 (63)	Iel-nls-TurboGreen-nls-ubiquitin	ubiquitin-nls-DsRed-nls-sv40	+0bp	+0bp
LA3596 (67)	Iel-nls-TurboGreen-nls-ubiquitin	ubiquitin-nls-DsRed-nls-sv40	+0bp	+0bp

Như được đề cập ở trên, nếu intron nằm ở vị trí 5' đối với với một vùng mã hoá protein (ORF-X), tốt hơn là sử dụng ubiquitin 3' với intron, 5' với ORP-X, nhờ đó tạo ra sự điều biến đặc hiệu con cái của ORF-X, mặc dù dẫn đến sự tách vật lý giữa trình tự này và intron tra, vì vậy làm giảm khả năng các trình tự thuộc ORF-X sẽ lồng vào khi cắt nối intron tra.

Các cấu trúc ghép và các trình tự ghép cũng được dự tính, ví dụ có dạng:

X-ubi-Y

với intron được cắt nối thay thế được lồng vào giữa vùng mã hoá X và vùng mã hoá ubiquitin (ubi), hoặc bên trong vùng mã hóa ubiquitin, hoặc giữa vùng mã hoá ubiquitin và vùng mã hoá Y. Vì vậy, X sẽ được biểu hiện bất kể có việc cắt nối intron hay không, trong khi Y sẽ chỉ được biểu hiện nếu intron được tách ở dạng thích hợp. Cấu hình và cách bố trí khác của kiểu thông thường sẽ là hiển nhiên với người có trình độ trung bình trong lĩnh vực kỹ thuật này. Một số

ví dụ về cấu hình như vậy là LA3014, LA3054, LA3056, LA3166, LA3488 và LA3596 mà đều sử dụng thể dung hợp ubiquitin theo cách này, chứng tỏ khả năng ý tưởng này được ứng dụng thành công trong Medfly chuyển gen. Các ví dụ khác ở muỗi chuyển gen gồm có LA3604 và LA3612, việc thể hiện khả năng ứng dụng sinh chủng tự nhiên của hệ này không chỉ trong các loài khác nhau (muỗi và Medfly), mà còn ở các phạm vi khác nhau gồm AaActin4, Aadsx và Cctra.

LA3596 (xem Hình. 67 và SEQ ID NO. 145) là một kiểu tương tự với LA3488, được dự định để sinh huỳnh quang màu xanh ở cả hai giới tính (bằng sự biểu hiện nhân được khoanh vùng protein huỳnh quang TurboGreen), nhưng chỉ sinh huỳnh quang màu đỏ ở con cái (bằng sự biểu hiện hạt nhân được khoanh vùng protein hùynh quang DsRed2). Điều này được thực hiện bởi sự hợp nhất hai loại protein này, được điều khiển bằng trình tự tăng cường/catxet khởi đầu Hr5-Iel, được liên kết với một đoạn liên kết 11 amino ngắn (đoạn liên kết SG4) và một vùng mã hoá gồm ubiquitin (với một đột biến điểm được dự định để làm ổn định protein tạo thành bằng cách làm giảm xu hướng phân hủy qua trung gian ubiquitincủa nó) và intron CcTRA để giới hạn sự biểu hiện DsRed2 ở các con cái. Medfly chuyển gen được tạo ra với cấu trúc này. Sự phát huỳnh quang màu đỏ bị giới hạn ở các con cái trong dòng này như được mong đợi, trong khi sự phát huỳnh quang màu xanh được quan sát ở cả con đực và con cái. Điều này có thể được sử dụng để phân chia giới tính bằng sàng lọc huỳnh quang đối với protein phát huỳnh quang cụ thể, trong trường hợp này sự phát huỳnh quang màu đỏ cho thấy sự biểu hiện của DsRed2.

Ví dụ 17: Ví dụ minh họa khác về Cctra

Sự tham chiếu cũng được tạo ra đối với LA3014 và LA3166 và dữ liệu kiểu hình từ đó trong các ví dụ khác.

Trước đây, các tác giả đã tạo ra, và thu được các gen chuyển với intron CcTRA trong protein chức năng khác ngoài tTAV, xem LA3014 và LA3166. LA3014 chứa một thể dung hợp ubiquitin-reaper^{KR} nằm xuôi chiều với intron

CcTRA. Số liệu kiểu hình cho thấy Medfly chuyển gen LA3014 tạo ra khả năng gây chết con cái có thể ức chế được. Phân tích RT-PCR trên ARN được chiết từ các con đực và các con cái con trưởng thành được nuôi không có tetraxyclin, sử dụng các đoạn mồi và ReaperKR, chứng tỏ rằng việc cắt nối chính xác xuất hiện ở các con cái (dải 508bp) và không có dải như vậy được tìm thấy ở các con đực (Hình. 37). LA3166 là một cấu trúc khác với intron CcTRA được đặt bên trong vùng mã hoá ubiquitin được dung hợp với reaper^{KR}, nhưng được đặt ở một vị trí khác trong ubiquitin. LA3166 cũng tạo ra một hiệu quả gây chết đặc hiệu con cái có thể ức chế được có tính trội trong Medfly.

LA1038 là một ví dụ mới về việc sử dụng intron CcTRA trong một phạm vi trình tự khác, ở đây được đặt trong một đoạn gồm NipplDm được gọi là 'nipper' mà cũng cắt nối chính xác trong Medfly chuyển gen nếu được phân tích bằng RT-PCR (Hình. 12). LA670 được quy định như một nguồn tTAV để điều kiển sự biểu hiện của nipper cắt nối thay thế.

Gần đây, các tác giả cũng tạo ra, và thu được các gen chuyển với, các cấu trúc 'chỉ chứa intron' dựa trên Cctra với intron trong một gen khác (nhiều ví dụ trong các ví dụ nêu trên, trừ khi được chỉ rõ theo cách khác, có trong tTAV hoặc một trong số các biến thể của nó, tTAV2 hoặc tTAV3 chẳng hạn). Các cấu trúc này hoạt động như được dự báo. Đây là kết quả quan trọng, vì vậy thể hiện rằng không có các trình tự exon thiết yếu trong Cctra mà các tác giả đã tình cờ sao lại một cách đơn giản (về mặt chức năng, nếu không nhất thiết phải về mặt trình tự) trong tTAV. Các tác giả cũng tạo ra các cấu trúc ubi-rpr^{KR} thuộc loại này (LA3014 và LA3166), mà cũng phù hợp, phương pháp dung hợp ubiquitin được mô tả ở trên. Phương pháp dung hợp ubiquitin còn được minh họa kỹ hơn bằng phương pháp phân tích RT-PCR đối với LA3054, LA3056 và LA3488 (Hình 11, 13, 14), như được mô tả trong ví dụ 16, nêu trên.

Tài liệu tham khảo

Allen ML, Christensen BM. Related 2004 Flight muscle-specific expression of act88F: GFP in transgenic Culex quinquefasciatus Say (Diptera: Culicidae). Parasitol Int. 55 (4): 307- 14. Bennett D_5 Szoor B, Gross S, VereshchaginaN, Alphey L. 2003 Ectopic expression of inhibitors of protein phosphatase type 1 (PPl) can be used to analyze roles of PPl in Drosophila development. Genetics. 164(1) :235-45.

Black, D. (2003). Mechanisms of alteRNAtive pre-messenger ARN splicing. Annu Rev Biochem 72, 291-336.

Burset, M., Seledtsov, L, and Solovyev, V. (2001). SpliceDB: database of canonical and non- canonical splice sites in mammalian bộ di truyềǹs. Nucleic Acids Research $2P_5$ 255-259. Caceres JF, Kornblihtt AR. 2002 AlteRNAtive splicing: multiple control mechanisms and involvement in human disease. Trends Genet. 18(4): 186-93.

Cande C_5 Cecconi F, Dessen P_5 Kroemer G. 2002 Apoptosis-inducing factor (AIF): key to the conserved caspase-independent pathways of cell death? J Cell Sci. 115(24):4727-34.

Cartegni, L., Chew, S., and Krainer, A. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nature Reviews Genetics 3, 285-298. Clark, F., and Thanaraj, T. (2002). Categorization and characterization of transcript-confirmed constitutively and alteRNAtively spliced introns and exons from human. Human Molecular Genetics 11, 451-464.

Funaguma, S., Suzuki, M., Tamura, T., and Shimada, T. (2005). The Bmdsx transgene including trimmed introns is sex-specifically spliced in tissues of the silkworm, Bombyx mori. J Insect Sci 5, 17.

George, EX., Ober, M.B. and Emerson Jr, CP. (1989). Functional domains of the Drosophila melanogaster muscle myosin heavy-chain gene are encoded by alteRNAtively spliced exons. MoL Cell Biol. P:2957-2974.

Graveley BR. 2001 AlteRNAtive splicing: increasing diversity in the proteomic world. Trends Genet. 17(2): 100-7.

Hammes, A.$_5$ Guo, J.K., Lutsch, G., Leheste, J.R., Landrock, D., Zeigler, U., Gubler, M.C. and Schedl, A. (2001). Two splice variants of the Wilms' Tumour 1 gene have distinct functions during sex determination and nephron formation. Cell 106:319-329.

Hastings, G.A. and Emerson Jr$_5$ CP (1991). Myosin functional domains encoded by alteRNAtive exons are expressed in specific thoracic muscles of Drosophila. J. Cell Biol. 114: 263-276.

Hedley, MX. and Maniatis (1991). Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to a tra-2 protein in vivo. Cell 65:579-586. Heinrich J. C. and Scott MJ. 2000 A repressible female-specific lethal genetic system for making transgenic insect strains suitable for a sterile-release program PNAS 97 (15): 8229-8232 Horn C, Wimmer EA. 2003 A transgene-based, embryo-specific lethality system for insect pest management. Nat Biotechnol. 2i(%): 64-70.

Hoshijima, K.K, Inoue, L., Higuchi, L, Sakamoto, H. and Shimura, Y. (1991). Control of doublesex alteRNAtive splicing by transformer and transformer-2 in Drosophila. Science 252:833-836.

Huang, Q., Deveraux, Q.L., Maeda, S., Salvesen, G.S., Stennicke, H.R., Hammock, B.D. and Reed, J.C. (2002). Evolutionary conservation of apoptosis mechanisms: Lepidopteran and baculoviral inhibitor of apoptosis proteins are inhibitor of mammalian caspase- 9.Agricultural Sciences 97(4): 1427-1432.

Ito, Y., Hirochicka, H. and Kurata, N. (2002). Organ-specific alteRNAtive transcripts of KNOX family class 2 homeobox genes of rice. Gene 288:41-47.

Johnson JM, Castle J_5 Garrett-Engele P, Kan Z_5 Loerch PM_5 Armour CD_5 Santos R_5 Schadt EE_5 Stoughton R, Shoemaker DD. 2003 Bộ di truyền-wide survey of human alteRNAtive pre-mRNA splicing with exon junction microarrays. Science. 302(5653):2141-4.

Jurica MS, Moore MJ. 2003 Pre-mRNA splicing: awash in a sea of proteins. MoI Cell. 12(1):5- 14.

Kazzaz JA_5 Rozek CE. 1989 Tissue-specific expression of the alteRNAtely processed Drosophila myosin heavy-chain messenger RNAs. Dev Biol. 133 (2): 550-61.

Maniatis, T., and Tasic, B. (2002). AlteRNAtive pre-mRNA splicing and proteome expansion in metazoans. Nature 418, 236-243.

Munoz, D., Jimenez, A., Marinotti, O., and James, A. (2004). Yhe.AeAct-4 gene is expressed in the developing flight muscles of females Aedes aegypti. Insect Molecular Biology 13, 563-568. Nishiyama, R., Mizuno, H., Okada, S., Yamaguchi, T., Takenaka, M., Fukuzawa, H. and Ohyama, K. (1999). Two mRNA species encoding calcium-dependent protein kinases are differentially expressed in sexual organs of Marchantia polymorpha through alteRNAtive splicing. Plant Cell Physiol.40(2J:205-212.

Nishiyama, R.,Yamato, K.T., Miura, K., Sakida, M., Okada, $S._5$ Kono, K., Takahama, M., Sone, T., Takenaka, M., Fukuzawa, H. and Ohyama, K. (2000). Comparison of expressed sequence tags from male and female sexual organs of Marchantia polymorpha. ADN Res. 7:165-174. Olson, M.R., Holley, C.L., Ji Yoo, S., Huh, J.R, Hay, B.A. and Kornbluth, S. (2003). Reaper is regulated by IAP-mediated Ubiquitination. J.Biol.Chem., 27Sf<5;:4028-4034.

Olson, M.R., Holley, CX., Gan, E.C., Colon-Ramos, D.A., Kaplan, B. and Kornbluth, S. (2003). A GH3-like domain in reaper is required for mitochondrial localization and induction of IAP degradation. J. Biol. Chem. 218(45). -44758-44768.

Pan, Q., Shai, O., Misquitta, C, Zhang, W., Saltzman, A., Mohammad, N., Babak, T., Siu, H., Hughes, T., Morris, Q., et al. (2004). Revealing global regulatory features of mammalian alteRNAtive splicing using a quantitative microarray platform. MoI Cell 16, 929-941. Pane, A., Salvemini, M., Delli Bovi, P., Polito, C, and Saccone, G. (2002). The transformer gene in Ceratitis capitata provides a genetic basis for selecting and remembering the sexual fate. Development 129, 3715-3725.

Park, J., Parisky, K., Celotto, A., Reenan, R., and Graveley, B. (2004). Identification of alteRNAtive splicing regulators by ARN interference in Drosophila. Proc Nat'l Acad Sci (USA) 101, 15974-15979.

Parker L, Gross S, Beullens M, Bollen M, Bennett D, Alphey L. 2002 Functional interaction between nuclear inhibitor of protein phosphatase type 1 (NIPPl) and protein phosphatase type 1 (PPl) in Drosophila: consequences of over-expression of NIPPl in flies and suppression by co- expression of PPl. Biochem J. 368 (3) :1 '89-97.

Raphael, K.A., Whyard, S., Shearman, D., An, X. and Frommer, M. (2004). Bactrocera tyroni and closely related pest-tephritids-molecular analysis and prospects for transgenic control strategies. Insect Biochem. MoI. Biol. 34:167-176.

Ryner, L. and Baker, B. S. (1991). Regulation of doublesex pre-mRNA processing occurs by 3'- splice site activation. Genes Dev. 5:2071-2085.

Saccone, G., Pane, A., and Polito, C. (2002). Sex determination in flies, fruitfles and butterflies. Genetica 116, 15-23.

Scali, C, Catteruccia, F., Li, Q., and Crisanti, A. (2005). Identification of sex-specific transcripts of the Anopheles gambiae doublesex gene. J Exp Biol 208, 3701-3709. Scott, M., Heinrich, J., and Li, X. (2004). Progress towards the development of a transgenic strain of the Australian sheep blowfly (Lucilia cuprina) suitable for a male-only sterile release program. Insect Biochem MoI Biol 34, 185-192.

Seo, S-J., Cheon, H-M., Sun, J., Sappington, T. W. and Raikhel, A.S. (2003). Tissue- and stage- specific expression of two lipophorin receptor variants with seven and eight ligand-binding repeats in the adult muõi. J. Biol. Chem. 275(¥5j:41954-41962.

Siebel CW, Fresco LD, Rio DC. 1992 The mechanism of somatic inhibition of Drosophila P- element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5' splice site control Ul snRNP binding. Genes Dev. (5^:1386-401.

Shivikrupa, Singh., R and Swarup, G. (1999). Identification of a novel splice variant of C3G which shows tissue-specific expression. ADN Cell Biol. /5: 701-708.

Smith, C, and Valcarcel, J. (2000). AlteRNAtive pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci 25, 381-388.

Stoss, O.$_5$ Stoilov, P., Hartmann, A. M., Nayler, O., and Stamm, S. (1999). The in vivo minigene approach to analyze tissue-specific splicing. Brain Research Protocols 4, 383-394.

Stoss, O., Olbrich, M, Hartmann, A. M., Konig, H., Memmott, J., Andreadis, A and Stamm, S.

(2001). The STAR/GSG family protein rSLM-2 regulates the selection of alteRNAtive splice sites. J. Biol. Chem. 276(12):8665-8673.

Streuli, M. and Saito, H. (1989). Regulation of tissue-specific alteRNAtive splicing: exon-specific cis-elements govern the splicing of leukocyte common antigen pre-mRNA. EMBO J. 8(3): 787-796.

Suzuki, M., Ohbayashi, F., Mita, K., and Shimada, T. (2001). The mechanism of sex-specific splicing at the doublesex gene is different between Drosophila melanogaster and Bombyx mori. Insect Biochem MoI Biol 31, 1201-1211.

Thanaraj, T., and Clark, F. (2001). Human GC-AG alteRNAtive intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Research 29, 2581-2593.

Thanaraj, T., Stamm, S., Clark, F., Reithoven, J., Le Texier, V., and Muilu, J. (2004). ASD: the AlteRNAtive Splicing Database. Nucleic Acids Research 32, D64-D69.

Varshavsky, A. (2000). Ubiquitin fusion technique and its descendants. Meth Enz 327.

Venables, J. (2002). AlteRNAtive splicing in the testes. Curr Opin Genet Dev 12, 615-619.

Venables JP. 2004 Aberrant and alteRNAtive splicing in cancer. Cancer Res. 64(21): 7647-54.

Vernooy, S.Y., Copeland, J., Ghaboosi, N., Griffin, E.E., Yoo, SJ. and Hay, BA. (2000). J. Cell Biol. 150(2):F69-¥75.

White, K., Tahoaglu, E. and Steller, H. (1996). Cell killing by the Drosophila gene reaper. Science 271 (5250): 805-807.

Wing, J.P., Zhou, L., Schwartz, L.M. and Nambu, J.R. (2001) Distinct cell killing properties of the Drosophila reaper, head involution defective, and grim genes. Cell Death Diffn 5(11): 930-939.

YaIi Chiu A., and Pin Ouyang, A.B.,(2006).Loss of Pnn expression attenuates expression levels of SR family splicing factors and modulates alteRNAtive pre-mRNA splicing in vivo. Bioch. Biophys.Res. Comm.iW:663-671.

Yoshimura, K., Yabuta, Y., Ishikawa, T. and Shigeoka, S. (2002). Idenitification of a cis element for tissue-specific alteRNAtive splicing of chloroplast Ascorbate Peroxidase pre-mENA in higher plants. J. Biol.Chem 277 (43,1:40623-40632.

Giải thích trình tự

<223> trình tự của pLA3359 (SED ID NO. 47). <***> jζey bao gồm:

1. Anopheles gambiae dsx (Agdsx) mini-gen, [mini-gen là trình tự tái tổ hợp bắt nguồn từ một gen nói riêng (ví dụ gen Agdsx) bằng cách nối các đoạn độc lập với nhau trong khi giữ lại các trình tự 5 '-3' ban đầu; đây là gen tương đương để loại bỏ một số đoạn bên trong từ đoạn dài hơn trong hệ gen xuất phát từ gen đó], (1-3135): bao gồm Agdsx phần exon3, exon 4a (con cái), exon 4b (con cái) và phần exon 5 (cả con đực và con cái).

<***> các exon xuất phát từ Agdsx tại các vị trí 426 đến 560 (phần exon 3); 1068 đến 2755 (bao gồm phần exon 4, được tìm thấy ở con cái); 1809 đến 2755 (bao gồm phần exon 4, được tìm thấy ở con cái); và 2914 đến 3135 (bao gồm phần exon 5, được tìm thấy ở con đực).

<***> Cắt nối thay thế phiên mã bắt đầu trong đoạn bắt nguồn từ baculovirut AcMNPV IeI (immediate early 1) tại vị trí -8031 (Đoạn IeI là từ vị trí 7431 đến 8060).

<***> Các điểm đặc trưng bao gồm:

1. Intron thêm vào bắt nguồn từ đoạn gen Drosophila ('các đoạn intron') ngược với trình tự Agdsx từ vị trí 8075 đến 8137.

<223> Trình tự pLA3433 (SED ID NO. 48). <***> điểm đặc biệt chủ yếu bao gồm:

1. Mini-gen Agdsx (778-4623): bao gồm Agdsx phần exon 2, exon 3, exon 4a (con cái), exon 4b (con cái) và phần exon 5 (con đực và con cái).

<***> Exons bắt nguồn từ Agdsx từ vị trí 778 to 908 (phần exon 2); 1913 đến 2048 (phần exon 3); 2556 đến 2642 (phần exon 4a); 3297 đến 4243 (phần exon 4b) và 4402 đến 4623 (phần exon 5).

<***> Cắt nối thay thế phiên mã bắt đầu trong đoạn bắt nguồn từ baculovirut AcMNPV IeI (immediate early 1) tại vị trí -606 (Đoạn IeI từ vị trí 6 đến 635).

<***> Điểm đặc trưng bao gồm:

1. intron thêm vào bắt nguồn từ đoạn gen Drosophila ('các đoạn intron') ngược với trình tự Agdsx từ vị trí 650 đến 712.

<223> Trình tự pLA3491. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Aedes aegypti dsx (Aadsx) mini-gen: bao gồm phần Aadsx exon 4, exon5a (con cái), exon 5b (con cái), và phần exonδ (con đực và con cái).

<***> Exons bắt nguồn từ Aadsx từ vị trí 1316 to 1450 (phần exon 4); 2626 đến 3761 (phần exon 5a); 3293 đến 3761 (phần exon 5b); và 5215 đến 5704 (phần exon 6). <***> Phần phiên mã Fl được dự đoán bao gồm các nucleotit ~1174-1450, 2626-3761, 5215—5850.

<***> Phần phiên mã F2 được dự đoán bao gồm các nucleotit ~1174-1450, 3293-3761, 5215—5850.

<***> Phần phiên mã F3 được dự đoán bao gồm các nucleotit ~1174-1450, 2626-3083, 3293-3761, 5215-5850.

<***> Phần phiên mã Ml được dự đoán bao gồm các nucleotit ~1174-1450, 5215—5850. <***> Cắt nối thay thế phiên mã bắt đầu trong đoạn bắt nguồn từ baculo virut AcMNPV IeI (immediate early 1) tại vị trí ~1174 (đoạn IeI từ vị trí 574 đến 1203).

<***> Điểm đặc trưng bao gồm:

1. intron thêm vào bắt nguồn từ đoạn gen Drosophila ('các đoạn intron') ngược dòng với trình tự Aadsx từ vị trí 1218 đến 1280.

<223> Trình tự pLA3646. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Mini-gen Aadsx (17218-11707): bao gồm phần Aadsx exon 4 từ vị trí 17113 đến 16979, exon 5a từ vị trí 15803 đến 15025 + 14010 đến 13650, exon 5b từ vị trí 15136 đến 15025 + 14010 đến 13650 và exon 6 từ vị trí 12196 đến 11707 (chú ý: sự định hướng ngược). <***> phần exon 4 chứa 4 điểm đột biến liên quan đến kiểu tự nhiên tại các vị trí 17087 (ATG- ACG), 17053 (ATG-ACG),

17050 (ATG-ACG) và 17041 (ATG-ACG) (chú ý: sự định hướng ngược); phần exon 5a và 5b chứa 3 điểm đột biến liên quan đến kiểu tự nhiên tại các vị trí 15129 (ATG-ATA), 15116 (ATG-ATA) và 15113 (ATG-ATA) (chú ý: sự định hướng ngược). Tất cả các đột biến được loại ra trình tự ATG.

<***> tTAV2 được lồng vào các exon 5a và 5b từ vị trí 15024 đến 14011 (chú ý: sự định hướng ngược).

<***> Cắt nối thay thế phiên mã bắt đầu bắt nguồn từ đoạn hsp70 tại vị trí ~17312 (đoạn hsp70 từ vị trí 17354 đến 17225); (chú ý: sự định hướng ngược).

<***> Điểm đặc trưng bao gồm:

1. Intron thêm vào bắt nguồn từ đoạn gen Drosophila ('các đoạn intron') ngược dòng với trình tự Aadsx từ vị trí 1107 to 1045 (chú ý: sự định hướng ngược)

Trình tự pLA3435 (SED ID NO. 46). <223> Các điểm đặc trưng quan trọng bao gồm:

1. Bombyx mori dsx (Bmdsx) minigene (1411-3161) bằng một cầu nối ngoại sinh giữa exon 3 và 4 hòa lẫn của con cái.

<***> đoạn chia đều exon 2(1411bp-1554bp)

<***> Phần đặc trưng của con cái exon 3 (2121bp-2202) hòa lẫn với phần exon 4 (2225bp-2290bp) bằng một cầu nối ngoại sinh (2203bp-2224bp) <* * *> đoạn chia đều exon 5 (3007bp-3161 bp)

<***> con cái ^{8x} sản phẩm cắt nối mini-gen được mã hóa bởi 1411-1554 + 2121-2290 +3007- 3161.

<***> con đực dsx sản phẩm cắt nối mini-gen được mã hóa bởi 1411-1554 +3007-3161. <***> sự phiên mã được dự đoán bắt đầu tại vị trí khoảng-1239 trong đó đoạn này bắt nguồn từ baculovirut AcMNPV IeI (immediate early 1) trình tự khởi đầu (639bp-1268bp).

<223> Trình tự pLA3534. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Mini-gen Aadsx (6996-4425): chứa Aadsx exon 4, phần exon 5a (con cái) và phần exon 5b (con cái), bao gồm các đoạn intron Aadsx.

<***> các exon bắt nguồn từ Aadsx từ vị trí 6968 đến 6834 (phần exon 4), 5462 đến 4425 (phần exon 5a) và 4795 đến 4425 (phần exon 5b); (chú ý sự định hướng ngược). <***> Phần phiên mã Fl được dự đoán bao gồm các nucleotit -7146-6834, 5462—4300 (chú ý: sự định hướng ngược).

<***> Phần phiên mã F2 được dự đoán bao gồm các nucleotit -7146-6834, 4795—4300 (chú ý: sự định hướng ngược).

<***> Phần phiên mã F3 được dự đoán bao gồm các nucleotit -7146-6834, 5462-5005, 4795—4300 (chú ý: sự định hướng ngược).

<***> Cắt nối thay thế phiên mã bắt đầu trong đoạn bắt nguồn từ baculovirut AcMNPV IeI

(immediate early 1) tại vị trí ~7146 (Đoạn IeI là từ vị trí 7746 đến 7117, sự định hướng ngược).

<223> Trình tự pLA3612. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Ubiquitin-vùng mã hóa tTAV2 được lồng vào exon của gen Aadsx ở con cái. <***> Ubiquitin-tTAV2 từ vị trí 15185-16429 trong Aadsx (ubiquitin là từ 15185- 15412; tTAV2 là từ 15413-16429), bao gồm codon khởi đầu và codon kết thúc. <***> trình tự bắt nguồn từ Aadsx: 13150-15184, 16438-18805.

<***> Aadsx-ubiquitin-tTAV2 cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 (đoạn hsp7Q là từ 13014-13143).

<223> Trình tự pLA3619. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Vùng mã hóa tTAV2 được lồng vào exon của gen Aadsx ở con cái. <***> Trình tự bắt nguồn từ Aadsx: 5635-364I_5 2610-243 (chú ý: sự định hướng ngược). <***> Aadsx-tTAV2 cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 từ 5642 đến 5771 (chú ý: sự định hướng ngược).

<***> tTAV2 được dự đoán dịch mã giữa 2619-3635, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<223> Trình tự pLA3545. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Vùng mã hóa AaActin4 và 5' UTR bao gồm intron đầu tiên điều chỉnh mức biểu hiện tTAV. <***> Trình tự bắt nguồn từ AaActin4 từ vị trí 923-4285. <***> Cắt nối thay thế phiên mã được dự đoán bắt đầu từ khoảng vị trí -2366. <***>intron đầu tiên từ AaActin4 (biến thể cắt nối ở con cái) là từ 2458-4259. <* * *> tTAV được dự đoán dịch mã giữa 4300-5316, bao gồm codon khởi đầu và codon kết thúc.

<223> Trình tự pLA3604. <***> Các điểm đặc trưng quan trọng bao gồm:

1. Trình tự khởi đầu AaActin4 và 5' UTR điều hòa mức biểu hiện ubiquitin-tTAV2. <***> Trình tự bắt nguồn từ AaActin4 từ vị trí 5795-2407 (chú ý: sự định hướng ngược). <***> Cắt nối thay thế phiên mã được dự đoán bắt đầu từ khoảng vị trí ~4353 (chú ý: sự định hướng ngược).

<***> Tj_{16} fust intron from AaActin4 (con cái cắt nối variant) là từ 2455-4254 (chú ý: sự định hướng ngược).

<***> bản phiên mã Ubquitin-tTAV2 được dự đoán dịch mã từ codon khởi đầu sắp đặt trong exon đầu tiên của Gen AaAct4 tại 4299-4297 (ubiquitin là từ 2406-2179; tTAV2 là từ 2178- 1162); (chú ý: sự định hướng ngược).

<223> Trình tự pLA3641. <**#> Các điểm đặc trưng quan trọng bao gồm:

1. tTAV vùng mã hóa được lồng vào exon của gen CodlingDsx ở con cái. <***> $_{tT}A\gamma$ từ vị trí 2731-3747 trong gen CodlingDsx.

<***> Dsx-tTAV cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 (đoạn hsρ70 là từ 4811-4940).

<***> tTAV phiên mã được dự đoán dịch mã giữa 2731-3747, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<223> Trình tự pLA3570 <***> $fç_e y$ các đặc trưng bao gồm:

1. tTAV vùng mã hóa được lồng vào exon của gen PBW-Dsx ở con cái. <***> tTAV vùng mã hóa là từ 2336-3352.

<***> Dsx-tTAV cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 (đoạn hsp70 là từ 4683-4812).

<***> tTAV phiên mã được dự đoán dịch mã giữa 2336-3352, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<223> Trình tự pLAl 188 (SED ID NO.49) <***> Các điểm đặc trưng quan trọng bao gồm:

1. tTAV vùng mã hóa lồng vào intron Cctra.

<***> intron CcTRA từ vị trí 3905-2561 trong tTAV (chú ý: sự định hướng ngược). <***> tTAV cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí 4217 (đoạn hsp70 là từ 4260-4131); (chú ý: sự định hướng ngược).

<***> tTAV Phiên mã Fl được dự đoán dịch mã giữa 4040- 1679 (chú ý: sự định hướng ngược). <* * *> điểm đặc trưng bao gồm:

1. Adh intron trong đó dự đoán phiên mã Fl từ vị trí 4118-4049 (chú ý: sự định hướng ngược).

<223> Trình tự pLA3077 (SED ID NO. 50). <***> Các điểm đặc trưng quan trọng bao gồm:

1. tTAV vùng mã hóa lồng vào intron Cctra.

<***> intron CcTRA từ vị trí 3975-2631 trong tTAV (chú ý: sự định hướng ngược). <***> tTAV cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí ~4217 (đoạn hsp70 là từ 4260-4131); (chú ý: sự định hướng ngược).

<***> tTAV Phiên mã Fl được dự đoán dịch mã giữa 4039-1678, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<***> Điểm đặc trưng bao gồm:

1. Adh intron trong đó dự đoán phiên mã Fl từ vị trí 4117-4048 (chú ý: sự định hướng ngược).

<223> Trình tự pLA3097 (SED ID NO. 51). <***> Các điểm đặc trưng quan trọng bao gồm:

1. tTAV vùng mã hóa lồng vào intron Cctra.

<* * *> intron CcTRA từ vị trí 3282-1938 trong tTAV (chú ý: sự định hướng ngược). <***> tTAV cắt nối thay thế phiên mã bắt trong đoạn hsp70 tại vị trí -3382 (đoạn hsp70 là từ 3425-3296); (chú ý: sự định hướng ngược).

<**#> tTAV Phiên mã Fl được dự đoán dịch mã giữa 3285-924, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<223> Trình tự pLA3233 (SED IDNO. 52). <***> Các điểm đặc trưng quan trọng bao gồm:

1. vùng mã hóa tTAV2 lồng vào intron Cctra.

<*##> intron CcTRA từ vị trí 3289-1945 trong tTAV2 (chú ý: sự định hướng ngược). <***> tTAV2 cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí -3389 (đoạn hsp70 là từ 3432-3303); (chú ý: sự định hướng ngược).

<* * *> tTAV2 Phiên mã Fl được dự đoán dịch mã giữa 3292-931, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<223> Trình tự pLA3014 (SED ID NO. 53). <***> Các điểm đặc trưng quan trọng bao gồm:

1. ubi-reaper[KR] vùng mã hóa lồng vào intron Cctra. <***> intron CcTRA từ vị trí 3356-4700 trong ubi-reaper[KR].

<***> ubi-reaper[KR] cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí -3234 (đoạn hsp70 là từ 3191-3320).

<***> ubi-reaper[KR] Phiên mã Fl được dự đoán dịch mã giữa 3331-5143, bao gồm codon khởi đầu và codon kết thúc (ubiquitin là từ 3331-3355, 4701-4948; reaper[KR] là từ 4949- 5143).

<223> Trình tự pLA3166 (SED ID NO. 54). <***> fζ;ey features include:

1. ubi-reaper[KR] vùng mã hóa lồng vào intron Cctra.

<***> intron CcTRA từ vị trí 9987-8643 trong ubi-reaper[KR] (chú ý: sự định hướng ngược). <***> ubi-reaper[KR] cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí -10227 (đoạn hsp70 là từ 10270-10141); (chú ý: sự định hướng ngược). <***> ubi-reaper[KR] Phiên mã Fl được dự đoán dịch mã giữa 10126-8359, bao gồm codon khởi đầu và codon kết thúc (ubiquitin là từ 10126-9988, 8642-8554; reaper[KR] là từ 8553- 8359); (chú ý: sự định hướng ngược).

<223> Trình tự pLA3376 (SED ID NO. 55). <***> Các điểm đặc trưng quan trọng bao gồm:

1. vùng mã hóa tTAV2 lồng vào intron Cctra.

2. tTAV3 vùng mã hóa lồng vào intron Bztra.

3. reaper[KR] vùng mã hóa lồng vào intron Bztra.

<**#> intron CcTRA từ vị trí 3289-1945 trong tTAV2 (chú ý: sự định hướng ngược).

<***> Bztra intron từ vị trí 5981-5014 trong tTAV3 (chú ý: sự định hướng ngược).

<***> Bztra intron từ vị trí 16391-17358 trong reaper[KR].

<***> tTAV2 cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí ~3389

(đoạn hsp70 là từ 3432-3303); (chú ý: sự định hướng ngược).

<***> tTAV3 cắt nối thay thế phiên mã bắt đầu trong đoạn sry-alpha tại vị trí-6019 (đoạn sry-alpha là từ 6243-5999); (chú ý: sự định hướng ngược).

<***> reaper [KR] cắt nối thay thế phiên mã bắt đầu trong đoạn hunchback tại vị trí -16339 (đoạn hunchback là từ 16289-16372).

<***> tTAV2 Phiên mã Fl được dự đoán dịch mã giữa 3292-931, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<* **> tTAV3 Phiên mã Fl được dự đoán dịch mã giữa 5984-4006, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<***> reaper[KR] Phiên mã Fl được dự đoán dịch mã giữa 16385-17550, bao gồm codon khởi đầu và codon kết thúc.

<223> Trình tự pLA3242 (SED ID NO. 56). <***> Các điểm đặc trưng quan trọng bao gồm:

1) tTAV vùng mã hóa lồng vào intron Cctra.

2) reaper[KR] vùng mã hóa lồng vào intron Cctra.

<***> intron CcTRA từ vị trí 3282-1938 trong tTAV (chú ý: sự định hướng ngược).

<***> Qrtj-a intron từ vị trí 5488-4180 trong reaperKR (chú ý: sự định hướng ngược).

<***> reaperKR cắt nối thay thế phiên mã bắt đầu trong đoạn hunchback tại vị trí ~5540 (đoạn hunchback là từ 5590-5507); (chú ý: sự định hướng ngược).

<***> tTAV cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí -3382

(đoạn hsp70 là từ 3425-3296); (chú ý: sự định hướng ngược).

<***> reaperKR Phiên mã Fl được dự đoán dịch mã giữa 4088-5494, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<***> tTAV Phiên mã Fl được dự đoán dịch mã giữa 924-3285, bao gồm codon khởi đầu và codon kết thúc (chú ý: sự định hướng ngược).

<223> Trình tự pLA1172 (SED ID NO. 106). <***> Các điểm đặc trưng quan trọng bao gồm:

1. tTAV vùng mã hóa giữa AaActin4 bắt nguồn từ đoạn. <***> AaActin4 bắt nguồn từ đoạn 7868-11257 và 12366-13100. <***> tTAV phiên mã được dự đoán dịch mã giữa 11342-12358, bao gồm codon khởi đầu và codon kết thúc.

<**#> AaActin4-tTAV phiên mã được dự đoán bắt đầu tại vị trí ~9312. <***> AaActin4 chứa intron (biến thể cắt nối ở con cái) từ vị trí 9403-11204.

<223> Trình tự pLA1038 (Hình 12). <***> Các điểm đặc trưng quan trọng bao gồm:

1. Đoạn NipplDm ('nipper') vùng mã hóa lồng vào intron Cctra ở bên cạnh trình tự gen chuyển tra exonic.

<***> intron CcTRA từ vị trí 3365-4709 trong nipper.

<***> intron CcTRA nằm ở bên cạnh trình tự exonic Cctra tại vị trí 3343-3364 và 4710-4729. <***> nipper cắt nối thay thế phiên mã bắt đầu trong đoạn hsp70 tại vị trí ~3243 (đoạn hsp70 là từ 3200-3329).

<***> nipper Phiên mã Fl được dự đoán dịch mã giữa 3340-5014, bao gồm codon khởi đầu và codon kết thúc.

<223> Trình tự pLA3054 (SED ID NO. 158). <***> Các điểm đặc trưng quan trọng bao gồm:

1. DsRed-ubi-tTAV vùng mã hóa lồng vào intron Cctra cạnh trình tự gen chuyển exonic.

<***> intron CcTRA từ vị trí 3509-2165 trong DsRed-ubi-tTAV (chú ý: sự định hướng ngược). <***> intron CcTRA cạnh trình tự gen chuyển Cctra exonic tại vị trí 3531-3510 và 2164-2145 (chú ý: sự định hướng ngược).

<***> DsRed-ubi-tTAV cắt nối thay thế phiên mã bắt đầu vừa bắt nguồn trong đoạn hsp70 tại vị trí -3243 (đoạn hsp70 là từ 4930-4801) vừa bắt nguồn trong đoạn Opie2 tại vị trí ~4353 (đoạn Opie2 fragment là từ 4795-4255); (chú ý: sự định hướng ngược).

<#***> DsRed-ubi-tTAV Phiên mã Fl được dự đoán dịch mã giữa 4320-888, bao gồm codon khởi đầu và codon kết thúc (DsRed là từ 4212-3538; ubiquitin là từ 2135-1908; tTAV là từ 1907-888); (chú ý: sự định hướng ngược).

<223> Trình tự pLA3056 (SED ID NO. 159). <***> Các điểm đặc trưng quan trọng bao gồm:

1. DsRed-ubi-tTAV vùng mã hóa lồng vào intron Cctra cạnh trình tự gen chuyển exonic.

<***> intron CcTRA từ vị trí 3731-2387 in DsRed-ubi-tTAV (chú ý: sự định hướng ngược). <***> intron CcTRA cạnh trình tự gen chuyển Cctra exonic tại các vị trí 3753-3732 và 2386-2145 (chú ý: sự định hướng ngược).

<***> DsRed-ubi-tTAV cắt nối thay thế phiên mã bắt đầu vừa bắt nguồn trong đoạn ihsp70 tại vị trí -5109 (đoạn hsp70 là từ 5152-5023) vừa bắt nguồn trong đoạn Opie2 tại vị trí -4575 (đoạn Opie2 là từ 5017-4477); (chú ý: sự định hướng ngược).

<***> DsRed-ubi-tTAV Phiên mã Fl được dự đoán dịch mã giữa 4542-888, bao gồm codon khởi đầu và codon kết thúc (DsRed là từ 4434-3760; ubiquitin là từ 2135-1908; tTAV là từ 1907-888); (chú ý: sự định hướng ngược). <***> Điểm đặc trưng bao gồm:

1. intron thêm vào bắt nguồn từ gen Cctra (intron thứ hai của Cctra Phiên mã F1) trong đó dự đoán Phiên mã Fl từ vị trí 2222-2168 (chú ý: sự định hướng ngược).

<223> Trình tự pLA3488 (SED ID NO. 160). <***> Các điểm đặc trưng quan trọng bao gồm:

1. TurboGreen-ubi-DsRed vùng mã hóa lồng vào intron Cctra. <***> intron CcTRA từ vị trí 2263-3607 in TurboGreen-ubi-DsRed. <***> TurboGreen-ubi-DsRed cắt nối thay thế phiên mã bắt đầu trong đoạn bắt nguồn từ baculovirut AcMNPV IeI (immediate early 1) tại vị trí ~1180 (Đoạn IeI là từ 580- 1209).

<***> TurboGreen-ubi-DsRed Phiên mã Fl được dự đoán dịch mã giữa 1311-4467, bao gồm codon khởi đầu và codon kết thúc (TurboGreen là từ 1311-2093; SG4 linker là từ 2094- 2123; ubiquitin là từ 2124-3696, bao gồm intron CcTRA; DsRed là từ 3697-4467).

<***> Điểm đặc trưng bao gồm:

1. intron thêm vào bắt nguồn từ đoạn gen Drosophila ('các đoạn intron') trong đó dự đoán phiên mã Fl từ vị trí 1224-1286.

<223> Trình tự pLA3596 (SED ID NO. 145). <***> Các điểm đặc trưng quan trọng bao gồm:

1. TurboGreen-ubi-DsRed2 vùng mã hóa lồng vào intron Cctra. <***> intron CcTRA từ vị trí 5947-7291 in TurboGreen-ubi-DsRed2. <***> TurboGreen-ubi-DsRed2 cắt nối thay thế phiên mã bắt đầu trong đoạn bắt nguồn từ baculovirut AcMNPV IeI (immediate early 1) tại vị trí ~4864 (Đoạn IeI là từ 4264- 4893).

<***> TurboGreen-ubi-DsRed2 Phiên mã Fl được dự đoán dịch mã giữa 4995-8148, bao gồm codon khởi đầu và codon kết thúc (TurboGreen là từ 4995-5777; SG4 linker là từ 5778- 5807; ubiquitin là từ 5808-7380, bao gồm intron CcTRA; DsRed2 là từ 7381-8151). <***> Điểm đặc trưng bao gồm:

1. intron thêm vào bắt nguồn từ đoạn gen Drosophila ('các đoạn intron') trong đó dự đoán phiên mã Fl từ vị trí 4908-4970.

2. Sự đột biến axit amin dùng để so sánh với LA3488 tại vị trí 7294-7296.

YÊU CẦU BẢO HỘ

1. Phương pháp biểu hiện protein chức năng ở côn trùng thông qua sự cắt nối thay thế ARN phiên mã đặc hiệu giới tính, trong đó phương pháp này bao gồm các bước:

a) phiên mã, trong côn trùng nêu trên, polynucleotit ngoại sinh từ hệ thống biểu hiện để tạo ra sản phẩm phiên mã ARN, trong đó hệ thống biểu hiện này bao gồm:

trình tự khởi đầu được liên kết một cách có kiểm soát với trình tự polynucleotit ngoại sinh mã hóa cho protein chức năng, nằm giữa codon khởi đầu và codon kết thúc, và

trình tự kiểm soát cắt nối intron bao gồm miền liên kết protein;

trong đó miền liên kết protein bao gồm trình tự liên ứng ADN được nêu trong SEQ ID NO:1 hoặc ARN tương ứng với nó, trong đó trình tự kiểm soát cắt nối intron bao gồm trình tự cho cắt nối GT ở đầu 5' của nó (5'-GT), và trong đó trình tự kiểm soát cắt nối intron được kẹp sườn bởi nucleotit guanin (G); và

trong đó trình tự kiểm soát cắt nối intron là từ đầu 3' đến codon khởi đầu ATG của trình tự polynucleotit ngoại sinh;

b) cắt nối thay thế sản phẩm phiên mã ARN của trình tự polynucleoti ngoại sinh, cùng hoạt động với vị trí cắt, để tạo ra sản phẩm ARN thông tin (mARN) được cắt nối thứ nhất, mà không bao gồm khung đọc mở liên tục kéo dài từ codon khởi đầu đến codon kết thúc, và hơn nữa sản phẩm ARN vận chuyển được cắt nối thay thế mà bao gồm khung đọc mở liên tục kéo dài từ codon khởi đầu đến codon kết thúc, xác định protein chức năng nêu trên đối với việc biểu hiện.

2. Phương pháp theo điểm 1, trong đó trình tự polynucleotit được biểu hiện bao gồm hai hoặc nhiều exon mã hóa cho protein chức năng.

3. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó protein này là chất đánh dấu hoặc có tác dụng gây chết, gây hại hoặc gây vô sinh.

4. Phương pháp theo điểm 3, trong đó protein có tác dụng gây chết dẫn đến vô sinh.

5. Phương pháp theo điểm 4, trong đó tác dụng gây chết của protein có thể kìm hãm được một cách có điều kiện.

6. Phương pháp theo điểm 3, trong đó protein được chọn từ nhóm bao gồm yếu tố kích thích gây chết tế bào theo chương trình, Hid, Reaper (Rpr) và Nipp1Dm.

7. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó hệ thống này bao gồm ít nhất một cơ chế phản hồi chủ động, có ít nhất một protein chức năng được có mức biểu hiện khác nhau, thông qua việc cắt nối, và có ít nhất một trình tự khởi đầu, trong đó một sản phẩm gen được biểu hiện đóng vai trò như một yếu tố kiểm soát phiên mã chủ động đối với ít nhất một trình tự khởi đầu, và nhờ đó có thể kiểm soát sản phẩm hoặc việc biểu hiện sản phẩm.

8. Phương pháp theo điểm 7, trong đó trình tự tăng cường được liên kết với trình tự khởi đầu, sản phẩm gen này giúp tăng cường hoạt tính của trình tự khởi đầu thông qua trình tự tăng cường.

9. Phương pháp theo điểm 8, trong đó yếu tố kiểm soát là sản phẩm của gen tTA hoặc thể tương đồng của nó, và trong đó một hoặc nhiều đơn vị khởi động tetO được liên kết một cách có kiểm soát với trình tự khởi đầu và trình tự tăng cường, trong đó tTA hoặc thể tương đồng của nó giúp tăng cường hoạt tính của trình tự khởi đầu thông qua tetO.

10. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó bản thân protein chức năng là chất hoạt hóa phiên mã, chẳng hạn như hệ thống tTAV, bao gồm tTAV, tTAV2 hoặc tTAV3.

11. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó trình tự khởi đầu được hoạt hóa bởi các điều kiện môi trường, chẳng hạn sự có mặt hoặc vắng mặt một yếu tố đặc biệt như tetraxyclin trong hệ thống TET hoặc sự thay đổi nhiệt độ môi trường.

12. Phương pháp theo điểm bất kỳ trong số các điểm từ 1 đến 10, trong đó trình tự khởi đầu được chọn từ nhóm bao gồm trình tự khởi đầu đặc hiệu phôi từ *Drosophila melanogaster* hoặc thể tương đồng của nó có mức độ tương đồng về

trình tự ít nhất 85% và gen *slow as molasses (slam)* ở *Drosophila* như hoặc thể tương đồng của nó có mức độ tương đồng về trình tự ít nhất 85%, trong đó mức tương đồng về mặt trình tự được xác định chính xác bằng cách sử dụng BLAST.

13. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó trình tự kiểm soát cắt nối có nguồn gốc từ intron *tra*.

14. Phương pháp theo điểm 13, trong đó trình tự kiểm soát cắt nối có nguồn gốc từ gen chuyển Medfly *Ccfra*, hoặc thể tương đồng cùng loài hoặc thể tương đồng khác loài của gen chuyển *Drosophila* có mức độ tương đồng về trình tự ít nhất 85%, trong đó thể tương đồng đã được xác định chính xác bằng cách sử dụng BLAST.

15. Phương pháp theo điểm 14, trong đó thể tương đồng cùng loài hoặc thể tương đồng khác loài của gen chuyển *Drosophila* là từ ruồi giấm tephritid.

16. Phương pháp theo điểm 15, trong đó, ruồi giấm tephritid là *C. rosa*, hoặc *B. zonata*.

17. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó trình tự kiểm soát cắt nối bao gồm nucleotit GU ở đầu 5' và AG ở đầu 3' trong ARN.

18. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó việc cắt nối thay thế được kiểm soát bởi sự gắn kết của protein TRA hoặc phức hợp protein TRA/TRA2.

19. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó côn trùng là thuộc bộ hai cánh.

20. Phương pháp theo điểm 19, trong đó côn trùng là ruồi giấm tephritid được chọn từ nhóm bao gồm: Medfly (*Ceratitis capitatd*), Mexfly (*Anastrepha ludens*), ruồi giấm Oriental (*Bactrocera dorsalis*), ruồi giấm Olive (*Bactrocera oleae*), ruồi Melon (*Bactrocera cucurbitae*), ruồi giấm Natal (*Ceratitis rosa*), ruồi giấm Cherry (*RJ-iagoletis cerasi*), ruồi giấm Queensland (*Bactrocera tyroni*), ruồi giấm Peach (*Bactrocera zonata*), ruồi giấm Caribbean (*Anastrepha suspensa*) và ruồi giấm tây Ấn Độ (*Anastrepha obliqua*).

21. Phương pháp theo điểm 19, trong đó côn trùng là muỗi từ chi *Stegomyia, Aedes, Anopheles* hoặc *Culex*.

22. Phương pháp theo điểm 21, trong đó muỗi được chọn từ nhóm bao gồm *Aedes aegypti, Aedes albopictus, Anopheles stephensi, Anopheles albimanus* và *Anopheles gambiae*.

23. Phương pháp theo điểm 19, trong đó côn trùng được chọn từ nhóm bao gồm: ruồi giun thế giới mới (*Cochliomyia hominivorax*), ruồi giun thế giới cũ (*Chrysomya bezzianct*) và nhặng cừu ở Úc (*Lucilia cuprina*), bướm đêm codling (*Cydia pomonella*), tằm (*Bombyx mori*), sâu hồng hại bông (*Pectinophora gossypiella*), sâu bướm diamondback (*Plutella xylostetta*), sâu bướm Gypsy (*Lymantria dispar*), sâu cam Navel (*Amyelois transitella*), sâu đục cành đào (*Anarsia lineatella*) và sâu đục thân lúa (*Trypoiyza incertulas*), sâu bướm cú, cụ thể là Heliothinae, xén tóc Nhật Bản (*Popilla japonica*), xén tóc vân trắng (*fimphognatus* spp.), bọ vòi voi hại bông (*Anthonomous grandis*), sâu đục rễ ngô (*Diabrotica* spp) và xén tóc hại khoai tây Colorado (*Leptinotarsa decemlineata*).

24. Phương pháp theo điểm bất kỳ trong số các điểm nêu trên, trong đó việc biểu hiện của trình tự polynucleotit ngoại lai dẫn đến biểu hiện kiểu hình ở côn trùng.

25. Phương pháp kiểm soát quần thể côn trùng trong môi trường tự nhiên, trong đó phương pháp này bao gồm các bước:

i) nhân giống ra một nguồn côn trùng, côn trùng này mang một hệ thống biểu hiện gen bao gồm hệ thống biểu hiện như được nêu trong điểm phương pháp bất kỳ trong số các điểm từ 1 đến 24 mà là hệ thống gen trội gây chết;

ii) phát tán nguồn sinh vật nêu trên vào môi trường tại địa điểm cần kiểm soát quần thể; và

iii) đạt được sự kiểm soát quần thể thông qua việc gây chết giai đoạn sớm bởi việc biểu hiện hệ thống gây chết ở thế hệ con được tạo ra từ việc lai giống giữa các cá thể từ nguồn đã nêu với các cá thể khác giới tính trong quần thể tự nhiên.

26. Phương pháp theo điểm 25, trong đó sự gây chết ở giai đoạn sớm là giai đoạn phôi hoặc trước giai đoạn thuần thục giới tính.

27. Phương pháp theo điểm 26, trong đó sự gây chết ở giai đoạn sớm xảy ra sớm trong quá trình phát triển.

28. Phương pháp theo điểm 25 hoặc 26, trong đó tác dụng gây chết của hệ thống gây chết là có điều kiện và xảy ra trong môi trường tự nhiên thông qua sự biểu hiện của gen gây chết, sự biểu hiện của gen gây chết này thuộc sự kiểm soát của một protein chuyển hoạt hóa có thể kìm hãm được, việc nhân giống này diễn ra trong điều kiện cho phép với sự có mặt của cơ chất, mà cơ chất này không có trong môi trường tự nhiên và có thể kìm hãm được chất chuyển hoạt hóa nêu trên.

29. Phương pháp phân chia giới tính, trong đó phương pháp này bao gồm các bước:

i) nhân giống một nguồn côn trùng đực và cái được chuyển nhiễm với hệ thống biểu hiện như được xác định theo điểm bất kỳ trong số các điểm từ 1 đến 24 trong điều kiện cho phép hoặc giới hạn, mà cho phép con đực và con cái sống sót, trong đó hệ thống biểu hiện này bao gồm polynucleotit khác loài là gen gây chết; và

ii) loại bỏ các điều kiện cho phép hoặc giới hạn này để kích thích tác dụng gây chết của gen gây chết trong một nhóm giới tính, nhưng không ảnh hưởng đến nhóm giới tính còn lại do việc cắt nối thay thế đặc hiệu giới tính của gen gây chết.

DANH MỤC TRÌNH TỰ

<110> Oxitec Limited

<120> Phương pháp biểu hiện protein chức năng ở côn trùng và phương pháp phân chia giới tính côn trùng

<130> USP92438

<150> PCT/GB2004/003263

<151> 2004-07-28

<160> 162

<170> PatentIn version 3.3

<210> 1
<211> 13
<212> ADN
<213> nhân tạo
<220>
<223> Trình tự liên ứng Ceratitis capitata tra
<400> 1
tcwwcratca aca 13

<210> 2
<211> 10
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự chặn đầu LA3097
<400> 2
agccaccatg 10

<210> 3
<211> 10
<212> ADN
<213> nhân tạo
<220>
<223> Trình tự chặn đầu LA3097
<400> 3
gtcagccgcc 10

<210> 4
<211> 21
<212> ADN
<213> nhân tạo
<220>
<223> đoạn mồi 688 - iel-transcr

<400> 4
gttgcaagtt gacactggcg g 21

<210> 5
<211> 21
<212> ADN
<213> nhân tạo
<220>
<223> đoạn mồi 790 - Aedsx-m-r2
<400> 5
ccactgtgta aggcttcctc c 21

<210> 6
<211> 21
<212> ADN
<213> nhân tạo
<220>
<223> đoạn mồi 761 - Aedsx-fem-r
<400> 6
ggatggttgg ttgaagatcc g 21

<210> 7
<211> 21
<212> ADN
<213> nhân tạo
<220>
<223> đoạn mồi AedsxR1
<400> 7
actgcgcaac tctacaccgt c 21

<210> 8
<211> 13
<212> ARN
<213> nhân tạo
<220>
<223> trình tự liên ứng Pane et al
<400> 8
ucwwcrauca aca 13

<210> 9
<211> 13
<212> ARN
<213> nhân tạo
<220>
<223> trình tự liên ứng Scali et al 2005
<400> 9
ucwwcaauca aca 13

<210> 10

<211> 13
<212> ADN
<213> Drosophila sp.
<400> 10
tcaacaagca aca 13

<210> 11
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 11
ttatcaaaca aca 13

<210> 12
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 12
tcatcaatta aaa 13

<210> 13
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 13
tcatcaatca aac 13

<210> 14
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 14
tcttcaacca acc 13

<210> 15
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 15
cctacaatct aca 13

<210> 16
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 16
tcttagatca aaa 13

<210> 17

<211> 13
<212> ADN
<213> Drosophila sp.
<400> 17
tcttcgatca tta 13

<210> 18
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 18
ccaacaatct aca 13

<210> 19
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 19
tcaaagatca cca 13

<210> 20
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 20
tcttcggtcg acg 13

<210> 21
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 21
tcgacaaaca aaa 13

<210> 22
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 22
tattcaaaca acg 13

<210> 23
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 23
ttttcgataa aaa 13

<210> 24

<211> 13
<212> ADN
<213> Drosophila sp.
<400> 24
tcttcagtct gca 13

<210> 25
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 25
gattcaatca tca 13

<210> 26
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 26
ttatcgagca aaa 13

<210> 27
<211> 13
<212>
<213> Drosophila sp.
<400> 27
tcataactca aga 13

<210> 28
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 28
tcagaaatca aaa 13

<210> 29
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 29
tctttaattt aca 13

<210> 30
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 30
tttacaatcc tca 13

<210> 31

<211> 13
<212> ADN
<213> Drosophila sp.
<400> 31
tcatagatca gga 13

<210> 32
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 32
acctcaaaca aca 13

<210> 33
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 33
tcatcgaaca ccc 13

<210> 34
<211> 1014
<212> ADN
<213> nhân tạo
<220>
<223> khung đọc mở của cấu trúc tTAV
<400> 34
atgggcagcc gcctggataa gtccaaagtc atcaactccg cgttggagct gttgaacgaa 60
gttggcattg agggactgac gacccgcaag ttggcgcaga agctgggcgt ggagcagccc 120
accctctact ggcacgtgaa gaataagcgg gcgctgctgg atgccctggc catcgagatg 180
ctcgaccgcc accacacgca tttttgcccg ttggaaggcg agtcctggca ggacttcctc 240
cgcaataacg ccaagtcgtt ccgctgcgct ctgctgtccc accgagacgg tgccaaagtc 300
catctcggca cgcgcccgac cgaaaagcaa tacgagacac tggagaacca gctcgcgttc 360
ctgtgccagc aaggcttcag cctggaaaat gctctctacg ctctgagcgc cgtcggtcac 420
tttaccctgg gctgcgtgct ggaggaccaa gagcatcaag tcgcaaaaga ggagcgcgag 480
accccaacaa ccgattcgat gcccccactg ctgcgtcagg caatcgagct gttcgatcat 540
caaggagccg agccggcatt cctgttcggc ttggagctga ttatctgcgg attggaaaag 600
caactgaaat gcgagtcggg ctcgggcccc gcgtacagcc gcgcgcgtac gaaaaacaat 660
tacgggtcta ccatcgaggg cctgctcgat ctcccggacg acgacgcccc cgaagaggcg 720
gggctggcgg ctccgcgcct gtcctttctc cccgcgggac acacgcgcag actgtcgacg 780
gccccccccga ccgatgtcag cctgggggac gagctccact tagacggcga ggacgtggcg 840
atggcgcatg ccgacgcgct agacgatttc gatctggaca tgttgggggga cggggattcc 900
ccgggtccgg gatttacccc ccacgactcc gccccctacg gcgctctgga tatggccgac 960
ttcgagtttg agcagatgtt taccgatgcc cttggaattg acgagtacgg tggg 1014

<210> 35
<211> 338
<212> PRT
<213> nhân tạo

<220>
<223> Trình tự protein của tTAV
<400> 35

```
Met Gly Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu
1               5                  10                  15
Leu Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala
            20                  25                  30
Gln Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn
        35                  40                  45
Lys Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His
    50                  55                  60
His Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu
65                  70                  75                  80
Arg Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp
                85                  90                  95
Gly Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu
            100                 105                 110
Thr Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu
        115                 120                 125
Glu Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly
    130                 135                 140
Cys Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu
145                 150                 155                 160
Thr Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu
                165                 170                 175
Leu Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu
            180                 185                 190
Leu Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser
        195                 200                 205
Gly Pro Ala Tyr Ser Arg Ala Arg Thr Lys Asn Asn Tyr Gly Ser Thr
    210                 215                 220
Ile Glu Gly Leu Leu Asp Leu Pro Asp Asp Asp Ala Pro Glu Glu Ala
225                 230                 235                 240
Gly Leu Ala Ala Pro Arg Leu Ser Phe Leu Pro Ala Gly His Thr Arg
                245                 250                 255
Arg Leu Ser Thr Ala Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu
            260                 265                 270
His Leu Asp Gly Glu Asp Val Ala Met Ala His Ala Asp Ala Leu Asp
        275                 280                 285
Asp Phe Asp Leu Asp Met Leu Gly Asp Gly Asp Ser Pro Gly Pro Gly
    290                 295                 300
Phe Thr Pro His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp
305                 310                 315                 320
Phe Glu Phe Glu Gln Met Phe Thr Asp Ala Leu Gly Ile Asp Glu Tyr
                325                 330                 335
Gly Gly
```

<210> 36

<211> 1014
<212> ADN
<213> nhân tạo
<220>
<223> Khung đọc mở của tTAV2
<400> 36

atgagccgcc tggataagtc caaagtcatc aactccgcgt tggagctgtt gaacgaagtt 60
ggcattgagg gactgacgac ccgcaagttg gcgcagaagc tgggcgtgga gcagcccacc 120
ctctactggc acgtgaagaa taagcgggcg ctgctggatg ccctggccat cgagatgctc 180
gaccgccacc acacgcattt ttgcccgttg gaaggcgagt cctggcagga cttcctccgc 240
aataacgcca agtcgttccg ctgcgctctg ctgtcccacc gagacggtgc caaagtccat 300
ctcggcacgc gcccgaccga aaagcaatac gagacactgg agaaccagct cgcgttcctg 360
tgccagcaag gcttcagcct ggaaaatgct ctctacgctc tgagcgccgt cggtcacttt 420
accctgggct gcgtgctgga ggaccaagag catcaagtcg caaaagagga gcgcgagacc 480
ccaacaaccg attcgatgcc cccactgctg cgtcaggcaa tcgagctgtt cgatcatcaa 540
ggagccgagc cggcattcct gttcggcttg gagctgatta tctgcggatt ggaaaagcaa 600
ctgaaatgcg agtcgggctc gggccccgcc tacagccgcg cccgcaccaa gaacaactac 660
ggcagcacca tcgagggcct gctggatctg ccggatgatg atgccccgga ggaggcgggc 720
ctggccgccc cgcgcctgag cttcctgccg gccggacaca cccgccgcct gtcgaccgcc 780
ccgccgaccg acgtgagcct gggcgatgag ctgcacctgg atggcgagga tgtggcgatg 840
gcccacgccg atgccctgga cgacttcgac ctggacatgc tgggcgatgg cgatagcccg 900
ggaccgggat tcacccccgca cgatagcgcc ccctacggcg ccctggatat ggccgatttc 960
gagttcgagc agatgttcac cgacgccctg ggcatcgatg agtacggcgg ctaa 1014

<210> 37
<211> 337
<212> PRT
<213> nhân tạo
<220>
<223> trình tự protein của tTAV2
<400> 37

Met Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu Leu
1 5 10 15
Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala Gln
20 25 30
Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn Lys
35 40 45
Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His His
50 55 60
Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu Arg
65 70 75 80
Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp Gly
85 90 95
Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu Thr
100 105 110
Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu Glu
115 120 125
Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly Cys
130 135 140

Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu Thr
145 150 155 160
Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu Leu
165 170 175
Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu Leu
180 185 190
Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser Gly
195 200 205
Pro Ala Tyr Ser Arg Ala Arg Thr Lys Asn Asn Tyr Gly Ser Thr Ile
210 215 220
Glu Gly Leu Leu Asp Leu Pro Asp Asp Asp Ala Pro Glu Glu Ala Gly
225 230 235 240
Leu Ala Ala Pro Arg Leu Ser Phe Leu Pro Ala Gly His Thr Arg Arg
245 250 255
Leu Ser Thr Ala Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu His
260 265 270
Leu Asp Gly Glu Asp Val Ala Met Ala His Ala Asp Ala Leu Asp Asp
275 280 285
Phe Asp Leu Asp Met Leu Gly Asp Gly Asp Ser Pro Gly Pro Gly Phe
290 295 300
Thr Pro His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp Phe
305 310 315 320
Glu Phe Glu Gln Met Phe Thr Asp Ala Leu Gly Ile Asp Glu Tyr Gly
325 330 335
Gly

<210> 38
<211> 1011
<212> ADN
<213> nhân tạo
<220>
<223> Khung đọc mở của tTAV3
<400> 38
atgggcagcc gcctggacaa gagcaaggtg atcaacagcg ccctggagct gctgaacgaa 60
gttggtatcg agggcctgac cacccgcaag ctggcccaga agctgggcgt ggaacagccg 120
accctgtact ggcacgtgaa gaacaagcgc gccctgctgg acgccctggc catcgaaatg 180
ctggatcgcc accacaccca cttctgcccg ctggagggcg agagctggca ggatttcctg 240
cgcaacaacg ccaagagctt ccgctgcgcc ctgctgtcgc accgcgatgg cgccaaggtg 300
cacctgggca cccgcccgac cgagaagcag tacgagaccc tggagaacca gctggccttc 360
ctgtgccagc agggcttcag cctggagaac gccctgtacg ccctgagcgc cgtgggccac 420
ttcaccctgg gctgtgtgct ggaggatcag gagcaccagg tggccaagga ggagcgcgag 480
accccgacca ccgatagcat gccgccgctg ctgcgccagg ccatcgagct gttcgatcac 540
cagggcgccg agccggcctt cctgttcggc ctggagctga tcatctgcgg cctggaaaag 600
cagctgaagt gcgagagcgg cagcgcctac agccgcgccc gtaccaagaa caactatggc 660
agcaccatcg agggactgct ggacctgccg gatgacgatg ccccggagga agccggcctg 720
gccgcccccc gcctgagctt cctgcccgcc ggacacacgc gccgcctgag caccgccccg 780
ccgaccgatg tgagcctggg cgacgagctg cacctggatg gagaggatgt ggcaatggcc 840
cacgccgacg ccctggacga tttcgacctg gatatgctgg gcgatggaga tagcccggga 900
ccgggcttca cgccccacga tagcgccccg tacggcgccc tggacatggc cgacttcgag 960

ttcgagcaaa tgttcaccga cgcgctgggc atcgatgagt atggcgggta g 1011

<210> 39
<211> 336
<212> PRT
<213> nhân tạo
<220>
<223> Trình tự protein của tTAV3
<400> 39

Met Gly Ser Arg Leu Asp Lys Ser Lys Val Ile Asn Ser Ala Leu Glu
1 5 10 15
Leu Leu Asn Glu Val Gly Ile Glu Gly Leu Thr Thr Arg Lys Leu Ala
20 25 30
Gln Lys Leu Gly Val Glu Gln Pro Thr Leu Tyr Trp His Val Lys Asn
35 40 45
Lys Arg Ala Leu Leu Asp Ala Leu Ala Ile Glu Met Leu Asp Arg His
50 55 60
His Thr His Phe Cys Pro Leu Glu Gly Glu Ser Trp Gln Asp Phe Leu
65 70 75 80
Arg Asn Asn Ala Lys Ser Phe Arg Cys Ala Leu Leu Ser His Arg Asp
85 90 95
Gly Ala Lys Val His Leu Gly Thr Arg Pro Thr Glu Lys Gln Tyr Glu
100 105 110
Thr Leu Glu Asn Gln Leu Ala Phe Leu Cys Gln Gln Gly Phe Ser Leu
115 120 125
Glu Asn Ala Leu Tyr Ala Leu Ser Ala Val Gly His Phe Thr Leu Gly
130 135 140
Cys Val Leu Glu Asp Gln Glu His Gln Val Ala Lys Glu Glu Arg Glu
145 150 155 160
Thr Pro Thr Thr Asp Ser Met Pro Pro Leu Leu Arg Gln Ala Ile Glu
165 170 175
Leu Phe Asp His Gln Gly Ala Glu Pro Ala Phe Leu Phe Gly Leu Glu
180 185 190
Leu Ile Ile Cys Gly Leu Glu Lys Gln Leu Lys Cys Glu Ser Gly Ser
195 200 205
Ala Tyr Ser Arg Ala Arg Thr Lys Asn Asn Tyr Gly Ser Thr Ile Glu
210 215 220
Gly Leu Leu Asp Leu Pro Asp Asp Asp Ala Pro Glu Glu Ala Gly Leu
225 230 235 240
Ala Ala Pro Arg Leu Ser Phe Leu Pro Ala Gly His Thr Arg Arg Leu
245 250 255
Ser Thr Ala Pro Pro Thr Asp Val Ser Leu Gly Asp Glu Leu His Leu
260 265 270
Asp Gly Glu Asp Val Ala Met Ala His Ala Asp Ala Leu Asp Asp Phe
275 280 285
Asp Leu Asp Met Leu Gly Asp Gly Asp Ser Pro Gly Pro Gly Phe Thr
290 295 300
Pro His Asp Ser Ala Pro Tyr Gly Ala Leu Asp Met Ala Asp Phe Glu
305 310 315 320

Phe Glu Gln Met Phe Thr Asp Ala Leu Gly Ile Asp Glu Tyr Gly Gly
325 330 335

<210> 40
<211> 568
<212> ADN
<213> Pectinophora gossypiella
<400> 40
gctagtggag aactgccaca aactgctgga aaagttccac tactcctggg aaatgatgcc 60
cctggtgctg gtcattctaa actacgccgg ctccgacctc gacgaggctt ctagaaaaat 120
tgatgaaggg aagatgatca tcaacgagta cgcgaggaag cacaatctga acatcttcga 180
tggccacgag ctaaggaact cgactcgcca gtacggactt taatacagta atattagttt 240
tctccaacaa cactaaacac gacataacac gctacacgca aaaaatacac gagtctttaa 300
tgttttacac gctcagtaaa ttattcactt acacgcttaa ctaaaatttt acacaatcgg 360
taaaaaaata caacaattta ttatcgtaaa aattacacaa aataaatgag atttaaatgt 420
cgtttaataa aataaaataa aaatagcatc gggaatatct tttcacctat tgccggagaa 480
cagtttaaat ggatactctc atttgaatca ttttaattgt agtagcattt tattttatta 540
ttaatagcaa taagtacaca aacataaa 568

<210> 41
<211> 610
<212> ADN
<213> Pectinophora gossypiella
<400> 41
gtagtggaga actgccacaa actgctggaa aagttccact actcctggga aatgatgccc 60
ctggtgctgg tcattctaaa ctacgccggc tccgacctcg acgaggcttc tagaaaaatt 120
gatgaaggga agatgatcat caacgagtac gcgaggaagc acaatctgaa catcttcgat 180
ggccacgagc tgaggaactc gactcgccag tacggacttt aatacagaaa atgctgagcg 240
aaattaataa tataagtggt gtactatcgt cgtccatgaa gttattttgc gaatgatact 300
ttgttttgta tgtgctgtgt gttgtgtgga cttttgctgt gcgttgctgt ttgcgatgga 360
aggactattg tgtcgtcgcc acgctggact attcgcacat tgggtggtcc accagtggcg 420
gatgtacgag cggtcgctgt gctcgctcct ggagctgcaa gcgcgcaaag ggacgtactc 480
ggtgtgctgc tcaccccgct acgtcatcgc gccccgagtac gcgtcacacc tgttgcctct 540
gccgcttacc acgcagagat catccccgcc gccccgcgcac ttgtagcgat gcgaacctgc 600
gccgcgggaa 610

<210> 42
<211> 449
<212> DNA
<213> Pectinophora gossypiella

<220>
<221> dấu hiệu hỗn tạp
<222> (26).. (26)
<223> n là a, c, g hoặc t
<400> 42
gctagtggag aactgccaca aactgntgga aaagttccac tactcctggg aaatgatgcc 60
cctggtgctg gtcattctaa actacgccgg ctccgacctc gacgaggctt ctagaaaaat 120
tgatgaagca cattgggtgg tccaccagtg gcggatgtac gagcggtcgc tgtgctcgct 180

cctggagctg caagcgcgca aagggacgta ctcggtgtgc tgctcacccc gctacgtcat 240
cgcgcccgag tgcgcgtcac acctgttgcc tctgccgctt accacgcaga gatcatcccc 300
gccgcccgcg cacttgtagc gatgcgaacc tgcgccgcgg gaagtaagta ctatttcatt 360
tattattctt tttatttttg gttttaaggt gctgacagac ttgaatttca agcaaatagt 420
gtctgacaaa gagctcaaaa tagacatgt 449

<210> 43
<211> 28774
<212> ADN
<213> Aedes aegypti
<400> 43
acagtgaaat ttgatcgatc actcatcgaa acgagatcac tttcgattga tcgtgacaat 60
tttttagaat ccatttcaca gtcgttggga ctgttgaccc tgtcacttta aactagctag 120
tgagtagctt tgctctagtg aaagctaact agcactgtta aaaaatctta ggtaaagtgt 180
cagcaaccct gacaactggg ccacctcttg ccgaccataa gcaaatgaaa tcaaatggtt 240
cgctacgaag gttaattggg tttcgatcta cttcgtccta agcgctattt ttcgtcatac 300
ggtggagaac ggctggtatt cgtttacttt agtttaccaa gcgatgcttc caattaaccc 360
aaagctagat gaagcaggat tcgcgataaa aagcagtatg cgaacttaaa atgttctact 420
acattacggc gggtattcaa atttacctgc cacataaatt tattttccaa gtataatttg 480
cgaaagctgc aatggttcat gcttgaattt tacaagatga tgtaatgccg cccataagtt 540
taaatggacg gtgtatttaa ataaaaggtt catattaaac gctttcgacg ttaccaagta 600
ccatttgtac acaaacatgt aataaaacta ttgtatttct ataaataact tcagttcaat 660
catccacttt gcacattttc accgaaatcg catggacgaa ggtaaacatg tgtttgtaca 720
ttattttgat aacataaaga tatttattga agtcaagtta gtaggtgaaa cgtgtaaaag 780
tggctttagc gtacctgctt gacgtaccga gcgaaatctg attagcggtc gactaagcca 840
taaaacttct acaattcaca aaattttgaa aaattccctc gctgccacga tactaatgca 900
ctgcatggct cgctttagac taatcgccag ctgattcggt attttgaaga tgttaagtgt 960
tttaaaactt tttaagggag cgacggtgct atgattacgt aatcaaatgt tctttctttt 1020
actttcagac caattgcaga acaagcttta tcctaatcca tctcattttg ggaacagcac 1080
tagccgcgac cattagccgt ttagtttaca agaaagaaaa tgaaagtctg gttaacgtct 1140
tgttcgaaat aggattaggt agagtaaaac ccttgtcgtg atcggcgctg gtaatcggca 1200
tctgcgtaga gaacatgttg tacttcctcg aggacgattg ctcgcgctcg cacggttctt 1260
attgctacca tggtgaaacc actagcgccg aggaagtgct agacgcatct cttgtacaac 1320
atgaaggagc tcctgctaac gagcgcgagc gtgccaagaa taacgatggt accactttgg 1380
tgatcgcggc tccttcacga taccgttgtg aaggttttct gaattgcgca tcgtctccga 1440
agggtgtgtc caggtgcatt gtctcccaac tgacctgttc ccgacaatat cgagcactaa 1500
atggcaacac ttccaaaaga cttaacgcgt agcagaggct tcccacacag gtccacgtaa 1560
cagagggttg actggacaag ggctgttata gctcgtgatt ggtttccatt agagagcagt 1620
atctcgtagt agcgtaggag agtccattag agtgcgatat tccgtgagtt tgtgtgaccg 1680
gcgatagaga agccctgacg ccaaaggtaa tctctcgtca tagagcatca tcgcatcctc 1740
tcaggtaatc tcacgctata aggcactcaa acacactggc cgctatctct tcgggactgc 1800
cgcgcttcaa gacgattgta actcggaaac tgacctgatt agtacataaa aagagaccta 1860
ttgcgtaagc ttataagaaa cgagtttgtc cacacggttg gcgcgaagtt ctgctaacat 1920
tgagcctttg actggactaa tcatgtattt ttctctggat aacgcattcg aatattcttt 1980
gctcaaacag gtgtgccaac atggtttcgc aagatcgctg gatggtaaag atgtccgagg 2040
cagggtacga taaccgggcg gatggcagtg gagcttccag cagcagcctg aacccgcgaa 2100
taccaaagcg ttctagcgac ctaccatttc tacaggctcc gtcccatgct attggcccgc 2160
ctaccgtcac ctcgaaggtc gtcgtcggac ttgggcgctt cgccgccgaa ctgtgcccgc 2220
tgccggaacc acggtcacaa gatcggcctg aagggacaca agcgctattg taagtatcgc 2280

aattgtacct gcgaaaagtg gcggcggctt gacacgggcg acggccttgg tgccagtgtt 2340
ctagccggac ttccctgtgt tcgcgataac attcatagcg ttaacatgga cgcttttcac 2400
ctgcctgacg gccgaacggc agcgggtcat ggccctgcag acggctctcc gaagggcgca 2460
aacccaggac gaacagcggt tgctggtaga cggagaggtg gacggactgc cggcttgccg 2520
tcgcccagta ccgggacgtc tgccgagagg cttcccgcgt ttgggtcctg cttgtcgcca 2580
acgaccatct gcctctccac cccgccgaac cggtacatag ccttcaaata ccaaaattgt 2640
ctgacctaaa agagatgatc cataattctc agcagaggtc gttgatcgac tgcgactcgt 2700
gggcggcttg gccatgtatc ggaagtttat ggttttaaca gactggattt tctctactag 2760
gtattaagag tcgtctccag caactagctg acgctgagca ccaccggctc gatgaactcc 2820
accccgggca gctcgttggt aacgctgtcc cagcaccgaa gatcaccctg ctccgccgcg 2880
tcggtccacc ccagcgaggc ggtggccgag ctacttgagg tggggcccgt cgagcaacca 2940
ttgcgacagg gtcgtggctt ctagtgggac gaggcggcgc agccaggtgg ggtcgctccg 3000
tcagcaaaac gttgcaggta ggtgtgaggc atatctattt cgttattctc tcaatgtttg 3060
tggagaaccg gccggaattc aacatcgaag tcggtttctg agtcgttttg caacgtccat 3120
ccacactccg tatagataaa gcaataagag agttacaaac acctcttggc cggccttaag 3180
ttgtagcttc agccaaagac ttctattgat ttatgataaa tttctctcaa atgtttgcgc 3240
ggagggtgga tttttgagag ctgagtggtg tagaaacgaa atgggcatca aacgttatgc 3300
aagataacta aatactattt aaagagagtt tacaaacgcg cctcccacct aaaaactctc 3360
gactcaccac atctttgctt tacccgtagt ttgcaatacg ggcgctgctt gaaacaggtt 3420
tatgttaggg gtttcctgtg tttcatacag tcaccccatt gttatgtata gcacacagat 3480
atggataaaa gttggattaa ccgcgacgaa ctttgtccaa atacaatccc caaaggacac 3540
aaagtatgtc agtggggtaa caatacatat cgtgtgtcta tacctatttt caacctaatt 3600
gcagtgaata tcccatcaaa tagagttgca attgagtaga acacatttta ccaacgtata 3660
aagcatcgta atcaattata atatacttaa gcaaaatata cgtcacttat agggtagttt 3720
atctcaacgt taactcatct tgtgtaaaat ggttgcatat ttcgtagcat tagttaatat 3780
tatatgaatt cgttttatgt atggggaaat aatttgtcaa ccacatttct agaaaagttg 3840
attcatacat gtgtgctttt gaaagccata taccacatta tgtttgattc atatctctta 3900
taccccttta ttaaacagtt ggtgtaaaga tcttttcaac taagtatgta cacacgaaaa 3960
ctttcggtat atggtgtaat acaaactaag tatagagaat taatatgagt cgatttatcg 4020
cgaaattttt caaaatgtcc tatgtaccaa tgaaagatac tctcttatct cgctctgttt 4080
tgaacataac aactgaaact attatactca gctaaatagc gctttaaaaa gttttacagg 4140
atacatggtt actttctatg agagaataga gcgagacaaa acttgtattg ttgactttga 4200
tttgggaagt ttttcactat agataaaaaa atgtccttga ctagcgtttc atacaaaaaa 4260
aaaaaaaaac gcaaccaaaa atgttaatgt ggttcagtga aaacccttca aaaagtgata 4320
tctatttttt tacaggaact gatcgcaaag tatgtttttt ttttttttg cgttggtttt 4380
tacaattaca ccaagtcact ttgattaaag aggaagtaaa ctaagatagt gtctcaatgt 4440
tggataggtc atttagaaaa ggtccgcgag attggatcca taataatgat tctcctctct 4500
aactaatttc tccttcattt gattctatca cagagttaca acctatccag taaatctttt 4560
ccaggcgctc taacctaggt attattacta agaggagaga cactgatccg catctgtggg 4620
atggacaacg tttgtaattt ctatcggtat cgaaaataat cgcgcatttt cgggcgtatt 4680
ccagaaaaca acaatgaaat gtgactaggc gtagacaccc tacctgttgc aaacattaaa 4740
gatagccata gcttttatta gcgcgtaaaa gcccgcataa ggtcttttgt tgttacttta 4800
atactgaagc aaatgtgcac aattttcatt acatgatatt attcaatggg gtaggtgggc 4860
gacaaaatag attcattaat gttggataat aggggcgttt tatgacttcg tttacacgtg 4920
ttaaaagtaa tgtactataa taagttaccc catccacccg ctgttttatc taagtaatta 4980
caacctatta tccccgcaaa gtcattatcc ctaaatgctc cacctcagct ggtggccccg 5040
tcagtcagtt gatcgggaaa gcagcaatca atccggagac aggtcgacct ccatcgaaca 5100
cagtaatagg gatttacgag gtggagtcga ccaccggggc agtcagtcaa ctagcccttt 5160
cgtcgttagt taggcctctg tccagctgga ggtagcttgt ggaaccgaac aacactagat 5220

gttcgatttc taacgaccga ctaagaacat cgtcggaagc gtctggttca ttcgacgagc 5280
cggaaggggt tcatctttcg ccttggcttg ttgtgatcta caagctaaag attgctggct 5340
gattcttgta gcagccttcg cagaccaagt aagctgctcg gccttcccca agtagaaagc 5400
ctcgtcgtcg aacgaatagc tgctgctaca cttcgcgtcg ttatcgtcgt cgggggattg 5460
gtgtttgtaa ctgcgcactc gtttatacat tgttgtttgc gagcagcagc ttgcttatcg 5520
acgacgatgt gaagcgcagc aatagcagca gccccctaac cacaaacatt gacgcgtgag 5580
caaatatgta acaacaaacg cgatcggcgg gcgctgtaac tgcctgcagt cacgcgttca 5640
ttcgcagtcg ttgtcgtagt catacacacg ccgtcgttcc tttgtatcag ctgtgtagca 5700
gctagccgcc cgcgacattg acggacgtca gtgcgcaagt aagcgtcagc aacagcatca 5760
gtatgtgtgc ggcagcaagg aaacatagtc gacacatcgt tttagtggtg ttacaacatt 5820
gagctacttt ttgcgtttcg ctttcgtgct gcggcggcgg cggcgggact tcgctgcact 5880
gataggaacg gaatgcatgc aaatcaccac aatgttgtaa ctcgatgaaa aacgcaaagc 5940
gaaagcacga cgccgccgcc gccgccctga agcgacgtga ctatccttgc cttacgtacg 6000
tgctccggtt gaagagagct ctgcgccact tgtggcgggt ttcactcaaa aggcatcgtc 6060
gcgtcgcaac aaagtgcgca cattcgacgc gtaactgtaa acgaggccaa cttctctcga 6120
gacgcggtga acaccgccca aagtgagttt tccgtagcag cgcagcgttg tttcacgcgt 6180
gtaagctgcg cattgacatt gtaaatagaa agactttggt gcgtttagaa aaagggtcac 6240
aaagggtggc aagtgagtat gtatgtgagc tcatttcatt ctcgatggca ttgagacgta 6300
catttatctt tctgaaacca cgcaaatctt tttcccagtg tttcccaccg ttcactcata 6360
catacactcg agtaaagtaa gagctaccgt aactctgcat atctattctg agaacgaaag 6420
ttcaatggat gcattttatg caatgccacc ggaattttcc tatgaactgc tttcacactt 6480
cttttaagaa aattttgcag tagataagac tcttgctttc aagttaccta cgtaaaatac 6540
gttacggtgg ccttaaaagg atacttgacg aaagtgtgaa gaaaattctt ttaaaacgtc 6600
atttaattta ttcactccat ttagttctga cgtaacattc cagataacac acttcaaagt 6660
catggtcagt tcatgttgaa cgaatgtgca ccgcgatcca taaattaaat aagtgaggta 6720
aatcaagact gcattgtaag gtctattgtg tgaagtttca gtaccagtca agtacaactt 6780
gcttacacgt ggcgctaggt cgcagaacga ttccatgtct taatgtcgtc acttatcata 6840
taatcaccca gtttttgccc cacttaaaaa aacgatgtcc acttttttatc tgagtttctt 6900
gcgtcttgct aaggtacaga attacagcag tgaatagtat attagtgggt caaaaacggg 6960
gtgaattttt ttgctacagg tgaaaaatag actcaaagaa tctcctctct tttcagccaa 7020
ccactccagc ggaacccctg aacccggaaa catggtacca ggtgagttcg ctgttgaaat 7080
actaatttgc agaaaacata agaggagaga aaagtcggtt ggtgaggtcg ccttggggac 7140
ttgggccttt gtaccatggt ccactcaagc gacaacttta tgattaaacg tcttttgtat 7200
agaaattttg ctaccgattt accataactg gaatcgaaga caatatgact tcatcacacc 7260
agcagtaaac acggcgtaaa aatgattcat caggacccgc tctttaaaac gatggctaaa 7320
tggtattgac cttagcttct gttatactga agtagtgtgg tcgtcatttg tgccgcattt 7380
ttactaagta gtcctgggcg tcaatagccc tgtttttcca cgctcatctt gggtttcaca 7440
tcggtgaaca ccacttggag acgttttcac acaatgttca tgttcttctt tgagtaaatg 7500
agttatcggg acaaaaaggt gcgagtagaa cccaaagtgt agccacttgt ggtgaacctc 7560
tgcaaaagtg tgttacaagt acaagaagaa actcatttac aagttatgcg tggtcccgtg 7620
ctcatcaaga tagtgtgcca cacataagaa ttatcttaat tgaggccttc tgcgggccgt 7680
gagcttgttt gctacgccct ttcaatacgc accagggcac gagtagttct atcacacggt 7740
gtgtattctt aatagaatta actccggaag acgcccggca ctcgaacaaa cgatgcggga 7800
tccttggcgt tgagttttag tttctttgac agagaaagac ttttgataat ctactttctg 7860
cagctacgac ctttctctga actatttgga aaattataac aggaaccgca actcaaaatc 7920
aaagaaactg tctctttctg aaaactatta gatgaaagac gtcgatgctg gaaagagact 7980
tgataaacct tttaatattg ttatgttgac aatatttatc ccttcgatta acaaaaaact 8040
tcaagccagg gaaacatcca gtgtgaaaac actaagcggc gcactttggt tcatttcatt 8100
aatacaactg ttataaatag ggaagctaat tgtttttga agttcggtcc ctttgtaggt 8160

cacacttttg tgattcgccg cgtgaaacca agtaaagtaa cgtatcgatc actcttaatt 8220
caagatgaca aagtggttga gtagtagagt acgtggctca caatcggaag gttcttggct 8280
cgaatctcaa tgtatgctat gcatagctag tgagaattaa gttctactgt ttcaccaact 8340
catcatctca tgcaccgagt gttagccttc caagaaccga gcttagagtt acatacgata 8400
ttttaacttt ttttttattt tgtcgatcat aaacggatgc gcgactcagc atttttggca 8460
tttgaatcat gattccgagt aatcagctac aaaaacctaa aaaattgaaa aaaaaataaa 8520
acagctagta tttgcctacg cgctgagtcg taaaaaccgt aaacttagta ctaaggctca 8580
ttagtcgatg tttttggatt cgcgtgtgtt gcgttacggc aatctgactc atgatatcat 8640
gagtccaaat catggtgtat tttcataaga cgaaaacacg ctggaatcat gatatcatga 8700
gcgcacacaa cgcaatgccg ttagactgag tactatagta ctcaggttta gtaccacata 8760
aaagtattct gcttttgtgc gaccttagta ctatagtact gtaataatct tgtttttgga 8820
ttctgatttc tacccgtgca tttctaaagt ttgcaaagaa ggaagcttca aaaaacttcc 8880
aaaagcttat gttacagaag cattattaga acaaaaacct aagactaaag atgggcacgt 8940
aaagatttca aacgtttctt ccttcgaagt tttttgaagg ttttcgaata caatgtcttc 9000
cttggaaagc ttaagttaca gcagtttccg taccagaacg ttggaaagct tatattacga 9060
aacagtaata gggtttctat gcggtggaag tgctgttata gaacctttcg aattcaatgt 9120
cgtcaaaggc atggtcttgc aacctttcga atataatgct ttgtcattat cccaaagata 9180
cgccaccttc acgacaatat tggcgtgtaa gcatttataa tacatctggg tatcatcgaa 9240
atcattagaa aaaatgcggt ataagtttca cttgaattca gatcagtgat cgattgttac 9300
accgcacatt cgtaaatatt atgtagaccc atagtagctt tagtaatctt ttttacgcca 9360
tattcaaagt gaacttaagt ctagtcacta gctaacaatg agttcaaata gatccaaata 9420
tatgagggtg aaacgtcatt gcgatccact gtgaactgca gttgattggc cgcaatttca 9480
aaatatgtac acccgagtga tcaagtttat ctaggtttat atactcccac tttgcagtaa 9540
cgctaggtga cacttgacgt caactaaccg gcgttaaagt tttatacatg tgggctcact 9600
tctgcacggc tgttcagctg acatccttca ttgtcccagt cgttcataca aacttgcccg 9660
tcaagatcaa ggaagttggc gcttgatcaa tgttctgttt agacgtgccg acaagtcgac 9720
tgtaggaagt aacagggtca gcaagtatgt ttgaacgggc agttctagtt ccttcaaccg 9780
cgaactagtt acaagacaaa catttctttt ttcttaagta gtattgggcg ctgcggtcac 9840
ctcatttatc tttttgaaat tgtttcggaa ataatgcacg agatgcaata acggttcttg 9900
gtaaagaaaa aagaattcat cataacccgc gacgccagtg gagtaaatag aaaaacttta 9960
acaaagcctt tattacgtgc tctacgttat tgccaagaac aacatagtca tgtagaacct 10020
tacaaatgat cagaattgat ttgatcaatt catttccagc tttcaaactg acgatcgccc 10080
aatgctaccg tccatcacga ttgtatcagt acatcttgga atgtttacta gtcttaacta 10140
aactagttaa gtaaaggtcg aaagtttgac tgctagcggg ttacgatggc aggtagtgct 10200
tattccacgc actggctgtc atgttccctg ccagatttac gtagtgttct tttgtaaagg 10260
caacactgct gcactgctcc aagtcactcc aagcttcatc ataaggtgcg tgaccgacag 10320
tacaagggac ggtctaaatg catcacaaga aaacatttcc gttgtgacga cgtgacgagg 10380
ttcagtgagg ttcgaagtag tgcgagttga agcaaactgt gaaggattga tattttgaat 10440
taaatcaagc tctcgcgttg caggcagctg taacttgcca ccaagtatga tcggtcttcc 10500
acgctcaact tcgtttgaca cttcctaact ataaaactta atttagttcg agagcgcaac 10560
gtccgtcgac attgaacggt ggttcatact agccagaagg gacttcgttc cataaaaagt 10620
ggaatgctcc tcgtccgatt tccagaaaca gtcggttatg caataaaaca ggatcaggtt 10680
cgatgactct tggcgatatc ctgaagcaag gtatttttca ccttacgagg agcaggctaa 10740
aggtctttgt cagccaatac gttattttgt cctagtccaa gctactgaga accgctatag 10800
tgaattggag tcgttaccta tcccccgata aagatatcct ctcgcaattc gaggggggatt 10860
aggattagaa accgtttgct gatatttgcg agatataaaa acttaacctc agcaatggat 10920
agggggctat ttctatagga gagcgttaag ctccccctaa tcctaatctt tggcaaacga 10980
ctataaacgc tctatatttt actaataaaa tcttcaattc gctaaaagca cttcaattct 11040
tgttttctct tctggtttca gttgaccccc atatgcgagt gcagcatcac ggaccggact 11100

tgattatttt agaagttaag cgattttcgt gaagttaaga acaaaagaga agaccaaagt 11160
caactggggg tatacgctca cgtcgtagtg cctggcctga caggaacagg tgcgtacttc 11220
cttaacttca ctatcaataa aaccgtacct cctccagtcc atcgaaacaa caataaaata 11280
ctgcaccgat cagctggaat gtccttgtcc acgcatgaag gaattgaagt gatagttatt 11340
ttggcatgga ggaggtcagg tagctttgtt gttattttat gacgtggcta gtcgacctta 11400
ttctatcccg ggaggtccaa tcgctacaat ttatgcacat ttaattccac tggagccatg 11460
tgcgttcggg catcttatca ggcgttcggg aattgaaact aagatagggc cctccaggtt 11520
agcgatgtta aatacgtgta aattaaggtg acctcggtac acgcaagccc gtagaatagt 11580
ccgcaagccc ttaactttga ttacgacctc atttgtcatt aacgggatgc attcgtacgc 11640
agtcagcgtc ttatcggcat atatgcggta gccccccgag tgacaattaa accatggagc 11700
aatgctggag taaacagtaa ttgccctacg taagcatgcg tcagtcgcag aatagccgta 11760
tatacgccat cgggggggctc actgttaatt tggtacctcg cgaaaccaat ttcacagcgg 11820
tccaccaact accgaatgcg atgcattttt atacgacagt ggcgttacta ggtgcttaac 11880
atatcaaaac ttggaagctt gctttggtta aagtgtcgcc aggtggttga tggcttacgc 11940
tacgtaaaaa tatgctgtca ccgcaatgat ccacgaattg tatagttttg aaccttcgaa 12000
cctttcaaaa gcttgcaaag cttccttcca ggagcttgga aagcttcctt ccaggagctt 12060
ggaaagcttc cttccaggag cttggaaagc ttccttccag ggaaagtttt cgaacgtttc 12120
gaaggaaggt cctcgaacct ttcgaaggaa ggtcctcgaa cctttcgaag gaaggtcctc 12180
gaacctttcg aaggaaggtc gagcttggaa agcttccttc caggagcttg gaaagcttcc 12240
ttccagtagc ttggaaagct tccttccagg agcttggaaa gcttccttcc aggagcttgg 12300
ctcgaacctt tcgaaggaag gtcctcgaac ctttcgaagg aaggtcatcg aacctttcga 12360
aggaaggtcc tcgaaccttt cgaaggaagg tcctcgaacc aaagcttcct tccaggagct 12420
tggaaagctt ccttccagga gcttggaaag cttccttcca ggagcttgga aagcttcctt 12480
ccaggagctt ggaaagcttc tttcgaagga aggtcctcga acctttcgaa ggaaggtcct 12540
cgaacctttc gaaggaaggt cctcgaacct ttcgaaggaa ggtcctcgaa cctttcgaag 12600
cttccaggag cttggaaagc ttccttccag gagcttggaa agcttccttc caggagcttg 12660
gaaagcttcc ttccaggagc ttggaaagct tccttccagg gaaggtcctc gaacctttcg 12720
aaggaaggtc ctcgaacctt tcgaaggaag gtcctcgaac ctttcgaagg aaggtcctcg 12780
aacctttcga aggaaggtcc agcttggaaa gcttccttcc aggagcttgg aaagcttcct 12840
tccaggagct tggaaagctt ccttccagga gcttggaaag cttccttcca ggagcttgga 12900
tcgaacctt cgaaggaagg tcctcgaacc tttcgaagga aggtcctcga acctttcgaa 12960
ggaaggtcct cgaacctttc gaaggaaggt cctcgaacct aagcttcctt ccaggagctt 13020
ggaaagcttc cttccaggag cttggaaagc ttccttccag gagcttggaa agcttccttc 13080
caggagcttg gaaagcttcc ttcgaaggaa ggtcctcgaa cctttcgaag gaaggtcctc 13140
gaacctttcg aaggaaggtc ctcgaacctt tcgaaggaag gtcctcgaac ctttcgaagg 13200
ttccaggagt ggaaaagatt cctgaaaagt acttggagaa attcctcgag ttatttcagt 13260
aaagattata ctggaggaac caatggtgga atcacttgag aaggtcctca ccttttctaa 13320
ggacttttca tgaacctctt taaggagctc aataaagtca tttctaatat gacctccttg 13380
gttaccacct tagtgaactc gcatttcggc agaaatcccc ggcaaaatcg ctatggaaaa 13440
atcccctgcaa aaaatcctgg aataatcctt gccggaatct catgaggaac tcctggtaaa 13500
cgtaaagccg tctttaggga ccgttttagc gataccttt tagggacgtt ttttaggacc 13560
ttattaggaa cggccttaga gtactccttg aggaccattt attctttaac aaatttctgt 13620
ttatttctc tacaaagtta cagctccttt accgtgccga ttggccagaa atgaccccaa 13680
agactcatgg ggtacgatct taagaaattg tttaaagaca aataaaagag atgtttcaat 13740
gtcgaggaaa tggcacggct aaccggtctt tactggggtt tctgagtacc ccatgctaga 13800
tatttctgcc aaatatactg tatgtttgtt tctttctgat atgcttttaa gctcaatttt 13860
ctttggaatg gtggagattt gttttggcct ccaatatact ataaagacgg tttatatgac 13920
atacaaacaa agaaagacta tacgaaaatt cgagttaaaa gaaaccttac cacctctaaa 13980
caaaaccgga ggttatatga tgctagctcg tagttcgtac ctgaagtcaa ctcctcaatt 14040

cctaaatgct acaataatat ataaaatttt aggaaataac tgcaaaatat tctgaaggcc 14100
acgatcgagc atcaagcatg gacttcagtt gaggagttaa ggatttacga tgttattata 14160
tattttaaaa tcctttattg acgttttata agacttccgg atgtcttgat ctatcttgat 14220
gtatctaata tgtaatccca gaagcattct agtttttct gataatctgt gaaataagtt 14280
gtttttacga actttgactt tacagaacta gatagaacta catagattat acattagggt 14340
cttcgtaaga tcaaaaaaga ctattagaca ctttattcaa caaaaatgct tgaaactgaa 14400
ttcgggattt gaggtacaag ctttcaaata tattggaggt tctgcgatat taacttcaat 14460
gaattattgg aaattagaaa tcgtcttgtg catacgggtt aagccctaaa ctccatgttc 14520
gaaagtttat ataacctcca agacgctata attgaagtta cttaataacc tttaatcttt 14580
agcagaacac gtatgcccaa aatcgatttt agtctctggt agatttcgag agggaatgtc 14640
tgaagaaatt ttctgaccta catgtgaagt attgtctgtc aaattcaaaa tattttctgt 14700
ttagctaaaa tcagagacca tctaaagctc tcccttacag acttctttaa aagactggat 14760
gtacacttca taacagacag tttaagtttt ataaaagaca aggaaattaa aatttttgg 14820
ggaaaactcg aaactccttg gatatccaag gaaacaaaaa aaaaagaaat atctgaagaa 14880
gtgcatcgtc ctttttcctt tcctttaatt ttaaaaaacc cctttgagc tttgaggaac 14940
ctataggttc ctttgttttt tttttcttta tagacttctt cacgtagcag gaaaaaggaa 15000
aattattgtt ttaattaact aatagttctg ctagaaaggt ttttggcaga accccaaaat 15060
gatattcaaa gcaactaaca gctcgatttc ccctcgtttc ttaataacaa aattaattga 15120
ttatcaagac gatctttcca aaaaccgtct tggggtttta ctataagttt cgttgattgt 15180
cgagctaaag gggagcaaag caatttcaga cgacgaactt gtcaaacgat ctcaatggct 15240
cctggagaag ctgcgatacc cctgggagat gatgcccctg atgtacgtga tactgaaagg 15300
gttaaagtct gctgcttgaa cagtttgcta gagttaccga ggacctcttc gacgctatgg 15360
ggaccctcta ctacggggac tacatgcact atgactttcc cgccgacgga gacgtcaata 15420
aagcgcgcca acggattgac gaaggtatgg gggttcttac cggttgggac tgtttccgag 15480
gtatcgatcg ggtgtcactc gcggctgcct ctgcagttat ttcgcgcggt tgcctaactg 15540
cttccatacc cccaagaatg gccaaccctg acaaaggctc catagctagc ccacagtgag 15600
acttcctggg tgctcccatt ttgtaactgc taacgcttat tattgagttt caggacatct 15660
gggatcttcg gtcgacggag tctattccca acagtgccct tgaaggaccc acgagggtaa 15720
aacattgacg attgcgaata ataactcaaa gtcctgtaga ccctagaagc cagctgcctc 15780
agataagggt tgtcacggga ggatcaaaca ctgccatcat gcagtttccg tagcctgttg 15840
ggctacgctc cccgacttga catcccccat tcttatcaaa caacaactca aggcctgaga 15900
cctagtttgt gacggtagta cgtcaaaggc atcggacaac ccgatgcgag gggctgaact 15960
gtaggggta agaatagttt gttgttgagt tccggactct caacgagtgg tggaatttgc 16020
gcacgaagtc attggtttgt cctggtaaaa gttaaaaggg ttaactggag ggttaattga 16080
cacggtttca actgatggcc gttgctcacc accttaaacg cgtgcttcag taaccaaaca 16140
ggaccatttt caatttccc aattgacctc ccaattaact gtgccaaagt tgactaccgg 16200
ttattgacac acggatgaaa gacttgcacg cttgaccttc tgtctgtact aataaaagtt 16260
acgttggctg ggttttgggg tcataatggc cccaaaatcg aataactgtg tgcctacttt 16320
ctgaacgtgc gaactggaag acagacatga ttattttcaa tgcaaccgac ccaaaacccc 16380
agtattaccg gggttttagc aatcgtcata acttcttgaa atacaactca cgtttaagac 16440
cattcaagag tattagatca tcgtctataa tagcagattt gaaatttact tcacatttcg 16500
ttagcagtat tgaagaactt tatgttgagt gcaaattctg gtaagttctc ataatctagt 16560
agcagatatt atcgtctaaa ctttaaatga agtgtaaagc gtattgcagt gccccttgct 16620
tccacaatgg aattagttaa agtttcgaga gcattgtcaa tatcaagtgt tgttagcaaa 16680
caaatgctaa catcaagatt cataacgtca cggggaacga aggtgttacc ttaatcaatt 16740
tcaaagctct cgtaacagtt atagttcaca acaatcgttt gtttacgatt gtagttctaa 16800
actatcgatg tttgattcac atgtattcca atcagctcgt aaaaaatgga aagtggagct 16860
gatagggttg agaatcgctt catgggataa ttggaaacag tgatagctac aaactaagtg 16920
tacataaggt tagtcgagca ttttttacct ttcacctcga ctatcccaac tcttagcgaa 16980

gtaccctatt aacctttgtc ggacatgatc agaatgaaaa tcagcgtgag taaccagttg 17040
actacaaaga tgactagagt cggttaagaa aaattcaagt agggctatca ggttattgaa 17100
cctgtactag tcttactttt agtcgcactc attggtcaac tgatgtttct actgatctca 17160
gccaattctt tttaagttca tcccgatagt ccaataactt ttgaaaaata tcccgaaggg 17220
ccctcatcaa ttaaaatttt gcctttggaa atgtttggca ttcaagtagc aaatttttaac 17280
atactgcgat tcgatttccg aacttttttat agggcttccc gggagtagtt aattttaaaa 17340
cggaaacctt tacaaaccgt aagttcatcg tttaaaattg tatgacgcta agctaaaggc 17400
caagttagtt tgaaacaaat taacttgcta cccagtgcat taaaaaggca agtaggcagc 17460
tttggaagta taaacttagc tgtgttttaa cagaagcact gttcaatcaa actttgttta 17520
attgaacgat gggtcacgta atttttccgt tcatccgtcg aaaccttcat atttgaatcg 17580
acacaaaatt gtcttcgtga cgcaagtttc aaaaattttg gtttcgaatg acaaaaaaag 17640
ttgatgttat atacgcctat tgaatgatga ttccagttga tcatttcgac aaacaaaaaa 17700
gcgttcaaag tttttaaaac caaagcttac tgtttttttc aactacaata tatgcggata 17760
acttactact aaggtcaact agtaaagctg tttgtttttt gaatctcttt tgatttcaga 17820
tccaggattc aaataacatt ccgttatcag ataaagggtt aatgccacaa tcgtgtggtc 17880
cattatcccc ggaaacttca cttagagaaa actaaagtct aggtcctaag tttattgtaa 17940
ggcaatagtc tatttcccaa ttacggtgtt agcacaccag gtaatagggg cctttgaagt 18000
caccgtcaca ctcgatccag atctgatgtg atctctgccg tcgggcgcct cagaagcgaa 18060
aaccacattc gcccgcgctc tccggaatta tgtcgtaaaa gtggcagtgt gagctaggtc 18120
tagactacac tagagacggc agcccgcgga gtcttcgctt ttggtgtaag cgggcgcgag 18180
aggccttaat acagcatttt taaaacttta caaccataat tattcagaac ttcgacgact 18240
gcgcgatgac ttggccgcgg tgtgcctgct tgggatggac ctccgagcac tgaaagcagt 18300
attttgaaat gttggtatta ataagtcttg aagctgctga cgcgctactg aaccggcgcc 18360
acacggacga accctacctg gaggctcgtg actttcgtca ggtttgtaca aattgaatgg 18420
gctatttgaa attaattggg ctgcgataac ttcaaagtgt gacatcaaaa tggtgtgagt 18480
tttttactgc acaaattcca ccaaacatgt ttaacttacc cgataaactt taattaaccc 18540
gacgctattg aagtttcaca ctgtagtttt accacactca aaaaatgacg tgtttaaggt 18600
agttatttcc tacttcatat caatcggagc tccaggagtg aagatccaaa ttaccaagct 18660
tggccatttc gtatgaaaaa cggcaaaatg atctttttttt tcaataaagg atgaagtata 18720
gttagcctcg aggtcctcac ttctaggttt aatggttcga accggtaaag catacttttt 18780
gccgttttac tagaaaaaaa cgccagtcac tgtatctcat gatccagatg agataaaaaa 18840
gttcgagtct tcgacaaagt tgttttggaa gtcatggaca ttcttaagca aacaacttag 18900
gcggtcagtg acatagagta ctaggtctac tctatttttt caagctcaga agctgtttca 18960
acaaaacctt cagtacctgt aagaattcgt ttgttgaatc ttttgccact aggtggcgcc 19020
agtaagcata ttcgtcatca aacgtcaaca tcccaccgca aaatcgctag tgtttggagg 19080
ggattttaac ctccaaattg aaaacggtga tccaccgcgg tcattcgtat aagcagtagt 19140
ttgcagttgt agggtggcgt tttagcgatc acaaacctcc cctaaaattg gaggtttaac 19200
ccaaataacc tccaaatcat cacctccaag ttagttctaa tacactccgt tatatgaaat 19260
atggtggtgc gtcgatcgtc gcaagtttat cgttaaacag ggtttattgg aggtttagta 19320
gtggaggttc aatcaagatt atgtgaggca atatacttta taccaccacg cagctagcag 19380
cgttcaaata gcaatttgtc tcaataaaat gagcatttta tatcgtgata catatgagaa 19440
gatagaggtt tcaattaaaa caaatccaca tggtgtcgct aataaaattg tgcattttaa 19500
agttatttta ctcgtaaaat atagcactat gtatactctt ctatctccaa agttaatttt 19560
gtttaggtgt accacagcga ttattttaac acgtaaaatt gcgagttata tcctctgatc 19620
aagataaaat agaaaattcg atttttgaat attcaattat aagagcctga ataactacaa 19680
catgtagtga atcgaaactg cgctcaatat aggagactag ttctatttta tcttttaagc 19740
taaaaactta taagttaata ttctcggact tattgatgtt gtacatcact tagctttgac 19800
atttatgacg gtttgtgaag gttacacgtc ctaagcattt ggattcaaga aaagcaagag 19860
atatgacgaa tgtaaacttt atcgtatcaa tgaagtaact taaatactgc caaacacttc 19920

caatgtgcag gattcgtaaa cctaagttct tttcgttctc tatactgctt acatttgaaa 19980
tagcatagtt acttcattga agcgtccaga acagtacaaa ccaacatcgt accgtcgtat 20040
tccactccgg tcgttgcaat atctctaggt ccaccgaaaa acactcatga ccaagatcgt 20100
tcgcaggtct tgtcatgttt ggttgtagca tggcagcata aggtgaggcc agcaacgtta 20160
tagagatcca ggtggctttt tgtgagtact ggttctagca gtcgtcgatc ttggtccacc 20220
gaaacaccga tgtccatatc gtttcgtcga acttggacca acgattcatg caactgatga 20280
caacgcggcc cccgggtcgt cagcagctag aaccaggtgg ctttgtggct acaggtatag 20340
caaagcagct tgaacctggt tgctaagtac gttgactact gttgcgccgg gggcccagca 20400
accaatatcc gaaaaatcca actgttcttc tctgcctcgc aggtcaagcc gtggtcaatg 20460
aatactcacg attgcacaat ctgaacatgt tcgacggtgt tggttatagg ctttttaggt 20520
tgacaagaag agacggagcg tccagttcgg caccagttac ttatgagtgc taacgtgtta 20580
gacttgtaca agctgccaca agagttgcgc agtacgacgc gccagtccgg atgatagact 20640
ttttacacga tcagcacgac ccactgcgct gcggcaaagg tcgaaccgaa acaagaataa 20700
tctcaacgcg tcatgctgcg cggtcaggcc tactatctga aaaatgtgct agtcgtgctg 20760
ggtgacgcga cgccgtttcc agcttggctt tgttcttatt accacgaaga tcagatcgat 20820
tcgacggaag aagcaatcga atgcaaagaa gaatcggaac gaagaaaact ctaaagcatc 20880
gcatatttac aaagcataac tggtgcttct agtctagcta agctgccttc ttcgttagct 20940
tacgtttctt cttagccttg cttcttttga gatttcgtag cgtataaatg tttcgtattg 21000
ggaaaacccg caagttcaaa ctagtgatta gtgtaagatg aagcaaagca gaaatgtagt 21060
atctagattt ttcgacgtta gtttacaaag ataaaaaatg ccttttgggc gttcaagttt 21120
gatcactaat cacattctac ttcgtttcgt ctttacatca tagatctaaa aagctgcaat 21180
caaatgtttc tatttttac aggttggaca tacaatcgtg ggtattcgtc tgagttcgtc 21240
acaactgcac cggaaactgt gaaacagaat agagccaacc tgtgcgcgga gaatgttgag 21300
tccaacctgt atgttagcac ccataagcag actcaagcag tgttgacgtg gcctttgaca 21360
ctttgtctta tctcggttgg acacgcgcct cttacaactc gtcattataa gcttccttag 21420
catccacggg tgaaagtcga tcgacggaag cctgcaagac tctgtcgatg ggctttcgtc 21480
ctagaagaat aagattaaac cagtaatatt cgaaggaatc gtaggtgccc actttcagct 21540
agctgccttc ggacgttctg agacagctac ccgaaagcag gatcttctta ttctaatttg 21600
ctgaaatgta ttctcccgtg gaatggtttc atttgagtaa ttctgtatct tctccttccc 21660
aattccacga acgcgacgaa ctctaataca aacaacataa gactttacat aagagggcac 21720
cttaccaaag taaactcatt aagacataga agaggaaggg ttaaggtgct tgcgctgctt 21780
gagattatgt ttgttgtatt tgaccacagt gcaaatgctg tttaacgata atagcgacat 21840
gcagccattc tggggctacc acgtgtagct ctacttgtga gacagcgttc ctaaagagtg 21900
actggtgtca cgtttacgac aaattgctat tatcgctgta cgtcggtaag accccgatgg 21960
tgcacatcga gatgaacact ctgtcgcaag gatttctcac tgaaagtgca aacaagtgat 22020
gaaaccaata gtgcaaagca agtttagagg gaaaatttaa aaaatgcaaa acagcagtag 22080
tacttaactt ttaagattgt actttcacgt ttgttcacta ctttggttat cacgtttcgt 22140
tcaaatctcc cttttaaatt ttttacgttt tgtcgtcatc atgaattgaa aattctaaca 22200
gtttcgaaag ccgaagtgtg ttccatctgc caccggaaaa aaacgacgac agcagaatca 22260
tcaacaagca acatccatcc gaaaaaatcc gggaaaccgg caaagctttc ggcttcacac 22320
aaggtagacg gtggcctttt tttgctgctg tcgtcttagt agttgttcgt tgtaggtagg 22380
ctttttttagg ccctttggcc atcttcaacc aaccatccta caatctacaa accagagatt 22440
atatctcttc aatcgtttcc gacatcggtc ggtttcggtg cccaaaatga tctgataaac 22500
tagaagttgg ttggtaggat gttagatgtt tggtctctaa tatagagaag ttagcaaagg 22560
ctgtagccag ccaaagccac gggttttact agactatttg acttatctct ctgtagcttg 22620
catgccattg cgagcgtatt ttggtagctg gccgttgcca aacggctccg acaggtactg 22680
ctattggagg ttgtgcacga tgaatagaga gacatcgaac gtacggtaac gctcgcataa 22740
aaccatcgac cggcaacggt ttgccgaggc tgtccatgac gataacctcc aacacgtgct 22800
ccacgttgag tttgcctttt gagttggaga gtgtgtcttt tcgtcatata tttggccttt 22860

tcaagggtga ttttcaggct gcgtaaagat tgtatagttt ggtgcaactc aaacggaaaa 22920
ctcaacctct cacacagaaa agcagtatat aaaccggaaa agttcccact aaaagtccga 22980
cgcatttcta acatatcaaa aaccagctaa aacatattga tgacaagttc tatttcagca 23040
ccacaaacaa gcctgttaat gtctctcacc gcaaccattg ttctgcgcgc gttataatca 23100
ttggtcgatt ttgtataact actgttcaag ataaagtcgt ggtgtttgtt cggacaatta 23160
cagagagtgg cgttggtaac aagacgcgcg caatattagt gcatagaagt ttatttttctt 23220
tgggatgatt caaatattac gtgacgcaaa gtttgccaat tttagaaccc ctccctcctc 23280
cacgtaacgg cttttgtgtg cgtatcttca aataaaagaa accctactaa gtttataatg 23340
cactgcgttt caaacggtta aaatcttggg gagggaggag gtgcattgcc gaaaacacac 23400
aaaaatttaa attttgtgta tagaccgtag catttcggaa gacccccctcc cttactctgt 23460
tgagttacgt aaaatttcaa cgatcctttt gtagttctga tttttaaatt taaaacacat 23520
atctggcatc gtaaagcctt ctgggggagg gaatgagaca actcaatgca ttttaaagtt 23580
gctaggaaaa catcaagact attttatatc agcgtgcagt gttatgaaga tatccacagt 23640
ataaaatatt attttatttt aaattctatg ctgattatca atgtgttact agtggctttt 23700
taaaatatag tcgcacgtca caatacttct ataggtgtca tattttataa taaaataaaa 23760
tttaagatac gactaatagt tacacaatga tcaccgaaaa catactcatg ttgcgagctc 23820
gatttggcgc acggggtcat ctacacctga tacctttagg gtcgttgggg gaccacttag 23880
cgtgcacgta cggacattca gtatgagtac aacgctcgag ctaaaccgcg tgccccagta 23940
gatgtggact atggaaatcc cagcaacccc ctggtgaatc gcacgtgcat gcctgtaagt 24000
aaatgttgtt caaatttttt tcttaccaag acgagcactt tacaatgaca aactctggct 24060
ctgctctggc tctgctctgg ctctgctctg gctctgctct tttacaacaa gtttaaaaaa 24120
agaatggttc tgctcgtgaa atgttactgt ttgagaccga gacgagaccg agacgagacc 24180
gagacgagac cgagacgaga ggctctgctc tggctctgct ctggctctgc tctggctctg 24240
ctctggctct gctctggctc tgctctggct ctgctctggc tctgctctgg ctctgctctg 24300
ccgagacgag accgagacga gaccgagacg agaccgagac gagaccgaga cgagaccgag 24360
acgagaccga gacgagaccg agacgagacc gagacgagac gctctgctct ggctctgctc 24420
tggctctgct ctggctctgc tctggctctg ctctggctct gctctggctc tgctctggct 24480
ctgctctggc tctgctctgg cgagacgaga ccgagacgag accgagacga gaccgagacg 24540
agaccgagac gagaccgaga cgagaccgag acgagaccga gacgagaccg agacgagacc 24600
ctctgctctg caaaatgctc tggattaatt tattgctcac actcttttgc tgttggacca 24660
ctattcattt caaatcttca atatgttcct attaccccca gagacgagac gttttacgag 24720
acctaattaa ataacgagtg tgagaaaacg acaacctggt gataagtaaa gtttagaagt 24780
tatacaagga taatgggggt aacacggtcc acacggatcg atttcaacta actccactct 24840
cgtatgcata ttttgtgtat aaatttttgaa taatcgaaaa gggttgctgc aaatgttaat 24900
ttgtgccagg tgtgcctagc taaagttgat tgaggtgaga gcatacgtat aaaacacata 24960
tttaaaactt attagctttt cccaacgacg tttacaatta atttttttccc tctacccccct 25020
cactctgtcg ttggcgttgg aaaaaaatca ccactgcata caaaacactc attggttggg 25080
tggaaggacg gtttagcaga taaaaaaggg agatggggga gtgagacagc aaccgcaacc 25140
ttttttttagt ggtgacgtat gttttgtgag taaccaaccc accttcctgc caaatcgtct 25200
gttgctaaat tttccatatc acgctgattg atttgtgatt aaaaataaat ataaatagaa 25260
aatgaataat tcccacatgt gtttcggtat taggcaccgg caacgattta aaaggtatag 25320
tgcgactaac taaacactaa tttttattta tatttatctt ttacttatta agggtgtaca 25380
caaagccata atccgtggcc catggggcgg cgaagtgcag acggttctag ttctcattat 25440
ttggcatcga ttggcggtca aactacaacc tccatggaga aacaggcccc atccgtactt 25500
gtaccccgcc gcttcacgtc tgccaagatc aagagtaata aaccgtagct aaccgccagt 25560
ttgatgttgg aggtacctct ttgtccgggg taggcatgaa agttattaat aaataacaat 25620
gatttgaatt tgaatcattc atgctgcggc gtggctgatt tcggtgaatt gttgttctct 25680
tagagaaaga gggggatttg tcaataatta tttattgtta ctaaacttaa acttagtaag 25740
tacgacgccg caccgactaa agccacttaa caacaagaga atctctttct ccccctaaac 25800

aatttggacg agtaaataac attgaatatt acactttatg actaatcacc agtaatgaaa 25860
caacacgggt gatgatttca aaagcttcat tctaaatgca ttaaacctgc tcatttattg 25920
taacttataa tgtgaaatac tgattagtgg tcattacttt gttgtgccca ctactaaagt 25980
tttcgaagta agatttacgt tggttcactt ttggtggcag atttaaaact cttatcttcc 26040
tcttttcttc aacaggtttc acgccatcaa agacgcttgg cagccgcttc catttgcgta 26100
accaagtgaa aaccaccgtc taaattttga gaatagaagg agaaaagaag ttgtccaaag 26160
tgcggtagtt tctgcgaacc gtcggcgaag gtaaacgcat gcaaacgtat gttaacctta 26220
ggttttaatg ttaaaagtat caccaaaaat caagtcccaa gacttctgca agaatggttt 26280
atgctgaatt tattcgaaat cgtttgcata caattggaat ccaaaattac aattttcata 26340
gtggttttta gttcagggtt ctgaagacgt tcttaccaaa tacgacttaa ataagcttta 26400
ggttttattt tcatcgaaac atgtgtgatg taggctacta ttttggtaaa accgttggca 26460
acgactgtat ttaaactcac aaaatttgaa ccaaacttat ccaaaataaa agtagctttg 26520
tacacactac atccgatgat aaaaccattt tggcaaccgt tgctgacata aatttgagtg 26580
ttttaaactt ggtttgaata aattgtaact tttaattgag taaacatagg cgaaagagag 26640
tgattcaaat gggattcgga atcgaacggt tcttctaagt aagacaaacg aaaaaaacaa 26700
ttaacattga aaattaactc atttgtatcc gctttctctc actaagttta ccctaagcct 26760
tagcttgcca agaagattca ttctgtttgc ttttttgtt ccaaacgagt caaagctgca 26820
aaaacttcaa gtttgaactg tgatatcaat gaaattaaat acgaactatg tatcaagatt 26880
acagtaaaat ttaaagaaga ggtttgctca gtttcgacgt ttttgaagtt caaacttgac 26940
actatagtta ctttaattta tgcttgatac atagttctaa tgtcatttta aatttcttct 27000
ctttcaacgc atgaaacagg agggtggcaa ccgaaaagtg actgaatcaa ttgcgggtta 27060
tcattcgaga tatccagggg ttgaattgtg agaaaacttc gaaagttgcg tactttgtcc 27120
tcccaccgtt ggcttttcac tgacttagtt aacgcccaat agtaagctct ataggtcccc 27180
aacttaacac tcttttgaag ttcttcttct tattcttggc aatacgtcct cactgggata 27240
gagtctgctt cctaacttca tgttcaatga ccacttccac agttattaac tgagagcttt 27300
aagaagaaga ataagaaccg ttatgcagga gtgaccctat ctcagacgaa ggattgaagt 27360
acaagttact ggtgaaggtg tcaataattg actctcgaaa ctttgccaaa gttgccattt 27420
tcgcattcgt atatcgtgtg gcagcagtgt tgtgaaaaac tcaatttctc ataactaacg 27480
cttgagattt ttcatgcgtg gaaacggttt caacggtaaa agcgtaagca tatagcacac 27540
cgtcgtcaca acacttttg agttaaagag tattgattgc gaactctaaa aagtacgcac 27600
agttgtcaat cacgcaactc agcagtcaaa attttccaca gtatacttac acacggcaat 27660
aatttcttgc tagtctggta aaattatagt aatcttttct tcaacagtta gtgcgttgag 27720
tcgtcagttt taaaaggtgt catatgaatg tgtgccgtta ttaaagaacg atcagaccat 27780
tttaatatca ttagaaaaga aacgtaaaca acaaaattcg ggtttcaaga gtttttgacg 27840
ggagcaagca aaataggatt tagaattttg catgagacga agtttgaaaa ttttattgtc 27900
ttgcatttgt tgttttaagc ccaaagttct caaaaactgc cctcgttcgt tttatcctaa 27960
atcttaaaac gtactctgct tcaaactttt aaaataacag aaatttagta tcggttcaat 28020
cgaattttcg aacacaattg taggctctat ataaactaca tttattccct tattttgcca 28080
gatacaatac tcgcataact tttaaatcat agccaagtta gcttaaaagc ttgtgttaac 28140
atccgagata tatttgatgt aaataaggga ataaaacggt ctatgttatg agcgtattga 28200
tgagatctcg cctaaaaagc cattggtaac cgagtgtgta gctctttgtt tctaagccaa 28260
ttaatggacc tggatgaaaa ctatcatcac tgggaaatag actctagagc ggatttttcg 28320
gtaaccattg gctcacacat cgagaaacaa agattcggtt aattacctgg acctactttt 28380
gatagtagtg acccttttatc aggaggaact tgtctttatc gtagcattgt taaataacgt 28440
gtaaacccat ttgtttcctc ggtagctgca agctacacac tcgattacca atggctttta 28500
tcctccttga acagaaatag catcgtaaca atttattgca catttgggta aacaaaggag 28560
ccatcgacgt tcgatgtgtg agctaatggt taccgaaaat gggcgagatc acaagttatg 28620
cgagaatact tcccgaaatc accacctttt acccttttaa ataacgaaat tactacaaac 28680
ttcgttaccc gctctagtgt tcaatacgct cttatgaagg gctttagtgg tggaaaatgg 28740

gaaaatttat tgctttaatg atgtttgaag caat 28774

<210> 44
<211> 3399
<212> ADN
<213> Cydia pomonella
<220>
<221> dấu hiệu hỗn tạp
<222> (1179).. (1184)
<223> n là a, c, g hoặc t
<400> 44

catcagacgg gcccaggctc aggatgaagc tagagcgcgg gcggcggacg cagggctcca 60
ccctcccggg atcgagctag atcggcctga gccgccagtg gtgaaagcgc cgaggagtcc 120
cgtgatcccg ccgccgccgc cgcgctccat gggatcggcg agctgcgact ccgttccggg 180
atcgcccggg gtatcgccgt atgcgccgaa cccgccgtcc gctccgcctc cgccgatgcc 240
gccgctcccg cctccgcaac cagtggccct ggactccctg gtagaaaact gccacaagct 300
gctggaaaaa ttccactaca gttgggagat gatgccgctc gtgctggtca tcctcaacta 360
cgccggctcc gacctggagg aggcctcgcg gaagattgac gaaggtaagt ttaaatttaa 420
gtacataaca atgcttacag acgaattgaa agggaatgtg actcggctaa tccaccagga 480
tataattttg tagagtgcgc taaagaattc tagcaacgga cgctgttatt ctgccaccgc 540
cgttgatgcc gccgtcttct gatagtgata ctttaagatc cgtatactac gctcacttcc 600
attcacttat gtcgtacgga gtattaatat gggtaaactc gcggacacga aacgattacg 660
aaaacgcaga gtacttagat tggagcaaag cccagggatt cgccgagact tttttgttac 720
ggaagattga tgaaggtaag caagttggga ctgtggcgag ttgacacatg aaacaagtca 780
aggtcacagc tggagttcca ttaaagctgg atgctaccgc tagtcatcct gaggccggct 840
ccgacttcgt gcaatgaggt attaagctgc tggaattgaa tggaatatag tggtgaaaca 900
ctactactag gtttaagcgt ttagttatat ggttgttttc ttatttttaa tttttaaatg 960
ctctgctaag ctaaaacggc waatgtctat ttttgattat aaagacttat ataaaacaac 1020
ttgtttagct tctttkacgt ctttttgtta agctgtgccc tggttttaaa wkgggcgaac 1080
acytcacgaa taagacgtaa ttttaaaaag aaaatagata tcggccctct tggttcgcat 1140
ttatacatat gtattgctgc ccgtgcgaat gttggggann nnnnaaacag tacccctagt 1200
gtaartaaat tcgatttcga aacgtgacgt acgcgtttgc gtttagtctc mwtttgtatt 1260
ggatttagaa agagcgcgcc aagcgggacg ttttggaaac tcaaaatcct atacaaaatg 1320
agacttaacg caaasgcgtt tcgtcacgtt atgatgtcga tcaaatttac actaggggta 1380
cagaggtatt gcagtaactg tacaaatact aaactaaatt aataaattag ctaaatctaa 1440
aatataccct tcaggcattg tactaaggat gctggcggaa ttacttgtgc gaggaagccg 1500
ccagcttttc ggtcaccatt tacgagtacg tataccaaac gcttcgttgc tgcaaaaaag 1560
tttcaacgcc aaatggtaca aaatgcttta tattgttctc tatatattat attaacacat 1620
cgttatttta acctaggtct tagttatgta caaggttaca taaaatagat gttcctagtc 1680
cattcctccg tgtatgttgt gtctattata aagcaaggct gcattttgta atcagtcaat 1740
ttcaatataa aaaagttgca tcgtttttttt ttactkttcg acaattaaat tcaagtagca 1800
aaaaataacc caccttaatt tgtcatggtc ataatgaaac aatgacaarg ttttttttat 1860
cgcccgatac atgtacgtgt tctccaaaat gcagtctccg cgccgccaag cgaacgttca 1920
aactgtgcga tttccgttgt ccccaggcaa aatgatcatc aacgattacg ccaggaagca 1980
taatctgaac atcttcgacg ggctcgagct gaggaactcg acacgccact ccatttcgga 2040
tggcgatgaa aaacgcccac cgcaacctaa gcaagtctca aagtaaggtt ccatttaaat 2100
catctcaaaa ccgttagaaa cactcaaaaa gaaaccaaaa ttctgttcgg aaaccgacct 2160
ttgtttttta cacacactta gaccgaattt gcaaatttta acccccttatt cctaaaacta 2220
gcaatggtaa gctcggctga atttcacata caaacggagt ttcgttctca ttataaaact 2280

gcgtgttgga ttgtaatgga actttgcaca tacaatgaca tgaggtatgt ctagggctga 2340
aattagttta tacttggtat ctgaggctac ataaactaat tacagcctta gacttggagg 2400
atttaacaac tggaaacacc ttgtctgtaa ttctctgtac aacgatttta cgggggagga 2460
gcaaatatgt cagttaaacg tcagtccaaa caatacatat gactattggc cgtggtattt 2520
cgacggaggg gtaataagct cttaaaggcg actccgatat gcctaatcct attgttagta 2580
caaagtttca gagcaattta gctagtcgtt ttaaaatgag agcgtaacta cgttagcttg 2640
ctcttcttcc tcctgctctt atcccacgtt atgtggggtc ggcacaacat gttcctctct 2700
tctcactcct ttctttctca tatcctcttt cacacaatcc atccatcgtt tacttacaac 2760
cgagcttgct ggggaccgtt aaggcgccgc gagttcaggt tcttctctca ctctcactct 2820
cactggtgtg agcggagcga gacagcgttt tattttcgcc ttatcgaggt tccactgtat 2880
tataaataac ttacatttat aaagacgctg taatcgataa gaagttgagt cacgcttacg 2940
tcgcttacgt actacgtata gtaacgtagc ctgccgttta caaacaatgt acggagctac 3000
aacgttgcaa gttcggtccc cacacaacac aatgtgtcat aacacattaa caacattgtt 3060
acacacccac acatacaaat ttgctaagtt gataaaagag tggtgtgtcc gacgaatcag 3120
aacatcacta acccagtcgt gatttcattt ccacagtgac cggacgaagg tggagaagtt 3180
cgaaatttaa aaaaagtgac cacattttat ttaatagtga tgtgcaagtg atactatttt 3240
tattttgttt ttcttttgta ggaaaatgct gagcgaaata aataatttta gtggtgtgct 3300
atcgtcatcg atgaagttgt tttgcgaatg atactatgtt cttcaagtgc tgtgttttgt 3360
ggactgtggg gtgactgttc ctgtaaataa gcttcgttg 3399

<210> 45
<211> 996
<212> ADN
<213> Cydia pomonella
<400> 45
catcagacgg gcccaggctc aggatgaagc tagagcgcgg gcggcggacg cagggctcca 60
ccctcccggg atcgagctag atcggcctga gccgccagtg gtgaaagcgc cgaggagtcc 120
cgtgatcccg ccgccgccgc cgcgctccat gggatcggcg agctgcgact ccgttccggg 180
atcgcccggg gtatcgccgt atgcgccgca cccgccgtcc gctccgcctc cgccgatgcc 240
gccgctcccg cctccgcaac cagtggcctt ggactccctg gtagaaaact gccacaagct 300
gctggaaaaa ttccactaca gttgggagat gatgccgctc gtgctggtca tcctcaacta 360
cgccggctcc gacctggagg aggcctcgcg gaagattgac gaagcctcct gggtggtgca 420
ccagtggcgg ctgtacgagc gctcactgtg ctcgctgctg gagctgcaag cgcgcaaaga 480
gtcgttttgc tgctcgccgc gctatgtgct gtcgcgcgag tacgcgccgc acctgcccgt 540
gccgctcatg cgctcgccgc cgccagcgca cttgtagccc cacaccgcgc cgcgacagac 600
ggcgcacgag cccactgagc catctacttc ggccaaaccc gagtaggccc gaggccgacc 660
cgagcccgac ccgagaggac ccgagtgggc tattccggac tttacctagt tttatatgtg 720
ctatacgtgt tacaacacgc atatttgtat attatcacgg acattaagtt ggagagcggt 780
taccttatct tgttaacccg gtccttgaag taattattcc cagatatatt aagaaaacca 840
gtgaatactt tgcctgatgt ataattaaca gttgttaagc aaccatgaga attatggtat 900
ttcttgtgga catgttgcag ctagaaattt catatcatcg gtgataaaat ttaaccacac 960
tgtggttggc ggaaaaccac attgtttgta atattg 996

<210> 46
<211> 6751
<212> ADN
<213> nhân tạo
<220>
<223> Trình tự của cấu trúc/plasmit pLA3435-Bombyx mori-dsx

<220>
<221> Dấu hiệu hỗn tạp
<222> (1617).. (1622)
<223> n là a, c, g hoặc t
<400> 46

ggccgcatgg tacccattgc ttgtcattta ttaatttgga tgatgtcatt tgtttttaaa 60
attgaactgg ctttacgagt agaattctac gcgtaaaaca caatcaagta tgagtcataa 120
tctgatgtca tgttttgtac acggctcata accgaactgg ctttacgagt agaattctac 180
ttgtaatgca cgatcagtgg atgatgtcat ttgtttttca aatcgagatg atgtcatgtt 240
ttgcacacgg ctcataaact cgctttacga gtagaattct acgtgtaacg cacgatcgat 300
tgatgagtca tttgttttgc aatatgatat catacaatat gactcatttg ttttcaaaa 360
ccgaacttga tttacgggta gaattctact tgtaaagcac aatcaaaaag atgatgtcat 420
ttgtttttca aaactgaact cgctttacga gtagaattct acgtgtaaaa cacaatcaag 480
aaatgatgtc atttgttata aaaataaaag ctgatgtcat gttttgcaca tggctcataa 540
ctaaactcgc tttacgggta gaattctacg cgtaaaacat gattgataat taaataattc 600
atttgcaagc tatacgttaa atcaaacgga cgctcgaggt tgcacaacac tattatcgat 660
ttgcagttcg ggacataaat gtttaaatat atcgatgtct ttgtgatgcg cgcgacattt 720
ttgtaggtta ttgataaaat gaacggatac gttgcccgac attatcatta aatccttggc 780
gtagaatttg tcgggtccat tgtccgtgtg cgctagcatg cccgtaacgg acctcgtact 840
tttggcttca aaggttttgc gcacagacaa aatgtgccac acttgcagct ctgcatgtgt 900
gcgcgttacc acaaatccca acggcgcagt gtacttgttg tatgcaaata aatctcgata 960
aaggcgcggc gcgcgaatgc agctgatcac gtacgctcct cgtgttccgt tcaaggacgg 1020
tgttatcgac ctcagattaa tgtttatcgg ccgactgttt tcgtatccgc tcaccaaacg 1080
cgtttttgca ttaacattgt atgtcggcgg atgttctata tctaatttga ataaataaac 1140
gataaccgcg ttggttttag agggcataat aaaagaaata ttgttatcgt gttcgccatt 1200
agggcagtat aaattgacgt tcatgttgga tattgtttca gttgcaagtt gacactggcg 1260
gcgacaagca attctaattg gggtaagttt tcccgttctt ttctgggttc ttccctttg 1320
ctcatccttg ctgcactacc ttcaggtgca agttgagatt caggccacca tgggagatcc 1380
caccccaccc aagaagaagc gcaaaccggt ccgtccctc ggagacgctt gtggagaact 1440
gtcacagact cctcgagaag ttccattact cgtgggagat gatgccgctt gtgctcgtca 1500
tcatgaacta cgcccgcagc gacttggatg aggcttcaag gaaaatctac gaaggtaccg 1560
aatgtgtaaa tacgagtgta gcgttgatta gaaaacggac attgttcgtg agtttannnn 1620
nnggtctctc tggccagcaa gacatttgaa acactgtaaa aaaattcatt gaaaaaaaag 1680
aacactgtaa tgaaaatatt ctgaatgctt aatctggtat ttcagggatt aaactgattg 1740
tgatgaaaag tgattaaact attttcttta agtaccaaat taaccgaaca ggtttgggtc 1800
tttcctttca gtaacaaaca aaatctatcg aaggtaagaa ataaacaaca ggatattttc 1860
ttttactaaa aatcaataag gagactgcac tatttcaatg ttcaacttcc tttatcgaat 1920
gcatgaaaaa tttaattgtc taaaaatcta aattactaat taacgcaaag gaacctttgc 1980
ctaaaaaaaa aaataagcta ttaaacgaat gcctaaaata cgtaacagtg ttgccagttg 2040
taaaaattgc gaatccgaga agtgcagttt cctgaaatgc ccagcgatac gaatttccta 2100
tgttagagtc ttgtccgcag ggaagatgat cgtcgacgag tacgcgagga agcacaactt 2160
gaacgtgttc gacggactag aactaaggaa ctcgacacgc caggcgcgcc ggatccggcc 2220
ggccgaaaat gctggaaatt aataatataa gtggtgtact gtcttcgtca atgaagttat 2280
tttgcgaatg atacttagtt ttacaagtgc cgtggtgtgt gttgacactt gctgtgcgat 2340
gctgtgcgaa tttcaacgga aatatttgtt gtcgtaacat tggatctatg ggtaagttta 2400
gtataataac tttactctgt tcacattagt gaaacataca tttgtaaaat ttgtgtttta 2460
ctaatgtgaa atttatttt ggaaattcac gttaacacta ttgaataaaa aaaaatcgat 2520
aatgtaattt aaaaaaaata caaaaatata gttttcgctt attgttagaa agaaaatttt 2580

acatacgcca ttttgaataa ttccttccgg gtacattggg ccctaaacca gcgatcgggg 2640
aacttttta attattaccc taaaatattt ttatgtaagt tgatattacc gatggcgaag 2700
aacaacaaaa aaaaaaacga aatcgcttct ttttagcatc tttcatatta tagaccccac 2760
gataatttta aatcacaacg attataaaga agtttcactt caatatatac tttttactca 2820
caaaagtttc atttttaccc catttgggat aatttagccc ggttcccccc ccgaccgctg 2880
gcctaaacgt atcaccgaca atagctaaaa taacaaggta cgttcgattt gccgagctga 2940
actaacatta cacagctttg cattattcat atgtacattg cgactgaaac gtccggaccg 3000
ttacaggtta ttggatgatg catcaatggc gattgcagca gtattcgttg tgctacggag 3060
cgctggagtt gtcggcgcgc aaggatgtgg ccgcgctatg ttgcctccga gatacgtgct 3120
ggcgcccgag gtcccgccgc gtctggtgcc cctccagctg atctagataa ctgatcataa 3180
tcagccatac cacatttgta gaggttttac ttgctttaaa aaacctccca cacctccccc 3240
tgaacctgaa acataaaatg aatgcaattg ttgttgttaa cttgtttatt gcagcttata 3300
atggttacaa ataaagcaat agcatcacaa atttcacaaa taaagcattt ttttcactgc 3360
attctagttg tggtttgtcc aaactcatca atgtatctta acgcgagtta attaagtgcg 3420
cgtaaattgt aagcgttaat attttgttaa aattcgcgtt aaatttttgt taaatcagct 3480
catttttaa ccaataggcc gaaatcggca aaatccctta taaatcaaaa gaatagaccg 3540
agatagggtt gagtgttgtt ccagtttgga acaagagtcc actattaaag aacgtggact 3600
ccaacgtcaa agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac 3660
cctaatcaag ttttttgggg tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga 3720
gcccccgatt tagagcttga cggggaaagc cggcgaacgt ggcgagaaag gaagggaaga 3780
aagcgaaagg agcgggcgct agggcgctgg caagtgtagc ggtcacgctg cgcgtaacca 3840
ccacacccgc cgcgcttaat gcgccgctac agggcgcgtc aggtggcact tttcggggaa 3900
atgtgcgcgg aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca 3960
tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt cctgaggcgg 4020
aaagaaccag ctgtggaatg tgtgtcagtt agggtgtgga aagtccccag gctccccagc 4080
aggcagaagt atgcaaagca tgcatctcaa ttagtcagca accaggtgtg gaaagtcccc 4140
aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag caaccatagt 4200
cccgccccta actccgccca tcccgccccc aactccgccc agttccgccc attctccgcc 4260
ccatggctga ctaatttttt ttatttatgc agaggccgag gccgcctcgg cctctgagct 4320
attccagaag tagtgaggag gcttttttgg aggcctaggc ttttgcaaag atcgatcaag 4380
agacaggatg aggatcgttt cgcatgattg aacaagatgg attgcacgca ggttctccgg 4440
ccgcttgggt ggagaggcta ttcggctatg actgggcaca acagacaatc ggctgctctg 4500
atgccgccgt gttccggctg tcagcgcagg ggcgcccggt tctttttgtc aagaccgacc 4560
tgtccggtgc cctgaatgaa ctgcaagacg aggcagcgcg gctatcgtgg ctggccacga 4620
cgggcgttcc ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg gactggctgc 4680
tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct gccgagaaag 4740
tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggct acctgcccat 4800
tcgaccacca agcgaaacat cgcatcgagc gagcacgtac tcggatggaa gccggtcttg 4860
tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa ctgttcgcca 4920
ggctcaaggc gagcatgccc gacggcgagg atctcgtcgt gacccatggc gatgcctgct 4980
tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt ggccggctgg 5040
gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct gaagagcttg 5100
gcggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cgccgctccc gattcgcagc 5160
gcatcgcctt ctatcgcctt cttgacgagt tcttctgagc gggactctgg ggttcgaaat 5220
gaccgaccaa gcgacgccca acctgccatc acgagatttc gattccaccg ccgccttcta 5280
tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc tccagcgcgg 5340
ggatctcatg ctggagttct tcgcccaccc tagggggagg ctaactgaaa cacggaagga 5400
gacaataccg gaaggaaccc gcgctatgac ggcaataaaa agacagaata aaacgcacgg 5460
tgttgggtcg tttgttcata aacgcggggt tcggtcccag ggctggcact ctgtcgatac 5520

cccaccgaga ccccattggg gccaatacgc ccgcgtttct tccttttccc caccccaccc 5580
cccaagttcg ggtgaaggcc cagggctcgc agccaacgtc ggggcggcag gccctgccat 5640
agcctcaggt tactcatata tactttagat tgatttaaaa cttcattttt aatttaaaag 5700
gatctaggtg aagatccttt ttgataatct catgaccaaa atcccttaac gtgagttttc 5760
gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag atccttttt 5820
tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt 5880
gccggatcaa gagctaccaa ctcttttttc gaaggtaact ggcttcagca gagcgcagat 5940
accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga actctgtagc 6000
accgcctaca tacctcgctc tgctaatcct gttaccagtg gctgctgcca gtggcgataa 6060
gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc agcggtcggg 6120
ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca ccgaactgag 6180
atacctacag cgtgagctat gagaaagcgc cacgcttccc gaagggagaa aggcggacag 6240
gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc cagggggaaa 6300
cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt 6360
gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg 6420
gttcctggcc ttttgctggc cttttgctca catgttcttt cctgcgttat cccctgattc 6480
tgtggataac cgtattaccg ccatgcatta gttattaata gtaatcaatt acggggtcat 6540
tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaat ggcccgcctg 6600
gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt cccatagtaa 6660
cgccaatagg gactttccat tgacgtcaat gggtggagta tttacggtaa actgcccact 6720
tggcagtaca tcaagtgtat catagcgatg c 6751

<210> 47
<211> 8183
<212> ADN
<213> nhân tạo
<220>
<223> Trình tự của cấu trúc pLA3359-Anopheles gambiae dsx
<400> 47
ccggtgctgc tgttgctgat gctacgatcc tcgacagtga ttggaaacgc ctggagatgg 60
tgggaaaaaa tcaaacacaa aaacggtcct aatgaacatc gtgtgttctc attcgctgcc 120
acgattgaca ccttcgataa gacgcacata atgagctaaa ggagagggga cagggtcttg 180
tctttgccac gagcgataag attgcaatca ctcgtgagcg tgtgctgctg ggctgaagaa 240
gaaacacttt ccacagcagt aggtgggaag tgggattgtg gaacgtggca ttgaaaagaa 300
cctatttttct aaagccccgag agcccgttct cgaactggaa aacgagatgc agaagttttt 360
tattgtcccc cgccaggaaa acaaatgtat ttaatgcttt ctctgccttt tccgccccgt 420
ttcagacgac gagctagtga agcgagccca atggctgttg gagaaactcg gctacccgtg 480
ggagatgatg cccctgatgt acgtcatact gaagagcgcc gatggcgatg tacaaaaagc 540
acaccagcgg atcgacgaag gtaagctggc gatgatggtg tcgttcgaca tcactttcat 600
caccgtgtca gacatctact gtgcctagca ccggtccagt ggtcacaggg tgtagcaaaa 660
acgtgttctt ttttgcgaga gactctacct catgatgcag ctgttaagga aaggtttcag 720
atgaagacaa tttttcctag gataatatga tcttaagtta cctgcgtatg agtgtttaac 780
attgtcgtct caactccaag gaatgtttta accgtctagg gctagtttat ttatactgtt 840
ctcattgaaa tgtcgttaaa tccaacatgt taagttagct agctcagaca cgagaagtta 900
ggagtatctg catcttgaag gtagcggcat atggtgttat gccacgttca ctgacttcaa 960
aattcgatac aaaaaaaaac aaaatcaaaa acaaaattgt gaattccgtc agccagcagc 1020
agtgaccttc aaagccttac ctttccattc atttatgttt aacacaggtc aagcggtggt 1080
caacgaatac tcacgattgc ataatctgaa catgtttgat ggcgtggagt tgcgcaatac 1140
cacccgtcag agtggatgat aaactttccg caccactgta actgtccgta tctttgtatg 1200

tgggtgtgtg tatgtgtgtt tggtgaaacg aattcaattg ttctgtgcta ttttaaatca 1260
agccgcgtgc gcaactgatg ccgataagtt caaactagtg tttaaggagt ggagagagag 1320
agccgcacca cggtacagaa gggcagcaga atgggtcggc agcctagctg cactggtgcg 1380
gtgcgtccgg cgtctcgggg ggagggcggg gaaattctag tgttaaatcg gagcagcaaa 1440
aacaaaacag tggtcgtccc gttcaagaaa cggcctgtac acacacagaa aacactgcag 1500
catgtttgta catagtagat cctagagcag gtggtcgttg ctcctcgaac gctctggacg 1560
cacggcttcg cgcgtacttg cgtagcgttc caccgatcgt gggtattcgt actgccacaa 1620
gcccgctttc tcccatgcaa tctctgcaac caaaccaaca aacaacaaca aaataccaat 1680
cgacacaatg aatcacaccc cttttgtatc atctgtatat tcttgttctt tgcgttcttt 1740
tccatgtggc ccacgccccg gcgggtacgt aattgcgtcg aaaaccccga aaaccccggc 1800
acatacagtg tacatacggt ttgaggacaa ctttgacctg cagcccttct ggggctgcca 1860
cgtgtagcta tacttgtgag atcgggcgcc gacggtgtaa agcgcgaatg gccgccacac 1920
agtgtgtcca ctccaacact acccctctgg aactaccccg tccagggatg caccggctcg 1980
gctcatgccc ctgcaaaaca gtccgggctc cactgtagta gctccggcgt tgctctgaga 2040
gaaggatgcc cttcgaagtg tcgaaagcgt gcattgggcg ttcaagtgtg tgtctgtgtt 2100
aggtttagcg agaaacagca gcagttgcgt gtgctgaaaa gcgaaggagt aatagagtgc 2160
ataatgaaaa tgaaaatgaa aatgaagcaa aagtagaagg cggaggagag caacctgtgt 2220
tccactagta gcgaatagtt tagtctagtt tcgtcaccaa tcaaccttcc aaccatcgtt 2280
caaccaatac ctgagtcaac atcgtcatcg ttatcgtgcc acaactttat taaaaatgaa 2340
ccttgtccgc gccaccgtag ggtgatctga ggcgaccttt cttacgggcg cgactcacat 2400
gccatcgtca ccttctccaa tcaaaaccaa cagcctgtac cgatggtgtg caattgtgcg 2460
tgcgtgtgtg ttattagcaa aaaaagagaa agagacggcg agagagagat agatcgagat 2520
cgagagtaca aaagagcagt agaaatgttc gttgtttgtt ttccgtaaca cagttgttta 2580
gccaaaatgg gaatttccaa taatcccggg ggcgggggaaa tgcgggaata ctgcgtacac 2640
acatacatca atcaaaaaga aaaatccttg cgctacatca ctaccgtttg cgcggtgctg 2700
atctagagca gaccactttc cacgccattc tacaatcaat caatctgtgc agaaggtatg 2760
gtaagacggc ctttgagcga gtcacggtcg ccaccataac gccgtccgac gagggctgaa 2820
tgcgaacttt gctaatcgat tttccgcttt ctttttatcc caccccccct tctctctctc 2880
tcttttgcac cgccccttgt aaccccccaaa aaggtaaacg acacattaag acctacgaag 2940
cgctggtgaa gtcatcgctc gatccgaaca gcgaccggct gacggaagac gacgacgagg 3000
acgagaacat ctcggtgacc cgcaccaact ccaccattcg gtcgaggtcc agctcgctgt 3060
cgcggtcccg gtcctgctcg cgccaggccg aaactccccg ggccgacgat cgggccctga 3120
accttgacac caaatagatc tcgacccaag aaaaagcgga aggtggagga cccgtaagat 3180
ccaccggatc tagataactg atcataatca gccataccac atttgtagag gttttacttg 3240
ctttaaaaaa cctcccacac ctccccctga acctgaaaca taaaatgaat gcaattgttg 3300
ttgttaactt gtttattgca gcttataatg gttacaaata aagcaatagc atcacaaatt 3360
tcacaaataa agcatttttt tcactgcatt ctagttgtgg tttgtccaaa ctcatcaatg 3420
tatcttaacg cgagttaatt aagtgcgcgt aaattgtaag cgttaatatt ttgttaaaat 3480
tcgcgttaaa tttttgttaa atcagctcat tttttaacca ataggccgaa atcggcaaaa 3540
tcccttataa atcaaaagaa tagaccgaga tagggttgag tgttgttcca gtttggaaca 3600
agagtccact attaaagaac gtggactcca acgtcaaagg gcgaaaaacc gtctatcagg 3660
gcgatggccc actacgtgaa ccatcaccct aatcaagttt tttggggtcg aggtgccgta 3720
aagcactaaa tcggaaccct aaagggagcc cccgatttag agcttgacgg ggaaagccgg 3780
cgaacgtggc gagaaaggaa gggaagaaag cgaaaggagc gggcgctagg gcgctggcaa 3840
gtgtagcggt cacgctgcgc gtaaccacca cacccgccgc gcttaatgcg ccgctacagg 3900
gcgcgtcagg tggcactttt cggggaaatg tgcgcggaac ccctatttgt ttatttttct 3960
aaatacattc aaatatgtat ccgctcatga gacaataacc ctgataaatg cttcaataat 4020
attgaaaaag gaagagtcct gaggcggaaa gaaccagctg tggaatgtgt gtcagttagg 4080
gtgtggaaag tccccaggct ccccagcagg cagaagtatg caaagcatgc atctcaatta 4140

gtcagcaacc aggtgtggaa agtccccagg ctccccagca ggcagaagta tgcaaagcat 4200
gcatctcaat tagtcagcaa ccatagtccc gcccctaact ccgcccatcc cgcccctaac 4260
tccgcccagt tccgcccatt ctccgcccca tggctgacta attttttta tttatgcaga 4320
ggccgaggcc gcctcggcct ctgagctatt ccagaagtag tgaggaggct tttttggagg 4380
cctaggcttt tgcaaagatc gatcaagaga caggatgagg atcgtttcgc atgattgaac 4440
aagatggatt gcacgcaggt tctccggccg cttgggtgga gaggctattc ggctatgact 4500
gggcacaaca gacaatcggc tgctctgatg ccgccgtgtt ccggctgtca gcgcaggggc 4560
gcccggttct ttttgtcaag accgacctgt ccggtgccct gaatgaactg caagacgagg 4620
cagcgcggct atcgtggctg gccacgacgg gcgttccttg cgcagctgtg ctcgacgttg 4680
tcactgaagc gggaagggac tggctgctat tgggcgaagt gccggggcag gatctcctgt 4740
catctcacct tgctcctgcc gagaaagtat ccatcatggc tgatgcaatg cggcggctgc 4800
atacgcttga tccggctacc tgcccattcg accaccaagc gaaacatcgc atcgagcgag 4860
cacgtactcg gatggaagcc ggtcttgtcg atcaggatga tctggacgaa gagcatcagg 4920
ggctcgcgcc agccgaactg ttcgccaggc tcaaggcgag catgcccgac ggcgaggatc 4980
tcgtcgtgac ccatggcgat gcctgcttgc cgaatatcat ggtggaaaat ggccgctttt 5040
ctggattcat cgactgtggc cggctgggtg tggcggaccg ctatcaggac atagcgttgg 5100
ctacccgtga tattgctgaa gagcttggcg gcgaatgggc tgaccgcttc ctcgtgcttt 5160
acggtatcgc cgctcccgat tcgcagcgca tcgccttcta tcgccttctt gacgagttct 5220
tctgagcggg actctggggt tcgaaatgac cgaccaagcg acgcccaacc tgccatcacg 5280
agatttcgat tccaccgccg ccttctatga aaggttgggc ttcggaatcg ttttccggga 5340
cgccggctgg atgatcctcc agcgcgggga tctcatgctg gagttcttcg cccaccctag 5400
ggggaggcta actgaaacac ggaaggagac aataccggaa ggaacccgcg ctatgacggc 5460
aataaaaaga cagaataaaa cgcacggtgt tgggtcgttt gttcataaac gcggggttcg 5520
gtcccagggc tggcactctg tcgatacccc accgagaccc cattggggcc aatacgcccg 5580
cgtttcttcc ttttccccac cccaccccc aagttcgggt gaaggcccag ggctcgcagc 5640
caacgtcggg gcggcaggcc ctgccatagc ctcaggttac tcatatatac tttagattga 5700
tttaaaactt catttttaat ttaaaaggat ctaggtgaag atccttttg ataatctcat 5760
gaccaaaatc ccttaacgtg agttttcgtt ccactgagcg tcagaccccg tagaaaagat 5820
caaaggatct tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa 5880
accaccgcta ccagcggtgg tttgtttgcc ggatcaagag ctaccaactc tttttccgaa 5940
ggtaactggc ttcagcagag cgcagatacc aaatactgtc cttctagtgt agccgtagtt 6000
aggccaccac ttcaagaact ctgtagcacc gcctacatac ctcgctctgc taatcctgtt 6060
accagtggct gctgccagtg gcgataagtc gtgtcttacc gggttggact caagacgata 6120
gttaccggat aaggcgcagc ggtcgggctg aacgggggt tcgtgcacac agcccagctt 6180
ggagcgaacg acctacaccg aactgagata cctacagcgt gagctatgag aaagcgccac 6240
gcttcccgaa gggagaaagg cggacaggta tccggtaagc ggcagggtcg gaacaggaga 6300
gcgcacgagg gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg 6360
ccacctctga cttgagcgtc gatttttgtg atgctcgtca gggggcggga gcctatggaa 6420
aaacgccagc aacgcggcct ttttacggtt cctggccttt tgctggcctt ttgctcacat 6480
gttctttcct gcgttatccc ctgattctgt ggataaccgt attaccgcca tgcattagtt 6540
attaatagta atcaattacg gggtcattag ttcatagccc atatatggag ttccgcgtta 6600
cataacttac ggtaaatggc ccgcctggct gaccgcccaa cgacccccgc ccattgacgt 6660
caataatgac gtatgttccc atagtaacgc caatagggac tttccattga cgtcaatggg 6720
tggagtattt acggtaaact gcccacttgg cagtacatca agtgtatcat agcgatgcgg 6780
ccgcatggta cccattgctt gtcatttatt aatttggatg atgtcatttg tttttaaaat 6840
tgaactggct ttacgagtag aattctacgc gtaaaacaca atcaagtatg agtcataatc 6900
tgatgtcatg ttttgtacac ggctcataac cgaactggct ttacgagtag aattctactt 6960
gtaatgcacg atcagtggat gatgtcattt gtttttcaaa tcgagatgat gtcatgtttt 7020
gcacacggct cataaactcg ctttacgagt agaattctac gtgtaacgca cgatcgattg 7080

atgagtcatt tgttttgcaa tatgatatca tacaatatga ctcatttgtt tttcaaaacc 7140
gaacttgatt tacgggtaga attctacttg taaagcacaa tcaaaaagat gatgtcattt 7200
gtttttcaaa actgaactcg ctttacgagt agaattctac gtgtaaaaca caatcaagaa 7260
atgatgtcat ttgttataaa aataaaagct gatgtcatgt tttgcacatg gctcataact 7320
aaactcgctt tacgggtaga attctacgcg taaaacatga ttgataatta aataattcat 7380
ttgcaagcta tacgttaaat caaacggacg ctcgaggttg cacaacacta ttatcgattt 7440
gcagttcggg acataaatgt ttaaatatat cgatgtcttt gtgatgcgcg cgacattttt 7500
gtaggttatt gataaaatga acggatacgt tgcccgacat tatcattaaa tccttggcgt 7560
agaatttgtc gggtccattg tccgtgtgcg ctagcatgcc cgtaacggac ctcgtacttt 7620
tggcttcaaa ggttttgcgc acagacaaaa tgtgccacac ttgcagctct gcatgtgtgc 7680
gcgttaccac aaatcccaac ggcgcagtgt acttgttgta tgcaaataaa tctcgataaa 7740
ggcgcggcgc gcgaatgcag ctgatcacgt acgctcctcg tgttccgttc aaggacggtg 7800
ttatcgacct cagattaatg tttatcggcc gactgttttc gtatccgctc accaaacgcg 7860
tttttgcatt aacattgtat gtcggcggat gttctatatc taatttgaat aaataaacga 7920
taaccgcgtt ggttttagag ggcataataa aagaaatatt gttatcgtgt tcgccattag 7980
ggcagtataa attgacgttc atgttggata ttgtttcagt tgcaagttga cactggcggc 8040
gacaagcaat tctaattggg gtaagttttc ccgttctttt ctgggttctt ccctttttgct 8100
catccttgct gcactacctt caggtgcaag ttgagattca ggccaccatg ggagatccca 8160
ccccacccaa gaagaagcgc aaa 8183

<210> 48
<211> 7342
<212> ADN
<213> nhân tạo
<220>
<223> Trình tự của cấu trúc pLA3433-Agdsx (Anopheles gambiae) với exon 2 được bao hàm
<400> 48
ctagtgtcga cgatgtaggt cacggtctcg aagccgcggt gcgggtgcca gggcgtgccc 60
ttgggctccc cgggcgcgta ctccacctca cccatctggt ccatcatgat gaacgggtcg 120
aggtggcggt agttgatccc ggcgaacgcg cggcgcaccg ggaagccctc gccctcgaaa 180
ccgctgggcg cggtggtcac ggtgagcacg ggacgtgcga cggcgtcggc gggtgcggat 240
acgcggggca gcgtcagcgg gttctcgacg gtcacggcgg gcatgtcgac cgccggcgcc 300
ttaattaact cgcgttaaga tacattgatg agtttggaca aaccacaact agaatgcagt 360
gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc tttatttgta accattataa 420
gctgcaataa acaagttaac aacaacaatt gcattcattt tatgtttcag gttcaggggg 480
aggtgtggga ggtttttttaa agcaagtaaa acctctacaa atgtggtatg gctgattatg 540
atcagttatc tagatccggt ggatcttacg ggtcctccac cttccgcttt ttcttgggtc 600
gagatctgag tccggaatcc tcgtcgctac cgatggcgct ggtgatgcgg ggcacgctgt 660
gggcgtaggt cacctcgcgc tggcacacgt ggtcgcgctt gtcgctggtg tccctcatct 720
gcttggtgat gatggtcacg aagtgggggc cggggatctt gatggcgcgg ctgccgttga 780
aggtcatctt gctgtcgaag tggcccatca tcaggccgcc gtcggcggtg gtgaagccga 840
tgaaggccag ctggcgcacg gcgttggggc cgtgggggaa catgtgggtc tcgttgggca 900
ggatgtccac cagctggtcg cgcatgatgg ggccgtcggg ctggaagccg tcgcagttca 960
cggtgatgcg gctgaccacg caggtgccgt ccagctcgta ggtgtggtgg ctggtcatgg 1020
tgccgtcgtt ctcgaagcgc acggtgcggt cgatgctcag gccctcgggg aagcactcct 1080
gggcgaagtg gctgatgccg ttggggtagc gggcgaagaa gggctcgccg tactggatca 1140
ggtggcagat gggcttccag ctcatgggca gcttgccggt ctcgcacacg gcgtgcacgt 1200
tgaagtcgcc gtgggggaac ttgctgctgc cgtcggccac gatggtgaac ttctggccgt 1260

tcacctcgcc gtcgatgaag attttgaagg tcatgtcgct ctggaacagg gcggggccgc 1320
cctctgaacc atcctcgtcc atggtggcga ccggtttgcg cttcttcttg ggtggggtgg 1380
gatccaccag agacaggttg cggcggcggt tggatggcgt gggcgcgttg gcgttgttgg 1440
accggctcat gttgtgtcgc tgtaacagat gctgttcaac tgtgtttacc agatcgttgc 1500
gggctgtatt tataggcgcg ataagcggga cgggcgcctc gtgtccggtc acgcgcatga 1560
gataacgcgc ggctgatatg gaggcgcgtc ctgttccgat aaggagttgc gtccggctgc 1620
ggttagcaac acaggaagct ggcgtcctgt cacgataaga caacactcgt ccggtccgat 1680
aatgtgattc gtacgtgaca ggacgcgacc cgataaggcc ggcctacgtg actgccgaca 1740
cgtacttttt tgcactgcaa aaaggttcaa tgtgtggtag tgtatttgga gcgtatacaa 1800
cggtgtagac tatttatgta aaatagtcta cgaaacgtag agtttgtact atgtatgggc 1860
ccgcgtgcaa aagcgtgttt ttttgcagtg caaaaaagtt ggtggtgggg aggccaccga 1920
gtatggtacc atgcggccgc gtacgcgccc ggggagccca agggcacgcc ctggcacccg 1980
tccggtgctt atctagagcg cgcttggcgt aatcatggtc atagctgttt cctgtgtgaa 2040
attgttatcc gctcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct 2100
ggggtgccta atgagtgagc taactcacat taattgcgtt gcgctcactg cccgctttcc 2160
agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg 2220
gtttgcgtat tgggcgctct tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc 2280
ggctgcggcg agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag 2340
gggataacgc aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa 2400
aggccgcgtt gctggcgttt ttccataggc tccgcccccc tgacgagcat cacaaaaatc 2460
gacgctcaag tcagaggtgg cgaaacccga caggactata aagataccag gcgtttcccc 2520
ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg 2580
cctttctccc ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt 2640
cggtgtaggt cgttcgctcc aagctgggct gtgtgcacga acccccccgtt cagcccgacc 2700
gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc 2760
cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 2820
agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg 2880
ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa 2940
ccaccgctgg tagcggtggt tttttgtttt gcaagcagca gattacgcgc agaaaaaaag 3000
gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact 3060
cacgttaagg gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa 3120
attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt 3180
accaatgctt aatcagtgag gcacctatct cagcgatctg tctatttcgt tcatccatag 3240
ttgcctgact ccccgtcgtg tagataacta cgatacggga gggcttacca tctggccccа 3300
gtgctgcaat gataccgcga gacccacgct caccggctcc agatttatca gcaataaacc 3360
agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc tccatccagt 3420
ctattaattg ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg 3480
ttgttgccat tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca 3540
gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg 3600
ttagctcctt cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca 3660
tggttatggc agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg 3720
tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct 3780
cttgcccggc gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgctca 3840
tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca 3900
gttcgatgta acccactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg 3960
tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac 4020
ggaaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcagggtt 4080
attgtctcat gagcggatac atatttgaat gtatttagaa aaataaacaa atagggggttc 4140
cgcgcacatt tccccgaaaa gtgccaccta aattgtaagc gttaatattt tgttaaaatt 4200

cgcgttaaat ttttgttaaa tcagctcatt ttttaaccaa taggccgaaa tcggcaaaat 4260
cccttataaa tcaaaagaat agaccgagat agggttgagt gttgttccag tttggaacaa 4320
gagtccacta ttaaagaacg tggactccaa cgtcaaaggg cgaaaaaccg tctatcaggg 4380
cgatggccca ctacgtgaac catcacccta atcaagtttt ttggggtcga ggtgccgtaa 4440
agcactaaat cggaacccta aagggagccc ccgatttaga gcttgacggg gaaagccggc 4500
gaacgtggcg agaaaggaag ggaagaaagc gaaaggagcg ggcgctaggg cgctggcaag 4560
tgtagcggtc acgctgcgcg taaccaccac acccgccgcg cttaatgcgc cgctacaggg 4620
cgcgtcccat tcgccattca ggctgcgcaa ctgttgggaa gggcgatcgg tgcgggcctc 4680
ttcgctatta cgccagctgg cgaaaggggg atgtgctgca aggcgattaa gttgggtaac 4740
gccagggttt tcccagtcac gacgttgtaa aacgacggcc agtgagcgcg ctagcgttta 4800
aacgagctct aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 4860
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 4920
ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggaggtgt 4980
gggaggtttt ttaaagcaag taaaacctct acaaatgtgg tatggctgat tatgatcctg 5040
cagctacgcc gctacgtctt ccgtgccgtc ctgggcgtcg tcttcgtcgt cgtcggtcgg 5100
cggcttcgcc cacgtgatcg aagcgcgctt ctcgatgggc gttccctgcc ccctgccgct 5160
agtcgacttc gtgacaacga tcttgtctac gaagagcccg acgaacacgc gcttgtcgtc 5220
tactgacgcg cgcccccacc acgacttagg gccggtcggg tcagcgtcgg cgtcttcggg 5280
gaaccattgg tcaaggggaa gcttcggggc ttcggcggct tcaagttcgg caagccgctc 5340
ttccgccccc tgctgccgga gcgtcagcgc tgcctgttgc ttccggaagt gcttcctgcc 5400
aacgggtccg tcgtacgcgc ctgccgcgcg gtcttcgtac agctcttcaa gggcgttcag 5460
ggcgtcggcg cgctccgcaa caaggttcgc ccgttcgccg ctcttctcag gcgcctcagt 5520
gagcttgccg aagcgtcggg cggcttccca cagaagcgcc aacgtctctt cgtcgccttc 5580
ggcgtgcctg atcttgttga agatgcgttc cgcaacgaac ttgtcgagtg ccgccatgct 5640
gacgttgcac gtgccttcgt gctgcccagg tgcggacggg tcgaccacct tccggcgacg 5700
gcagcggtaa gagtccttga tcgattcttc cccgcgcttc gaagtcatga cggcgccaca 5760
ctcgcagtac agcttgtcca tggcggacag aatggcttgc ccccgggaaa gccccttgcc 5820
gcgcccccctg ccgtccaacc acgcctgaag ctcataccac tcagcgggct cgatgatcgg 5880
tccgcaatca agctcgaccg gccggagcgt gatcgggtcg cgctgaatgc ggtaaccctc 5940
aatcttcgtg gtcggcgtgc cgtccggctt cttcttgtag atcacctcag cggcgaagcc 6000
cgcaatacgc gggtcccgaa ggattcgcat aacggttgcc gggtcccagg cgcttgaagc 6060
ggtcttcttc ccaatcgtct cgccccgggt cggcacggcg tcagcgtcca tgcgcttaca 6120
aagccccgtg atgctgcccg ggtgaatggc ggcttgactg cccggcttga agggaaggtg 6180
tttgtgcgtc ttgatctcac gccaccacca ccggattacg tcgggctcga actcgaaggg 6240
tccggtaagg ggagtggtcg agtgcgcaag cttgttgatg acgacattga ccattcggcc 6300
gttgcgcgtg atctccttcg tctccgaaac aagctcgaag ccgtaaggcg ccttcccgcc 6360
gacgtacccg cccaattcgc gctgaaggtt cttcgtgtcg agaatcttcg ccgacttcag 6420
cgaagattct ttgtgcgacg cgtcgagccg cataatcagg tgaatcaggt ccatgacgtt 6480
tccctgccgg aagacgcctt cctgagtgga aacaatcgtc acgcccaggg cgagcaattc 6540
cgagacaatc ggaatcgcgt ccatgacctt caggcgcgag aagcgcgaca cgtcatagac 6600
aatgatcatg ttgagccgcc cggcgcggca ttcgttcagg atgcgttcga actccgggcg 6660
ctccgccgtc ccgaacgccg acgtgccccgg cgcttcgctg aaatgcccga cgaacctgaa 6720
ccggcccccg tcgcgctcga cttcgcgctg aaggtcggcc gccttgtctt cgttggcgct 6780
acgctgtgtc gctgggcttg ctgcgctcga attctcgcgc tcgcgcgact gacggtcgta 6840
agcacccgcg tacgtgtcca tggcggatcc gtgtcgctgt aacagatgct gttcaactgt 6900
gtttaccaga tcgttgcggg ctgtatttat aggcgcgata agcgggacgg gcgcctcgtg 6960
tccggtcacg cgcatgagat aacgcgcggc tgatatggag gcgcgtcctg ttccgataag 7020
gagttgcgtc cggctgcggt tagcaacaca ggaagctggc gtcctgtcac gataagacaa 7080
cactcgtccg gtccgataat gtgattcgta cgtgacagga cgcgacccga taaggccggc 7140

ctacgtgact gccgacacgt actttttttgc actgcaaaaa ggttcaatgt gtggtagtgt 7200
atttggagcg tatacaacgg tgtagactat ttatgtaaaa tagtctacga aacgtagagt 7260
ttgtactatg tatgggcccg cgtgcaaaag cgtgtttttt tgcagtgcaa aaaagttggt 7320
ggtggggagg ccaccgagta ta 7342

<210> 49
<211> 11868
<212> ADN
<213> nhân tạo
<220>
<223> trình tự 49 của cấu trúc intron pLA1188-cctra
<400> 49
gtggtttttg tcaaacgaag attctatgac gtgtttaaag tttaggtcga gtaaagcgca 60
aatcttttttt aaccctagaa agatagtctg cgtaaaattg acgcatgcat tcttgaaata 120
ttgctctctc tttctaaata gcgcgaatcc gtcgctgtgc atttaggaca tctcagtcgc 180
cgcttggagc tcccgtgagg cgtgcttgtc aatgcggtaa gtgtcactga ttttgaacta 240
taacgaccgc gtgagtcaaa atgacgcatg attatctttt acgtgacttt taagatttaa 300
ctcatacgat aattatattg ttatttcatg ttctacttac gtgataactt attatatata 360
tattttcttg ttatagatat cgtgactaat atataataaa atgggtagtt ctttagacga 420
tgagcatatc ctctctgctc ttctgcaaag cgatgacgag cttgttggtg aggattctga 480
cagtgaaata tcagatcacg taagtgaaga tgacgtccag agcgatacag aagaagcgtt 540
tatagatgag gtacatgaag tgcagccaac gtcaagcggt agtgaaatat tagacgaaca 600
aaatgttatt gaacaaccag gttcttcatt ggcttctaac agaatcttga ccttgccaca 660
gaggactatt agaggtaaga ataaacattg ttggtcaact tcaaagtcca cgaggcgtag 720
ccgagtctct gcactgaaca ttgtcagatc ggcccgggcg gccgtttttc ttgaaatatt 780
gctctctctt tctaaatagc gcgaatccgt cgctgtgcat ttaggacatc tcagtcgccg 840
cttggagctc ccaaacgcgc cagtggtagt acacagtact gtgggtgttc agtttgaaat 900
cctcttgctt ctccattgtc tcggttacct ttggtcaaat ccatgggttc tattgcctat 960
atactcttgc gattaccagt gattgcgcta ttagctatta gatggattgt tggccaaact 1020
tgtcgcttaa gtggctggga attgtaaccg taggcccgag tgtaatgatc ccccataaaa 1080
agttttcgca atgcctttat tttttgttgc aaatctctct ttattctgcg gtattcttca 1140
ttattgcggg gatggggaaa gtgtttatat agaagcaact tacgattgaa cccaaatgca 1200
cctgacaagc aaggtcaaag ggccagattt ttaaatatat tatttagtct taggactctc 1260
tatttgcaat taaattactt tgctacctga gggttaaatc ttccccattg ataataataa 1320
ttccactata tgttcaattg ggtttcaccg cgcttagtta catgacgagc cctaatgagc 1380
cgtcggtggt ctataaactg tgccttacaa atacttgcaa ctcttctcgt tttgaagtca 1440
gcagagttat tgctaattgc taattgctaa ttgcttttaa ctgatttctt cgaaattggt 1500
gctatgttta tggcgctatt aacaagtatg aatgtcaggt ttaaccaggg gatgcttaat 1560
tgtgttctca acttcaaagg cagaaatgtt tactcttgac catgggttta ggtataatgt 1620
tatcaagctc ctcgagttaa cgttacgtta acgttaacgt tcgaggtcga ctctagaact 1680
acccaccgta ctcgtcaatt ccaagggcat cggtaaacat ctgctcaaac tcgaagtcgg 1740
ccatatccag agcgccgtag ggggcggagt cgtgggggggt aaatcccgga cccggggaat 1800
ccccgtcccc caacatgtcc agatcgaaat cgtctagcgc gtcggcatgc gccatcgcca 1860
cgtcctcgcc gtctaagtgg agctcgtccc ccaggctgac atcggtcggg ggggccgtcg 1920
acagtctgcg cgtgtgtccc gcgggggagaa aggacaggcg cggagccgcc agccccgcct 1980
cttcgggggc gtcgtcgtcc gggagatcga gcaggccctc gatggtagac ccgtaattgt 2040
ttttcgtacg cgcgcggctg tacgcggggc ccgagccccga ctcgcatttc agttgctttt 2100
ccaatccgca gataatcagc tccaagccga acaggaatgc cggctcggct ccttgatgat 2160
cgaacagctc gattgcctga cgcagcagtg ggggcatcga atcggttgtt ggggtctcgc 2220

gctcctcttt tgcgacttga tgctcttggt cctccagcac gcagcccagg gtaaagtgac 2280
cgacggcgct cagagcgtag agagcatttt ccaggctgaa gccttgctgg cacaggaacg 2340
cgagctggtt ctccagtgtc tcgtattgct tttcggtcgg gcgcgtgccg agatggactt 2400
tggcaccgtc tcggtgggac agcagagcgc agcggaacga cttggcgtta ttgcggagga 2460
agtcctgcca ggactcgcct tccaacgggc aaaaatgcgt gtggtggcgg tcgagcatct 2520
cgatggccag ggcatccagc agcgcccgct tattcttcac ctatagatac catagatgta 2580
tggattagta tcatatacat acaaaggcta tttttgggac atattaatat taacaatttc 2640
cgtgatagtt ttcaccattt ttgttgaatg ttacgttgaa aatttaaatt tgttttaaat 2700
taattttacc agtcatgtgt tcttaaaagt ttttatgatt gaaacggcat aaagtggttc 2760
aaaaatttat caagaaaggc tttccttttt taaatcttat ctttttctct taaaaatcac 2820
tagtcaattc attattaatt tgttaacttg aatttggaat gtctatttac tttcagataa 2880
attaaagcaa gaaacttaat attcgaaaaa aattgattct aaatggaatt tcacttgatc 2940
ttcatgtatg catatcaatt tttatttaca ttgtataata agtttcgagt tgattgttgt 3000
aatccacagg tgtcccagag aattaaattc caaattaccc aagtttattg aatgttgatt 3060
gtagtttcag ttgctttgtt gctgcaacaa tggcttgttg attgtagata ttttcccttt 3120
ccttggttta cttattacat agactgaaaa agaggtttac ttttttgata cttatgaaaa 3180
atttctatta gtgattacta accaatcgct atatgtttac tagaaaacaa ataaactctt 3240
tacattaaca ttcaataatg tttgctctgt aaccgacaat tgaaggcgtt acagcaacag 3300
taatataact agcttcttaa ccctcatcta ttaaccccat cgtttaaaac actatgttaa 3360
atggtctaac aaatctagat actaatagat gtcttattac ttagcagcca cagctgcaac 3420
atccaagaca atttttgaaa cttcttattg agctcttggc agcagaaatg ttggtatttt 3480
tcacagcttt ctgaaagacc ggcaccttcc tccggttccc gtttctgaat tcaagaggat 3540
ttccgacccc caattaatcc cgaaacaaat aaggtatatt caaaatgatg gaaaagtcat 3600
ggctgctgac cttatttta ttcctattga tagaatatta ttccccttt aaatacactg 3660
tactaagagg tccggctata attttactca cttgtcgatt atcccataga atgttgattg 3720
tagttggttg cttttccagg tgagagttga tcaagtcaca aaagttagcg tgtgttgatt 3780
gtagatttga aggtaaaata atttttgcac ccattcatcg ggtaaaacgt tctccataga 3840
atacatttcc atcgataatt gataacttat gaatttcaaa gaaaaaaata tgcttttaaa 3900
attacgtgcc agtagagggt gggctgctcc acgcccagct tctgcgccaa cttgcgggtc 3960
gtcagtccct caatgccaac ttcgttcaac agctccaacg cggagttgat gactttggac 4020
ttatccaggc ggctgcccat ggtggtttct aaaggtgtta taaatcaaat tagttttgtt 4080
ttttcttgaa aactttgcgt ttcctttgat caacttaccg ccagggtacc gcagattgtt 4140
tagcttgttc agctgcgctt gtttatttgc ttagctttcg cttagcgacg tgttcacttt 4200
gcttgtttga attgaattgt cgctccgtag acgaagcgcc tctatttata ctccggcgct 4260
cgttttcgag tttaccactc cctatcagtg atagagaaaa gtgaaagtcg agtttaccac 4320
tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc agtgatagag 4380
aaaagtgaaa gtcgagttta ccactcccta tcagtgatag agaaaagtga aagtcgagtt 4440
taccactccc tatcagtgat agagaaaagt gaaagtcgag tttaccactc cctatcagtg 4500
atagagaaaa gtgaaagtcg agtttaccac tccctatcag tgatagagaa aagtgaaagt 4560
cgaaacctgg cgcgccccgg ccatcgagaa agagagagag aagagaagag agagaacatt 4620
cgagaaagag agagagaaga gaagagagag aacatactcc ctatcagtga tagagaagtc 4680
cctatcagtg atagagatgt ccctatcagt gatagagagt tccctatcag tgatagagac 4740
gtccctatca gtgatagaga agtccctatc agtgatagag agatccctat cagtgataga 4800
gatttcccta tcagtgatag agaggtccct atcagtgata gagacttccc tatcagtgat 4860
agagaaatcc ctatcagtga tagagacatc cctatcagtg atagagaact ccctatcagt 4920
gatagagacc tccctatcag tgatagagat cgatgcggcc gcatggtacc cattgcttgt 4980
catttattaa tttggatgat gtcatttgtt tttaaaattg aactggcttt acgagtagaa 5040
ttctacgcgt aaaacacaat caagtatgag tcataatctg atgtcatgtt ttgtacacgg 5100
ctcataaccg aactggcttt acgagtagaa ttctacttgt aatgcacgat cagtggatga 5160

tgtcatttgt ttttcaaatc gagatgatgt catgttttgc acacggctca taaactcgct 5220
ttacgagtag aattctacgt gtaacgcacg atcgattgat gagtcatttg ttttgcaata 5280
tgatatcata caatatgact catttgtttt tcaaaaccga acttgattta cgggtagaat 5340
tctacttgta aagcacaatc aaaaagatga tgtcatttgt ttttcaaaac tgaactcgct 5400
ttacgagtag aattctacgt gtaaaacaca atcaagaaat gatgtcattt gttataaaaa 5460
taaaagctga tgtcatgttt tgcacatggc tcataactaa actcgcttta cgggtagaat 5520
tctacgcgta aaacatgatt gataattaaa taattcattt gcaagctata cgttaaatca 5580
aacggacgct cgaggttgca caacactatt atcgatttgc agttcgggac ataaatgttt 5640
aaatatatcg atgtctttgt gatgcgcgcg acatttttgt aggttattga taaaatgaac 5700
ggatacgttg cccgacatta tcattaaatc cttggcgtag aatttgtcgg gtccattgtc 5760
cgtgtgcgct agcatgcccg taacggacct cgtacttttg gcttcaaagg ttttgcgcac 5820
agacaaaatg tgccacactt gcagctctgc atgtgtgcgc gttaccacaa atcccaacgg 5880
cgcagtgtac ttgttgtatg caaataaatc tcgataaagg cgcggcgcgc gaatgcagct 5940
gatcacgtac gctcctcgtg ttccgttcaa ggacggtgtt atcgacctca gattaatgtt 6000
tatcggccga ctgttttcgt atccgctcac caaacgcgtt tttgcattaa cattgtatgt 6060
cggcggatgt tctatatcta atttgaataa ataaacgata accgcgttgg ttttagaggg 6120
cataataaaa gaaatattgt tatcgtgttc gccattaggg cagtataaat tgacgttcat 6180
gttggatatt gtttcagttg caagttgaca ctggcggcga caagcaattc taattggggt 6240
aagttttccc gttcttttct gggttcttcc cttttgctca tccttgctgc actaccttca 6300
ggtgcaagtt gagattcagg ccaccatggg agatcccacc ccacccaaga agaagcgcaa 6360
accggtcgcc accatggcct cctccgagaa cgtcatcacc gagttcatgc gcttcaaggt 6420
gcgcatggag ggcaccgtga acggccacga gttcgagatc gagggcgagg gcgagggccg 6480
cccctacgag ggccacaaca ccgtgaagct gaaggtgacc aagggcggcc ccctgccctt 6540
cgcctgggac atcctgtccc cccagttcca gtacggctcc aaggtgtacg tgaagcaccc 6600
cgccgacatc cccgactaca agaagctgtc cttccccgag ggcttcaagt gggagcgcgt 6660
gatgaacttc gaggacggcg gcgtggcgac cgtgacccag gactcctccc tgcaggacgg 6720
ctgcttcatc tacaaggtga agttcatcgg cgtgaacttc ccctccgacg gccccgtgat 6780
gcagaagaag accatgggct gggaggcctc caccgagcgc ctgtaccccc gcgacggcgt 6840
gctgaagggc gagacccaca aggccctgaa gctgaaggac ggcggccact acctggtgga 6900
gttcaagtcc atctacatgg ccaagaagcc cgtgcagctg cccggctact actacgtgga 6960
cgccaagctg gacatcacct cccacaacga ggactacacc atcgtggagc agtacgagcg 7020
caccgagggc cgccaccacc tgttcctgag atctcgaccc aagaaaaagc ggaaggtgga 7080
ggacccgtaa gatccaccgg atctagataa ctgatcataa tcagccatac cacatttgta 7140
gaggttttac ttgctttaaa aaacctccca cacctccccc tgaacctgaa acataaaatg 7200
aatgcaattg ttgttgttaa cttgtttatt gcagcttata atggttacaa ataaagcaat 7260
agcatcacaa atttcacaaa taaagcattt ttttcactgc attctagttg tggtttgtcc 7320
aaactcatca atgtatctta acgcgagtta attaaggccg ctcatttaaa tctggccggc 7380
cgcaaccatt gtgggaaccg tgcgatcaaa caaacgcgag ataccggaag tactgaaaaa 7440
cagtcgctcc aggccagtgg gaacatcgat gttttgtttt gacggacccc ttactctcgt 7500
ctcatataaa ccgaagccag ctaagatggt atacttatta tcatcttgtg atgaggatgc 7560
ttctatcaac gaaagtaccg gtaaaccgca aatggttatg tattataatc aaactaaagg 7620
cggagtggac acgctagacc aaatgtgttc tgtgatgacc tgcagtagga agacgaatag 7680
gtggcctatg gcattattgt acggaatgat aaacattgcc tgcataaatt cttttattat 7740
atacagccat aatgtcagta gcaagggaga aaaggtccaa agtcgcaaaa aatttatgag 7800
aaacctttac atgagcctga cgtcatcgtt tatgcgtaag cgtttagaag ctcctacttt 7860
gaagagatat ttgcgcgata atatctctaa tattttgcca aatgaagtgc ctggtacatc 7920
agatgacagt actgaagagc cagtaatgaa aaaacgtact tactgtactt actgcccctc 7980
taaaataagg cgaaaggcaa atgcatcgtg caaaaaatgc aaaaaagtta tttgtcgaga 8040
gcataatatt gatatgtgcc aaagttgttt ctgactgact aataagtata atttgtttct 8100

attatgtata agttaagcta attacttatt ttataataca acatgactgt ttttaaagta 8160
caaaataagt ttatttttgt aaaagagaga atgtttaaaa gttttgttac tttatagaag 8220
aaattttgag tttttgtttt tttttaataa ataaataaac ataaataaat tgtttgttga 8280
atttattatt agtatgtaag tgtaaatata ataaaactta atatctattc aaattaataa 8340
ataaacctcg atatacagac cgataaaaca catgcgtcaa ttttacgcat gattatcttt 8400
aacgtacgtc acaatatgat tatctttcta gggttaaata atagtttcta atttttttat 8460
tattcagcct gctgtcgtga ataccgtata tctcaacgct gtctgtgaga ttgtcgtatt 8520
ctagcctttt tagtttttcg ctcatcgact tgatattgtc cgacacattt tcgtcgattt 8580
gcgttttgat caaagacttg agcagagaca cgttaatcaa ctgttcaaat tgatccatat 8640
taacgatatc aacccgatgc gtatatggtg cgtaaaatat atttttaac cctcttatac 8700
tttgcactct gcgttaatac gcgttcgtgt acagacgtaa tcatgttttc ttttttggat 8760
aaaactccta ctgagtttga cctcatatta gaccctcaca agttgcaaaa cgtggcattt 8820
tttaccaatg aagaatttaa agttatttta aaaaatttca tcacagattt aaagaagaac 8880
caaaaattaa attatttcaa cagtttaatc gaccagttaa tcaacgtgta cacagacgcg 8940
tcggcaaaaa acacgcagcc cgacgtgttg gctaaaatta ttaaatcaac ttgtgttata 9000
gtcacggatt tgccgtccaa cgtgttcctc aaaaagttga agaccaacaa gtttacggac 9060
actattaatt atttgatttt gccccacttc attttgtggg atcacaattt tgttatattt 9120
taaacaaagc ttggcactgg ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt 9180
tacccaactt aatcgccttg cagcacatcc cccttcgcc agctggcgta atagcgaaga 9240
ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg aatggcgaat ggcgcctgat 9300
gcggtatttt ctccttacgc atctgtgcgg tatttcacac cgcatatggt gcactctcag 9360
tacaatctgc tctgatgccg catagttaag ccagccccga cacccgccaa cacccgctga 9420
cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg tgaccgtctc 9480
cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga gacgaaaggg 9540
cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt cttagacgtc 9600
aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca 9660
ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 9720
aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt 9780
ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 9840
gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 9900
ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc 9960
ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac actattctca 10020
gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 10080
aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct 10140
gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt 10200
aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 10260
caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 10320
tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 10380
acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga 10440
gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt 10500
agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga 10560
gataggtgcc tcactgatta agcattggta actgtcagac caagtttact catatatact 10620
ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga tccttttttga 10680
taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 10740
agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 10800
aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 10860
ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta 10920
gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 10980
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc 11040

aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggttt cgtgcacaca 11100
gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agcattgaga 11160
aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg 11220
aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt 11280
cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag 11340
cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt 11400
tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt 11460
tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt cagtgagcga 11520
ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta 11580
atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa 11640
tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc cggctcgtat 11700
gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta 11760
cgaatttcga cctgcaggca tgcaagcttg catgcctgca ggtcgacgct cgcgcgactt 11820
ggtttgccat tctttagcgc gcgtcgcgtc acacagcttg gccacaat 11868

<210> 50
<211> 11868
<212> ADN
<213> nhân tạo
<220>
<223> Trình tự của cấu trúc pLA3077-a Cctra intron-tTAV
<400> 50
gtggtttttg tcaaacgaag attctatgac gtgtttaaag tttaggtcga gtaaagcgca 60
aatctttttt aaccctagaa agatagtctg cgtaaaattg acgcatgcat tcttgaaata 120
ttgctctctc tttctaaata gcgcgaatcc gtcgctgtgc atttaggaca tctcagtcgc 180
cgcttggagc tcccgtgagg cgtgcttgtc aatgcggtaa gtgtcactga ttttgaacta 240
taacgaccgc gtgagtcaaa atgacgcatg attatctttt acgtgacttt taagatttaa 300
ctcatacgat aattatattg ttatttcatg ttctacttac gtgataactt attatatata 360
tattttcttg ttatagatat cgtgactaat atataataaa atgggtagtt ctttagacga 420
tgagcatatc ctctctgctc ttctgcaaag cgatgacgag cttgttggtg aggattctga 480
cagtgaaata tcagatcacg taagtgaaga tgacgtccag agcgatacag aagaagcgtt 540
tatagatgag gtacatgaag tgcagccaac gtcaagcggt agtgaaatat tagacgaaca 600
aaatgttatt gaacaaccag gttcttcatt ggcttctaac agaatcttga ccttgccaca 660
gaggactatt agaggtaaga ataaacattg ttggtcaact tcaaagtcca cgaggcgtag 720
ccgagtctct gcactgaaca ttgtcagatc ggcccgggcg ccgtttttct tgaaatattg 780
ctctctcttt ctaaatagcg cgaatccgtc gctgtgcatt taggacatct cagtcgccgc 840
ttggagctcc caaacgcgcc agtggtagta cacagtactg tgggtgttca gtttgaaatc 900
ctcttgcttc tccattgtct cggttacctt tggtcaaatc catgggttct attgcctata 960
tactcttgcg attaccagtg attgcgctat tagctattag atggattgtt ggccaaactt 1020
gtcgcttaag tggctgggaa ttgtaaccgt aggcccgagt gtaatgatcc cccataaaaa 1080
gttttcgcaa tgcctttatt ttttgttgca aatctctctt tattctgcgg tattcttcat 1140
tattgcgggg atggggaaag tgtttatata gaagcaactt acgattgaac ccaaatgcac 1200
ctgacaagca aggtcaaagg gccagatttt taaatatatt atttagtctt aggactctct 1260
atttgcaatt aaattacttt gctacctgag ggttaaatct tccccattga taataataat 1320
tccactatat gttcaattgg gtttcaccgc gcttagttac atgacgagcc ctaatgagcc 1380
gtcggtggtc tataaactgt gccttacaaa tacttgcaac tcttctcgtt ttgaagtcag 1440
cagagttatt gctaattgct aattgctaat tgcttttaac tgatttcttc gaaattggtg 1500
ctatgtttat ggcgctatta acaagtatga atgtcaggtt taaccagggg atgcttaatt 1560
gtgttctcaa cttcaaaggc agaaatgttt actcttgacc atgggtttag gtataatgtt 1620

atcaagctcc tcgagttaac gttacgttaa cgttaacgtt cgaggtcgac tctagaacta 1680
cccaccgtac tcgtcaattc caagggcatc ggtaaacatc tgctcaaact cgaagtcggc 1740
catatccaga gcgccgtagg gggcggagtc gtgggggggta aatcccggac ccgggggaatc 1800
cccgtccccc aacatgtcca gatcgaaatc gtctagcgcg tcggcatgcg ccatcgccac 1860
gtcctcgccg tctaagtgga gctcgtcccc caggctgaca tcggtcgggg gggccgtcga 1920
cagtctgcgc gtgtgtcccg cggggagaaa ggacaggcgc ggagccgcca gccccgcctc 1980
ttcgggggcg tcgtcgtccg ggagatcgag caggccctcg atggtagacc cgtaattgtt 2040
tttcgtacgc gcgcggctgt acgcggggcc cgagcccgac tcgcatttca gttgcttttc 2100
caatccgcag ataatcagct ccaagccgaa caggaatgcc ggctcggctc cttgatgatc 2160
gaacagctcg attgcctgac gcagcagtgg gggcatcgaa tcggttgttg gggtctcgcg 2220
ctcctctttt gcgacttgat gctcttggtc ctccagcacg cagcccaggg taaagtgacc 2280
gacggcgctc agagcgtaga gagcattttc caggctgaag ccttgctggc acaggaacgc 2340
gagctggttc tccagtgtct cgtattgctt ttcggtcggg cgcgtgccga gatggacttt 2400
ggcaccgtct cggtgggaca gcagagcgca gcggaacgac ttggcgttat tgcggaggaa 2460
gtcctgccag gactcgcctt ccaacgggca aaaatgcgtg tggtggcggt cgagcatctc 2520
gatggccagg gcatccagca gcgcccgctt attcttcacg tgccagtaga gggtgggctg 2580
ctccacgccc agcttctgcg ccaacttgcg ggtcgtcagt ccctcaatac ctatagatac 2640
catagatgta tggattagta tcatatacat acaaaggcta tttttgggac atattaatat 2700
taacaatttc cgtgatagtt ttcaccattt ttgttgaatg ttacgttgaa aatttaaatt 2760
tgttttaaat taattttacc agtcatgtgt tcttaaaagt ttttatgatt gaaacggcat 2820
aaagtggttc aaaaatttat caagaaaggc tttccttttt taaatcttat ctttttctct 2880
taaaaatcac tagtcaattc attattaatt tgttaacttg aatttggaat gtctatttac 2940
tttcagataa attaaagcaa gaaacttaat attcgaaaaa aattgattct aaatggaatt 3000
tcacttgatc ttcatgtatg catatcaatt tttatttaca ttgtataata agtttcgagt 3060
tgattgttgt aatccacagg tgtcccagag aattaaattc caaattaccc aagtttattg 3120
aatgttgatt gtagtttcag ttgctttgtt gctgcaacaa tggcttgttg attgtagata 3180
ttttcccttt ccttggttta cttattacat agactgaaaa agaggtttac ttttttgata 3240
cttatgaaaa atttctatta gtgattacta accaatcgct atatgtttac tagaaaacaa 3300
ataaactctt tacattaaca ttcaataatg tttgctctgt aaccgacaat tgaaggcgtt 3360
acagcaacag taatataact agcttcttaa ccctcatcta ttaaccccat cgtttaaaac 3420
actatgttaa atggtctaac aaatctagat actaatagat gtcttattac ttagcagcca 3480
cagctgcaac atccaagaca atttttgaaa cttcttattg agctcttggc agcagaaatg 3540
ttggtatttt tcacagcttt ctgaaagacc ggcaccttcc tccggttccc gtttctgaat 3600
tcaagaggat ttccgacccc caattaatcc cgaaacaaat aaggtatatt caaaatgatg 3660
gaaaagtcat ggctgctgac cttatttttа ttcctattga tagaatatta ttccccttt 3720
aaatacactg tactaagagg tccggctata attttactca cttgtcgatt atcccataga 3780
atgttgattg tagttggttg cttttccagg tgagagttga tcaagtcaca aaagttagcg 3840
tgtgttgatt gtagatttga aggtaaaata atttttgcac ccattcatcg ggtaaaacgt 3900
tctccataga atacatttcc atcgataatt gataacttat gaatttcaaa gaaaaaaata 3960
tgcttttaaa attaccaact tcgttcaaca gctccaacgc ggagttgatg actttggact 4020
tatccaggcg gctgcccatg gtggtttcta aaggtgttat aaatcaaatt agttttgttt 4080
tttcttgaaa actttgcgtt tcctttgatc aacttaccgc cagggtacct gcagattgtt 4140
tagcttgttc agctgcgctt gtttatttgc ttagctttcg cttagcgacg tgttcacttt 4200
gcttgtttga attgaattgt cgctccgtag acgaagcgcc tctatttata ctccggcgct 4260
cgttttcgag tttaccactc cctatcagtg atagagaaaa gtgaaagtcg agtttaccac 4320
tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc agtgatagag 4380
aaaagtgaaa gtcgagttta ccactcccta tcagtgatag agaaaagtga aagtcgagtt 4440
taccactccc tatcagtgat agagaaaagt gaaagtcgag tttaccactc cctatcagtg 4500
atagagaaaa gtgaaagtcg agtttaccac tccctatcag tgatagagaa aagtgaaagt 4560

cgaaacctgg cgcgccccgg ccatcgagaa agagagagag aagagaagag agagaacatt 4620
cgagaaagag agagagaaga gaagagagag aacatactcc ctatcagtga tagagaagtc 4680
cctatcagtg atagagatgt ccctatcagt gatagagagt tccctatcag tgatagagac 4740
gtccctatca gtgatagaga agtccctatc agtgatagag agatccctat cagtgataga 4800
gatttcccta tcagtgatag agaggtccct atcagtgata gagacttccc tatcagtgat 4860
agagaaatcc ctatcagtga tagagacatc cctatcagtg atagagaact ccctatcagt 4920
gatagagacc tccctatcag tgatagagat cgatgcggcc gcatggtacc cattgcttgt 4980
catttattaa tttggatgat gtcatttgtt tttaaaattg aactggcttt acgagtagaa 5040
ttctacgcgt aaaacacaat caagtatgag tcataatctg atgtcatgtt ttgtacacgg 5100
ctcataaccg aactggcttt acgagtagaa ttctacttgt aatgcacgat cagtggatga 5160
tgtcatttgt ttttcaaatc gagatgatgt catgttttgc acacggctca taaactcgct 5220
ttacgagtag aattctacgt gtaacgcacg atcgattgat gagtcatttg ttttgcaata 5280
tgatatcata caatatgact catttgtttt tcaaaaccga acttgattta cgggtagaat 5340
tctacttgta aagcacaatc aaaaagatga tgtcatttgt ttttcaaaac tgaactcgct 5400
ttacgagtag aattctacgt gtaaaacaca atcaagaaat gatgtcattt gttataaaaa 5460
taaaagctga tgtcatgttt tgcacatggc tcataactaa actcgcttta cgggtagaat 5520
tctacgcgta aaacatgatt gataattaaa taattcattt gcaagctata cgttaaatca 5580
aacggacgct cgaggttgca caacactatt atcgatttgc agttcgggac ataaatgttt 5640
aaatatatcg atgtctttgt gatgcgcgcg acatttttgt aggttattga taaaatgaac 5700
ggatacgttg cccgacatta tcattaaatc cttggcgtag aatttgtcgg gtccattgtc 5760
cgtgtgcgct agcatgcccg taacggacct cgtacttttg gcttcaaagg ttttgcgcac 5820
agacaaaatg tgccacactt gcagctctgc atgtgtgcgc gttaccacaa atcccaacgg 5880
cgcagtgtac ttgttgtatg caaataaatc tcgataaagg cgcggcgcgc gaatgcagct 5940
gatcacgtac gctcctcgtg ttccgttcaa ggacggtgtt atcgacctca gattaatgtt 6000
tatcggccga ctgttttcgt atccgctcac caaacgcgtt tttgcattaa cattgtatgt 6060
cggcggatgt tctatatcta atttgaataa ataaacgata accgcgttgg ttttagaggg 6120
cataataaaa gaaatattgt tatcgtgttc gccattaggg cagtataaat tgacgttcat 6180
gttggatatt gtttcagttg caagttgaca ctggcggcga caagcaattc taattggggt 6240
aagttttccc gttcttttct gggttcttcc cttttgctca tccttgctgc actaccttca 6300
ggtgcaagtt gagattcagg ccaccatggg agatcccacc ccacccaaga agaagcgcaa 6360
accggtcgcc accatggcct cctccgagaa cgtcatcacc gagttcatgc gcttcaaggt 6420
gcgcatggag ggcaccgtga acggccacga gttcgagatc gagggcgagg gcgagggccg 6480
cccctacgag ggccacaaca ccgtgaagct gaaggtgacc aagggcggcc ccctgccctt 6540
cgcctgggac atcctgtccc cccagttcca gtacggctcc aaggtgtacg tgaagcaccc 6600
cgccgacatc cccgactaca agaagctgtc cttccccgag ggcttcaagt gggagcgcgt 6660
gatgaacttc gaggacggcg gcgtggcgac cgtgacccag gactcctccc tgcaggacgg 6720
ctgcttcatc tacaaggtga agttcatcgg cgtgaacttc ccctccgacg gccccgtgat 6780
gcagaagaag accatgggct gggaggcctc caccgagcgc ctgtaccccc gcgacggcgt 6840
gctgaagggc gagacccaca aggccctgaa gctgaaggac ggcggccact acctggtgga 6900
gttcaagtcc atctacatgg ccaagaagcc cgtgcagctg cccggctact actacgtgga 6960
cgccaagctg gacatcacct cccacaacga ggactacacc atcgtggagc agtacgagcg 7020
caccgagggc cgccaccacc tgttcctgag atctcgaccc aagaaaaagc ggaaggtgga 7080
ggacccgtaa gatccaccgg atctagataa ctgatcataa tcagccatac cacatttgta 7140
gaggttttac ttgctttaaa aaacctccca cacctccccc tgaacctgaa acataaaatg 7200
aatgcaattg ttgttgttaa cttgtttatt gcagcttata atggttacaa ataaagcaat 7260
agcatcacaa atttcacaaa taaagcattt ttttcactgc attctagttg tggtttgtcc 7320
aaactcatca atgtatctta acgcgagtta attaaggccg ctcatttaaa tctggccggc 7380
cgcaaccatt gtgggaaccg tgcgatcaaa caaacgcgag ataccggaag tactgaaaaa 7440
cagtcgctcc aggccagtgg gaacatcgat gttttgtttt gacggacccc ttactctcgt 7500

ctcatataaa ccgaagccag ctaagatggt atacttatta tcatcttgtg atgaggatgc 7560
ttctatcaac gaaagtaccg gtaaaccgca aatggttatg tattataatc aaactaaagg 7620
cggagtggac acgctagacc aaatgtgttc tgtgatgacc tgcagtagga agacgaatag 7680
gtggcctatg gcattattgt acggaatgat aaacattgcc tgcataaatt cttttattat 7740
atacagccat aatgtcagta gcaagggaga aaaggtccaa agtcgcaaaa aatttatgag 7800
aaacctttac atgagcctga cgtcatcgtt tatgcgtaag cgtttagaag ctcctacttt 7860
gaagagatat ttgcgcgata atatctctaa tattttgcca aatgaagtgc ctggtacatc 7920
agatgacagt actgaagagc cagtaatgaa aaaacgtact tactgtactt actgcccctc 7980
taaaataagg cgaaaggcaa atgcatcgtg caaaaaatgc aaaaaagtta tttgtcgaga 8040
gcataatatt gatatgtgcc aaagttgttt ctgactgact aataagtata atttgtttct 8100
attatgtata agttaagcta attacttatt ttataataca acatgactgt ttttaaagta 8160
caaaataagt ttatttttgt aaaagagaga atgtttaaaa gttttgttac tttatagaag 8220
aaattttgag tttttgtttt tttttaataa ataaataaac ataaataaat tgtttgttga 8280
atttattatt agtatgtaag tgtaaatata ataaaactta atatctattc aaattaataa 8340
ataaacctcg atatacagac cgataaaaca catgcgtcaa ttttacgcat gattatcttt 8400
aacgtacgtc acaatatgat tatctttcta gggttaaata atagtttcta atttttttat 8460
tattcagcct gctgtcgtga ataccgtata tctcaacgct gtctgtgaga ttgtcgtatt 8520
ctagcctttt tagtttttcg ctcatcgact tgatattgtc cgacacattt tcgtcgattt 8580
gcgttttgat caaagacttg agcagagaca cgttaatcaa ctgttcaaat tgatccatat 8640
taacgatatc aacccgatgc gtatatggtg cgtaaaatat attttttaac cctcttatac 8700
tttgcactct gcgttaatac gcgttcgtgt acagacgtaa tcatgttttc ttttttggat 8760
aaaactccta ctgagtttga cctcatatta gaccctcaca agttgcaaaa cgtggcattt 8820
tttaccaatg aagaatttaa agttatttta aaaaatttca tcacagattt aaagaagaac 8880
caaaaattaa attatttcaa cagtttaatc gaccagttaa tcaacgtgta cacagacgcg 8940
tcggcaaaaa acacgcagcc cgacgtgttg gctaaaatta ttaaatcaac ttgtgttata 9000
gtcacggatt tgccgtccaa cgtgttcctc aaaaagttga agaccaacaa gtttacggac 9060
actattaatt atttgatttt gccccacttc attttgtggg atcacaattt tgttatattt 9120
taaacaaagc ttggcactgg ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt 9180
tacccaactt aatcgccttg cagcacatcc cccttccgcc agctggcgta atagcgaaga 9240
ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg aatggcgaat ggcgcctgat 9300
gcggtatttt ctccttacgc atctgtgcgg tatttcacac cgcatatggt gcactctcag 9360
tacaatctgc tctgatgccg catagttaag ccagccccga cacccgccaa cacccgctga 9420
cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg tgaccgtctc 9480
cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga gacgaaaggg 9540
cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt cttagacgtc 9600
aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca 9660
ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa 9720
aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt 9780
ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca 9840
gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag 9900
ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc 9960
ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac actattctca 10020
gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt 10080
aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct 10140
gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt 10200
aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga 10260
caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact 10320
tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc 10380
acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga 10440

gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt 10500
agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga 10560
gataggtgcc tcactgatta agcattggta actgtcagac caagtttact catatatact 10620
ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga tcctttttga 10680
taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt cagaccccgt 10740
agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct gctgcttgca 10800
aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc taccaactct 10860
ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc ttctagtgta 10920
gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc tcgctctgct 10980
aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg ggttggactc 11040
aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggttt cgtgcacaca 11100
gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg agcattgaga 11160
aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg gcagggtcgg 11220
aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt atagtcctgt 11280
cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag gggggcggag 11340
cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt gctggccttt 11400
tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt 11460
tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt cagtgagcga 11520
ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta 11580
atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca acgcaattaa 11640
tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc cggctcgtat 11700
gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg accatgatta 11760
cgaatttcga cctgcaggca tgcaagcttg catgcctgca ggtcgacgct cgcgcgactt 11820
ggtttgccat tctttagcgc gcgtcgcgtc acacagcttg gccacaat 11868

<210> 51
<211> 11788
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự 51 của cấu trúc pLA3097-a Cctra intron-tTAV
<400> 51
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt tacctttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcgggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gaactaccca ccgtactcgt caattccaag ggcatcggta 960
aacatctgct caaactcgaa gtcggccata tccagagcgc cgtagggggc ggagtcgtgg 1020

ggggtaaatc ccggacccgg ggaatccccg tcccccaaca tgtccagatc gaaatcgtct 1080
agcgcgtcgg catgcgccat cgccacgtcc tcgccgtcta agtggagctc gtcccccagg 1140
ctgacatcgg tcgggggggc cgtcgacagt ctgcgcgtgt gtcccgcggg gagaaaggac 1200
aggcgcggag ccgccagccc cgcctcttcg ggggcgtcgt cgtccgggag atcgagcagg 1260
ccctcgatgg tagacccgta attgttttte gtacgcgcgc ggctgtacgc ggggcccgag 1320
cccgactcgc atttcagttg cttttccaat ccgcagataa tcagctccaa gccgaacagg 1380
aatgccggct cggctccttg atgatcgaac agctcgattg cctgacgcag cagtgggggc 1440
atcgaatcgg ttgttggggt ctcgcgctcc tcttttgcga cttgatgctc ttggtcctcc 1500
agcacgcagc ccagggtaaa gtgaccgacg gcgctcagag cgtagagagc attttccagg 1560
ctgaagcctt gctggcacag gaacgcgagc tggttctcca gtgtctcgta ttgcttttcg 1620
gtcgggcgcg tgccgagatg gactttggca ccgtctcggt gggacagcag agcgcagcgg 1680
aacgacttgg cgttattgcg gaggaagtcc tgccaggact cgccttccaa cgggcaaaaa 1740
tgcgtgtggt ggcggtcgag catctcgatg gccagggcat ccagcagcgc ccgcttattc 1800
ttcacgtgcc agtagagggt gggctgctcc acgcccagct tctgcgccaa cttgcgggtc 1860
gtcagtccct caatgccaac ttcgttcaac agctccaacg cggagttgat gactttggac 1920
ttatccaggc ggctgaccta tagataccat agatgtatgg attagtatca tatacataca 1980
aaggctattt ttgggacata ttaatattaa caatttccgt gatagttttc accatttttg 2040
ttgaatgtta cgttgaaaat ttaaatttgt tttaaattaa ttttaccagt catgtgttct 2100
taaaagtttt tatgattgaa acggcataaa gtggttcaaa aatttatcaa gaaaggcttt 2160
ccttttttaa atcttatctt tttctcttaa aaatcactag tcaattcatt attaatttgt 2220
taacttgaat ttggaatgtc tatttacttt cagataaatt aaagcaagaa acttaatatt 2280
cgaaaaaaat tgattctaaa tggaatttca cttgatcttc atgtatgcat atcaattttt 2340
atttacattg tataataagt ttcgagttga ttgttgtaat ccacaggtgt cccagagaat 2400
taaattccaa attacccaag tttattgaat gttgattgta gtttcagttg ctttgttgct 2460
gcaacaatgg cttgttgatt gtagatattt tcccttccct tggtttactt attacataga 2520
ctgaaaaaga ggtttacttt tttgatactt atgaaaaatt tctattagtg attactaacc 2580
aatcgctata tgtttactag aaaacaaata aactctttac attaacattc aataatgttt 2640
gctctgtaac cgacaattga aggcgttaca gcaacagtaa tataactagc ttcttaaccc 2700
tcatctatta accccatcgt ttaaaacact atgttaaatg gtctaacaaa tctagatact 2760
aatagatgtc ttattactta gcagccacag ctgcaacatc caagacaatt tttgaaactt 2820
cttattgagc tcttggcagc agaaatgttg gtatttttca cagctttctg aaagaccggc 2880
accttcctcc ggttcccgtt tctgaattca agaggatttc cgacccccaa ttaatcccga 2940
aacaaataag gtatattcaa aatgatggaa aagtcatggc tgctgacctt atttttattc 3000
ctattgatag aatattattc ccctttttaaa tacactgtac taagaggtcc ggctataatt 3060
ttactcactt gtcgattatc ccatagaatg ttgattgtag ttggttgctt ttccaggtga 3120
gagttgatca agtcacaaaa gttagcgtgt gttgattgta gatttgaagg taaaataatt 3180
tttgcaccca ttcatcgggt aaaacgttct ccatagaata catttccatc gataattgat 3240
aacttatgaa tttcaaagaa aaaaatatgc ttttaaaatt accatggtgg ctagcgcaga 3300
ttgtttagct tgttcagctg cgcttgttta tttgcttagc tttcgcttag cgacgtgttc 3360
actttgcttg tttgaattga attgtcgctc cgtagacgaa gcgcctctat ttatactccg 3420
gcgctcgttt tcgagtttac cactccctat cagtgataga gaaaagtgaa agtcgagttt 3480
accactccct atcagtgata gagaaaagtg aaagtcgagt ttaccactcc ctatcagtga 3540
tagagaaaag tgaaagtcga gtttaccact ccctatcagt gatagagaaa agtgaaagtc 3600
gagtttacca ctccctatca gtgatagaga aaagtgaaag tcgagtttac cactccctat 3660
cagtgataga gaaaagtgaa agtcgagttt accactccct atcagtgata gagaaaagtg 3720
aaagtcgaaa cctggcgcgc cccggccatc gagaaagaga gagagaagag aagagagaga 3780
acattcgaga aagagagaga gaagagaaga gagagaacat actccctatc agtgatagag 3840
aagtccctat cagtgataga gatgtcccta tcagtgatag agagttccct atcagtgata 3900
gagacgtccc tatcagtgat agagaagtcc ctatcagtga tagagagatc cctatcagtg 3960

atagagattt ccctatcagt gatagagagg tccctatcag tgatagagac ttccctatca 4020
gtgatagaga aatccctatc agtgatagag acatccctat cagtgataga gaactcccta 4080
tcagtgatag agacctccct atcagtgata gagatcgatg cggccgcatg gtacccattg 4140
cttgtcattt attaatttgg atgatgtcat ttgtttttaa aattgaactg gctttacgag 4200
tagaattcta cgcgtaaaac acaatcaagt atgagtcata atctgatgtc atgttttgta 4260
cacggctcat aaccgaactg gctttacgag tagaattcta cttgtaatgc acgatcagtg 4320
gatgatgtca tttgtttttc aaatcgagat gatgtcatgt tttgcacacg gctcataaac 4380
tcgctttacg agtagaattc tacgtgtaac gcacgatcga ttgatgagtc atttgttttg 4440
caatatgata tcatacaata tgactcattt gtttttcaaa accgaacttg atttacgggt 4500
agaattctac ttgtaaagca caatcaaaaa gatgatgtca tttgtttttc aaaactgaac 4560
tcgctttacg agtagaattc tacgtgtaaa acacaatcaa gaaatgatgt catttgttat 4620
aaaaataaaa gctgatgtca tgttttgcac atggctcata actaaactcg ctttacgggt 4680
agaattctac gcgtaaaaca tgattgataa ttaaataatt catttgcaag ctatacgtta 4740
aatcaaacgg acgctcgagg ttgcacaaca ctattatcga tttgcagttc gggacataaa 4800
tgtttaaata tatcgatgtc tttgtgatgc gcgcgacatt tttgtaggtt attgataaaa 4860
tgaacggata cgttgcccga cattatcatt aaatccttgg cgtagaattt gtcgggtcca 4920
ttgtccgtgt gcgctagcat gcccgtaacg gacctcgtac ttttggcttc aaaggttttg 4980
cgcacagaca aaatgtgcca cacttgcagc tctgcatgtg tgcgcgttac cacaaatccc 5040
aacggcgcag tgtacttgtt gtatgcaaat aaatctcgat aaaggcgcgg cgcgcgaatg 5100
cagctgatca cgtacgctcc tcgtgttccg ttcaaggacg gtgttatcga cctcagatta 5160
atgtttatcg gccgactgtt ttcgtatccg ctcaccaaac gcgttttgc attaacattg 5220
tatgtcggcg gatgttctat atctaatttg aataaataaa cgataaccgc gttggtttta 5280
gagggcataa taaaagaaat attgttatcg tgttcgccat tagggcagta taaattgacg 5340
ttcatgttgg atattgtttc agttgcaagt tgacactggc ggcgacaagc aattctaatt 5400
ggggtaagtt ttcccgttct tttctgggtt cttccctttt gctcatcctt gctgcactac 5460
cttcaggtgc aagttgagat tcaggccacc atgggagatc ccaccccacc caagaagaag 5520
cgcaaaccgg tcgccaccat ggcctcctcc gagaacgtca tcaccgagtt catgcgcttc 5580
aaggtgcgca tggagggcac cgtgaacggc cacgagttcg agatcgaggg cgagggcgag 5640
ggccgccccc acgagggcca caacaccgtg aagctgaagg tgaccaaggg cggccccctg 5700
cccttcgcct gggacatcct gtcccccccag ttccagtacg gctccaaggt gtacgtgaag 5760
caccccgccg acatccccga ctacaagaag ctgtccttcc ccgagggctt caagtgggag 5820
cgcgtgatga acttcgagga cggcggcgtg gcgaccgtga cccaggactc ctccctgcag 5880
gacggctgct tcatctacaa ggtgaagttc atcggcgtga acttcccctc cgacggcccc 5940
gtgatgcaga agaagaccat gggctgggag gcctccaccg agcgcctgta cccccgcgac 6000
ggcgtgctga agggcgagac ccacaaggcc ctgaagctga aggacggcgg ccactacctg 6060
gtggagttca agtccatcta catggccaag aagcccgtgc agctgcccgg ctactactac 6120
gtggacgcca agctggacat cacctcccac aacgaggact acaccatcgt ggagcagtac 6180
gagcgcaccg agggccgcca ccacctgttc ctgagatctc gacccaagaa aaagcggaag 6240
gtggaggacc cgtaagatcc accggatcta gataactgat cataatcagc ctaccacat 6300
ttgtagaggt tttacttgct ttaaaaaacc tcccacacct ccccctgaac ctgaaacata 6360
aaatgaatgc aattgttgtt gttaacttgt ttattgcagc ttataatggt tacaaataaa 6420
gcaatagcat cacaaatttc acaaataaag catttttttc actgcattct agttgtggtt 6480
tgtccaaact catcaatgta tcttaacgcg agttaattaa ggccgctcat ttaaatctgg 6540
ccggccgcaa ccattgtggg aaccgtgcga tcaaacaaac gcgagatacc ggaagtactg 6600
aaaaacagtc gctccaggcc agtgggaaca tcgatgtttt gttttgacgg accccttact 6660
ctcgtctcat ataaaccgaa gccagctaag atggtatact tattatcatc ttgtgatgag 6720
gatgcttcta tcaacgaaag taccggtaaa ccgcaaatgg ttatgtatta taatcaaact 6780
aaaggcggag tggacacgct agaccaaatg tgttctgtga tgacctgcag taggaagacg 6840
aataggtggc ctatggcatt attgtacgga atgataaaca ttgcctgcat aaattctttt 6900

attatataca gccataatgt cagtagcaag ggagaaaagg tccaaagtcg caaaaaattt 6960
atgagaaacc tttacatgag cctgacgtca tcgtttatgc gtaagcgttt agaagctcct 7020
actttgaaga gatatttgcg cgataatatc tctaatattt tgccaaatga agtgcctggt 7080
acatcagatg acagtactga agagccagta atgaaaaaac gtacttactg tacttactgc 7140
ccctctaaaa taaggcgaaa ggcaaatgca tcgtgcaaaa aatgcaaaaa agttatttgt 7200
cgagagcata atattgatat gtgccaaagt tgtttctgac tgactaataa gtataatttg 7260
tttctattat gtataagtta agctaattac ttattttata atacaacatg actgttttta 7320
aagtacaaaa taagtttatt tttgtaaaag agagaatgtt taaaagtttt gttactttat 7380
agaagaaatt ttgagttttt gttttttttt aataaataaa taaacataaa taaattgttt 7440
gttgaattta ttattagtat gtaagtgtaa atataataaa acttaatatc tattcaaatt 7500
aataaataaa cctcgatata cagaccgata aaacacatgc gtcaatttta cgcatgatta 7560
tctttaacgt acgtcacaat atgattatct ttctagggtt aaataatagt ttctaatttt 7620
tttattattc agcctgctgt cgtgaatacc gtatatctca acgctgtctg tgagattgtc 7680
gtattctagc ctttttagtt tttcgctcat cgacttgata ttgtccgaca cattttcgtc 7740
gatttgcgtt ttgatcaaag acttgagcag agacacgtta atcaactgtt caaattgatc 7800
catattaacg atatcaaccc gatgcgtata tggtgcgtaa aatatatttt ttaaccctct 7860
tatactttgc actctgcgtt aatacgcgtt cgtgtacaga cgtaatcatg ttttcttttt 7920
tggataaaac tcctactgag tttgacctca tattagaccc tcacaagttg caaaacgtgg 7980
catttttac caatgaagaa tttaaagtta ttttaaaaaa tttcatcaca gatttaaaga 8040
agaaccaaaa attaaattat ttcaacagtt taatcgacca gttaatcaac gtgtacacag 8100
acgcgtcggc aaaaaacacg cagcccgacg tgttggctaa aattattaaa tcaacttgtg 8160
ttatagtcac ggatttgccg tccaacgtgt tcctcaaaaa gttgaagacc aacaagttta 8220
cggacactat taattatttg attttgcccc acttcatttt gtgggatcac aattttgtta 8280
tattttaaac aaagcttggc actggccgtc gttttacaac gtcgtgactg ggaaaaccct 8340
ggcgttaccc aacttaatcg ccttgcagca catccccctt tcgccagctg gcgtaatagc 8400
gaagaggccc gcaccgatcg ccccttcccaa cagttgcgca gcctgaatgg cgaatggcgc 8460
ctgatgcggt attttctcct tacgcatctg tgcggtattt cacaccgcat atggtgcact 8520
ctcagtacaa tctgctctga tgccgcatag ttaagccagc cccgacaccc gccaacaccc 8580
gctgacgcgc cctgacgggc ttgtctgctc ccggcatccg cttacagaca agctgtgacc 8640
gtctccggga gctgcatgtg tcagaggttt tcaccgtcat caccgaaacg cgcgagacga 8700
aagggcctcg tgatacgcct atttttatag gttaatgtca tgataataat ggtttcttag 8760
acgtcaggtg gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa 8820
atacattcaa atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat 8880
tgaaaaagga agagtatgag tattcaacat ttccgtgtcg cccttattcc ctttttgcg 8940
gcattttgcc ttcctgtttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa 9000
gatcagttgg gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt 9060
gagagttttc gccccgaaga acgttttcca atgatgagca cttttaaagt tctgctatgt 9120
ggcgcggtat tatcccgtat tgacgccggg caagagcaac tcggtcgccg catacactat 9180
tctcagaatg acttggttga gtactcacca gtcacagaaa agcatcttac ggatggcatg 9240
acagtaagag aattatgcag tgctgccata accatgagtg ataacactgc ggccaactta 9300
cttctgacaa cgatcggagg accgaaggag ctaaccgctt ttttgcacaa catgggggat 9360
catgtaactc gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag 9420
cgtgacacca cgatgcctgt agcaatggca acaacgttgc gcaaactatt aactggcgaa 9480
ctacttactc tagcttcccg gcaacaatta atagactgga tggaggcgga taaagttgca 9540
ggaccacttc tgcgctcggc ccttccggct ggctggttta ttgctgataa atctggagcc 9600
ggtgagcgtg ggtctcgcgg tatcattgca gcactggggc cagatggtaa gccctcccgt 9660
atcgtagtta tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc 9720
gctgagatag gtgcctcact gattaagcat tggtaactgt cagaccaagt ttactcatat 9780
atactttaga ttgatttaaa acttcatttt taatttaaaa ggatctaggt gaagatcctt 9840

tttgataatc tcatgaccaa aatcccttaa cgtgagtttt cgttccactg agcgtcagac 9900
cccgtagaaa agatcaaagg atcttcttga gatccttttt ttctgcgcgt aatctgctgc 9960
ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt tgccggatca agagctacca 10020
actcttttte cgaaggtaac tggcttcagc agagcgcaga taccaaatac tgtccttcta 10080
gtgtagccgt agttaggcca ccacttcaag aactctgtag caccgcctac atacctcgct 10140
ctgctaatcc tgttaccagt ggctgctgcc agtggcgata agtcgtgtct taccgggttg 10200
gactcaagac gatagttacc ggataaggcg cagcggtcgg gctgaacggg gggttcgtgc 10260
acacagccca gcttggagcg aacgacctac accgaactga gatacctaca gcgtgagcat 10320
tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca ggtatccggt aagcggcagg 10380
gtcggaacag gagagcgcac gagggagctt ccagggggaa acgcctggta tctttatagt 10440
cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt tgtgatgctc gtcaggggggg 10500
cggagcctat ggaaaaacgc cagcaacgcg gccttttttac ggttcctggc cttttgctgg 10560
ccttttgctc acatgttctt tcctgcgtta tcccctgatt ctgtggataa ccgtattacc 10620
gcctttgagt gagctgatac cgctcgccgc agccgaacga ccgagcgcag cgagtcagtg 10680
agcgaggaag cggaagagcg cccaatacgc aaaccgcctc tccccgcgcg ttggccgatt 10740
cattaatgca gctggcacga caggtttccc gactggaaag cgggcagtga gcgcaacgca 10800
attaatgtga gttagctcac tcattaggca ccccaggctt tacactttat gcttccggct 10860
cgtatgttgt gtggaattgt gagcggataa caatttcaca caggaaacag ctatgaccat 10920
gattacgaat ttcgacctgc aggcatgcaa gcttgcatgc ctgcaggtcg acgctcgcgc 10980
gacttggttt gccattcttt agcgcgcgtc gcgtcacaca gcttggccac aatgtggttt 11040
ttgtcaaacg aagattctat gacgtgttta aagtttaggt cgagtaaagc gcaaatcttt 11100
tttaacccta gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa atattgctct 11160
ctctttctaa atagcgcgaa tccgtcgctg tgcatttagg acatctcagt cgccgcttgg 11220
agctcccgtg aggcgtgctt gtcaatgcgg taagtgtcac tgattttgaa ctataacgac 11280
cgcgtgagtc aaaatgacgc atgattatct tttacgtgac ttttaagatt taactcatac 11340
gataattata ttgttatttc atgttctact tacgtgataa cttattatat atatattttc 11400
ttgttataga tatcgtgact aatatataat aaaatgggta gttctttaga cgatgagcat 11460
atcctctctg ctcttctgca aagcgatgac gagcttgttg gtgaggattc tgacagtgaa 11520
atatcagatc acgtaagtga agatgacgtc cagagcgata cagaagaagc gtttatagat 11580
gaggtacatg aagtgcagcc aacgtcaagc ggtagtgaaa tattagacga acaaaatgtt 11640
attgaacaac caggttcttc attggcttct aacagaatct tgaccttgcc acagaggact 11700
attagaggta agaataaaca ttgttggtca acttcaaagt ccacgaggcg tagccgagtc 11760
tctgcactga acattgtcag atcggccc 11788

<210> 52
<211> 13292
<212> ADN
<213> Nhân tao
<220> Trình tự của cấu trúc pLA3233-Cctra-intron-tTAV2
<223>
<400> 52
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt taccttttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480

tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gacaccggtg ttagccgccg tactcatcga tgcccagggc 960
gtcggtgaac atctgctcga actcgaaatc ggccatatcc agggcgccgt agggggcgct 1020
atcgtgcggg gtgaatcccg gtcccgggct atcgccatcg cccagcatgt ccaggtcgaa 1080
gtcgtccagg gcatcggcgt gggccatcgc cacatcctcg ccatccaggt gcagctcatc 1140
gcccaggctc acgtcggtcg gcggggcggt cgacaggcgg cgggtgtgtc cggccggcag 1200
gaagctcagg cgcgggggcgg ccaggccccgc ctcctccggg gcatcatcat ccggcagatc 1260
cagcaggccc tcgatggtgc tgccgtagtt gttcttggtg cgggcgcggc tgtaggcggg 1320
gcccgagccc gactcgcatt tcagttgctt ttccaatccg cagataatca gctccaagcc 1380
gaacaggaat gccggctcgg ctccttgatg atcgaacagc tcgattgcct gacgcagcag 1440
tgggggcatc gaatcggttg ttggggtctc gcgctcctct tttgcgactt gatgctcttg 1500
gtcctccagc acgcagccca gggtaaagtg accgacggcg ctcagagcgt agagagcatt 1560
ttccaggctg aagccttgct ggcacaggaa cgcgagctgg ttctccagtg tctcgtattg 1620
cttttcggtc gggcgcgtgc cgagatggac tttggcaccg tctcggtggg acagcagagc 1680
gcagcggaac gacttggcgt tattgcggag gaagtcctgc caggactcgc cttccaacgg 1740
gcaaaaatgc gtgtggtggc ggtcgagcat ctcgatggcc agggcatcca gcagcgcccg 1800
cttattcttc acgtgccagt agagggtggg ctgctccacg cccagcttct gcgccaactt 1860
gcgggtcgtc agtccctcaa tgccaacttc gttcaacagc tccaacgcgg agttgatgac 1920
tttggactta tccaggcggc tgacctatag ataccataga tgtatggatt agtatcatat 1980
acatacaaag gctatttttg ggacatatta atattaacaa tttccgtgat agttttcacc 2040
atttttgttg aatgttacgt tgaaaattta aatttgtttt aaattaattt taccagtcat 2100
gtgttcttaa aagtttttat gattgaaacg gcataaagtg gttcaaaaat ttatcaagaa 2160
aggctttcct tttttaaatc ttatcttttt ctcttaaaaa tcactagtca attcattatt 2220
aatttgttaa cttgaatttg gaatgtctat ttactttcag ataaattaaa gcaagaaact 2280
taatattcga aaaaaattga ttctaaatgg aatttcactt gatcttcatg tatgcatatc 2340
aatttttatt tacattgtat aataagtttc gagttgattg ttgtaatcca caggtgtccc 2400
agagaattaa attccaaatt acccaagttt attgaatgtt gattgtagtt tcagttgctt 2460
tgttgctgca acaatggctt gttgattgta gatattttcc ctttccttgg tttacttatt 2520
acatagactg aaaaagaggt ttacttttttt gatacttatg aaaaatttct attagtgatt 2580
actaaccaat cgctatatgt ttactagaaa acaaataaac tctttacatt aacattcaat 2640
aatgtttgct ctgtaaccga caattgaagg cgttacagca acagtaatat aactagcttc 2700
ttaaccctca tctattaacc ccatcgttta aaacactatg ttaaatggtc taacaaatct 2760
agatactaat agatgtctta ttacttagca gccacagctg caacatccaa gacaattttt 2820
gaaacttctt attgagctct tggcagcaga aatgttggta tttttcacag ctttctgaaa 2880
gaccggcacc ttcctccggt tcccgtttct gaattcaaga ggatttccga cccccaatta 2940
atcccgaaac aaataaggta tattcaaaat gatggaaaag tcatggctgc tgaccttatt 3000
tttattccta ttgatagaat attattcccc ttttaaatac actgtactaa gaggtccggc 3060
tataatttta ctcacttgtc gattatccca tagaatgttg attgtagttg gttgcttttc 3120
caggtgagag ttgatcaagt cacaaaagtt agcgtgtgtt gattgtagat ttgaaggtaa 3180
aataatttttt gcacccattc atcgggtaaa acgttctcca tagaatacat ttccatcgat 3240
aattgataac ttatgaattt caaagaaaaa aatatgcttt taaaattacc atggtggcta 3300
gcgcagattg tttagcttgt tcagctgcgc ttgtttattt gcttagcttt cgcttagcga 3360
cgtgttcact ttgcttgttt gaattgaatt gtcgctccgt agacgaagcg cctctattta 3420

tactccggcg ctcgttttcg agtttaccac tccctatcag tgatagagaa aagtgaaagt 3480
cgagtttacc actccctatc agtgatagag aaaagtgaaa gtcgagttta ccactcccta 3540
tcagtgatag agaaaagtga aagtcgagtt taccactccc tatcagtgat agagaaaagt 3600
gaaagtcgag tttaccactc cctatcagtg atagagaaaa gtgaaagtcg agtttaccac 3660
tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc agtgatagag 3720
aaaagtgaaa gtcgaaacct ggcgcgcccc ggccatcgag aaagagagag agaagagaag 3780
agagagaaca ttcgagaaag agagagagaa gagaagagag agaacatact ccctatcagt 3840
gatagagaag tccctatcag tgatagagat gtccctatca gtgatagaga gttccctatc 3900
agtgatagag acgtccctat cagtgataga gaagtcccta tcagtgatag agagatccct 3960
atcagtgata gagatttccc tatcagtgat agagaggtcc ctatcagtga tagagacttc 4020
cctatcagtg atagagaaat ccctatcagt gatagagaca tccctatcag tgatagagaa 4080
ctccctatca gtgatagaga cctccctatc agtgatagag atcgatgcgg ccgcatggta 4140
cccattgctt gtcatttatt aatttggatg atgtcatttg tttttaaaat tgaactggct 4200
ttacgagtag aattctacgc gtaaaacaca atcaagtatg agtcataatc tgatgtcatg 4260
ttttgtacac ggctcataac cgaactggct ttacgagtag aattctactt gtaatgcacg 4320
atcagtggat gatgtcattt gtttttcaaa tcgagatgat gtcatgtttt gcacacggct 4380
cataaactcg ctttacgagt agaattctac gtgtaacgca cgatcgattg atgagtcatt 4440
tgttttgcaa tatgatatca tacaatatga ctcatttgtt tttcaaaacc gaacttgatt 4500
tacgggtaga attctacttg taaagcacaa tcaaaaagat gatgtcattt gtttttcaaa 4560
actgaactcg ctttacgagt agaattctac gtgtaaaaca caatcaagaa atgatgtcat 4620
ttgttataaa aataaaagct gatgtcatgt tttgcacatg gctcataact aaactcgctt 4680
tacgggtaga attctacgcg taaaacatga ttgataatta aataattcat ttgcaagcta 4740
tacgttaaat caaacggacg ctcgaggttg cacaacacta ttatcgattt gcagttcggg 4800
acataaatgt ttaaatatat cgatgtcttt gtgatgcgcg cgacattttt gtaggttatt 4860
gataaaatga acggatacgt tgcccgacat tatcattaaa tccttggcgt agaatttgtc 4920
gggtccattg tccgtgtgcg ctagcatgcc cgtaacggac ctcgtacttt tggcttcaaa 4980
ggttttgcgc acagacaaaa tgtgccacac ttgcagctct gcatgtgtgc gcgttaccac 5040
aaatcccaac ggcgcagtgt acttgttgta tgcaaataaa tctcgataaa ggcgcggcgc 5100
gcgaatgcag ctgatcacgt acgctcctcg tgttccgttc aaggacggtg ttatcgacct 5160
cagattaatg tttatcggcc gactgttttc gtatccgctc accaaacgcg tttttgcatt 5220
aacattgtat gtcggcggat gttctatatc taatttgaat aaataaacga taaccgcgtt 5280
ggttttagag ggcataataa aagaaatatt gttatcgtgt tcgccattag ggcagtataa 5340
attgacgttc atgttggata ttgtttcagt tgcaagttga cactggcggc gacaagcaat 5400
tctaattggg gtaagttttc ccgttctttt ctgggttctt ccctttttgct catccttgct 5460
gcactacctt caggtgcaag ttgagattca ggccaccatg ggagatccca ccccacccaa 5520
gaagaagcgc aaaccggtcg ccaccatgga cgaggatggt tcagagggcg gccccgccct 5580
gttccagagc gacatgacct tcaaaatctt catcgacggc gaggtgaacg gccagaagtt 5640
caccatcgtg gccgacggca gcagcaagtt cccccacggc gacttcaacg tgcacgccgt 5700
gtgcgagacc ggcaagctgc ccatgagctg gaagcccatc tgccacctga tccagtacgg 5760
cgagcccttc ttcgcccgct accccaacgg catcagccac ttcgcccagg agtgcttccc 5820
cgagggcctg agcatcgacc gcaccgtgcg cttcgagaac gacggcacca tgaccagcca 5880
ccacacctac gagctggacg gcacctgcgt ggtcagccgc atcaccgtga actgcgacgg 5940
cttccagccc gacggcccca tcatgcgcga ccagctggtg gacatcctgc ccaacgagac 6000
ccacatgttc ccccacggcc ccaacgccgt gcgccagctg gccttcatcg gcttcaccac 6060
cgccgacggc ggcctgatga tgggccactt cgacagcaag atgaccttca acggcagccg 6120
cgccatcaag atccccggcc cccacttcgt gaccatcatc accaagcaga tgagggacac 6180
cagcgacaag cgcgaccacg tgtgccagcg cgaggtgacc tacgcccaca gcgtgccccg 6240
catcaccagc gccatcggta gcgacgagga ttccggactc agatctcgac ccaagaaaaa 6300
gcggaaggtg gaggacccgt aagatccacc ggatctagat aactgatcat aatcagccat 6360

accacatttg tagaggtttt acttgcttta aaaaacctcc cacacctccc cctgaacctg 6420
aaacataaaa tgaatgcaat tgttgttgtt aacttgttta ttgcagctta taatggttac 6480
aaataaagca atagcatcac aaatttcaca aataaagcat ttttttcact gcattctagt 6540
tgtggtttgt ccaaactcat caatgtatct taacgcgagt taattaacac cgaaatcgta 6600
attcacggca tcattacaaa atattttgac gttttggacc tcgtccctaa tgacaccata 6660
acggtggcct tgaagtatat ttaaccctag aaagatagtc tgcgtaaaat tgacgcatgc 6720
attcttgaaa tattgctctc tctttctaaa tagcgcgaat ccgtcgctgt gcatttagga 6780
catctcagtc gccgcttgga gctcccgtga ggcgtgcttg tcaatgcggt aagtgtcact 6840
gattttgaac tataacgacc gcgtgagtca aaatgacgca tgattatctt ttacgtgact 6900
tttaagattt aactcatacg ataattatat tgttatttca tgttctactt acgtgataac 6960
ttattatata tatattttct tgttatagat atcgtgacta atatataata aaatgggtag 7020
ttctttagac gatgagcata tcctctctgc tcttctgcaa agcgatgacg agcttgttgg 7080
tgaggattct gacagtgaaa tatcagatca cgtaagtgaa gatgacgtcc aggaaatctg 7140
gccggccgca accattgtgg gaaccgtgcg atcaaacaaa cgcgagatac cggaagtact 7200
gaaaaacagt cgctccaggc cagtgggaac atcgatgttt tgttttgacg gaccccttac 7260
tctcgtctca tataaaccga agccagctaa gatggtatac ttattatcat cttgtgatga 7320
ggatgcttct atcaacgaaa gtaccggtaa accgcaaatg gttatgtatt ataatcaaac 7380
taaaggcgga gtggacacgc tagaccaaat gtgttctgtg atgacctgca gtaggaagac 7440
gaataggtgg cctatggcat tattgtacgg aatgataaac attgcctgca taaattcttt 7500
tattatatac agccataatg tcagtagcaa gggagaaaag gtccaaagtc gcaaaaaatt 7560
tatgagaaac ctttacatga gcctgacgtc atcgtttatg cgtaagcgtt tagaagctcc 7620
tactttgaag agatatttgc gcgataatat ctctaatatt ttgccaaatg aagtgcctgg 7680
tacatcagat gacagtactg aagagccagt aatgaaaaaa cgtacttact gtacttactg 7740
cccctctaaa ataaggcgaa aggcaaatgc atcgtgcaaa aaatgcaaaa aagttatttg 7800
tcgagagcat aatattgata tgtgccaaag ttgtttctga ctgactaata agtataattt 7860
gtttctatta tgtataagtt aagctaatta cttattttat aatacaacat gactgttttt 7920
aaagtacaaa ataagtttat ttttgtaaaa gagagaatgt ttaaaagttt tgttacttta 7980
tagaagaaat tttgagtttt tgtttttttt taataaataa ataaacataa ataaattgtt 8040
tgttgaattt attattagta tgtaagtgta aatataataa aacttaatat ctattcaaat 8100
taataaataa acctcgatat acagaccgat aaaacacatg cgtcaatttt acgcatgatt 8160
atctttaacg tacgtcacaa tatgattatc tttctagggt taaataatag tttctaattt 8220
ttttattatt cagcctgctg tcgtgaatac cgtatatctc aacgctgtct gtgagattgt 8280
cgtattctag ccttttttagt ttttcgctca tcgacttgat attgtccgac acattttcgt 8340
cgatttgcgt tttgatcaaa gacttgagca gagacacgtt aatcaactgt tcaaattgat 8400
ccatattaac gatatcaacc cgatgcgtat atggtgcgta aaatatattt tttaaccctc 8460
ttatactttg cactctgcgt taatacgcgt tcgtgtacag acgtaatcat gttttctttt 8520
ttggataaaa ctcctactga gtttgacctc atattagacc ctcacaagtt gcaaaacgtg 8580
gcatttttta ccaatgaaga atttaaagtt attttaaaaa atttcatcac agatttaaag 8640
aagaaccaaa aattaaatta tttcaacagt ttaatcgacc agttaatcaa cgtgtacaca 8700
gacgcgtcgg caaaaaacac gcagcccgac gtgttggcta aaattattaa atcaacttgt 8760
gttatagtca cggatttgcc gtccaacgtg ttcctcaaaa agttgaagac caacaagttt 8820
acggacacta ttaattattt gattttgccc cacttcattt tgtgggatca caattttgtt 8880
atattttaaa caaagcttgg cactggccgt cgttttacaa cgtcgtgact gggaaaaccc 8940
tggcgttacc caacttaatc gccttgcagc acatcccccc ttcgccagct ggcgtaatag 9000
cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg 9060
cctgatgcgg tattttctcc ttacgcatct gtgcggtatt tcacaccgca tatggtgcac 9120
tctcagtaca atctgctctg atgccgcata gttaagccag ccccgacacc cgccaacacc 9180
cgctgacgcg ccctgacggg cttgtctgct cccggcatcc gcttacagac aagctgtgac 9240
cgtctccggg agctgcatgt gtcagaggtt ttcaccgtca tcaccgaaac gcgcgagacg 9300

aaagggcctc gtgatacgcc tatttttata ggttaatgtc atgataataa tggtttctta 9360
gacgtcaggt ggcacttttc ggggaaatgt gcgcggaacc cctatttgtt tatttttcta 9420
aatacattca aatatgtatc cgctcatgag acaataaccc tgataaatgc ttcaataata 9480
ttgaaaaagg aagagtatga gtattcaaca tttccgtgtc gcccttattc cctttttgc 9540
ggcattttgc cttcctgttt ttgctcaccc agaaacgctg gtgaaagtaa aagatgctga 9600
agatcagttg ggtgcacgag tgggttacat cgaactggat ctcaacagcg gtaagatcct 9660
tgagagtttt cgccccgaag aacgttttcc aatgatgagc acttttaaag ttctgctatg 9720
tggcgcggta ttatcccgta ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta 9780
ttctcagaat gacttggttg agtactcacc agtcacagaa aagcatctta cggatggcat 9840
gacagtaaga gaattatgca gtgctgccat aaccatgagt gataacactg cggccaactt 9900
acttctgaca acgatcggag gaccgaagga gctaaccgct tttttgcaca acatgggggа 9960
tcatgtaact cgccttgatc gttgggaacc ggagctgaat gaagccatac caaacgacga 10020
gcgtgacacc acgatgcctg tagcaatggc aacaacgttg cgcaaactat taactggcga 10080
actacttact ctagcttccc ggcaacaatt aatagactgg atggaggcgg ataaagttgc 10140
aggaccactt ctgcgctcgg cccttccggc tggctggttt attgctgata aatctggagc 10200
cggtgagcgt gggtctcgcg gtatcattgc agcactgggg ccagatggta agccctcccg 10260
tatcgtagtt atctacacga cggggagtca ggcaactatg gatgaacgaa atagacagat 10320
cgctgagata ggtgcctcac tgattaagca ttggtaactg tcagaccaag tttactcata 10380
tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg tgaagatcct 10440
ttttgataat ctcatgacca aaatcccttа acgtgagttt tcgttccact gagcgtcaga 10500
ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg taatctgctg 10560
cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc aagagctacc 10620
aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata ctgtccttct 10680
agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta catacctcgc 10740
tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc ttaccgggtt 10800
ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg ggggttcgtg 10860
cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac agcgtgagca 10920
ttgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg taagcggcag 10980
ggtcggaaca ggagagcgca cgagggagct tccagggggа aacgcctggt atctttatag 11040
tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct cgtcaggggg 11100
gcggagccta tggaaaaacg ccagcaacgc ggcctttttа cggttcctgg ccttttgctg 11160
gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata accgtattac 11220
cgcctttgag tgagctgata ccgctcgccg cagccgaacg accgagcgca gcgagtcagt 11280
gagcgaggaa gcggaagagc gcccaatacg caaaccgcct ctccccgcgc gttggccgat 11340
tcattaatgc agctggcacg acaggtttcc cgactggaaa gcgggcagtg agcgcaacgc 11400
aattaatgtg agttagctca ctcattaggc accccaggct ttacacttta tgcttccggc 11460
tcgtatgttg tgtggaattg tgagcggata acaatttcac acaggaaaca gctatgacca 11520
tgattacgaa tttcgacctg caggcatgca agcttgcatg cctgcaggtc gacgctcgcg 11580
cgacttggtt tgccattctt tagcgcgcgt cgcgtcacac agcttggcca caatgtggtt 11640
tttgtcaaac gaagattcta tgacgtgttt aaagtttagg tcgagtaaag cgcaaatctt 11700
ttttaaccct agaaagatag tctgcgtaaa attgacgcat gcattcttga aatattgctc 11760
tctctttcta aatagcgcga atccgtcgct gtgcatttag gacatctcag tcgccgcttg 11820
gagctcccgt gaggcgtgct tgtcaatgcg gtaagtgtca ctgattttga actataacga 11880
ccgcgtgagt caaaatgacg catgattatc ttttacgtga cttttaagat ttaactcata 11940
cgataattat attgttattt catgttctac ttacgtgata acttattata tatatatttt 12000
cttgttatag atatcgtgac taatatataa taaaatgggt agttctttag acgatgagca 12060
tatcctctct gctcttctgc aaagcgatga cgagcttgtt ggtgaggatt ctgacagtga 12120
aatatcagat cacgtaagtg aagatgacgt ccagagcgat acagaagaag cgtttataga 12180
tgaggtacat gaagtgcagc caacgtcaag cggtagtgaa atattagacg aacaaaatgt 12240

tattgaacaa ccaggttctt cattggcttc taacagaatc ttgaccttgc cacagaggac 12300
tattagaggt aagaataaac attgttggtc aacttcaaag tccacgaggc gtagccgagt 12360
ctctgcactg aacattgtca gatcggcccg gcggagtgga cacgctagac caaatgtgtt 12420
ctgtgatgac ctgcagtagg aagacgaata ggtggcctat ggcattattg tacggaatga 12480
taaacattgc ctgcataaat tcttttatta tatacagcca taatgtcagt agcaagggag 12540
aaaaggtcca aagtcgcaaa aaatttatga gaaaccttta catgagcctg acgtcatcgt 12600
ttatgcgtaa gcgtttagaa gctcctactt tgaagagata tttgcgcgat aatatctcta 12660
atattttgcc aaatgaagtg cctggtacat cagatgacag tactgaagag ccagtaatga 12720
aaaaacgtac ttactgtact tactgccccc ctaaaataag gcgaaaggca aatgcatcgt 12780
gcaaaaaatg caaaaaagtt atttgtcgag agcataatat tgatatgtgc caaagttgtt 12840
tctgactgac taataagtat aatttgtttc tattatgtat aagttaagct aattacttat 12900
tttataatac aacatgactg tttttaaagt acaaaataag tttatttttg taaaagagag 12960
aatgtttaaa agttttgtta ctttatagaa gaaattttga gtttttgttt ttttttaata 13020
aataaataaa cataaataaa ttgtttgttg aatttattat tagtatgtaa gtgtaaatat 13080
aataaaactt aatatctatt caaattaata aataaacctc gatatacaga ccgataaaac 13140
acatgcgtca attttacgca tgattatctt taacgtacgt cacaatatga ttatctttct 13200
agggttaaaa tgaatgtaag cactttatta acgaaatctt tgggaatatt tcgctcatca 13260
gcattttatt tgagcaggag tccgagatgc cc 13292

<210> 53
<211> 14713
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự của cấu trúc pLA3014-Cctra-intron-Ubiquitin-reaperKR
<400> 53
cgcgccggac gcggcaagtc tgcgagctta tatttacgtg gatctccggt gtgtccatga 60
ttcggcatca tatcataaac gacgaattcc aataaaaact ttgcttgttg ataacacctg 120
atgttcagag atgcccgata aaatcacagc tgttctggtt cacagtcacc agaaataaaa 180
aatattggaa ttgagatgta cacaattaac gatatttata aatatcttcc gatagtctat 240
cgtccggtta atcaaaataa agtgcgacga attaacatat tttcaaaatt aagacgcttt 300
gatagatgta tttgtataga gatagaaatt aaggttaaaa taacataaat gccaaagttt 360
agagcactat tcaataattc tcttgatttc aaattgaaat aatacacaat ataacatttt 420
ctaacactac aaagtcacga tattcttcca ccaaccgata gtatcgcaca cttgccattc 480
gcctcatcac gcacacgccc gcttcacaat tcaaacgaac ggcattttat tttcacagga 540
tcccgggagt cgtgaatgtt ttacccaata tcgactttca ttgttaactg accaaaattg 600
taatctgttc tgttagttgt cgagtgcctg tgccgcgatc gctatgggca tatgttgcca 660
aactctaaac caaatactca ttctgatgtt ttaaatgatt tgccctccca tatgtccttc 720
cgagtgagag acacaaaaaa ttccaacaca ctattgcaat gaaaataaat ttcctttatt 780
agccagaagt cagatgctca aggggcttca tgatgtcccc ataatttttg gcagagggaa 840
aaagatctca gtggtatttg tgagccaggg cattggccac accagccacc accttctgat 900
aggcagcctg cacctgagga gtgaattctt tgccaaaatg atgagacagc acaacaacca 960
gcacgttgcc caggagctgt aggaaagaga agaaggcatg aacatggtta gcagaggggc 1020
ccggtttgga ctcagagtat tttatcctca tctcaaacag tgtatatcat tgtaaccata 1080
aagagaaagg caggatgatg accagggtgt agttgtttct accaataaga atatttccac 1140
gccagccaga atttatatgc agaaatattc taccttatca tttaattata acaattgttc 1200
tctaaaactg tgctgaagta caatataata taccctgatt gccttgaaaa aaaagtgatt 1260
agagaaagta cttacaatct gacaaataaa caaaagtgaa tttaaaaatt cgttacaaat 1320
gcaagctaaa gtttaacgaa aaagttacag aaaatgaaaa gaaaataaga ggagacaatg 1380

gttgtcaaca gagtagaaag tgaaagaaac aaaattatca tgagggtcca tggtgataca 1440
agggacatct tcccattcta aacaacaccc tgaaaacttt gccccctcca tataacatga 1500
attttacaat agcgaaaaag aaagaacaat caagggtccc caaactcacc ctgaagttct 1560
cagctctaga cgcgtttcac tacccaccgt actcgtcaat tccaagggca tcggtaaaca 1620
tctgctcaaa ctcgaagtcg gccatatcca gagcgccgta gggggcggag tcgtgggggg 1680
taaatcccgg acccggggaa tccccgtccc ccaacatgtc cagatcgaaa tcgtctagcg 1740
cgtcggcatg cgccatcgcc acgtcctcgc cgtctaagtg gagctcgtcc cccaggctga 1800
catcggtcgg gggggccgtc gacagtctgc gcgtgtgtcc cgcggggaga aaggacaggc 1860
gcggagccgc cagccccgcc tcttcggggg cgtcgtcgtc cgggagatcg agcaggccct 1920
cgatggtaga cccgtaattg tttttcgtac gcgcgcggct gtacgcggac ccactttcac 1980
atttaagttg tttttctaat ccgcatatga tcaattcaag gccgaataag aaggctggct 2040
ctgcaccttg gtgatcaaat aattcgatag cttgtcgtaa taatggcggc atactatcag 2100
tagtaggtgt ttccctttct tctttagcga cttgatgctc ttgatcttcc aatacgcaac 2160
ctaaagtaaa atgccccaca gcgctgagtg catataatgc attctctagt gaaaaacctt 2220
gttggcataa aaaggctaat tgattttcga gagtttcata ctgtttttct gtaggccgtg 2280
tacctaaatg tacttttgct ccatcgcgat gacttagtaa agcacatcta aaacttttag 2340
cgttattacg taaaaaatct tgccagcttt ccccttctaa agggcaaaag tgagtatggt 2400
gcctatctaa catctcaatg gctaaggcgt cgagcaaagc ccgcttattt tttacatgcc 2460
aatacaatgt aggctgctct acacctagct tctgggcgag tttacgggtt gttaaacctt 2520
cgattccgac ctcattaagc agctctaatg cgctgttaat cactttactt ttatctaatc 2580
tcaattccat ggtggcaacc tgcaaggcga atgaataaac aagattgtgg cgaacagtgt 2640
aatgcgaaga acccacctct gctccaattc ccaattccct attcagctcg agcggggatc 2700
cccgggtacc gagctcgaat tcgggggcgc ggaggctgga tcggtcccgg tgtcttctat 2760
ggaggtcaaa acagcgtgga tggcgtctcc aggcgatctg acggttcact aaacgagctc 2820
tgcttatata ggcctcccac cgtacacgcc tacctcgacc cgggtaccga gctcgacttt 2880
cacttttctc tatcactgat agggagtggt aaactcgact ttcacttttc tctatcactg 2940
atagggagtg gtaaactcga ctttcacttt tctctatcac tgatagggag tggtaaactc 3000
gactttcact tttctctatc actgataggg agtggtaaac tcgactttca cttttctcta 3060
tcactgatag ggagtggtaa actcgacttt cacttttctc tatcactgat agggagtggt 3120
aaactcgact ttcacttttc tctatcactg atagggagtg gtaaactcga aatgtcgact 3180
atgcggaccg agcgccggag tataaataga ggcgcttcgt ctacggagcg acaattcaat 3240
tcaaacaagc aaagtgaaca cgtcgctaag cgaaagctaa gcaaataaac aagcgcagct 3300
gaacaagcta aacaatctgc gctagccacc atggttgtta ttaaacgtag atttggtaat 3360
tttaaaagca tatttttttc tttgaaattc ataagttatc aattatcgat ggaaatgtat 3420
tctatggaga acgttttacc cgatgaatgg gtgcaaaaat tattttacct tcaaatctac 3480
aatcaacaca cgctaacttt tgtgacttga tcaactctca cctggaaaag caaccaacta 3540
caatcaacat tctatgggat aatcgacaag tgagtaaaat tatagccgga cctcttagta 3600
cagtgtattt aaaaggggaa taatattcta tcaataggaa taaaaataag gtcagcagcc 3660
atgacttttc catcattttg aatatacctt atttgtttcg ggattaattg ggggtcggaa 3720
atcctcttga attcagaaac gggaaccgga ggaaggtgcc ggtctttcag aaagctgtga 3780
aaaataccaa catttctgct gccaagagct caataagaag tttcaaaaat tgtcttggat 3840
gttgcagctg tggctgctaa gtaataagac atctattagt atctagattt gttagaccat 3900
ttaacatagt gttttaaacg atggggttaa tagatgaggg ttaagaagct agttatatta 3960
ctgttgctgt aacgccttca attgtcggtt acagagcaaa cattattgaa tgttaatgta 4020
aagagtttat ttgttttcta gtaaacatat agcgattggt tagtaatcac taatagaaat 4080
ttttcataag tatcaaaaaa gtaaacctct ttttcagtct atgtaataag taaaccaagg 4140
aaagggaaaa tatctacaat caacaagcca ttgttgcagc aacaaagcaa ctgaaactac 4200
aatcaacatt caataaactt gggtaatttg gaatttaatt ctctgggaca cctgtggatt 4260
acaacaatca actcgaaact tattatacaa tgtaaataaa aattgatatg catacatgaa 4320

gatcaagtga aattccattt agaatcaatt tttttcgaat attaagtttc ttgctttaat 4380
ttatctgaaa gtaaatagac attccaaatt caagttaaca aattaataat gaattgacta 4440
gtgattttta agagaaaaag ataagattta aaaaaggaaa gcctttcttg ataaattttt 4500
gaaccacttt atgccgtttc aatcataaaa acttttaaga acacatgact ggtaaaatta 4560
atttaaaaca aatttaaatt ttcaacgtaa cattcaacaa aaatggtgaa aactatcacg 4620
gaaattgtta atattaatat gtcccaaaaa tagcctttgt atgtatatga tactaatcca 4680
tacatctatg gtatctatag gtgaaggctc aaagcctctg atgcagatct ttgtgaagac 4740
tttgaccgga aagaccatca ccctcgaggt agagccatcg gacaccattg agaatgtaaa 4800
ggccaagatt caggataagg agggaatccc cccagatcag cagcgtctga tcttcgctgg 4860
caagcaactg gaagacggac gcaccctgtc cgattacaac atccagaagg agtccaccct 4920
tcacttggtc cttcgtctcc gtggtggcgc cgtggccttc tacatcccgg atcaggccac 4980
cctgctgcgc gaggccgagc agcgcgagca gcagatcctg cgcctgcgcg agagccagtg 5040
gcgcttcctg gccaccgtgg tgctggagac cctgcgccag tacaccagct gccacccgcg 5100
caccggccgc cgcagcggcc gttaccgccg tccgagccag taacaccggt gatcataatc 5160
agccatacca catttgtaga ggttttactt gctttaaaaa acctcccaca cctccccctg 5220
aacctgaaac ataaaatgaa tgcaattgtt gttgttaact tgtttattgc agcttataat 5280
ggttacaaat aaagcaatag catcacaaat ttcacaaata aagcattttt ttcactgcat 5340
tctagttgtg gtttgtccaa actcatcaat gtatcttaac gcgagtttaa acgcgtccgc 5400
atacgtccgc tcacgttaag ttccgcagag agaagttgtt gaaaacataa acagaatcac 5460
ttgttgcact ctttgagaaa actggggcta ttgcggaaaa aaccaactaa aaatattgca 5520
ggttaggggt actacgctcg attggcgtac ggccaccact tttgcgactt cactgttaac 5580
cgctaccttc atagagactt ttacccgata aatgttatgt agtttgactt tctctgttaa 5640
tcacaagaaa aaatattgtg gaaattaaaa ttatctcaaa ctcaataagg aaataataat 5700
atatacacct atgttttata gaagtcaaca gtaaataagt tatttggaaa accattgtag 5760
ccgtttaaat aaatctcctt gagtgtgttt taaataacgg tcattaagta tattacttgg 5820
ccctctgaat ttcttgaatt acaccatttt ttgaaataaa tcaatccaaa agactacttt 5880
ttggtggcaa atgaactgca taaaaagtaa caaaagaaat atgtttttga aataacagta 5940
tagctgaagt gtattaaaaa ataccgtcat atgagcgacc cgctgttacc gcttcgctgc 6000
gaatgacaaa acgggctgag caagaaaatg gcgtagaagg cgacgaaaat tcgtttcact 6060
cgtgaagaaa acctcgataa ctgaggaata cagctgggat ttaaagagca tattcgaact 6120
acaagcagag atgtttcctg gtggaaacgg aaacgccgat ttgggctaca acaagcatgc 6180
ccacgtccat ggacttggac aacatggcca tgggcacaac cataatcaca atcagttcct 6240
gcgcagcccc caccaccccc cacacatttt tcactgccct ccgggggcgg tcagggcatg 6300
gtgacgccca tggtagccgc cggcctgccg ctcgccatgc agggtggcgt tggcatcgat 6360
tggcgcagct cgcccagcaa tggattaatt aactcgcgtt aagatacatt gatgagtttg 6420
gacaaaccac aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta 6480
ttgctttatt tgtaaccatt ataagctgca ataaacaagt taacaacaac aattgcattc 6540
attttatgtt tcaggttcag ggggaggtgt gggaggtttt ttaaagcaag taaaacctct 6600
acaaatgtgg tatggctgat tatgatcagt tatctagatc cggtggatct tacgggtcct 6660
ccaccttccg ctttttcttg ggtcgagatc tcaggaacag gtggtggcgg ccctcggtgc 6720
gctcgtactg ctccacgatg gtgtagtcct cgttgtggga ggtgatgtcc agcttggcgt 6780
ccacgtagta gtagccgggc agctgcacgg gcttcttggc catgtagatg gacttgaact 6840
ccaccaggta gtggccgccg tccttcagct tcagggcctt gtgggtctcg cccttcagca 6900
cgccgtcgcg ggggtacagg cgctcggtgg aggcctccca gcccatggtc ttcttctgca 6960
tcacggggcc gtcggagggg aagttcacgc cgatgaactt caccttgtag atgaagcagc 7020
cgtcctgcag ggaggagtcc tgggtcacgg tcgccacgcc gccgtcctcg aagttcatca 7080
cgcgctccca cttgaagccc tcgggggaagg acagcttctt gtagtcgggg atgtcggcgg 7140
ggtgcttcac gtacaccttg gagccgtact ggaactgggg ggacaggatg tcccaggcga 7200
agggcagggg gccgcccttg gtcaccttca gcttcacggt gttgtggccc tcgtaggggc 7260

ggccctcgcc ctcgccctcg atctcgaact cgtggccgtt cacggtgccc tccatgcgca 7320
ccttgaagcg catgaactcg gtgatgacgt tctcggagga ggccatggtg gcgaccggtt 7380
tgcgcttctt cttgggtggg gtgggatccc cgatctgcat tttggattat tctgcgggtc 7440
aaaatagaga tgtggaaaat tagtacgaaa tcaaatgagt ttcgttgaaa ttacaaaact 7500
attgaaacta acttcctggc tggggaataa aaatgggaaa cttatttatc gacgccaact 7560
ttgttgagaa acccctatta accctctacg aatattggaa caaaggaaag cgaagaaaca 7620
ggaacaaagg tagttgagaa acctgttccg ttgctcgtca tcgttttcat aatgcgagtg 7680
tgtgcatgta tatatacaca gctgaaacgc atgcatacac attattttgt gtgtatatgg 7740
tgacgtcaca actactaagc aataagaaat tttccagacg tggctttcgt ttcaagcaac 7800
ctactctatt tcagctaaaa ataagtggat ttcgttggta aaatacttca attaagcaaa 7860
gaactaacta actaataaca tgcacacaaa tgctcgagtg cgttcgtgat ttctcgaatt 7920
ttcaaatgcg tcactgcgaa tttcacaatt tgccaataaa tcttggcgaa aatcaacacg 7980
caagttttat ttatagattt gtttgcgttt tgatgccaat tgattgggaa aacaagatgc 8040
gtggctgcca atttcttatt ttgtaattac gtagagcgtt gaataaaaaa aaaatggccg 8100
aacaaagacc ttgaaatgca gtttttcttg aaattactca acgtcttgtt gctcttatta 8160
ctaattggta acagcgagtt aaaaacttac gtttcttgtg actttcgaga atgttctttt 8220
aattgtactt taatcaccaa caattaagta taaatttttc gctgattgcg ctttactttc 8280
tgcttgtact tgctgctgca aatgtcaatt ggttttgaag gcgaccgttc gcgaacgctg 8340
tttatatacc ttcggtgtcc gttgaaaatc actaaaaaat accgtagtgt tcgtaacact 8400
ttagtacaga gaaaaaaaat tgtgccgaaa tgtttttgat acgtacgaat accttgtatt 8460
aaaatttttt atgatttctg tgtatcactt ttttttgtg ttttcgttt aaactcacca 8520
cagtacaaaa caataaaata tttttaagac aatttcaaat tgagaccttt ctcgtactga 8580
cttgaccggc tgaatgagga tttctaccta gacgacctac ttcttaccat gacattgaat 8640
gcaatgccac ctttgatcta aacttacaaa agtccaaggc ttgttaggat tggtgtttat 8700
ttagtttgct tttgaaatag cactgtcttc tctaccggct ataatttga aactcgcagc 8760
ttgactggaa atttaaaaag taattctgtg taggtaaagg gtgttttaaa agtgtgatgt 8820
gttgagcgtt gcggcaacga ctgctattta tgtatatatt ttcaaaactt attgtttttg 8880
aagtgtttta aatggagcta tctggcaacg ctgcgcataa tcttacacaa gcttttctta 8940
atccattttt aagtgaaatt tgtttttact ctttcggcaa ataattgtta aatcgcttta 9000
agtgggctta catctggata agtaatgaaa acctgcatat tataatatta aaacatataa 9060
tccactgtgc tttccccgtg tgtggccata tacctaaaaa agtttatttt cgcagagccc 9120
cgcacggtca cactacggtt cggcgatttt cgattttgga cagtactgat tgcaagcgca 9180
ccgaaagcaa aatggagctg gagattttga acgcgaagaa cagcaagccg tacggcaagg 9240
tgaaggtgcc ctccggcgcc acgcccatcg gcgatctgcg cgccctaatt cacaagaccc 9300
tgaagcagac cccacacgcg aatcgccagt cgcttcgtct ggaactgaag ggcaaaagcc 9360
tgaaagatac ggacacattg gaatctctgt cgctgcgttc cggcgacaag atcggggtac 9420
catgcggccg ctcatttaaa tctggccggc ctggccgatc tgacaatgtt cagtgcagag 9480
actcggctac gcctcgtgga ctttgaagtt gaccaacaat gtttattctt acctctaata 9540
gtcctctgtg gcaaggtcaa gattctgtta gaagccaatg aagaacctgg ttgttcaata 9600
acattttgtt cgtctaatat ttcactaccg cttgacgttg gctgcacttc atgtacctca 9660
tctataaacg cttcttctgt atcgctctgg acgtcatctt cacttacgtg atctgatatt 9720
tcactgtcag aatcctcacc aacaagctcg tcatcgcttt gcagaagagc agagaggata 9780
tgctcatcgt ctaaagaact acccatttta ttatatatta gtcacgatat ctataacaag 9840
aaaatatata tataataagt tatcacgtaa gtagaacatg aaataacaat ataattatcg 9900
tatgagttaa atcttaaaag tcacgtaaaa gataatcatg cgtcatttg actcacgcgg 9960
tcgttatagt tcaaaatcag tgacacttac cgcattgaca agcacgcctc acgggagctc 10020
caagcggcga ctgagatgtc ctaaatgcac agcgacggat tcgcgctatt tagaaagaga 10080
gagcaatatt tcaagaatgc atgcgtcaat tttacgcaga ctatctttct agggttaaaa 10140
aagatttgcg ctttactcga cctaaacttt aaacacgtca tagaatcttc gtttgacaaa 10200

aaccacattg tggccaagct gtgtgacgcg acgcgcgcta aagaatggca aaccaagtcg 10260
cgcgagcgtc gacctgcagg catgcaagct tgcatgcctg caggtcgaaa ttcgtaatca 10320
tggtcatagc tgtttcctgt gtgaaattgt tatccgctca caattccaca caacatacga 10380
gccggaagca taaagtgtaa agcctggggt gcctaatgag tgagctaact cacattaatt 10440
gcgttgcgct cactgcccgc tttccagtcg ggaaacctgt cgtgccagct gcattaatga 10500
atcggccaac gcgcggggag aggcggtttg cgtattgggc gctcttccgc ttcctcgctc 10560
actgactcgc tgcgctcggt cgttcggctg cggcgagcgg tatcagctca ctcaaaggcg 10620
gtaatacggt tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc 10680
cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg cgtttttcca taggctccgc 10740
cccctgacg agcatcacaa aaatcgacgc tcaagtcaga ggtggcgaaa cccgacagga 10800
ctataaagat accaggcgtt tccccctgga agctccctcg tgcgctctcc tgttccgacc 10860
ctgccgctta ccggatacct gtccgccttt ctcccttcgg gaagcgtggc gctttctcaa 10920
tgctcacgct gtaggtatct cagttcggtg taggtcgttc gctccaagct gggctgtgtg 10980
cacgaacccc ccgttcagcc cgaccgctgc gccttatccg gtaactatcg tcttgagtcc 11040
aacccggtaa gacacgactt atcgccactg gcagcagcca ctggtaacag gattagcaga 11100
gcgaggtatg taggcggtgc tacagagttc ttgaagtggt ggcctaacta cggctacact 11160
agaaggacag tatttggtat ctgcgctctg ctgaagccag ttaccttcgg aaaaagagtt 11220
ggtagctctt gatccggcaa acaaaccacc gctggtagcg gtggtttttt tgtttgcaag 11280
cagcagatta cgcgcagaaa aaaaggatct caagaagatc ctttgatctt ttctacgggg 11340
tctgacgctc agtggaacga aaactcacgt taagggattt tggtcatgag attatcaaaa 11400
aggatcttca cctagatcct tttaaattaa aaatgaagtt ttaaatcaat ctaaagtata 11460
tatgagtaaa cttggtctga cagttaccaa tgcttaatca gtgaggcacc tatctcagcg 11520
atctgtctat ttcgttcatc catagttgcc tgactccccg tcgtgtagat aactacgata 11580
cgggagggct taccatctgg ccccagtgct gcaatgatac cgcgagaccc acgctcaccg 11640
gctccagatt tatcagcaat aaaccagcca gccggaaggg ccgagcgcag aagtggtcct 11700
gcaactttat ccgcctccat ccagtctatt aattgttgcc gggaagctag agtaagtagt 11760
tcgccagtta atagtttgcg caacgttgtt gccattgcta caggcatcgt ggtgtcacgc 11820
tcgtcgtttg gtatggcttc attcagctcc ggttcccaac gatcaaggcg agttacatga 11880
tcccccatgt tgtgcaaaaa agcggttagc tccttcggtc ctccgatcgt tgtcagaagt 11940
aagttggccg cagtgttatc actcatggtt atggcagcac tgcataattc tcttactgtc 12000
atgccatccg taagatgctt ttctgtgact ggtgagtact caaccaagtc attctgagaa 12060
tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa tacgggataa taccgcgcca 12120
catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt cttcgggggcg aaaactctca 12180
aggatcttac cgctgttgag atccagttcg atgtaaccca ctcgtgcacc caactgatct 12240
tcagcatctt ttactttcac cagcgtttct gggtgagcaa aaacaggaag gcaaaatgcc 12300
gcaaaaaagg gaataagggc gacacggaaa tgttgaatac tcatactctt cctttttcaa 12360
tattattgaa gcatttatca gggttattgt ctcatgagcg gatacatatt tgaatgtatt 12420
tagaaaaata aacaaatagg ggttccgcgc acatttcccc gaaaagtgcc acctgacgtc 12480
taagaaacca ttattatcat gacattaacc tataaaaata ggcgtatcac gaggcccttt 12540
cgtctcgcgc gtttcggtga tgacggtgaa aacctctgac acatgcagct cccggagacg 12600
gtcacagctt gtctgtaagc ggatgccggg agcagacaag cccgtcaggg cgcgtcagcg 12660
ggtgttggcg ggtgtcgggg ctggcttaac tatgcggcat cagagcagat tgtactgaga 12720
gtgcaccata tatgcggtgt gaaataccgc acagatgcgt aaggagaaaa taccgcatca 12780
ggcgccattc gccattcagg ctgcgcaact gttgggaagg gcgatcggtg cgggcctctt 12840
cgctattacg ccagctggcg aaaggggat gtgctgcaag gcgattaagt tgggtaacgc 12900
cagggtttc ccagtcacga cgttgtaaaa cgacggccag tgccaagctt tgtttaaaat 12960
ataacaaaat tgtgatccca caaaatgaag tggggcaaaa tcaaataatt aatagtgtcc 13020
gtaaacttgt tggtcttcaa ctttttgagg aacacgttgg acggcaaatc cgtgactata 13080
acacaagttg atttaataat tttagccaac acgtcgggct gcgtgttttt tgccgacgcg 13140

tctgtgtaca cgttgattaa ctggtcgatt aaactgttga aataatttaa tttttggttc 13200
ttctttaaat ctgtgatgaa attttttaaa ataactttaa attcttcatt ggtaaaaaat 13260
gccacgtttt gcaacttgtg agggtctaat atgaggtcaa actcagtagg agttttatcc 13320
aaaaaagaaa acatgattac gtctgtacac gaacgcgtat taacgcagag tgcaaagtat 13380
aagagggtta aaaaatatat tttacgcacc atatacgcat cgggttgata tcgttaatat 13440
ggatcaattt gaacagttga ttaacgtgtc tctgctcaag tctttgatca aaacgcaaat 13500
cgacgaaaat gtgtcggaca atatcaagtc gatgagcgaa aaactaaaaa ggctagaata 13560
cgacaatctc acagacagcg ttgagatata cggtattcac gacagcaggc tgaataataa 13620
aaaaattaga aactattatt taaccctaga aagataatca tattgtgacg tacgttaaag 13680
ataatcatgc gtaaaattga cgcatgtgtt ttatcggtct gtatatcgag gtttatttat 13740
taatttgaat agatattaag ttttattata tttacactta catactaata ataaattcaa 13800
caaacaattt atttatgttt atttatttat taaaaaaaaa caaaaactca aaatttcttc 13860
tataaagtaa caaaactttt aaacattctc tcttttacaa aaataaactt attttgtact 13920
ttaaaaacag tcatgttgta ttataaaata agtaattagc ttaacttata cataatagaa 13980
acaaattata cttattagtc agtcagaaac aactttggca catatcaata ttatgctctc 14040
gacaaataac ttttttgcat tttttgcacg atgcatttgc ctttcgcctt attttagagg 14100
ggcagtaagt acagtaagta cgtttttttca ttactggctc ttcagtactg tcatctgatg 14160
taccaggcac ttcatttggc aaaatattag agatattatc gcgcaaatat ctcttcaaag 14220
taggagcttc taaacgctta cgcataaacg atgacgtcag gctcatgtaa aggtttctca 14280
taaatttttt gcgactttgg acctttttctc ccttgctact gacattatgg ctgtatataa 14340
taaaagaatt tatgcaggca atgtttatca ttccgtacaa taatgccata ggccacctat 14400
tcgtcttcct actgcaggtc atcacagaac acatttggtc tagcgtgtcc actccgcctt 14460
tagtttgatt ataatacata accatttgcg gtttaccggt actttcgttg atagaagcat 14520
cctcatcaca agatgataat aagtatacca tcttagctgg cttcggttta tatgagacga 14580
gagtaagggg tccgtcaaaa caaaacatcg atgttcccac tggcctggag cgactgtttt 14640
tcagtacttc cggtatctcg cgtttgtttg atcgcacggt tcccacaatg gttgcggcca 14700
gcccgggcta tgg 14713

<210> 54
<211> 15848
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự của cấu trúc pLA3166-Cctra intron-Ubiquitin-reaperKR
<400> 54
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt taccttttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcgggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840

ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gaactaccca ccgtactcgt caattccaag ggcatcggta 960
aacatctgct caaactcgaa gtcggccata tccagagcgc cgtagggggc ggagtcgtgg 1020
ggggtaaatc ccggacccgg ggaatccccg tcccccaaca tgtccagatc gaaatcgtct 1080
agcgcgtcgg catgcgccat cgccacgtcc tcgccgtcta agtggagctc gtcccccagg 1140
ctgacatcgg tcgggggggc cgtcgacagt ctgcgcgtgt gtcccgcggg gagaaaggac 1200
aggcgcggag ccgccagccc cgcctcttcg ggggcgtcgt cgtccgggag atcgagcagg 1260
ccctcgatgg tagacccgta attgtttttc gtacgcgcgc ggctgtacgc ggggcccgag 1320
cccgactcgc atttcagttg cttttccaat ccgcagataa tcagctccaa gccgaacagg 1380
aatgccggct cggctccttg atgatcgaac agctcgattg cctgacgcag cagtgggggc 1440
atcgaatcgg ttgttggggt ctcgcgctcc tcttttgcga cttgatgctc ttggtcctcc 1500
agcacgcagc ccagggtaaa gtgaccgacg gcgctcagag cgtagagagc attttccagg 1560
ctgaagcctt gctggcacag gaacgcgagc tggttctcca gtgtctcgta ttgcttttcg 1620
gtcgggcgcg tgccgagatg gactttggca ccgtctcggt gggacagcag agcgcagcgg 1680
aacgacttgg cgttattgcg gaggaagtcc tggaaatggg atagatattg gtgttattgt 1740
tcatgtggca tataaaggac aagcaacaaa aaacgaacat aacatgagag atggttctga 1800
atcagaactt ctgaatatta tcctcccaaa agggttaaag tttttattaa gcatattacg 1860
ttttatacca cttccttatg taaaatttcc ttcgtagttt aatatcatgt gaaatcatat 1920
ataatttcta tcgaacgttt gttcaaattg aatgatgtca ttttttgaat aattggttat 1980
aattttataa catctcccga cttcgacatg tggttggtac taatgattgc gaaatcgccc 2040
tccgagaatg agaacaaccg aggtccaccg tctggtcgag attaaaacac ttgaggagtg 2100
ctttggtgac tcgatcaata ggtacagggc tcgttgccaa caatctggcc agctggacat 2160
ccgggacctc gttccccccc ggggtatcaa aatttttgta gtgtaaatag tagtacactc 2220
ttaaaaataa tgaaaattac tgcggacgta attcacatta tgattgaatg acactatcat 2280
tgacatttcc cgaatcagac accatcgtat ttaaaatgtg acacaaattc acctcatttg 2340
gctcgcttct tttatgtgca tccaaaagac gtaaaatcgc atgatttttt cggagtgtgt 2400
agtaagattg tcaaatttta attttaaata accagagccc ataaagcaaa gcaacactag 2460
gaaaaaaccc acaaactcaa cctgtccaaa aaaaaatata acaatcaaag ttgagggaat 2520
cggggtcaaa cgtcatgtaa aaatattttt tgtaaaaacc aaaccaggaa taaatatgaa 2580
tttaatcgga aaaaattgca aaatcgcata atttaatcct ccaactgtac tttatccagc 2640
ctgttgcaga aatgatgttt aaaggttcta atctgtaatt gttattagcc ttcaatactg 2700
atgtagtatt tatttcttat tgaaacattg agagctttat tttccaaagt tgtcattttc 2760
tcattcgtat atcgtaatat gtatattcgt aaatggcaag cacaatgata cttagggtag 2820
tcaaggatat ttcaattacg aaaagatcct gaaacgaccg ggaatcgaac ccttcagcat 2880
ggttttgctt tgtagctgct gaatctaacc actaggctga tgaagatccc attttagggt 2940
tgcaagttct caaagagcaa gaatgccaaa atagtgtcaa aagaagccct atttgacgat 3000
atacctttta gtctctacgt taatttgcta tgataattta tcatcaatta attggcaaag 3060
cctgatgcac gaaaagatct tcttctaaaa tttcagttgt tcttttcaac acattatgta 3120
atcataaaat ttaattaata aacctttttt ttttgtaact atccacagtt gatcaggcat 3180
aattttcttg gaaagtaaag tccatattta ggttgatgtt gaataaaaaa actttcaatt 3240
cactcttctg tttcacttca gaacttacgt aatacgacat tatgcatggt gcacacggaa 3300
caggataaga cgttcacaag ggatcaacat cacatcggat cgtaatcact ggatctggaa 3360
cacatatgac gccacaagac agcacatttt acacgatcac cagacgtgaa caaggaactg 3420
gatccacaag acgtcacagg aagacggcac atttccaacg gcttcgatgg aacttttctc 3480
gagtcttttt ccaccaatca taaacaccga cctgccagga ctcgccttcc aacgggcaaa 3540
aatgcgtgtg gtggcggtcg agcatctcga tggccagggc atccagcagc gcccgcttat 3600
tcttcacgtg ccagtagagg gtgggctgct ccacgcccag cttctgcgcc aacttgcggg 3660
tcgtcagtcc ctcaatgcca acttcgttca acagctccaa cgcggagttg atgactttgg 3720
acttatccag gcggctgccc atggtggttt ctaaaggtgt tataaatcaa attagttttg 3780

ttttttcttg aaaactttgc gtttcctttg atcaacttac cgccagggta ccgcagattg 3840
tttagcttgt tcagctgcgc ttgtttattt gcttagcttt cgcttagcga cgtgttcact 3900
ttgcttgttt gaattgaatt gtcgctccgt agacgaagcg cctctattta tactccggcg 3960
ctcgttttcg agtttaccac tccctatcag tgatagagaa aagtgaaagt cgagtttacc 4020
actccctatc agtgatagag aaaagtgaaa gtcgagttta ccactcccta tcagtgatag 4080
agaaaagtga aagtcgagtt taccactccc tatcagtgat agagaaaagt gaaagtcgag 4140
tttaccactc cctatcagtg atagagaaaa gtgaaagtcg agtttaccac tccctatcag 4200
tgatagagaa aagtgaaagt cgagtttacc actccctatc agtgatagag aaaagtgaaa 4260
gtcgaaacct ggcgcgcccc ggccatcgag aaagagagag agaagagaag agagagaaca 4320
ttcgagaaag agagagagaa gagaagagag agaacatact ccctatcagt gatagagaag 4380
tccctatcag tgatagagat gtccctatca gtgatagaga gttccctatc agtgatagag 4440
acgtccctat cagtgataga gaagtcccta tcagtgatag agagatccct atcagtgata 4500
gagatttccc tatcagtgat agagaggtcc ctatcagtga tagagacttc cctatcagtg 4560
atagagaaat ccctatcagt gatagagaca tccctatcag tgatagagaa ctccctatca 4620
gtgatagaga cctccctatc agtgatagag atcgatgcgg ccgcatggta cccattgctt 4680
gtcatttatt aatttggatg atgtcatttg tttttaaaat tgaactggct ttacgagtag 4740
aattctacgc gtaaaacaca atcaagtatg agtcataatc tgatgtcatg ttttgtacac 4800
ggctcataac cgaactggct ttacgagtag aattctactt gtaatgcacg atcagtggat 4860
gatgtcattt gtttttcaaa tcgagatgat gtcatgtttt gcacacggct cataaactcg 4920
ctttacgagt agaattctac gtgtaacgca cgatcgattg atgagtcatt tgttttgcaa 4980
tatgatatca tacaatatga ctcatttgtt tttcaaaacc gaacttgatt tacgggtaga 5040
attctacttg taaagcacaa tcaaaaagat gatgtcattt gtttttcaaa actgaactcg 5100
ctttacgagt agaattctac gtgtaaaaca caatcaagaa atgatgtcat ttgttataaa 5160
aataaaagct gatgtcatgt tttgcacatg gctcataact aaactcgctt tacgggtaga 5220
attctacgcg taaaacatga ttgataatta aataattcat ttgcaagcta tacgttaaat 5280
caaacggacg ctcgaggttg cacaacacta ttatcgattt gcagttcggg acataaatgt 5340
ttaaatatat cgatgtcttt gtgatgcgcg cgacattttt gtaggttatt gataaaatga 5400
acggatacgt tgcccgacat tatcattaaa tccttggcgt agaatttgtc gggtccattg 5460
tccgtgtgcg ctagcatgcc cgtaacggac ctcgtacttt tggcttcaaa ggttttgcgc 5520
acagacaaaa tgtgccacac ttgcagctct gcatgtgtgc gcgttaccac aaatcccaac 5580
ggcgcagtgt acttgttgta tgcaaataaa tctcgataaa ggcgcggcgc gcgaatgcag 5640
ctgatcacgt acgctcctcg tgttccgttc aaggacggtg ttatcgacct cagattaatg 5700
tttatcggcc gactgttttc gtatccgctc accaaacgcg tttttgcatt aacattgtat 5760
gtcggcggat gttctatatc taatttgaat aaataaacga taaccgcgtt ggttttagag 5820
ggcataataa aagaaatatt gttatcgtgt tcgccattag ggcagtataa attgacgttc 5880
atgttggata ttgtttcagt tgcaagttga cactggcggc gacaagcaat tctaattggg 5940
gtaagttttc ccgttctttt ctgggttctt ccctttgct catccttgct gcactacctt 6000
caggtgcaag ttgagattca ggccaccatg ggagatccca ccccacccaa gaagaagcgc 6060
aaaccggtcg ccaccatgga cgaggatggt tcagagggcg gccccgccct gttccagagc 6120
gacatgacct tcaaaatctt catcgacggc gaggtgaacg gccagaagtt caccatcgtg 6180
gccgacggca gcagcaagtt ccccccacggc gacttcaacg tgcacgccgt gtgcgagacc 6240
ggcaagctgc ccatgagctg gaagcccatc tgccacctga tccagtacgg cgagcccttc 6300
ttcgccccgct accccaacgg catcagccac ttcgcccagg agtgcttccc cgagggcctg 6360
agcatcgacc gcaccgtgcg cttcgagaac gacggcacca tgaccagcca ccacacctac 6420
gagctggacg gcacctgcgt ggtcagccgc atcaccgtga actgcgacgg cttccagccc 6480
gacggccccca tcatgcgcga ccagctggtg gacatcctgc ccaacgagac ccacatgttc 6540
ccccacggcc ccaacgccgt gcgccagctg gccttcatcg gcttcaccac cgccgacggc 6600
ggcctgatga tgggccactt cgacagcaag atgaccttca acggcagccg cgccatcaag 6660
atccccggcc cccacttcgt gaccatcatc accaagcaga tgagggacac cagcgacaag 6720

cgcgaccacg tgtgccagcg cgaggtgacc tacgcccaca gcgtgccccg catcaccagc 6780
gccatcggta gcgacgagga ttccggactc agatctcgac ccaagaaaaa gcggaaggtg 6840
gaggacccgt aagatccacc ggatctagat aactgatcat aatcagccat accacatttg 6900
tagaggtttt acttgcttta aaaaacctcc cacacctccc cctgaacctg aaacataaaa 6960
tgaatgcaat tgttgttgtt aacttgttta ttgcagctta taatggttac aaataaagca 7020
atagcatcac aaatttcaca aataaagcat ttttttcact gcattctagt tgtggtttgt 7080
ccaaactcat caatgtatct taacgcgagt taattaatcc attgctgggc gagctgcgcc 7140
aatcgatgcc aacgccaccc tgcatggcga gcggcaggcc ggcggctacc atgggcgtca 7200
ccatgccctg accgcccccg gagggcagtg aaaaatgtgt gggggggtggt gggggctgcg 7260
caggaactga ttgtgattat ggttgtgccc atggccatgt tgtccaagtc catggacgtg 7320
ggcatgcttg ttgtagccca aatcggcgtt tccgtttcca ccaggaaaca tctctgcttg 7380
tagttcgaat atgctcttta aatcccagct gtattcctca gttatcgagg ttttcttcac 7440
gagtgaaacg aattttcgtc gccttctacg ccatttttctt gctcagcccg ttttgtcatt 7500
cgcagcgaag cggtaacagc gggtcgctca tatgacggta ttttttaata cacttcagct 7560
atactgttat ttcaaaaaca tatttctttt gttacttttt atgcagttca tttgccacca 7620
aaaagtagtc ttttggattg atttatttca aaaaatggtg taattcaaga aattcagagg 7680
gccaagtaat atacttaatg accgttattt aaaacacact caaggagatt tatttaaacg 7740
gctacaatgg ttttccaaat aacttattta ctgttgactt ctataaaaca taggtgtata 7800
tattattatt tccttattga gtttgagata attttaattt ccacaatatt ttttcttgtg 7860
attaacagag aaagtcaaac tacataacat ttatcgggta aaagtctcta tgaaggtagc 7920
ggttaacagt gaagtcgcaa aagtggtggc cgtacgccaa tcgagcgtag taccccctaac 7980
ctgcaatatt tttagttggt tttttccgca atagccccag ttttctcaaa gagtgcaaca 8040
agtgattctg tttatgtttt caacaacttc tctctgcgga acttaacgtg agcggacgta 8100
tgcggacgcg tttaaactcg cgttaagata cattgatgag tttggacaaa ccacaactag 8160
aatgcagtga aaaaaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac 8220
cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt 8280
tcaggggggag gtgtgggagg ttttttaaag caagtaaaac ctctacaaat gtggtatggc 8340
tgattatgat caccggtgtt actggctcgg acggcggtaa cggccgctgc ggcggccggt 8400
gcgcgggtgg cagctggtgt actggcgcag ggtctccagc accacggtgg ccaggaagcg 8460
ccactggctc tcgcgcaggc gcaggatctg ctgctcgcgc tgctcggcct cgcgcagcag 8520
ggtggcctga tccgggatgt agaaggccac ggcgccacca cggagacgaa ggaccaagtg 8580
aagggtggac tccttctgga tgttgtaatc ggacagggtg cgtccgtctt ccagttgctt 8640
acctatagat accatagatg tatggattag tatcatatac atacaaaggc tatttttggg 8700
acatattaat attaacaatt tccgtgatag ttttcaccat ttttgttgaa tgttacgttg 8760
aaaatttaaa tttgttttaa attaatttta ccagtcatgt gttcttaaaa gtttttatga 8820
ttgaaacggc ataaagtggt tcaaaaattt atcaagaaag gctttccttt tttaaatctt 8880
atctttttct cttaaaaatc actagtcaat tcattattaa tttgttaact tgaatttgga 8940
atgtctattt actttcagat aaattaaagc aagaaactta atattcgaaa aaaattgatt 9000
ctaaatggaa tttcacttga tcttcatgta tgcatatcaa tttttattta cattgtataa 9060
taagtttcga gttgattgtt gtaatccaca ggtgtcccag agaattaaat tccaaattac 9120
ccaagtttat tgaatgttga ttgtagtttc agttgctttg ttgctgcaac aatggcttgt 9180
tgattgtaga tattttccct ttccttggtt tacttattac atagactgaa aaagaggttt 9240
actttttga tacttatgaa aaatttctat tagtgattac taaccaatcg ctatatgttt 9300
actagaaaac aaataaactc tttacattaa cattcaataa tgtttgctct gtaaccgaca 9360
attgaaggcg ttacagcaac agtaatataa ctagcttctt aaccctcatc tattaacccc 9420
atcgtttaaa acactatgtt aaatggtcta acaaatctag atactaatag atgtcttatt 9480
acttagcagc cacagctgca acatccaaga caatttttga aacttcttat tgagctcttg 9540
gcagcagaaa tgttggtatt tttcacagct ttctgaaaga ccggcacctt cctccggttc 9600
ccgtttctga attcaagagg atttccgacc cccaattaat cccgaaacaa ataaggtata 9660

ttcaaaatga tggaaaagtc atggctgctg accttatttt tattcctatt gatagaatat 9720
tattcccctt ttaaatacac tgtactaaga ggtccggcta taattttact cacttgtcga 9780
ttatcccata gaatgttgat tgtagttggt tgcttttcca ggtgagagtt gatcaagtca 9840
caaaagttag cgtgtgttga ttgtagattt gaaggtaaaa taatttttgc acccattcat 9900
cgggtaaaac gttctccata gaatacattt ccatcgataa ttgataactt atgaatttca 9960
aagaaaaaaa tatgctttta aaattaccag cgaagatcag acgctgctga tctgggggga 10020
ttccctcctt atcctgaatc ttggccttta cattctcaat ggtgtccgat ggctctacct 10080
cgagggtgat ggtctttccg gtcaaagtct tcacaaagat ctgcattttg gattgctagc 10140
gcagattgtt tagcttgttc agctgcgctt gtttatttgc ttagctttcg cttagcgacg 10200
tgttcacttt gcttgtttga attgaattgt cgctccgtag acgaagcgcc tctatttata 10260
ctccggcgct cggtccgcat agtcgacatt tcgagtttac cactccctat cagtgataga 10320
gaaaagtgaa agtcgagttt accactccct atcagtgata gagaaaagtg aaagtcgagt 10380
ttaccactcc ctatcagtga tagagaaaag tgaaagtcga gtttaccact ccctatcagt 10440
gatagagaaa agtgaaagtc gagtttacca ctccctatca gtgatagaga aaagtgaaag 10500
tcgagtttac cactccctat cagtgataga gaaaagtgaa agtcgagttt accactccct 10560
atcagtgata gagaaaagtg aaagtcgagc tcggtacccg ggtcgaggta ggcgtgtacg 10620
gtgggaggaa atctggccgg ccgcaaccat tgtgggaacc gtgcgatcaa acaaacgcga 10680
gataccggaa gtactgaaaa acagtcgctc caggccagtg ggaacatcga tgttttgttt 10740
tgacggaccc cttactctcg tctcatataa accgaagcca gctaagatgg tatacttatt 10800
atcatcttgt gatgaggatg cttctatcaa cgaaagtacc ggtaaaccgc aaatggttat 10860
gtattataat caaactaaag gcggagtgga cacgctagac caaatgtgtt ctgtgatgac 10920
ctgcagtagg aagacgaata ggtggcctat ggcattattg tacggaatga taaacattgc 10980
ctgcataaat tcttttatta tatacagcca taatgtcagt agcaagggag aaaaggtcca 11040
aagtcgcaaa aaatttatga gaaaccttta catgagcctg acgtcatcgt ttatgcgtaa 11100
gcgtttagaa gctcctactt tgaagagata tttgcgcgat aatatctcta atattttgcc 11160
aaatgaagtg cctggtacat cagatgacag tactgaagag ccagtaatga aaaaacgtac 11220
ttactgtact tactgccccc ctaaaataag gcgaaaggca aatgcatcgt gcaaaaaatg 11280
caaaaaagtt atttgtcgag agcataatat tgatatgtgc caaagttgtt tctgactgac 11340
taataagtat aatttgtttc tattatgtat aagttaagct aattacttat tttataatac 11400
aacatgactg tttttaaagt acaaaataag tttatttttg taaaagagag aatgtttaaa 11460
agttttgtta ctttatagaa gaaattttga gtttttgttt ttttttaata aataaataaa 11520
cataaataaa ttgtttgttg aatttattat tagtatgtaa gtgtaaatat aataaaactt 11580
aatatctatt caaattaata aataaacctc gatatacaga ccgataaaac acatgcgtca 11640
attttacgca tgattatctt taacgtacgt cacaatatga ttatctttct agggttaaat 11700
aatagtttct aatttttta ttattcagcc tgctgtcgtg aataccgtat atctcaacgc 11760
tgtctgtgag attgtcgtat tctagccttt ttagttttc gctcatcgac ttgatattgt 11820
ccgacacatt ttcgtcgatt tgcgttttga tcaaagactt gagcagagac acgttaatca 11880
actgttcaaa ttgatccata ttaacgatat caacccgatg cgtatatggt gcgtaaaata 11940
tatttttaa ccctcttata ctttgcactc tgcgttaata cgcgttcgtg tacagacgta 12000
atcatgtttt ctttttgga taaaactcct actgagtttg acctcatatt agaccctcac 12060
aagttgcaaa acgtggcatt ttttaccaat gaagaattta aagttatttt aaaaaatttc 12120
atcacagatt taaagaagaa ccaaaaatta aattatttca acagtttaat cgaccagtta 12180
atcaacgtgt acacagacgc gtcggcaaaa aacacgcagc ccgacgtgtt ggctaaaatt 12240
attaaatcaa cttgtgttat agtcacggat ttgccgtcca acgtgttcct caaaaagttg 12300
aagaccaaca agtttacgga cactattaat tatttgattt tgccccactt cattttgtgg 12360
gatcacaatt ttgttatatt ttaaacaaag cttggcactg gccgtcgttt tacaacgtcg 12420
tgactgggaa aaccctggcg ttacccaact taatcgcctt gcagcacatc cccctttcgc 12480
cagctggcgt aatagcgaag aggcccgcac cgatcgccct tcccaacagt tgcgcagcct 12540
gaatggcgaa tggcgcctga tgcggtattt tctccttacg catctgtgcg gtatttcaca 12600

ccgcatatgg tgcactctca gtacaatctg ctctgatgcc gcatagttaa gccagccccg 12660
acacccgcca acacccgctg acgcgccctg acgggcttgt ctgctcccgg catccgctta 12720
cagacaagct gtgaccgtct ccgggagctg catgtgtcag aggttttcac cgtcatcacc 12780
gaaacgcgcg agacgaaagg gcctcgtgat acgcctattt ttataggtta atgtcatgat 12840
aataatggtt tcttagacgt caggtggcac ttttcgggga aatgtgcgcg gaacccctat 12900
ttgtttattt ttctaaatac attcaaatat gtatccgctc atgagacaat aaccctgata 12960
aatgcttcaa taatattgaa aaaggaagag tatgagtatt caacatttcc gtgtcgccct 13020
tattcccttt tttgcggcat tttgccttcc tgtttttgct cacccagaaa cgctggtgaa 13080
agtaaaagat gctgaagatc agttgggtgc acgagtgggt tacatcgaac tggatctcaa 13140
cagcggtaag atccttgaga gttttcgccc cgaagaacgt tttccaatga tgagcacttt 13200
taaagttctg ctatgtggcg cggtattatc ccgtattgac gccgggcaag agcaactcgg 13260
tcgccgcata cactattctc agaatgactt ggttgagtac tcaccagtca cagaaaagca 13320
tcttacggat ggcatgacag taagagaatt atgcagtgct gccataacca tgagtgataa 13380
cactgcggcc aacttacttc tgacaacgat cggaggaccg aaggagctaa ccgcttttt 13440
gcacaacatg ggggatcatg taactcgcct tgatcgttgg gaaccggagc tgaatgaagc 13500
cataccaaac gacgagcgtg acaccacgat gcctgtagca atggcaacaa cgttgcgcaa 13560
actattaact ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga 13620
ggcggataaa gttgcaggac cacttctgcg ctcggccctt ccggctggct ggtttattgc 13680
tgataaatct ggagccggtg agcgtgggtc tcgcggtatc attgcagcac tggggccaga 13740
tggtaagccc tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga 13800
acgaaataga cagatcgctg agataggtgc ctcactgatt aagcattggt aactgtcaga 13860
ccaagtttac tcatatatac tttagattga tttaaaactt cattttaat ttaaaaggat 13920
ctaggtgaag atccttttg ataatctcat gaccaaaatc ccttaacgtg agttttcgtt 13980
ccactgagcg tcagaccccg tagaaaagat caaaggatct tcttgagatc ctttttttct 14040
gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta ccagcggtgg tttgtttgcc 14100
ggatcaagag ctaccaactc tttttccgaa ggtaactggc ttcagcagag cgcagatacc 14160
aaatactgtc cttctagtgt agccgtagtt aggccaccac ttcaagaact ctgtagcacc 14220
gcctacatac ctcgctctgc taatcctgtt accagtggct gctgccagtg gcgataagtc 14280
gtgtcttacc gggttggact caagacgata gttaccggat aaggcgcagc ggtcgggctg 14340
aacgggggt tcgtgcacac agcccagctt ggagcgaacg acctacaccg aactgagata 14400
cctacagcgt gagcattgag aaagcgccac gcttcccgaa gggagaaagg cggacaggta 14460
tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg gagcttccag ggggaaacgc 14520
ctggtatctt tatagtcctg tcgggtttcg ccacctctga cttgagcgtc gatttttgtg 14580
atgctcgtca gggggggcgga gcctatggaa aaacgccagc aacgcggcct ttttacggtt 14640
cctggccttt tgctggcctt ttgctcacat gttctttcct gcgttatccc ctgattctgt 14700
ggataaccgt attaccgcct ttgagtgagc tgataccgct cgccgcagcc gaacgaccga 14760
gcgcagcgag tcagtgagcg aggaagcgga agagcgccca atacgcaaac cgcctctccc 14820
cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg tttcccgact ggaaagcggg 14880
cagtgagcgc aacgcaatta atgtgagtta gctcactcat taggcacccc aggctttaca 14940
ctttatgctt ccggctcgta tgttgtgtgg aattgtgagc ggataacaat ttcacacagg 15000
aaacagctat gaccatgatt acgaatttcg acgctcgcgc gacttggttt gccattcttt 15060
agcgcgcgtc gcgtcacaca gcttggccac aatgtggttt ttgtcaaacg aagattctat 15120
gacgtgttta aagtttaggt cgagtaaagc gcaaatcttt tttaacccta gaaagatagt 15180
ctgcgtaaaa ttgacgcatg cattcttgaa atattgctct ctctttctaa atagcgcgaa 15240
tccgtcgctg tgcatttagg acatctcagt cgccgcttgg agctcccgtg aggcgtgctt 15300
gtcaatgcgg taagtgtcac tgattttgaa ctataacgac cgcgtgagtc aaaatgacgc 15360
atgattatct tttacgtgac ttttaagatt taactcatac gataattata ttgttatttc 15420
atgttctact tacgtgataa cttattatat atatattttc ttgttataga tatcgtgact 15480
aatatataat aaaatgggta gttctttaga cgatgagcat atcctctctg ctcttctgca 15540

aagcgatgac gagcttgttg gtgaggattc tgacagtgaa atatcagatc acgtaagtga 15600
agatgacgtc cagagcgata cagaagaagc gtttatagat gaggtacatg aagtgcagcc 15660
aacgtcaagc ggtagtgaaa tattagacga acaaaatgtt attgaacaac caggttcttc 15720
attggcttct aacagaatct tgaccttgcc acagaggact attagaggta agaataaaca 15780
ttgttggtca acttcaaagt ccacgaggcg tagccgagtc tctgcactga acattgtcag 15840
atcggccc 15848

<210> 55
<211> 17802
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự của pLA3376-Bztra intron-reaperKR và Bztra-intron-tTAV3.
<400> 55
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt tacctttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gacaccggtg ttagccgccg tactcatcga tgcccagggc 960
gtcggtgaac atctgctcga actcgaaatc ggccatatcc agggcgccgt agggggcgct 1020
atcgtgcggg gtgaatcccg gtcccgggct atcgccatcg cccagcatgt ccaggtcgaa 1080
gtcgtccagg gcatcggcgt gggccatcgc cacatcctcg ccatccaggt gcagctcatc 1140
gcccaggctc acgtcggtcg gcggggcggt cgacaggcgg cgggtgtgtc cggccggcag 1200
gaagctcagg cgcggggcgg ccaggcccgc ctcctccggg gcatcatcat ccggcagatc 1260
cagcaggccc tcgatggtgc tgccgtagtt gttcttggtg cgggcgcggc tgtaggcggg 1320
gcccgagccc gactcgcatt tcagttgctt ttccaatccg cagataatca gctccaagcc 1380
gaacaggaat gccggctcgg ctccttgatg atcgaacagc tcgattgcct gacgcagcag 1440
tgggggcatc gaatcggttg ttggggtctc gcgctcctct tttgcgactt gatgctcttg 1500
gtcctccagc acgcagccca gggtaaagtg accgacggcg ctcagagcgt agagagcatt 1560
ttccaggctg aagccttgct ggcacaggaa cgcgagctgg ttctccagtg tctcgtattg 1620
cttttcggtc gggcgcgtgc cgagatggac tttggcaccg tctcggtggg acagcagagc 1680
gcagcggaac gacttggcgt tattgcggag gaagtcctgc caggactcgc cttccaacgg 1740
gcaaaaatgc gtgtggtggc ggtcgagcat ctcgatggcc agggcatcca gcagcgcccg 1800
cttattcttc acgtgccagt agagggtggg ctgctccacg cccagcttct gcgccaactt 1860
gcgggtcgtc agtccctcaa tgccaacttc gttcaacagc tccaacgcgg agttgatgac 1920
tttggactta tccaggcggc tgacctatag ataccataga tgtatggatt agtatcatat 1980
acatacaaag gctatttttg ggacatatta atattaacaa tttccgtgat agttttcacc 2040
atttttgttg aatgttacgt tgaaaattta aatttgtttt aaattaattt taccagtcat 2100

gtgttcttaa aagtttttat gattgaaacg gcataaagtg gttcaaaaat ttatcaagaa 2160
aggctttcct tttttaaatc ttatcttttt ctcttaaaaa tcactagtca attcattatt 2220
aatttgttaa cttgaatttg gaatgtctat ttactttcag ataaattaaa gcaagaaact 2280
taatattcga aaaaaattga ttctaaatgg aatttcactt gatcttcatg tatgcatatc 2340
aatttttatt tacattgtat aataagtttc gagttgattg ttgtaatcca caggtgtccc 2400
agagaattaa attccaaatt acccaagttt attgaatgtt gattgtagtt tcagttgctt 2460
tgttgctgca acaatggctt gttgattgta gatattttcc ctttccttgg tttacttatt 2520
acatagactg aaaaagaggt ttacttttt gatacttatg aaaaatttct attagtgatt 2580
actaaccaat cgctatatgt ttactagaaa acaaataaac tctttacatt aacattcaat 2640
aatgtttgct ctgtaaccga caattgaagg cgttacagca acagtaatat aactagcttc 2700
ttaaccctca tctattaacc ccatcgttta aaacactatg ttaaatggtc taacaaatct 2760
agatactaat agatgtctta ttacttagca gccacagctg caacatccaa gacaattttt 2820
gaaacttctt attgagctct tggcagcaga aatgttggta tttttcacag ctttctgaaa 2880
gaccggcacc ttcctccggt tcccgtttct gaattcaaga ggatttccga ccccaattta 2940
atcccgaaac aaataaggta tattcaaaat gatggaaaag tcatggctgc tgaccttatt 3000
tttattccta ttgatagaat attattcccc ttttaaatac actgtactaa gaggtccggc 3060
tataatttta ctcacttgtc gattatccca tagaatgttg attgtagttg gttgcttttc 3120
caggtgagag ttgatcaagt cacaaaagtt agcgtgtgtt gattgtagat ttgaaggtaa 3180
aataattttt gcacccattc atcgggtaaa acgttctcca tagaatacat ttccatcgat 3240
aattgataac ttatgaattt caaagaaaaa aatatgcttt taaaattacc atggtggcta 3300
gcgcagattg tttagcttgt tcagctgcgc ttgtttattt gcttagcttt cgcttagcga 3360
cgtgttcact ttgcttgttt gaattgaatt gtcgctccgt agacgaagcg cctctattta 3420
tactccggcg ctcgttttcg agtttaccac tccctatcag tgatagagaa aagtgaaagt 3480
cgagtttacc actccctatc agtgatagag aaaagtgaaa gtcgagttta ccactcccta 3540
tcagtgatag agaaaagtga aagtcgagtt taccactccc tatcagtgat agagaaaagt 3600
gaaagtcgag tttaccactc cctatcagtg atagagaaaa gtgaaagtcg agtttaccac 3660
tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc agtgatagag 3720
aaaagtgaaa gtcgaaacct gcgcgccgtt taaactcgcg ttaagataca ttgatgagtt 3780
tggacaaacc acaactagaa tgcagtgaaa aaaatgcttt atttgtgaaa tttgtgatgc 3840
tattgcttta tttgtaacca ttataagctg caataaacaa gttaacaaca acaattgcat 3900
tcattttatg tttcaggttc agggggaggt gtgggaggtt ttttaaagca agtaaaacct 3960
ctacaaatgt ggtatggctg attatgatcg ctctagacac cggtgctacc cgccatactc 4020
atcgatgccc agcgcgtcgg tgaacatttg ctcgaactcg aagtcggcca tgtccagggc 4080
gccgtacggg gcgctatcgt ggggcgtgaa gcccggtccc gggctatctc catcgcccag 4140
catatccagg tcgaaatcgt ccagggcgtc ggcgtgggcc attgccacat cctctccatc 4200
caggtgcagc tcgtcgccca ggctcacatc ggtcggcggg gcggtgctca ggcggcgcgt 4260
gtgtccggcg ggcaggaagc tcaggcgggg ggcggccagg ccggcttcct ccggggcatc 4320
gtcatccggc aggtccagca gtcgctcgat ggtgctgcca tagttgttct tggtacgggc 4380
gcggctgtag gcgctgccgc tctcgcactt cagctgcttt tccaggccgc agatgatcag 4440
ctccaggccg aacaggaagg ccggctcggc gccctggtga tcgaacagct cgatggcctg 4500
gcgcagcagc ggcggcatgc tatcggtggt cggggtctcg cgctcctcct tggccacctg 4560
gtgctcctga tcctccagca cacagcccag ggtgaagtgg cccacggcgc tcagggcgta 4620
cagggcgttc tccaggctga agccctgctg gcacaggaag gccagctggt tctccagggt 4680
ctcgtactgc ttctcggtcg ggcgggtgcc caggtgcacc ttggcgccat cgcggtgcga 4740
cagcagggcg cagcggaagc tcttggcgtt gttgcgcagg aaatcctgcc agctctcgcc 4800
ctccagcggg cagaagtggg tgtggtggcg atccagcatt tcgatggcca gggcgtccag 4860
cagggcgcgc ttgttcttca cgtgccagta cagggtcggc tgttccacgc ccagcttctg 4920
ggccagcttg cgggtggtca ggccctcgat accaacttcg ttcagcagct ccagggcgct 4980
gttgatcacc ttgctcttgt ccaggcggct gacctgtgaa tacggttaat gtcactatta 5040

gtgatttata aaaataaatt tgatttatat atcaacaatt tttcatcgca gccttcagct 5100
ttttgttgaa taattataat gatatttttt acgattcaaa tcatttaatt gttactcaac 5160
gaaataagtt taattcaaat tttaaaacaa gattatatat taagattaga ataagaaaga 5220
actttgttag attatttaat taaaaagatt aaaatttaag tctccagtca ctatttaaag 5280
atcatctttc aaacgttaaa gtgaattcaa acgagacgtt caaatttcga ttaaacagta 5340
attaactcta aatttctatc acgaattaag ttattgaata tgaaggttta tatttattta 5400
catcatctaa taggtttgag ttgattgttg taatccgcat gtgccagaag atatcaattt 5460
ccaaattgtc cgagttcatg gaatgttgat tgttgtttgt gttgctttgt aattgttgca 5520
gggagtattt atggtttgtt gattgtagta taaggctgtt tctaaaggct agaaaataat 5580
tttatttatt tgaaaataag taaatataca taatattact aacaataggt cgtcctattt 5640
tttgatattc tgcacaaatt tttaaaacac aaagattgca atacttttag acactaatac 5700
tgcacactct gaaaaattat taaattattt ttaaaaactt accttaatac tttagagaaa 5760
aatattatac cgcacctttc tactttatac tcactttatt ataccagttg catgttgatt 5820
gtagttcttt gacaagaaaa tattccatat tgctccaaat tatcttggta agttgattgg 5880
tgcgtcattt gagcaagcta acaccttgtc tcatttaagt tcgcctcaag atctcatagc 5940
atttttaaat atcactatat ttagtaagta attagaatta ccatggtggt ttgctagccg 6000
ttctatcaga tgtgctccgg gaaacagaaa tgttcaacta agttctggcg gacgacgcaa 6060
caccttttata tactttgcca agcgcacagg tagaaaggac ctattttggg gattaaaaaa 6120
catctgcctg ttttattgcc atacccgcga aaattcgcga aatccgctac tttacctact 6180
ggggttcctg gaaaatgggc gaagaacggc aaagaactgg tactttccgt caataattgt 6240
ttagaagaga gagaacatac tccctatcag tgatagagaa gtccctatca gtgatagaga 6300
tgtccctatc agtgatagag agttccctat cagtgataga gacgtcccta tcagtgatag 6360
agaagtccct atcagtgata gagagatccc tatcagtgat agagatttcc ctatcagtga 6420
tagagaggtc cctatcagtg atagagactt ccctatcagt gatagagaaa tccctatcag 6480
tgatagagac atccctatca gtgatagaga actccctatc agtgatagag acctccctat 6540
cagtgataga gatcgatgcg gccgcatggt acccattgct tgtcatttat taatttggat 6600
gatgtcattt gtttttaaaa ttgaactggc tttacgagta gaattctacg cgtaaaacac 6660
aatcaagtat gagtcataat ctgatgtcat gttttgtaca cggctcataa ccgaactggc 6720
tttacgagta gaattctact tgtaatgcac gatcagtgga tgatgtcatt tgtttttcaa 6780
atcgagatga tgtcatgttt tgcacacggc tcataaactc gctttacgag tagaattcta 6840
cgtgtaacgc acgatcgatt gatgagtcat ttgttttgca atatgatatc atacaatatg 6900
actcatttgt ttttcaaaac cgaacttgat ttacgggtag aattctactt gtaaagcaca 6960
atcaaaaaga tgatgtcatt tgtttttcaa aactgaactc gctttacgag tagaattcta 7020
cgtgtaaaac acaatcaaga aatgatgtca tttgttataa aaataaaagc tgatgtcatg 7080
ttttgcacat ggctcataac taaactcgct ttacgggtag aattctacgc gtaaaacatg 7140
attgataatt aaataattca tttgcaagct atacgttaaa tcaaacggac gctcgaggtt 7200
gcacaacact attatcgatt tgcagttcgg gacataaatg tttaaatata tcgatgtctt 7260
tgtgatgcgc gcgacatttt tgtaggttat tgataaaatg aacggatacg ttgcccgaca 7320
ttatcattaa atccttggcg tagaatttgt cgggtccatt gtccgtgtgc gctagcatgc 7380
ccgtaacgga cctcgtactt ttggcttcaa aggttttgcg cacagacaaa atgtgccaca 7440
cttgcagctc tgcatgtgtg cgcgttacca caaatcccaa cggcgcagtg tacttgttgt 7500
atgcaaataa atctcgataa aggcgcggcg cgcgaatgca gctgatcacg tacgctcctc 7560
gtgttccgtt caaggacggt gttatcgacc tcagattaat gtttatcggc cgactgtttt 7620
cgtatccgct caccaaacgc gtttttgcat taacattgta tgtcggcgga tgttctatat 7680
ctaatttgaa taaataaacg ataaccgcgt tggttttaga gggcataata aaagaaatat 7740
tgttatcgtg ttcgccatta gggcagtata aattgacgtt catgttggat attgtttcag 7800
ttgcaagttg acactggcgg cgacaagcaa ttctaattgg ggtaagtttt cccgttcttt 7860
tctgggttct tccctttgc tcatccttgc tgcactacct tcaggtgcaa gttgagattc 7920
aggccaccat gggagatccc accccaccca agaagaagcg caaaccggtc gccaccatgg 7980

agagcgacga gagcggcctg cccgccatgg agatcgagtg ccgcatcacc ggcaccctga 8040
acggcgtgga gttcgagctg gtgggcggcg gagagggcac ccccgagcag ggccgcatga 8100
ccaacaagat gaagagcacc aaaggcgccc tgaccttcag cccctacctg ctgagccacg 8160
tgatgggcta cggcttctac cacttcggca cctaccccag cggctacgag aacccccttcc 8220
tgcacgccat caacaacggc ggctacacca acacccgcat cgagaagtac gaggacggcg 8280
gcgtgctgca cgtgagcttc agctaccgct acgaggccgg ccgcgtgatc ggcgacttca 8340
aggtgatggg caccggcttc cccgaggaca gcgtgatctt caccgacaag atcatccgca 8400
gcaacgccac cgtggagcac ctgcacccca tgggcgataa cgatctggat ggcagcttca 8460
cccgcacctt cagcctgcgc gacggcggct actacagctc cgtggtggac agccacatgc 8520
acttcaagag cgccatccac cccagcatcc tgcagaacgg gggccccatg ttcgccttcc 8580
gccgcgtgga ggaggatcac agcaacaccg agctgggcat cgtggagtac cagcacgcct 8640
tcaagacccc ggatgcagat gccggtgaag aaagatctcg acccaagaaa aagcggaagg 8700
tggaggaccc gtaagatcca ccggatctag ataactgatc ataatcagcc ataccacatt 8760
tgtagaggtt ttacttgctt taaaaaacct cccacacctc cccctgaacc tgaaacataa 8820
aatgaatgca attgttgttg ttaacttgtt tattgcagct tataatggtt acaaataaag 8880
caatagcatc acaaatttca caaataaagc attttttttca ctgcattcta gttgtggttt 8940
gtccaaactc atcaatgtat cttaacgcga gttatcgcgc tcgcgcgact gacggtcgta 9000
agcacccgcg tacgtgtcca ccccggtcac aacccccttgt gtcatgtcgg cgaccctacg 9060
cccccaactg agagaactca aaggttaccc cagttggggc actactcccg aaaaccgctt 9120
ctgacctggg aaaacgtgaa gccccggggc atccgctgag ggttgccgcc ggggcttcgg 9180
tgtgtccgtc agtacttaat taacaccgaa atcgtaattc acggcatcat tacaaaatat 9240
tttgacgttt tggacctcgt ccctaatgac accataacgg tggccttgaa gtatatttaa 9300
ccctagaaag atagtctgcg taaaattgac gcatgcattc ttgaaatatt gctctctctt 9360
tctaaatagc gcgaatccgt cgctgtgcat ttaggacatc tcagtcgccg cttggagctc 9420
ccgtgaggcg tgcttgtcaa tgcggtaagt gtcactgatt ttgaactata acgaccgcgt 9480
gagtcaaaat gacgcatgat tatcttttac gtgactttta agatttaact catacgataa 9540
ttatattgtt atttcatgtt ctacttacgt gataacttat tatatatata ttttcttgtt 9600
atagatatcg tgactaatat ataataaaat gggtagttct ttagacgatg agcatatcct 9660
ctctgctctt ctgcaaagcg atgacgagct tgttggtgag gattctgaca gtgaaatatc 9720
agatcacgta agtgaagatg acgtccagga aatctggccg gccgcaacca ttgtgggaac 9780
cgtgcgatca aacaaacgcg agataccgga agtactgaaa aacagtcgct ccaggccagt 9840
gggaacatcg atgttttgtt ttgacggacc ccttactctc gtctcatata aaccgaagcc 9900
agctaagatg gtatacttat tatcatcttg tgatgaggat gcttctatca acgaaagtac 9960
cggtaaaccg caaatggtta tgtattataa tcaaactaaa ggcggagtgg acacgctaga 10020
ccaaatgtgt tctgtgatga cctgcagtag gaagacgaat aggtggccta tggcattatt 10080
gtacggaatg ataaacattg cctgcataaa ttcttttatt atatacagcc ataatgtcag 10140
tagcaaggga gaaaaggtcc aaagtcgcaa aaaatttatg agaaaccttt acatgagcct 10200
gacgtcatcg tttatgcgta agcgtttaga agctcctact ttgaagagat atttgcgcga 10260
taatatctct aatattttgc caaatgaagt gcctggtaca tcagatgaca gtactgaaga 10320
gccagtaatg aaaaaacgta cttactgtac ttactgcccc tctaaaataa ggcgaaaggc 10380
aaatgcatcg tgcaaaaaat gcaaaaaagt tatttgtcga gagcataata ttgatatgtg 10440
ccaaagttgt ttctgactga ctaataagta taatttgttt ctattatgta taagttaagc 10500
taattactta ttttataata caacatgact gtttttaaag tacaaaataa gtttatttttt 10560
gtaaaagaga gaatgtttaa aagttttgtt actttataga agaaattttg agtttttgtt 10620
ttttttttaat aaataaataa acataaataa attgtttgtt gaatttatta ttagtatgta 10680
agtgtaaata taataaaact taatatctat tcaaattaat aaataaacct cgatatacag 10740
accgataaaa cacatgcgtc aattttacgc atgattatct ttaacgtacg tcacaatatg 10800
attatctttc tagggttaaa taatagtttc taatttttttt attattcagc ctgctgtcgt 10860
gaataccgta tatctcaacg ctgtctgtga gattgtcgta ttctagcctt tttagttttt 10920

cgctcatcga cttgatattg tccgacacat tttcgtcgat ttgcgttttg atcaaagact 10980
tgagcagaga cacgttaatc aactgttcaa attgatccat attaacgata tcaacccgat 11040
gcgtatatgg tgcgtaaaat atatttttta accctcttat actttgcact ctgcgttaat 11100
acgcgttcgt gtacagacgt aatcatgttt tctttttgg ataaaactcc tactgagttt 11160
gacctcatat tagaccctca caagttgcaa aacgtggcat tttttaccaa tgaagaattt 11220
aaagttattt taaaaaattt catcacagat ttaaagaaga accaaaaatt aaattatttc 11280
aacagtttaa tcgaccagtt aatcaacgtg tacacagacg cgtcggcaaa aaacacgcag 11340
cccgacgtgt tggctaaaat tattaaatca acttgtgtta tagtcacgga tttgccgtcc 11400
aacgtgttcc tcaaaaagtt gaagaccaac aagtttacgg acactattaa ttatttgatt 11460
ttgccccact tcattttgtg ggatcacaat tttgttatat tttaaacaaa gcttggcact 11520
ggccgtcgtt ttacaacgtc gtgactggga aaaccctggc gttacccaac ttaatcgcct 11580
tgcagcacat ccccctttcg ccagctggcg taatagcgaa gaggcccgca ccgatcgccc 11640
ttcccaacag ttgcgcagcc tgaatggcga atggcgcctg atgcggtatt ttctccttac 11700
gcatctgtgc ggtatttcac accgcatatg gtgcactctc agtacaatct gctctgatgc 11760
cgcatagtta agccagcccc gacacccgcc aacacccgct gacgcgccct gacgggcttg 11820
tctgctcccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca 11880
gaggttttca ccgtcatcac cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt 11940
tttataggtt aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg 12000
aaatgtgcgc ggaaccccta tttgtttatt tttctaaata cattcaaata tgtatccgct 12060
catgagacaa taaccctgat aaatgcttca ataatattga aaaaggaaga gtatgagtat 12120
tcaacatttc cgtgtcgccc ttattccctt ttttgcggca ttttgccttc ctgtttttgc 12180
tcacccagaa acgctggtga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg 12240
ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg 12300
ttttccaatg atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga 12360
cgccgggcaa gagcaactcg gtcgccgcat acactattct cagaatgact tggttgagta 12420
ctcaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc 12480
tgccataacc atgagtgata acactgcggc caacttactt ctgacaacga tcggaggacc 12540
gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg 12600
ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc 12660
aatggcaaca acgttgcgca aactattaac tggcgaacta cttactctag cttcccggca 12720
acaattaata gactggatgg aggcggataa agttgcagga ccacttctgc gctcggccct 12780
tccggctggc tggtttattg ctgataaatc tggagccggt gagcgtgggt ctcgcggtat 12840
cattgcagca ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg 12900
gagtcaggca actatggatg aacgaaatag acagatcgct gagataggtg cctcactgat 12960
taagcattgg taactgtcag accaagttta ctcatatata ctttagattg atttaaaact 13020
tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaaat 13080
cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc 13140
ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct 13200
accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga aggtaactgg 13260
cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt taggccacca 13320
cttcaagaac tctgtagcac cgcctacata cctcgctctg ctaatcctgt taccagtggc 13380
tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga 13440
taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct tggagcgaac 13500
gacctacacc gaactgagat acctacagcg tgagcattga gaaagcgcca cgcttcccga 13560
agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag 13620
ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg 13680
acttgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga aaaacgccag 13740
caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgctcaca tgttctttcc 13800
tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc 13860

tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aagagcgccc 13920
aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag 13980
gtttcccgac tggaaagcgg gcagtgagcg caacgcaatt aatgtgagtt agctcactca 14040
ttaggcaccc caggctttac actttatgct tccggctcgt atgttgtgtg gaattgtgag 14100
cggataacaa tttcacacag gaaacagcta tgaccatgat tacgaatttc gacctgcagg 14160
catgcaagct tgcatgcctg caggtcgacg ctcgcgcgac ttggtttgcc attctttagc 14220
gcgcgtcgcg tcacacagct tggccacaat gtggtttttg tcaaacgaag attctatgac 14280
gtgtttaaag tttaggtcga gtaaagcgca aatctttttt aaccctagaa agatagtctg 14340
cgtaaaattg acgcatgcat tcttgaaata ttgctctctc tttctaaata gcgcgaatcc 14400
gtcgctgtgc atttaggaca tctcagtcgc cgcttggagc tcccgtgagg cgtgcttgtc 14460
aatgcggtaa gtgtcactga ttttgaacta taacgaccgc gtgagtcaaa atgacgcatg 14520
attatctttt acgtgacttt taagatttaa ctcatacgat aattatattg ttatttcatg 14580
ttctacttac gtgataactt attatatata tattttcttg ttatagatat cgtgactaat 14640
atataataaa atgggtagtt ctttagacga tgagcatatc ctctctgctc ttctgcaaag 14700
cgatgacgag cttgttggtg aggattctga cagtgaaata tcagatcacg taagtgaaga 14760
tgacgtccag agcgatacag aagaagcgtt tatagatgag gtacatgaag tgcagccaac 14820
gtcaagcggt agtgaaatat tagacgaaca aaatgttatt gaacaaccag gttcttcatt 14880
ggcttctaac agaatcttga ccttgccaca gaggactatt agaggtaaga ataaacattg 14940
ttggtcaact tcaaagtcca cgaggcgtag ccgagtctct gcactgaaca ttgtcagatc 15000
ggcccggcgg agtggacacg ctagaccaaa tgtgttctgt gatgacctgc agtaggaaga 15060
cgaataggtg gcctatggca ttattgtacg gaatgataaa cattgcctgc ataaattctt 15120
ttattatata cagccataat gtcagtagca agggagaaaa ggtccaaagt cgcaaaaaat 15180
ttatgagaaa cctttacatg agcctgacgt catcgtttat gcgtaagcgt ttagaagctc 15240
ctactttgaa gagatatttg cgcgataata tctctaatat tttgccaaat gaagtgcctg 15300
gtacatcaga tgacagtact gaagagccag taatgaaaaa acgtacttac tgtacttact 15360
gcccctctaa aataaggcga aaggcaaatg catcgtgcaa aaaatgcaaa aaagttattt 15420
gtcgagagca taatattgat atgtgccaaa gttgtttctg actgactaat aagtataatt 15480
tgtttctatt atgtataagt taagctaatt acttatttta taatacaaca tgactgtttt 15540
taaagtacaa aataagttta tttttgtaaa agagagaatg tttaaaagtt ttgttacttt 15600
atagaagaaa ttttgagttt ttgttttttt ttaataaata aataaacata aataaattgt 15660
ttgttgaatt tattattagt atgtaagtgt aaatataata aaacttaata tctattcaaa 15720
ttaataaata aacctcgata tacagaccga taaaacacat gcgtcaattt tacgcatgat 15780
tatctttaac gtacgtcaca atatgattat ctttctaggg ttaaaatgaa tgtaagcact 15840
ttattaacga aatctttggg aatatttcgc tcatcagcat tttatttgag caggagtccg 15900
agatgcccgg ccgcgccggc catcgagaaa gagagagaga agagaagaga gagaacattc 15960
gagaaagaga gagagaagag aagagagaga acatactccc tatcagtgat agagaagtcc 16020
ctatcagtga tagagatgtc cctatcagtg atagagagtt ccctatcagt gatagagacg 16080
tccctatcag tgatagagaa gtccctatca gtgatagaga gatccctatc agtgatagag 16140
atttccctat cagtgataga gaggtcccta tcagtgatag agacttccct atcagtgata 16200
gagaaatccc tatcagtgat agagacatcc ctatcagtga tagagaactc cctatcagtg 16260
atagagacct ccctatcagt gatagagatc gatccgtcta cctgagcgat atataaacta 16320
atgcctgttg caattgttca gtcagtcacg agtttgttac cactgcgaca agctagcaac 16380
caccatggcg gtaattctaa ttacttacta aatatagtga tatttaaaaa tgctatgaga 16440
tcttgaggcg aacttaaatg agacaaggtg ttagcttgct caaatgacgc accaatcaac 16500
ttaccaagat aatttggagc aatatggaat attttcttgt caaagaacta caatcaacat 16560
gcaactggta taataaagtg agtataaagt agaaaggtgc ggtataatat ttttctctaa 16620
agtattaagg taagttttta aaaataattt aataattttt cagagtgtgc agtattagtg 16680
tctaaaagta ttgcaatctt tgtgttttaa aaatttgtgc agaatatcaa aaaataggac 16740
gacctattgt tagtaatatt atgtatattt acttattttc aaataaataa aattattttc 16800

tagcctttag aaacagcctt atactacaat caacaaacca taaatactcc ctgcaacaat 16860
tacaaagcaa cacaaacaac aatcaacatt ccatgaactc ggacaatttg gaaattgata 16920
tcttctggca catgcggatt acaacaatca actcaaacct attagatgat gtaaataaat 16980
ataaaccttc atattcaata acttaattcg tgatagaaat ttagagttaa ttactgttta 17040
atcgaaattt gaacgtctcg tttgaattca ctttaacgtt tgaaagatga tctttaaata 17100
gtgactggag acttaaattt taatcttttt aattaaataa tctaacaaag ttctttctta 17160
ttctaatctt aatatataat cttgttttaa aatttgaatt aaacttattt cgttgagtaa 17220
caattaaatg atttgaatcg taaaaaatat cattataatt attcaacaaa aagctgaagg 17280
ctgcgatgaa aaattgttga tatataaatc aaatttattt ttataaatca ctaatagtga 17340
cattaaccgt attcacaggt ggccttctac atcccggatc aggccaccct gctgcgcgag 17400
gccgagcagc gcgagcagca gatcctgcgc ctgcgcgaga gccagtggcg cttcctggcc 17460
accgtggtgc tggagaccct gcgccagtac accagctgcc acccgcgcac cggccgccgc 17520
agcggccgtt accgccgtcc gagccagtaa caccggtgat cataatcagc cataccacat 17580
ttgtagaggt tttacttgct ttaaaaaacc tcccacacct ccccctgaac ctgaaacata 17640
aaatgaatgc aattgttgtt gttaacttgt ttattgcagc ttataatggt tacaaataaa 17700
gcaatagcat cacaaatttc acaaataaag catttttttc actgcattct agttgtggtt 17760
tgtccaaact catcaatgta tcttaacgcg agtttaggcg cg 17802

<210> 56
<211> 15134
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự của cấu trúc pLA3242-Crtra intron-reaperKR
<400> 56
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt taccttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatcccccа taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gaactaccca ccgtactcgt caattccaag ggcatcggta 960
aacatctgct caaactcgaa gtcggccata tccagagcgc cgtagggggc ggagtcgtgg 1020
ggggtaaatc ccggacccgg ggaatccccg tcccccaaca tgtccagatc gaaatcgtct 1080
agcgcgtcgg catgcgccat cgccacgtcc tcgccgtcta agtggagctc gtcccccagg 1140
ctgacatcgg tcgggggggc cgtcgacagt ctgcgcgtgt gtcccgcggg gagaaaggac 1200
aggcgcggag ccgccagccc cgcctcttcg ggggcgtcgt cgtccgggag atcgagcagg 1260
ccctcgatgg tagacccgta attgttttтc gtacgcgcgc ggctgtacgc ggggcccgag 1320
cccgactcgc atttcagttg cttttccaat ccgcagataa tcagctccaa gccgaacagg 1380
aatgccggct cggctccttg atgatcgaac agctcgattg cctgacgcag cagtgggggc 1440

atcgaatcgg ttgttggggt ctcgcgctcc tcttttgcga cttgatgctc ttggtcctcc 1500
agcacgcagc ccagggtaaa gtgaccgacg gcgctcagag cgtagagagc attttccagg 1560
ctgaagcctt gctggcacag gaacgcgagc tggttctcca gtgtctcgta ttgcttttcg 1620
gtcgggcgcg tgccgagatg gactttggca ccgtctcggt gggacagcag agcgcagcgg 1680
aacgacttgg cgttattgcg gaggaagtcc tgccaggact cgccttccaa cgggcaaaaa 1740
tgcgtgtggt ggcggtcgag catctcgatg gccagggcat ccagcagcgc ccgcttattc 1800
ttcacgtgcc agtagagggt gggctgctcc acgcccagct tctgcgccaa cttgcgggtc 1860
gtcagtccct caatgccaac ttcgttcaac agctccaacg cggagttgat gactttggac 1920
ttatccaggc ggctgaccta tagataccat agatgtatgg attagtatca tatacataca 1980
aaggctattt ttgggacata ttaatattaa caatttccgt gatagttttc accatttttg 2040
ttgaatgtta cgttgaaaat ttaaatttgt tttaaattaa ttttaccagt catgtgttct 2100
taaaagtttt tatgattgaa acggcataaa gtggttcaaa aatttatcaa gaaaggcttt 2160
ccttttttaa atcttatctt tttctcttaa aaatcactag tcaattcatt attaatttgt 2220
taacttgaat ttggaatgtc tatttacttt cagataaatt aaagcaagaa acttaatatt 2280
cgaaaaaaat tgattctaaa tggaatttca cttgatcttc atgtatgcat atcaattttt 2340
atttacattg tataataagt ttcgagttga ttgttgtaat ccacaggtgt cccagagaat 2400
taaattccaa attacccaag tttattgaat gttgattgta gtttcagttg ctttgttgct 2460
gcaacaatgg cttgttgatt gtagatattt tccctttcct tggtttactt attacataga 2520
ctgaaaaaga ggtttacttt tttgatactt atgaaaaatt tctattagtg attactaacc 2580
aatcgctata tgtttactag aaaacaaata aactctttac attaacattc aataatgttt 2640
gctctgtaac cgacaattga aggcgttaca gcaacagtaa tataactagc ttcttaaccc 2700
tcatctatta accccatcgt ttaaaacact atgttaaatg gtctaacaaa tctagatact 2760
aatagatgtc ttattactta gcagccacag ctgcaacatc caagacaatt tttgaaactt 2820
cttattgagc tcttggcagc agaaatgttg gtatttttca cagctttctg aaagaccggc 2880
accttcctcc ggttcccgtt tctgaattca agaggatttc cgacccccaa ttaatcccga 2940
aacaaataag gtatattcaa aatgatggaa aagtcatggc tgctgacctt atttttattc 3000
ctattgatag aatattattc ccctttaaa tacactgtac taagaggtcc ggctataatt 3060
ttactcactt gtcgattatc ccatagaatg ttgattgtag ttggttgctt ttccaggtga 3120
gagttgatca agtcacaaaa gttagcgtgt gttgattgta gatttgaagg taaaataatt 3180
tttgcaccca ttcatcgggt aaaacgttct ccatagaata catttccatc gataattgat 3240
aacttatgaa tttcaaagaa aaaaatatgc ttttaaaatt accatggtgg ctagcgcaga 3300
ttgtttagct tgttcagctg cgcttgttta tttgcttagc tttcgcttag cgacgtgttc 3360
actttgcttg tttgaattga attgtcgctc cgtagacgaa gcgcctctat ttatactccg 3420
gcgctcgttt tcgagtttac cactccctat cagtgataga gaaaagtgaa agtcgagttt 3480
accactccct atcagtgata gagaaaagtg aaagtcgagt ttaccactcc ctatcagtga 3540
tagagaaaag tgaaagtcga gtttaccact ccctatcagt gatagagaaa agtgaaagtc 3600
gagtttacca ctccctatca gtgatagaga aaagtgaaag tcgagtttac cactccctat 3660
cagtgataga gaaaagtgaa agtcgagttt accactccct atcagtgata gagaaaagtg 3720
aaagtcgaaa cctggcgcgc ctaaactcgc gttaagatac attgatgagt ttggacaaac 3780
cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt 3840
atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca ttcattttat 3900
gtttcaggtt caggggggag tgtgggaggt tttttaaagc aagtaaaacc tctacaaatg 3960
tggtatggct gattatgatc accggtgtta ctggctcgga cggcggtaac ggcgcgctgcg 4020
gcggccggtg cgcgggtggc agctggtgta ctggcgcagg gtctccagca ccacggtggc 4080
caggaagcgc cactggctct cgcgcaggcg caggatctgc tgctcgcgct gctcggcctc 4140
gcgcagcagg gtggcctgat ccgggatgta gaaggccacc taaagatacc atggatgtat 4200
gaattagtat catatacata taaatgcttt ttttttggc atattaatgt taaaaatatc 4260
aacaatttcc gtgatagttt ttaccatttt tgttgaatgt ttactttgaa aacttaaata 4320
ttttttaact aattttacca gtcatgtgtt attaaaagta tttatgaata aaactgcaag 4380

taaagcgttt caaaaattta tcaagtaaaa ctttactttt tttaaatctt aactgtcaat 4440
tcattattaa tttattaatt taaatttgca atgtctattt actttaagac aaattaaagc 4500
aagaaactaa atattcgaat caattctttt ttaaatgaaa ttttacttca tcatcatgta 4560
tgtgtgtatc aatttttatt tacattgtat aataagtttc gagttgattg ttgtaatccg 4620
caggtgtccc gaagtattaa attccgaatt cccaagttta ttgaatgttg attgtagttt 4680
cagttgtttt gttattgcaa caatggcttg ttgattggag atattttcct tttccttggt 4740
ttacttacta catagactga aaaagatgtt tgacttttt gatactattg taaaatttct 4800
attagtgatt actaaccaat cgctataagt ttaatagaaa acaaataaac tctttgcatc 4860
cagatatacc tagcttctta accettatct attaactcca ttgcttgtaa caaatctaga 4920
tattaataga tgtctaatta cttagcaaaa cttctttttg attaagcagc cacagctgtc 4980
gattttggtc atatttaaag gaaataaatg cgtttaaaat aataattaat ataagttttg 5040
aaacttttta ctaacacttg gcagcaggaa gtaggtgttt ttcacagctt tctgaaccac 5100
cggcaccttc cccggtctcc gttgtcggag ttcagcagga tttccggccc ccaattaacc 5160
ccgaaacaaa acatgtctta ttaataaggt gtattcaaaa tagtgggaat gtcatgactg 5220
ctgaccttat ttttattcct attgtaagtg ttccggctat aattttactc acttgtccat 5280
tatcccatag aatgttatgt tgattgtagt tgtttgcttt tccaggtgag agttgatcaa 5340
gtcgcaaaag ttagcgtgtg ttgattgtag atttgaaggt aaaataattt tgtacacatt 5400
catcaggcaa aacgttctcc atcgaataaa cttccatcga taattgatag cttatgaatt 5460
tcaaaaaaaa atatgctttt aaaattaccg ccatggtggt tgctagcttg tcgcagtggt 5520
aacaaactcg tgactgactg aacaattgca acaggcatta gtttatatat cgctcaggta 5580
gacggatcga tctctatcac tgatagggag gtctctatca ctgataggga gttctctatc 5640
actgataggg atgtctctat cactgatagg gatttctcta tcactgatag ggaagtctct 5700
atcactgata gggacctctc tatcactgat agggaaatct ctatcactga tagggatctc 5760
tctatcactg atagggactt ctctatcact gatagggacg tctctatcac tgatagggaa 5820
ctctctatca ctgataggga catctctatc actgataggg acttctctat cactgatagg 5880
gagtatgttc tctctcttct cttctctctc tctttctcga atgttctctc tcttctcttc 5940
tctctctctt tctcgatggc cggcctggct taattaactc gcgttaagat acattgatga 6000
gtttggacaa accacaacta gaatgcagtg aaaaaaatgc tttatttgtg aaatttgtga 6060
tgctattgct ttatttgtaa ccattataag ctgcaataaa caagttaaca acaacaattg 6120
cattcatttt atgtttcagg ttcaggggga ggtgtgggag gtttttaaa gcaagtaaaa 6180
cctctacaaa tgtggtatgg ctgattatga tcagttatct agatccggtg gatcttacgg 6240
gtcctccacc ttccgctttt tcttgggtcg agatctgagt ccggaatcct cgtcgctacc 6300
gatggcgctg gtgatgcggg gcacgctgtg ggcgtaggtc acctcgcgct ggcacacgtg 6360
gtcgcgcttg tcgctggtgt ccctcatctg cttggtgatg atggtcacga agtgggggcc 6420
ggggatcttg atggcgcggc tgccgttgaa ggtcatcttg ctgtcgaagt ggcccatcat 6480
caggccgccg tcggcggtgg tgaagccgat gaaggccagc tggcgcacgg cgttggggcc 6540
gtgggggaac atgtgggtct cgttgggcag gatgtccacc agctggtcgc gcatgatggg 6600
gccgtcgggc tggaagccgt cgcagttcac ggtgatgcgg ctgaccacgc aggtgccgtc 6660
cagctcgtag gtgtggtggc tggtcatggt gccgtcgttc tcgaagcgca cggtgcggtc 6720
gatgctcagg ccctcgggga agcactcctg ggcgaagtgg ctgatgccgt tggggtagcg 6780
ggcgaagaag ggctcgccgt actggatcag gtggcagatg ggcttccagc tcatgggcag 6840
cttgccggtc tcgcacacgg cgtgcacgtt gaagtcgccg tgggggaact tgctgctgcc 6900
gtcggccacg atggtgaact tctggccgtt cacctcgccg tcgatgaaga ttttgaaggt 6960
catgtcgctc tggaacaggg cggggccgcc ctctgaacca tcctcgtcca tggtggcgac 7020
cggtttgcgc ttcttcttgg gtggggtggg atctcccatg gtggcctgaa tctcaacttg 7080
cacctgaagg tagtgcagca aggatgagca aaagggaaga acccagaaaa gaacgggaaa 7140
acttacccca attagaattg cttgtcgccg ccagtgtcaa cttgcaactg aaacaatatc 7200
caacatgaac gtcaatttat actgccctaa tggcgaacac gataacaata tttcttttat 7260
tatgccctct aaaaccaacg cggttatcgt ttatttattc aaattagata tagaacatcc 7320

gccgacatac aatgttaatg caaaaacgcg tttggtgagc ggatacgaaa acagtcggcc 7380
gataaacatt aatctgaggt cgataacacc gtccttgaac ggaacacgag gagcgtacgt 7440
gatcagctgc attcgcgcgc cgcgccttta tcgagattta tttgcataca acaagtacac 7500
tgcgccgttg ggatttgtgg taacgcgcac acatgcagag ctgcaagtgt ggcacatttt 7560
gtctgtgcgc aaaaccttg aagccaaaag tacgaggtcc gttacgggca tgctagcgca 7620
cacggacaat ggacccgaca aattctacgc caaggattta atgataatgt cgggcaacgt 7680
atccgttcat tttatcaata acctacaaaa atgtcgcgcg catcacaaag acatcgatat 7740
atttaaacat ttatgtcccg aactgcaaat cgataatagt gttgtgcaac ctcgagcgtc 7800
cgtttgattt aacgtatagc ttgcaaatga attatttaat tatcaatcat gttttacgcg 7860
tagaattcta cccgtaaagc gagtttagtt atgagccatg tgcaaaacat gacatcagct 7920
tttatttta taacaaatga catcatttct tgattgtgtt ttacacgtag aattctactc 7980
gtaaagcgag ttcagttttg aaaaacaaat gacatcatct ttttgattgt gctttacaag 8040
tagaattcta cccgtaaatc aagttcggtt ttgaaaaaca aatgagtcat attgtatgat 8100
atcatattgc aaaacaaatg actcatcaat cgatcgtgcg ttacacgtag aattctactc 8160
gtaaagcgag tttatgagcc gtgtgcaaaa catgacatca tctcgatttg aaaaacaaat 8220
gacatcatcc actgatcgtg cattacaagt agaattctac tcgtaaagcc agttcggtta 8280
tgagccgtgt acaaaacatg acatcagatt atgactcata cttgattgtg ttttacgcgt 8340
agaattctac tcgtaaagcc agttcaattt taaaaacaaa tgacatcatc caaattaata 8400
aatgacaagc aatgggtacc atgcggccgc accgaaatcg taattcacgg catcattaca 8460
aaatattttg acgttttgga cctcgtccct aatgacacca taacggtggc cttgaagtat 8520
atttaaccct agaaagatag tctgcgtaaa attgacgcat gcattcttga aatattgctc 8580
tctctttcta aatagcgcga atccgtcgct gtgcatttag gacatctcag tcgccgcttg 8640
gagctcccgt gaggcgtgct tgtcaatgcg gtaagtgtca ctgattttga actataacga 8700
ccgcgtgagt caaaatgacg catgattatc ttttacgtga cttttaagat ttaactcata 8760
cgataattat attgttattt catgttctac ttacgtgata acttattata tatatatttt 8820
cttgttatag atatcgtgac taatatataa taaaatgggt agttctttag acgatgagca 8880
tatcctctct gctcttctgc aaagcgatga cgagcttgtt ggtgaggatt ctgacagtga 8940
aatatcagat cacgtaagtg aagatgacgt ccaggaaatc tggccggccg caaccattgt 9000
gggaaccgtg cgatcaaaca aacgcgagat accggaagta ctgaaaaaca gtcgctccag 9060
gccagtggga acatcgatgt tttgttttga cggacccctt actctcgtct catataaacc 9120
gaagccagct aagatggtat acttattatc atcttgtgat gaggatgctt ctatcaacga 9180
aagtaccggt aaaccgcaaa tggttatgta ttataatcaa actaaaggcg gagtggacac 9240
gctagaccaa atgtgttctg tgatgacctg cagtaggaag acgaataggt ggcctatggc 9300
attattgtac ggaatgataa acattgcctg cataaattct tttattatat acagccataa 9360
tgtcagtagc aagggagaaa aggtccaaag tcgcaaaaaa tttatgagaa acctttacat 9420
gagcctgacg tcatcgttta tgcgtaagcg tttagaagct cctactttga agagatattt 9480
gcgcgataat atctctaata ttttgccaaa tgaagtgcct ggtacatcag atgacagtac 9540
tgaagagcca gtaatgaaaa aacgtactta ctgtacttac tgcccctcta aaataaggcg 9600
aaaggcaaat gcatcgtgca aaaaatgcaa aaaagttatt tgtcgagagc ataatattga 9660
tatgtgccaa agttgtttct gactgactaa taagtataat ttgtttctat tatgtataag 9720
ttaagctaat tacttatttt ataatacaac atgactgttt ttaaagtaca aaataagttt 9780
atttttgtaa aagagagaat gtttaaaagt tttgttactt tatagaagaa attttgagtt 9840
tttgtttttt tttaataaat aaataaacat aaataaattg tttgttgaat ttattattag 9900
tatgtaagtg taaatataat aaaacttaat atctattcaa attaataaat aaacctcgat 9960
atacagaccg ataaaacaca tgcgtcaatt ttacgcatga ttatctttaa cgtacgtcac 10020
aatatgatta tctttctagg gttaaataat agtttctaat ttttttatta ttcagcctgc 10080
tgtcgtgaat accgtatatc tcaacgctgt ctgtgagatt gtcgtattct agccttttta 10140
gtttttcgct catcgacttg atattgtccg acacattttc gtcgatttgc gttttgatca 10200
aagacttgag cagagacacg ttaatcaact gttcaaattg atccatatta acgatatcaa 10260

cccgatgcgt atatggtgcg taaaatatat tttttaaccc tcttatactt tgcactctgc 10320
gttaatacgc gttcgtgtac agacgtaatc atgttttctt ttttggataa aactcctact 10380
gagtttgacc tcatattaga ccctcacaag ttgcaaaacg tggcattttt taccaatgaa 10440
gaatttaaag ttattttaaa aaatttcatc acagatttaa agaagaacca aaaattaaat 10500
tatttcaaca gtttaatcga ccagttaatc aacgtgtaca cagacgcgtc ggcaaaaaac 10560
acgcagcccg acgtgttggc taaaattatt aaatcaactt gtgttatagt cacggatttg 10620
ccgtccaacg tgttcctcaa aaagttgaag accaacaagt ttacggacac tattaattat 10680
ttgattttgc cccacttcat tttgtgggat cacaattttg ttatatttta aacaaagctt 10740
ggcactggcc gtcgttttac aacgtcgtga ctgggaaaac cctggcgtta cccaacttaa 10800
tcgccttgca gcacatcccc ctttcgccag ctggcgtaat agcgaagagg cccgcaccga 10860
tcgcccttcc caacagttgc gcagcctgaa tggcgaatgg cgcctgatgc ggtattttct 10920
ccttacgcat ctgtgcggta tttcacaccg catatggtgc actctcagta caatctgctc 10980
tgatgccgca tagttaagcc agccccgaca cccgccaaca cccgctgacg cgccctgacg 11040
ggcttgtctg ctcccggcat ccgcttacag acaagctgtg accgtctccg ggagctgcat 11100
gtgtcagagg ttttcaccgt catcaccgaa acgcgcgaga cgaaagggcc tcgtgatacg 11160
cctatttta taggttaatg tcatgataat aatggtttct tagacgtcag gtggcacttt 11220
tcgggaaat gtgcgcggaa cccctatttg tttatttttc taaatacatt caaatatgta 11280
tccgctcatg agacaataac cctgataaat gcttcaataa tattgaaaaa ggaagagtat 11340
gagtattcaa catttccgtg tcgcccttat tccctttttt gcggcatttt gccttcctgt 11400
ttttgctcac ccagaaacgc tggtgaaagt aaaagatgct gaagatcagt tgggtgcacg 11460
agtgggttac atcgaactgg atctcaacag cggtaagatc cttgagagtt ttcgccccga 11520
agaacgtttt ccaatgatga gcacttttaa agttctgcta tgtggcgcgg tattatcccg 11580
tattgacgcc gggcaagagc aactcggtcg ccgcatacac tattctcaga atgacttggt 11640
tgagtactca ccagtcacag aaaagcatct tacggatggc atgacagtaa gagaattatg 11700
cagtgctgcc ataaccatga gtgataacac tgcggccaac ttacttctga caacgatcgg 11760
aggaccgaag gagctaaccg ctttttgca caacatgggg gatcatgtaa ctcgccttga 11820
tcgttgggaa ccggagctga atgaagccat accaaacgac gagcgtgaca ccacgatgcc 11880
tgtagcaatg gcaacaacgt tgcgcaaact attaactggc gaactactta ctctagcttc 11940
ccggcaacaa ttaatagact ggatggaggc ggataaagtt gcaggaccac ttctgcgctc 12000
ggcccttccg gctggctggt ttattgctga taaatctgga gccggtgagc gtgggtctcg 12060
cggtatcatt gcagcactgg ggccagatgg taagccctcc cgtatcgtag ttatctacac 12120
gacggggagt caggcaacta tggatgaacg aaatagacag atcgctgaga taggtgcctc 12180
actgattaag cattggtaac tgtcagacca agtttactca tatatacttt agattgattt 12240
aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atctcatgac 12300
caaaatccct taacgtgagt tttcgttcca ctgagcgtca gaccccgtag aaaagatcaa 12360
aggatcttct tgagatcctt ttttctgcg cgtaatctgc tgcttgcaaa caaaaaaacc 12420
accgctacca gcggtggttt gtttgccgga tcaagagcta ccaactcttt ttccgaaggt 12480
aactggcttc agcagagcgc agataccaaa tactgtcctt ctagtgtagc cgtagttagg 12540
ccaccacttc aagaactctg tagcaccgcc tacatacctc gctctgctaa tcctgttacc 12600
agtggctgct gccagtggcg ataagtcgtg tcttaccggg ttggactcaa gacgatagtt 12660
accggataag gcgcagcggt cgggctgaac gggggttcg tgcacacagc ccagcttgga 12720
gcgaacgacc tacaccgaac tgagatacct acagcgtgag cattgagaaa gcgccacgct 12780
tcccgaaggg agaaaggcgg acaggtatcc ggtaagcggc agggtcggaa caggagagcg 12840
cacgagggag cttccagggg gaaacgcctg gtatctttat agtcctgtcg ggtttcgcca 12900
cctctgactt gagcgtcgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa 12960
cgccagcaac gcggcctttt tacggttcct ggccttttgc tggccttttg ctcacatgtt 13020
ctttcctgcg ttatcccctg attctgtgga taaccgtatt accgcctttg agtgagctga 13080
taccgctcgc cgcagccgaa cgaccgagcg cagcgagtca gtgagcgagg aagcggaaga 13140
gcgcccaata cgcaaaccgc ctctccccgc gcgttggccg attcattaat gcagctggca 13200

cgacaggttt cccgactgga aagcgggcag tgagcgcaac gcaattaatg tgagttagct 13260
cactcattag gcaccccagg ctttacactt tatgcttccg gctcgtatgt tgtgtggaat 13320
tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg aatttcgacc 13380
tgcaggcatg caagcttgca tgcctgcagg tcgacgctcg cgcgacttgg tttgccattc 13440
tttagcgcgc gtcgcgtcac acagcttggc cacaatgtgg tttttgtcaa acgaagattc 13500
tatgacgtgt ttaaagttta ggtcgagtaa agcgcaaatc ttttttaacc ctagaaagat 13560
agtctgcgta aaattgacgc atgcattctt gaaatattgc tctctctttc taaatagcgc 13620
gaatccgtcg ctgtgcattt aggacatctc agtcgccgct tggagctccc gtgaggcgtg 13680
cttgtcaatg cggtaagtgt cactgatttt gaactataac gaccgcgtga gtcaaaatga 13740
cgcatgatta tcttttacgt gacttttaag atttaactca tacgataatt atattgttat 13800
ttcatgttct acttacgtga taacttatta tatatatatt ttcttgttat agatatcgtg 13860
actaatatat aataaaatgg gtagttcttt agacgatgag catatcctct ctgctcttct 13920
gcaaagcgat gacgagcttg ttggtgagga ttctgacagt gaaatatcag atcacgtaag 13980
tgaagatgac gtccagagcg atacagaaga agcgtttata gatgaggtac atgaagtgca 14040
gccaacgtca agcggtagtg aaatattaga cgaacaaaat gttattgaac aaccaggttc 14100
ttcattggct tctaacagaa tcttgacctt gccacagagg actattagag gtaagaataa 14160
acattgttgg tcaacttcaa agtccacgag gcgtagccga gtctctgcac tgaacattgt 14220
cagatcggcc cggcggagtg gacacgctag accaaatgtg ttctgtgatg acctgcagta 14280
ggaagacgaa taggtggcct atggcattat tgtacggaat gataaacatt gcctgcataa 14340
attcttttat tatatacagc cataatgtca gtagcaaggg agaaaaggtc caaagtcgca 14400
aaaaatttat gagaaacctt tacatgagcc tgacgtcatc gtttatgcgt aagcgtttag 14460
aagctcctac tttgaagaga tatttgcgcg ataatatctc taatattttg ccaaatgaag 14520
tgcctggtac atcagatgac agtactgaag agccagtaat gaaaaaacgt acttactgta 14580
cttactgccc ctctaaaata aggcgaaagg caaatgcatc gtgcaaaaaa tgcaaaaaag 14640
ttatttgtcg agagcataat attgatatgt gccaaagttg tttctgactg actaataagt 14700
ataatttgtt tctattatgt ataagttaag ctaattactt attttataat acaacatgac 14760
tgtttttaaa gtacaaaata agtttatttt tgtaaaagag agaatgttta aaagttttgt 14820
tactttatag aagaaatttt gagtttttgt tttttttaa taaataaata aacataaata 14880
aattgtttgt tgaatttatt attagtatgt aagtgtaaat ataataaaac ttaatatcta 14940
ttcaaattaa taaataaacc tcgatataca gaccgataaa acacatgcgt caattttacg 15000
catgattatc tttaacgtac gtcacaatat gattatcttt ctagggttaa aatgaatgta 15060
agcactttat taacgaaatc tttgggaata tttcgctcat cagcatttta tttgagcagg 15120
agtccgagat gccc 15134

<210> 57
<211> 1403
<212> ADN
<213> Nhân tạo
<220>
<223> SEQ ID NO. 57 Trình tự một phần của sản phẩm phiêm mã giống đực được tạo ra ở Drosophila melanogaster từ các thể biến nạp LA3077 khác với trình tự được tạo ra từ các dòng Medfly LA3077
<400> 57

ggccagatct gttgttatta aacgtagatt tggtaatttt aaaagcatat ttttttcttt 60
gaaattcata agttatcaat tatcgatgga aatgtattct atggagaacg ttttacccga 120
tgaatgggtg caaaaattat tttaccttca aatctacaat caacacacgc taacttttgt 180
gacttgatca actctcacct ggaaaagcaa ccaactacaa tcaacattct atgggataat 240
cgacaagtga gtaaaattat agccggacct cttagtacag tgtatttaaa aggggaataa 300
tattctatca ataggaataa aaataaggtc agcagccatg acttttccat cattttgaat 360

ataccttatt tgtttcggga ttaattgggg gtcggaaatc ctcttgaatt cagaaacggg 420
aaccggagga aggtgccggt ctttcagaaa gctgtgaaaa ataccaacat ttctgctgcc 480
aagagctcaa taagaagttt caaaaattgt cttggatgtt gcagctgtgg ctgctaagta 540
ataagacatc tattagtatc tagatttgtt agaccattta acatagtgtt ttaaacgatg 600
gggttaatag atgagggtta agaagctagt tatattactg ttgctgtaac gccttcaatt 660
gtcggttaca gagcaaacat tattgaatgt taatgtaaag agtttatttg ttttctagta 720
aacatatagc gattggttag taatcactaa tagaaatttt tcataagtat caaaaaagta 780
aacctctttt tcagtctatg taataagtaa accaaggaaa gggaaaatat ctacaatcaa 840
caagccattg ttgcagcaac aaagcaactg aaactacaat caacattcaa taaacttggg 900
taatttggaa tttaattctc tgggacacct gtggattaca acaatcaact cgaaacttat 960
tatacaatgt aaataaaaat tgatatgcat acatgaagat caagtgaaat tccatttaga 1020
atcaattttt ttcgaatatt aagtttcttg ctttaattta tctgaaagta aatagacatt 1080
ccaaattcaa gttaacaaat taataatgaa ttgactagtg atttttaaga gaaaaagata 1140
agatttaaaa aaggaaagcc tttcttgata aatttttgaa ccactttatg ccgtttcaat 1200
cataaaaact tttaagaaca catgactggt aaaattaatt taaaacaaat ttaaattttc 1260
aacgtaacat tcaacaaaaa tggtgaaaac tatcacggaa attgttaata ttaatatgtc 1320
ccaaaaatag cctttgtatg tatatgatac taatccatac atctatggta tctataggtg 1380
aaggctcaaa gcctctggct agc 1403

<210> 58
<211> 972
<212> ADN
<213> Bactrocera zonata
<400> 58
cggtaattct aattacttac taaatatagt gatatttaaa aatgctatga gatcttgagg 60
cgaacttaaa tgagacaagg tgttagcttg ctcaaatgac gcaccaatca acttaccaag 120
ataatttgga gcaatatgga atattttctt gtcaaagaac tacaatcaac atgcaactgg 180
tataataaag tgagtataaa gtagaaaggt gcggtataat atttttctct aaagtattaa 240
ggtaagtttt taaaaataat ttaataattt ttcagagtgt gcagtattag tgtctaaaag 300
tattgcaatc tttgtgtttt aaaaatttgt gcagaatatc aaaaaatagg acgacctatt 360
gttagtaata ttatgtatat ttacttattt tcaaataaat aaaattattt tctagccttt 420
agaaacagcc ttatactaca atcaacaaac cataaatact ccctgcaaca attacaaagc 480
aacacaaaca acaatcaaca ttccatgaac tcggacaatt tggaaattga tatcttctgg 540
cacatgcgga ttacaacaat caactcaaac ctattagatg atgtaaataa atataaacct 600
tcatattcaa taacttaatt cgtgatagaa atttagagtt aattactgtt taatcgaaat 660
ttgaacgtct cgtttgaatt cactttaacg tttgaaagat gatctttaaa tagtgactgg 720
agacttaaat tttaatcttt ttaattaaat aatctaacaa agttctttct tattctaatc 780
ttaatatata atcttgtttt aaaatttgaa ttaaacttat ttcgttgagt aacaattaaa 840
tgatttgaat cgtaaaaaat atcattataa ttattcaaca aaaagctgaa ggctgcgatg 900
aaaaattgtt gatatataaa tcaaatttat ttttataaat cactaatagt gacattaacc 960
gtattcacag gt 972

<210> 59
<211> 1312
<212> ADN
<213> Ceratitis rosa
<400> 59
tggtaatttt aaaagcatat tttttttga aattcataag ctatcaatta tcgatggaag 60
tttattcgat ggagaacgtt ttgcctgatg aatgtgtaca aaattatttt accttcaaat 120

ctacaatcaa cacacgctaa cttttgcgac ttgatcaact ctcacctgga aaagcaaaca 180
actacaatca acataacatt ctatgggata atggacaagt gagtaaaatt atagccggaa 240
cacttacaat aggaataaaa ataaggtcag cagtcatgac attcccacta ttttgaatac 300
accttattaa taagacatgt tttgtttcgg ggttaattgg gggccggaaa tcctgctgaa 360
ctccgacaac ggagaccggg gaaggtgccg gtggttcaga aagctgtgaa aaacacctac 420
ttcctgctgc caagtgttag taaaaagttt caaaacttat attaattatt attttaaacg 480
catttatttc ctttaaatat gaccaaaatc gacagctgtg gctgcttaat caaaaagaag 540
ttttgctaag taattagaca tctattaata tctagatttg ttacaagcaa tggagttaat 600
agataagggt taagaagcta ggtatatctg gatgcaaaga gtttatttgt tttctattaa 660
acttatagcg attggttagt aatcactaat agaaatttta caatagtatc aaaaaagtca 720
aacatctttt tcagtctatg tagtaagtaa accaaggaaa aggaaaatat ctccaatcaa 780
caagccattg ttgcaataac aaaacaactg aaactacaat caacattcaa taaacttggg 840
aattcggaat ttaatacttc gggacacctg cggattacaa caatcaactc gaaacttatt 900
atacaatgta aataaaaatt gatacacaca tacatgatga tgaagtaaaa tttcatttaa 960
aaaagaattg attcgaatat ttagtttctt gctttaattt gtcttaaagt aaatagacat 1020
tgcaaattta aattaataaa ttaataatga attgacagtt aagatttaaa aaaagtaaag 1080
ttttacttga taaatttttg aaacgcttta cttgcagttt tattcataaa tacttttaat 1140
aacacatgac tggtaaaatt agttaaaaaa tatttaagtt ttcaaagtaa acattcaaca 1200
aaaatggtaa aaactatcac ggaaattgtt gatattttta acattaatat gccaaaaaaa 1260
aaagcattta tatgtatatg atactaattc atacatccat ggtatcttta gg 1312

<210> 60
<211> 21
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi spl-agdsx-e3
<400> 60
cgagcccaat ggctgttgga g 21

<210> 61
<211> 22
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi spl-agdsx-m
<400> 61
gtcaaggttc agggcccgat cg 22

<210> 62
<211> 21
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi spl-agdsx-e3
<400> 62
cgagcccaat ggctgttgga g 21

<210> 63

<211> 22
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi spl-agdsx-m
<400> 63
gtcaaggttc agggcccgat cg 22

<210> 64
<211> 20
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi aedesxF1
<400> 64
tcaatggctc ctggagaagc 20

<210> 65
<211> 25
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi aedesxR5
<400> 65
accattcttg cagaagtctt gggac 25

<210> 66
<211> 19
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi aedesxR2
<400> 66
aacattctcc gcgcacagg 19

<210> 67
<211> 23
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi Agexon1
<400> 67
gacgctcgct ctggtacagt tcg 23

<210> 68
<211> 20
<212> ADN
<213> Nhân tạo
<220>

<223> Đoạn mồi Tra (tTAV) seq+
<400> 68
cctgccagga ctcgccttcc 20

<210> 69
<211> 23
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi Agexon1
<400> 69
gacgctcgct ctggtacagt tcg 23

<210> 70
<211> 26
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi Exon 3
<400> 70
gttgtcgctt tgactggcaa tgtcgc 26

<210> 71
<211> 632
<212> ADN
<213> Pectinophora gossypiella
<400> 71
gaactgccac aaactgctgg aaaagttcca ctactcctgg gaaatgatgc ccctggtgct 60
ggtcattcta aactacgccg gctccgacct cgacgaggct tctagaaaaa ttgatgaagg 120
gaagatgatc atcaacgagt acgcgaggga gcacaatctg aacatcttcg atggccacga 180
gctgaggaac tcgactcgcc agaaaatgct gagcgaaatt aataatataa gtggtgtact 240
atcgtcgtcc atgaagttat tttgcgaatg atactttgtt ttgtatgtgc tgtgtgttgt 300
gtggactttt gctgtgcgtt gctgtttgcg atggaaggac tattgtgtcg tcgccacgct 360
ggactattcg cacattgggt ggtccaccag tggcggatgt acgagcggtc gctgtgctcg 420
ctcctggagc tgcaagcgcg caaagggacg tactcggtgt gctgctcacc ccgctacgtc 480
atcgcgcccg agtacgcgtc acacctgttg cctctgccgc ttaccacgca gagatcatcc 540
ccgccgcccg cgcacttgta gcgatgcgaa cctgcgccgc gggaagcggc gcaagaaccc 600
gccgatgccc cggcgtcgtc gtcgggtgcc ac 632

<210> 72
<211> 222
<212> ADN
<213> Drosophila melanogaster
<400> 72
atgcagatct ttgtgaagac tttgaccgga aagaccatca ccctcgaggt agagccatcg 60
gacaccattg agaatgtaaa ggccaagatt caggataagg agggaatccc cccagatcag 120
cagcgtctga tcttcgctgg caagcaactg gaagacggac gcaccctgtc cgattacaac 180
atccagaagg agtccaccct tcacttggtc cttcgtctcc gt 222

<210> 73
<211> 74
<212> PRT
<213> Drosophila melanogaster
<400> 73
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu
1 5 10 15
Val Glu Pro Ser Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp
20 25 30
Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys
35 40 45
Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu
50 55 60
Ser Thr Leu His Leu Val Leu Arg Leu Arg
65 70

<210> 74
<211> 34
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 74
caagcaaagt gaacacgtcg ctaagcgaaa gcta 34

<210> 75
<211> 22
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 75
gcgggtggca gctggtgtac tg 22

<210> 76
<211> 34
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 76
caagcaaagt gaacacgtcg ctaagcgaaa gcta 34

<210> 77
<211> 24
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi

<400> 77
gcggaacgac ttggcgttat tgcg 24

<210> 78
<211> 28
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 78
ggaagggtcc ttacgctata gagcgcag 28

<210> 79
<211> 31
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 79
ccaggcgaag ttgttattaa gcgtagattt g 31

<210> 80
<211> 33
<212> DNA
<213> artificial
<220>
<223> primer
<400> 80
cgtcgctttg aaacagaggc tttgagcctt ctc 33

<210> 81
<211> 48
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 81
gctagcaacc accatggcgg taattctaat tacttactaa atatagtg 48

<210> 82
<211> 41
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 82
ccgggatgta gaaggccacc tgtgaatacg gttaatgtca c 41

<210> 83

<211> 31
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 83
cagtcagtca cgagtttgtt accactgcga c 31

<210> 84
<211> 22
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 84
gcgggtggca gctggtgtac tg 22

<210> 85
<211> 21
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 85
cggagcacat ctgatagaac g 21

<210> 86
<211> 23
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 86
cgcggctgta ggcgctgccg ctc 23

<210> 87
<211> 31
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 87
ccaggcgaag ttgttattaa gcgtagattt g 31

<210> 88
<211> 33
<212> ADN
<213> Nhân tạo
<220>

<223> Đoạn mồi
<400> 88
cgtcgctttg aaacagaggc tttgagcctt ctc 33

<210> 89
<211> 52
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 89
gctagcaacc accatggcgg taattttaaa agcatatttt tttttgaaat tc 52

<210> 90
<211> 41
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 90
ccgggatgta gaaggccacc taaagatacc atggatgtat g 41

<210> 91
<211> 31
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 91
cagtcagtca cgagtttgtt accactgcga c 31

<210> 92
<211> 22
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 92
gcgggtggca gctggtgtac tg 22

<210> 93
<211> 21
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 93
gttgcaagtt gacactggcg g 21

<210> 94
<211> 23
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 94
aggtgtggga ggttttttaa agc 23

<210> 95
<211> 52
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 95
cctgtaatac gactcactat agggcgtttt ttttttttt ttttttttt tt 52

<210> 96
<211> 33
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 96
gcaaacggca atcagacggg cccaggctca gga 33

<210> 97
<211> 28
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 97
cctgtaatac gactcactat agggcgtt 28

<210> 98
<211> 37
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 98
gggatcgagc tagatcggcc tgagccgcca gtggtga 37

<210> 99
<211> 28
<212> ADN
<213> Nhân tạo

<220>
<223> Đoạn mồi
<400> 99
cctgtaatac gactcactat agggcgtt 28

<210> 100
<211> 32
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 100
cgctccatgg gatcggcgag ctgcgactcc gt 32

<210> 101
<211> 27
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 101
gcaacaacca gcggtgtccc ttgaaac 27

<210> 102
<211> 28
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 102
cctgtaatac gactcactat agggcgtt 28

<210> 103
<211> 28
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 103
gctagtggag aactgccaca aactgctg 28

<210> 104
<211> 34
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 104
caagcaaagt gaacacgtcg ctaagcgaaa gcta 34

<210> 105
<211> 25
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi
<400> 105
gccctcgatg gtagacccgt aattg 25

<210> 106
<211> 14874
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự nucleotit LA1172, bao gồm bộ khung plasmit
<400> 106
gggctggccg caaccattgt gggaaccgtg cgatcaaaca aacgcgagat accggaagta 60
ctgaaaaaca gtcgctccag gccagtggga acatcgatgt tttgttttga cggacccctt 120
actctcgtct catataaacc gaagccagct aagatggtat acttattatc atcttgtgat 180
gaggatgctt ctatcaacga aagtaccggt aaaccgcaaa tggttatgta ttataatcaa 240
actaaaggcg gagtggacac gctagaccaa atgtgttctg tgatgacctg cagtaggaag 300
acgaataggt ggcctatggc attattgtac ggaatgataa acattgcctg cataaattct 360
tttattatat acagccataa tgtcagtagc aagggagaaa aggtccaaag tcgcaaaaaa 420
tttatgagaa acctttacat gagcctgacg tcatcgttta tgcgtaagcg tttagaagct 480
cctactttga agagatattt gcgcgataat atctctaata ttttgccaaa tgaagtgcct 540
ggtacatcag atgacagtac tgaagagcca gtaatgaaaa aacgtactta ctgtacttac 600
tgcccctcta aaataaggcg aaaggcaaat gcatcgtgca aaaaatgcaa aaaagttatt 660
tgtcgagagc ataatattga tatgtgccaa agttgtttct gactgactaa taagtataat 720
ttgtttctat tatgtataag ttaagctaat tacttatttt ataatacaac atgactgttt 780
ttaaagtaca aaataagttt atttttgtaa aagagagaat gtttaaaagt tttgttactt 840
tatagaagaa attttgagtt tttgtttttt tttaataaat aaataaacat aaataaattg 900
tttgttgaat ttattattag tatgtaagtg taaatataat aaaacttaat atctattcaa 960
attaataaat aaacctcgat atacagaccg ataaaacaca tgcgtcaatt ttacgcatga 1020
ttatctttaa cgtacgtcac aatatgatta tctttctagg gttaaataat agtttctaat 1080
ttttttatta ttcagcctgc tgtcgtgaat accgtatatc tcaacgctgt ctgtgagatt 1140
gtcgtattct agccttttta gtttttcgct catcgacttg atattgtccg acacattttc 1200
gtcgatttgc gttttgatca aagacttgag cagagacacg ttaatcaact gttcaaattg 1260
atccatatta acgatatcaa cccgatgcgt atatggtgcg taaaatatat tttttaaccc 1320
tcttatactt tgcactctgc gttaatacgc gttcgtgtac agacgtaatc atgttttctt 1380
ttttggataa aactcctact gagtttgacc tcatattaga ccctcacaag ttgcaaaacg 1440
tggcattttt taccaatgaa gaatttaaag ttattttaaa aaatttcatc acagatttaa 1500
agaagaacca aaaattaaat tatttcaaca gtttaatcga ccagttaatc aacgtgtaca 1560
cagacgcgtc ggcaaaaaac acgcagcccg acgtgttggc taaaattatt aaatcaactt 1620
gtgttatagt cacggatttg ccgtccaacg tgttcctcaa aaagttgaag accaacaagt 1680
ttacggacac tattaattat ttgattttgc cccacttcat tttgtgggat cacaattttg 1740
ttatatttta aacaaagctt ggcactggcc gtcgttttac aacgtcgtga ctgggaaaac 1800
cctggcgtta cccaacttaa tcgccttgca gcacatcccc ctttcgccag ctggcgtaat 1860
agcgaagagg cccgcaccga tcgcccttcc caacagttgc gcagcctgaa tggcgaatgg 1920

cgcctgatgc ggtattttct ccttacgcat ctgtgcggta tttcacaccg catatatggt 1980
gcactctcag tacaatctgc tctgatgccg catagttaag ccagccccga cacccgccaa 2040
cacccgctga cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg 2100
tgaccgtctc cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga 2160
gacgaaaggg cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt 2220
cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aacccctatt tgtttatttt 2280
tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat 2340
aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 2400
ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 2460
ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 2520
tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc 2580
tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac 2640
actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 2700
gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca 2760
acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg 2820
gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg 2880
acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 2940
gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 3000
ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 3060
gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 3120
cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 3180
agatcgctga gataggtgcc tcactgatta agcattggta actgtcagac caagtttact 3240
catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 3300
tccttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 3360
cagaccccgt agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct 3420
gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc 3480
taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 3540
ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc 3600
tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 3660
ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acggggggtt 3720
cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 3780
agcattgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 3840
gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt 3900
atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 3960
gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt 4020
gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 4080
ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt 4140
cagtgagcga ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 4200
cgattcatta atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca 4260
acgcaattaa tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc 4320
cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg 4380
accatgatta cgaatttcga cctgcaggca tgcaagcttg catgcctgca ggtcgacgct 4440
cgcgcgactt ggtttgccat tctttagcgc gcgtcgcgtc acacagcttg gccacaatgt 4500
ggtttttgtc aaacgaagat tctatgacgt gtttaaagtt taggtcgagt aaagcgcaaa 4560
tctttttaa ccctagaaag atagtctgcg taaaattgac gcatgcattc ttgaaatatt 4620
gctctctctt tctaaatagc gcgaatccgt cgctgtgcat ttaggacatc tcagtcgccg 4680
cttggagctc ccgtgaggcg tgcttgtcaa tgcggtaagt gtcactgatt ttgaactata 4740
acgaccgcgt gagtcaaaat gacgcatgat tatcttttac gtgactttta agatttaact 4800
catacgataa ttatattgtt atttcatgtt ctacttacgt gataacttat tatatatata 4860

ttttcttgtt atagatatcg tgactaatat ataataaaat gggtagttct ttagacgatg 4920
agcatatcct ctctgctctt ctgcaaagcg atgacgagct tgttggtgag gattctgaca 4980
gtgaaatatc agatcacgta agtgaagatg acgtccagag cgatacagaa gaagcgttta 5040
tagatgaggt acatgaagtg cagccaacgt caagcggtag tgaaatatta gacgaacaaa 5100
atgttattga acaaccaggt tcttcattgg cttctaacag aatcttgacc ttgccacaga 5160
ggactattag aggtaagaat aaacattgtt ggtcaacttc aaagtccacg aggcgtagcc 5220
gagtctctgc actgaacatt gtcagatcgg ccaggccggc cagatttaaa tgagcggccg 5280
catggtacca tactcggtgg cctccccacc accaactttt ttgcactgca aaaaaacacg 5340
cttttgcacg cgggcccata catagtacaa actctacgtt tcgtagacta ttttacataa 5400
atagtctaca ccgttgtata cgctccaaat acactaccac acattgaacc tttttgcagt 5460
gcaaaaaagt acgtgtcggc agtcacgtag gccggcctta tcgggtcgcg tcctgtcacg 5520
tacgaatcac attatcggac cggacgagtg ttgtcttatc gtgacaggac gccagcttcc 5580
tgtgttgcta accgcagccg gacgcaactc cttatcggaa caggacgcgc ctccatatca 5640
gccgcgcgtt atctcatgcg cgtgaccgga cacgaggcgc ccgtcccgct tatcgcgcct 5700
ataaatacag cccgcaacga tctggtaaac acagttgaac agcatctgtt acagcgacac 5760
aacatgagcc ggtccaacaa cgccaacgcg cccacgccat ccaaccgccg ccgcaacctg 5820
tctctggtgg atcccacccc acccaagaag aagcgcaaac cggtcgccac catggcctcc 5880
tccgagaacg tcatcaccga gttcatgcgc ttcaaggtgc gcatggaggg caccgtgaac 5940
ggccacgagt tcgagatcga gggcgagggc gagggccgcc cctacgaggg ccacaacacc 6000
gtgaagctga aggtgaccaa gggcggcccc ctgcccttcg cctgggacat cctgtccccc 6060
cagttccagt acggctccaa ggtgtacgtg aagcaccccg ccgacatccc cgactacaag 6120
aagctgtcct tccccgaggg cttcaagtgg gagcgcgtga tgaacttcga ggacggcggc 6180
gtggcgaccg tgacccagga ctcctccctg caggacggct gcttcatcta caaggtgaag 6240
ttcatcggcg tgaacttccc ctccgacggc cccgtgatgc agaagaagac catgggctgg 6300
gaggcctcca ccgagcgcct gtacccccgc gacggcgtgc tgaagggcga gacccacaag 6360
gccctgaagc tgaaggacgg cggccactac ctggtggagt tcaagtccat ctacatggcc 6420
aagaagcccg tgcagctgcc cggctactac tacgtggacg ccaagctgga catcacctcc 6480
cacaacgagg actacaccat cgtggagcag tacgagcgca ccgagggccg ccaccacctg 6540
ttcctgagat ctcgacccaa gaaaaagcgg aaggtggagg acccgtaaga tccaccggat 6600
ctagataact gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa 6660
acctcccaca cctccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact 6720
tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata 6780
aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttaac 6840
gcgagttaat taatccattg ctgggcgagc tgcgccaatc gatgccaacg ccaccctgca 6900
tggcgagcgg caggccggcg gctaccatgg gcgtcaccat gccctgaccg cccccggagg 6960
gcagtgaaaa atgtgtgggg ggtggtgggg gctgcgcagg aactgattgt gattatggtt 7020
gtgcccatgg ccatgttgtc caagtccatg gacgtgggca tgcttgttgt agcccaaatc 7080
ggcgtttccg tttccaccag gaaacatctc tgcttgtagt tcgaatatgc tctttaaatc 7140
ccagctgtat tcctcagtta tcgaggtttt cttcacgagt gaaacgaatt ttcgtcgcct 7200
tctacgccat tttcttgctc agcccgtttt gtcattcgca gcgaagcggt aacagcgggt 7260
cgctcatatg acggtatttt ttaatacact tcagctatac tgttatttca aaaacatatt 7320
tcttttgtta ctttttatgc agttcatttg ccaccaaaaa gtagtctttt ggattgattt 7380
atttcaaaaa atggtgtaat tcaagaaatt cagagggcca agtaatatac ttaatgaccg 7440
ttatttaaaa cacactcaag gagatttatt taaacggcta caatggtttt ccaaataact 7500
tatttactgt tgacttctat aaaacatagg tgtatatatt attatttcct tattgagttt 7560
gagataattt taatttccac aatatttttt cttgtgatta acagagaaag tcaaactaca 7620
taacatttat cgggtaaaag tctctatgaa ggtagcggtt aacagtgaag tcgcaaaagt 7680
ggtggccgta cgccaatcga gcgtagtacc cctaacctgc aatattttta gttggttttt 7740
tccgcaatag ccccagtttt ctcaaagagt gcaacaagtg attctgttta tgttttcaac 7800

aacttctctc tgcggaactt aacgtgagcg gacgtatgcg gacgcgccat ggtttaaact 7860
cgctagcact gggaagttga cgttgatata gagccgaatt gaacttcacc gctgcttggt 7920
aattactcta caagttcatt taggagaacc ggattcgaaa gatgatttte cagcgtttag 7980
ctttcagatg gccgcataca ttttgcacca ccaaaccgaa actcactagc gtatccaatc 8040
gttcgttttt tggtgccggt gtgttacgaa ctttagctat caagctaaag caatttgctc 8100
tggtcttccg tgctaaaaag aaaaaaaaac tgtttttttt ttggttttga tatttgcgct 8160
atttttactt gggccttaat tgaacaaact tttgaaagtt tccacagcga aatcgttttc 8220
gacgatgcca tttttggtaa catttgcatt ttcttgctca aattgcttgc aaaacccgtg 8280
aaagacatta atattcgata gtgtcatcca aaatcacgaa aatgattgtt gcaaaacgtt 8340
gaacaattta cacatgtaaa aaacaaccat cgattaatgt ttattcaaac tttttacaag 8400
aagggttatt ctgatcaatg tcacccccgct gatgaatgtt accccggatt acacttctcg 8460
aaaagtggtt caaaatgcta cttgagaatt tttatctgtc aaaggaagca aattcgagtc 8520
gaattaaatg gtatagtcct gaattaggtt tccatttact tacaggtatt ccactaaata 8580
gctggaagat ttattttaca caataatgat aattcgtacc ccaaagagtg tagccctact 8640
ttttttctctc ttttttttttt gtaaattttc atcgctgcgt gccagcttac cgacatgtcg 8700
cgacagcata aagagcctgt caagagatga agaaaaatga caaggagtca gtggtcaggt 8760
ctctgtatca atatttgacg tcctgacttt ccaatatacc tttccttaaa gagtagagat 8820
catgcgatac gtgaataaat atcgtttgga cttcgaaata gaacataatt taaggtagct 8880
gatcagtagt tgaacatctt cagacttctg ggacaagaag tgtttttttg tttgtagaaa 8940
aggtttttgt taaattatat ttgtaagata attcaatgaa tatatctctg attcagtaat 9000
caatccgtac cacgcaccgt ttaagaaaca ccctgtaggt ttgcatcacg tctcagacaa 9060
aagtgtatcg atgtgcgaac actgcatacc ggcgctttgc aaataatgcc aaatttagat 9120
atgcattaca ttgtcacttc gcaaaacaca cactcccaaa tgcgtcggaa acctcacccg 9180
aacgcacgat cgtaacgcga tcgatcgccg attgattgat cggaattaac tatctcaatc 9240
gatccttcta tggactgatg catgggccgg cacttccgag tataaaaccc cggtaaaccc 9300
aaggaatcac tcacaatcgg attttgacgc tcgctctggt acagttcgat acggtctagt 9360
gaaaccgagg ataacgacga aggtttttcc ccattgatcc aggtcggtgt ttatgattgg 9420
tggaaaaaga ctcgagaaaa gttccatcga agccgttgga aatgtgccgt cttcctgtga 9480
cgtcttgtgg atccagttcc ttgttcacgt ctggtgatcg tgtaaaatgt gctgtcttgt 9540
ggcgtcatat gtgttccaga tccagtgatt acgatccgat gtgatgttga tcccttgtga 9600
acgtcttatc ctgttccgtg tgcaccatgc ataatgtcgt attacgtaag ttctgaagtg 9660
aaacagaaga gtgaattgaa agtttttttа ttcaacatca acctaaatat ggactttact 9720
ttccaagaaa attatgcctg atcaactgtg gatagttaca aaaaaaaaag gtttattaat 9780
taaattttat gattacataa tgtgttgaaa agaacaactg aaattttaga agaagatctt 9840
ttcgtgcatc aggctttgcc aattaattga tgataaatta tcatagcaaa ttaacgtaga 9900
gactaaaagg tatatcgtca aatagggctt cttttgacac tattttggca ttcttgctct 9960
ttgagaactt gcaaccctaa aatgggatct tcatcagcct agtggttaga ttcagcagct 10020
acaaagcaaa accatgctga agggttcgat tcccggtcgt ttcaggatct tttcgtaatt 10080
gaaatatcct tgactaccct aagtatcatt gtgcttgcca tttacgaata tacatattac 10140
gatatacgaa tgagaaaatg acaactttgg aaaataaagc tctcaatgtt tcaataagaa 10200
ataaatacta catcagtatt gaaggctaat aacaattaca gattagaacc tttaaacatc 10260
atttctgcaa caggctggat aaagtacagt tggaggatta aattatgcga ttttgcaatt 10320
ttttccgatt aaattcatat ttattcctgg tttggttttt acaaaaaata tttttacatg 10380
acgtttgacc ccgattccct caactttgat tgttatattt ttttttggac aggttgagtt 10440
tgtgggtttt ttcctagtgt tgctttgctt tatgggctct ggttatttaa aattaaaatt 10500
tgacaatctt actacacact ccgaaaaaat catgcgattt tacgtctttt ggatgcacat 10560
aaaagaagcg agccaaatga ggtgaatttg tgtcacattt taaatacgat ggtgtctgat 10620
tcgggaaatg tcaatgatag tgtcattcaa tcataatgtg aattacgtcc gcagtaattt 10680
tcattatttt taagagtgta ctactattta cactacaaaa attttgatac cccaggggggg 10740

aacgaggtcc cggatgtcca gctggccaga ttgttggcaa cgagccctgt acctattgat 10800
cgagtcacca aagcactcct caagtgtttt aatctcgacc agacggtgga cctcggttgt 10860
tctcattctc ggagggcgat ttcgcaatca ttagtaccaa ccacatgtcg aagtcgggag 10920
atgttataaa attataacca attattcaaa aaatgacatc attcaatttg aacaaacgtt 10980
cgatagaaat tatatatgat ttcacatgat attaaactac gaagaaaatt ttacataagg 11040
aagtggtata aaacgtaata tgcttaataa aaactttaac ccttttggga ggataatatt 11100
cagaagttct gattcagaac catctctcat gttatgttcg ttttttgttg cttgtccttt 11160
atatgccaca tgaacaataa caccaatatc tatcccattt ccaggaccta acggaccttg 11220
aagcggcgcc aaaacgtgtg acgatgatgc tggtaccctg gcggtaagtt gatcaaagga 11280
aacgcaaagt tttcaagaaa aaacaaaact aatttgattt ataacacctt tagaaaccac 11340
catgggcagc cgcctggata agtccaaagt catcaactcc gcgttggagc tgttgaacga 11400
agttggcatt gagggactga cgacccgcaa gttggcgcag aagctgggcg tggagcagcc 11460
cacctctac tggcacgtga agaataagcg ggcgctgctg gatgccctgg ccatcgagat 11520
gctcgaccgc caccacacgc atttttgccc gttggaaggc gagtcctggc aggacttcct 11580
ccgcaataac gccaagtcgt tccgctgcgc tctgctgtcc caccgagacg gtgccaaagt 11640
ccatctcggc acgcgcccga ccgaaaagca atacgagaca ctggagaacc agctcgcgtt 11700
cctgtgccag caaggcttca gcctggaaaa tgctctctac gctctgagcg ccgtcggtca 11760
ctttaccctg ggctgcgtgc tggaggacca agagcatcaa gtcgcaaaag aggagcgcga 11820
gaccccaaca accgattcga tgcccccact gctgcgtcag gcaatcgagc tgttcgatca 11880
tcaaggagcc gagccggcat tcctgttcgg cttggagctg attatctgcg gattggaaaa 11940
gcaactgaaa tgcgagtcgg gctcgggccc cgcgtacagc cgcgcgcgta cgaaaaacaa 12000
ttacgggtct accatcgagg gcctgctcga tctcccggac gacgacgccc ccgaagaggc 12060
ggggctggcg gctccgcgcc tgtcctttct ccccgcggga cacacgcgca gactgtcgac 12120
ggcccccccg accgatgtca gcctgggggа cgagctccac ttagacggcg aggacgtggc 12180
gatggcgcat gccgacgcgc tagacgattt cgatctggac atgttggggg acggggattc 12240
cccgggtccg ggatttaccc cccacgactc cgccccctac ggcgctctgg atatggccga 12300
cttcgagttt gagcagatgt ttaccgatgc ccttggaatt gacgagtacg gtgggtagtt 12360
ctagaattgt ccaccgcaag tgcttctaag ccgatcccga ttgtactgat taccataagc 12420
gacattgcca gtgaaagcga caacagcagc atcaaagtac atttgtcata ctgattcggc 12480
tactaccacc atccggaatc agcttgcatc gaacatcaaa tcacgttatt caatgtatct 12540
gtcatccagc tcagacaagt cggagctttt ccagtcgcga aaatctgcga ctccagcgga 12600
aagcaccgaa ccacagagag gactcgtatg aaagccaggg aagaaaccat cattcacctt 12660
gcagcaaata ggaaaaaaaa cggacatctt caacaaacaa aagcccatgc gctaacttgg 12720
tttaggagtt tagtgtgaca ccatgacccc gctgatgatc tttacttagc acaccataac 12780
caccttttatg cgttcgttca tccaaaatct acaggatatc actgcagccg cgagaagaac 12840
tcgtgaacca tcctgttttc ttttttatta tattcttact tttaacttca aattattttc 12900
agtaataaaa cgtctcaaaa taataagttc ataatgagtt taattttacg gaataagaac 12960
aaccatttaa gttattaaat ccttagattt aatggaatta gattgattat atggaaccca 13020
gacttggtaa aaaataaact ccacgttaaa tttctttctg agacttaaaa ttctttcggg 13080
aaagctggga gcaattctcg caccggtgct agggccgcat agtcgacatt tcgagtttac 13140
cactccctat cagtgataga gaaaagtgaa agtcgagttt accactccct atcagtgata 13200
gagaaaagtg aaagtcgagt ttaccactcc ctatcagtga tagagaaaag tgaaagtcga 13260
gtttaccact ccctatcagt gatagagaaa agtgaaagtc gagtttacca ctccctatca 13320
gtgatagaga aaagtgaaag tcgagtttac cactccctat cagtgataga gaaaagtgaa 13380
agtcgagttt accactccct atcagtgata gagaaaagtg aaagtcgagc tcggtacccg 13440
ggtcgaggta ggcgtgtacg gtgggaggcc tatataagca gagctcgttt agtgaaccgt 13500
cagatcgcct ggagacgcca tccacgctgt tttgacctcc atagaagaca ccgggaccga 13560
tccagcctcc gcggccccga attcgagctc ggtacccggg gatccccgct cgaccaccat 13620
gggcgctctc ctgggcctgc ccgaaagcca aacggagctt gataatctta cagaatacaa 13680

cacggcccac aatcggcgca tctcaatgct gggcatcgat gatgatacca atatgcgaaa 13740
gcaaaacgcc ttgaaacagg gacggcgcac tcgaaatgtc acatttaacg atgaggagat 13800
tgtcatcaat cctgaggatg tggatcctaa tgtgggacgc ttcaggaact tggtacaaac 13860
cactgtggtg cccgccaaga gggctcgctg cgacgtcaac cattagtgat aacgcgtcta 13920
gctagagctg agaacttcag ggtgagtttg gggacccttg attgttcttt ctttttcgct 13980
attgtaaaat tcatgttata tggaggggge aaagttttca gggtgttgtt tagaatggga 14040
agatgtccct tgtatcacca tggaccctca tgataatttt gtttctttca ctttctactc 14100
tgttgacaac cattgtctcc tcttattttc ttttcatttt ctgtaacttt ttcgttaaac 14160
tttagcttgc atttgtaacg aatttttaaa ttcacttttg tttatttgtc agattgtaag 14220
tactttctct aatcactttt ttttcaaggc aatcagggta tattatattg tacttcagca 14280
cagttttaga gaacaattgt tataattaaa tgataaggta gaatatttct gcatataaat 14340
tctggctggc gtggaaatat tcttattggt agaaacaact acaccctggt catcatcctg 14400
cctttctctt tatggttaca atgatataca ctgtttgaga tgaggataaa atactctgag 14460
tccaaaccgg gcccctctgc taaccatgtt catgccttct tctctttcct acagctcctg 14520
ggcaacgtgc tggttgttgt gctgtctcat cattttggca aagaattcac tcctcaggtg 14580
caggctgcct atcagaaggt ggtggctggt gtggccaatg ccctggctca caaataccac 14640
tgagatcttt ttccctctgc caaaaattat ggggacatca tgaagccccc tgagcatctg 14700
acttctggct aataaaggaa atttatttte attgcaatag tgtgttggaa ttttttgtgt 14760
ctctcactcg gaaggacata tgggagggca aatcatttaa aacatcagaa tgagtatttg 14820
gtttagagtt tggcaacata tgcccatagc ggccctagcg gcgcgccata gccc 14874

<210> 107
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 107
tcaataatcg tca 13

<210> 108
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 108
tcatcaaacg tca 13

<210> 109
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 109
ttatcgttaa aca 13

<210> 110
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 110
taaacagtca ata 13

<210> 111
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 111
tacacgatca gca 13

<210> 112
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 112
aatacaaaca aca 13

<210> 113
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 113
tcatcaacaa gca 13

<210> 114
<211> 13
<212> DNA
<213> Drosophila sp.
<400> 114
tctacaaacc aga 13

<210> 115
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 115
acatcgattc aca 13

<210> 116
<211> 13
<212> DNA
<213> Drosophila sp.
<400> 116
cgctcaatca aca 13

<210> 117
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 117
tctaccataa aaa 13

<210> 118
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 118
aaatgaatca aca 13

<210> 119
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 119
acatcgttca acg 13

<210> 120
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 120
tcttgattca cca 13

<210> 121
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 121
tctgcagaca aca 13

<210> 122
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 122
tcttcggtaa tca 13

<210> 123
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 123
tctataaaca ata 13

<210> 124
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 124
taaacaataa ata 13

<210> 125
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 125
taaacaagca aaa 13

<210> 126
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 126
tcaacgatcg gcg 13

<210> 127
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 127
tgatccatca tca 13

<210> 128
<211> 13
<212>
<213> Drosophila sp.
<400> 128
tcaacatgca aga 13

<210> 129
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 129
tcttaaataa aga 13

<210> 130
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 130
tcaaagatct ata 13

<210> 131
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 131
taatgaatta aca 13

<210> 132
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 132
tttaccatca act 13

<210> 133
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 133
taatgaaaca aca 13

<210> 134
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 134
gtttcaatta aaa 13

<210> 135
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 135
tattcaatta taa 13

<210> 136
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 136
tcttcaatcg ttt 13

<210> 137
<211> 13
<212> ADN
<213> Drosophila sp.
<400> 137
tcaacgatcc ttt 13

<210> 138
<211> 13
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự liên ứng bảng 2
<400> 138

tcwwcratca aca 13

<210> 139
<211> 34
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi HSP
<400> 139
caagcaaagt gaacacgtcg ctaagcgaaa gcta 34

<210> 140
<211> 25
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi VP16
<400> 140
gccctcgatg gtagacccgt aattg 25

<210> 141
<211> 24
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi Agexon1F
<400> 141
ggaaaccgag gataacgacg aagg 24

<210> 142
<211> 24
<212> ADN
<213> Nhân tạo
<220>
<223> Đoạn mồi TETRR1
<400> 142
gcggaacgac ttggcgttat tgcg 24

<210> 143
<211> 6243
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3576
<400> 143
cctctacaaa tgtggtatgg ctgattatga tcagttatct agatccggtg gatcttacgg 60
gtcctccacc ttccgctttt tcttgggtcg agatctcagg aacaggtggt ggcggccctc 120
ggtgcgctcg tactgctcca cgatggtgta gtcctcgttg tgggaggtga tgtccagctt 180
ggcgtccacg tagtagtagc cgggcagctg cacgggcttc ttggccatgt agatggactt 240

gaactccacc aggtagtggc cgccgtcctt cagcttcagg gccttgtggg tctcgccctt 300
cagcacgccg tcgcgggggt acaggcgctc ggtggaggcc tcccagccca tggtcttctt 360
ctgcatcacg gggccgtcgg aggggaagtt cacgccgatg aacttcacct tgtagatgaa 420
gcagccgtcc tgcagggagg agtcctgggt cacggtcgcc acgccgccgt cctcgaagtt 480
catcacgcgc tcccacttga agccctcggg gaaggacagc ttcttgtagt cggggatgtc 540
ggcggggtgc ttcacgtaca ccttggagcc gtactggaac tgggggggaca ggatgtccca 600
ggcgaagggc agggggccgc ccttggtcac cttcagcttc acggtgttgt ggccctcgta 660
ggggcggccc tcgccctcgc cctcgatctc gaactcgtgg ccgttcacgg tgccctccat 720
gcgcaccttg aagcgcatga actcggtgat gacgttctcg gaggaggcca tggtggcgac 780
cggtttgcgc ttcttcttgg gtggggtggg atctcccatg gtggcctgaa tctcaacttg 840
cacctggcga tcgcctaaag gtgttataaa tcaaattagt tttgtttttt cttgaaaact 900
ttgcgtttcc tttgatcaac ttaccgccag ggtacctgca gattgtttag cttgttcagc 960
tgcgcttgtt tatttgctta gctttcgctt agcgacgtgt tcactttgct tgtttgaatt 1020
gaattgtcgc tccgtagacg aagcgcctct atttatactc cggcgctgtt taaacatcca 1080
ccatgcgccc gcatcgatct ctatcactga tagggaggtc tctatcactg atagggagtt 1140
ctctatcact gatagggatg tctctatcac tgatagggat ttctctatca ctgataggga 1200
agtctctatc actgataggg acctctctat cactgatagg gaaatctcta tcactgatag 1260
ggatctctct atcactgata gggacttctc tatcactgat agggacgtct ctatcactga 1320
tagggaactc tctatcactg atagggacat ctctatcact gatagggact tctctatcac 1380
tgatagggag tatgttctct ctcttctctt ctctctctct ttctcgaatg ttctctctct 1440
tctcttctct ctctctttct cgatggccgg ggcgcgccag gtttcgactt tcacttttct 1500
ctatcactga tagggagtgg taaactcgac tttcactttt ctctatcact gatagggagt 1560
ggtaaactcg actttcactt ttctctatca ctgataggga gtggtaaact cgactttcac 1620
ttttctctat cactgatagg gagtggtaaa ctcgactttc acttttctct atcactgata 1680
gggagtggta aactcgactt tcacttttct ctatcactga tagggagtgg taaactcgac 1740
tttcactttt ctctatcact gatagggagt ggtaaactcg agcggccgcc accgcggtgg 1800
agctccagct tttgttccct ttagtgaggg ttaattgcgc gcttggcgta atcatggtca 1860
tagctgtttc ctgtgtgaaa ttgttatccg ctcacaattc cacacaacat acgagccgga 1920
agcataaagt gtaaagcctg gggtgcctaa tgagtgagct aactcacatt aattgcgttg 1980
cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc agctgcatta atgaatcggc 2040
caacgcgcgg ggagaggcgg tttgcgtatt gggcgctctt ccgcttcctc gctcactgac 2100
tcgctgcgct cggtcgttcg gctgcggcga gcggtatcag ctcactcaaa ggcggtaata 2160
cggttatcca cagaatcagg ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa 2220
aaggccagga accgtaaaaa ggccgcgttg ctggcgtttt tccataggct ccgccccct 2280
gacgagcatc acaaaaatcg acgctcaagt cagaggtggc gaaacccgac aggactataa 2340
agataccagg cgtttccccc tggaagctcc ctcgtgcgct ctcctgttcc gaccctgccg 2400
cttaccggat acctgtccgc ctttctccct tcgggaagcg tggcgctttc tcatagctca 2460
cgctgtaggt atctcagttc ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa 2520
ccccccgttc agcccgaccg ctgcgccttа tccggtaact atcgtcttga gtccaacccg 2580
gtaagacacg acttatcgcc actggcagca gccactggta acaggattag cagagcgagg 2640
tatgtaggcg gtgctacaga gttcttgaag tggtggccta actacggcta cactagaagg 2700
acagtatttg gtatctgcgc tctgctgaag ccagttacct tcggaaaaag agttggtagc 2760
tcttgatccg gcaaacaaac caccgctggt agcggtggtt tttttgtttg caagcagcag 2820
attacgcgca gaaaaaaagg atctcaagaa gatcctttga tcttttctac ggggtctgac 2880
gctcagtgga acgaaaactc acgttaaggg attttggtca tgagattatc aaaaaggatc 2940
ttcacctaga tcctttttaaa ttaaaaatga agttttaaat caatctaaag tatatatgag 3000
taaacttggt ctgacagtta ccaatgctta atcagtgagg cacctatctc agcgatctgt 3060
ctatttcgtt catccatagt tgcctgactc cccgtcgtgt agataactac gatacgggag 3120
ggcttaccat ctggccccag tgctgcaatg ataccgcgag acccacgctc accggctcca 3180

gatttatcag caataaacca gccagccgga agggccgagc gcagaagtgg tcctgcaact 3240
ttatccgcct ccatccagtc tattaattgt tgccgggaag ctagagtaag tagttcgcca 3300
gttaatagtt tgcgcaacgt tgttgccatt gctacaggca tcgtggtgtc acgctcgtcg 3360
tttggtatgg cttcattcag ctccggttcc caacgatcaa ggcgagttac atgatccccc 3420
atgttgtgca aaaaagcggt tagctccttc ggtcctccga tcgttgtcag aagtaagttg 3480
gccgcagtgt tatcactcat ggttatggca gcactgcata attctcttac tgtcatgcca 3540
tccgtaagat gcttttctgt gactggtgag tactcaacca agtcattctg agaatagtgt 3600
atgcggcgac cgagttgctc ttgcccggcg tcaatacggg ataatacgcg ccacatagc 3660
agaactttaa aagtgctcat cattggaaaa cgttcttcgg ggcgaaaact ctcaaggatc 3720
ttaccgctgt tgagatccag ttcgatgtaa cccactcgtg cacccaactg atcttcagca 3780
tcttttactt tcaccagcgt ttctgggtga gcaaaaacag gaaggcaaaa tgccgcaaaa 3840
aagggaataa gggcgacacg gaaatgttga atactcatac tcttcctttt tcaatattat 3900
tgaagcattt atcagggtta ttgtctcatg agcggatacа tatttgaatg tatttagaaa 3960
aataaacaaa taggggttcc gcgcacattt ccccgaaaag tgccacctaa attgtaagcg 4020
ttaatatttt gttaaaattc gcgttaaatt tttgttaaat cagctcattt tttaaccaat 4080
aggccgaaat cggcaaaatc ccttataaat caaaagaata gaccgagata gggttgagtg 4140
ttgttccagt ttggaacaag agtccactat taaagaacgt ggactccaac gtcaaagggc 4200
gaaaaaccgt ctatcagggc gatggcccac tacgtgaacc atcaccctaa tcaagttttt 4260
tggggtcgag gtgccgtaaa gcactaaatc ggaaccctaa agggagcccc cgatttagag 4320
cttgacgggg aaagccggcg aacgtggcga gaaaggaagg gaagaaagcg aaaggagcgg 4380
gcgctagggc gctggcaagt gtagcggtca cgctgcgcgt aaccaccaca cccgccgcgc 4440
ttaatgcgcc gctacagggc gcgtcccatt cgccattcag gctgcgcaac tgttgggaag 4500
ggcgatcggt gcgggcctct tcgctattac gccagctggc gaaaggggga tgtgctgcaa 4560
ggcgattaag ttgggtaacg ccagggtttt cccagtcacg acgttgtaaa acgacggcca 4620
gtgagcgcgc gtaatacgac tcactatagg gcgaattggg taccgggccc ccctcgagg 4680
tcgacgatgt aggtcacggt ctcgaagccg cggtgcgggt gccagggcgt gccccttgggc 4740
tccccgggcg cgtactccac ctcacccatc tggtccatca tgatgaacgg gtcgaggtgg 4800
cggtagttga tcccggcgaa cgcgcggcgc accgggaagc cctcgccctc gaaaccgctg 4860
ggcgcggtgg tcacggtgag cacgggacgt gcgacggcgt cggcgggtgc ggatacgcgg 4920
ggcagcgtca gcgggttctc gacggtcacg gcgggcatgt cgacggtatc gataagcttg 4980
ggcccccccc cgaggttccc acaatggtta attcgagctc gccccggggat ctaattcaat 5040
tagagactaa ttcaattaga gctaattcaa ttaggatcca agcttatcga tttcgaaccc 5100
tcgaccgccg gagtataaat agaggcgctt cgtctacgga gcgacaattc aattcaaaca 5160
agcaaagtga acacgtcgct aagcgaaagc taagcaaata aacaagcgca gctgaacaag 5220
ctaaacaatc ggggtaccgc tagagtcgat cccaccccac ccaagaagaa gcgcaaaccg 5280
gtaccatggc ctcctccgag aacgtcatca ccgagttcat gcgcttcaag gtgcgcatgg 5340
agggcaccgt gaacggccac gagttcgaga tcgagggcga gggcgagggc cgcccctacg 5400
agggccacaa caccgtgaag ctgaaggtga ccaagggcgg ccccctgccc ttcgcctggg 5460
acatcctgtc cccccagttc cagtacggct ccaaggtgta cgtgaagcac cccgccgaca 5520
tccccgacta caagaagctg tccttccccg agggcttcaa gtgggagcgc gtgatgaact 5580
tcgaggacgg cggcgtggcg accgtgaccc aggactcctc cctgcaggac ggcgtgttca 5640
tctacaaggt gaagttcatc ggcgtgaact tcccctccga cggccccgtg atgcagaaga 5700
agaccatggg ctgggaggcc tccaccgagc gcctgtaccc ccgcgacggc gtgctgaagg 5760
gcgagaccca caaggccctg aagctgaagg acggcggcca ctacctggtg gagttcaagt 5820
ccatctacat ggccaagaag cccgtgcagc tgcccggcta ctactacgtg gacgccaagc 5880
tggacatcac ctcccacaac gaggactaca ccatcgtgga gcagtacgag cgcaccgagg 5940
gccgccacca cctgttcctg tgatgatcat aatcagccat accacatttg tagaggtttt 6000
acttgcttta aaaaacctcc cacacctccc cctgaacctg aaacataaaa tgaatgcaat 6060
tgttgttgtt aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac 6120

aaatttcaca aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat 6180
caatgtatct taacgcgagt taattaaggc cgctcattta tcagcgcttt aaatttgcgc 6240
atg 6243

<210> 144
<211> 5746
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3582
<400> 144
cgcctaaagg tgttataaat caaattagtt ttgtttttc ttgaaaactt tgcgtttcct 60
ttgatcaact taccgccagg gtacctgcag attgtttagc ttgttcagct gcgcttgttt 120
atttgcttag ctttcgctta gcgacgtgtt cactttgctt gtttgaattg aattgtcgct 180
ccgtagacga agcgcctcta tttatactcc ggcgctgttt aaacatccac catgcgcccg 240
catcgatctc tatcactgat agggaggtct ctatcactga tagggagttc tctatcactg 300
atagggatgt ctctatcact gatagggatt tctctatcac tgatagggaa gtctctatca 360
ctgataggga cctctctatc actgataggg aaatctctat cactgatagg gatctctcta 420
tcactgatag ggacttctct atcactgata gggacgtctc tatcactgat agggaactct 480
ctatcactga tagggacatc tctatcactg atagggactt ctctatcact gatagggagt 540
atgttctctc tcttctcttc tctctctctt tctcgaatgt tctctctctt ctcttctctc 600
tctctttctc gatggccggg gcgcgccagg tttcgacttt cacttttctc tatcactgat 660
agggagtggt aaactcgact ttcacttttc tctatcactg atagggagtg gtaaactcga 720
ctttcacttt tctctatcac tgatagggag tggtaaactc gactttcact tttctctatc 780
actgataggg agtggtaaac tcgactttca cttttctcta tcactgatag ggagtggtaa 840
actcgacttt cacttttctc tatcactgat agggagtggt aaactcgact ttcacttttc 900
tctatcactg atagggagtg gtaaactcga gcggccgcca ccgcggtgga gctccagctt 960
ttgttccctt tagtgagggt taattgcgcg cttggcgtaa tcatggtcat agctgtttcc 1020
tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagtg 1080
taaagcctgg ggtgcctaat gagtgagcta actcacatta attgcgttgc gctcactgcc 1140
cgctttccag tcgggaaacc tgtcgtgcca gctgcattaa tgaatcggcc aacgcgcggg 1200
gagaggcggt ttgcgtattg ggcgctcttc cgcttcctcg ctcactgact cgctgcgctc 1260
ggtcgttcgg ctgcggcgag cggtatcagc tcactcaaag gcggtaatac ggttatccac 1320
agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa 1380
ccgtaaaaag gccgcgttgc tggcgttttt ccataggctc cgcccccctg acgagcatca 1440
caaaaatcga cgctcaagtc agaggtggcg aaacccgaca ggactataaa gataccaggc 1500
gtttcccccc ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata 1560
cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta 1620
tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt gtgcacgaac cccccgttca 1680
gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag tccaacccgg taagacacga 1740
cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg 1800
tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg 1860
tatctgcgct ctgctgaagc cagttacctt cggaaaaaga gttggtagct cttgatccgg 1920
caaacaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag 1980
aaaaaaagga tctcaagaag atcctttgat cttttctacg gggtctgacg ctcagtggaa 2040
cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat 2100
cctttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc 2160
tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc 2220
atccatagtt gcctgactcc ccgtcgtgta gataactacg atacgggagg gcttaccatc 2280

tggccccagt gctgcaatga taccgcgaga cccacgctca ccggctccag atttatcagc 2340
aataaaccag ccagccggaa gggccgagcg cagaagtggt cctgcaactt tatccgcctc 2400
catccagtct attaattgtt gccgggaagc tagagtaagt agttcgccag ttaatagttt 2460
gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca cgctcgtcgt ttggtatggc 2520
ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatcccccа tgttgtgcaa 2580
aaaagcggtt agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt 2640
atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg 2700
cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgta tgcggcgacc 2760
gagttgctct tgcccggcgt caatacggga taataccgcg ccacatagca gaactttaaa 2820
agtgctcatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt 2880
gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt 2940
caccagcgtt tctgggtgag caaaaacagg aaggcaaaat gccgcaaaaa agggaataag 3000
ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta 3060
tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat 3120
aggggttccg cgcacatttc cccgaaaagt gccacctaaa ttgtaagcgt taatattttg 3180
ttaaaattcg cgttaaattt ttgttaaatc agctcatttt ttaaccaata ggccgaaatc 3240
ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 3300
tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc 3360
tatcagggcg atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg 3420
tgccgtaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga 3480
aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcg 3540
ctggcaagtg tagcggtcac gctgcgcgta accaccacac ccgccgcgct taatgcgccg 3600
ctacagggcg cgtcccattc gccattcagg ctgcgcaact gttgggaagg gcgatcggtg 3660
cgggcctctt cgctattacg ccagctggcg aaaggggggat gtgctgcaag gcgattaagt 3720
tgggtaacgc cagggttttc ccagtcacga cgttgtaaaa cgacggccag tgagcgcgcg 3780
taatacgact cactataggg cgaattgggt accgggcccc cccctcgaggt cgacgatgta 3840
ggtcacggtc tcgaagccgc ggtgcgggtg ccagggcgtg cccttgggct ccccgggcgc 3900
gtactccacc tcacccatct ggtccatcat gatgaacggg tcgaggtggc ggtagttgat 3960
cccggcgaac gcgcggcgca ccgggaagcc ctcgccctcg aaaccgctgg gcgcggtggt 4020
cacggtgagc acgggacgtg cgacggcgtc ggcgggtgcg gatacgcggg gcagcgtcag 4080
cgggttctcg acggtcacgg cgggcatgtc gacggtatcg ataagcttgg gccccccctc 4140
gaggttccca caatggttaa ttcgagctcg cccggggatc taattcaatt agagactaat 4200
tcaattagag ctaattcaat taggatccaa gcttatcgat ttcgaaccct cgaccgccgg 4260
agtataaata gaggcgcttc gtctacggag cgacaattca attcaaacaa gcaaagtgaa 4320
cacgtcgcta agcgaaagct aagcaaataa acaagcgcag ctgaacaagc taaacaatcg 4380
gggtaccgct agagtcgatc ccaccccacc caagaagaag cgcaaaccgg taccatggcc 4440
tcctccgaga acgtcatcac cgagttcatg cgcttcaagg tgcgcatgga gggcaccgtg 4500
aacggccacg agttcgagat cgagggcgag ggcgagggcc gccctacga gggccacaac 4560
accgtgaagc tgaaggtgac caagggcggc ccctgccct tcgcctggga catcctgtcc 4620
ccccagttcc agtacggctc caaggtgtac gtgaagcacc ccgccgacat ccccgactac 4680
aagaagctgt ccttccccga gggcttcaag tgggagcgcg tgatgaactt cgaggacggc 4740
ggcgtggcga ccgtgaccca ggactcctcc ctgcaggacg gctgcttcat ctacaaggtg 4800
aagttcatcg gcgtgaactt ccctccgac ggccccgtga tgcagaagaa gaccatgggc 4860
tgggaggcct ccaccgagcg cctgtacccc cgcgacggcg tgctgaaggg cgagacccac 4920
aaggccctga agctgaagga cggcggccac tacctggtgg agttcaagtc catctacatg 4980
gccaagaagc ccgtgcagct gccccggctac tactacgtgg acgccaagct ggacatcacc 5040
tcccacaacg aggactacac catcgtggag cagtacgagc gcaccgaggg ccgccaccac 5100
ctgttcctgt gatgatcata atcagccata ccacatttgt agaggtttta cttgctttaa 5160
aaaacctccc acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgtta 5220

acttgtttat tgcagcttat aatggttaca ataaaagcaa tagcatcaca aatttcacaa 5280
ataaagcatt ttttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt 5340
aacgcgagtt aattaaggcc gctcatttat cagcgcttta aatttgcgca tgctagttta 5400
atacaccttt cgcagcaagt agtatagctt gttgcacagc agacatcggg aacgggttgg 5460
gttatttttct tgagcgtgac ggaacagaat ctcatgaaag gcctgcacca gatggtagcg 5520
gttgtggtga aggctgactt gcgtcatcgt cggagtcagt ggaggagttg gtggaattga 5580
ctccgttgga cttgttggcg acggtggtgg cgaactgaat tggttctgat tttgctgttg 5640
ttgcattaaa atctgctgct gctgttgcat catttgcaac tgatactgct tctcgatttc 5700
atcatcgatg gcgggaatgt agaatgcgat tgccatggtg ggcgat 5746

<210> 145
<211> 15121
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3596
<400> 145
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt taccttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatcccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gacaccggtg ctacccgcca tactcatcga tgcccagcgc 960
gtcggtgaac atttgctcga actcgaagtc ggccatgtcc agggcgccgt acggggcgct 1020
atcgtggggc gtgaagcccg gtcccgggct atctccatcg cccagcatat ccaggtcgaa 1080
atcgtccagg gcgtcggcgt gggccattgc cacatcctct ccatccaggt gcagctcgtc 1140
gcccaggctc acatcggtcg gcggggcggt gctcaggcgg cgcgtgtgtc cggcgggcag 1200
gaagctcagg cgggggcgcg ccaggccggc ttcctccggg gcatcgtcat ccggcaggtc 1260
cagcagtccc tcgatggtgc tgccatagtt gttcttggta cgggcgcggc tgtaggcgct 1320
gccgctctcg cacttcagct gcttttccag gccgcagatg atcagctcca ggccgaacag 1380
gaaggccggc tcggcgccct ggtgatcgaa cagctcgatg gcctggcgca gcagcggcgg 1440
catgctatcg gtggtcgggg tctcgcgctc ctccttggcc acctggtgct cctgatcctc 1500
cagcacacag cccagggtga agtggcccac ggcgctcagg gcgtacaggg cgttctccag 1560
gctgaagccc tgctggcaca ggaaggccag ctggttctcc agggtctcgt actgcttctc 1620
ggtcgggcgg gtgccccaggt gcaccttggc gccatcgcgg tgcgacagca gggcgcagcg 1680
gaagctcttg gcgttgttgc gcaggaaatc ctgccagctc tcgccctcca gcgggcagaa 1740
gtgggtgtgg tggcgatcca gcatttcgat ggccagggcg tccagcaggg cgcgcttgtt 1800
cttcacgtgc cagtacaggg tcggctgttc cacgcccagc ttctgggcca gcttgcgggt 1860
ggtcaggccc tcgataccaa cttcgttcag cagctccagg gcgctgttga tcaccttgct 1920

cttgtccagg cggctgacct gtgaatacgg ttaatgtcac tattagtgat ttataaaaat 1980
aaatttgatt tatatatcaa caatttttca tcgcagcctt cagctttttg ttgaataatt 2040
ataatgatat tttttacgat tcaaatcatt taattgttac tcaacgaaat aagtttaatt 2100
caaattttaa aacaagatta tatattaaga ttagaataag aaagaacttt gttagattat 2160
ttaattaaaa agattaaaat ttaagtctcc agtcactatt taaagatcat ctttcaaacg 2220
ttaaagtgaa ttcaaacgag acgttcaaat ttcgattaaa cagtaattaa ctctaaattt 2280
ctatcacgaa ttaagttatt gaatatgaag gtttatattt atttacatca tctaataggt 2340
ttgagttgat tgttgtaatc cgcatgtgcc agaagatatc aatttccaaa ttgtccgagt 2400
tcatggaatg ttgattgttg tttgtgttgc tttgtaattg ttgcagggag tatttatggt 2460
ttgttgattg tagtataagg ctgtttctaa aggctagaaa ataatttttat ttatttgaaa 2520
ataagtaaat atacataata ttactaacaa taggtcgtcc tatttttga tattctgcac 2580
aaatttttaa aacacaaaga ttgcaatact tttagacact aatactgcac actctgaaaa 2640
attattaaat tatttttaaa aacttacctt aatactttag agaaaaatat tataccgcac 2700
ctttctactt tatactcact ttattatacc agttgcatgt tgattgtagt tctttgacaa 2760
gaaaatattc catattgctc caaattatct tggtaagttg attggtgcgt catttgagca 2820
agctaacacc ttgtctcatt taagttcgcc tcaagatctc atagcatttt taaatatcac 2880
tatatttagt aagtaattag aattaccatg gtggtttgct agcggtacct gcagattgtt 2940
tagcttgttc agctgcgctt gtttatttgc ttagctttcg cttagcgacg tgttcacttt 3000
gcttgtttga attgaattgt cgctccgtag acgaagcgcc tctatttata ctccggcgct 3060
cggtccgact ctctatcact gatagggagt attgtcctct ctatcactga tagggaatgc 3120
tatgttctct atcactgata gggaagttga tagtctctat cactgatagg gagtggtaat 3180
ttctctatca ctgataggga ttagtgatgt ctctatcact gatagggatt ggaatattct 3240
ctatcactga tagggagtgg taatatctct atcactgata gggactggag ttttctctat 3300
cactgatagg gacacgctga ctctctatca ctgataggga taagcttact ctctatcact 3360
gatagggagt attgtcctct ctatcactga tagggaatgc tatgttctct atcactgata 3420
gggaagttga tagtctctat cactgatagg gagtggtaat ttctctatca ctgataggga 3480
ttagtgatgt ctctatcact gatagggatt ggaatattct ctatcactga tagggagtgg 3540
taatatctct atcactgata gggactggag ttttctctat cactgatagg gacacgctga 3600
ctctctatca ctgataggga taagcggccg catggtaccc attgcttgtc atttattaat 3660
ttggatgatg tcatttgttt ttaaaattga actggcttta cgagtagaat tctacgcgta 3720
aaacacaatc aagtatgagt cataatctga tgtcatgttt tgtacacggc tcataaccga 3780
actggcttta cgagtagaat tctacttgta atgcacgatc agtggatgat gtcatttgtt 3840
tttcaaatcg agatgatgtc atgttttgca cacggctcat aaactcgctt tacgagtaga 3900
attctacgtg taacgcacga tcgattgatg agtcatttgt tttgcaatat gatatcatac 3960
aatatgactc atttgttttt caaaaccgaa cttgatttac gggtagaatt ctacttgtaa 4020
agcacaatca aaaagatgat gtcatttgtt tttcaaaact gaactcgctt tacgagtaga 4080
attctacgtg taaaacacaa tcaagaaatg atgtcatttg ttataaaaat aaaagctgat 4140
gtcatgtttt gcacatggct cataactaaa ctcgctttac gggtagaatt ctacgcgtaa 4200
aacatgattg ataattaaat aattcatttg caagctatac gttaaatcaa acggacgctc 4260
gaggttgcac aacactatta tcgatttgca gttcgggaca taaatgttta aatatatcga 4320
tgtctttgtg atgcgcgcga catttttgta ggttattgat aaaatgaacg gatacgttgc 4380
ccgacattat cattaaatcc ttggcgtaga atttgtcggg tccattgtcc gtgtgcgcta 4440
gcatgcccgt aacggacctc gtacttttgg cttcaaaggt tttgcgcaca gacaaaatgt 4500
gccacacttg cagctctgca tgtgtgcgcg ttaccacaaa tcccaacggc gcagtgtact 4560
tgttgtatgc aaataaatct cgataaaggc gcggcgcgcg aatgcagctg atcacgtacg 4620
ctcctcgtgt tccgttcaag gacggtgtta tcgacctcag attaatgttt atcggccgac 4680
tgttttcgta tccgctcacc aaacgcgttt ttgcattaac attgtatgtc ggcggatgtt 4740
ctatatctaa tttgaataaa taaacgataa ccgcgttggt tttagagggc ataataaaag 4800
aaatattgtt atcgtgttcg ccattagggc agtataaatt gacgttcatg ttggatattg 4860

tttcagttgc aagttgacac tggcggcgac aagcaattct aattggggta agttttcccg 4920
ttcttttctg ggttcttccc ttttgctcat ccttgctgca ctaccttcag gtgcaagttg 4980
agattcaggc caccatggga gatcccaccc cacccaagaa gaagcgcaaa ccggtcgcca 5040
ccatggagag cgacgagagc ggcctgcccg ccatggagat cgagtgccgc atcaccggca 5100
ccctgaacgg cgtggagttc gagctggtgg gcggcggaga gggcaccccc gagcagggcc 5160
gcatgaccaa caagatgaag agcaccaaag gcgccctgac cttcagcccc tacctgctga 5220
gccacgtgat gggctacggc ttctaccact tcggcaccta ccccagcggc tacgagaacc 5280
ccttcctgca cgccatcaac aacggcggct acaccaacac ccgcatcgag aagtacgagg 5340
acggcggcgt gctgcacgtg agcttcagct accgctacga ggccggccgc gtgatcggcg 5400
acttcaaggt gatgggcacc ggcttccccg aggacagcgt gatcttcacc gacaagatca 5460
tccgcagcaa cgccaccgtg gagcacctgc accccatggg cgataacgat ctggatggca 5520
gcttcacccg caccttcagc ctgcgcgacg gcggctacta cagctccgtg gtggacagcc 5580
acatgcactt caagagcgcc atccacccca gcatcctgca gaacgggggc cccatgttcg 5640
ccttccgccg cgtggaggag gatcacagca acaccgagct gggcatcgtg gagtaccagc 5700
acgccttcaa gaccccggat gcagatgccg gtgaagaaag atctcgaccc aagaaaaagc 5760
ggaaggtgga ggacccgtct ggaggcggtg gatccggcgg tggaggcatg cagatctttg 5820
tgaagacttt gaccggaaag accatcaccc tcgaggtaga gccatcggac accattgaga 5880
atgtaaaggc caagattcag gataaggagg gaatcccccc agatcagcag cgtctgatct 5940
tcgctggtaa ttttaaaagc atattttttt ctttgaaatt cataagttat caattatcga 6000
tggaaatgta ttctatggag aacgttttac ccgatgaatg ggtgcaaaaa ttattttacc 6060
ttcaaatcta caatcaacac acgctaactt ttgtgacttg atcaactctc acctggaaaa 6120
gcaaccaact acaatcaaca ttctatggga taatcgacaa gtgagtaaaa ttatagccgg 6180
acctcttagt acagtgtatt taaaagggga ataatattct atcaatagga ataaaaataa 6240
ggtcagcagc catgactttt ccatcatttt gaatatacct tatttgtttc gggattaatt 6300
gggggtcgga aatcctcttg aattcagaaa cgggaaccgg aggaaggtgc cggtctttca 6360
gaaagctgtg aaaaatacca acatttctgc tgccaagagc tcaataagaa gtttcaaaaa 6420
ttgtcttgga tgttgcagct gtggctgcta agtaataaga catctattag tatctagatt 6480
tgttagacca tttaacatag tgttttaaac gatggggtta atagatgagg gttaagaagc 6540
tagttatatt actgttgctg taacgccttc aattgtcggt tacagagcaa acattattga 6600
atgttaatgt aaagagttta tttgttttct agtaaacata tagcgattgg ttagtaatca 6660
ctaatagaaa tttttcataa gtatcaaaaa agtaaacctc tttttcagtc tatgtaataa 6720
gtaaaccaag gaaagggaaa atatctacaa tcaacaagcc attgttgcag caacaaagca 6780
actgaaacta caatcaacat tcaataaact tgggtaattt ggaatttaat tctctgggac 6840
acctgtggat tacaacaatc aactcgaaac ttattataca atgtaaataa aaattgatat 6900
gcatacatga agatcaagtg aaattccatt tagaatcaat ttttttcgaa tattaagttt 6960
cttgctttaa tttatctgaa agtaaataga cattccaaat tcaagttaac aaattaataa 7020
tgaattgact agtgattttt aagagaaaaa gataagattt aaaaaaggaa agcctttctt 7080
gataaatttt tgaaccactt tatgccgttt caatcataaa aacttttaag aacacatgac 7140
tggtaaaatt aatttaaaac aaatttaaat tttcaacgta acattcaaca aaaatggtga 7200
aaactatcac ggaaattgtt aatattaata tgtcccaaaa atagcctttg tatgtatatg 7260
atactaatcc atacatctat ggtatctata ggtcgccaac tggaagacgg acgcaccctg 7320
tccgattaca acatccagaa ggagtccacc cttcacttgg tccttcgtct ccgcggtggc 7380
atgcagatcg gggatcccac cccacccaag aagaagcgca aaccggtcgc caccatggcc 7440
tcctccgaga acgtcatcac cgagttcatg cgcttcaagg tgcgcatgga gggcaccgtg 7500
aacggccacg agttcgagat cgagggcgag ggcgagggcc gccccctacga gggccacaac 7560
accgtgaagc tgaaggtgac caagggcggc cccctgccct tcgcctggga catcctgtcc 7620
ccccagttcc agtacggctc caaggtgtac gtgaagcacc ccgccgacat ccccgactac 7680
aagaagctgt ccttccccga gggcttcaag tgggagcgcg tgatgaactt cgaggacggc 7740
ggcgtggcga ccgtgaccca ggactcctcc ctgcaggacg gctgcttcat ctacaaggtg 7800

aagttcatcg gcgtgaactt cccctccgac ggccccgtga tgcagaagaa gaccatgggc 7860
tgggaggcct ccaccgagcg cctgtacccc cgcgacggcg tgctgaaggg cgagacccac 7920
aaggccctga agctgaagga cggcggccac tacctggtgg agttcaagtc catctacatg 7980
gccaagaagc ccgtgcagct gcccggctac tactacgtgg acgccaagct ggacatcacc 8040
tcccacaacg aggactacac catcgtggag cagtacgagc gcaccgaggg ccgccaccac 8100
ctgttcctga gatctcgacc caagaaaaag cggaaggtgg aggacccgta agatccaccg 8160
gatctagata actgatcata atcagccata ccacatttgt agaggtttta cttgctttaa 8220
aaaacctccc acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgtta 8280
acttgtttat tgcagcttat aatggttaca ataaaagcaa tagcatcaca aatttcacaa 8340
ataaagcatt tttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt 8400
aacgcgagtt aattaacacc gaaatcgtaa ttcacggcat cattacaaaa tattttgacg 8460
ttttggacct cgtccctaat gacaccataa cggtggcctt gaagtatatt taaccctaga 8520
aagatagtct gcgtaaaatt gacgcatgca ttcttgaaat attgctctct ctttctaaat 8580
agcgcgaatc cgtcgctgtg catttaggac atctcagtcg ccgcttggag ctcccgtgag 8640
gcgtgcttgt caatgcggta agtgtcactg attttgaact ataacgaccg cgtgagtcaa 8700
aatgacgcat gattatcttt tacgtgactt ttaagattta actcatacga taattatatt 8760
gttatttcat gttctactta cgtgataact tattatatat atattttctt gttatagata 8820
tcgtgactaa tatataataa aatgggtagt tctttagacg atgagcatat cctctctgct 8880
cttctgcaaa gcgatgacga gcttgttggt gaggattctg acagtgaaat atcagatcac 8940
gtaagtgaag atgacgtcca ggaaatctgg ccggccgcaa ccattgtggg aaccgtgcga 9000
tcaaacaaac gcgagatacc ggaagtactg aaaaacagtc gctccaggcc agtgggaaca 9060
tcgatgtttt gttttgacgg accccttact ctcgtctcat ataaaccgaa gccagctaag 9120
atggtatact tattatcatc ttgtgatgag gatgcttcta tcaacgaaag taccggtaaa 9180
ccgcaaatgg ttatgtatta taatcaaact aaaggcggag tggacacgct agaccaaatg 9240
tgttctgtga tgacctgcag taggaagacg aataggtggc ctatggcatt attgtacgga 9300
atgataaaca ttgcctgcat aaattctttt attatataca gccataatgt cagtagcaag 9360
ggagaaaagg tccaaagtcg caaaaaattt atgagaaacc tttacatgag cctgacgtca 9420
tcgtttatgc gtaagcgttt agaagctcct actttgaaga gatatttgcg cgataatatc 9480
tctaatattt tgccaaatga agtgcctggt acatcagatg acagtactga agagccagta 9540
atgaaaaaac gtacttactg tacttactgc ccctctaaaa taaggcgaaa ggcaaatgca 9600
tcgtgcaaaa aatgcaaaaa agttatttgt cgagagcata atattgatat gtgccaaagt 9660
tgtttctgac tgactaataa gtataatttg tttctattat gtataagtta agctaattac 9720
ttattttata atacaacatg actgttttta aagtacaaaa taagtttatt tttgtaaaag 9780
agagaatgtt taaaagtttt gttactttat agaagaaatt ttgagttttt gttttttttt 9840
aataaataaa taaacataaa taaattgttt gttgaattta ttattagtat gtaagtgtaa 9900
atataataaa acttaatatc tattcaaatt aataaataaa cctcgatata cagaccgata 9960
aaacacatgc gtcaatttta cgcatgatta tctttaacgt acgtcacaat atgattatct 10020
ttctagggtt aaataatagt ttctaatttt tttattattc agcctgctgt cgtgaatacc 10080
gtatatctca acgctgtctg tgagattgtc gtattctagc ctttttagtt tttcgctcat 10140
cgacttgata ttgtccgaca cattttcgtc gatttgcgtt ttgatcaaag acttgagcag 10200
agacacgtta atcaactgtt caaattgatc catattaacg atatcaaccc gatgcgtata 10260
tggtgcgtaa aatatatttt ttaaccctct tatactttgc actctgcgtt aatacgcgtt 10320
cgtgtacaga cgtaatcatg ttttcttttt tggataaaac tcctactgag tttgacctca 10380
tattagaccc tcacaagttg caaaacgtgg catttttttac caatgaagaa tttaaagtta 10440
ttttaaaaaa tttcatcaca gatttaaaga agaaccaaaa attaaattat ttcaacagtt 10500
taatcgacca gttaatcaac gtgtacacag acgcgtcggc aaaaaacacg cagcccgacg 10560
tgttggctaa aattattaaa tcaacttgtg ttatagtcac ggatttgccg tccaacgtgt 10620
tcctcaaaaa gttgaagacc aacaagttta cggacactat taattatttg attttgcccc 10680
acttcatttt gtgggatcac aattttgtta tattttaaac aaagcttggc actggccgtc 10740

gttttacaac gtcgtgactg ggaaaaccct ggcgttaccc aacttaatcg ccttgcagca 10800
catccccctt tcgccagctg gcgtaatagc gaagaggccc gcaccgatcg ccccttcccaa 10860
cagttgcgca gcctgaatgg cgaatggcgc ctgatgcggt attttctcct tacgcatctg 10920
tgcggtattt cacaccgcat atggtgcact ctcagtacaa tctgctctga tgccgcatag 10980
ttaagccagc cccgacaccc gccaacaccc gctgacgcgc cctgacgggc ttgtctgctc 11040
ccggcatccg cttacagaca agctgtgacc gtctccgggg gctgcatgtg tcagaggttt 11100
tcaccgtcat caccgaaacg cgcgagacga aagggcctcg tgatacgcct atttttatag 11160
gttaatgtca tgataataat ggtttcttag acgtcaggtg gcacttttcg gggaaatgtg 11220
cgcggaaccc ctatttgttt atttttctaa atacattcaa atatgtatcc gctcatgaga 11280
caataaccct gataaatgct tcaataatat tgaaaaagga agagtatgag tattcaacat 11340
ttccgtgtcg cccttattcc ctttttttgcg gcattttgcc ttcctgtttt tgctcaccca 11400
gaaacgctgg tgaaagtaaa agatgctgaa gatcagttgg gtgcacgagt gggttacatc 11460
gaactggatc tcaacagcgg taagatcctt gagagtttte gccccgaaga acgttttcca 11520
atgatgagca cttttaaagt tctgctatgt ggcgcggtat tatcccgtat tgacgccggg 11580
caagagcaac tcggtcgccg catacactat tctcagaatg acttggttga gtactcacca 11640
gtcacagaaa agcatcttac ggatggcatg acagtaagag aattatgcag tgctgccata 11700
accatgagtg ataacactgc ggccaactta cttctgacaa cgatcggagg accgaaggag 11760
ctaaccgctt ttttgcacaa catgggggat catgtaactc gccttgatcg ttgggaaccg 11820
gagctgaatg aagccatacc aaacgacgag cgtgacacca cgatgcctgt agcaatggca 11880
acaacgttgc gcaaactatt aactggcgaa ctacttactc tagcttcccg gcaacaatta 11940
atagactgga tggaggcgga taaagttgca ggaccacttc tgcgctcggc ccttccggct 12000
ggctggttta ttgctgataa atctggagcc ggtgagcgtg ggtctcgcgg tatcattgca 12060
gcactggggc cagatggtaa gccctcccgt atcgtagtta tctacacgac ggggagtcag 12120
gcaactatgg atgaacgaaa tagacagatc gctgagatag gtgcctcact gattaagcat 12180
tggtaactgt cagaccaagt ttactcatat atactttaga ttgatttaaa acttcatttt 12240
taatttaaaa ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa 12300
cgtgagtttt cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga 12360
gatccttttt ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg 12420
gtggtttgtt tgccggatca agagctacca actcttttte cgaaggtaac tggcttcagc 12480
agagcgcaga taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag 12540
aactctgtag caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc 12600
agtggcgata agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg 12660
cagcggtcgg gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac 12720
accgaactga gatacctaca gcgtgagcat tgagaaagcg ccacgcttcc cgaagggaga 12780
aaggcggaca ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt 12840
ccaggggaa acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag 12900
cgtcgatttt tgtgatgctc gtcaggggg cggagcctat ggaaaaacgc cagcaacgcg 12960
gccttttac ggttcctggc cttttgctgg ccttttgctc acatgttctt tcctgcgtta 13020
tcccctgatt ctgtggataa ccgtattacc gcctttgagt gagctgatac cgctcgccgc 13080
agccgaacga ccgagcgcag cgagtcagtg agcgaggaag cggaagagcg cccaatacgc 13140
aaaccgcctc tccccgcgcg ttggccgatt cattaatgca gctggcacga caggtttccc 13200
gactggaaag cgggcagtga gcgcaacgca attaatgtga gttagctcac tcattaggca 13260
ccccaggctt tacactttat gcttccggct cgtatgttgt gtggaattgt gagcggataa 13320
caatttcaca caggaaacag ctatgaccat gattacgaat ttcgacctgc aggcatgcaa 13380
gcttgcatgc ctgcaggtcg acgctcgcgc gacttggttt gccattcttt agcgcgcgtc 13440
gcgtcacaca gcttggccac aatgtggttt ttgtcaaacg aagattctat gacgtgttta 13500
aagtttaggt cgagtaaagc gcaaatcttt tttaacccta gaaagatagt ctgcgtaaaa 13560
ttgacgcatg cattcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 13620
tgcatttagg acatctcagt cgccgcttgg agctcccgtg aggcgtgctt gtcaatgcgg 13680

taagtgtcac tgattttgaa ctataacgac cgcgtgagtc aaaatgacgc atgattatct 13740
tttacgtgac ttttaagatt taactcatac gataattata ttgttatttc atgttctact 13800
tacgtgataa cttattatat atatattttc ttgttataga tatcgtgact aatatataat 13860
aaaatgggta gttctttaga cgatgagcat atcctctctg ctcttctgca aagcgatgac 13920
gagcttgttg gtgaggattc tgacagtgaa atatcagatc acgtaagtga agatgacgtc 13980
cagagcgata cagaagaagc gtttatagat gaggtacatg aagtgcagcc aacgtcaagc 14040
ggtagtgaaa tattagacga acaaaatgtt attgaacaac caggttcttc attggcttct 14100
aacagaatct tgaccttgcc acagaggact attagaggta agaataaaca ttgttggtca 14160
acttcaaagt ccacgaggcg tagccgagtc tctgcactga acattgtcag atcggcccgg 14220
cggagtggac acgctagacc aaatgtgttc tgtgatgacc tgcagtagga agacgaatag 14280
gtggcctatg gcattattgt acggaatgat aaacattgcc tgcataaatt cttttattat 14340
atacagccat aatgtcagta gcaagggaga aaaggtccaa agtcgcaaaa aatttatgag 14400
aaacctttac atgagcctga cgtcatcgtt tatgcgtaag cgtttagaag ctcctacttt 14460
gaagagatat ttgcgcgata atatctctaa tattttgcca aatgaagtgc ctggtacatc 14520
agatgacagt actgaagagc cagtaatgaa aaaacgtact tactgtactt actgccccta 14580
taaaataagg cgaaaggcaa atgcatcgtg caaaaaatgc aaaaaagtta tttgtcgaga 14640
gcataatatt gatatgtgcc aaagttgttt ctgactgact aataagtata atttgtttct 14700
attatgtata agttaagcta attacttatt ttataataca acatgactgt ttttaaagta 14760
caaaataagt ttatttttgt aaaagagaga atgtttaaaa gttttgttac tttatagaag 14820
aaattttgag tttttgtttt tttttaataa ataaataaac ataaataaat tgtttgttga 14880
atttattatt agtatgtaag tgtaaatata ataaaactta atatctattc aaattaataa 14940
ataaacctcg atatacagac cgataaaaca catgcgtcaa ttttacgcat gattatcttt 15000
aacgtacgtc acaatatgat tatctttcta gggttaaaat gaatgtaagc actttattaa 15060
cgaaatcttt gggaatattt cgctcatcag cattttattt gagcaggagt ccgagatgcc 15120
c 15121

<210> 146
<211> 533
<212> ADN
<213> Nhân tạo
<220>
<223> Mảnh 146 PBW dsx (Fig 6)
<400> 146
gtccaatcga tcacaatgta tcacaacgtt gcgaattcag tttcacaatc acacgcaaca 60
aacrcgrcac gttacaaatt agttactttg aatcgatcga ttatgatgcg gccgactcar 120
cggccccccgg cagcactaac cagtagtgat ttccactttg cagtgaccgg accaaaactt 180
cgaaattcga attgtaaagt gacagttcat ttcccgccaa gtgttgtgcc agtgtcatgt 240
cgatatttat tttatttttct tttttgtagg aaaatgctga gcgaaattaa taatataagt 300
ggtgtactat cgtcgtccat gaagttattt tgcgaatgat actttgtttt gtatgtgctg 360
tgtgttgtgt ggacttttgc tgtgcgttgc tgtttgcgat ggaaggacta ttgtgtcgtc 420
gccacgctgg actattcggt gagtggtaga ataatatttt atctatttca tcgcggtaca 480
attgactttt tattactact cactgctatg gaggaatctc aggaacatcg taa 533

<210> 147
<211> 611
<212> ADN
<213> Nhân tạo
<220>
<223> Mảnh Bombyx-dsx (Fig 6)

<400> 147
gcgattattt aattctatat atttttcaaa ttcagtttct attccactaa caatgtacac 60
tacacgtaca catacacaca acaaaatagt gaatcgataa attagtgtgt cataacacat 120
taacaacatt gttacacacc cacacataca aatttgctaa gttgatagtc gaataatcgg 180
aatggttcgc atcacactac taaccagtcg tgatttccac tttacagtga ccggacgaag 240
gtggagaaat tcgaaattta aatataaaag tgacaattcg aatttccacg cgcgcgctct 300
agtgatgtgc cagtgtgtga atatcaatat tatttttat tttctttttt gtaggaaaat 360
gctggaaatt aataatataa gtggtgtact gtcttcgtca atgaagttat tttgcgaatg 420
atacttagtt ttacaagtgc cgtggtgtgt gttgacactt gctgtgcgat gctgtgcgaa 480
tttcaacgga aatatttgtt gtcgtaacat tggatctatg ggtaagttta gtataataac 540
tttactctgt tcacattagt gaaacataca tttgtaaaat ttgtgtttta ctaatgtgaa 600
atttatttt g 611

<210> 148
<211> 570
<212> ADN
<213> Nhân tạo
<220>
<223> mảnh codling-dsx (Fig 6)
<400> 148
ttacaaacaa tgtacggagc tacaacgttg caagttcggt ccccacacaa cacaatgtgt 60
cataacacat taacaacatt gttacacacc cacacataca aatttgctaa gttgataaaa 120
gagtggtgtg tccgacgaat cagaacatca ctaacccagt cgtgatttca tttccacagt 180
gaccggacga aggtggagaa gttcgaaatt taaaaaaagt gaccacattt tatttaatag 240
tgatgtgcaa gtgatactat ttttatttg ttttttcttt gtaggaaaat gctgagcgaa 300
ataaataatt ttagtggtgt gctatcgtca tcgatgaagt tgttttgcga atgatactat 360
gttcttcaag tgctgtgttt tgtggactgt ggggtgactg ttcctgtaaa taagcttcgt 420
tggacattgt gtctcacaca tcggatctca tggtaagtgc tagtgctagc atyrmaactt 480
aactctctga gcgaattcct ttgactctaa agtcacacgr acagccatac aatcaaagct 540
acgctctaat tttaagatga cawtctgtaa 570

<210> 149
<211> 4389
<212> ADN
<213> Nhân tạo
<220>
<223> Cấu trúc DSX Minigenel rom LA3491
<400> 149
acgacgaact tgtcaaacga tctcaatggc tcctggagaa gctgcgatac ccctgggaga 60
tgatgcccct gatgtacgtg atactgaaag gcgccgacgg agacgtcaat aaagcgcgcc 120
aacggattga cgaaggtatg ggggttctta ccggttggga ctgtttccga ggtatcgatc 180
gggtgtcact cacttcctgg gtgctcccat tttgtaactg ctaacgctta ttattgagtt 240
tcaggacatc tgggatcttc ggtcgacgga gtctattccc aacagtgccc tggatcaaac 300
actgccatca tgcagtttcc gtagcctgtt gggctacgct ccccgacttg acatccccca 360
ttcttatcaa acaacaactc aaggcctgag acaacgagtg gtggaatttg cgcacgaagt 420
cattggtttg tcctggtaaa agttaaaagg gttaactgga gggttaattg acacggtttc 480
aactgatggc cttattgaca cacggatgaa agacttgcac gcttgacctt ctgtctgtac 540
taataaaagt tacgttggct gggttttggg gtcataatgg ccccaaaatc gaatcgtcat 600
aacttcttga aatacaactc acgtttaaga ccattcaaga gtattagatc atcgtctata 660

atagcagatt tgaaatttac ttcacatttc ggtattgcag tgcccttgc ttccacaatg 720
gaattaggtc ggtggtgcgt cgatcgtcgc aagtttatcg ttaaacagtc aataaaatga 780
gcattttata tcgtgataca tatgagaaga tagaggtttc aattaaaaca aatccacatg 840
gtgtcgctaa taaaattgtg cattttaagc gagttatatc ctctgatcaa gataaaatag 900
aaaattcgat ttttgaatat tcaattataa gagcctgaat aactacaaca tgtagtgaat 960
cgaaactgat ttatgacggt ttgtgaaggt tacacgtcct aagcatttgg attcaagaaa 1020
agcaagagat atgacgaatg taaactttat cgtatcaatg aagtaactag cgtccagaac 1080
agtacaaacc aacatcgtac cgtcgtattc cactccggtc gttgcaatat ctctaggtcc 1140
accgaaaaac actcatgacc aagatcgtgt cgtcgatctt ggtccaccga aacaccgatg 1200
tccatatcgt ttcgtcgaac ttggaccaac gattcatgca actgatgaca acgcggcccc 1260
cgggtcgtac caatatccga aaaatccaac tgttcttctc tgcctcgcag gtcaagccgt 1320
ggtcaatgaa tactcacgat tgcacaatct gaacatgttc gacggtgtag agttgcgcag 1380
tacgacgcgc cagtccggat gatagacttt ttacacgatc agcacgaccc actgcgctgc 1440
ggcaaaggtc gaaccgaaac aagaataaac cacgaagatc agatcgattc gacggaagaa 1500
gcaatcgaat gcaaagaaga atcggaacga agaaaactct aaagcatcgc atatttacaa 1560
agcataacgg aaaacccgca agttcaaact agtgattagt gtaagatgaa gcaaagcaga 1620
aatgtagtat ctagattttt cgacgttagt ttacaaagat aaaaaatgag gttggacata 1680
caatcgtggg tattcgtctg agttcgtcac aactgcaccg gaaactgtga aacagaatag 1740
agccaacctg tgcgcggaga atgttgaggt cattataagc ttccttagca tccacgggtg 1800
aaagtcgatc gacggaagcc tgcaagactc tgtcgatggg ctttcgtcct agaagaataa 1860
gattaaacct gaaatgtatt ctcccgtgga atggtttcat ttgagtaatt ctgtatcttc 1920
tccttcccaa ttccacgaac gcgacgaact ctaatacaaa caacataatg accacagtgc 1980
aaatgctgtt taacgataat agcgacatgc agccattctg gggctaccac gtgtagctct 2040
acttgtgaga cagcgttcct aaagagtgtg aaagtgcaaa caagtgatga aaccaatagt 2100
gcaaagcaag tttagaggga aaatttaaaa aatgcaaaac agcagtagta cttaactttt 2160
aagattgtgt ttcgaaagcc gaagtgaggc tgttccatct gccaccggaa aaaaacgacg 2220
acagcagaat catcaacaag caacatccat ccgaaaaaat ccgggaaacc ggatcttcaa 2280
ccaaccatcc tacaatctac aaaccagaga ttatatctct tcaatcgttt ccgacatcgg 2340
tcggtttcgg tgcccaaaat gatctgataa acacttatct ctctgtagct tgcatgccat 2400
tgcgagcgta ttttggtagc tggccgttgc caaacggctc cgacaggtac tgctattgga 2460
ggttgtgcac gaccacgttg agtttgcctt ttgagttgga gagtgtgtct tttcgtcata 2520
tattcggcct tttcaagggt gattttcagg ctacgtaatg attgtatagt ttaaccagct 2580
aaaacatatt gatgacaagt tctatttcag caccacaaac aagcctgtta atgtctctca 2640
ccgcaaccat tgttctgcgc gcgttataat cagcatagaa gtttattttc tttgggatga 2700
ttcaaatatt acgtgacgca aagtttgcca attttagaac ccctccctcc tccacgtaac 2760
ggcttttgtg tgaaaaattt aaattttgtg tatagaccgt agcatttcgg aagacccccc 2820
cccttactct gttgagttac gtaaaatttc aacgatcctt ttgtagttct gaattttata 2880
tcagcgtgca gtgttatgaa gatatccaca gtataaaata ttatttttatt ttaaattcta 2940
tgctgattat caatgtgtta ctagtggctt ttcatactca tgttgcgagc tcgatttggc 3000
gcacggtact tatcaaggca tgtatgtatg ttgtttgaag caactgtata actgtttgaa 3060
actatctaat tggtgagctc gtttcattta gtatataata atgataattg ctatggagac 3120
gttatttact agcaagtgat ttgacgacct gaaatcggaa caaatagaca acgtttttat 3180
aaatacaata aatcagaact ttccattatt gggtacaaag agttgcgcta tttcgatact 3240
gtcagatcag attttccagc acaacgatac cttgatatgc gataacttag aattagacct 3300
tcaaatccat ctctccagct atgaacagtc atatagataa agccaatggc gttatgaggt 3360
agcggaaagc gtcatctttc caatgctatc taagtacata atttgctata gctttctatt 3420
aatcgtagtt tgagagatgc aaagtcagtt atctcgtatc aaggtttgat tgttttggaa 3480
attagctaaa cagttgacat tatcacccgt ctttagggga taagcgcata caaatgtgta 3540
tttagttgtt cattgaagta acgtaagata ggcaagtatg gaaacgagct caccaaacgt 3600

cgaaatacgt ctaataaatt tgtgttcagc aggatggttc aaaatttatt tgcatcacct 3660
caaaattaca gtacctagtg ctgtttgtga caaacatcaa aaggtaaaat caaactcgtg 3720
gcgtcgtgca atctccatag aatgaacaat ttctaaccgt atttgatgga aagacattga 3780
gtctactatc ctcttaacag cattgcactt gtctataaac aataaataat ttgttctttt 3840
ttacattttc tttccccact ttcgcccccc ccccccccaa aaatcaatcc ctcaaacagg 3900
atacgacatt tgttgcatct actttccgaa gcgttccagc agacacagac actggccgga 3960
cgaggagaac atctccgtca cccgcactcc gtctgcgtca cggtcgccat gtgccgattt 4020
tcgtacccgg tcacagtcca gctcgccgga taacaacggt ggcgcgctca atctggacac 4080
gaaatctacc aaagcgacga ccgccaccac cgacgacgaa gaggttatgt acgagaaacg 4140
cagcccgaag tccattgaat ctaccgagtt gcggtgccgt ctggaggaag ccttacacag 4200
tggcgctgct gctgctgcgg ctgctgaaga acctctggcg ggcggaagcg gttcccactg 4260
gaagagagaa agtttcggct ctacggagga gattcccact cgacccgctc acagtgaacc 4320
ggaagataat ggatttgaaa acggattgga agcgcaccag tcccatattc tgcacagcat 4380
acatcggaa 4389

<210> 150
<211> 2572
<212> ADN
<213> Nhân tạo
<220>
<223> DSX Minigene2 từ cấu trúc LA3534
<400> 150
ctcgatttcc cctcgtttcc aatttcagac gacgaacttg tcaaacgatc tcaatggctc 60
ctggagaagc tgcgataccc ctgggagatg atgcccctga tgtacgtgat actgaaaggc 120
gccgacggag acgtcaataa agcgcgccaa cggattgacg aaggtatggg ggttcttacc 180
ggttgggact gtttccgagg tatcgatcgg gtgtcactca cttcctgggt gctcccattt 240
tgtaactgct aacgcttatt attgagtttc aggacatctg ggatcttcgg tcgacggagt 300
ctattcccaa cagtgccctg gatcaaacac tgccatcatg cagtttccgt agcctgttgg 360
gctacgctcc ccgacttgac atcccccatt cttatcaaac aacaactcaa ggcctgagac 420
aacgagtggt ggaatttgcg cacgaagtca ttggtttgtc ctggtaaaag ttaaaagggt 480
taactggagg gttaattgac acggtttcaa ctgatggcct tattgacaca cggatgaaag 540
acttgcacgc ttgaccttct gtctgtacta ataaaagtta cgttggctgg gttttggggt 600
cataatggcc ccaaaatcga atcgtcataa cttcttgaaa tacaactcac gtttaagacc 660
attcaagagt attagatcat cgtctataat agcagatttg aaatttactt cacatttcgg 720
tattgcagtg cccttgctt ccacaatgga attagttaaa gtttcgagag cattgtcaat 780
atcaagtgtt gttagcaaac aaatgctaac atcaagatta ctatcgatgt ttgattcaca 840
tgtattccaa tcagctcgta aaaaatggaa agtggagctg atagggttga ggtctcacgt 900
gctccaaatc atcacctcca agttagttct aatacactcc gttatatgaa atatggtggt 960
gcgtcgatcg tcgcaagttt atcgttaaac agtcaataaa atgagcattt tatatcgtga 1020
tacatatgag aagatagagg tttcaattaa aacaaatcca catggtgtcg ctaataaaat 1080
tgtgcatttt aagcgagtta tatcctctga tcaagataaa atagaaaatt cgatttttga 1140
atattcaatt ataagagcct gaataactac aacatgtagt gaatcgaaac tgatttatga 1200
cggtttgtga aggttacacg tcctaagcat ttggattcaa gaaaagcaag agatatgacg 1260
aatgtaaact ttatcgtatc aatgaagtaa ctagcgtcca gaacagtaca aaccaacatc 1320
gtaccgtcgt attccactcc ggtcgttgca atatctctag gtccaccgaa aaacactcat 1380
gaccaagatc gtgtcgtcga tcttggtcca ccgaaacacc gatgtccata tcgtttcgtc 1440
gaacttggac caacgattca tgcaactgat gacaacgcgg cccccgggtc gtaccaatat 1500
ccgaaaaatc caactgttct tctctgcctc gcaggtcaag ccgtggtcaa tgaatactca 1560
cgattgcaca atctgaacat gttcgacggt gtagagttgc gcagtacgac gcgccagtcc 1620

ggatgataga ctttttacac gatcagcacg acccactgcg ctgcggcaaa ggtcgaaccg 1680
aaacaagaat aaaccacgaa gatcagatcg attcgacgga agaagcaatc gaatgcaaag 1740
aagaatcgga atgaagaaaa ctctaaagca tcgcatattt acaaagcata acggaaaacc 1800
cgcaagttca aactagtgat tagtgtaaga tgaagcaaag cagaaatgta gtatctagat 1860
ttttcgacgt tagtttacaa agataagaaa tgaggttgga catacaatcg tgggtattcg 1920
tctgagttcg tcacaactgc accggaaact gtgaaacaga atagagccaa cctgtgcgcg 1980
gagaatgttg aggtcattat aagcttcctt agcatccacg ggtgaaagtc gatcgacgga 2040
agcctgcaag actctgtcga tgggctttcg tcctagaaga ataagattaa acctgaaatg 2100
tattctcccg tggaatggtt tcatttgagt aattctgtat cttctccttc ccaattccac 2160
gaacgcgacg aactctaata caaacaacat aatgaccaca gtgcaaatgc tgtttaacga 2220
taatagcgac atgcagccat tctggggcta ccacgtgtag ctctacttgt gagacagcgt 2280
tcctaaagag tgtgaaagtg caaacaagtg atgaaaccaa tagtgcaaag caagtttaga 2340
gggaaaattt aaaaaatgca aaacagcagt agtacttaac ttttaagatt gtgtttcgaa 2400
agccgaagtg tgttccatct gccaccggaa aaaaacgacg acagcagaat catcaacaag 2460
caacatccat ccgaaaaaat ccgggaaacc ggatcttcaa ccaaccatcc tacaatctac 2520
aaaccagaga ttatatctct tcaatcgttt ccgacatcgg tcggtttcgg tg 2572

<210> 151
<211> 18790
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit nguyên vẹn LA3619
<400> 151
cgcgcctaag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat 60
gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata 120
aacaagttaa caacaacaat tgcattcatt ttatgtttca ggttcagggg gaggtgtggg 180
aggttttta aagcaagtaa aacctctaca aatgtggtat ggctgattat gatcgttgca 240
cattccgatg tatgctgtgc agaatatggg actggtgcgc ttccaatccg ttttcaaatc 300
cattatcttc cggttcactg tgagcgggtc gagtgggaat ctcctccgta gagccgaaac 360
tttctctctt ccagtgggaa ccgcttccgc ccgccagagg ttcttcagca gccgcagcag 420
cagcagcgcc actgtgtaag gcttcctcca gacggcaccg caactcggta gattcaatgg 480
acttcgggct gcgtttctcg tacataacct cttcgtcgtc ggtggtggcg gtcgtcgctt 540
tggtagattt cgtgtccaga ttgagcgcgc caccgttgtt atccggcgag ctggactgtg 600
accgggtacg aaaatcggca catggcgacc gtgacgcaga cggagtgcgg gtgacggaga 660
tgttctcctc gtccggccag tgtctgtgtc tgctggaacg cttcggaaag tagatgcaac 720
aaatgtcgta tcctgtttga gggattgatt tttgggggggg gggggggcga aagtggggaa 780
agaaaatgta aaaaagaaca aattatttat tgtttataga caagtgcaat gctgttaaga 840
ggatagtaga ctcaatgtct ttccatcaaa tacggttaga aattgttcat tctatggaga 900
ttgcacgacg ccacgagttt gattttacct tttgatgttt gtcacaaaca gcactaggta 960
ctgtaatttt gaggtgatgc aaataaattt tgaaccatcc tgctgaacac aaatttatta 1020
gacgtatttc gacgtttggt gagctcgttt ccatacttgc ctatcttacg ttacttcaat 1080
gaacaactaa atacacattt gtatgcgctt atcccctaaa gacgggtgat aatgtcaact 1140
gtttagctaa tttccaaaac aatcaaacct tgatacgaga taactgactt tgcatctctc 1200
aaactacgat taatagaaag ctatagcaaa ttatgtactt agatagcatt ggaaagatga 1260
cgctttccgc tacctcataa cgccattggc tttatctata tgactgttca tagctggaga 1320
gatggatttg aaggtctaat tctaagttat cgcatatcaa ggtatcgttg tgctggaaaa 1380
tctgatctga cagtatcgaa atagcgcaac tctttgtacc caataatgga aagttctgat 1440
ttattgtatt tataaaaacg ttgtctattt gttccgattt caggtcgtca aatcacttgc 1500

tagtaaataa cgtctccata gcaattatca ttattatata ctaaatgaaa cgagctcacc 156(
aattagatag tttcaaacag ttatacagtt gcttcaaaca acatacatac atgccttgat 162(
aagtaccgtg cgccaaatcg agctcgcaac atgagtatga aaagccacta gtaacacatt 168(
gataatcagc atagaattta aaataaaata atattttata ctgtggatat cttcataaca 174(
ctgcacgctg atataaaatt cagaactaca aaaggatcgt tgaaatttta cgtaactcaa 180(
cagagtaagg gagggggtct tccgaaatgc tacggtctat acacaaaatt taaatttttc 186(
acacaaaagc cgttacgtgg aggagggagg ggttctaaaa ttggcaaact ttgcgtcacg 192(
taatatttga atcatcccaa agaaaataaa cttctatgct gattataacg cgcgcagaac 198(
aatggttgcg gtgagagaca ttaacaggct tgtttgtggt gctgaaatag aacttgtcat 204(
caatatgttt tagctggtta aactatacaa tcattacgta gcctgaaaat caccccttgaa 210(
aaggccgaat atatgacgaa aagacacact ctccaactca aaaggcaaac tcaacgtggt 2160
cgtgcacaac ctccaatagc agtacctgtc ggagccgttt ggcaacggcc agctaccaaa 2220
atacgctcgc aatggcatgc aagctacaga gagataagtg tttatcagat cattttgggc 2280
accgaaaccg accgatgtcg gaaacgattg aagagatata atctctggtt tgtagattgt 2340
aggatggttg gttgaagatc cggtttcccg gatttttttcg gatggatgtt gcttgttgat 2400
gattctgctg tcgtcgtttt tttccggtgg cagatggaac agcctcactt cggctttcga 2460
aacacaatct taaaagttaa gtactactgc tgttttgcat tttttaaatt ttccctctaa 2520
acttgctttg cactattggt ttcatcactt gtttgcactt tcacactctt taggaacgct 2580
gtctcacaag tagagcttgc ggtggacaat caccggtgtt agccgccgta ctcatcgatg 2640
cccagggcgt cggtgaacat ctgctcgaac tcgaaatcgg ccatatccag ggcgccgtag 2700
ggggcgctat cgtgcggggt gaatcccggt cccgggctat cgccatcgcc cagcatgtcc 2760
aggtcgaagt cgtccagggc atcggcgtgg gccatcgcca catcctcgcc atccaggtgc 2820
agctcatcgc ccaggctcac gtcggtcggc ggggcggtcg acaggcggcg ggtgtgtccg 2880
gccggcagga agctcaggcg cggggcggcc aggcccgcct cctccggggc atcatcatcc 2940
ggcagatcca gcaggccctc gatggtgctg ccgtagttgt tcttggtgcg ggcgcggctg 3000
taggcggggc ccgagcccga ctcgcatttc agttgctttt ccaatccgca gataatcagc 3060
tccaagccga acaggaatgc cggctcggct ccttgatgat cgaacagctc gattgcctga 3120
cgcagcagtg ggggcatcga atcggttgtt ggggtctcgc gctcctcttt tgcgacttga 3180
tgctcttggt cctccagcac gcagcccagg gtaaagtgac cgacggcgct cagagcgtag 3240
agagcatttt ccaggctgaa gccttgctgg cacaggaacg cgagctggtt ctccagtgtc 3300
tcgtattgct tttcggtcgg gcgcgtgccg agatggactt tggcaccgtc tcggtgggac 3360
agcagagcgc agcggaacga cttggcgtta ttgcggagga agtcctgcca ggactcgcct 3420
tccaacgggc aaaaatgcgt gtggtggcgg tcgagcatct cgatggccag ggcatccagc 3480
agcgcccgct tattcttcac gtgccagtag agggtgggct gctccacgcc cagcttctgc 3540
gccaacttgc gggtcgtcag tccctcaatg ccaacttcgt tcaacagctc caacgcggag 3600
ttgatgactt tggacttatc caggcggctg cccatggtgg ttttccagtg gcgccgcttc 3660
acgtggtagc cccagaatgg ctgcatgtcg ctattatcgt taaacagcat ttgcactgtg 3720
gtcattatgt tgtttgtatt agagttcgtc gcgttcgtgg aattgggaag gagaagatac 3780
agaattactc aaatgaaacc attccacggg agaatacatt tcaggtttaa tcttattctt 3840
ctaggacgaa agcccatcga cagagtcttg caggcttccg tcgatcgact ttcacccgtg 3900
gatgctaagg aagcttataa tgacctcaac attctccgcg cacaggttgg ctctattctg 3960
tttcacagtt tccggtgcag ttgtgacgaa ctcagacgaa tacccacgat tgtatgtcca 4020
acctcatttt ttatctttgt aaactaacgt cgaaaaatct agatactaca tttctgcttt 4080
gcttcatctt acactaatca ctagtttgaa cttgcgggtt ttccgttatg ctttgtaaat 4140
atgcgatgct ttagagtttt cttcgttccg attcttcttt gcattcgatt gcttcttccg 4200
tcgaatcgat ctgatcttcg tggtttattc ttgtttcggt tcgaccttg ccgcagcgca 4260
gtgggtcgtg ctgatcgtgt aaaaagtcta tcatccggac tggcgcgtcg tactgcgcaa 4320
ctctacaccg tcgaacatgt tcagattgtg caatcgtgag tattcattga ccacggcttg 4380
acctgcgagg cagagaagaa cagttggatt tttcggatat tggtacgacc cgggggccgc 4440

gttgtcatca gttgcatgaa tcgttggtcc aagttcgacg aaacgatatg gacatcggtg 4500
tttcggtgga ccaagatcga cgacacgatc ttggtcatga gtgtttttcg gtggacctag 4560
agatattgca acgaccggag tggaatacga cggtacgatg ttggtttgta ctgttctgga 4620
cgctagttac ttcattgata cgataaagtt tacattcgtc atatctcttg cttttcttga 4680
atccaaatgc ttaggacgtg taaccttcac aaaccgtcat aaatcagttt cgattcacta 4740
catgttgtag ttattcaggc tcttataatt gaatattcaa aaatcgaatt ttctatttta 4800
tcttgatcag aggatataac tcgcttaaaa tgcacaattt tattagcgac accatgtgga 4860
tttgttttaa ttgaaacctc tatcttctca tatgtatcac gatataaaat gctcatttta 4920
ttgactgttt aacgataaac ttgcgacgat cgacgcacca ccgacctaat tccattgtgg 4980
aagcaagggg cactgcaata ccgaaatgtg aagtaaattt caaatctgct attatagacg 5040
atgatctaat actcttgaat ggtcttaaac gtgagttgta tttcaagaag ttatgacgat 5100
tcgattttgg ggccattatg accccaaaac ccagccaacg taacttttat tagtacagac 5160
agaaggtcaa gcgtgcaagt ctttcatccg tgtgtcaata aggccatcag ttgaaaccgt 5220
gtcaattaac cctccagtta acccttttaa cttttaccag gacaaaccaa tgacttcgtg 5280
cgcaaattcc accactcgtt gtctcaggcc ttgagttgtt gtttgataag aatgggggat 5340
gtcaagtcgg ggagcgtagc ccaacaggct acggaaactg catgatggca gtgtttgatc 5400
cagggcactg ttgggaatag actccgtcga ccgaagatcc cagatgtcct gaaactcaat 5460
aataagcgtt agcagttaca aaatgggagc acccaggaag tgagtgacac ccgatcgata 5520
cctcggaaac agtcccaacc ggtaagaacc cccatacctt cgtcaatccg ttggcgcgct 5580
ttattgacgt ctccgtcggc gcctttcagt atcacgtaca tcaggggcac cacctcctag 5640
ggcagattgt ttagcttgtt cagctgcgct tgtttatttg cttagctttc gcttagcgac 5700
gtgttcactt tgcttgtttg aattgaattg tcgctccgta gacgaagcgc ctctatttat 5760
actccggcgc tcgttttcga gtttaccact ccctatcagt gatagagaaa agtgaaagtc 5820
gagtttacca ctccctatca gtgatagaga aaagtgaaag tcgagtttac cactccctat 5880
cagtgataga gaaaagtgaa agtcgagttt accactccct atcagtgata gagaaaagtg 5940
aaagtcgagt ttaccactcc ctatcagtga tagagaaaag tgaaagtcga gtttaccact 6000
ccctatcagt gatagagaaa agtgaaagtc gagtttacca ctccctatca gtgatagaga 6060
aaagtgaaag tcgaaacctg gcgcgccccg gccatcgaga aagagagaga gaagagaaga 6120
gagagaacat tcgagaaaga gagagagaag agaagagaga gaacatactc cctatcagtg 6180
atagagaagt ccctatcagt gatagagatg tccctatcag tgatagagag ttccctatca 6240
gtgatagaga cgtccctatc agtgatagag aagtccctat cagtgataga gagatcccta 6300
tcagtgatag agatttccct atcagtgata gagaggtccc tatcagtgat agagacttcc 6360
ctatcagtga tagagaaatc cctatcagtg atagagacat ccctatcagt gatagagaac 6420
tccctatcag tgatagagac ctccctatca gtgatagaga tcgatgcggc cgcgagcgcc 6480
ggagtataaa tagaggcgct tcgtctacgg agcgacaatt caattcaaac aagcaaagtg 6540
aacacgtcgc taagcgaaag ctaagcaaat aaacaagcgc agctgaacaa gctaaacaat 6600
ctgcaggtac cctggcggta agttgatcaa aggaaacgca aagttttcaa gaaaaaacaa 6660
aactaatttg atttataaca cctttagaaa gcggggctag ccaccatggg cagcgcctac 6720
agccgcgccc gtaccaagaa caactatggc agcaccatcg agggactgct ggacctgccg 6780
gatgacgatg ccccggagga agccggcctg gccgcccccc gcctgagctt cctgcccgcc 6840
ggacacacgc gccgcctgag caccgccccg ccgaccgatg tgagcctggg cgacgagctg 6900
cacctggatg gagaggatgt ggcaatggcc cacgccgacg ccctggacga tttcgacctg 6960
gatatgctgg gcgatggaga tagcccggga ccgggcttca cgccccacga tagcgccccg 7020
tacggcgccc tggacatggc cgacttcgag ttcgagcaaa tgttcaccga cgcgctgggc 7080
atcgatgagt atggcgggta ggtttaaact cgcgttaaga tacattgatg agtttggaca 7140
aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc 7200
tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt 7260
tatgtttcag gttcaggggg aggtgtggga ggtttttttaa agcaagtaaa acctctacaa 7320
atgtggtatg gctgattatg atcagttatc tagatccggt ggatcttacg ggtcctccac 7380

cttccgcttt ttcttgggtc gagatctcag gaacaggtgg tggcggccct cggtgcgctc 7440
gtactgctcc acgatggtgt agtcctcgtt gtgggaggtg atgtccagct tggcgtccac 7500
gtagtagtag ccgggcagct gcacgggctt cttggccatg tagatggact tgaactccac 7560
caggtagtgg ccgccgtcct tcagcttcag ggccttgtgg gtctcgccct tcagcacgcc 7620
gtcgcggggg tacaggcgct cggtggaggc ctcccagccc atggtcttct tctgcatcac 7680
ggggccgtcg gaggggaagt tcacgccgat gaacttcacc ttgtagatga agcagccgtc 7740
ctgcagggag gagtcctggg tcacggtcgc cacgccgccg tcctcgaagt tcatcacgcg 7800
ctcccacttg aagccctcgg ggaaggacag cttcttgtag tcggggatgt cggcggggtg 7860
cttcacgtac accttggagc cgtactggaa ctgggggggac aggatgtccc aggcgaaggg 7920
caggggccg ccttggtca ccttcagctt cacggtgttg tggccctcgt aggggcggcc 7980
ctcgccctcg ccctcgatct cgaactcgtg gccgttcacg gtgccctcca tgcgcacctt 8040
gaagcgcatg aactcggtga tgacgttctc ggaggaggcc atggtggcga ccggtttgcg 8100
cttcttcttg ggtggggtgg gatctcccat ggtggcctga atctcaactt gcacctgaag 8160
gtagtgcagc aaggatgagc aaaagggaag aacccagaaa agaacgggaa aacttacccc 8220
aattagaatt gcttgtcgcc gccagtgtca acttgcaact gaaacaatat ccaacatgaa 8280
cgtcaattta tactgcccta atggcgaaca cgataacaat atttctttta ttatgccctc 8340
taaaaccaac gcggttatcg tttatttatt caaattagat atagaacatc cgccgacata 8400
caatgttaat gcaaaaacgc gtttggtgag cggatacgaa aacagtcggc cgataaacat 8460
taatctgagg tcgataacac cgtccttgaa cggaacacga ggagcgtacg tgatcagctg 8520
cattcgcgcg ccgcgccttt atcgagattt atttgcatac aacaagtaca ctgcgccgtt 8580
gggatttgtg gtaacgcgca cacatgcaga gctgcaagtg tggcacattt tgtctgtgcg 8640
caaaaccttt gaagccaaaa gtacgaggtc cgttacgggc atgctactag cgcacacgga 8700
caatggaccc gacaaattct acgccaagga tttaatgata atgtcgggca acgtatccgt 8760
tcattttatc aataacctac aaaaatgtcg cgcgcatcac aaagacatcg atatatttaa 8820
acatttatgt cccgaactgc aaatcgataa tagtgttgtg caacctcgag cgtccgtttg 8880
atttaacgta tagcttgcaa atgaattatt taattatcaa tcatgtttta cgcgtagaat 8940
tctacccgta aagcgagttt agttatgagc catgtgcaaa acatgacatc agcttttatt 9000
tttataacaa atgacatcat ttcttgattg tgttttacac gtagaattct actcgtaaag 9060
cgagttcagt tttgaaaaac aaatgacatc atctttttga ttgtgcttta caagtagaat 9120
tctacccgta aatcaagttc ggttttgaaa aacaaatgag tcatattgta tgatatcata 9180
ttgcaaaaca aatgactcat caatcgatcg tgcgttacac gtagaattct actcgtaaag 9240
cgagtttatg agccgtgtgc aaaacatgac atcatctcga tttgaaaaac aaatgacatc 9300
atccactgat cgtgcattac aagtagaatt ctactcgtaa agccagttcg gttatgagcc 9360
gtgtacaaaa catgacatca gattatgact catacttgat tgtgttttac gcgtagaatt 9420
ctactcgtaa agccagttca attttaaaaa caaatgacat catccaaatt aataaatgac 9480
aagcaatggg taccatgcgg cctggcctcg cgctcgcgcg actgacggtc gtaagcaccc 9540
gcgtacgtgt ccaccccggt cacaaccccct tgtgtcatgt cggcgaccct acgcccccaa 9600
ctgagagaac tcaaaggtta ccccagttgg ggcactactc ccgaaaaccg cttctgacct 9660
gggaaaacgt gaagccccgg ggcatccgct gagggttgcc gccggggctt cggtgtgtcc 9720
gtcagtactt aattaacacc gaaatcgtaa ttcacggcat cattacaaaa tattttgacg 9780
ttttggacct cgtccctaat gacaccataa cggtggcctt gaagtatatt taaccctaga 9840
aagatagtct gcgtaaaatt gacgcatgca ttcttgaaat attgctctct ctttctaaat 9900
agcgcgaatc cgtcgctgtg catttaggac atctcagtcg ccgcttggag ctcccgtgag 9960
gcgtgcttgt caatgcggta agtgtcactg attttgaact ataacgaccg cgtgagtcaa 10020
aatgacgcat gattatcttt tacgtgactt ttaagattta actcatacga taattatatt 10080
gttatttcat gttctactta cgtgataact tattatatat atattttctt gttatagata 10140
tcgtgactaa tatataataa aatgggtagt tctttagacg atgagcatat cctctctgct 10200
cttctgcaaa gcgatgacga gcttgttggt gaggattctg acagtgaaat atcagatcac 10260
gtaagtgaag atgacctcga ggatccaagc ttatcgattt cgaaccctcg accgccggag 10320

tataaataga ggcgcttcgt ctacggagcg acaattcaat tcaaacaagc aaagtgaaca 10380
cgtcgctaag cgaaagctaa gcaaataaac aagcgcagct gaacaagcta aacaatcggg 10440
gtaccgctag agtcgatccc accccaccca agaagaagcg caaaccggta ccatggcctc 10500
ctccgagaac gtcatcaccg agttcatgcg cttcaaggtg cgcatggagg gcaccgtgaa 10560
cggccacgag ttcgagatcg agggcgaggg cgagggccgc ccctacgagg gccacaacac 10620
cgtgaagctg aaggtgacca agggcggccc cctgcccttc gcctgggaca tcctgtcccc 10680
ccagttccag tacggctcca aggtgtacgt gaagcacccc gccgacatcc ccgactacaa 10740
gaagctgtcc ttccccgagg gcttcaagtg ggagcgcgtg atgaacttcg aggacggcgg 10800
cgtggcgacc gtgacccagg actcctcccc gcaggacggc tgcttcatct acaaggtgaa 10860
gttcatcggc gtgaacttcc cctccgacgg ccccgtgatg cagaagaaga ccatgggctg 10920
ggaggcctcc accgagcgcc tgtacccccg cgacggcgtg ctgaagggcg agacccacaa 10980
ggccctgaag ctgaaggacg gcggccacta cctggtggag ttcaagtcca tctacatggc 11040
caagaagccc gtgcagctgc ccggctacta ctacgtggac gccaagctgg acatcacctc 11100
ccacaacgag gactacacca tcgtggagca gtacgagcgc accgagggcc gccaccacct 11160
gttcctgtga tgatcataat cagccatacc acatttgtag aggttttact tgctttaaaa 11220
aacctcccac acctcccccc gaacctgaaa cataaaatga atgcaattgt tgttgttaac 11280
ttgtttattg cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat 11340
aaagcatttt tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa 11400
cgcgagttaa ttacggccgc tcatttaaat ctggccggcc gcaaccattg tgggaaccgt 11460
gcgatcaaac aaacgcgaga taccggaagt actgaaaaac agtcgctcca ggccagtggg 11520
aacatcgatg ttttgttttg acggacccct tactctcgtc tcatataaac cgaagccagc 11580
taagatggta tacttattat catcttgtga tgaggatgct tctatcaacg aaagtaccgg 11640
taaaccgcaa atggttatgt attataatca aactaaaggc ggagtggaca cgctagacca 11700
aatgtgttct gtgatgacct gcagtaggaa gacgaatagg tggcctatgg cattattgta 11760
cggaatgata aacattgcct gcataaattc ttttattata tacagccata atgtcagtag 11820
caagggagaa aaggtccaaa gtcgcaaaaa atttatgaga aacctttaca tgagcctgac 11880
gtcatcgttt atgcgtaagc gtttagaagc tcctactttg aagagatatt tgcgcgataa 11940
tatctctaat attttgccaa atgaagtgcc tggtacatca gatgacagta ctgaagagcc 12000
agtaatgaaa aaacgtactt actgtactta ctgcccctct aaaataaggc gaaaggcaaa 12060
tgcatcgtgc aaaaaatgca aaaaagttat ttgtcgagag cataatattg atatgtgcca 12120
aagttgtttc tgactgacta ataagtataa tttgtttcta ttatgtataa gttaagctaa 12180
ttacttattt tataatacaa catgactgtt tttaaagtac aaaataagtt tatttttgta 12240
aaagagagaa tgtttaaaag ttttgttact ttatagaaga aattttgagt ttttgttttt 12300
ttttaataaa taaataaaca taaataaatt gtttgttgaa tttattatta gtatgtaagt 12360
gtaaatataa taaaacttaa tatctattca aattaataaa taaacctcga tatacagacc 12420
gataaaacac atgcgtcaat tttacgcatg attatcttta acgtacgtca caatatgatt 12480
atctttctag ggttaaataa tagtttctaa tttttttatt attcagcctg ctgtcgtgaa 12540
taccgtatat ctcaacgctg tctgtgagat tgtcgtattc tagccttttt agtttttcgc 12600
tcatcgactt gatattgtcc gacacatttt cgtcgatttg cgttttgatc aaagacttga 12660
gcagagacac gttaatcaac tgttcaaatt gatccatatt aacgatatca acccgatgcg 12720
tatatggtgc gtaaaatata ttttttaacc ctcttatact ttgcactctg cgttaatacg 12780
cgttcgtgta cagacgtaat catgttttct tttttggata aaactcctac tgagtttgac 12840
ctcatattag accctcacaa gttgcaaaac gtggcatttt ttaccaatga agaatttaaa 12900
gttattttaa aaaatttcat cacagattta aagaagaacc aaaaattaaa ttatttcaac 12960
agtttaatcg accagttaat caacgtgtac acagacgcgt cggcaaaaaa cacgcagccc 13020
gacgtgttgg ctaaaattat taaatcaact tgtgttatag tcacggattt gccgtccaac 13080
gtgttcctca aaaagttgaa gaccaacaag tttacggaca ctattaatta tttgattttg 13140
ccccacttca ttttgtggga tcacaatttt gttatatttt aaacaaagct tggcactggc 13200
cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt acccaactta atcgccttgc 13260

agcacatccc cctttcgcca gctggcgtaa tagcgaagag gcccgcaccg atcgcccttc 13320
ccaacagttg cgcagcctga atggcgaatg gcgcctgatg cggtatttte tccttacgca 13380
tctgtgcggt atttcacacc gcatatggtg cactctcagt acaatctgct ctgatgccgc 13440
atagttaagc cagccccgac acccgccaac acccgctgac gcgccctgac gggcttgtct 13500
gctcccggca tccgcttaca gacaagctgt gaccgtctcc gggagctgca tgtgtcagag 13560
gttttcaccg tcatcaccga aacgcgcgag acgaaagggc ctcgtgatac gcctattttt 13620
ataggttaat gtcatgataa taatggtttc ttagacgtca ggtggcactt ttcggggaaa 13680
tgtgcgcgga acccctattt gtttatttttt ctaaatacat tcaaatatgt atccgctcat 13740
gagacaataa ccctgataaa tgcttcaata atattgaaaa aggaagagta tgagtattca 13800
acatttccgt gtcgcccttа ttcccttttt tgcggcattt tgccttcctg tttttgctca 13860
cccagaaacg ctggtgaaag taaaagatgc tgaagatcag ttgggtgcac gagtgggtta 13920
catcgaactg gatctcaaca gcggtaagat ccttgagagt tttcgccccg aagaacgttt 13980
tccaatgatg agcacttttа aagttctgct atgtggcgcg gtattatccc gtattgacgc 14040
cgggcaagag caactcggtc gccgcataca ctattctcag aatgacttgg ttgagtactc 14100
accagtcaca gaaaagcatc ttacggatgg catgacagta agagaattat gcagtgctgc 14160
cataaccatg agtgataaca ctgcggccaa cttacttctg acaacgatcg gaggaccgaa 14220
ggagctaacc gcttttttgc acaacatggg ggatcatgta actcgccttg atcgttggga 14280
accggagctg aatgaagcca taccaaacga cgagcgtgac accacgatgc ctgtagcaat 14340
ggcaacaacg ttgcgcaaac tattaactgg cgaactactt actctagctt cccggcaaca 14400
attaatagac tggatggagg cggataaagt tgcaggacca cttctgcgct cggcccttcc 14460
ggctggctgg tttattgctg ataaatctgg agccggtgag cgtgggtctc gcggtatcat 14520
tgcagcactg gggccagatg gtaagccctc ccgtatcgta gttatctaca cgacggggag 14580
tcaggcaact atggatgaac gaaatagaca gatcgctgag ataggtgcct cactgattaa 14640
gcattggtaa ctgtcagacc aagtttactc atatatactt tagattgatt taaaacttca 14700
tttttaattt aaaaggatct aggtgaagat ccttttttgat aatctcatga ccaaaatccc 14760
ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca aaggatcttc 14820
ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac caccgctacc 14880
agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg taactggctt 14940
cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag gccaccactt 15000
caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac cagtggctgc 15060
tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt taccggataa 15120
ggcgcagcgg tcgggctgaa cgggggttc gtgcacacag cccagcttgg agcgaacgac 15180
ctacaccgaa ctgagatacc tacagcgtga gcattgagaa agcgccacgc ttcccgaagg 15240
gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc gcacgaggga 15300
gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc acctctgact 15360
tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa acgccagcaa 15420
cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt tctttcctgc 15480
gttatcccct gattctgtgg ataaccgtat taccgccttt gagtgagctg ataccgctcg 15540
ccgcagccga acgaccgagc gcagcgagtc agtgagcgag gaagcggaag agcgcccaat 15600
acgcaaaccg cctctccccg cgcgttggcc gattcattaa tgcagctggc acgacaggtt 15660
tcccgactgg aaagcgggca gtgagcgcaa cgcaattaat gtgagttagc tcactcatta 15720
ggcaccccag gctttacact ttatgcttcc ggctcgtatg ttgtgtggaa ttgtgagcgg 15780
ataacaattt cacacaggaa acagctatga ccatgattac gaatttcgac ctgcaggcat 15840
gcaagcttgc atgcctgcag gtcgacgctc gcgcgacttg gtttgccatt ctttagcgcg 15900
cgtcgcgtca cacagcttgg ccacaatgtg gtttttgtca aacgaagatt ctatgacgtg 15960
tttaaagttt aggtcgagta aagcgcaaat ctttttttaac cctagaaaga tagtctgcgt 16020
aaaattgacg catgcattct tgaaatattg ctctctcttt ctaaatagcg cgaatccgtc 16080
gctgtgcatt taggacatct cagtcgccgc ttggagctcc cgtgaggcgt gcttgtcaat 16140
gcggtaagtg tcactgattt tgaactataa cgaccgcgtg agtcaaaatg acgcatgatt 16200

atcttttacg tgacttttaa gatttaactc atacgataat tatattgtta tttcatgttc 16260
tacttacgtg ataacttatt atatatatat tttcttgtta tagatatcgt gactaatata 16320
taataaaatg ggtagttctt tagacgatga gcatatcctc tctgctcttc tgcaaagcga 16380
tgacgagctt gttggtgagg attctgacag tgaaatatca gatcacgtaa gtgaagatga 16440
cgtccagagc gatacagaag aagcgtttat agatgaggta catgaagtgc agccaacgtc 16500
aagcggtagt gaaatattag acgaacaaaa tgttattgaa caaccaggtt cttcattggc 16560
ttctaacaga atcttgacct tgccacagag gactattaga ggtaagaata aacattgttg 16620
gtcaacttca aagtccacga ggcgtagccg agtctctgca ctgaacattg tcagatcggc 16680
ccgctcgccc ggggaactag ttcaattaga gactaattca attagagcta attcaattag 16740
gatccaagct tatcgatttc gaaccctcga ccgccggagt ataaatagag gcgcttcgtc 16800
tacggagcga caattcaatt caaacaagca aagtgaacac gtcgctaagc gaaagctaag 16860
caaataaaca agcgcagctg aacaagctaa acaatcgggg taccgctaga gtcgatccca 16920
ccccacccaa gaagaagcgc aaaccggtcg ccaccatggc cctgtccaac aagttcatcg 16980
gcgacgacat gaagatgacc taccacatgg acggctgcgt gaacggccac tacttcaccg 17040
tgaagggcga gggcagcggc aagccctacg agggcaccca gacctccacc ttcaaggtga 17100
ccatggccaa cggcggcccc ctggccttct ccttcgacat cctgtccacc gtgttcatgt 17160
acggcaaccg ctgcttcacc gcctacccca ccagcatgcc cgactacttc aagcaggcct 17220
tccccgacgg catgtcctac gagagaacct tcacctacga ggacggcggc gtggccaccg 17280
ccagctggga gatcagcctg aagggcaact gcttcgagca caagtccacc ttccacggcg 17340
tgaacttccc cgccgacggc cccgtgatgg ccaagaagac caccggctgg gacccctcct 17400
tcgagaagat gaccgtgtgc gacggcatct tgaagggcga cgtgaccgcc ttcctgatgc 17460
tgcagggcgg cggcaactac agatgccagt tccacacctc ctacaagacc aagaagcccg 17520
tgaccatgcc ccccaaccac gtggtggagc accgcatcgc cagaaccgac ctggacaagg 17580
gcggcaacag cgtgcagctg accgagcacg ccgtggccca catcacctcc gtggtgcccc 17640
tctccggact cagatcataa tcagccatac cacatttgta gaggttttac ttgctttaaa 17700
aaacctccca cacctccccc tgaacctgaa acataaaatg aatgcaattg ttgttgttaa 17760
cttgtttatt gcagcttata atggttacaa ataaagcaat agcatcacaa atttcacaaa 17820
taaagcattt ttttcactgc attctagttg tggtttgtcc aaactcatca atgtatctta 17880
ccgcggagtg gacacgctag accaaatgtg ttctgtgatg acctgcagta ggaagacgaa 17940
taggtggcct atggcattat tgtacggaat gataaacatt gcctgcataa attcttttat 18000
tatatacagc cataatgtca gtagcaaggg agaaaaggtc caaagtcgca aaaaatttat 18060
gagaaacctt tacatgagcc tgacgtcatc gtttatgcgt aagcgtttag aagctcctac 18120
tttgaagaga tatttgcgcg ataatatctc taatattttg ccaaatgaag tgcctggtac 18180
atcagatgac agtactgaag agccagtaat gaaaaaacgt acttactgta cttactgccc 18240
ctctaaaata aggcgaaagg caaatgcatc gtgcaaaaaa tgcaaaaaag ttatttgtcg 18300
agagcataat attgatatgt gccaaagttg tttctgactg actaataagt ataatttgtt 18360
tctattatgt ataagttaag ctaattactt attttataat acaacatgac tgtttttaaa 18420
gtacaaaata agtttatttt tgtaaaagag agaatgttta aaagttttgt tactttatag 18480
aagaaatttt gagtttttgt tttttttaa taaataaata aacataaata aattgtttgt 18540
tgaatttatt attagtatgt aagtgtaaat ataataaaac ttaatatcta ttcaaattaa 18600
taaataaacc tcgatataca gaccgataaa acacatgcgt caattttacg catgattatc 18660
tttaacgtac gtcacaatat gattatcttt ctagggttaa aatgaatgta agcactttat 18720
taacgaaatc tttgggaata tttcgctcat cagcatttta tttgagcagg agtccgagat 18780
gcccgggcgg 18790

<210> 152
<211> 19053
<212> ADN
<213> Nhân tạo

<220>
<223> Trình tự plasmit nguyên vẹn LA3612
<400> 152

gggcatctcg gactcctgct caaataaaat gctgatgagc gaaatattcc caaagatttc 6
gttaataaag tgcttacatt cattttaacc ctagaaagat aatcatattg tgacgtacgt 12
taaagataat catgcgtaaa attgacgcat gtgttttatc ggtctgtata tcgaggttta 18
tttattaatt tgaatagata ttaagtttta ttatatttac acttacatac taataataaa 24
ttcaacaaac aatttattta tgtttattta tttattaaaa aaaaacaaaa actcaaaatt 30
tcttctataa agtaacaaaa cttttaaaca ttctctcttt tacaaaaata aacttatttt 36
gtactttaaa aacagtcatg ttgtattata aaataagtaa ttagcttaac ttatacataa 42
tagaaacaaa ttatacttat tagtcagtca gaaacaactt tggcacatat caatattatg 48
ctctcgacaa ataacttttt tgcatttttt gcacgatgca tttgcctttc gccttatttt 54
agaggggcag taagtacagt aagtacgttt tttcattact ggctcttcag tactgtcatc 60
tgatgtacca ggcacttcat ttggcaaaat attagagata ttatcgcgca aatatctctt 66
caaagtagga gcttctaaac gcttacgcat aaacgatgac gtcaggctca tgtaaaggtt 72
tctcataaat tttttgcgac tttggacctt ttctcccttg ctactgacat tatggctgta 78
tataataaaa gaatttatgc aggcaatgtt tatcattccg tacaataatg ccataggcca 84
cctattcgtc ttcctactgc aggtcatcac agaacacatt tggtctagcg tgtccactcc 90
gcggtaagat acattgatga gtttggacaa accacaacta gaatgcagtg aaaaaaatgc 960
tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag ctgcaataaa 1020
caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga ggtgtgggag 1080
gtttttttaaa gcaagtaaaa cctctacaaa tgtggtatgg ctgattatga tctgagtccg 1140
gagaagggca ccacggaggt gatgtgggcc acggcgtgct cggtcagctg cacgctgttg 1200
ccgcccttgt ccaggtcggt tctggcgatg cggtgctcca ccacgtggtt ggggggcatg 1260
gtcacgggct tcttggtctt gtaggaggtg tggaactggc atctgtagtt gccgccgccc 1320
tgcagcatca ggaaggcggt cacgtcgccc ttcaagatgc cgtcgcacac ggtcatcttc 1380
tcgaaggagg ggtcccagcc ggtggtcttc ttggccatca cggggccgtc ggcggggaag 1440
ttcacgccgt ggaaggtgga cttgtgctcg aagcagttgc ccttcaggct gatctcccag 1500
ctggcggtgg ccacgccgcc gtcctcgtag gtgaaggttc tctcgtagga catgccgtcg 1560
gggaaggcct gcttgaagta gtcgggcatg ctggtggggt aggcggtgaa gcagcggttg 1620
ccgtacatga acacggtgga caggatgtcg aaggagaagg ccaggggccc gccgttggcc 1680
atggtcacct tgaaggtgga ggtctgggtg ccctcgtagg gcttgccgct gccctcgccc 1740
ttcacggtga agtagtggcc gttcacgcag ccgtccatgt ggtaggtcat cttcatgtcg 1800
tcgccgatga acttgttgga cagggccatg gtggcgaccg gtttgcgctt cttcttgggt 1860
ggggtgggat cgactctagc ggtaccccga ttgtttagct tgttcagctg cgcttgttta 1920
tttgcttagc tttcgcttag cgacgtgttc actttgcttg tttgaattga attgtcgctc 1980
cgtagacgaa gcgcctctat ttatactccg gcggtcgagg gttcgaaatc gataagcttg 2040
gatcctaatt gaattagctc taattgaatt agtctctaat tgaactagtt ccccgggcga 2100
gcgggccgat ctgacaatgt tcagtgcaga gactcggcta cgcctcgtgg actttgaagt 2160
tgaccaacaa tgtttattct tacctctaat agtcctctgt ggcaaggtca agattctgtt 2220
agaagccaat gaagaacctg gttgttcaat aacattttgt tcgtctaata tttcactacc 2280
gcttgacgtt ggctgcactt catgtacctc atctataaac gcttcttctg tatcgctctg 2340
gacgtcatct tcacttacgt gatctgatat ttcactgtca gaatcctcac caacaagctc 2400
gtcatcgctt tgcagaagag cagagaggat atgctcatcg tctaaagaac tacccatttt 2460
attatatatt agtcacgata tctataacaa gaaaatatat atataataag ttatcacgta 2520
agtagaacat gaaataacaa tataattatc gtatgagtta aatcttaaaa gtcacgtaaa 2580
agataatcat gcgtcatttt gactcacgcg gtcgttatag ttcaaaatca gtgacactta 2640
ccgcattgac aagcacgcct cacgggagct ccaagcggcg actgagatgt cctaaatgca 2700
cagcgacgga ttcgcgctat ttagaaagag agagcaatat ttcaagaatg catgcgtcaa 2760

ttttacgcag actatctttc tagggttaaa aaagatttgc gctttactcg acctaaactt 282(
taaacacgtc atagaatctt cgtttgacaa aaaccacatt gtggccaagc tgtgtgacgc 288(
gacgcgcgct aaagaatggc aaaccaagtc gcgcgagcgt cgacctgcag gcatgcaagc 294(
ttgcatgcct gcaggtcgaa attcgtaatc atggtcatag ctgtttcctg tgtgaaattg 300(
ttatccgctc acaattccac acaacatacg agccggaagc ataaagtgta aagcctgggg 306(
tgcctaatga gtgagctaac tcacattaat tgcgttgcgc tcactgcccg ctttccagtc 312(
gggaaacctg tcgtgccagc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt 318(
gcgtattggg cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct 324(
gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga 330(
taacgcagga aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc 336(
cgcgttgctg gcgtttttcc ataggctccg cccccctgac gagcatcaca aaaatcgacg 342(
ctcaagtcag aggtggcgaa acccgacagg actataaaga taccaggcgt ttcccccctgg 3480
aagctccctc gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt 3540
tctcccttcg ggaagcgtgg cgctttctca atgctcacgc tgtaggtatc tcagttcggt 3600
gtaggtcgtt cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg 3660
cgccttatcc ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact 3720
ggcagcagcc actggtaaca ggattagcag agcgaggtat gtaggcggtg ctacagagtt 3780
cttgaagtgg tggcctaact acggctacac tagaaggaca gtatttggta tctgcgctct 3840
gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 3900
cgctggtagc ggtggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 3960
tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg 4020
ttaagggatt ttggtcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta 4080
aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa acttggtctg acagttacca 4140
atgcttaatc agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttgc 4200
ctgactcccc gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc 4260
tgcaatgata ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc 4320
agccggaagg gccgagcgca gaagtggtcc tgcaacttta tccgcctcca tccagtctat 4380
taattgttgc cgggaagcta gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt 4440
tgccattgct acaggcatcg tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc 4500
cggttcccaa cgatcaaggc gagttacatg atcccccatg ttgtgcaaaa aagcggttag 4560
ctccttcggt cctccgatcg ttgtcagaag taagttggcc gcagtgttat cactcatggt 4620
tatggcagca ctgcataatt ctcttactgt catgccatcc gtaagatgct tttctgtgac 4680
tggtgagtac tcaaccaagt cattctgaga atagtgtatg cggcgaccga gttgctcttg 4740
cccggcgtca atacgggata ataccgcgcc acatagcaga actttaaaag tgctcatcat 4800
tggaaaacgt tcttcggggc gaaaactctc aaggatctta ccgctgttga gatccagttc 4860
gatgtaaccc actcgtgcac ccaactgatc ttcagcatct tttactttca ccagcgtttc 4920
tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa 4980
atgttgaata ctcatactct tccttttca atattattga agcatttatc agggttattg 5040
tctcatgagc ggatacatat ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg 5100
cacatttccc cgaaaagtgc cacctgacgt ctaagaaacc attattatca tgacattaac 5160
ctataaaaat aggcgtatca cgaggccctt tcgtctcgcg cgtttcggtg atgacggtga 5220
aaacctctga cacatgcagc tcccggagac ggtcacagct tgtctgtaag cggatgccgg 5280
gagcagacaa gcccgtcagg gcgcgtcagc gggtgttggc gggtgtcggg gctggcttaa 5340
ctatgcggca tcagagcaga ttgtactgag agtgcaccat atgcggtgtg aaataccgca 5400
cagatgcgta aggagaaaat accgcatcag gcgccattcg ccattcaggc tgcgcaactg 5460
ttgggaaggg cgatcggtgc gggcctcttc gctattacgc cagctggcga aaggggatg 5520
tgctgcaagg cgattaagtt gggtaacgcc agggttttcc cagtcacgac gttgtaaaac 5580
gacggccagt gccaagcttt gtttaaaata taacaaaatt gtgatcccac aaaatgaagt 5640
ggggcaaaat caaataatta atagtgtccg taaacttgtt ggtcttcaac tttttgagga 5700

acacgttgga cggcaaatcc gtgactataa cacaagttga tttaataatt ttagccaaca 5760
cgtcgggctg cgtgtttttt gccgacgcgt ctgtgtacac gttgattaac tggtcgatta 5820
aactgttgaa ataatttaat ttttggttct tctttaaatc tgtgatgaaa ttttttaaaa 5880
taactttaaa ttcttcattg gtaaaaaatg ccacgttttg caacttgtga gggtctaata 5940
tgaggtcaaa ctcagtagga gttttatcca aaaaagaaaa catgattacg tctgtacacg 6000
aacgcgtatt aacgcagagt gcaaagtata agagggttaa aaaatatatt ttacgcacca 6060
tatacgcatc gggttgatat cgttaatatg gatcaatttg aacagttgat taacgtgtct 6120
ctgctcaagt ctttgatcaa aacgcaaatc gacgaaaatg tgtcggacaa tatcaagtcg 6180
atgagcgaaa aactaaaaag gctagaatac gacaatctca cagacagcgt tgagatatac 6240
ggtattcacg acagcaggct gaataataaa aaaattagaa actattattt aaccctagaa 6300
agataatcat attgtgacgt acgttaaaga taatcatgcg taaaattgac gcatgtgttt 6360
tatcggtctg tatatcgagg tttatttatt aatttgaata gatattaagt tttattatat 6420
ttacacttac atactaataa taaattcaac aaacaattta tttatgttta tttatttatt 6480
aaaaaaaaac aaaaactcaa aatttcttct ataaagtaac aaaactttta aacattctct 6540
cttttacaaa aataaactta ttttgtactt taaaaacagt catgttgtat tataaaataa 6600
gtaattagct taacttatac ataatagaaa caaattatac ttattagtca gtcagaaaca 6660
actttggcac atatcaatat tatgctctcg acaaataact tttttgcatt ttttgcacga 6720
tgcatttgcc tttcgcctta ttttagaggg gcagtaagta cagtaagtac gttttttcat 6780
tactggctct tcagtactgt catctgatgt accaggcact tcatttggca aaatattaga 6840
gatattatcg cgcaaatatc tcttcaaagt aggagcttct aaacgcttac gcataaacga 6900
tgacgtcagg ctcatgtaaa ggtttctcat aaatttttttg cgactttgga ccttttctcc 6960
cttgctactg acattatggc tgtatataat aaaagaattt atgcaggcaa tgtttatcat 7020
tccgtacaat aatgccatag gccacctatt cgtcttccta ctgcaggtca tcacagaaca 7080
catttggtct agcgtgtcca ctccgccttt agtttgatta taatacataa ccatttgcgg 7140
tttaccggta ctttcgttga tagaagcatc ctcatcacaa gatgataata agtataccat 7200
cttagctggc ttcggtttat atgagacgag agtaaggggt ccgtcaaaac aaaacatcga 7260
tgttcccact ggcctggagc gactgttttt cagtacttcc ggtatctcgc gtttgtttga 7320
tcgcacggtt cccacaatgg ttgcggccgg ccagatttaa atgagcggcc gtaattaact 7380
cgcgttaaga tacattgatg agtttggaca aaccacaact agaatgcagt gaaaaaaatg 7440
ctttatttgt gaaatttgtg atgctattgc tttatttgta accattataa gctgcaataa 7500
acaagttaac aacaacaatt gcattcattt tatgtttcag gttcaggggg aggtgtggga 7560
ggtttttttaa agcaagtaaa acctctacaa atgtggtatg gctgattatg atcatcacag 7620
gaacaggtgg tggcggccct cggtgcgctc gtactgctcc acgatggtgt agtcctcgtt 7680
gtgggaggtg atgtccagct tggcgtccac gtagtagtag ccgggcagct gcacgggctt 7740
cttggccatg tagatggact tgaactccac caggtagtgg ccgccgtcct tcagcttcag 7800
ggccttgtgg gtctcgccct tcagcacgcc gtcgcggggg tacaggcgct cggtggaggc 7860
ctcccagccc atggtcttct tctgcatcac ggggccgtcg gaggggaagt tcacgccgat 7920
gaacttcacc ttgtagatga agcagccgtc ctgcagggag gagtcctggg tcacggtcgc 7980
cacgccgccg tcctcgaagt tcatcacgcg ctcccacttg aagccctcgg ggaaggacag 8040
cttcttgtag tcgcgggatgt cggcgggtg cttcacgtac accttggagc cgtactggaa 8100
ctgggggac aggatgtccc aggcgaaggg caggggccg cccttggtca ccttcagctt 8160
cacggtgttg tggccctcgt aggggcggcc ctcgccctcg ccctcgatct cgaactcgtg 8220
gccgttcacg gtgccctcca tgcgcacctt gaagcgcatg aactcggtga tgacgttctc 8280
ggaggaggcc atggtaccgg tttgcgcttc ttcttgggtg gggtgggatc gactctagcg 8340
gtaccccgat tgtttagctt gttcagctgc gcttgtttat ttgcttagct ttcgcttagc 8400
gacgtgttca ctttgcttgt ttgaattgaa ttgtcgctcc gtagacgaag cgcctctatt 8460
tatactccgg cggtcgaggg ttcgaaatcg ataagcttgg atcctcgagg tcatcttcac 8520
ttacgtgatc tgatatttca ctgtcagaat cctcaccaac aagctcgtca tcgctttgca 8580
gaagagcaga gaggatatgc tcatcgtcta aagaactacc cattttatta tatattagtc 8640

acgatatcta taacaagaaa atatatatat aataagttat cacgtaagta gaacatgaaa 8700
taacaatata attatcgtat gagttaaatc ttaaaagtca cgtaaaagat aatcatgcgt 8760
cattttgact cacgcggtcg ttatagttca aaatcagtga cacttaccgc attgacaagc 8820
acgcctcacg ggagctccaa gcggcgactg agatgtccta aatgcacagc gacggattcg 8880
cgctatttag aaagagagag caatatttca agaatgcatg cgtcaatttt acgcagacta 8940
tctttctagg gttaaatata cttcaaggcc accgttatgg tgtcattagg gacgaggtcc 9000
aaaacgtcaa aatattttgt aatgatgccg tgaattacga tttcggtgtt aattaagtac 9060
tgacggacac accgaagccc cggcggcaac cctcagcgga tgccccgggg cttcacgttt 9120
tcccaggtca gaagcggttt tcgggagtag tgccccaact ggggtaacct ttgagttctc 9180
tcagttgggg gcgtagggtc gccgacatga cacaaggggt tgtgaccggg gtggacacgt 9240
acgcgggtgc ttacgaccgt cagtcgcgcg agcgcgaggc caggccgcat ggtacccatt 9300
gcttgtcatt tattaatttg gatgatgtca tttgttttta aaattgaact ggctttacga 9360
gtagaattct acgcgtaaaa cacaatcaag tatgagtcat aatctgatgt catgttttgt 9420
acacggctca taaccgaact ggctttacga gtagaattct acttgtaatg cacgatcagt 9480
ggatgatgtc atttgttttt caaatcgaga tgatgtcatg ttttgcacac ggctcataaa 9540
ctcgctttac gagtagaatt ctacgtgtaa cgcacgatcg attgatgagt catttgtttt 9600
gcaatatgat atcatacaat atgactcatt tgtttttcaa aaccgaactt gatttacggg 9660
tagaattcta cttgtaaagc acaatcaaaa agatgatgtc atttgttttt caaaactgaa 9720
ctcgctttac gagtagaatt ctacgtgtaa aacacaatca agaaatgatg tcatttgtta 9780
taaaaataaa agctgatgtc atgttttgca catggctcat aactaaactc gctttacggg 9840
tagaattcta cgcgtaaaac atgattgata attaaataat tcatttgcaa gctatacgtt 9900
aaatcaaacg gacgctcgag gttgcacaac actattatcg atttgcagtt cgggacataa 9960
atgtttaaat atatcgatgt ctttgtgatg cgcgcgacat ttttgtaggt tattgataaa 10020
atgaacggat acgttgcccg acattatcat taaatccttg gcgtagaatt tgtcgggtcc 10080
attgtccgtg tgcgctagta gcatgcccgt aacggacctc gtacttttgg cttcaaaggt 10140
tttgcgcaca gacaaaatgt gccacacttg cagctctgca tgtgtgcgcg ttaccacaaa 10200
tcccaacggc gcagtgtact tgttgtatgc aaataaatct cgataaaggc gcggcgcgcg 10260
aatgcagctg atcacgtacg ctcctcgtgt tccgttcaag gacggtgtta tcgacctcag 10320
attaatgttt atcggccgac tgttttcgta tccgctcacc aaacgcgttt ttgcattaac 10380
attgtatgtc ggcggatgtt ctatatctaa tttgaataaa taaacgataa ccgcgttggt 10440
tttagagggc ataataaaag aaatattgtt atcgtgttcg ccattagggc agtataaatt 10500
gacgttcatg ttggatattg tttcagttgc aagttgacac tggcggcgac aagcaattct 10560
aattggggta agttttcccg ttcttttctg ggttcttccc ttttgctcat ccttgctgca 10620
ctaccttcag gtgcaagttg agattcaggc caccatggga gatcccaccc cacccaagaa 10680
gaagcgcaaa ccggtcgcca ccatggcctc ctccgagaac gtcatcaccg agttcatgcg 10740
cttcaaggtg cgcatggagg gcaccgtgaa cggccacgag ttcgagatcg agggcgaggg 10800
cgagggccgc ccctacgagg gccacaacac cgtgaagctg aaggtgacca agggcggccc 10860
cctgcccttc gcctgggaca tcctgtcccc ccagttccag tacggctcca aggtgtacgt 10920
gaagcacccc gccgacatcc ccgactacaa gaagctgtcc ttccccgagg gcttcaagtg 10980
ggagcgcgtg atgaacttcg aggacggcgg cgtggcgacc gtgacccagg actcctccct 11040
gcaggacggc tgcttcatct acaaggtgaa gttcatcggc gtgaacttcc cctccgacgg 11100
ccccgtgatg cagaagaaga ccatgggctg gaggcctcc accgagcgcc tgtacccccg 11160
cgacggcgtg ctgaagggcg agacccacaa ggccctgaag ctgaaggacg gcggccacta 11220
cctggtggag ttcaagtcca tctacatggc caagaagccc gtgcagctgc ccggctacta 11280
ctacgtggac gccaagctgg acatcacctc ccacaacgag gactacacca tcgtggagca 11340
gtacgagcgc accgagggcc gccaccacct gttcctgaga tctcgaccca agaaaaagcg 11400
gaaggtggag gacccgtaag atccaccgga tctagataac tgatcataat cagccatacc 11460
acatttgtag aggttttact tgctttaaaa aacctcccac acctcccctt gaacctgaaa 11520
cataaaatga atgcaattgt tgttgttaac ttgtttattg cagcttataa tggttacaaa 11580

taaagcaata gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt 11640
ggtttgtcca aactcatcaa tgtatcttaa cgcgagttta aacctacccg ccatactcat 11700
cgatgcccag cgcgtcggtg aacatttgct cgaactcgaa gtcggccatg tccagggcgc 11760
cgtacggggc gctatcgtgg ggcgtgaagc ccggtcccgg gctatctcca tcgcccagca 11820
tatccaggtc gaaatcgtcc agggcgtcgg cgtgggccat tgccacatcc tctccatcca 11880
ggtgcagctc gtcgccccagg ctcacatcgg tcggcggggc ggtgctcagg cggcgcgtgt 11940
gtccggcggg caggaagctc aggcgggggg cggccaggcc ggcttcctcc ggggcatcgt 12000
catccggcag gtccagcagt ccctcgatgg tgctgccata gttgttcttg gtacgggcgc 12060
ggctgtaggc gctgcccatg gtggctagcc ccgctttcta aaggtgttat aaatcaaatt 12120
agttttgttt tttcttgaaa actttgcgtt tcctttgatc aacttaccgc cagggtacct 12180
gcagattgtt tagcttgttc agctgcgctt gtttatttgc ttagctttcg cttagcgacg 12240
tgttcacttt gcttgtttga attgaattgt cgctccgtag acgaagcgcc tctatttata 12300
ctccggcgct cgcggccgca tcgatctcta tcactgatag ggaggtctct atcactgata 12360
gggagttctc tatcactgat agggatgtct ctatcactga tagggatttc tctatcactg 12420
atagggaagt ctctatcact gatagggacc tctctatcac tgatagggaa atctctatca 12480
ctgataggga tctctctatc actgataggg acttctctat cactgatagg gacgtctcta 12540
tcactgatag ggaactctct atcactgata gggacatctc tatcactgat agggacttct 12600
ctatcactga tagggagtat gttctctctc ttctcttctc tctctctttc tcgaatgttc 12660
tctctcttct cttctctctc tctttctcga tggccggggc gcgccaggtt tcgactttca 12720
cttttctcta tcactgatag ggagtggtaa actcgacttt cacttttctc tatcactgat 12780
agggagtggt aaactcgact ttcacttttc tctatcactg atagggagtg gtaaactcga 12840
ctttcacttt tctctatcac tgatagggag tggtaaactc gactttcact tttctctatc 12900
actgataggg agtggtaaac tcgactttca cttttctcta tcactgatag ggagtggtaa 12960
actcgacttt cacttttctc tatcactgat agggagtggt aaactcgaaa acgagcgccg 13020
gagtataaat agaggcgctt cgtctacgga gcgacaattc aattcaaaca agcaaagtga 13080
acacgtcgct aagcgaaagc taagcaaata aacaagcgca gctgaacaag ctaaacaatc 13140
tgccctagga tctcagtggc tcctggagaa gctgcgatac ccctgggaga tgatgcccct 13200
gatgtacgtg atactgaaag gcgccgacgg agacgtcaat aaagcgcgcc aacggattga 13260
cgaaggtatg ggggttctta ccggttggga ctgtttccga ggtatcgatc gggtgtcact 13320
cacttcctgg gtgctcccat tttgtaactg ctaacgctta ttattgagtt tcaggacatc 13380
tgggatcttc ggtcgacgga gtctattccc aacagtgccc tggatcaaac actgccatca 13440
tgcagtttcc gtagcctgtt gggctacgct ccccgacttg acatccccca ttcttatcaa 13500
acaacaactc aaggcctgag acaacgagtg gtggaatttg cgcacgaagt cattggtttg 13560
tcctggtaaa agttaaaagg gttaactgga gggttaattg acacggtttc aactgatggc 13620
cttattgaca cacggatgaa agacttgcac gcttgacctt ctgtctgtac taataaaagt 13680
tacgttggct gggttttggg gtcataatgg ccccaaaatc gaatcgtcat aacttcttga 13740
aatacaactc acgtttaaga ccattcaaga gtattagatc atcgtctata atagcagatt 13800
tgaaatttac ttcacatttc ggtattgcag tgcccccttgc ttccacaatg gaattaggtc 13860
ggtggtgcgt cgatcgtcgc aagtttatcg ttaaacagtc aataaaatga gcattttata 13920
tcgtgataca tatgagaaga tagaggtttc aattaaaaca aatccacatg gtgtcgctaa 13980
taaaattgtg cattttaagc gagttatatc ctctgatcaa gataaaatag aaaattcgat 14040
ttttgaatat tcaattataa gagcctgaat aactacaaca tgtagtgaat cgaaactgat 14100
ttatgacggt ttgtgaaggt tacacgtcct aagcatttgg attcaagaaa agcaagagat 14160
atgacgaatg taaactttat cgtatcaatg aagtaactag cgtccagaac agtacaaacc 14220
aacatcgtac cgtcgtattc cactccggtc gttgcaatat ctctaggtcc accgaaaaac 14280
actcatgacc aagatcgtgt cgtcgatctt ggtccaccga aacaccgatg tccatatcgt 14340
ttcgtcgaac ttggaccaac gattcatgca actgatgaca acgcggcccc cgggtcgtac 14400
caatatccga aaaatccaac tgttcttctc tgcctcgcag gtcaagccgt ggtcaatgaa 14460
tactcacgat tgcacaatct gaacatgttc gacggtgtag agttgcgcag tacgacgcgc 14520

cagtccggat gatagacttt ttacacgatc agcacgaccc actgcgctgc ggcaaaggtc 14580
gaaccgaaac aagaataaac cacgaagatc agatcgattc gacggaagaa gcaatcgaat 14640
gcaaagaaga atcggaacga agaaaactct aaagcatcgc atatttacaa agcataacgg 14700
aaaacccgca agttcaaact agtgattagt gtaagatgaa gcaaagcaga aatgtagtat 14760
ctagattttt cgacgttagt ttacaaagat aaaaaatgag gttggacata caatcgtggg 14820
tattcgtctg agttcgtcac aactgcaccg gaaactgtga aacagaatag agccaacctg 14880
tgcgcggaga atgttgaggt cattataagc ttccttagca tccacgggtg aaagtcgatc 14940
gacggaagcc tgcaagactc tgtcgatggg ctttcgtcct agaagaataa gattaaacct 15000
gaaatgtatt ctcccgtgga atggtttcat ttgagtaatt ctgtatcttc tccttcccaa 15060
ttccacgaac gcgacgaact ctaatacaaa caacataatg accacagtgc aaatgctgtt 15120
taacgataat agcgacatgc agccattctg gggctaccac gtgtggctct acttgcgatc 15180
caaaatgcag atcttcgtca agaccctgac cggcaagacc atcaccctgg aggtggagcc 15240
gagcgatacc atcgagaacg tgaaggccaa gatccaggac aaggagggca tcccgccgga 15300
tcagcagcgc ctgatcttcg ccggacgcca gctggaggat ggccgcaccc tgagcgacta 15360
caacatccag aaggagagca ccctgcacct ggtgctgcgc ctgcgcggtg gtatggtcag 15420
ccgcctggat aagtccaaag tcatcaactc cgcgttggag ctgttgaacg aagttggcat 15480
tgagggactg acgacccgca agttggcgca gaagctgggc gtggagcagc ccaccctcta 15540
ctggcacgtg aagaataagc gggcgctgct ggatgccctg gccatcgaga tgctcgaccg 15600
ccaccacacg cattttgcc cgttggaagg cgagtcctgg caggacttcc tccgcaataa 15660
cgccaagtcg ttccgctgcg ctctgctgtc ccaccgagac ggtgccaaag tccatctcgg 15720
cacgcgcccg accgaaaagc aatacgagac actggagaac cagctcgcgt tcctgtgcca 15780
gcaaggcttc agcctggaaa atgctctcta cgctctgagc gccgtcggtc actttaccct 15840
gggctgcgtg ctggaggacc aagagcatca agtcgcaaaa gaggagcgcg agaccccaac 15900
aaccgattcg atgcccccac tgctgcgtca ggcaatcgag ctgttcgatc atcaaggagc 15960
cgagccggca ttcctgttcg gcttggagct gattatctgc ggattggaaa agcaactgaa 16020
atgcgagtcg ggctcgggcc ccgcctacag ccgcgcccgc accaagaaca actacggcag 16080
caccatcgag ggcctgctgg atctgccgga tgatgatgcc ccggaggagg cgggcctggc 16140
cgccccgcgc ctgagcttcc tgccggccgg acacacccgc cgcctgtcga ccgccccgcc 16200
gaccgacgtg agcctgggcg atgagctgca cctggatggc gaggatgtgg cgatggccca 16260
cgccgatgcc ctggacgact tcgacctgga catgctgggc gatggcgata gcccgggacc 16320
gggattcacc ccgcacgata gcgccccctа cggcgccctg gatatggccg atttcgagtt 16380
cgagcagatg ttcaccgacg ccctgggcat cgatgagtac ggcggctaac accggtgatt 16440
gtccaccgca agctctactt gtgagacagc gttcctaaag agtgtgaaag tgcaaacaag 16500
tgatgaaacc aatagtgcaa agcaagttta gagggaaaat ttaaaaaatg caaaacagca 16560
gtagtactta acttttaaga ttgtgtttcg aaagccgaag tgaggctgtt ccatctgcca 16620
ccggaaaaaa acgacgacag cagaatcatc aacaagcaac atccatccga aaaaatccgg 16680
gaaaccggat cttcaaccaa ccatcctaca atctacaaac cagagattat atctcttcaa 16740
tcgtttccga catcggtcgg tttcggtgcc caaaatgatc tgataaacac ttatctctct 16800
gtagcttgca tgccattgcg agcgtatttt ggtagctggc cgttgccaaa cggctccgac 16860
aggtactgct attggaggtt gtgcacgacc acgttgagtt tgccttttga gttggagagt 16920
gtgtcttttc gtcatatatt cggcctttтc aagggtgatt ttcaggctac gtaatgattg 16980
tatagtttaa ccagctaaaa catattgatg acaagttcta tttcagcacc acaaacaagc 17040
ctgttaatgt ctctcaccgc aaccattgtt ctgcgcgcgt tataatcagc atagaagttt 17100
attttctttg ggatgattca aatattacgt gacgcaaagt ttgccaattt tagaacccct 17160
ccctcctcca cgtaacggct tttgtgtgaa aaatttaaat tttgtgtata gaccgtagca 17220
tttcggaaga ccccctccct tactctgttg agttacgtaa aatttcaacg atcctttgt 17280
agttctgaat tttatatcag cgtgcagtgt tatgaagata tccacagtat aaaatattat 17340
tttattttaa attctatgct gattatcaat gtgttactag tggcttttca tactcatgtt 17400
gcgagctcga tttggcgcac ggtacttatc aaggcatgta tgtatgttgt ttgaagcaac 17460

tgtataactg tttgaaacta tctaattggt gagctcgttt catttagtat ataataatga 17520
taattgctat ggagacgtta tttactagca agtgatttga cgacctgaaa tcggaacaaa 17580
tagacaacgt ttttataaat acaataaatc agaactttcc attattgggt acaaagagtt 17640
gcgctatttc gatactgtca gatcagattt tccagcacaa cgataccttg atatgcgata 17700
acttagaatt agaccttcaa atccatctct ccagctatga acagtcatat agataaagcc 17760
aatggcgtta tgaggtagcg gaaagcgtca tctttccaat gctatctaag tacataattt 17820
gctatagctt tctattaatc gtagtttgag agatgcaaag tcagttatct cgtatcaagg 17880
tttgattgtt ttggaaatta gctaaacagt tgacattatc acccgtcttt aggggataag 17940
cgcatacaaa tgtgtattta gttgttcatt gaagtaacgt aagataggca agtatggaaa 18000
cgagctcacc aaacgtcgaa atacgtctaa taaatttgtg ttcagcagga tggttcaaaa 18060
tttatttgca tcacctcaaa attacagtac ctagtgctgt ttgtgacaaa catcaaaagg 18120
taaaatcaaa ctcgtggcgt cgtgcaatct ccatagaatg aacaatttct aaccgtattt 18180
gatggaaaga cattgagtct actatcctct taacagcatt gcacttgtct ataaacaata 18240
aataatttgt tctttttac attttctttc cccactttcg cccccccccc ccccaaaaat 18300
caatccctca aacaggatac gacatttgtt gcatctactt tccgaagcgt tccagcagac 18360
acagacactg gccggacgag gagaacatct ccgtcacccg cactccgtct gcgtcacggt 18420
cgccatgtgc cgattttcgt acccggtcac agtccagctc gccggataac aacggtggcg 18480
cgctcaatct ggacacgaaa tctaccaaag cgacgaccgc caccaccgac gacgaagagg 18540
ttatgtacga gaaacgcagc ccgaagtcca ttgaatctac cgagttgcgg tgccgtctgg 18600
aggaagcctt acacagtggc gctgctgctg ctgcggctgc tgaagaacct ctggcgggcg 18660
gaagcggttc ccactggaag agagaaagtt tcggctctac ggaggagatt cccactcgac 18720
ccgctcacag tgaaccggaa gataatggat ttgaaaacgg attggaagcg caccagtccc 18780
atattctgca cagcatacat cggaatgtgc aacgatcata atcagccata ccacatttgt 18840
agaggtttta cttgctttaa aaaacctccc acacctcccc ctgaacctga aacataaaat 18900
gaatgcaatt gttgttgtta acttgtttat tgcagcttat aatggttaca aataaagcaa 18960
tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt gtggtttgtc 19020
caaactcatc aatgtatctt aggcgcgccg ccc 19053

<210> 153
<211> 10540
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3491
<400> 153
ctaggcttta cgagtagaat tctacgcgta aaacacaatc aagtatgagt cataatctga 60
tgtcatgttt tgtacacggc tcataaccga actggcttta cgagtagaat tctacttgta 120
atgcacgatc agtggatgat gtcatttgtt tttcaaatcg agatgatgtc atgttttgca 180
cacggctcat aaactcgctt tacgagtaga attctacgtg taacgcacga tcgattgatg 240
agtcatttgt tttgcaatat gatatcatac aatatgactc atttgttttt caaaaccgaa 300
cttgatttac gggtagaatt ctacttgtaa agcacaatca aaaagatgat gtcatttgtt 360
tttcaaaact gaactcgctt tacgagtaga attctacgtg taaaacacaa tcaagaaatg 420
atgtcatttg ttataaaaat aaaagctgat gtcatgtttt gcacatggct cataactaaa 480
ctcgctttac gggtagaatt ctacgcgtaa aacatgattg ataattaaat aattcatttg 540
caagctatac gttaaatcaa acggacgctc gaggttgcac aacactatta tcgatttgca 600
gttcgggaca taaatgttta aatatatcga tgtctttgtg atgcgcgcga catttttgta 660
ggttattgat aaaatgaacg gatacgttgc ccgacattat cattaaatcc ttggcgtaga 720
atttgtcggg tccattgtcc gtgtgcgcta gcatgcccgt aacggacctc gtacttttgg 780
cttcaaaggt tttgcgcaca gacaaaatgt gccacacttg cagctctgca tgtgtgcgcg 840

ttaccacaaa tcccaacggc gcagtgtact tgttgtatgc aaataaatct cgataaaggc 900
gcggcgcgcg aatgcagctg atcacgtacg ctcctcgtgt tccgttcaag gacggtgtta 960
tcgacctcag attaatgttt atcggccgac tgttttcgta tccgctcacc aaacgcgttt 1020
ttgcattaac attgtatgtc ggcggatgtt ctatatctaa tttgaataaa taaacgataa 1080
ccgcgttggt ttagagggc ataataaaag aaatattgtt atcgtgttcg ccattagggc 1140
agtataaatt gacgttcatg ttggatattg tttcagttgc aagttgacac tggcggcgac 1200
aagcaattct aattggggta agttttcccg ttcttttctg ggttcttccc ttttgctcat 1260
ccttgctgca ctaccttcag gtgcaagttg agattcaggc caccatggga gcttcacgac 1320
gaacttgtca aacgatctca atggctcctg gagaagctgc gataccctg ggagatgatg 1380
cccctgatgt acgtgatact gaaaggcgcc gacggagacg tcaataaagc gcgccaacgg 1440
attgacgaag gtatgggggt tcttaccggt tgggactgtt tccgaggtat cgatcgggtg 1500
tcactcactt cctgggtgct cccattttgt aactgctaac gcttattatt gagtttcagg 1560
acatctggga tcttcggtcg acggagtcta ttcccaacag tgccctggat caaacactgc 1620
catcatgcag tttccgtagc ctgttgggct acgctcccg acttgacatc ccccattctt 1680
atcaaacaac aactcaaggc ctgagacaac gagtggtgga atttgcgcac gaagtcattg 1740
gtttgtcctg gtaaaagtta aaagggttaa ctggagggtt aattgacacg gtttcaactg 1800
atggccttat tgacacacgg atgaaagact tgcacgcttg accttctgtc tgtactaata 1860
aaagttacgt tggctgggtt ttggggtcat aatggcccca aaatcgaatc gtcataactt 1920
cttgaaatac aactcacgtt taagaccatt caagagtatt agatcatcgt ctataatagc 1980
agatttgaaa tttacttcac atttcggtat tgcagtgccc cttgcttcca caatggaatt 2040
aggtcggtgg tgcgtcgatc gtcgcaagtt tatcgttaaa cagtcaataa aatgagcatt 2100
ttatatcgtg atacatatga gaagatagag gtttcaatta aaacaaatcc acatggtgtc 2160
gctaataaaa ttgtgcattt taagcgagtt atatcctctg atcaagataa aatagaaaat 2220
tcgatttttg aatattcaat tataagagcc tgaataacta caacatgtag tgaatcgaaa 2280
ctgatttatg acggtttgtg aaggttacac gtcctaagca tttggattca agaaaagcaa 2340
gagatatgac gaatgtaaac tttatcgtat caatgaagta actagcgtcc agaacagtac 2400
aaaccaacat cgtaccgtcg tattccactc cggtcgttgc aatatctcta ggtccaccga 2460
aaaacactca tgaccaagat cgtgtcgtcg atcttggtcc accgaaacac cgatgtccat 2520
atcgtttcgt cgaacttgga ccaacgattc atgcaactga tgacaacgcg gcccccgggt 2580
cgtaccaata tccgaaaaat ccaactgttc ttctctgcct cgcaggtcaa gccgtggtca 2640
atgaatactc acgattgcac aatctgaaca tgttcgacgg tgtagagttg cgcagtacga 2700
cgcgccagtc cggatgatag acttttttaca cgatcagcac gacccactgc gctgcggcaa 2760
aggtcgaacc gaaacaagaa taaaccacga agatcagatc gattcgacgg aagaagcaat 2820
cgaatgcaaa gaagaatcgg aacgaagaaa actctaaagc atcgcatatt tacaaagcat 2880
aacggaaaac ccgcaagttc aaactagtga ttagtgtaag atgaagcaaa gcagaaatgt 2940
agtatctaga tttttcgacg ttagtttaca aagataaaaa atgaggttgg acatacaatc 3000
gtgggtattc gtctgagttc gtcacaactg caccggaaac tgtgaaacag aatagagcca 3060
acctgtgcgc ggagaatgtt gaggtcatta taagcttcct tagcatccac gggtgaaagt 3120
cgatcgacgg aagcctgcaa gactctgtcg atgggctttc gtcctagaag aataagatta 3180
aacctgaaat gtattctccc gtggaatggt ttcatttgag taattctgta tcttctcctt 3240
cccaattcca cgaacgcgac gaactctaat acaaacaaca taatgaccac agtgcaaatg 3300
ctgtttaacg ataatagcga catgcagcca ttctgggggct accacgtgta gctctacttg 3360
tgagacagcg ttcctaaaga gtgtgaaagt gcaaacaagt gatgaaacca atagtgcaaa 3420
gcaagtttag agggaaaatt taaaaaatgc aaaacagcag tagtacttaa cttttaagat 3480
tgtgtttcga aagccgaagt gaggctgttc catctgccac cggaaaaaaa cgacgacagc 3540
agaatcatca acaagcaaca tccatccgaa aaaatccggg aaaccggatc ttcaaccaac 3600
catcctacaa tctacaaacc agagattata tctcttcaat cgtttccgac atcggtcggt 3660
ttcggtgccc aaaatgatct gataaacact tatctctctg tagcttgcat gccattgcga 3720
gcgtattttg gtagctggcc gttgccaaac ggctccgaca ggtactgcta ttggaggttg 3780

tgcacgacca cgttgagttt gccttttgag ttggagagtg tgtcttttcg tcatatattc 3840
ggcctttca agggtgattt tcaggctacg taatgattgt atagtttaac cagctaaaac 3900
atattgatga caagttctat ttcagcacca caaacaagcc tgttaatgtc tctcaccgca 3960
accattgttc tgcgcgcgtt ataatcagca tagaagttta ttttctttgg gatgattcaa 4020
atattacgtg acgcaaagtt tgccaatttt agaacccctc cctcctccac gtaacggctt 4080
ttgtgtgaaa aatttaaatt ttgtgtatag accgtagcat ttcggaagac cccctccctt 4140
actctgttga gttacgtaaa atttcaacga tccttttgta gttctgaatt ttatatcagc 4200
gtgcagtgtt atgaagatat ccacagtata aaatattatt ttattttaaa ttctatgctg 4260
attatcaatg tgttactagt ggcttttcat actcatgttg cgagctcgat ttggcgcacg 4320
gtacttatca aggcatgtat gtatgttgtt tgaagcaact gtataactgt ttgaaactat 4380
ctaattggtg agctcgtttc atttagtata taataatgat aattgctatg gagacgttat 4440
ttactagcaa gtgatttgac gacctgaaat cggaacaaat agacaacgtt tttataaata 4500
caataaatca gaactttcca ttattgggta caaagagttg cgctatttcg atactgtcag 4560
atcagatttt ccagcacaac gataccttga tatgcgataa cttagaatta gaccttcaaa 4620
tccatctctc cagctatgaa cagtcatata gataaagcca atggcgttat gaggtagcgg 4680
aaagcgtcat ctttccaatg ctatctaagt acataatttg ctatagcttt ctattaatcg 4740
tagtttgaga gatgcaaagt cagttatctc gtatcaaggt ttgattgttt tggaaattag 4800
ctaaacagtt gacattatca cccgtcttta ggggataagc gcatacaaat gtgtatttag 4860
ttgttcattg aagtaacgta agataggcaa gtatggaaac gagctcacca aacgtcgaaa 4920
tacgtctaat aaatttgtgt tcagcaggat ggttcaaaat ttatttgcat cacctcaaaa 4980
ttacagtacc tagtgctgtt tgtgacaaac atcaaaaggt aaaatcaaac tcgtggcgtc 5040
gtgcaatctc catagaatga acaatttcta accgtatttg atggaaagac attgagtcta 5100
ctatcctctt aacagcattg cacttgtcta taaacaataa ataatttgtt ctttttaca 5160
ttttctttcc ccactttcgc ccccccccc cccaaaaatc aatccctcaa acaggatacg 5220
acatttgttg catctacttt ccgaagcgtt ccagcagaca cagacactgg ccggacgagg 5280
agaacatctc cgtcacccgc actccgtctg cgtcacggtc gccatgtgcc gattttcgta 5340
cccggtcaca gtccagctcg ccggataaca acggtggcgc gctcaatctg gacacgaaat 5400
ctaccaaagc gacgaccgcc accaccgacg acgaagaggt tatgtacgag aaacgcagcc 5460
cgaagtccat tgaatctacc gagttgcggt gccgtctgga ggaagcctta cacagtggcg 5520
ctgctgctgc tgcggctgct gaagaacctc tggcgggcgg aagcggttcc cactggaaga 5580
gagaaagttt cggctctacg gaggagattc ccactcgacc cgctcacagt gaaccggaag 5640
ataatggatt tgaaaacgga ttggaagcgc accagtccca tattctgcac agcatacatc 5700
ggaatgtgca acgatcataa tcagccatac cacatttgta gaggttttac ttgctttaaa 5760
aaacctccca cacctccccc tgaacctgaa acataaaatg aatgcaattg ttgttgttaa 5820
cttgtttatt gcagcttata atggttacaa ataaagcaat agcatcacaa atttcacaaa 5880
taaagcattt ttttcactgc attctagttg tggtttgtcc aaactcatca atgtatctta 5940
gggccgccac cgcggtggag ctccagcttt tgttcccttt agtgagggtt aattgcgcgc 6000
ttggcgtaat catggtcata gctgtttcct gtgtgaaatt gttatccgct cacaattcca 6060
cacaacatac gagccggaag cataaagtgt aaagcctggg gtgcctaatg agtgagctaa 6120
ctcacattaa ttgcgttgcg ctcactgccc gctttccagt cgggaaacct gtcgtgccag 6180
ctgcattaat gaatcggcca acgcgcgggg agaggcggtt tgcgtattgg gcgctcttcc 6240
gcttcctcgc tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct 6300
cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg 6360
tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc 6420
cataggctcc gcccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga 6480
aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct 6540
cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg 6600
gcgctttctc atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag 6660
ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttatc cggtaactat 6720

cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac 6780
aggattagca gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtggcctaac 6840
tacggctaca ctagaaggac agtatttggt atctgcgctc tgctgaagcc agttaccttc 6900
ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt 6960
tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc 7020
ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttggtcatg 7080
agattatcaa aaaggatctt cacctagatc cttttaaatt aaaaatgaag ttttaaatca 7140
atctaaagta tatatgagta aacttggtct gacagttacc aatgcttaat cagtgaggca 7200
cctatctcag cgatctgtct atttcgttca tccatagttg cctgactccc cgtcgtgtag 7260
ataactacga tacgggaggg cttaccatct ggccccagtg ctgcaatgat accgcgagac 7320
ccacgctcac cggctccaga tttatcagca ataaaccagc cagccggaag ggccgagcgc 7380
agaagtggtc ctgcaacttt atccgcctcc atccagtcta ttaattgttg ccgggaagct 7440
agagtaagta gttcgccagt taatagtttg cgcaacgttg ttgccattgc tacaggcatc 7500
gtggtgtcac gctcgtcgtt tggtatggct tcattcagct ccggttccca acgatcaagg 7560
cgagttacat gatcccccat gttgtgcaaa aaagcggtta gctccttcgg tcctccgatc 7620
gttgtcagaa gtaagttggc cgcagtgtta tcactcatgg ttatggcagc actgcataat 7680
tctcttactg tcatgccatc cgtaagatgc ttttctgtga ctggtgagta ctcaaccaag 7740
tcattctgag aatagtgtat gcggcgaccg agttgctctt gcccggcgtc aatacgggat 7800
aataccgcgc cacatagcag aactttaaaa gtgctcatca ttggaaaacg ttcttcgggg 7860
cgaaaactct caaggatctt accgctgttg agatccagtt cgatgtaacc cactcgtgca 7920
cccaactgat cttcagcatc ttttactttc accagcgttt ctgggtgagc aaaaacagga 7980
aggcaaaatg ccgcaaaaaa gggaataagg gcgacacgga aatgttgaat actcatactc 8040
ttcctttttc aatattattg aagcatttat cagggttatt gtctcatgag cggatacata 8100
tttgaatgta tttagaaaaa taaacaaata ggggttccgc gcacatttcc ccgaaaagtg 8160
ccacctaaat tgtaagcgtt aatattttgt taaaattcgc gttaaatttt tgttaaatca 8220
gctcattttt taaccaatag gccgaaatcg gcaaaatccc ttataaatca aaagaataga 8280
ccgagatagg gttgagtgtt gttccagttt ggaacaagag tccactatta aagaacgtgg 8340
actccaacgt caaagggcga aaaaccgtct atcagggcga tggcccacta cgtgaaccat 8400
caccctaatc aagttttttg gggtcgaggt gccgtaaagc actaaatcgg aaccctaaag 8460
ggagcccccg atttagagct tgacggggaa agccggcgaa cgtggcgaga aaggaaggga 8520
agaaagcgaa aggagcgggc gctagggcgc tggcaagtgt agcggtcacg ctgcgcgtaa 8580
ccaccacacc cgccgcgctt aatgcgccgc tacagggcgc gtcccattcg ccattcaggc 8640
tgcgcaactg ttgggaaggg cgatcggtgc gggcctcttc gctattacgc cagctggcga 8700
aaggggatg tgctgcaagg cgattaagtt gggtaacgcc agggttttcc cagtcacgac 8760
gttgtaaaac gacggccagt gagcgcgcgt aatacgactc actatagggc gaattgggta 8820
ccgggcccaa gcttatcgat accgtcgaca tgcccgccgt gaccgtcgag aacccgctga 8880
cgctgccccg cgtatccgca cccgccgacg ccgtcgcacg tcccgtgctc accgtgacca 8940
ccgcgcccag cggtttcgag ggcgagggct tcccggtgcg ccgcgcgttc gccgggatca 9000
actaccgcca cctcgacccg ttcatcatga tggaccagat gggtgaggtg gagtacgcgc 9060
ccggggagcc caagggcacg ccctggcacc cgcaccgcgg cttcgagacc gtgacctaca 9120
tcgtcgacct cgagggggag ccccccctcg aggttcccac aatggttaat tcgagctcgc 9180
ccggggatct aattcaatta gagactaatt caattagagc taattcaatt aggatccaag 9240
cttatcgatt tcgaaccctc gaccgccgga gtataaatag aggcgcttcg tctacggagc 9300
gacaattcaa ttcaaacaag caaagtgaac acgtcgctaa gcgaaagcta agcaaataaa 9360
caagcgcagc tgaacaagct aaacaatcgg ggtaccgcta gagtcgatcc caccccaccc 9420
aagaagaagc gcaaaccggt cgccaccatg gcctcctccg agaacgtcat caccgagttc 9480
atgcgcttca aggtgcgcat ggagggcacc gtgaacggcc acgagttcga gatcgagggc 9540
gagggcgagg gccgccccta cgagggccac aacaccgtga agctgaaggt gaccaagggc 9600
ggcccccctgc ccttcgcctg ggacatcctg tcccccccagt tccagtacgg ctccaaggtg 9660

tacgtgaagc accccgccga catccccgac tacaagaagc tgtccttccc cgagggcttc 9720
aagtgggagc gcgtgatgaa cttcgaggac ggcggcgtgg cgaccgtgac ccaggactcc 9780
tccctgcagg acggctgctt catctacaag gtgaagttca tcggcgtgaa cttcccctcc 9840
gacggccccg tgatgcagaa gaagaccatg ggctgggagg cctccaccga gcgcctgtac 9900
ccccgcgacg gcgtgctgaa gggcgagacc cacaaggccc tgaagctgaa ggacggcggc 9960
cactacctgg tggagttcaa gtccatctac atggccaaga agcccgtgca gctgcccggc 10020
tactactacg tggacgccaa gctggacatc acctcccaca acgaggacta caccatcgtg 10080
gagcagtacg agcgcaccga gggccgccac cacctgttcc tgagatctcg acccaagaaa 10140
aagcggaagg tggaggaccc gtaagatcca ccggatctag ataactgatc ataatcagcc 10200
ataccacatt tgtagaggtt ttacttgctt taaaaaacct cccacacctc cccctgaacc 10260
tgaaacataa aatgaatgca attgttgttg ttaacttgtt tattgcagct tataatggtt 10320
acaaataaag caatagcatc acaaatttca caaataaagc attttttttca ctgcattcta 10380
gttgtggttt gtccaaactc atcaatgtat cttatcatgt ctggatcccg tttgacggta 10440
tcgataagct tgatggggat ccggaaccct taattaccgt tcgtataatg tatgctatac 10500
gaagttatta ggtccctcga cctgcagccc gggggatcca 10540

<210> 154
<211> 4446
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3515
<400> 154
ggccgccacc gcggtggagc tccagctttt gttcccttta gtgagggtta attgcgcgct 60
tggcgtaatc atggtcatag ctgtttcctg tgtgaaattg ttatccgctc acaattccac 120
acaacatacg agccggaagc ataaagtgta aagcctgggg tgcctaatga gtgagctaac 180
tcacattaat tgcgttgcgc tcactgcccg ctttccagtc gggaaacctg tcgtgccagc 240
tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg cgctcttccg 300
cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc 360
actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt 420
gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc 480
ataggctccg cccccctgac gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa 540
acccgacagg actataaaga taccaggcgt ttcccctgg aagctccctc gtgcgctctc 600
ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg ggaagcgtgg 660
cgctttctca tagctcacgc tgtaggtatc tcagttcggt gtaggtcgtt cgctccaagc 720
tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc ggtaactatc 780
gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc actggtaaca 840
ggattagcag agcgaggtat gtaggcggtg ctacagagtt cttgaagtgg tggcctaact 900
acggctacac tagaaggaca gtatttggta tctgcgctct gctgaagcca gttaccttcg 960
gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc ggtggttttt 1020
ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat cctttgatct 1080
tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt ttggtcatga 1140
gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt tttaaatcaa 1200
tctaaagtat atatgagtaa acttggtctg acagttacca atgcttaatc agtgaggcac 1260
ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc gtcgtgtaga 1320
taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata ccgcgagacc 1380
cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg gccgagcgca 1440
gaagtggtcc tgcaacttta tccgcctcca tccagtctat taattgttgc cgggaagcta 1500
gagtaagtag ttcgccagtt aatagtttgc gcaacgttgt tgccattgct acaggcatcg 1560

tggtgtcacg ctcgtcgttt ggtatggctt cattcagctc cggttcccaa cgatcaaggc 1620
gagttacatg atccccccatg ttgtgcaaaa aagcggttag ctccttcggt cctccgatcg 1680
ttgtcagaag taagttggcc gcagtgttat cactcatggt tatggcagca ctgcataatt 1740
ctcttactgt catgccatcc gtaagatgct tttctgtgac tggtgagtac tcaaccaagt 1800
cattctgaga atagtgtatg cggcgaccga gttgctcttg cccggcgtca atacgggata 1860
ataccgcgcc acatagcaga actttaaaag tgctcatcat tggaaaacgt tcttcggggc 1920
gaaaactctc aaggatctta ccgctgttga gatccagttc gatgtaaccc actcgtgcac 1980
ccaactgatc ttcagcatct tttactttca ccagcgtttc tgggtgagca aaaacaggaa 2040
ggcaaaatgc cgcaaaaaag ggaataaggg cgacacggaa atgttgaata ctcatactct 2100
tcctttttca atattattga agcatttatc agggttattg tctcatgagc ggatacatat 2160
ttgaatgtat ttagaaaaat aaacaaatag gggttccgcg cacatttccc cgaaaagtgc 2220
cacctaaatt gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag 2280
ctcatttttt aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac 2340
cgagataggg ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga 2400
ctccaacgtc aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc 2460
accctaatca agtttttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg 2520
gagcccccga tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa 2580
gaaagcgaaa ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac 2640
caccacaccc gccgcgctta atgcgccgct acagggcgcg tcccattcgc cattcaggct 2700
gcgcaactgt tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa 2760
aggggatgt gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg 2820
ttgtaaaacg acggccagtg agcgcgcgta atacgactca ctatagggcg aattgggtac 2880
cgggcccccc ctcgaggtcg acgatgtagg tcacggtctc gaagccgcgg tgcgggtgcc 2940
agggcgtgcc cttgggctcc ccgggcgcgt actccacctc acccatctgg tccatcatga 3000
tgaacgggtc gaggtggcgg tagttgatcc cggcgaacgc gcggcgcacc gggaagccct 3060
cgccctcgaa accgctgggc gcggtggtca cggtgagcac gggacgtgcg acggcgtcgg 3120
cgggtgcgga tacgcggggc agcgtcagcg ggttctcgac ggtcacggcg ggcatgtcga 3180
cggtatcgat aagcttgggc ccccctcga ggttcccaca atggttaatt cgagctcgcc 3240
cggggatcta attcaattag agactaattc aattagagct aattcaatta ggatccaagc 3300
ttatcgattt cgaaccctcg accgccggag tataaataga ggcgcttcgt ctacggagcg 3360
acaattcaat tcaaacaagc aaagtgaaca cgtcgctaag cgaaagctaa gcaaataaac 3420
aagcgcagct gaacaagcta aacaatcggg gtaccgctag agtcgatccc accccaccca 3480
agaagaagcg caaaccggta ccatggcctc ctccgagaac gtcatcaccg agttcatgcg 3540
cttcaaggtg cgcatggagg gcaccgtgaa cggccacgag ttcgagatcg agggcgaggg 3600
cgagggccgc ccctacgagg gccacaacac cgtgaagctg aaggtgacca agggcggccc 3660
cctgcccttc gcctgggaca tcctgtcccc ccagttccag tacggctcca aggtgtacgt 3720
gaagcacccc gccgacatcc ccgactacaa gaagctgtcc ttccccgagg gcttcaagtg 3780
ggagcgcgtg atgaacttcg aggacggcgg cgtggcgacc gtgacccagg actcctccct 3840
gcaggacggc tgcttcatct acaaggtgaa gttcatcggc gtgaacttcc cctccgacgg 3900
ccccgtgatg cagaagaaga ccatgggctg ggaggcctcc accgagcgcc tgtacccccg 3960
cgacggcgtg ctgaagggcg agacccacaa ggccctgaag ctgaaggacg gcggccacta 4020
cctggtggag ttcaagtcca tctacatggc caagaagccc gtgcagctgc ccggctacta 4080
ctacgtggac gccaagctgg acatcacctc ccacaacgag gactacacca tcgtggagca 4140
gtacgagcgc accgagggcc gccaccacct gttcctgtga tgatcataat cagccatacc 4200
acatttgtag aggttttact tgctttaaaa aacctcccac acctcccccct gaacctgaaa 4260
cataaaatga atgcaattgt tgttgttaac ttgtttattg cagcttataa tggttacaaa 4320
taaagcaata gcatcacaaa tttcacaaat aaagcatttt tttcactgca ttctagttgt 4380
ggtttgtcca aactcatcaa tgtatcttaa cgcgagttaa ttaaggccgc tcatttaaat 4440
ctggcc 4446

<210> 155
<211> 12991
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3545
<400> 155

gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt taccttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900
aacgttcgag gtcgactcta gcactgggaa gttgacgttg atatagagcc gaattgaact 960
tcaccgctgc ttggtaatta ctctacaagt tcatttagga gaaccggatt cgaaagatga 1020
ttttccagcg tttagctttc agatggccgc atacattttg caccaccaaa ccgaaactca 1080
ctagcgtatc caatcgttcg ttttttggtg ccggtgtgtt acgaacttta gctatcaagc 1140
taaagcaatt tgctctggtc ttccgtgcta aaaagaaaaa aaaactgttt ttttttggt 1200
tttgatattt gcgctatttt tacttgggcc ttaattgaac aaacttttga aagtttccac 1260
agcgaaatcg ttttcgacga tgccattttt ggtaacattt gcattttctt gctcaaattg 1320
cttgcaaaac ccgtgaaaga cattaatatt cgatagtgtc atccaaaatc acgaaaatga 1380
ttgttgcaaa acgttgaaca atttacacat gtaaaaaaca accatcgatt aatgtttatt 1440
caaacttttt acaagaaggg ttattctgat caatgtcacc ccgctgatga atgttacccc 1500
ggattacact tctcgaaaag tggttcaaaa tgctacttga gaatttttat ctgtcaaagg 1560
aagcaaattc gagtcgaatt aaatggtata gtcctgaatt aggtttccat ttacttacag 1620
gtattccact aaatagctgg aagatttatt ttacacaata atgataattc gtaccccaaa 1680
gagtgtagcc ctacttttttt ctctcttttt tttttgtaaa ttttcatcgc tgcgtgccag 1740
cttaccgaca tgtcgcgaca gcataaagag cctgtcaaga gatgaagaaa aatgacaagg 1800
agtcagtggt caggtctctg tatcaatatt tgacgtcctg actttccaat atacctttcc 1860
ttaaagagta gagatcatgc gatacgtgaa taaatatcgt ttggacttcg aaatagaaca 1920
taatttaagg tagctgatca gtagttgaac atcttcagac ttctgggaca agaagtgttt 1980
ttttgtttgt agaaaaggtt tttgttaaat tatatttgta agataattca atgaatatat 2040
ctctgattca gtaatcaatc cgtaccacgc accgtttaag aaacaccctg taggtttgca 2100
tcacgtctca gacaaaagtg tatcgatgtg cgaacactgc ataccggcgc tttgcaaata 2160
atgccaaatt tagatatgca ttacattgtc acttcgcaaa acacacactc ccaaatgcgt 2220
cggaaacctc acccgaacgc acgatcgtaa cgcgatcgat cgccgattga ttgatcggaa 2280
ttaactatct caatcgatcc ttctatggac tgatgcatgg gccggcactt ccgagtataa 2340
aaccccggta aacccaagga atcactcaca atcggatttt gacgctcgct ctggtacagt 2400
tcgatacggt ctagtgaaac cgaggataac gacgaaggtt tttccccatt gatccaggtc 2460

ggtgtttatg attggtggaa aaagactcga gaaaagttcc atcgaagccg ttggaaatgt 252C
gccgtcttcc tgtgacgtct tgtggatcca gttccttgtt cacgtctggt gatcgtgtaa 258C
aatgtgctgt cttgtggcgt catatgtgtt ccagatccag tgattacgat ccgatgtgat 264C
gttgatccct tgtgaacgtc ttatcctgtt ccgtgtgcac catgcataat gtcgtattac 2700
gtaagttctg aagtgaaaca gaagagtgaa ttgaaagttt ttttattcaa catcaaccta 2760
aatatggact ttactttcca agaaaattat gcctgatcaa ctgtggatag ttacaaaaaa 2820
aaaaggttta ttaattaaat tttatgatta cataatgtgt tgaaaagaac aactgaaatt 2880
ttagaagaag atcttttcgt gcatcaggct ttgccaatta attgatgata aattatcata 2940
gcaaattaac gtagagacta aaaggtatat cgtcaaatag ggcttctttt gacactattt 3000
tggcattctt gctctttgag aacttgcaac cctaaaatgg gatcttcatc agcctagtgg 3060
ttagattcag cagctacaaa gcaaaaccat gctgaagggt tcgattcccg gtcgtttcag 3120
gatcttttcg taattgaaat atccttgact accctaagta tcattgtgct tgccatttac 3180
gaatatacat attacgatat acgaatgaga aaatgacaac tttggaaaat aaagctctca 3240
atgtttcaat aagaaataaa tactacatca gtattgaagg ctaataacaa ttacagatta 3300
gaacctttaa acatcatttc tgcaacaggc tggataaagt acagttggag gattaaatta 3360
tgcgattttg caatttttc cgattaaatt catatttatt cctggtttgg tttttacaaa 3420
aaatattttt acatgacgtt tgaccccgat tccctcaact ttgattgtta tatttttttt 3480
tggacaggtt gagtttgtgg gttttttcct agtgttgctt tgctttatgg gctctggtta 3540
tttaaaatta aaatttgaca atcttactac acactccgaa aaaatcatgc gattttacgt 3600
cttttggatg cacataaaag aagcgagcca aatgaggtga atttgtgtca cattttaaat 3660
acgatggtgt ctgattcggg aaatgtcaat gatagtgtca ttcaatcata atgtgaatta 3720
cgtccgcagt aattttcatt atttttaaga gtgtactact atttacacta caaaaatttt 3780
gataccccag gggggaacga ggtcccggat gtccagctgg ccagattgtt ggcaacgagc 3840
cctgtaccta ttgatcgagt caccaaagca ctcctcaagt gttttaatct cgaccagacg 3900
gtggacctcg gttgttctca ttctcggagg gcgatttcgc aatcattagt accaaccaca 3960
tgtcgaagtc gggagatgtt ataaaattat aaccaattat tcaaaaaatg acatcattca 4020
atttgaacaa acgttcgata gaaattatat atgatttcac atgatattaa actacgaaga 4080
aaattttaca taaggaagtg gtataaaacg taatatgctt aataaaaact ttaacccttt 4140
tgggaggata atattcagaa gttctgattc agaaccatct ctcatgttat gttcgttttt 4200
tgttgcttgt cctttatatg ccacatgaac aataacacca atatctatcc catttccagg 4260
acctaacgga ccttgaagcg gcgccactag taaaccacca tgggcagccg cctggataag 4320
tccaaagtca tcaactccgc gttggagctg ttgaacgaag ttggcattga gggactgacg 4380
acccgcaagt tggcgcagaa gctgggcgtg gagcagccca ccctctactg gcacgtgaag 4440
aataagcggg cgctgctgga tgccctggcc atcgagatgc tcgaccgcca ccacacgcat 4500
ttttgcccgt tggaaggcga gtcctggcag gacttcctcc gcaataacgc caagtcgttc 4560
cgctgcgctc tgctgtccca ccgagacggt gccaaagtcc atctcggcac gcgcccgacc 4620
gaaaagcaat acgagacact ggagaaccag ctcgcgttcc tgtgccagca aggcttcagc 4680
ctggaaaatg ctctctacgc tctgagcgcc gtcggtcact ttaccctggg ctgcgtgctg 4740
gaggaccaag agcatcaagt cgcaaaagag gagcgcgaga ccccaacaac cgattcgatg 4800
cccccactgc tgcgtcaggc aatcgagctg ttcgatcatc aaggagccga gccggcattc 4860
ctgttcggct tggagctgat tatctgcgga ttggaaaagc aactgaaatg cgagtcgggc 4920
tcgggccccg cctacagccg cgcccgcacc aagaacaact acggcagcac catcgagggc 4980
ctgctggatc tgccggatga tgatgccccg gaggaggcgg gcctggccgc cccgcgcctg 5040
agcttcctgc cggccggaca cacccgccgc ctgtcgaccg ccccgccgac cgacgtgagc 5100
ctgggcgatg agctgcacct ggatggcgag gatgtggcga tggcccacgc cgatgccctg 5160
gacgacttcg acctggacat gctgggcgat ggcgatagcc cgggaccggg attcaccccg 5220
cacgatagcg ccccctacgg cgccctggat atggccgatt tcgagttcga gcagatgttc 5280
accgacgccc tgggcatcga tgagtacggc ggctaacacc ggaaactcgc gttaagatac 5340
attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa 5400

atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca agttaacaac 5460
aacaattgca ttcattttat gtttcaggtt caggggggagg tgtgggaggt tttttaaagc 5520
aagtaaaacc tctacaaatg tggtatggct gattatgatc agttatctag atccggtgga 5580
tcttacgggt cctccacctt ccgcttttttc ttgggtcgag atctcaggaa caggtggtgg 5640
cggccctcgg tgcgctcgta ctgctccacg atggtgtagt cctcgttgtg ggaggtgatg 5700
tccagcttgg cgtccacgta gtagtagccg ggcagctgca cgggcttctt ggccatgtag 5760
atggacttga actccaccag gtagtggccg ccgtccttca gcttcagggc cttgtgggtc 5820
tcgcccttca gcacgccgtc gcgggggtac aggcgctcgg tggaggcctc ccagcccatg 5880
gtcttcttct gcatcacggg gccgtcggag gggaagttca cgccgatgaa cttcaccttg 5940
tagatgaagc agccgtcctg cagggaggag tcctgggtca cggtcgccac gccgccgtcc 6000
tcgaagttca tcacgcgctc ccacttgaag ccctcgggga aggacagctt cttgtagtcg 6060
gggatgtcgg cggggtgctt cacgtacacc ttggagccgt actggaactg gggggacagg 6120
atgtcccagg cgaagggcag gggccgccc ttggtcacct tcagcttcac ggtgttgtgg 6180
ccctcgtagg ggcggccctc gccctcgccc tcgatctcga actcgtggcc gttcacggtg 6240
ccctccatgc gcaccttgaa gcgcatgaac tcggtgatga cgttctcgga ggaggccatg 6300
gtggcgaccg gtttgcgctt cttcttgggt ggggtgggat ctcccatggt ggcctgaatc 6360
tcaacttgca cctgaaggta gtgcagcaag gatgagcaaa agggaagaac ccagaaaaga 6420
acgggaaaac ttaccccaat tagaattgct tgtcgccgcc agtgtcaact tgcaactgaa 6480
acaatatcca acatgaacgt caatttatac tgccctaatg gcgaacacga taacaatatt 6540
tcttttatta tgccctctaa aaccaacgcg gttatcgttt atttattcaa attagatata 6600
gaacatccgc cgacatacaa tgttaatgca aaaacgcgtt tggtgagcgg atacgaaaac 6660
agtcggccga taaacattaa tctgaggtcg ataacaccgt ccttgaacgg aacacgagga 6720
gcgtacgtga tcagctgcat tcgcgcgccg cgcctttatc gagatttatt tgcatacaac 6780
aagtacactg cgccgttggg atttgtggta acgcgcacac atgcagagct gcaagtgtgg 6840
cacattttgt ctgtgcgcaa aacctttgaa gccaaaagta cgaggtccgt tacgggcatg 6900
ctagcgcaca cggacaatgg acccgacaaa ttctacgcca aggatttaat gataatgtcg 6960
ggcaacgtat ccgttcattt tatcaataac ctacaaaaat gtcgcgcgca tcacaaagac 7020
atcgatatat ttaaacattt atgtcccgaa ctgcaaatcg ataatagtgt tgtgcaacct 7080
cgagcgtccg tttgatttaa cgtatagctt gcaaatgaat tatttaatta tcaatcatgt 7140
tttacgcgta gaattctacc cgtaaagcga gtttagttat gagccatgtg caaaacatga 7200
catcagcttt tatttttata acaaatgaca tcatttcttg attgtgtttt acacgtagaa 7260
ttctactcgt aaagcgagtt cagttttgaa aaacaaatga catcatcttt ttgattgtgc 7320
tttacaagta gaattctacc cgtaaatcaa gttcggtttt gaaaaacaaa tgagtcatat 7380
tgtatgatat catattgcaa aacaaatgac tcatcaatcg atcgtgcgtt acacgtagaa 7440
ttctactcgt aaagcgagtt tatgagccgt gtgcaaaaca tgacatcatc tcgatttgaa 7500
aaacaaatga catcatccac tgatcgtgca ttacaagtag aattctactc gtaaagccag 7560
ttcggttatg agccgtgtac aaaacatgac atcagattat gactcatact tgattgtgtt 7620
ttacgcgtag aattctactc gtaaagccag ttcaatttta aaaacaaatg acatcatcca 7680
aattaataaa tgacaagcaa tgggtaccat gcggccgctc atttaaatct ggccggcctg 7740
gccgatctga caatgttcag tgcagagact cggctacgcc tcgtggactt tgaagttgac 7800
caacaatgtt tattcttacc tctaatagtc ctctgtggca aggtcaagat tctgttagaa 7860
gccaatgaag aacctggttg ttcaataaca ttttgttcgt ctaatatttc actaccgctt 7920
gacgttggct gcacttcatg tacctcatct ataaacgctt cttctgtatc gctctggacg 7980
tcatcttcac ttacgtgatc tgatatttca ctgtcagaat cctcaccaac aagctcgtca 8040
tcgctttgca gaagagcaga gaggatatgc tcatcgtcta aagaactacc cattttatta 8100
tatattagtc acgatatcta taacaagaaa atatatatat aataagttat cacgtaagta 8160
gaacatgaaa taacaatata attatcgtat gagttaaatc ttaaaagtca cgtaaaagat 8220
aatcatgcgt cattttgact cacgcggtcg ttatagttca aaatcagtga cacttaccgc 8280
attgacaagc acgcctcacg ggagctccaa gcggcgactg agatgtccta aatgcacagc 8340

gacggattcg cgctatttag aaagagagag caatatttca agaatgcatg cgtcaatttt 8400
acgcagacta tctttctagg gttaaaaaag atttgcgctt tactcgacct aaactttaaa 8460
cacgtcatag aatcttcgtt tgacaaaaac cacattgtgg ccaagctgtg tgacgcgacg 8520
cgcgctaaag aatggcaaac caagtcgcgc gagcgtcgac ctgcaggcat gcaagcttgc 8580
atgcctgcag gtcgaaattc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat 8640
ccgctcacaa ttccacacaa catacgagcc ggaagcataa agtgtaaagc ctggggtgcc 8700
taatgagtga gctaactcac attaattgcg ttgcgctcac tgcccgcttt ccagtcggga 8760
aacctgtcgt gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt 8820
attgggcgct cttccgcttc ctcgctcact gactcgctgc gctcggtcgt tcggctgcgg 8880
cgagcggtat cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac 8940
gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg 9000
ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca 9060
agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc cccctggaagc 9120
tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc 9180
ccttcgggaa gcgtggcgct ttctcaatgc tcacgctgta ggtatctcag ttcggtgtag 9240
gtcgttcgct ccaagctggg ctgtgtgcac gaacccccgc ttcagcccga ccgctgcgcc 9300
ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca 9360
gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg 9420
aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg 9480
aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct 9540
ggtagcggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa 9600
gaagatcctt tgatcttttc tacggggtct gacgctcagt ggaacgaaaa ctcacgttaa 9660
gggattttgg tcatgagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa 9720
tgaagtttta aatcaatcta aagtatatat gagtaaactt ggtctgacag ttaccaatgc 9780
ttaatcagtg aggcacctat ctcagcgatc tgtctatttc gttcatccat agttgcctga 9840
ctccccgtcg tgtagataac tacgatacgg gagggcttac catctggccc cagtgctgca 9900
atgataccgc gagacccacg ctcaccggct ccagatttat cagcaataaa ccagccagcc 9960
ggaagggccg agcgcagaag tggtcctgca actttatccg cctccatcca gtctattaat 10020
tgttgccggg aagctagagt aagtagttcg ccagttaata gtttgcgcaa cgttgttgcc 10080
attgctacag gcatcgtggt gtcacgctcg tcgtttggta tggcttcatt cagctccggt 10140
tcccaacgat caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc ggttagctcc 10200
ttcggtcctc cgatcgttgt cagaagtaag ttggccgcag tgttatcact catggttatg 10260
gcagcactgc ataattctct tactgtcatg ccatccgtaa gatgcttttc tgtgactggt 10320
gagtactcaa ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg 10380
gcgtcaatac gggataatac cgcgccacat agcagaactt taaaagtgct catcattgga 10440
aaacgttctt cggggcgaaa actctcaagg atcttaccgc tgttgagatc cagttcgatg 10500
taacccactc gtgcacccaa ctgatcttca gcatctttta ctttcaccag cgtttctggg 10560
tgagcaaaaa caggaaggca aaatgccgca aaaaagggaa taagggcgac acggaaatgt 10620
tgaatactca tactcttcct ttttcaatat tattgaagca tttatcaggg ttattgtctc 10680
atgagcggat acatatttga atgtatttag aaaaataaac aaatagggggt tccgcgcaca 10740
tttccccgaa aagtgccacc tgacgtctaa gaaaccatta ttatcatgac attaacctat 10800
aaaaataggc gtatcacgag gccctttcgt ctcgcgcgtt tcggtgatga cggtgaaaac 10860
ctctgacaca tgcagctccc ggagacggtc acagcttgtc tgtaagcgga tgccgggagc 10920
agacaagccc gtcagggcgc gtcagcgggt gttggcgggt gtcggggctg gcttaactat 10980
gcggcatcag agcagattgt actgagagtg caccatatat gcggtgtgaa ataccgcaca 11040
gatgcgtaag gagaaaatac cgcatcaggc gccattcgcc attcaggctg cgcaactgtt 11100
gggaagggcg atcggtgcgg gcctcttcgc tattacgcca gctggcgaaa gggggatgtg 11160
ctgcaaggcg attaagttgg gtaacgccag ggttttccca gtcacgacgt tgtaaaacga 11220
cggccagtgc caagctttgt ttaaaatata acaaaattgt gatcccacaa aatgaagtgg 11280

ggcaaaatca aataattaat agtgtccgta aacttgttgg tcttcaactt tttgaggaac 11340
acgttggacg gcaaatccgt gactataaca caagttgatt taataatttt agccaacacg 11400
tcgggctgcg tgtttttttgc cgacgcgtct gtgtacacgt tgattaactg gtcgattaaa 11460
ctgttgaaat aatttaattt ttggttcttc tttaaatctg tgatgaaatt ttttaaaata 11520
actttaaatt cttcattggt aaaaaatgcc acgttttgca acttgtgagg gtctaatatg 11580
aggtcaaact cagtaggagt tttatccaaa aaagaaaaca tgattacgtc tgtacacgaa 11640
cgcgtattaa cgcagagtgc aaagtataag agggttaaaa aatatatttt acgcaccata 11700
tacgcatcgg gttgatatcg ttaatatgga tcaatttgaa cagttgatta acgtgtctct 11760
gctcaagtct ttgatcaaaa cgcaaatcga cgaaaatgtg tcggacaata tcaagtcgat 11820
gagcgaaaaa ctaaaaaggc tagaatacga caatctcaca gacagcgttg agatatacgg 11880
tattcacgac agcaggctga ataataaaaa aattagaaac tattatttaa ccctagaaag 11940
ataatcatat tgtgacgtac gttaaagata atcatgcgta aaattgacgc atgtgtttta 12000
tcggtctgta tatcgaggtt tatttattaa tttgaataga tattaagttt tattatattt 12060
acacttacat actaataata aattcaacaa acaatttatt tatgtttatt tatttattaa 12120
aaaaaaacaa aaactcaaaa tttcttctat aaagtaacaa aacttttaaa cattctctct 12180
tttacaaaaa taaacttatt ttgtacttta aaaacagtca tgttgtatta taaaataagt 12240
aattagctta acttatacat aatagaaaca aattatactt attagtcagt cagaaacaac 12300
tttggcacat atcaatatta tgctctcgac aaataacttt tttgcatttt ttgcacgatg 12360
catttgcctt tcgccttatt ttagaggggc agtaagtaca gtaagtacgt tttttcatta 12420
ctggctcttc agtactgtca tctgatgtac caggcacttc atttggcaaa atattagaga 12480
tattatcgcg caaatatctc ttcaaagtag gagcttctaa acgcttacgc ataaacgatg 12540
acgtcaggct catgtaaagg tttctcataa atttttttgcg actttggacc ttttctccct 12600
tgctactgac attatggctg tatataataa aagaatttat gcaggcaatg tttatcattc 12660
cgtacaataa tgccataggc cacctattcg tcttcctact gcaggtcatc acagaacaca 12720
tttggtctag cgtgtccact ccgcctttag tttgattata atacataacc atttgcggtt 12780
taccggtact ttcgttgata gaagcatcct catcacaaga tgataataag tataccatct 12840
tagctggctt cggtttatat gagacgagag taaggggtcc gtcaaaacaa aacatcgatg 12900
ttcccactgg cctggagcga ctgtttttca gtacttccgg tatctcgcgt ttgtttgatc 12960
gcacggttcc cacaatggtt gcggccagcc c 12991

<210> 156
<211> 18411
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3604
<400> 156
ttaaaatgaa tgtaagcact ttattaacga aatctttggg aatatttcgc tcatcagcat 60
tttatttgag caggagtccg agatgcccgg gcggcgcgaa actcccctgc aggataactt 120
cgtatagcat acattatacg aagttatcct agggaagttc ctatactttc tagagaatag 180
gaacttcgga ataggaactt cttcgaacgg ccaaaaaggc cggccggggc acgggcgccg 240
tttttcttga aatattgctc tctctttcta aatagcgcga atccgtcgct gtgcatttag 300
gacatctcag tcgccgcttg gagctcccaa acgcgccagt ggtagtacac agtactgtgg 360
gtgttcagtt tgaaatcctc ttgcttctcc attgtctcgg ttacctttgg tcaaatccat 420
gggttctatt gcctatatac tcttgcgatt accagtgatt gcgctattag ctattagatg 480
gattgttggc caaacttgtc gcttaagtgg ctgggaattg taaccgtagg cccgagtgta 540
atgatccccc ataaaaagtt ttcgcaatgc ctttatttttt tgttgcaaat ctctctttat 600
tctgcggtat tcttcattat tgcggggatg gggaaagtgt ttatatagaa gcaacttacg 660
attgaaccca aatgcacctg acaagcaagg tcaaagggcc agatttttaa atatattatt 720

tagtcttagg actctctatt tgcaattaaa ttactttgct acctgagggt taaatcttcc 780
ccattgataa taataattcc actatatgtt caattgggtt tcaccgcgct tagttacatg 840
acgagcccta atgagccgtc ggtggtctat aaactgtgcc ttacaaatac ttgcaactct 900
tctcgttttg aagtcagcag agttattgct aattgctaat tgctaattgc ttttaactga 960
tttcttcgaa attggtgcta tgtttatggc gctattaaca agtatgaatg tcaggtttaa 1020
ccaggggatg cttaattgtg ttctcaactt caaaggcaga aatgtttact cttgaccatg 1080
ggtttaggta taatgttatc aagctcctcg agttaacgtt acgttaacgt taacgttcga 1140
ggtcgactct agacaccggt gttagccgcc gtactcatcg atgcccaggg cgtcggtgaa 1200
catctgctcg aactcgaaat cggccatatc cagggcgccg taggggcgcc tatcgtgcgg 1260
ggtgaatccc ggtcccgggc tatcgccatc gcccagcatg tccaggtcga agtcgtccag 1320
ggcatcggcg tgggccatcg ccacatcctc gccatccagg tgcagctcat cgcccaggct 1380
cacgtcggtc ggcggggcgg tcgacaggcg gcgggtgtgt ccggccggca ggaagctcag 1440
gcgcggggcg gccaggcccg cctcctccgg ggcatcatca tccggcagat ccagcaggcc 1500
ctcgatggtg ctgccgtagt tgttcttggt gcgggcgcgg ctgtaggcgg ggcccgagcc 1560
cgactcgcat ttcagttgct tttccaatcc gcagataatc agctccaagc cgaacaggaa 1620
tgccggctcg gctccttgat gatcgaacag ctcgattgcc tgacgcagca gtgggggcat 1680
cgaatcggtt gttggggtct cgcgctcctc ttttgcgact tgatgctctt ggtcctccag 1740
cacgcagccc agggtaaagt gaccgacggc gctcagagcg tagagagcat tttccaggct 1800
gaagccttgc tggcacagga acgcgagctg gttctccagt gtctcgtatt gcttttcggt 1860
cgggcgcgtg ccgagatgga ctttggcacc gtctcggtgg gacagcagag cgcagcggaa 1920
cgacttggcg ttattgcgga ggaagtcctg ccaggactcg ccttccaacg ggcaaaaatg 1980
cgtgtggtgg cggtcgagca tctcgatggc cagggcatcc agcagcgccc gcttattctt 2040
cacgtgccag tagagggtgg gctgctccac gcccagcttc tgcgccaact tgcgggtcgt 2100
cagtccctca atgccaactt cgttcaacag ctccaacgcg gagttgatga ctttggactt 2160
atccaggcgg ctgaccatac caccgcgcag gcgcagcacc aggtgcaggg tgctctcctt 2220
ctggatgttg tagtcgctca gggtgcggcc atcctccagc tggcgtccgg cgaagatcag 2280
gcgctgctga tccggcggga tgccctcctt gtcctggatc ttggccttca cgttctcgat 2340
ggtatcgctc ggctccacct ccagggtgat ggtcttgccg gtcagggtct tgacgaagat 2400
ctgcatcgag ctagccgtca cacgttttgg cgccgcttca aggtccgtta ggtcctggaa 2460
atgggataga tattggtgtt attgttcatg tggcatataa aggacaagca acaaaaaacg 2520
aacataacat gagagatggt tctgaatcag aacttctgaa tattatcctc ccaaaagggt 2580
taaagttttt attaagcata ttacgtttta taccacttcc ttatgtaaaa ttttcttcgt 2640
agtttaatat catgtgaaat catatataat ttctatcgaa cgtttgttca aattgaatga 2700
tgtcattttt tgaataattg gttataattt tataacatct cccgacttcg acatgtggtt 2760
ggtactaatg attgcgaaat cgccctccga gaatgagaac aaccgaggtc caccgtctgg 2820
tcgagattaa aacacttgag gagtgctttg gtgactcgat caataggtac agggctcgtt 2880
gccaacaatc tggccagctg gacatccggg acctcgttcc cccctggggt atcaaaattt 2940
ttgtagtgta aatagtagta cactcttaaa aataatgaaa attactgcgg acgtaattca 3000
cattatgatt gaatgacact atcattgaca tttcccgaat cagacaccat cgtatttaaa 3060
atgtgacaca aattcacctc atttggctcg cttcttttat gtgcatccaa aagacgtaaa 3120
atcgcatgat ttttttcggag tgtgtagtaa gattgtcaaa ttttaatttt aaataaccag 3180
agcccataaa gcaaagcaac actaggaaaa aacccacaaa ctcaacctgt ccaaaaaaaa 3240
atataacaat caaagttgag ggaatcgggg tcaaacgtca tgtaaaaata ttttttgtaa 3300
aaaccaaacc aggaataaat atgaatttaa tcggaaaaaa ttgcaaaatc gcataattta 3360
atcctccaac tgtactttat ccagcctgtt gcagaaatga tgtttaaagg ttctaatctg 3420
taattgttat tagccttcaa tactgatgta gtatttattt cttattgaaa cattgagagc 3480
tttatttttcc aaagttgtca ttttctcatt cgtatatcgt aatatgtata ttcgtaaatg 3540
gcaagcacaa tgatactcag ggcagtcaag gatatttcaa ttacgaaaag atcctgaaac 3600
gaccgggaat cgaacccttc agcatggctt tgctttgtag ctgctgaatc taaccactag 3660

gctgatgaag atcccatttt agggttgcaa gttctcaaag agcaagaatg ccaaaatagt 3720
gtcaaaagaa gccctatttg acgatatacc ttttagtctc tacgttaatt tgctatgata 3780
atttatcatc aattaattgg caaagcctga tgcacgaaaa gatcttcttc taaaatttca 3840
gttgttcttt tcaacacatt atgtaatcat aaaatttata ataaaccttt tttttttgta 3900
actatccaca gttgatcagg cataatttte ttggaaagta aagtccatat ttaggttgat 3960
gttgaataaa aaaactttca attcactctt ctgtttcact tcagaactta cgtaatacga 4020
cattatgcat ggtgcacacg gaacaggata agacgttcac aagggatcaa catcacatcg 4080
gatcgtaatc actggatctg gaacacatat gacgccacaa gacagcacat tttacacgat 4140
caccagacgt gaacaaggaa ctggatccac aagacgtcac aggaagacgg cacatttcca 4200
acggcttcga tggaactttt ctcgagtctt tttccaccaa tcataaacac cgacctggat 4260
caatggggaa aaaccttcgt cgttatcctc ggtttccatg gtggcggtcc gtatcgaact 4320
gtaccagagc gagcgtcaaa atccgattgt gagtgattcc ttgggtttac cggggtttta 4380
tactcggaag tgccggccca tgcatcagtc catagaagga tcgattgaga tagttaattc 4440
cgatcaatca atcggcgatc gatcgcgtta cgatcgtgcg ttcgggtgag gtttccgacg 4500
catttgggag tgtgtgtttt gcgaagtgac aatgtaatgc atatctaaat ttggcattat 4560
ttgcaaagcg ccggtatgca gtgttcgcac atcgatacac ttttgtctga gacgtgatgc 4620
aaacctacag ggtgtttctt aaacggtgcg tggtacggat tgattactga atcagagata 4680
tattcattga attatcttac aaatataatt taacaaaaac cttttctaca aacaaaaaaa 4740
cacttcttgt cccagaagtc tgaagatgtt caactactga tcagctacct taaattatgt 4800
tctatttcga agtccaaacg atatttattc acgtatcgca tgatctctac tctttaagga 4860
aaggtatatt ggaaagtcag gacgtcaaat attgatacag agacctgacc actgactcct 4920
tgtcattttt cttcatctct tgacaggctc tttatgctgt cgcgacatgt cggtaagctg 4980
gcacgcagcg atgaaaattt acaaaaaaaa aagagagaaa aaagtagggc tacactcttt 5040
ggggtacgaa ttatcattat tgtgtaaaat aaatcttcca gctatttagt ggaatacctg 5100
taagtaaatg gaaacctaat tcaggactat accatttaat tcgactcgaa tttgcttcct 5160
ttgacagata aaaattctca agtagcattt tgaaccactt ttcgagaagt gtaatccggg 5220
gtaacattca tcagcggggt gacattgatc agaataaccc ttcttgtaaa aagtttgaat 5280
aaacattaat cgatggttgt tttttacatg tgtaaattgt tcaacgtttt gcaacaatca 5340
ttttcgtgat tttggatgac actatcgaat attaatgtct ttcacgggtt ttgcaagcaa 5400
tttgagcaag aaaatgcaaa tgttaccaaa aatggcatcg tcgaaaacga tttcgctgtg 5460
gaaactttca aaagtttgtt caattaaggc ccaagtaaaa atagcgcaaa tatcaaaacc 5520
aaaaaaaaaa cagttttttt ttcttttttag cacggaagac cagagcaaat tgctttagct 5580
tgatagctaa agttcgtaac acaccggcac caaaaaacga acgattggat acgctagcga 5640
gtttcggttt ggtggtgcaa aatgtatgcg gccatctgaa agctaaacgc tggaaaatca 5700
tctttcgaat ccggttctcc taaatgaact tgtagagtaa ttaccaagca gcggtgaagt 5760
tcaattcggc tctatatcaa cgtcaacttc ccagtgcgcg ccccggccat cgagaaagag 5820
agagagaaga gaagagagag aacattcgag aaagagagag agaagagaag agagagaaca 5880
tactccctat cagtgataga gaagtcccta tcagtgatag agatgtccct atcagtgata 5940
gagagttccc tatcagtgat agagacgtcc ctatcagtga tagagaagtc cctatcagtg 6000
atagagagat ccctatcagt gatagagatt tccctatcag tgatagagag gtccctatca 6060
gtgatagaga cttccctatc agtgatagag aaatccctat cagtgataga gacatcccta 6120
tcagtgatag agaactccct atcagtgata gagacctccc tatcagtgat agagatcgat 6180
gcggccgcga gcgccggagt ataaatagag gcgcttcgtc tacggagcga caattcaatt 6240
caaacaagca aagtgaacac gtcgctaagc gaaagctaag caaataaaca agcgcagctg 6300
aacaagctaa acaatctgca ggtaccctgg cggtaagttg atcaaaggaa acgcaaagtt 6360
ttcaagaaaa aacaaaacta atttgattta taacaccttt agaaagcggg gctagccacc 6420
atgggcagcg cctacagccg cgcccgtacc aagaacaact atggcagcac catcgaggga 6480
ctgctggacc tgccggatga cgatgccccg gaggaagccg gcctggccgc ccccgcctg 6540
agcttcctgc ccgccggaca cacgcgccgc ctgagcaccg ccccgccgac cgatgtgagc 6600

ctgggcgacg agctgcacct ggatggagag gatgtggcaa tggcccacgc cgacgccctg 6660
gacgatttcg acctggatat gctgggcgat ggagatagcc cgggaccggg cttcacgccc 6720
cacgatagcg ccccgtacgg cgccctggac atggccgact tcgagttcga gcaaatgttc 6780
accgacgcgc tgggcatcga tgagtatggc gggtaggttt aaactcgcgt taagatacat 6840
tgatgagttt ggacaaacca caactagaat gcagtgaaaa aaatgcttta tttgtgaaat 6900
ttgtgatgct attgctttat ttgtaaccat tataagctgc aataaacaag ttaacaacaa 6960
caattgcatt cattttatgt ttcaggttca gggggaggtg tgggaggttt tttaaagcaa 7020
gtaaaacctc tacaaatgtg gtatggctga ttatgatcag ttatctagat ccggtggatc 7080
ttacgggtcc tccaccttcc gcttttttctt gggtcgagat ctcaggaaca ggtggtggcg 7140
gccctcggtg cgctcgtact gctccacgat ggtgtagtcc tcgttgtggg aggtgatgtc 7200
cagcttggcg tccacgtagt agtagccggg cagctgcacg ggcttcttgg ccatgtagat 7260
ggacttgaac tccaccaggt agtggccgcc gtccttcagc ttcagggcct tgtgggtctc 7320
gcccttcagc acgccgtcgc gggggtacag gcgctcggtg gaggcctccc agcccatggt 7380
cttcttctgc atcacggggc cgtcggaggg gaagttcacg ccgatgaact tcaccttgta 7440
gatgaagcag ccgtcctgca gggaggagtc ctgggtcacg gtcgccacgc cgccgtcctc 7500
gaagttcatc acgcgctccc acttgaagcc ctcggggaag gacagcttct tgtagtcggg 7560
gatgtcggcg gggtgcttca cgtacacctt ggagccgtac tggaactggg gggacaggat 7620
gtcccaggcg aagggcaggg ggccgccctt ggtcaccttc agcttcacgg tgttgtggcc 7680
ctcgtagggg cggccctcgc cctcgccctc gatctcgaac tcgtggccgt tcacggtgcc 7740
ctccatgcgc accttgaagc gcatgaactc ggtgatgacg ttctcggagg aggccatggt 7800
ggcgaccggt ttgcgcttct tcttgggtgg ggtgggatct cccatggtgg cctgaatctc 7860
aacttgcacc tgaaggtagt gcagcaagga tgagcaaaag ggaagaaccc agaaaagaac 7920
gggaaaactt accccaatta gaattgcttg tcgccgccag tgtcaacttg caactgaaac 7980
aatatccaac atgaacgtca atttatactg ccctaatggc gaacacgata acaatatttc 8040
ttttattatg ccctctaaaa ccaacgcggt tatcgtttat ttattcaaat tagatataga 8100
acatccgccg acatacaatg ttaatgcaaa aacgcgtttg gtgagcggat acgaaaacag 8160
tcggccgata aacattaatc tgaggtcgat aacaccgtcc ttgaacggaa cacgaggagc 8220
gtacgtgatc agctgcattc gcgcgccgcg cctttatcga gatttatttg catacaacaa 8280
gtacactgcg ccgttgggat ttgtggtaac gcgcacacat gcagagctgc aagtgtggca 8340
cattttgtct gtgcgcaaaa cctttgaagc caaaagtacg aggtccgtta cgggcatgct 8400
actagcgcac acggacaatg gacccgacaa attctacgcc aaggatttaa tgataatgtc 8460
gggcaacgta tccgttcatt ttatcaataa cctacaaaaa tgtcgcgcgc atcacaaaga 8520
catcgatata tttaaacatt tatgtcccga actgcaaatc gataatagtg ttgtgcaacc 8580
tcgagcgtcc gtttgattta acgtatagct tgcaaatgaa ttatttaatt atcaatcatg 8640
ttttacgcgt agaattctac ccgtaaagcg agtttagtta tgagccatgt gcaaaacatg 8700
acatcagctt ttatttttat aacaaatgac atcatttctt gattgtgttt tacacgtaga 8760
attctactcg taaagcgagt tcagttttga aaaacaaatg acatcatctt tttgattgtg 8820
ctttacaagt agaattctac ccgtaaatca agttcggttt tgaaaaacaa atgagtcata 8880
ttgtatgata tcatattgca aaacaaatga ctcatcaatc gatcgtgcgt tacacgtaga 8940
attctactcg taaagcgagt ttatgagccg tgtgcaaaac atgacatcat ctcgatttga 9000
aaaacaaatg acatcatcca ctgatcgtgc attacaagta gaattctact cgtaaagcca 9060
gttcggttat gagccgtgta caaaacatga catcagatta tgactcatac ttgattgtgt 9120
tttacgcgta gaattctact cgtaaagcca gttcaatttt aaaaacaaat gacatcatcc 9180
aaattaataa atgacaagca atgggtacca tgcggcctgg cctcgcgctc gcgcgactga 9240
cggtcgtaag cacccgcgta cgtgtccacc ccggtcacaa ccccttgtgt catgtcggcg 9300
accctacgcc cccaactgag agaactcaaa ggttacccca gttggggcac tactcccgaa 9360
aaccgcttct gacctgggaa aacgtgaagc cccggggcat ccgctgaggg ttgccgccgg 9420
ggcttcggtg tgtccgtcag tacttaatta acaccgaaat cgtaattcac ggcatcatta 9480
caaaatattt tgacgttttg gacctcgtcc ctaatgacac cataacggtg gccttgaagt 9540

atatttaacc ctagaaagat agtctgcgta aaattgacgc atgcattctt gaaatattgc 9600
tctctctttc taaatagcgc gaatccgtcg ctgtgcattt aggacatctc agtcgccgct 9660
tggagctccc gtgaggcgtg cttgtcaatg cggtaagtgt cactgatttt gaactataac 9720
gaccgcgtga gtcaaaatga cgcatgatta tcttttacgt gacttttaag atttaactca 9780
tacgataatt atattgttat ttcatgttct acttacgtga taacttatta tatatatatt 9840
ttcttgttat agatatcgtg actaatatat aataaaatgg gtagttcttt agacgatgag 9900
catatcctct ctgctcttct gcaaagcgat gacgagcttg ttggtgagga ttctgacagt 9960
gaaatatcag atcacgtaag tgaagatgac ctcgaggatc caagcttatc gatttcgaac 10020
cctcgaccgc cggagtataa atagaggcgc ttcgtctacg gagcgacaat tcaattcaaa 10080
caagcaaagt gaacacgtcg ctaagcgaaa gctaagcaaa taaacaagcg cagctgaaca 10140
agctaaacaa tcggggtacc gctagagtcg atcccacccc acccaagaag aagcgcaaac 10200
cggtaccatg gcctcctccg agaacgtcat caccgagttc atgcgcttca aggtgcgcat 10260
ggagggcacc gtgaacggcc acgagttcga gatcgagggc gagggcgagg gccgccccta 10320
cgagggccac aacaccgtga agctgaaggt gaccaagggc ggccccctgc ccttcgcctg 10380
ggacatcctg tcccccccagt tccagtacgg ctccaaggtg tacgtgaagc acccgccga 10440
catccccgac tacaagaagc tgtccttccc cgagggcttc aagtgggagc gcgtgatgaa 10500
cttcgaggac ggcggcgtgg cgaccgtgac ccaggactcc tccctgcagg acggctgctt 10560
catctacaag gtgaagttca tcggcgtgaa cttcccctcc gacggccccg tgatgcagaa 10620
gaagaccatg ggctgggagg cctccaccga gcgcctgtac ccccgcgacg gcgtgctgaa 10680
gggcgagacc cacaaggccc tgaagctgaa ggacggcggc cactacctgg tggagttcaa 10740
gtccatctac atggccaaga agcccgtgca gctgcccggc tactactacg tggacgccaa 10800
gctggacatc acctcccaca acgaggacta caccatcgtg gagcagtacg agcgcaccga 10860
gggccgccac cacctgttcc tgtgatgatc ataatcagcc ataccacatt tgtagaggtt 10920
ttacttgctt taaaaaacct cccacacctc cccctgaacc tgaaacataa aatgaatgca 10980
attgttgttg ttaacttgtt tattgcagct tataatggtt acaaataaag caatagcatc 11040
acaaatttca caaataaagc attttttttca ctgcattcta gttgtggttt gtccaaactc 11100
atcaatgtat cttaacgcga gttaattacg gccgctcatt taaatctggc cggccgcaac 11160
cattgtggga accgtgcgat caaacaaacg cgagataccg gaagtactga aaaacagtcg 11220
ctccaggcca gtgggaacat cgatgttttg ttttgacgga ccccttactc tcgtctcata 11280
taaaccgaag ccagctaaga tggtatactt attatcatct tgtgatgagg atgcttctat 11340
caacgaaagt accggtaaac cgcaaatggt tatgtattat aatcaaacta aaggcggagt 11400
ggacacgcta gaccaaatgt gttctgtgat gacctgcagt aggaagacga ataggtggcc 11460
tatggcatta ttgtacggaa tgataaacat tgcctgcata aattctttta ttatatacag 11520
ccataatgtc agtagcaagg gagaaaaggt ccaaagtcgc aaaaaattta tgagaaacct 11580
ttacatgagc ctgacgtcat cgtttatgcg taagcgttta gaagctccta ctttgaagag 11640
atatttgcgc gataatatct ctaatatttt gccaaatgaa gtgcctggta catcagatga 11700
cagtactgaa gagccagtaa tgaaaaaacg tacttactgt acttactgcc cctctaaaat 11760
aaggcgaaag gcaaatgcat cgtgcaaaaa atgcaaaaaa gttatttgtc gagagcataa 11820
tattgatatg tgccaaagtt gtttctgact gactaataag tataatttgt ttctattatg 11880
tataagttaa gctaattact tattttataa tacaacatga ctgtttttaa agtacaaaat 11940
aagtttattt ttgtaaaaga gagaatgttt aaaagttttg ttactttata gaagaaattt 12000
tgagtttttg ttttttttta ataaataaat aaacataaat aaattgtttg ttgaatttat 12060
tattagtatg taagtgtaaa tataataaaa cttaatatct attcaaatta ataaataaac 12120
ctcgatatac agaccgataa aacacatgcg tcaattttac gcatgattat ctttaacgta 12180
cgtcacaata tgattatctt tctagggtta aataatagtt tctaattttt ttattattca 12240
gcctgctgtc gtgaataccg tatatctcaa cgctgtctgt gagattgtcg tattctagcc 12300
tttttagttt ttcgctcatc gacttgatat tgtccgacac attttcgtcg atttgcgttt 12360
tgatcaaaga cttgagcaga gacacgttaa tcaactgttc aaattgatcc atattaacga 12420
tatcaacccg atgcgtatat ggtgcgtaaa atatattttt taaccctctt atactttgca 12480

ctctgcgtta atacgcgttc gtgtacagac gtaatcatgt tttcttttt ggataaaact 12540
cctactgagt ttgacctcat attagaccct cacaagttgc aaaacgtggc attttttacc 12600
aatgaagaat ttaaagttat tttaaaaaat ttcatcacag atttaaagaa gaaccaaaaa 12660
ttaaattatt tcaacagttt aatcgaccag ttaatcaacg tgtacacaga cgcgtcggca 12720
aaaaacacgc agcccgacgt gttggctaaa attattaaat caacttgtgt tatagtcacg 12780
gatttgccgt ccaacgtgtt cctcaaaaag ttgaagacca acaagtttac ggacactatt 12840
aattatttga ttttgcccca cttcattttg tgggatcaca attttgttat attttaaaca 12900
aagcttggca ctggccgtcg ttttacaacg tcgtgactgg gaaaaccctg gcgttaccca 12960
acttaatcgc cttgcagcac atccccctttt cgccagctgg cgtaatagcg aagaggcccg 13020
caccgatcgc ccttcccaac agttgcgcag cctgaatggc gaatggcgcc tgatgcggta 13080
ttttctcctt acgcatctgt gcggtatttc acaccgcata tggtgcactc tcagtacaat 13140
ctgctctgat gccgcatagt taagccagcc ccgacacccg ccaacacccg ctgacgcgcc 13200
ctgacgggct tgtctgctcc cggcatccgc ttacagacaa gctgtgaccg tctccgggag 13260
ctgcatgtgt cagaggtttt caccgtcatc accgaaacgc gcgagacgaa agggcctcgt 13320
gatacgccta tttttatagg ttaatgtcat gataataatg gtttcttaga cgtcaggtgg 13380
cacttttcgg ggaaatgtgc gcggaacccc tatttgttta ttttctaaa tacattcaaa 13440
tatgtatccg ctcatgagac aataaccctg ataaatgctt caataatatt gaaaaaggaa 13500
gagtatgagt attcaacatt tccgtgtcgc ccttattccc ttttttgcgg catttttgcct 13560
tcctgttttt gctcacccag aaacgctggt gaaagtaaaa gatgctgaag atcagttggg 13620
tgcacgagtg ggttacatcg aactggatct caacagcggt aagatccttg agagttttcg 13680
ccccgaagaa cgttttccaa tgatgagcac ttttaaagtt ctgctatgtg gcgcggtatt 13740
atcccgtatt gacgccgggc aagagcaact cggtcgccgc atacactatt ctcagaatga 13800
cttggttgag tactcaccag tcacagaaaa gcatcttacg gatggcatga cagtaagaga 13860
attatgcagt gctgccataa ccatgagtga taacactgcg gccaacttac ttctgacaac 13920
gatcggagga ccgaaggagc taaccgcttt tttgcacaac atgggggatc atgtaactcg 13980
ccttgatcgt tgggaaccgg agctgaatga agccatacca aacgacgagc gtgacaccac 14040
gatgcctgta gcaatggcaa caacgttgcg caaactatta actggcgaac tacttactct 14100
agcttcccgg caacaattaa tagactggat ggaggcggat aaagttgcag gaccacttct 14160
gcgctcggcc cttccggctg gctggtttat tgctgataaa tctggagccg gtgagcgtgg 14220
gtctcgcggt atcattgcag cactggggcc agatggtaag ccctcccgta tcgtagttat 14280
ctacacgacg gggagtcagg caactatgga tgaacgaaat agacagatcg ctgagatagg 14340
tgcctcactg attaagcatt ggtaactgtc agaccaagtt tactcatata tactttagat 14400
tgatttaaaa cttcattttt aatttaaaag gatctaggtg aagatccttt ttgataatct 14460
catgaccaaa atcccttaac gtgagttttc gttccactga gcgtcagacc ccgtagaaaa 14520
gatcaaagga tcttcttgag atcctttttt tctgcgcgta atctgctgct tgcaaacaaa 14580
aaaaccaccg ctaccagcgg tggtttgttt gccggatcaa gagctaccaa ctcttttcc 14640
gaaggtaact ggcttcagca gagcgcagat accaaatact gtccttctag tgtagccgta 14700
gttaggccac cacttcaaga actctgtagc accgcctaca tacctcgctc tgctaatcct 14760
gttaccagtg gctgctgcca gtggcgataa gtcgtgtctt accgggttgg actcaagacg 14820
atagttaccg gataaggcgc agcggtcggg ctgaacgggg ggttcgtgca cacagcccag 14880
cttggagcga acgacctaca ccgaactgag atacctacag cgtgagcatt gagaaagcgc 14940
cacgcttccc gaagggagaa aggcggacag gtatccggta agcggcaggg tcggaacagg 15000
agagcgcacg agggagcttc cagggggaaa cgcctggtat ctttatagtc ctgtcgggtt 15060
tcgccacctc tgacttgagc gtcgattttt gtgatgctcg tcaggggggc ggagcctatg 15120
gaaaaacgcc agcaacgcgg cctttttacg gttcctggcc ttttgctggc cttttgctca 15180
catgttcttt cctgcgttat cccctgattc tgtggataac cgtattaccg cctttgagtg 15240
agctgatacc gctcgccgca gccgaacgac cgagcgcagc gagtcagtga gcgaggaagc 15300
ggaagagcgc ccaatacgca aaccgcctct ccccgcgcgt tggccgattc attaatgcag 15360
ctggcacgac aggtttcccg actggaaagc gggcagtgag cgcaacgcaa ttaatgtgag 15420

ttagctcact cattaggcac cccaggcttt acactttatg cttccggctc gtatgttgtg 15480
tggaattgtg agcggataac aatttcacac aggaaacagc tatgaccatg attacgaatt 15540
tcgacctgca ggcatgcaag cttgcatgcc tgcaggtcga cgctcgcgcg acttggtttg 15600
ccattcttta gcgcgcgtcg cgtcacacag cttggccaca atgtggtttt tgtcaaacga 15660
agattctatg acgtgtttaa agtttaggtc gagtaaagcg caaatctttt ttaaccctag 15720
aaagatagtc tgcgtaaaat tgacgcatgc attcttgaaa tattgctctc tctttctaaa 15780
tagcgcgaat ccgtcgctgt gcatttagga catctcagtc gccgcttgga gctcccgtga 15840
ggcgtgcttg tcaatgcggt aagtgtcact gattttgaac tataacgacc gcgtgagtca 15900
aaatgacgca tgattatctt ttacgtgact tttaagattt aactcatacg ataattatat 15960
tgttatttca tgttctactt acgtgataac ttattatata tatattttct tgttatagat 16020
atcgtgacta atatataata aaatgggtag ttctttagac gatgagcata tcctctctgc 16080
tcttctgcaa agcgatgacg agcttgttgg tgaggattct gacagtgaaa tatcagatca 16140
cgtaagtgaa gatgacgtcc agagcgatac agaagaagcg tttatagatg aggtacatga 16200
agtgcagcca acgtcaagcg gtagtgaaat attagacgaa caaaatgtta ttgaacaacc 16260
aggttcttca ttggcttcta acagaatctt gaccttgcca cagaggacta ttagaggtaa 16320
gaataaacat tgttggtcaa cttcaaagtc cacgaggcgt agccgagtct ctgcactgaa 16380
cattgtcaga tcggcccgct cgcccgggga actagttcaa ttagagacta attcaattag 16440
agctaattca attaggatcc aagcttatcg atttcgaacc ctcgaccgcc ggagtataaa 16500
tagaggcgct tcgtctacgg agcgacaatt caattcaaac aagcaaagtg aacacgtcgc 16560
taagcgaaag ctaagcaaat aaacaagcgc agctgaacaa gctaaacaat cggggtaccg 16620
ctagagtcga tcccacccca cccaagaaga agcgcaaacc ggtcgccacc atggccctgt 16680
ccaacaagtt catcggcgac gacatgaaga tgacctacca catggacggc tgcgtgaacg 16740
gccactactt caccgtgaag ggcgagggca gcggcaagcc ctacgagggc acccagacct 16800
ccaccttcaa ggtgaccatg gccaacggcg gccccctggc cttctccttc gacatcctgt 16860
ccaccgtgtt catgtacggc aaccgctgct tcaccgccta ccccaccagc atgcccgact 16920
acttcaagca ggccttcccc gacggcatgt cctacgagag aaccttcacc tacgaggacg 16980
gcggcgtggc caccgccagc tgggagatca gcctgaaggg caactgcttc gagcacaagt 17040
ccaccttcca cggcgtgaac ttccccgccg acggccccgt gatggccaag aagaccaccg 17100
gctgggaccc ctccttcgag aagatgaccg tgtgcgacgg catcttgaag ggcgacgtga 17160
ccgccttcct gatgctgcag ggcggcggca actacagatg ccagttccac acctcctaca 17220
agaccaagaa gcccgtgacc atgcccccca accacgtggt ggagcaccgc atcgccagaa 17280
ccgacctgga caagggcggc aacagcgtgc agctgaccga gcacgccgtg gcccacatca 17340
cctccgtggt gcccttctcc ggactcagat cataatcagc cataccacat ttgtagaggt 17400
tttacttgct ttaaaaaacc tcccacacct ccccctgaac ctgaaacata aaatgaatgc 17460
aattgttgtt gttaacttgt ttattgcagc ttataatggt tacaaataaa gcaatagcat 17520
cacaaatttc acaaataaag catttttttc actgcattct agttgtggtt tgtccaaact 17580
catcaatgta tcttaccgcg gagtggacac gctagaccaa atgtgttctg tgatgacctg 17640
cagtaggaag acgaataggt ggcctatggc attattgtac ggaatgataa acattgcctg 17700
cataaattct tttattatat acagccataa tgtcagtagc aagggagaaa aggtccaaag 17760
tcgcaaaaaa tttatgagaa acctttacat gagcctgacg tcatcgttta tgcgtaagcg 17820
tttagaagct cctactttga agagatattt gcgcgataat atctctaata ttttgccaaa 17880
tgaagtgcct ggtacatcag atgacagtac tgaagagcca gtaatgaaaa aacgtactta 17940
ctgtacttac tgcccctcta aaataaggcg aaaggcaaat gcatcgtgca aaaaatgcaa 18000
aaaagttatt tgtcgagagc ataatattga tatgtgccaa agttgtttct gactgactaa 18060
taagtataat ttgtttctat tatgtataag ttaagctaat tacttatttt ataatacaac 18120
atgactgttt ttaaagtaca aaataagttt atttttgtaa aagagagaat gtttaaaagt 18180
tttgttactt tatagaagaa attttgagtt tttgtttttt tttaataaat aaataaacat 18240
aaataaattg tttgttgaat ttattattag tatgtaagtg taaatataat aaaacttaat 18300
atctattcaa attaataaat aaacctcgat atacagaccg ataaaacaca tgcgtcaatt 18360

ttacgcatga ttatctttaa cgtacgtcac aatatgatta tctttctagg g 18411

<210> 157
<211> 18073
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3646
<400> 157

ctaggtaaga tacattgatg agtttggaca aaccacaact agaatgcagt gaaaaaaatg 60
ctttatttgt gaaatttgtg atgctattgc tttatttgta accattataa gctgcaataa 120
acaagttaac aacaacaatt gcattcattt tatgtttcag gttcaggggg aggtgtggga 180
ggttttttaa agcaagtaaa acctctacaa atgtggtatg gctgattatg atcagttatc 240
tagatccggt ggatcttacg ggtcctccac cttccgcttt ttcttgggtc gagatctcag 300
gaacaggtgg tggcggccct cggtgcgctc gtactgctcc acgatggtgt agtcctcgtt 360
gtgggaggtg atgtccagct tggcgtccac gtagtagtag ccgggcagct gcacgggctt 420
cttggccatg tagatggact tgaactccac caggtagtgg ccgccgtcct tcagcttcag 480
ggccttgtgg gtctcgccct tcagcacgcc gtcgcggggg tacaggcgct cggtggaggc 540
ctcccagccc atggtcttct tctgcatcac ggggccgtcg gaggggaagt tcacgccgat 600
gaacttcacc ttgtagatga agcagccgtc ctgcagggag gagtcctggg tcacggtcgc 660
cacgccgccg tcctcgaagt tcatcacgcg ctcccacttg aagccctcgg ggaaggacag 720
cttcttgtag tcggggatgt cggcggggtg cttcacgtac accttggagc cgtactggaa 780
ctgggggggac aggatgtccc aggcgaaggg caggggggcg cccttggtca ccttcagctt 840
cacggtgttg tggccctcgt aggggcggcc ctcgccctcg ccctcgatct cgaactcgtg 900
gccgttcacg gtgccctcca tgcgcacctt gaagcgcatg aactcggtga tgacgttctc 960
ggaggaggcc atggtggcga ccggtttgcg cttcttcttg ggtggggtgg gatctcccat 1020
ggtggcctga atctcaactt gcacctgaag gtagtgcagc aaggatgagc aaaagggaag 1080
aacccagaaa agaacgggaa aacttacccc aattagaatt gcttgtcgcc gccagtgtca 1140
acttgcaact gaaacaatat ccaacatgaa cgtcaattta tactgcccta atggcgaaca 1200
cgataacaat atttctttta ttatgccctc taaaaccaac gcggttatcg tttatttatt 1260
caaattagat atagaacatc cgccgacata caatgttaat gcaaaaacgc gtttggtgag 1320
cggatacgaa aacagtcggc cgataaacat taatctgagg tcggtaacac cgtccttgaa 1380
cggaacacga ggagcgtacg tgatcagctg cattcgcgcg ccgcgccttt atcgagattt 1440
atttgcatac aacaagtaca ctgcgccgtt gggatttgtg gtaacgcgca cacatgcaga 1500
gctgcaagtg tggcacattt tgtctgtgcg caaaaccttt gaagccaaaa gtacgaggtc 1560
cgttacgggc atgctagcgc acacggacaa tggacccgac aaattctacg ccaaggattt 1620
aatgataatg tcgggcaacg tatccgttca ttttatcaat aacctacaaa aatgtcgcgc 1680
gcatcacaaa gacatcgata tatttaaaca tttatgtccc gaactgcaaa tcgataatag 1740
tgttgtgcaa cctcgagcgt ccgtttgatt taacgtatag cttgcaaatg aattatttaa 1800
ttatcaatca tgttttacgc gtagaattct acccgtaaag cgagtttagt tatgagccat 1860
gtgcaaaaca tgacatcagc ttttatttttt ataacaaatg acatcatttc ttgattgtgt 1920
tttacacgta gaattctact cgtaaagcga gttcagtttt gaaaaacaaa tgacatcatc 1980
tttttgattg tgctttacaa gtagaattct acccgtaaat caagttcggt tttgaaaaac 2040
aaatgagtca tattgtatga tatcatattg caaaacaaat gactcatcaa tcgatcgtgc 2100
gttacacgta gaattctact cgtaaagcga gtttatgagc cgtgtgcaaa acatgacatc 2160
atctcgattt gaaaaacaaa tgacatcatc cactgatcgt gcattacaag tagaattcta 2220
ctcgtaaagc cagttcggtt atgagccgtg tacaaaacat gacatcagat tatgactcat 2280
acttgattgt gttttacgcg tagaattcta ctcgtaaagc cagttcaatt ttaaaaacaa 2340
atgacgcggc cgcattaaca ccgaaatcgt aattcacggc atcattacaa aatattttga 2400

cgttttggac ctcgtcccta atgacaccat aacggtggcc ttgaagtata tttaacccta 2460
gaaagatagt ctgcgtaaaa ttgacgcatg cattcttgaa atattgctct ctctttctaa 2520
atagcgcgaa tccgtcgctg tgcatttagg acatctcagt cgccgcttgg agctcccgtg 2580
aggcgtgctt gtcaatgcgg taagtgtcac tgattttgaa ctataacgac cgcgtgagtc 2640
aaaatgacgc atgattatct tttacgtgac ttttaagatt taactcatac gataattata 2700
ttgttatttc atgttctact tacgtgataa cttattatat atatattttc ttgttataga 2760
tatcgtgact aatatataat aaaatgggta gttctttaga cgatgagcat atcctctctg 2820
ctcttctgca aagcgatgac gagcttgttg gtgaggattc tgacagtgaa atatcagatc 2880
acgtaagtga agatgacctc gaggatccaa gcttatcgat ttcgaaccct cgaccgccgg 2940
agtataaata gaggcgcttc gtctacggag cgacaattca attcaaacaa gcaaagtgaa 3000
cacgtcgcta agcgaaagct aagcaaataa acaagcgcag ctgaacaagc taaacaatcg 3060
gggtaccgct agagtcgatc ccaccccacc caagaagaag cgcaaaccgg taccatggcc 3120
tcctccgaga acgtcatcac cgagttcatg cgcttcaagg tgcgcatgga gggcaccgtg 3180
aacggccacg agttcgagat cgagggcgag ggcgagggcc gccctacga gggccacaac 3240
accgtgaagc tgaaggtgac caagggcggc ccctgccct tcgcctggga catcctgtcc 3300
ccccagttcc agtacggctc caaggtgtac gtgaagcacc ccgccgacat ccccgactac 3360
aagaagctgt ccttccccga gggcttcaag tgggagcgcg tgatgaactt cgaggacggc 3420
ggcgtggcga ccgtgaccca ggactcctcc ctgcaggacg gctgcttcat ctacaaggtg 3480
aagttcatcg gcgtgaactt ccctccgac ggccccgtga tgcagaagaa gaccatgggc 3540
tgggaggcct ccaccgagcg cctgtacccc cgcgacggcg tgctgaaggg cgagacccac 3600
aaggccctga agctgaagga cggcggccac tacctggtgg agttcaagtc catctacatg 3660
gccaagaagc ccgtgcagct gccccggctac tactacgtgg acgccaagct ggacatcacc 3720
tcccacaacg aggactacac catcgtggag cagtacgagc gcaccgaggg ccgccaccac 3780
ctgttcctgt gatgatcata atcagccata ccacatttgt agaggtttta cttgctttaa 3840
aaaacctccc acacctcccc ctgaacctga aacataaaat gaatgcaatt gttgttgtta 3900
acttgtttat tgcagcttat aatggttaca ataaagcaa tagcatcaca aatttcacaa 3960
ataaagcatt tttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt 4020
aacgcgagtt aattacggcc gctcatttaa atctggccgg ccgcaaccat tgtgggaacc 4080
gtgcgatcaa acaaacgcga gataccggaa gtactgaaaa acagtcgctc caggccagtg 4140
ggaacatcga tgttttgttt tgacggaccc cttactctcg tctcatataa accgaagcca 4200
gctaagatgg tatacttatt atcatcttgt gatgaggatg cttctatcaa cgaaagtacc 4260
ggtaaaccgc aaatggttat gtattataat caaactaaag gcggagtgga cacgctagac 4320
caaatgtgtt ctgtgatgac ctgcagtagg aagacgaata ggtggcctat ggcattattg 4380
tacggaatga taaacattgc ctgcataaat tcttttatta tatacagcca taatgtcagt 4440
agcaagggag aaaaggtcca aagtcgcaaa aaatttatga gaaaccttta catgagcctg 4500
acgtcatcgt ttatgcgtaa gcgtttagaa gctcctactt tgaagagata tttgcgcgat 4560
aatatctcta atattttgcc aaatgaagtg cctggtacat cagatgacag tactgaagag 4620
ccagtaatga aaaaacgtac ttactgtact tactgccccc ctaaaataag gcgaaaggca 4680
aatgcatcgt gcaaaaaatg caaaaaagtt atttgtcgag agcataatat tgatatgtgc 4740
caaagttgtt tctgactgac taataagtat aatttgtttc tattatgtat aagttaagct 4800
aattacttat tttataatac aacatgactg tttttaaagt acaaaataag tttatttttg 4860
taaaagagag aatgtttaaa agttttgtta ctttatagaa gaaattttga gtttttgttt 4920
ttttttaata aataaataaa cataaataaa ttgtttgttg aatttattat tagtatgtaa 4980
gtgtaaatat aataaaactt aatatctatt caaattaata aataaacctc gatatacaga 5040
ccgataaaac acatgcgtca attttacgca tgattatctt taacgtacgt cacaatatga 5100
ttatctttct agggttaaat aatagtttct aatttttta ttattcagcc tgctgtcgtg 5160
aataccgtat atctcaacgc tgtctgtgag attgtcgtat tctagccttt ttagtttttc 5220
gctcatcgac ttgatattgt ccgacacatt ttcgtcgatt tgcgttttga tcaaagactt 5280
gagcagagac acgttaatca actgttcaaa ttgatccata ttaacgatat caacccgatg 5340

cgtatatggt gcgtaaaata tatttttaa ccctcttata ctttgcactc tgcgttaata 5400
cgcgttcgtg tacagacgta atcatgtttt ctttttgga taaaactcct actgagtttg 5460
acctcatatt agaccctcac aagttgcaaa acgtggcatt ttttaccaat gaagaattta 5520
aagttatttt aaaaaatttc atcacagatt taaagaagaa ccaaaaatta aattatttca 5580
acagtttaat cgaccagtta atcaacgtgt acacagacgc gtcggcaaaa aacacgcagc 5640
ccgacgtgtt ggctaaaatt attaaatcaa cttgtgttat agtcacggat ttgccgtcca 5700
acgtgttcct caaaaagttg aagaccaaca agtttacgga cactattaat tatttgattt 5760
tgccccactt cattttgtgg gatcacaatt ttgttatatt ttaaacaaag cttggcactg 5820
gccgtcgttt tacaacgtcg tgactgggaa aaccctggcg ttacccaact taatcgcctt 5880
gcagcacatc ccccttccgc cagctggcgt aatagcgaag aggcccgcac cgatcgccct 5940
tcccaacagt tgcgcagcct gaatggcgaa tggcgcctga tgcggtattt tctccttacg 6000
catctgtgcg gtatttcaca ccgcatatgg tgcactctca gtacaatctg ctctgatgcc 6060
gcatagttaa gccagccccg acacccgcca acacccgctg acgcgccctg acgggcttgt 6120
ctgctcccgg catccgctta cagacaagct gtgaccgtct ccgggagctg catgtgtcag 6180
aggttttcac cgtcatcacc gaaacgcgcg agacgaaagg gcctcgtgat acgcctattt 6240
ttataggtta atgtcatgat aataatggtt tcttagacgt caggtggcac ttttcgggga 6300
aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaatat gtatccgctc 6360
atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tatgagtatt 6420
caacatttcc gtgtcgccct tattcccttt tttgcggcat tttgccttcc tgtttttgct 6480
cacccagaaa cgctggtgaa agtaaaagat gctgaagatc agttgggtgc acgagtgggt 6540
tacatcgaac tggatctcaa cagcggtaag atccttgaga gttttcgccc cgaagaacgt 6600
tttccaatga tgagcacttt taaagttctg ctatgtggcg cggtattatc ccgtattgac 6660
gccgggcaag agcaactcgg tcgccgcata cactattctc agaatgactt ggttgagtac 6720
tcaccagtca cagaaaagca tcttacggat ggcatgacag taagagaatt atgcagtgct 6780
gccataacca tgagtgataa cactgcggcc aacttacttc tgacaacgat cggaggaccg 6840
aaggagctaa ccgcttttt gcacaacatg ggggatcatg taactcgcct tgatcgttgg 6900
gaaccggagc tgaatgaagc cataccaaac gacgagcgtg acaccacgat gcctgtagca 6960
atggcaacaa cgttgcgcaa actattaact ggcgaactac ttactctagc ttcccggcaa 7020
caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg ctcggccctt 7080
ccggctggct ggtttattgc tgataaatct ggagccggtg agcgtgggtc tcgcggtatc 7140
attgcagcac tggggccaga tggtaagccc tcccgtatcg tagttatcta cacgacgggg 7200
agtcaggcaa ctatggatga acgaaataga cagatcgctg agataggtgc ctcactgatt 7260
aagcattggt aactgtcaga ccaagtttac tcatatatac tttagattga tttaaaactt 7320
catttttaat ttaaaaggat ctaggtgaag atccttttttg ataatctcat gaccaaaatc 7380
ccttaacgtg agttttcgtt ccactgagcg tcagaccccg tagaaaagat caaaggatct 7440
tcttgagatc ctttttttct gcgcgtaatc tgctgcttgc aaacaaaaaa accaccgcta 7500
ccagcggtgg tttgtttgcc ggatcaagag ctaccaactc tttttccgaa ggtaactggc 7560
ttcagcagag cgcagatacc aaatactgtc cttctagtgt agccgtagtt aggccaccac 7620
ttcaagaact ctgtagcacc gcctacatac ctcgctctgc taatcctgtt accagtggct 7680
gctgccagtg gcgataagtc gtgtcttacc gggttggact caagacgata gttaccggat 7740
aaggcgcagc ggtcgggctg aacggggggt tcgtgcacac agcccagctt ggagcgaacg 7800
acctacaccg aactgagata cctacagcgt gagcattgag aaagcgccac gcttcccgaa 7860
gggagaaagg cggacaggta tccggtaagc ggcagggtcg gaacaggaga gcgcacgagg 7920
gagcttccag ggggaaacgc ctggtatctt tatagtcctg tcgggtttcg ccacctctga 7980
cttgagcgtc gatttttgtg atgctcgtca ggggggcgga gcctatggaa aaacgccagc 8040
aacgcggcct ttttacggtt cctggccttt tgctggcctt ttgctcacat gttctttcct 8100
gcgttatccc ctgattctgt ggataaccgt attaccgcct ttgagtgagc tgataccgct 8160
cgccgcagcc gaacgaccga gcgcagcgag tcagtgagcg aggaagcgga agagcgccca 8220
atacgcaaac cgcctctccc cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg 8280

tttcccgact ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gctcactcat 8340
taggcacccc aggctttaca ctttatgctt ccggctcgta tgttgtgtgg aattgtgagc 8400
ggataacaat ttcacacagg aaacagctat gaccatgatt acgaatttcg acctgcaggc 8460
atgcaagctt gcatgcctgc aggtcgacgc tcgcgcgact tggtttgcca ttctttagcg 8520
cgcgtcgcgt cacacagctt ggccacaatg tggtttttgt caaacgaaga ttctatgacg 8580
tgtttaaagt ttaggtcgag taaagcgcaa atctttttta accctagaaa gatagtctgc 8640
gtaaaattga cgcatgcatt cttgaaatat tgctctctct ttctaaatag cgcgaatccg 8700
tcgctgtgca tttaggacat ctcagtcgcc gcttggagct cccgtgaggc gtgcttgtca 8760
atgcggtaag tgtcactgat tttgaactat aacgaccgcg tgagtcaaaa tgacgcatga 8820
ttatctttta cgtgactttt aagatttaac tcatacgata attatattgt tatttcatgt 8880
tctacttacg tgataactta ttatatatat attttcttgt tatagatatc gtgactaata 8940
tataataaaa tgggtagttc tttagacgat gagcatatcc tctctgctct tctgcaaagc 9000
gatgacgagc ttgttggtga ggattctgac agtgaaatat cagatcacgt aagtgaagat 9060
gacgtccaga gcgatacaga agaagcgttt atagatgagg tacatgaagt gcagccaacg 9120
tcaagcggta gtgaaatatt agacgaacaa aatgttattg aacaaccagg ttcttcattg 9180
gcttctaaca gaatcttgac cttgccacag aggactatta gaggtaagaa taaacattgt 9240
tggtcaactt caaagtccac gaggcgtagc cgagtctctg cactgaacat tgtcagatcg 9300
gcccgctcgc ccggggaact agttcaatta gagactaatt caattagagc taattcaatt 9360
aggatccaag cttatcgatt tcgaaccctc gaccgccgga gtataaatag aggcgcttcg 9420
tctacggagc gacaattcaa ttcaaacaag caaagtgaac acgtcgctaa gcgaaagcta 9480
agcaaataaa caagcgcagc tgaacaagct aaacaatcgg ggtaccgcta gagtcgatcc 9540
caccccaccc aagaagaagc gcaaaccggt cgccaccatg gccctgtcca acaagttcat 9600
cggcgacgac atgaagatga cctaccacat ggacggctgc gtgaacggcc actacttcac 9660
cgtgaagggc gagggcagcg gcaagcccta cgagggcacc cagacctcca ccttcaaggt 9720
gaccatggcc aacggcggcc ccctggcctt ctccttcgac atcctgtcca ccgtgttcat 9780
gtacggcaac cgctgcttca ccgcctaccc caccagcatg cccgactact tcaagcaggc 9840
cttccccgac ggcatgtcct acgagagaac cttcacctac gaggacggcg gcgtggccac 9900
cgccagctgg gagatcagcc tgaagggcaa ctgcttcgag cacaagtcca ccttccacgg 9960
cgtgaacttc cccgccgacg gccccgtgat ggccaagaag accaccggct gggacccctc 10020
cttcgagaag atgaccgtgt gcgacggcat cttgaagggc gacgtgaccg ccttcctgat 10080
gctgcagggc ggcggcaact acagatgcca gttccacacc tcctacaaga ccaagaagcc 10140
cgtgaccatg ccccccaacc acgtggtgga gcaccgcatc gccagaaccg acctggacaa 10200
gggcggcaac agcgtgcagc tgaccgagca cgccgtggcc cacatcacct ccgtggtgcc 10260
cttctccgga ctcagatcat aatcagccat accacatttg tagaggtttt acttgcttta 10320
aaaaacctcc cacacctccc cctgaacctg aaacataaaa tgaatgcaat tgttgttgtt 10380
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 10440
aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 10500
taccgcggag tggacacgct agaccaaatg tgttctgtga tgacctgcag taggaagacg 10560
aataggtggc ctatggcatt attgtacgga atgataaaca ttgcctgcat aaattctttt 10620
attatataca gccataatgt cagtagcaag ggagaaaagg tccaaagtcg caaaaaattt 10680
atgagaaacc tttacatgag cctgacgtca tcgtttatgc gtaagcgttt agaagctcct 10740
actttgaaga gatatttgcg cgataatatc tctaatattt tgccaaatga agtgcctggt 10800
acatcagatg acagtactga agagccagta atgaaaaaac gtacttactg tacttactgc 10860
ccctctaaaa taaggcgaaa ggcaaatgca tcgtgcaaaa aatgcaaaaa agttatttgt 10920
cgagagcata atattgatat gtgccaaagt tgtttctgac tgactaataa gtataatttg 10980
tttctattat gtataagtta agctaattac ttattttata atacaacatg actgttttta 11040
aagtacaaaa taagtttatt tttgtaaaag agagaatgtt taaaagtttt gttacttttat 11100
agaagaaatt ttgagttttt gttttttttt aataaataaa taaacataaa taaattgttt 11160
gttgaattta ttattagtat gtaagtgtaa atataataaa acttaatatc tattcaaatt 11220

aataaataaa cctcgatata cagaccgata aaacacatgc gtcaatttta cgcatgatta 11280
tctttaacgt acgtcacaat atgattatct ttctagggtt aaaatgaatg taagcacttt 11340
attaacgaaa tctttgggaa tatttcgctc atcagcattt tatttgagca ggagtccgag 11400
atgcccgggc ggcgcgccga attcttaatt aacgccctag ccgcgatcgc atccgccgcg 11460
gtggcggccc taagatacat tgatgagttt ggacaaacca caactagaat gcagtgaaaa 11520
aaatgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccat tataagctgc 11580
aataaacaag ttaacaacaa caattgcatt cattttatgt ttcaggttca gggggaggtg 11640
tgggaggttt tttaaagcaa gtaaaacctc tacaaatgtg gtatggctga ttatgatcgt 11700
tgcacattcc gatgtatgct gtgcagaata tgggactggt gcgcttccaa tccgttttca 11760
aatccattat cttccggttc actgtgagcg ggtcgagtgg gaatctcctc cgtagagccg 11820
aaactttctc tcttccagtg ggaaccgctt ccgcccgcca gaggttcttc agcagccgca 11880
gcagcagcag cgccactgtg taaggcttcc tccagacggc accgcaactc ggtagattca 11940
atggacttcg ggctgcgttt ctcgtacata acctcttcgt cgtcggtggt ggcggtcgtc 12000
gctttggtag atttcgtgtc cagattgagc gcgccaccgt tgttatccgg cgagctggac 12060
tgtgaccggg tacgaaaatc ggcacatggc gaccgtgacg cagacggagt gcgggtgacg 12120
gagatgttct cctcgtccgg ccagtgtctg tgtctgctgg aacgcttcgg aaagtagatg 12180
caacaaatgt cgtatcctgt ttgagggatt gatttttggg gggggggggg gcgaaagtgg 12240
ggaaagaaaa tgtaaaaaag aacaaattat ttattgttta tagacaagtg caatgctgtt 12300
aagaggatag tagactcaat gtctttccat caaatacggt tagaaattgt tcattctatg 12360
gagattgcac gacgccacga gtttgatttt acctttttgat gtttgtcaca aacagcacta 12420
ggtactgtaa ttttgaggtg atgcaaataa attttgaacc atcctgctga acacaaattt 12480
attagacgta tttcgacgtt tggtgagctc gtttccatac ttgcctatct tacgttactt 12540
caatgaacaa ctaaatacac atttgtatgc gcttatcccc taaagacggg tgataatgtc 12600
aactgtttag ctaatttcca aaacaatcaa accttgatac gagataactg actttgcatc 12660
tctcaaacta cgattaatag aaagctatag caaattatgt acttagatag cattggaaag 12720
atgacgcttt ccgctacctc ataacgccat tggctttatc tatatgactg ttcatagctg 12780
gagagatgga tttgaaggtc taattctaag ttatcgcata tcaaggtatc gttgtgctgg 12840
aaaatctgat ctgacagtat cgaaatagcg caactctttg tacccaataa tggaaagttc 12900
tgatttattg tatttataaa aacgttgtct atttgttccg atttcaggtc gtcaaatcac 12960
ttgctagtaa ataacgtctc catagcaatt atcattatta tatactaaat gaaacgagct 13020
caccaattag atagtttcaa acagttatac agttgcttca aacaacatac atacatgcct 13080
tgataagtac cgtgcgccaa atcgagctcg caacatgagt atgaaaagcc actagtaaca 13140
cattgataat cagcatagaa tttaaaataa aataatattt tatactgtgg atatcttcat 13200
aacactgcac gctgatataa aattcagaac tacaaaagga tcgttgaaat tttacgtaac 13260
tcaacagagt aagggagggg gtcttccgaa atgctacggt ctatacacaa aatttaaatt 13320
tttcacacaa aagccgttac gtggaggagg gaggggttct aaaattggca aactttgcgt 13380
cacgtaatat ttgaatcatc ccaaagaaaa taaacttcta tgctgattat aacgcgcgca 13440
gaacaatggt tgcggtgaga gacattaaca ggcttgtttg tggtgctgaa atagaacttg 13500
tcatcaatat gttttagctg gttaaactat acaatcatta cgtagcctga aaatcaccct 13560
tgaaaaggcc gaatatatga cgaaaagaca cactctccaa ctcaaaaggc aaactcaacg 13620
tggtcgtgca caacctccaa tagcagtacc tgtcggagcc gtttggcaac ggccagctac 13680
caaaatacgc tcgcaatggc atgcaagcta cagagagata agtgtttatc agatcatttt 13740
gggcaccgaa accgaccgat gtcggaaacg attgaagaga tataatctct ggtttgtaga 13800
ttgtaggatg gttggttgaa gatccggttt cccggatttt ttcggatgga tgttgcttgt 13860
tgatgattct gctgtcgtcg ttttttttccg gtggcagatg gaacagcctc acttcggctt 13920
tcgaaacaca atcttaaaag ttaagtacta ctgctgtttt gcatttttta aattttccct 13980
ctaaacttgc tttgcactat tggtttcata gccgccgtac tcatcgatgc ccagggcgtc 14040
ggtgaacatc tgctcgaact cgaaatcggc catatccagg gcgccgtagg gggcgctatc 14100
gtgcggggtg aatcccggtc ccgggctatc gccatcgccc agcatgtcca ggtcgaagtc 14160

gtccagggca tcggcgtggg ccatcgccac atcctcgcca tccaggtgca gctcatcgcc 14220
caggctcacg tcggtcggcg gggcggtcga caggcggcgg gtgtgtccgg ccggcaggaa 14280
gctcaggcgc ggggcggcca ggcccgcctc ctccggggca tcatcatccg gcagatccag 14340
caggccctcg atggtgctgc cgtagttgtt cttggtgcgg gcgcggctgt aggcggggcc 14400
cgagcccgac tcgcatttca gttgcttttc caatccgcag ataatcagct ccaagccgaa 14460
caggaatgcc ggctcggctc cttgatgatc gaacagctcg attgcctgac gcagcagtgg 14520
gggcatcgaa tcggttgttg gggtctcgcg ctcctctttt gcgacttgat gctcttggtc 14580
ctccagcacg cagcccaggg taaagtgacc gacggcgctc agagcgtaga gagcattttc 14640
caggctgaag ccttgctggc acaggaacgc gagctggttc tccagtgtct cgtattgctt 14700
ttcggtcggg cgcgtgccga gatggacttt ggcaccgtct cggtgggaca gcagagcgca 14760
gcggaacgac ttggcgttat tgcggaggaa gtcctgccag gactcgcctt ccaacgggca 14820
aaaatgcgtg tggtggcggt cgagcatctc gatggccagg gcatccagca gcgcccgctt 14880
attcttcacg tgccagtaga gggtgggctg ctccacgccc agcttctgcg ccaacttgcg 14940
ggtcgtcagt ccctcaatgc caacttcgtt caacagctcc aacgcggagt tgatgacttt 15000
ggacttatcc aggcggctgc ccatggtcac ttgtttgcac tttcacactc tttaggaacg 15060
ctgtctcaca agtagagcta cacgtggtag ccccagaatg gctgtatgtc gctattatcg 15120
ttaaacagta tttgcactgt ggtcattatg ttgtttgtat tagagttcgt cgcgttcgtg 15180
gaattgggaa ggagaagata cagaattact caaatgaaac cattccacgg gagaatacat 15240
ttcaggttta atcttattct tctaggacga aagcccatcg acagagtctt gcaggcttcc 15300
gtcgatcgac tttcacccgt ggatgctaag gaagcttata atgacctcaa cattctccgc 15360
gcacaggttg gctctattct gtttcacagt ttccggtgca gttgtgacga actcagacga 15420
atacccacga ttgtatgtcc aacctcattt tttatctttg taaactaacg tcgaaaaatc 15480
tagatactac atttctgctt tgcttcatct tacactaatc actagtttga acttgcgggt 15540
tttccgttat gctttgtaaa tatgcgatgc tttagagttt tcttcgttcc gattcttctt 15600
tgcattcgat tgcttcttcc gtcgaatcga tctgatcttc gtggtttatt cttgtttcgg 15660
ttcgaccttt gccgcagcgc agtgggtcgt gctgatcgtg taaaaagtct atcatccgga. 15720
ctggcgcgtc gtactgcgca actctacacc gtcgaacatg ttcagattgt gcaatcgtga 15780
gtattcattg accacggctt gacctgcgag gcagagaaga acagttggat ttttcggata 15840
ttggtacgac ccgggggccg cgttgtcatc agttgcatga atcgttggtc caagttcgac 15900
gaaacgatat ggacatcggt gtttcggtgg accaagatcg acgacacgat cttggtcatg 15960
agtgtttttc ggtggaccta gagatattgc aacgaccgga gtggaatacg acggtacgat 16020
gttggtttgt actgttctgg acgctagtta cttcattgat acgataaagt ttacattcgt 16080
catatctctt gcttttcttg aatccaaatg cttaggacgt gtaaccttca caaaccgtca 16140
taaatcagtt tcgattcact acatgttgta gttattcagg ctcttataat tgaatattca 16200
aaaatcgaat tttctatttt atcttgatca gaggatataa ctcgcttaaa atgcacaatt 16260
ttattagcga caccatgtgg atttgtttta attgaaacct ctatcttctc atatgtatca 16320
cgatataaaa tgctcatttt attgactgtt taacgataaa cttgcgacga tcgacgcacc 16380
accgacctaa ttccattgtg gaagcaaggg gcactgcaat accgaaatgt gaagtaaatt 16440
tcaaatctgc tattatagac gatgatctaa tactcttgaa tggtcttaaa cgtgagttgt 16500
atttcaagaa gttatgacga ttcgattttg gggccattat gaccccaaaa cccagccaac 16560
gtaactttta ttagtacaga cagaaggtca agcgtgcaag tctttcatcc gtgtgtcaat 16620
aaggccatca gttgaaaccg tgtcaattaa ccctccagtt aacccttta actttacca 16680
ggacaaacca atgacttcgt gcgcaaattc caccactcgt tgtctcaggc cttgagttgt 16740
tgtttgataa gaatggggga tgtcaagtcg gggagcgtag cccaacaggc tacggaaact 16800
gcatgatggc agtgtttgat ccagggcact gttgggaata gactccgtcg accgaagatc 16860
ccagatgtcc tgaaactcaa taataagcgt tagcagttac aaaatgggag cacccaggaa 16920
gtgagtgaca cccgatcgat acctcggaaa cagtcccaac cggtaagaac ccccatacct 16980
tcgtcaatcc gttggcgcgc tttattgacg tctccgtcgg cgcctttcag tatcacgtac 17040
gtcaggggcg tcgtctccca ggggtatcgc agcttctcca ggagccgttg agatcgtttg 17100

acaagttcgt cgtggtacct ggcctgaatc tcaacttgca cctgaaggta gtgcagcaag 17160
gatgagcaaa agggaagaac ccagaaaaga acgggaaaac ttaccccaat tagaattggc 17220
tagcgcagat tgtttagctt gttcagctgc gcttgtttat ttgcttagct ttcgcttagc 17280
gacgtgttca ctttgcttgt ttgaattgaa ttgtcgctcc gtagacgaag cgcctctatt 17340
tatactccgg cgctcgtttt cgagtttacc actccctatc agtgatagag aaaagtgaaa 17400
gtcgagttta ccactcccta tcagtgatag agaaaagtga aagtcgagtt taccactccc 17460
tatcagtgat agagaaaagt gaaagtcgag tttaccactc cctatcagtg atagagaaaa 17520
gtgaaagtcg agtttaccac tccctatcag tgatagagaa aagtgaaagt cgagtttacc 17580
actccctatc agtgatagag aaaagtgaaa gtcgagttta ccactcccta tcagtgatag 17640
agaaaagtga aagtcgaaac ctggcgcgcc ccggccatcg agaaagagag agagaagaga 17700
agagagagaa cattcgagaa agagagagag aagagaagag agagaacata ctccctatca 17760
gtgatagaga agtccctatc agtgatagag atgtccctat cagtgataga gagttcccta 17820
tcagtgatag agacgtccct atcagtgata gagaagtccc tatcagtgat agagagatcc 17880
ctatcagtga tagagatttc cctatcagtg atagagaggt ccctatcagt gatagagact 17940
tccctatcag tgatagagaa atccctatca gtgatagaga catccctatc agtgatagag 18000
aactccctat cagtgataga gacctcccta tcagtgatag agatcgatgc ggccgcggcg 18060
gatgcgatcg cgg 18073

<210> 158
<211> 13293
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3054
<400> 158
gggcgccgtt tttcttgaaa tattgctctc tctttctaaa tagcgcgaat ccgtcgctgt 60
gcatttagga catctcagtc gccgcttgga gctcccaaac gcgccagtgg tagtacacag 120
tactgtgggt gttcagtttg aaatcctctt gcttctccat tgtctcggtt acctttggtc 180
aaatccatgg gttctattgc ctatatactc ttgcgattac cagtgattgc gctattagct 240
attagatgga ttgttggcca aacttgtcgc ttaagtggct gggaattgta accgtaggcc 300
cgagtgtaat gatcccccat aaaaagtttt cgcaatgcct ttatttttttg ttgcaaatct 360
ctctttattc tgcggtattc ttcattattg cggggatggg gaaagtgttt atatagaagc 420
aacttacgat tgaacccaaa tgcacctgac aagcaaggtc aaagggccag atttttaaat 480
atattattta gtcttaggac tctctatttg caattaaatt actttgctac ctgagggtta 540
aatcttcccc attgataata ataattccac tatatgttca attgggtttc accgcgctta 600
gttacatgac gagccctaat gagccgtcgg tggtctataa actgtgcctt acaaatactt 660
gcaactcttc tcgttttgaa gtcagcagag ttattgctaa ttgctaattg ctaattgctt 720
ttaactgatt tcttcgaaat tggtgctatg tttatggcgc tattaacaag tatgaatgtc 780
aggtttaacc aggggatgct taattgtgtt ctcaacttca aaggcagaaa tgtttactct 840
tgaccatggg tttaggtata atgttatcaa gctcctcgac gcgcctctta ctagaactac 900
ccaccgtact cgtcaattcc aagggcatcg gtaaacatct gctcaaactc gaagtcggcc 960
atatccagag cgccgtaggg ggcggagtcg tgggggtaa atcccggacc cggggaatcc 1020
ccgtcccccca acatgtccag atcgaaatcg tctagcgcgt cggcatgcgc catcgccacg 1080
tcctcgccgt ctaagtggag ctcgtccccc aggctgacat cggtcggggg ggccgtcgac 1140
agtctgcgcg tgtgtcccgc ggggagaaag gacaggcgcg gagccgccag ccccgcctct 1200
tcgggggcgt cgtcgtccgg gagatcgagc aggccctcga tggtagaccc gtaattgttt 1260
ttcgtacgcg cgcggctgta cgcggggccc gagcccgact cgcatttcag ttgcttttcc 1320
aatccgcaga taatcagctc caagccgaac aggaatgccg gctcggctcc ttgatgatcg 1380
aacagctcga ttgcctgacg cagcagtggg ggcatcgaat cggttgttgg ggtctcgcgc 1440

tcctcttttg cgacttgatg ctcttggtcc tccagcacgc agcccagggt aaagtgaccg 1500
acggcgctca gagcgtagag agcattttcc aggctgaagc cttgctggca caggaacgcg 1560
agctggttct ccagtgtctc gtattgcttt tcggtcgggc gcgtgccgag atggactttg 1620
gcaccgtctc ggtgggacag cagagcgcag cggaacgact tggcgttatt gcggaggaag 1680
tcctgccagg actcgccttc caacgggcaa aaatgcgtgt ggtggcggtc gagcatctcg 1740
atggccaggg catccagcag cgcccgctta ttcttcacgt gccagtagag ggtgggctgc 1800
tccacgccca gcttctgcgc caacttgcgg gtcgtcagtc cctcaatgcc aacttcgttc 1860
aacagctcca acgcggagtt gatgactttg gacttatcca ggcggctgcc accacggaga 1920
cgaaggacca agtgaagggt ggactccttc tggatgttgt aatcggacag ggtgcgtcca 1980
tcctcaagct gcttgccggc gaagatcaga cgctgctgat ctgggggggat tccctcctta 2040
tcctgaatct tggccttcac attctcaatg gtgtccgaag gctctacctc gagggtgatg 2100
gtctttccgg tcaaagtctt cacgaaaatc tgcatcgagc tagccagagg ctttgagcct 2160
tcacctatag ataccataga tgtatggatt agtatcatat acatacaaag gctatttttg 2220
ggacatatta atattaacaa tttccgtgat agttttcacc atttttgttg aatgttacgt 2280
tgaaaattta aatttgtttt aaattaattt taccagtcat gtgttcttaa aagtttttat 2340
gattgaaacg gcataaagtg gttcaaaaat ttatcaagaa aggctttcct tttttaaatc 2400
ttatcttttt ctcttaaaaa tcactagtca attcattatt aatttgttaa cttgaatttg 2460
gaatgtctat ttactttcag ataaattaaa gcaagaaact taatattcga aaaaaattga 2520
ttctaaatgg aatttcactt gatcttcatg tatgcatatc aatttttatt tacattgtat 2580
aataagtttc gagttgattg ttgtaatcca caggtgtccc agagaattaa attccaaatt 2640
acccaagttt attgaatgtt gattgtagtt tcagttgctt tgttgctgca acaatggctt 2700
gttgattgta gatattttcc ctttccttgg tttacttatt acatagactg aaaaagaggt 2760
ttacttttt gatacttatg aaaaatttct attagtgatt actaaccaat cgctatatgt 2820
ttactagaaa acaaataaac tctttacatt aacattcaat aatgtttgct ctgtaaccga 2880
caattgaagg cgttacagca acagtaatat aactagcttc ttaaccctca tctattaacc 2940
ccatcgttta aaacactatg ttaaatggtc taacaaatct agatactaat agatgtctta 3000
ttacttagca gccacagctg caacatccaa gacaattttt gaaacttctt attgagctct 3060
tggcagcaga aatgttggta tttttcacag ctttctgaaa gaccggcacc ttcctccggt 3120
tcccgtttct gaattcaaga ggatttccga cccccaatta atcccgaaac aaataaggta 3180
tattcaaaat gatggaaaag tcatggctgc tgaccttatt tttattccta ttgatagaat 3240
attattcccc ttttaaatac actgtactaa gaggtccggc tataatttta ctcacttgtc 3300
gattatccca tagaatgttg attgtagttg gttgcttttc caggtgagag ttgatcaagt 3360
cacaaaagtt agcgtgtgtt gattgtagat ttgaaggtaa aataattttt gcacccattc 3420
atcgggtaaa acgttctcca tagaatacat ttccatcgat aattgataac ttatgaattt 3480
caaagaaaaa aatatgcttt taaaattacc aaatctacgt ttaataacaa cagatctcag 3540
gaacaggtgg tggcggccct cggtgcgctc gtactgctcc acgatggtgt agtcctcgtt 3600
gtgggaggtg atgtccagct tggcgtccac gtagtagtag ccgggcagct gcacgggctt 3660
cttggccatg tagatggact tgaactccac caggtagtgg ccgccgtcct tcagcttcag 3720
ggccttgtgg gtctcgccct tcagcacgcc gtcgcggggg tacaggcgct cggtggaggc 3780
ctcccagccc atggtcttct tctgcatcac ggggccgtcg gaggggaagt tcacgccgat 3840
gaacttcacc ttgtagatga agcagccgtc ctgcagggag gagtcctggg tcacggtcgc 3900
cacgccgccg tcctcgaagt tcatcacgcg ctcccacttg aagccctcgg ggaaggacag 3960
cttcttgtag tcggggatgt cggcggggtg cttcacgtac accttggagc cgtactggaa 4020
ctgggggggac aggatgtccc aggcgaaggg caggggggcg cccttggtca ccttcagctt 4080
cacggtgttg tggccctcgt aggggcggcc ctcgccctcg ccctcgatct cgaactcgtg 4140
gccgttcacg gtgccctcca tgcgcacctt gaagcgcatg aactcggtga tgacgttctc 4200
ggaggaggcc atggtggcga ccggtttgcg cttcttcttg ggtggggtgg gatccaccag 4260
agacaggttg cggcggcggt tggatggcgt gggcgcgttg gcgttgttgg accggctcat 4320
gttgtgtcgc tgtaacagat gctgttcaac tgtgtttacc agatcgttgc gggctgtatt 4380

tataggcgcg ataagcggga cgggcgcctc gtgtccggtc acgcgcatga gataacgcgc 4440
ggctgatatg gaggcgcgtc ctgttccgat aaggagttgc gtccggctgc ggttagcaac 4500
acaggaagct ggcgtcctgt cacgataaga caacactcgt ccggtccgat aatgtgattc 4560
gtacgtgaca ggacgcgacc cgataaggcc ggcctacgtg actgccgaca cgtacttttt 4620
tgcactgcaa aaaggttcaa tgtgtggtag tgtatttgga gcgtatacaa cggtgtagac 4680
tatttatgta aaatagtcta cgaaacgtag agtttgtact atgtatgggc ccgcgtgcaa 4740
aagcgtgttt ttttgcagtg caaaaaagtt ggtggtgggg aggccaccga gtatggtacc 4800
gcagattgtt tagcttgttc agctgcgctt gtttatttgc ttagctttcg cttagcgacg 4860
tgttcacttt gcttgtttga attgaattgt cgctccgtag acgaagcgcc tctatttata 4920
ctccggcgct cgttttcgag tttaccactc cctatcagtg atagagaaaa gtgaaagtcg 4980
agtttaccac tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc 5040
agtgatagag aaaagtgaaa gtcgagttta ccactcccta tcagtgatag agaaaagtga 5100
aagtcgagtt taccactccc tatcagtgat agagaaaagt gaaagtcgag tttaccactc 5160
cctatcagtg atagagaaaa gtgaaagtcg agtttaccac tccctatcag tgatagagaa 5220
aagtgaaagt cgaaacctgg cgcgccccgg ccatcgagaa agagagagag aagagaagag 5280
agagaacatt cgagaaagag agagagaaga gaagagagag aacatactcc ctatcagtga 5340
tagagaagtc cctatcagtg atagagatgt ccctatcagt gatagagagt tccctatcag 5400
tgatagagac gtcccctatca gtgatagaga agtccctatc agtgatagag agatccctat 5460
cagtgataga gatttcccta tcagtgatag agaggtccct atcagtgata gagacttccc 5520
tatcagtgat agagaaatcc ctatcagtga tagagacatc cctatcagtg atagagaact 5580
ccctatcagt gatagagacc tccctatcag tgatagagat cgatgcggcc gcatggtacc 5640
cattgcttgt catttattaa tttggatgat gtcatttgtt tttaaaattg aactggcttt 5700
acgagtagaa ttctacgcgt aaaacacaat caagtatgag tcataatctg atgtcatgtt 5760
ttgtacacgg ctcataaccg aactggcttt acgagtagaa ttctacttgt aatgcacgat 5820
cagtggatga tgtcatttgt ttttcaaatc gagatgatgt catgttttgc acacggctca 5880
taaactcgct ttacgagtag aattctacgt gtaacgcacg atcgattgat gagtcatttg 5940
ttttgcaata tgatatcata caatatgact catttgtttt tcaaaaccga acttgattta 6000
cgggtagaat tctacttgta aagcacaatc aaaaagatga tgtcatttgt ttttcaaaac 6060
tgaactcgct ttacgagtag aattctacgt gtaaaacaca atcaagaaat gatgtcattt 6120
gttataaaaa taaaagctga tgtcatgttt tgcacatggc tcataactaa actcgcttta 6180
cgggtagaat tctacgcgta aaacatgatt gataattaaa taattcattt gcaagctata 6240
cgttaaatca aacggacgct cgaggttgca caacactatt atcgatttgc agttcgggac 6300
ataaatgttt aaatatatcg atgtctttgt gatgcgcgcg acatttttgt aggttattga 6360
taaaatgaac ggatacgttg cccgacatta tcattaaatc cttggcgtag aatttgtcgg 6420
gtccattgtc cgtgtgcgct agcatgcccg taacggacct cgtacttttg gcttcaaagg 6480
ttttgcgcac agacaaaatg tgccacactt gcagctctgc atgtgtgcgc gttaccacaa 6540
atcccaacgg cgcagtgtac ttgttgtatg caaataaatc tcgataaagg cgcggcgcgc 6600
gaatgcagct gatcacgtac gctcctcgtg ttccgttcaa ggacggtgtt atcgacctca 6660
gattaatgtt tatcggccga ctgttttcgt atccgctcac caaacgcgtt tttgcattaa 6720
cattgtatgt cggcggatgt tctatatcta atttgaataa ataaacgata accgcgttgg 6780
ttttagaggg cataataaaa gaaatattgt tatcgtgttc gccattaggg cagtataaat 6840
tgacgttcat gttggatatt gtttcagttg caagttgaca ctggcggcga caagcaattc 6900
taattggggt aagttttccc gttcttttct gggttcttcc cttttgctca tccttgctgc 6960
actaccttca ggtgcaagtt gagattcagg ccaccatggg agatcccacc ccacccaaga 7020
agaagcgcaa accggtcgcc accatggcct cctccgagaa cgtcatcacc gagttcatgc 7080
gcttcaaggt gcgcatggag ggcaccgtga acggccacga gttcgagatc gagggcgagg 7140
gcgagggccg cccctacgag ggccacaaca ccgtgaagct gaaggtgacc aagggcggcc 7200
ccctgccctt cgcctgggac atcctgtccc cccagttcca gtacggctcc aaggtgtacg 7260
tgaagcaccc cgccgacatc cccgactaca agaagctgtc cttccccgag ggcttcaagt 7320

gggagcgcgt gatgaacttc gaggacggcg gcgtggcgac cgtgacccag gactcctccc 7380
tgcaggacgg ctgcttcatc tacaaggtga agttcatcgg cgtgaacttc ccctccgacg 7440
gccccgtgat gcagaagaag accatgggct gggaggcctc caccgagcgc ctgtaccccc 7500
gcgacggcgt gctgaagggc gagacccaca aggccctgaa gctgaaggac ggcggccact 7560
acctggtgga gttcaagtcc atctacatgg ccaagaagcc cgtgcagctg cccggctact 7620
actacgtgga cgccaagctg gacatcacct cccacaacga ggactacacc atcgtggagc 7680
agtacgagcg caccgagggc cgccaccacc tgttcctgag atctcgaccc aagaaaaagc 7740
ggaaggtgga ggacccgtaa gatccaccgg atctagataa ctgatcataa tcagccatac 7800
cacatttgta gaggttttac ttgctttaaa aaacctccca cacctccccc tgaacctgaa 7860
acataaaatg aatgcaattg ttgttgttaa cttgtttatt gcagcttata atggttacaa 7920
ataaagcaat agcatcacaa atttcacaaa taaagcattt ttttcactgc attctagttg 7980
tggtttgtcc aaactcatca atgtatctta acgcgagtta attaaggccg ctcatttaaa 8040
tctggccggc cgcaaccatt gtgggaaccg tgcgatcaaa caaacgcgag ataccggaag 8100
tactgaaaaa cagtcgctcc aggccagtgg gaacatcgat gttttgtttt gacggacccc 8160
ttactctcgt ctcatataaa ccgaagccag ctaagatggt atacttatta tcatcttgtg 8220
atgaggatgc ttctatcaac gaaagtaccg gtaaaccgca aatggttatg tattataatc 8280
aaactaaagg cggagtggac acgctagacc aaatgtgttc tgtgatgacc tgcagtagga 8340
agacgaatag gtggcctatg gcattattgt acggaatgat aaacattgcc tgcataaatt 8400
cttttattat atacagccat aatgtcagta gcaagggaga aaaggtccaa agtcgcaaaa 8460
aatttatgag aaacctttac atgagcctga cgtcatcgtt tatgcgtaag cgtttagaag 8520
ctcctacttt gaagagatat ttgcgcgata atatctctaa tattttgcca aatgaagtgc 8580
ctggtacatc agatgacagt actgaagagc cagtaatgaa aaaacgtact tactgtactt 8640
actgcccctc taaaataagg cgaaaggcaa atgcatcgtg caaaaaatgc aaaaaagtta 8700
tttgtcgaga gcataatatt gatatgtgcc aaagttgttt ctgactgact aataagtata 8760
atttgtttct attatgtata agttaagcta attacttatt ttataataca acatgactgt 8820
ttttaaagta caaaataagt ttatttttgt aaaagagaga atgtttaaaa gttttgttac 8880
tttatagaag aaattttgag tttttgtttt tttttaataa ataaataaac ataaataaat 8940
tgtttgttga atttattatt agtatgtaag tgtaaatata ataaaactta atatctattc 9000
aaattaataa ataaacctcg atatacagac cgataaaaca catgcgtcaa ttttacgcat 9060
gattatcttt aacgtacgtc acaatatgat tatctttcta gggttaaata atagtttcta 9120
atttttttat tattcagcct gctgtcgtga ataccgtata tctcaacgct gtctgtgaga 9180
ttgtcgtatt ctagcctttt tagtttttcg ctcatcgact tgatattgtc cgacacattt 9240
tcgtcgattt gcgttttgat caaagacttg agcagagaca cgttaatcaa ctgttcaaat 9300
tgatccatat taacgatatc aacccgatgc gtatatggtg cgtaaaatat atttttaac 9360
cctcttatac tttgcactct gcgttaatac gcgttcgtgt acagacgtaa tcatgttttc 9420
ttttttggat aaaactccta ctgagtttga cctcatatta gaccctcaca agttgcaaaa 9480
cgtggcattt tttaccaatg aagaatttaa agttatttta aaaaatttca tcacagattt 9540
aaagaagaac caaaaattaa attatttcaa cagtttaatc gaccagttaa tcaacgtgta 9600
cacagacgcg tcggcaaaaa acacgcagcc cgacgtgttg gctaaaatta ttaaatcaac 9660
ttgtgttata gtcacggatt tgccgtccaa cgtgttcctc aaaaagttga agaccaacaa 9720
gtttacggac actattaatt atttgatttt gccccacttc attttgtggg atcacaattt 9780
tgttatattt taaacaaagc ttggcactgg ccgtcgtttt acaacgtcgt gactgggaaa 9840
accctggcgt tacccaactt aatcgccttg cagcacatcc cccttcgcca agctggcgta 9900
atagcgaaga ggcccgcacc gatcgccctt cccaacagtt gcgcagcctg aatggcgaat 9960
ggcgcctgat gcggtatttt ctccttacgc atctgtgcgg tatttcacac cgcatatggt 10020
gcactctcag tacaatctgc tctgatgccg catagttaag ccagccccga cacccgccaa 10080
cacccgctga cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg 10140
tgaccgtctc cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga 10200
gacgaaaggg cctcgtgata cgcctatttt tataggttaa tgtcatgata ataatggttt 10260

cttagacgtc aggtggcact ttcggggaa atgtgcgcgg aacccctatt tgtttatttt 10320
tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat 10380
aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 10440
ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 10500
ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 10560
tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc 10620
tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggt cgccgcatac 10680
actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 10740
gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca 10800
acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg 10860
gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg 10920
acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 10980
gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 11040
ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 11100
gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 11160
cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 11220
agatcgctga gataggtgcc tcactgatta agcattggta actgtcagac caagtttact 11280
catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga 11340
tccttttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt 11400
cagaccccgt agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct 11460
gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg gatcaagagc 11520
taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc 11580
ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc 11640
tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg 11700
ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggttt 11760
cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg 11820
agcattgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg 11880
gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggtatcttt 11940
atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag 12000
gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc ctggcctttt 12060
gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg gataaccgta 12120
ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag cgcagcgagt 12180
cagtgagcga ggaagcggaa gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc 12240
cgattcatta atgcagctgg cacgacaggt ttcccgactg gaaagcgggc agtgagcgca 12300
acgcaattaa tgtgagttag ctcactcatt aggcacccca ggctttacac tttatgcttc 12360
cggctcgtat gttgtgtgga attgtgagcg gataacaatt tcacacagga aacagctatg 12420
accatgatta cgaatttcga cctgcaggca tgcaagcttg catgcctgca ggtcgacgct 12480
cgcgcgactt ggtttgccat tctttagcgc gcgtcgcgtc acacagcttg gccacaatgt 12540
ggtttttgtc aaacgaagat tctatgacgt gtttaaagtt taggtcgagt aaagcgcaaa 12600
tctttttaa ccctagaaag atagtctgcg taaaattgac gcatgcattc ttgaaatatt 12660
gctctctctt tctaaatagc gcgaatccgt cgctgtgcat ttaggacatc tcagtcgccg 12720
cttggagctc ccgtgaggcg tgcttgtcaa tgcggtaagt gtcactgatt ttgaactata 12780
acgaccgcgt gagtcaaaat gacgcatgat tatcttttac gtgactttta agatttaact 12840
catacgataa ttatattgtt atttcatgtt ctacttacgt gataacttat tatatatata 12900
ttttcttgtt atagatatcg tgactaatat ataataaaat gggtagttct ttagacgatg 12960
agcatatcct ctctgctctt ctgcaaagcg atgacgagct tgttggtgag gattctgaca 13020
gtgaaatatc agatcacgta agtgaagatg acgtccagag cgatacagaa gaagcgttta 13080
tagatgaggt acatgaagtg cagccaacgt caagcggtag tgaaatatta gacgaacaaa 13140
atgttattga acaaccaggt tcttcattgg cttctaacag aatcttgacc ttgccacaga 13200

ggactattag aggtaagaat aaacattgtt ggtcaacttc aaagtccacg aggcgtagcc 13260
gagtctctgc actgaacatt gtcagatcgg ccc 13293

<210> 159
<211> 13515
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3056
<400> 159
gggcgccgtt tttcttgaaa tattgctctc tctttctaaa tagcgcgaat ccgtcgctgt 60
gcatttagga catctcagtc gccgcttgga gctcccaaac gcgccagtgg tagtacacag 120
tactgtgggt gttcagtttg aaatcctctt gcttctccat tgtctcggtt acctttggtc 180
aaatccatgg gttctattgc ctatatactc ttgcgattac cagtgattgc gctattagct 240
attagatgga ttgttggcca aacttgtcgc ttaagtggct gggaattgta accgtaggcc 300
cgagtgtaat gatcccccat aaaaagtttt cgcaatgcct ttatttttg ttgcaaatct 360
ctctttattc tgcggtattc ttcattattg cggggatggg gaaagtgttt atatagaagc 420
aacttacgat tgaacccaaa tgcacctgac aagcaaggtc aaagggccag atttttaaat 480
atattattta gtcttaggac tctctatttg caattaaatt actttgctac ctgagggtta 540
aatcttcccc attgataata ataattccac tatatgttca attgggtttc accgcgctta 600
gttacatgac gagccctaat gagccgtcgg tggtctataa actgtgcctt acaaatactt 660
gcaactcttc tcgttttgaa gtcagcagag ttattgctaa ttgctaattg ctaattgctt 720
ttaactgatt tcttcgaaat tggtgctatg tttatggcgc tattaacaag tatgaatgtc 780
aggtttaacc aggggatgct taattgtgtt ctcaacttca aaggcagaaa tgtttactct 840
tgaccatggg tttaggtata atgttatcaa gctcctcgac gcgcctctta ctagaactac 900
ccaccgtact cgtcaattcc aagggcatcg gtaaacatct gctcaaactc gaagtcggcc 960
atatccagag cgccgtaggg ggcggagtcg tgggggggtaa atcccggacc cggggaatcc 1020
ccgtcccccca acatgtccag atcgaaatcg tctagcgcgt cggcatgcgc catcgccacg 1080
tcctcgccgt ctaagtggag ctcgtccccc aggctgacat cggtcggggg ggccgtcgac 1140
agtctgcgcg tgtgtcccgc ggggagaaag gacaggcgcg gagccgccag ccccgcctct 1200
tcgggggcgt cgtcgtccgg gagatcgagc aggccctcga tggtagaccc gtaattgttt 1260
ttcgtacgcg cgcggctgta cgcggggccc gagcccgact cgcatttcag ttgcttttcc 1320
aatccgcaga taatcagctc caagccgaac aggaatgccg gctcggctcc ttgatgatcg 1380
aacagctcga ttgcctgacg cagcagtggg ggcatcgaat cggttgttgg ggtctcgcgc 1440
tcctcttttg cgacttgatg ctcttggtcc tccagcacgc agcccagggt aaagtgaccg 1500
acggcgctca gagcgtagag agcattttcc aggctgaagc cttgctggca caggaacgcg 1560
agctggttct ccagtgtctc gtattgcttt tcggtcgggc gcgtgccgag atggactttg 1620
gcaccgtctc ggtgggacag cagagcgcag cggaacgact tggcgttatt gcggaggaag 1680
tcctgccagg actcgccttc caacgggcaa aaatgcgtgt ggtggcggtc gagcatctcg 1740
atggccaggg catccagcag cgcccgctta ttcttcacgt gccagtagag ggtgggctgc 1800
tccacgccca gcttctgcgc caacttgcgg gtcgtcagtc cctcaatgcc aacttcgttc 1860
aacagctcca acgcggagtt gatgactttg gacttatcca ggcggctgcc accacggaga 1920
cgaaggacca agtgaaggct ggactccttc tggatgttgt aatcggacag ggtgcgtcca 1980
tcctcaagct gcttgccggc gaagatcaga cgctgctgat ctgggggggat tccctcctta 2040
tcctgaatct tggccttcac attctcaatg gtgtccgaag gctctacctc gagggtgatg 2100
gtctttccgg tcaaagtctt cacgaaaatc tgcatcgagc tagcaaatcg ttctgggctg 2160
ctggaatcct tttaaaaaaa atgatttttt ttttgctata aagctatgaa gtagttcact 2220
tactgtcgat ttgtgacgct ctttgcgcca ttgatttcaa cctcctcttt actgttgtta 2280
ctccgatctt taggctgtgt ttcaaaatga gcacccacat tacttacaac attatcaggg 2340

tttacaacga tgtcgtcgcg ttgaaacaga ggctttgagc cttcacctat agataccata 2400
gatgtatgga ttagtatcat atacatacaa aggctatttt tgggacatat taatattaac 2460
aatttccgtg atagttttca ccatttttgt tgaatgttac gttgaaaatt taaatttgtt 2520
ttaaattaat tttaccagtc atgtgttctt aaaagttttt atgattgaaa cggcataaag 2580
tggttcaaaa atttatcaag aaaggctttc ctttttaaa tcttatcttt ttctcttaaa 2640
aatcactagt caattcatta ttaatttgtt aacttgaatt tggaatgtct atttactttc 2700
agataaatta aagcaagaaa cttaatattc gaaaaaaatt gattctaaat ggaatttcac 2760
ttgatcttca tgtatgcata tcaattttta tttacattgt ataataagtt tcgagttgat 2820
tgttgtaatc cacaggtgtc ccagagaatt aaattccaaa ttacccaagt ttattgaatg 2880
ttgattgtag tttcagttgc tttgttgctg caacaatggc ttgttgattg tagatatttt 2940
ccctttcctt ggtttactta ttacatagac tgaaaaagag gtttactttt ttgatactta 3000
tgaaaaattt ctattagtga ttactaacca atcgctatat gtttactaga aaacaaataa 3060
actctttaca ttaacattca ataatgtttg ctctgtaacc gacaattgaa ggcgttacag 3120
caacagtaat ataactagct tcttaaccct catctattaa ccccatcgtt taaaacacta 3180
tgttaaatgg tctaacaaat ctagatacta atagatgtct tattacttag cagccacagc 3240
tgcaacatcc aagacaattt ttgaaacttc ttattgagct cttggcagca gaaatgttgg 3300
tatttttcac agctttctga aagaccggca ccttcctccg gttcccgttt ctgaattcaa 3360
gaggatttcc gacccccaat taatcccgaa acaaataagg tatattcaaa atgatggaaa 3420
agtcatggct gctgaccttа tttttattcc tattgataga atattattcc ccttttaaat 3480
acactgtact aagaggtccg gctataattt tactcacttg tcgattatcc catagaatgt 3540
tgattgtagt tggttgcttt tccaggtgag agttgatcaa gtcacaaaag ttagcgtgtg 3600
ttgattgtag atttgaaggt aaaataattt ttgcacccat tcatcgggta aaacgttctc 3660
catagaatac atttccatcg ataattgata acttatgaat ttcaaagaaa aaaatatgct 3720
tttaaaatta ccaaatctac gtttaataac aacagatctc aggaacaggt ggtggcggcc 3780
ctcggtgcgc tcgtactgct ccacgatggt gtagtcctcg ttgtgggagg tgatgtccag 3840
cttggcgtcc acgtagtagt agccgggcag ctgcacgggc ttcttggcca tgtagatgga 3900
cttgaactcc accaggtagt ggccgccgtc cttcagcttc agggccttgt gggtctcgcc 3960
cttcagcacg ccgtcgcggg ggtacaggcg ctcggtggag gcctcccagc ccatggtctt 4020
cttctgcatc acggggccgt cggaggggaa gttcacgccg atgaacttca ccttgtagat 4080
gaagcagccg tcctgcaggg aggagtcctg ggtcacggtc gccacgccgc cgtcctcgaa 4140
gttcatcacg cgctcccact tgaagccctc ggggaaggac agcttcttgt agtcggggat 4200
gtcggcgggg tgcttcacgt acaccttgga gccgtactgg aactgggggg acaggatgtc 4260
ccaggcgaag ggcagggggc cgcccttggt caccttcagc ttcacggtgt tgtggccctc 4320
gtaggggcgg ccctcgccct cgccctcgat ctcgaactcg tggccgttca cggtgccctc 4380
catgcgcacc ttgaagcgca tgaactcggt gatgacgttc tcggaggagg ccatggtggc 4440
gaccggtttg cgcttcttct tgggtggggt gggatccacc agagacaggt tgcggcggcg 4500
gttggatggc gtgggcgcgt tggcgttgtt ggaccggctc atgttgtgtc gctgtaacag 4560
atgctgttca actgtgttta ccagatcgtt gcgggctgta tttataggcg cgataagcgg 4620
gacgggcgcc tcgtgtccgg tcacgcgcat gagataacgc gcggctgata tggaggcgcg 4680
tcctgttccg ataaggagtt gcgtccggct gcggttagca acacaggaag ctggcgtcct 4740
gtcacgataa gacaacactc gtccggtccg ataatgtgat tcgtacgtga caggacgcga 4800
cccgataagg ccggcctacg tgactgccga cacgtacttt tttgcactgc aaaaaggttc 4860
aatgtgtggt agtgtatttg gagcgtatac aacggtgtag actatttatg taaaatagtc 4920
tacgaaacgt agagtttgta ctatgtatgg gcccgcgtgc aaaagcgtgt ttttttgcag 4980
tgcaaaaaag ttggtggtgg ggaggccacc gagtatggta ccgcagattg tttagcttgt 5040
tcagctgcgc ttgtttattt gcttagcttt cgcttagcga cgtgttcact ttgcttgttt 5100
gaattgaatt gtcgctccgt agacgaagcg cctctattta tactccggcg ctcgttttcg 5160
agtttaccac tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc 5220
agtgatagag aaaagtgaaa gtcgagttta ccactcccta tcagtgatag agaaaagtga 5280

aagtcgagtt taccactccc tatcagtgat agagaaaagt gaaagtcgag tttaccactc 5340
cctatcagtg atagagaaaa gtgaaagtcg agtttaccac tccctatcag tgatagagaa 5400
aagtgaaagt cgagtttacc actccctatc agtgatagag aaaagtgaaa gtcgaaacct 5460
ggcgcgcccc ggccatcgag aaagagagag agaagagaag agagagaaca ttcgagaaag 5520
agagagagaa gagaagagag agaacatact ccctatcagt gatagagaag tccctatcag 5580
tgatagagat gtccctatca gtgatagaga gttccctatc agtgatagag acgtccctat 5640
cagtgataga gaagtcccta tcagtgatag agagatccct atcagtgata gagatttccc 5700
tatcagtgat agagaggtcc ctatcagtga tagagacttc cctatcagtg atagagaaat 5760
ccctatcagt gatagagaca tccctatcag tgatagagaa ctccctatca gtgatagaga 5820
cctccctatc agtgatagag atcgatgcgg ccgcatggta cccattgctt gtcatttatt 5880
aatttggatg atgtcatttg tttttaaaat tgaactggct ttacgagtag aattctacgc 5940
gtaaaacaca atcaagtatg agtcataatc tgatgtcatg ttttgtacac ggctcataac 6000
cgaactggct ttacgagtag aattctactt gtaatgcacg atcagtggat gatgtcattt 6060
gtttttcaaa tcgagatgat gtcatgtttt gcacacggct cataaactcg ctttacgagt 6120
agaattctac gtgtaacgca cgatcgattg atgagtcatt tgttttgcaa tatgatatca 6180
tacaatatga ctcatttgtt tttcaaaacc gaacttgatt tacgggtaga attctacttg 6240
taaagcacaa tcaaaaagat gatgtcattt gtttttcaaa actgaactcg ctttacgagt 6300
agaattctac gtgtaaaaca caatcaagaa atgatgtcat ttgttataaa aataaaagct 6360
gatgtcatgt tttgcacatg gctcataact aaactcgctt tacgggtaga attctacgcg 6420
taaaacatga ttgataatta aataattcat ttgcaagcta tacgttaaat caaacggacg 6480
ctcgaggttg cacaacacta ttatcgattt gcagttcggg acataaatgt ttaaatatat 6540
cgatgtcttt gtgatgcgcg cgacattttt gtaggttatt gataaaatga acggatacgt 6600
tgcccgacat tatcattaaa tccttggcgt agaatttgtc gggtccattg tccgtgtgcg 6660
ctagcatgcc cgtaacggac ctcgtacttt tggcttcaaa ggttttgcgc acagacaaaa 6720
tgtgccacac ttgcagctct gcatgtgtgc gcgttaccac aaatcccaac ggcgcagtgt 6780
acttgttgta tgcaaataaa tctcgataaa ggcgcggcgc gcgaatgcag ctgatcacgt 6840
acgctcctcg tgttccgttc aaggacggtg ttatcgacct cagattaatg tttatcggcc 6900
gactgttttc gtatccgctc accaaacgcg tttttgcatt aacattgtat gtcggcggat 6960
gttctatatc taatttgaat aaataaacga taaccgcgtt ggttttagag ggcataataa 7020
aagaaatatt gttatcgtgt tcgccattag ggcagtataa attgacgttc atgttggata 7080
ttgtttcagt tgcaagttga cactggcggc gacaagcaat tctaattggg gtaagttttc 7140
ccgttctttt ctgggttctt ccctttttgct catccttgct gcactacctt caggtgcaag 7200
ttgagattca ggccaccatg ggagatccca ccccacccaa gaagaagcgc aaaccggtcg 7260
ccaccatggc ctcctccgag aacgtcatca ccgagttcat gcgcttcaag gtgcgcatgg 7320
agggcaccgt gaacggccac gagttcgaga tcgagggcga gggcgagggc cgcccctacg 7380
agggccacaa caccgtgaag ctgaaggtga ccaagggcgg ccccctgccc ttcgcctggg 7440
acatcctgtc cccccagttc cagtacggct ccaaggtgta cgtgaagcac cccgccgaca 7500
tccccgacta caagaagctg tccttccccg agggcttcaa gtgggagcgc gtgatgaact 7560
tcgaggacgg cggcgtggcg accgtgaccc aggactcctc cctgcaggac ggctgcttca 7620
tctacaaggt gaagttcatc ggcgtgaact tcccctccga cggccccgtg atgcagaaga 7680
agaccatggg ctgggaggcc tccaccgagc gcctgtaccc ccgcgacggc gtgctgaagg 7740
gcgagaccca caaggccctg aagctgaagg acggcggcca ctacctggtg gagttcaagt 7800
ccatctacat ggccaagaag cccgtgcagc tgcccggcta ctactacgtg gacgccaagc 7860
tggacatcac ctcccacaac gaggactaca ccatcgtgga gcagtacgag cgcaccgagg 7920
gccgccacca cctgttcctg agatctcgac ccaagaaaaa gcggaaggtg gaggacccgt 7980
aagatccacc ggatctagat aactgatcat aatcagccat accacatttg tagaggtttt 8040
acttgcttta aaaaacctcc cacacctccc cctgaacctg aaacataaaa tgaatgcaat 8100
tgttgttgtt aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac 8160
aaatttcaca aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat 8220

caatgtatct taacgcgagt taattaaggc cgctcattta aatctggccg gccgcaacca 8280
ttgtgggaac cgtgcgatca aacaaacgcg agataccgga agtactgaaa aacagtcgct 8340
ccaggccagt gggaacatcg atgttttgtt ttgacggacc ccttactctc gtctcatata 8400
aaccgaagcc agctaagatg gtatacttat tatcatcttg tgatgaggat gcttctatca 8460
acgaaagtac cggtaaaccg caaatggtta tgtattataa tcaaactaaa ggcggagtgg 8520
acacgctaga ccaaatgtgt tctgtgatga cctgcagtag gaagacgaat aggtggccta 8580
tggcattatt gtacggaatg ataaacattg cctgcataaa ttcttttatt atatacagcc 8640
ataatgtcag tagcaaggga gaaaaggtcc aaagtcgcaa aaaatttatg agaaaccttt 8700
acatgagcct gacgtcatcg tttatgcgta agcgtttaga agctcctact ttgaagagat 8760
atttgcgcga taatatctct aatattttgc caaatgaagt gcctggtaca tcagatgaca 8820
gtactgaaga gccagtaatg aaaaaacgta cttactgtac ttactgcccc tctaaaataa 8880
ggcgaaaggc aaatgcatcg tgcaaaaaat gcaaaaaagt tatttgtcga gagcataata 8940
ttgatatgtg ccaaagttgt ttctgactga ctaataagta taatttgttt ctattatgta 9000
taagttaagc taattactta ttttataata caacatgact gtttttaaag tacaaaataa 9060
gtttatttt gtaaaagaga gaatgtttaa aagttttgtt actttataga agaaattttg 9120
agtttttgtt ttttttaat aaataaataa acataaataa attgtttgtt gaatttatta 9180
ttagtatgta agtgtaaata taataaaact taatatctat tcaaattaat aaataaacct 9240
cgatatacag accgataaaa cacatgcgtc aattttacgc atgattatct ttaacgtacg 9300
tcacaatatg attatctttc tagggttaaa taatagtttc taatttttt attattcagc 9360
ctgctgtcgt gaataccgta tatctcaacg ctgtctgtga gattgtcgta ttctagcctt 9420
tttagttttt cgctcatcga cttgatattg tccgacacat tttcgtcgat ttgcgttttg 9480
atcaaagact tgagcagaga cacgttaatc aactgttcaa attgatccat attaacgata 9540
tcaacccgat gcgtatatgg tgcgtaaaat atatttttta accctcttat actttgcact 9600
ctgcgttaat acgcgttcgt gtacagacgt aatcatgttt tctttttgg ataaaactcc 9660
tactgagttt gacctcatat tagaccctca caagttgcaa aacgtggcat tttttaccaa 9720
tgaagaattt aaagttattt taaaaaattt catcacagat ttaaagaaga accaaaaatt 9780
aaattatttc aacagtttaa tcgaccagtt aatcaacgtg tacacagacg cgtcggcaaa 9840
aaacacgcag cccgacgtgt tggctaaaat tattaaatca acttgtgtta tagtcacgga 9900
tttgccgtcc aacgtgttcc tcaaaaagtt gaagaccaac aagtttacgg acactattaa 9960
ttatttgatt ttgccccact tcattttgtg ggatcacaat tttgttatat tttaaacaaa 10020
gcttggcact ggccgtcgtt ttacaacgtc gtgactggga aaaccctggc gttacccaac 10080
ttaatcgcct tgcagcacat ccccctttcg ccagctggcg taatagcgaa gaggcccgca 10140
ccgatcgccc ttcccaacag ttgcgcagcc tgaatggcga atggcgcctg atgcggtatt 10200
ttctccttac gcatctgtgc ggtatttcac accgcatatg gtgcactctc agtacaatct 10260
gctctgatgc cgcatagtta agccagcccc gacacccgcc aacacccgct gacgcgccct 10320
gacgggcttg tctgctcccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct 10380
gcatgtgtca gaggttttca ccgtcatcac cgaaacgcgc gagacgaaag ggcctcgtga 10440
tacgcctatt tttataggtt aatgtcatga taataatggt ttcttagacg tcaggtggca 10500
cttttcgggg aaatgtgcgc ggaacccCta tttgtttatt tttctaaata cattcaaata 10560
tgtatccgct catgagacaa taaccctgat aaatgcttca ataatattga aaaaggaaga 10620
gtatgagtat tcaacatttc cgtgtcgccc ttattcccct ttttgcggca ttttgccttc 10680
ctgtttttgc tcacccagaa acgctggtga aagtaaaaga tgctgaagat cagttgggtg 10740
cacgagtggg ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc 10800
ccgaagaacg ttttccaatg atgagcactt ttaaagttct gctatgtggc gcggtattat 10860
cccgtattga cgccgggcaa gagcaactcg gtcgccgcat acactattct cagaatgact 10920
tggttgagta ctcaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat 10980
tatgcagtgc tgccataacc atgagtgata acactgcggc caacttactt ctgacaacga 11040
tcggaggacc gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc 11100
ttgatcgttg ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga 11160

tgcctgtagc aatggcaaca acgttgcgca aactattaac tggcgaacta cttactctag 11220
cttcccggca acaattaata gactggatgg aggcggataa agttgcagga ccacttctgc 11280
gctcggccct tccggctggc tggtttattg ctgataaatc tggagccggt gagcgtgggt 11340
ctcgcggtat cattgcagca ctggggccag atggtaagcc ctcccgtatc gtagttatct 11400
acacgacggg gagtcaggca actatggatg aacgaaatag acagatcgct gagataggtg 11460
cctcactgat taagcattgg taactgtcag accaagttta ctcatatata ctttagattg 11520
atttaaaact tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca 11580
tgaccaaaat cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga 11640
tcaaaggatc ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa 11700
aaccaccgct accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga 11760
aggtaactgg cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt 11820
taggccacca cttcaagaac tctgtagcac cgcctacata cctcgctctg ctaatcctgt 11880
taccagtggc tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat 11940
agttaccgga taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct 12000
tggagcgaac gacctacacc gaactgagat acctacagcg tgagcattga gaaagcgcca 12060
cgcttcccga agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag 12120
agcgcacgag ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc 12180
gccacctctg acttgagcgt cgatttttgt gatgctcgtc aggggggcgg agcctatgga 12240
aaaacgccag caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgctcaca 12300
tgttctttcc tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag 12360
ctgataccgc tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg 12420
aagagcgccc aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagct 12480
ggcacgacag gtttcccgac tggaaagcgg gcagtgagcg caacgcaatt aatgtgagtt 12540
agctcactca ttaggcaccc caggctttac actttatgct tccggctcgt atgttgtgtg 12600
gaattgtgag cggataacaa tttcacacag gaaacagcta tgaccatgat tacgaatttc 12660
gacctgcagg catgcaagct tgcatgcctg caggtcgacg ctcgcgcgac ttggtttgcc 12720
attctttagc gcgcgtcgcg tcacacagct tggccacaat gtggtttttg tcaaacgaag 12780
attctatgac gtgtttaaag tttaggtcga gtaaagcgca aatctttttt aaccctagaa 12840
agatagtctg cgtaaaattg acgcatgcat tcttgaaata ttgctctctc tttctaaata 12900
gcgcgaatcc gtcgctgtgc atttaggaca tctcagtcgc cgcttggagc tcccgtgagg 12960
cgtgcttgtc aatgcggtaa gtgtcactga ttttgaacta taacgaccgc gtgagtcaaa 13020
atgacgcatg attatctttt acgtgacttt taagatttaa ctcatacgat aattatattg 13080
ttatttcatg ttctacttac gtgataactt attatatata tattttcttg ttatagatat 13140
cgtgactaat atataataaa atgggtagtt ctttagacga tgagcatatc ctctctgctc 13200
ttctgcaaag cgatgacgag cttgttggtg aggattctga cagtgaaata tcagatcacg 13260
taagtgaaga tgacgtccag agcgatacag aagaagcgtt tatagatgag gtacatgaag 13320
tgcagccaac gtcaagcggt agtgaaatat tagacgaaca aaatgttatt gaacaaccag 13380
gttcttcatt ggcttctaac agaatcttga ccttgccaca gaggactatt agaggtaaga 13440
ataaacattg ttggtcaact tcaaagtcca cgaggcgtag ccgagtctct gcactgaaca 13500
ttgtcagatc ggccc 13515

<210> 160
<211> 9423
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3488
<400> 160
cggcgcgccg gctttacgag tagaattcta cgcgtaaaac acaatcaagt atgagtcata 60

atctgatgtc atgttttgta cacggctcat aaccgaactg gctttacgag tagaattcta 120
cttgtaatgc acgatcagtg gatgatgtca tttgtttttc aaatcgagat gatgtcatgt 180
tttgcacacg gctcataaac tcgctttacg agtagaattc tacgtgtaac gcacgatcga 240
ttgatgagtc atttgttttg caatatgata tcatacaata tgactcattt gtttttcaaa 300
accgaacttg atttacgggt agaattctac ttgtaaagca caatcaaaaa gatgatgtca 360
tttgtttttc aaaactgaac tcgctttacg agtagaattc tacgtgtaaa acacaatcaa 420
gaaatgatgt catttgttat aaaaataaaa gctgatgtca tgttttgcac atggctcata 480
actaaactcg ctttacgggt agaattctac gcgtaaaaca tgattgataa ttaaataatt 540
catttgcaag ctatacgtta aatcaaacgg acgctcgagg ttgcacaaca ctattatcga 600
tttgcagttc gggacataaa tgtttaaata tatcgatgtc tttgtgatgc gcgcgacatt 660
tttgtaggtt attgataaaa tgaacggata cgttgcccga cattatcatt aaatccttgg 720
cgtagaattt gtcgggtcca ttgtccgtgt gcgctagcat gcccgtaacg gacctcgtac 780
ttttggcttc aaaggttttg cgcacagaca aaatgtgcca cacttgcagc tctgcatgtg 840
tgcgcgttac cacaaatccc aacggcgcag tgtacttgtt gtatgcaaat aaatctcgat 900
aaaggcgcgg cgcgcgaatg cagctgatca cgtacgctcc tcgtgttccg ttcaaggacg 960
gtgttatcga cctcagatta atgtttatcg gccgactgtt ttcgtatccg ctcaccaaac 1020
gcgtttttgc attaacattg tatgtcggcg gatgttctat atctaatttg aataaataaa 1080
cgataaccgc gttggtttta gagggcataa taaaagaaat attgttatcg tgttcgccat 1140
tagggcagta taaattgacg ttcatgttgg atattgtttc agttgcaagt tgacactggc 1200
ggcgacaagc aattctaatt ggggtaagtt ttcccgttct tttctgggtt cttccctttt 1260
gctcatcctt gctgcactac cttcaggtgc aagttgagat tcaggccacc atgggagatc 1320
ccaccccacc caagaagaag cgcaaaccgg tcgccaccat ggagagcgac gagagcggcc 1380
tgcccgccat ggagatcgag tgccgcatca ccggcaccct gaacggcgtg gagttcgagc 1440
tggtgggcgg cggagagggc acccccgagc agggccgcat gaccaacaag atgaagagca 1500
ccaaaggcgc cctgaccttc agcccctacc tgctgagcca cgtgatgggc tacggcttct 1560
accacttcgg cacctacccc agcggctacg agaacccctt cctgcacgcc atcaacaacg 1620
gcggctacac caacacccgc atcgagaagt acgaggacgg cggcgtgctg cacgtgagct 1680
tcagctaccg ctacgaggcc ggccgcgtga tcggcgactt caaggtgatg ggcaccggct 1740
tccccgagga cagcgtgatc ttcaccgaca agatcatccg cagcaacgcc accgtggagc 1800
acctgcaccc catgggcgat aacgatctgg atggcagctt cacccgcacc ttcagcctgc 1860
gcgacggcgg ctactacagc tccgtggtgg acagccacat gcacttcaag agcgccatcc 1920
accccagcat cctgcagaac gggggcccca tgttcgcctt ccgccgcgtg gaggaggatc 1980
acagcaacac cgagctgggc atcgtggagt accagcacgc cttcaagacc ccggatgcag 2040
atgccggtga agaaagatct cgacccaaga aaaagcggaa ggtggaggac ccgtctggag 2100
gcggtggatc cggcggtgga ggcatgcaga tctttgtgaa gactttgacc ggaaagacca 2160
tcaccctcga ggtagagcca tcggacacca ttgagaatgt aaaggccaag attcaggata 2220
aggagggaat ccccccagat cagcagcgtc tgatcttcgc tggtaatttt aaaagcatat 2280
ttttttcttt gaaattcata agttatcaat tatcgatgga aatgtattct atggagaacg 2340
ttttacccga tgaatgggtg caaaaattat tttaccttca aatctacaat caacacacgc 2400
taacttttgt gacttgatca actctcacct ggaaaagcaa ccaactacaa tcaacattct 2460
atgggataat cgacaagtga gtaaaattat agccggacct cttagtacag tgtatttaaa 2520
aggggaataa tattctatca ataggaataa aaataaggtc agcagccatg acttttccat 2580
cattttgaat ataccttatt tgtttcggga ttaattgggg gtcggaaatc ctcttgaatt 2640
cagaaacggg aaccggagga aggtgccggt ctttcagaaa gctgtgaaaa ataccaacat 2700
ttctgctgcc aagagctcaa taagaagttt caaaaattgt cttggatgtt gcagctgtgg 2760
ctgctaagta ataagacatc tattagtatc tagatttgtt agaccattta acatagtgtt 2820
ttaaacgatg gggttaatag atgagggtta agaagctagt tatattactg ttgctgtaac 2880
gccttcaatt gtcggttaca gagcaaacat tattgaatgt taatgtaaag agtttatttg 2940
ttttctagta aacatatagc gattggttag taatcactaa tagaaatttt tcataagtat 3000

caaaaaagta aacctctttt tcagtctatg taataagtaa accaaggaaa gggaaaatat 3060
ctacaatcaa caagccattg ttgcagcaac aaagcaactg aaactacaat caacattcaa 3120
taaacttggg taatttggaa tttaattctc tgggacacct gtggattaca acaatcaact 3180
cgaaacttat tatacaatgt aaataaaaat tgatatgcat acatgaagat caagtgaaat 3240
tccatttaga atcaattttt ttcgaatatt aagtttcttg ctttaattta tctgaaagta 3300
aatagacatt ccaaattcaa gttaacaaat taataatgaa ttgactagtg atttttaaga 3360
gaaaaagata agatttaaaa aaggaaagcc tttcttgata aatttttgaa ccactttatg 3420
ccgtttcaat cataaaaact tttaagaaca catgactggt aaaattaatt taaaacaaat 3480
ttaaattttc aacgtaacat tcaacaaaaa tggtgaaaac tatcacggaa attgttaata 3540
ttaatatgtc ccaaaaatag cctttgtatg tatatgatac taatccatac atctatggta 3600
tctataggta agcaactgga agacggacgc accctgtccg attacaacat ccagaaggag 3660
tccacccttc acttggtcct tcgtctccgc ggtggcatgc agatcgggga tcccacccca 3720
cccaagaaga agcgcaaacc ggtcgccacc atggcctcct ccgagaacgt catcaccgag 3780
ttcatgcgct tcaaggtgcg catggagggc accgtgaacg gccacgagtt cgagatcgag 3840
ggcgagggcg agggccgccc ctacgagggc cacaacaccg tgaagctgaa ggtgaccaag 3900
ggcggccccc tgcccttcgc ctgggacatc ctgtcccccc agttccagta cggctccaag 3960
gtgtacgtga agcaccccgc cgacatcccc gactacaaga agctgtcctt ccccgagggc 4020
ttcaagtggg agcgcgtgat gaacttcgag gacggcggcg tggcgaccgt gacccaggac 4080
tcctccctgc aggacggctg cttcatctac aaggtgaagt tcatcggcgt gaacttcccc 4140
tccgacggcc ccgtgatgca gaagaagacc atgggctggg aggcctccac cgagcgcctg 4200
taccccccgcg acggcgtgct gaagggcgag acccacaagg ccctgaagct gaaggacggc 4260
ggccactacc tggtggagtt caagtccatc tacatggcca agaagccccgt gcagctgccc 4320
ggctactact acgtggacgc caagctggac atcacctccc acaacgagga ctacaccatc 4380
gtggagcagt acgagcgcac cgagggccgc caccacctgt tcctgagatc tcgacccaag 4440
aaaaagcgga aggtggagga cccgtaagat ccaccgggtc tagataactg atcataatca 4500
gccataccac atttgtagag gttttacttg ctttaaaaaa cctcccacac ctccccctga 4560
acctgaaaca taaaatgaat gcaattgttg ttgttaactt gtttattgca gcttataatg 4620
gttacaaata aagcaatagc atcacaaatt tcacaaataa agcatttttt tcactgcatt 4680
ctagttgtgg tttgtccaaa ctcatcaatg tatcttaacg cgagttaatt aagaggcgcg 4740
gtaaaccgca aatggttatg tattataatc aaactaaagg cggagtggac acgctagacc 4800
aaatgtgttc tgtgatgacc tgcagtagga agacgaatag gtggcctatg gcattattgt 4860
acggaatgat aaacattgcc tgcataaatt cttttattat atacagccat aatgtcagta 4920
gcaagggaga aaaggtccaa agtcgcaaaa aatttatgag aaacctttac atgagcctga 4980
cgtcatcgtt tatgcgtaag cgtttagaag ctcctacttt gaagagatat ttgcgcgata 5040
atatctctaa tattttgcca aatgaagtgc ctggtacatc agatgacagt actgaagagc 5100
cagtaatgaa aaaacgtact tactgtactt actgcccctc taaaataagg cgaaaggcaa 5160
atgcatcgtg caaaaaatgc aaaaaagtta tttgtcgaga gcataatatt gatatgtgcc 5220
aaagttgttt ctgactgact aataagtata atttgtttct attatgtata agttaagcta 5280
attacttatt ttataataca acatgactgt ttttaaagta caaaataagt ttatttttgt 5340
aaaagagaga atgtttaaaa gttttgttac tttatagaag aaattttgag tttttgtttt 5400
tttttaataa ataaataaac ataaataaat tgtttgttga atttattatt agtatgtaag 5460
tgtaaatata ataaaactta atatctattc aaattaataa ataaacctcg atatacagac 5520
cgataaaaca catgcgtcaa ttttacgcat gattatcttt aacgtacgtc acaatatgat 5580
tatctttcta gggttaaata atagtttcta atttttttat tattcagcct gctgtcgtga 5640
ataccgtata tctcaacgct gtctgtgaga ttgtcgtatt ctagcctttt tagtttttcg 5700
ctcatcgact tgatattgtc cgacacattt tcgtcgattt gcgttttgat caaagacttg 5760
agcagagaca cgttaatcaa ctgttcaaat tgatccatat taacgatatc aacccgatgc 5820
gtatatggtg cgtaaaatat atttttttaac cctcttatac tttgcactct gcgttaatac 5880
gcgttcgtgt acagacgtaa tcatgttttc ttttttggat aaaactccta ctgagtttga 5940

cctcatatta gaccctcaca agttgcaaaa cgtggcattt tttaccaatg aagaatttaa 6000
agttatttta aaaaatttca tcacagattt aaagaagaac caaaaattaa attatttcaa 6060
cagtttaatc gaccagttaa tcaacgtgta cacagacgcg tcggcaaaaa acacgcagcc 6120
cgacgtgttg gctaaaatta ttaaatcaac ttgtgttata gtcacggatt tgccgtccaa 6180
cgtgttcctc aaaaagttga agaccaacaa gtttacggac actattaatt atttgatttt 6240
gccccacttc attttgtggg atcacaattt tgttatattt taaacaaagc ttggcactgg 6300
ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt tacccaactt aatcgccttg 6360
cagcacatcc cctttcgcc agctggcgta atagcgaaga ggcccgcacc gatcgccctt 6420
cccaacagtt gcgcagcctg aatggcgaat ggcgcctgat gcggtatttt ctccttacgc 6480
atctgtgcgg tatttcacac cgcatatatg gtgcactctc agtacaatct gctctgatgc 6540
cgcatagtta agccagcccc gacacccgcc aacacccgct gacgcgccct gacgggcttg 6600
tctgctcccg gcatccgctt acagacaagc tgtgaccgtc tccgggagct gcatgtgtca 6660
gaggttttca ccgtcatcac cgaaacgcgc gagacgaaag ggcctcgtga tacgcctatt 6720
tttataggtt aatgtcatga taataatggt ttcttagacg tcaggtggca cttttcgggg 6780
aaatgtgcgc ggaacccctа tttgtttatt tttctaaata cattcaaata tgtatccgct 6840
catgagacaa taaccctgat aaatgcttca ataatattga aaaaggaaga gtatgagtat 6900
tcaacatttc cgtgtcgccc ttattccctt ttttgcggca ttttgccttc ctgtttttgc 6960
tcacccagaa acgctggtga aagtaaaaga tgctgaagat cagttgggtg cacgagtggg 7020
ttacatcgaa ctggatctca acagcggtaa gatccttgag agttttcgcc ccgaagaacg 7080
ttttccaatg atgagcactt ttaaagttct gctatgtggc gcggtattat cccgtattga 7140
cgccgggcaa gagcaactcg gtcgccgcat acactattct cagaatgact tggttgagta 7200
ctcaccagtc acagaaaagc atcttacgga tggcatgaca gtaagagaat tatgcagtgc 7260
tgccataacc atgagtgata acactgcggc caacttactt ctgacaacga tcggaggacc 7320
gaaggagcta accgcttttt tgcacaacat gggggatcat gtaactcgcc ttgatcgttg 7380
ggaaccggag ctgaatgaag ccataccaaa cgacgagcgt gacaccacga tgcctgtagc 7440
aatggcaaca acgttgcgca aactattaac tggcgaacta cttactctag cttcccggca 7500
acaattaata gactggatgg aggcggataa agttgcagga ccacttctgc gctcggccct 7560
tccggctggc tggtttattg ctgataaatc tggagccggt gagcgtgggt ctcgcggtat 7620
cattgcagca ctggggccag atggtaagcc ctcccgtatc gtagttatct acacgacggg 7680
gagtcaggca actatggatg aacgaaatag acagatcgct gagataggtg cctcactgat 7740
taagcattgg taactgtcag accaagttta ctcatatata ctttagattg atttaaaact 7800
tcatttttaa tttaaaagga tctaggtgaa gatccttttt gataatctca tgaccaaaat 7860
cccttaacgt gagttttcgt tccactgagc gtcagacccc gtagaaaaga tcaaaggatc 7920
ttcttgagat cctttttttc tgcgcgtaat ctgctgcttg caaacaaaaa aaccaccgct 7980
accagcggtg gtttgtttgc cggatcaaga gctaccaact ctttttccga aggtaactgg 8040
cttcagcaga gcgcagatac caaatactgt ccttctagtg tagccgtagt taggccacca 8100
cttcaagaac tctgtagcac cgcctacata cctcgctctg ctaatcctgt taccagtggc 8160
tgctgccagt ggcgataagt cgtgtcttac cgggttggac tcaagacgat agttaccgga 8220
taaggcgcag cggtcgggct gaacgggggg ttcgtgcaca cagcccagct tggagcgaac 8280
gacctacacc gaactgagat acctacagcg tgagcattga gaaagcgcca cgcttcccga 8340
agggagaaag gcggacaggt atccggtaag cggcagggtc ggaacaggag agcgcacgag 8400
ggagcttcca gggggaaacg cctggtatct ttatagtcct gtcgggtttc gccacctctg 8460
acttgagcgt cgatttttgt gatgctcgtc agggggggcgg agcctatgga aaaacgccag 8520
caacgcggcc tttttacggt tcctggcctt ttgctggcct tttgctcaca tgttctttcc 8580
tgcgttatcc cctgattctg tggataaccg tattaccgcc tttgagtgag ctgataccgc 8640
tcgccgcagc cgaacgaccg agcgcagcga gtcagtgagc gaggaagcgg aagagcgccc 8700
aatacgcaaa ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag 8760
gtttcccgac tggaaagcgg gcagtgagcg caacgcaatt aatgtgagtt agctcactca 8820
ttaggcaccc caggctttac actttatgct tccggctcgt atgttgtgtg gaattgtgag 8880

cggataacaa tttcacacag gaaacagcta tgaccatgat tacgaatttc gacgctcgcg 8940
cgacttggtt tgccattctt tagcgcgcgt cgcgtcacac agcttggcca caatgtggat 9000
gtcgacttaa ccctagaaag atagtctgcg taaaattgac gcatgcattc ttgaaatatt 9060
gctctctctt tctaaatagc gcgaatccgt cgctgtgcat ttaggacatc tcagtcgccg 9120
cttggagctc ccgtgaggcg tgcttgtcaa tgcggtaagt gtcactgatt ttgaactata 9180
acgaccgcgt gagtcaaaat gacgcatgat tatcttttac gtgactttta agatttaact 9240
catacgataa ttatattgtt atttcatgtt ctacttacgt gataacttat tatatatata 9300
ttttcttgtt atagatatct accggtcata ctcggtggcc tccccaccac caacttttt 9360
gcactgcaaa aaaacacgct tttgcacgcg ggcccggcgc gccatctgcc ggccgcatgg 9420
tac 9423

<210> 161
<211> 17781
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3641
<400> 161
ttaaaatgaa tgtaagcact ttattaacga aatctttggg aatatttcgc tcatcagcat 60
tttatttgag caggagtccg agatgccccc ttcccttaag tcaatattac aaacaatgtg 120
gttttccgcc aaccacagtg tggttaaatt ttatcaccga tgatatgaaa tttctagctg 180
caacatgtcc acaagaaata ccataattct catggttgct taacaactgt taattataca 240
tcaggcaaag tattcactgg ttttcttaat atatctggga ataattactt caaggaccgg 300
gttaacaaga taaggtaacc gctctccaac ttaatgtccg tgataatata caaatatgcg 360
tgttgtaaca cgtatagcac atataaaact aggtaaagtc cggaatagcc cactcgggtc 420
ctctcgggtc gggctcgggt cggcctcggg cctactcggg tttggccgaa gtagatggct 480
cagtgggctc gtgcgccgtc tgtcgcggcg cggtgtgggg ctacaagtgc gctggcggcg 540
gcgagcgcat gagcggcacg ggcaggtgcg gcgcgtactc gcgcgacagc acatagcgcg 600
gcgagcagca aaacgactct ttgcgcgctt gcagctccag cagcgagcac agtgagcgct 660
cgtacagccg ccactggtgc accacccagg aggctggaat taacaagaca ggtttaaata 720
aaaactacac aaaaaacaaa taccctgcct acatgcccac caagcacttc cacgtgacct 780
gggaaaacta gaaagccaat ttgaatgtac attttgatat ctaaattatg taattttgtt 840
attttgtatt aaatatgcat aacacattac aaattataac aaaatgacct tgttgatgtc 900
acaacgtaag aattggccgt cgtatcgcgt tatcacgttt ttcgatttaa atcgaggctt 960
taacgtagrc ttaggtacga aagaagcact ctaagcgaat taccttgacg ctgacgcttg 1020
cagcgttgag cgtaaaaact ataactgagt cacacgaact cgccgcgaaa aagctggcct 1080
aatacaagaa aacagagtga gtagaagttt tgtagtcact ctaatttctt tgtattacaa 1140
tttgtagcaa cgaattgtat tatctatatc cagatataga tatatgttac atgaatattc 1200
ctgtcaaatt gttacagawt gtcatcttaa aattagagcg tagctttgat tgtatggctg 1260
tycgtgtgac tttagagtca aaggaattcg ctcagagagt taagttkyra tgctagcact 1320
agcacttacc atgagatccg atgtgtgaga cacaatgtcc aacgaagctt atttacagga 1380
acagtcaccc cacagtccac aaaacacagc acttgaagaa catagtatca ttcgcaaaac 1440
aacttcatcg atgacgatag cacaccacta aaattattta tttcgctcag cattttccta 1500
caaaagaaaa acaaaataaa aatagtatca cttgcacatc actattaaat aaaatgtggt 1560
cactttttt aaatttcgaa cttctccacc ttcgtccggt cactgtggaa atgaaatcac 1620
gactgggtta gtgatgttct gattcgtcgg acacaccact cttttatcaa cttagcaaat 1680
ttgtatgtgt gggtgtgtaa caatgttgtt aatgtgttat gacacattgt gttgtgtggg 1740
gaccgaactt gcaacgttgt agctccgtac attgtttgta aacggcaggc tacgttacta 1800
tacgtagtac gtaagcgacg taagcgtgac tcaacttctt atcgattaca gcgtctttat 1860

aaatgtaagt tatttataat acagtggaac ctcgataagg cgaaaataaa acgctgtctc 1920
gctccgctca caccagtgag agtgagagtg agagaagaac ctgaactcgc ggcgccttaa 1980
cggtccccag caagctcggt tgtaagtaaa cgatggatgg attgtgtgaa agaggatatg 2040
agaaagaaag gagtgagaag agaggaacat gttgtgccga ccccacataa cgtgggataa 2100
gagcaggagg aagaagagca agctaacgta gttacgctct cattttaaaa cgactagcta 2160
aattgctctg aaactttgta ctaacaatag gattaggcat atcggagtcg cctttaagag 2220
cttattaccc ctccgtcgaa ataccacggc caatagtcat atgtattgtt tggactgacg 2280
tttaactgac atatttgctc ctcccccgta aaatcgttgt acagagaatt acagacaagg 2340
tgtttccagt tgttaaatcc tccaagtcta aggctgtaat tagtttatgt agcctcagat 2400
accaagtata aactaatttc agccctagac atacctcatg tcattgtatg tgcaaagttc 2460
cattacaatc caacacgcag ttttataatg agaacgaaac tccgtttgta tgtgaaattc 2520
agccgagctt accattgcta gttttaggaa taaggggtta aaatttgcaa attcggtcta 2580
agtgtgtgta aaaaacaaag gtcggtttcc gaacagaatt ttggtttctt tttgagtgtt 2640
tctaacggtt ttgagatgat ttaaatggaa ccttactttg agacttgctt aggttgcggt 2700
gggcgttttt catcgccatc cgaaatggag ttagccgccg tattcatcga tgcccagggc 2760
gtcggtgaac atctgctcga actcgaaatc ggccatatcc agggcgccgt agggggcgct 2820
atcgtgcggg gtgaatcccg gtcccgggct atcgccatcg cccagcatgt ccaggtcgaa 2880
gtcgtccagg gcatcggcgt gggccatcgc cacatcctcg ccatccaggt gcagctcatc 2940
gcccaggctc acgtcggtcg gcgggggcggt cgacaggcgg cgggtgtgtc cggccggcag 3000
gaagctcagg cgcgggggcgg ccaggccccgc ctcctccggg gcatcatcat ccggcagatc 3060
cagcaggccc tcgatggtgc tgccgtagtt gttcttggtg cgggcgcggc tgtaggcggg 3120
gcccgagccc gactcgcatt tcagttgctt ttccaatccg cagataatca gctccaagcc 3180
gaacaggaat gccggctcgg ctccttgatg atcgaacagc tcgattgcct gacgcagcag 3240
tgggggcatc gaatcggttg ttggggtctc gcgctcctct tttgcgactt gatgctcttg 3300
gtcctccagc acgcagccca gggtaaagtg accgacggcg ctcagagcgt agagagcatt 3360
ttccaggctg aagccttgct ggcacaggaa cgcgagctgg ttctccagtg tctcgtattg 3420
cttttcggtc gggcgcgtgc cgagatggac tttggcaccg tctcggtggg acagcagagc 3480
gcagcggaac gacttggcgt tattgcggag gaagtcctgc caggactcgc cttccaacgg 3540
gcaaaaatgc gtgtggtggc ggtcgagcat ctcgatggcc agggcatcca gcagcgcccg 3600
cttattcttc acgtgccagt agagggtggg ctgctccacg cccagcttct gcgccaactt 3660
gcgggtcgtc agtccctcaa tgccaacttc gttcaacagc tccaacgcgg agttgatgac 3720
tttggactta tccaggcggc tgaccatttt gcctggggac aacggaaatc gcacagtttg 3780
aacgttcgct tggcggcgcg gagactgcat tttggagaac acgtacatgt atcgggcgat 3840
aaaaaaaacy ttgtcattgt ttcattatga ccatgacaaa ttaaggtggg ttatttttg 3900
ctacttgaat ttaattgtcg aamagtaaaa aaaaacgatg caacttttt atattgaaat 3960
tgactgatta caaaatgcag ccttgcttta taatagacac aacatacacg gaggaatgga 4020
ctaggaacat ctattttatg taaccttgta cataactaag acctaggtta aaataacgat 4080
gtgttaatat aatatataga gaacaatata aagcattttg taccatttgg cgttgaaact 4140
tttttgcagc aacgaagcgt ttggtatacg tactcgtaaa tggtgaccga aaagctggcg 4200
gcttcctcgc acaagtaatt ccgccagcat ccttagtaca atgcctgaag ggtatatttt 4260
agatttagct aatttattaa tttagtttag tatttgtaca gttactgcaa tacctctgta 4320
ccggaactcc agctgtgacc ttgacttgtt tcatgtgtca actcgccaca gtcccaactt 4380
gcttaccttc atcaatcttc cgtaacaaaa aagtctcggc gaatccctgg gctttgctcc 4440
aatctaagta ctctgcgttt tcgtaatcgt ttcgtgtccg cgagtttacc catattaata 4500
ctccgtacga cataagtgaa tggaagtgag cgtagtatac ggatcttaaa gtatcactat 4560
cagaagacgg cggcatcaac ggcggtggca gaataacagc gtccgttgct agaattcttt 4620
agcgcactct acaaaattat atcctggtgg attagccgag tcacattccc tttcaattcg 4680
tctgtaagca ttgttatgta cttaaattta aacttacctt cgtcaatctt ccgcgaggcc 4740
tcctccaggt cggagccggc gtagttgagg atgaccagca cgagcggcac cacctcccaa 4800

ctgttctagg gcagattgtt tagcttgttc agctgcgctt gtttatttgc ttagctttcg 4860
cttagcgacg tgttcacttt gcttgtttga attgaattgt cgctccgtag acgaagcgcc 4920
tctatttata ctccggcgct cgttttcgag tttaccactc cctatcagtg atagagaaaa 4980
gtgaaagtcg agtttaccac tccctatcag tgatagagaa aagtgaaagt cgagtttacc 5040
actccctatc agtgatagag aaaagtgaaa gtcgagttta ccactcccta tcagtgatag 5100
agaaaagtga aagtcgagtt taccactccc tatcagtgat agagaaaagt gaaagtcgaa 5160
acctggcgcg ccccggccat cgagaaagag agagagaaga gaagagagag aacattcgag 5220
aaagagagag agaagagaag agagagaaca tactccctat cagtgataga gaagtcccta 5280
tcagtgatag agatgtccct atcagtgata gagagttccc tatcagtgat agagacgtcc 5340
ctatcagtga tagagaagtc cctatcagtg atagagagat ccctatcagt gatagagatt 5400
tccctatcag tgatagagag gtccctatca gtgatagaga cttccctatc agtgatagag 5460
aaatccctat cagtgataga gacatcccta tcagtgatag agaactccct atcagtgata 5520
gagacctccc tatcagtgat agagatcgat gcggccgcga gcgccggagt ataaatagag 5580
gcgcttcgtc tacggagcga caattcaatt caaacaagca aagtgaacac gtcgctaagc 5640
gaaagctaag caaataaaca agcgcagctg aacaagctaa acaatctgca ggtaccctgg 5700
cggtaagttg atcaaaggaa acgcaaagtt ttcaagaaaa aacaaaacta atttgattta 5760
taacaccttt agaaagcggg gctagccacc atgggcagcg cctacagccg cgcccgtacc 5820
aagaacaact atggcagcac catcgaggga ctgctggacc tgccggatga cgatgccccg 5880
gaggaagccg gcctggccgc cccccgcctg agcttcctgc ccgccggaca cacgcgccgc 5940
ctgagcaccg ccccgccgac cgatgtgagc ctgggcgacg agctgcacct ggatggagag 6000
gatgtggcaa tggcccacgc cgacgccctg gacgatttcg acctggatat gctgggcgat 6060
ggagatagcc cgggaccggg cttcacgccc cacgatagcg ccccgtacgg cgccctggac 6120
atggccgact tcgagttcga gcaaatgttc accgacgcgc tgggcatcga tgagtatggc 6180
gggtaggttt aaactcgcgt taagatacat tgatgagttt ggacaaacca caactagaat 6240
gcagtgaaaa aaatgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccat 6300
tataagctgc aataaacaag ttaacaacaa caattgcatt cattttatgt ttcaggttca 6360
gggggaggtg tgggaggttt tttaaagcaa gtaaaacctc tacaaatgtg gtatggctga 6420
ttatgatcag ttatctagat ccggtggatc ttacgggtcc tccaccttcc gctttttctt 6480
gggtcgagat ctcaggaaca ggtggtggcg gccctcggtg cgctcgtact gctccacgat 6540
ggtgtagtcc tcgttgtggg aggtgatgtc cagcttggcg tccacgtagt agtagccggg 6600
cagctgcacg ggcttcttgg ccatgtagat ggacttgaac tccaccaggt agtggccgcc 6660
gtccttcagc ttcagggcct tgtgggtctc gcccttcagc acgccgtcgc gggggtacag 6720
gcgctcggtg gaggcctccc agcccatggt cttcttctgc atcacggggc cgtcggaggg 6780
gaagttcacg ccgatgaact tcaccttgta gatgaagcag ccgtcctgca gggaggagtc 6840
ctgggtcacg gtcgccacgc cgccgtcctc gaagttcatc acgcgctccc acttgaagcc 6900
ctcggggaag gacagcttct tgtagtcggg gatgtcggcg gggtgcttca cgtacacctt 6960
ggagccgtac tggaactggg gggacaggat gtcccaggcg aagggcaggg ggccgccctt 7020
ggtcaccttc agcttcacgg tgttgtggcc ctcgtagggg cggccctcgc cctcgccctc 7080
gatctcgaac tcgtggccgt tcacggtgcc ctccatgcgc accttgaagc gcatgaactc 7140
ggtgatgacg ttctcggagg aggccatggt ggcgaccggt ttgcgcttct tcttgggtgg 7200
ggtgggatct cccatggtgg cctgaatctc aacttgcacc tgaaggtagt gcagcaagga 7260
tgagcaaaag ggaagaaccc agaaaagaac gggaaaactt accccaatta gaattgcttg 7320
tcgccgccag tgtcaacttg caactgaaac aatatccaac atgaacgtca atttatactg 7380
ccctaatggc gaacacgata acaatatttc ttttattatg ccctctaaaa ccaacgcggt 7440
tatcgtttat ttattcaaat tagatataga acatccgccg acatacaatg ttaatgcaaa 7500
aacgcgtttg gtgagcggat acgaaaacag tcggccgata aacattaatc tgaggtcgat 7560
aacaccgtcc ttgaacggaa cacgaggagc gtacgtgatc agctgcattc gcgcgccgcg 7620
cctttatcga gatttatttg catacaacaa gtacactgcg ccgttgggat ttgtggtaac 7680
gcgcacacat gcagagctgc aagtgtggca cattttgtct gtgcgcaaaa cctttgaagc 7740

caaaagtacg aggtccgtta cgggcatgct actagcgcac acggacaatg gacccgacaa 7800
attctacgcc aaggatttaa tgataatgtc gggcaacgta tccgttcatt ttatcaataa 7860
cctacaaaaa tgtcgcgcgc atcacaaaga catcgatata tttaaacatt tatgtcccga 7920
actgcaaatc gataatagtg ttgtgcaacc tcgagcgtcc gtttgattta acgtatagct 7980
tgcaaatgaa ttatttaatt atcaatcatg ttttacgcgt agaattctac ccgtaaagcg 8040
agtttagtta tgagccatgt gcaaaacatg acatcagctt ttatttttat aacaaatgac 8100
atcatttctt gattgtgttt tacacgtaga attctactcg taaagcgagt tcagttttga 8160
aaaacaaatg acatcatctt tttgattgtg ctttacaagt agaattctac ccgtaaatca 8220
agttcggttt tgaaaaacaa atgagtcata ttgtatgata tcatattgca aaacaaatga 8280
ctcatcaatc gatcgtgcgt tacacgtaga attctactcg taaagcgagt ttatgagccg 8340
tgtgcaaaac atgacatcat ctcgatttga aaaacaaatg acatcatcca ctgatcgtgc 8400
attacaagta gaattctact cgtaaagcca gttcggttat gagccgtgta caaaacatga 8460
catcagatta tgactcatac ttgattgtgt tttacgcgta gaattctact cgtaaagcca 8520
gttcaatttt aaaaacaaat gacatcatcc aaattaataa atgacaagca atgggtacca 8580
tgcggcctgg cctcgcgctc gcgcgactga cggtcgtaag cacccgcgta cgtgtccacc 8640
ccggtcacaa ccccttgtgt catgtcggcg accctacgcc cccaactgag agaactcaaa 8700
ggttacccca gttggggcac tactcccgaa aaccgcttct gacctgggaa aacgtgaagc 8760
cccggggcat ccgctgaggg ttgccgccgg ggcttcggtg tgtccgtcag tacttaatta 8820
acaccgaaat cgtaattcac ggcatcatta caaaatattt tgacgttttg gacctcgtcc 8880
ctaatgacac cataacggtg gccttgaagt atatttaacc ctagaaagat agtctgcgta 8940
aaattgacgc atgcattctt gaaatattgc tctctctttc taaatagcgc gaatccgtcg 9000
ctgtgcattt aggacatctc agtcgccgct tggagctccc gtgaggcgtg cttgtcaatg 9060
cggtaagtgt cactgatttt gaactataac gaccgcgtga gtcaaaatga cgcatgatta 9120
tcttttacgt gacttttaag atttaactca tacgataatt atattgttat ttcatgttct 9180
acttacgtga taacttatta tatatatatt ttcttgttat agatatcgtg actaatatat 9240
aataaaatgg gtagttcttt agacgatgag catatcctct ctgctcttct gcaaagcgat 9300
gacgagcttg ttggtgagga ttctgacagt gaaatatcag atcacgtaag tgaagatgac 9360
ctcgaggatc caagcttatc gatttcgaac cctcgaccgc cggagtataa atagaggcgc 9420
ttcgtctacg gagcgacaat tcaattcaaa caagcaaagt gaacacgtcg ctaagcgaaa 9480
gctaagcaaa taaacaagcg cagctgaaca agctaaacaa tcggggtacc gctagagtcg 9540
atcccacccc acccaagaag aagcgcaaac cggtaccatg gcctcctccg agaacgtcat 9600
caccgagttc atgcgcttca aggtgcgcat ggagggcacc gtgaacggcc acgagttcga 9660
gatcgagggc gagggcgagg gccgccccta cgagggccac aacaccgtga agctgaaggt 9720
gaccaagggc ggcccccctgc ccttcgcctg ggacatcctg tcccccccagt tccagtacgg 9780
ctccaaggtg tacgtgaagc accccgccga catccccgac tacaagaagc tgtccttccc 9840
cgagggcttc aagtgggagc gcgtgatgaa cttcgaggac ggcggcgtgg cgaccgtgac 9900
ccaggactcc tccctgcagg acggctgctt catctacaag gtgaagttca tcggcgtgaa 9960
cttcccctcc gacggccccg tgatgcagaa gaagaccatg ggctgggagg cctccaccga 10020
gcgcctgtac ccccgcgacg gcgtgctgaa gggcgagacc cacaaggccc tgaagctgaa 10080
ggacggcggc cactacctgg tggagttcaa gtccatctac atggccaaga agcccgtgca 10140
gctgcccggc tactactacg tggacgccaa gctggacatc acctcccaca acgaggacta 10200
caccatcgtg gagcagtacg agcgcaccga gggccgccac cacctgttcc tgtgatgatc 10260
ataatcagcc ataccacatt tgtagaggtt ttacttgctt taaaaaacct cccacacctc 10320
cccctgaacc tgaaacataa aatgaatgca attgttgttg ttaacttgtt tattgcagct 10380
tataatggtt acaaataaag caatagcatc acaaatttca caaataaagc attttttttca 10440
ctgcattcta gttgtggttt gtccaaactc atcaatgtat cttaacgcga gttaattacg 10500
gccgctcatt taaatctggc cggccgcaac cattgtggga accgtgcgat caaacaaacg 10560
cgagataccg gaagtactga aaaacagtcg ctccaggcca gtgggaacat cgatgttttg 10620
ttttgacgga ccccttactc tcgtctcata taaaccgaag ccagctaaga tggtatactt 10680

attatcatct tgtgatgagg atgcttctat caacgaaagt accggtaaac cgcaaatggt 10740
tatgtattat aatcaaacta aaggcggagt ggacacgcta gaccaaatgt gttctgtgat 10800
gacctgcagt aggaagacga ataggtggcc tatggcatta ttgtacggaa tgataaacat 10860
tgcctgcata aattctttta ttatatacag ccataatgtc agtagcaagg gagaaaaggt 10920
ccaaagtcgc aaaaaattta tgagaaacct ttacatgagc ctgacgtcat cgtttatgcg 10980
taagcgttta gaagctccta ctttgaagag atatttgcgc gataatatct ctaatatttt 11040
gccaaatgaa gtgcctggta catcagatga cagtactgaa gagccagtaa tgaaaaaacg 11100
tacttactgt acttactgcc cctctaaaat aaggcgaaag gcaaatgcat cgtgcaaaaa 11160
atgcaaaaaa gttatttgtc gagagcataa tattgatatg tgccaaagtt gtttctgact 11220
gactaataag tataatttgt ttctattatg tataagttaa gctaattact tattttataa 11280
tacaacatga ctgtttttaa agtacaaaat aagtttattt ttgtaaaaga gagaatgttt 11340
aaaagttttg ttactttata gaagaaattt tgagtttttg ttttttttta ataaataaat 11400
aaacataaat aaattgtttg ttgaatttat tattagtatg taagtgtaaa tataataaaa 11460
cttaatatct attcaaatta ataaataaac ctcgatatac agaccgataa aacacatgcg 11520
tcaattttac gcatgattat ctttaacgta cgtcacaata tgattatctt tctagggtta 11580
aataatagtt tctaattttt ttattattca gcctgctgtc gtgaataccg tatatctcaa 11640
cgctgtctgt gagattgtcg tattctagcc tttttagttt ttcgctcatc gacttgatat 11700
tgtccgacac attttcgtcg atttgcgttt tgatcaaaga cttgagcaga gacacgttaa 11760
tcaactgttc aaattgatcc atattaacga tatcaacccg atgcgtatat ggtgcgtaaa 11820
atatattttt taaccctctt atactttgca ctctgcgtta atacgcgttc gtgtacagac 11880
gtaatcatgt tttcttttct ggataaaact cctactgagt ttgacctcat attagaccct 11940
cacaagttgc aaaacgtggc atttttttacc aatgaagaat ttaaagttat tttaaaaaat 12000
ttcatcacag atttaaagaa gaaccaaaaa ttaaattatt tcaacagttt aatcgaccag 12060
ttaatcaacg tgtacacaga cgcgtcggca aaaaacacgc agcccgacgt gttggctaaa 12120
attattaaat caacttgtgt tatagtcacg gatttgccgt ccaacgtgtt cctcaaaaag 12180
ttgaagacca acaagtttac ggacactatt aattatttga ttttgcccca cttcattttg 12240
tgggatcaca attttgttat attttaaaca aagcttggca ctggccgtcg ttttacaacg 12300
tcgtgactgg gaaaaccctg gcgttaccca acttaatcgc cttgcagcac atccccctttt 12360
cgccagctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac agttgcgcag 12420
cctgaatggc gaatggcgcc tgatgcggta ttttctcctt acgcatctgt gcggtatttc 12480
acaccgcata tggtgcactc tcagtacaat ctgctctgat gccgcatagt taagccagcc 12540
ccgacacccg ccaacacccg ctgacgcgcc ctgacgggct tgtctgctcc cggcatccgc 12600
ttacagacaa gctgtgaccg tctccgggag ctgcatgtgt cagaggtttt caccgtcatc 12660
accgaaacgc gcgagacgaa agggcctcgt gatacgccta tttttatagg ttaatgtcat 12720
gataataatg gtttcttaga cgtcaggtgg cacttttcgg ggaaatgtgc gcggaacccc 12780
tatttgttta ttttttctaaa tacattcaaa tatgtatccg ctcatgagac aataaccctg 12840
ataaatgctt caataatatt gaaaaaggaa gagtatgagt attcaacatt tccgtgtcgc 12900
ccttattccc ttttttgcgg cattttgcct tcctgttttt gctcacccag aaacgctggt 12960
gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg ggttacatcg aactggatct 13020
caacagcggt aagatccttg agagttttcg ccccgaagaa cgttttccaa tgatgagcac 13080
ttttaaagtt ctgctatgtg gcgcggtatt atcccgtatt gacgccgggc aagagcaact 13140
cggtcgccgc atacactatt ctcagaatga cttggttgag tactcaccag tcacagaaaa 13200
gcatcttacg gatggcatga cagtaagaga attatgcagt gctgccataa ccatgagtga 13260
taacactgcg gccaacttac ttctgacaac gatcggagga ccgaaggagc taaccgcttt 13320
tttgcacaac atgggggatc atgtaactcg ccttgatcgt tgggaaccgg agctgaatga 13380
agccatacca aacgacgagc gtgacaccac gatgcctgta gcaatggcaa caacgttgcg 13440
caaactatta actggcgaac tacttactct agcttcccgg caacaattaa tagactggat 13500
ggaggcggat aaagttgcag gaccacttct gcgctcggcc cttccggctg gctggtttat 13560
tgctgataaa tctggagccg gtgagcgtgg gtctcgcggt atcattgcag cactggggcc 13620

agatggtaag ccctcccgta tcgtagttat ctacacgacg gggagtcagg caactatgga 13680
tgaacgaaat agacagatcg ctgagatagg tgcctcactg attaagcatt ggtaactgtc 13740
agaccaagtt tactcatata tactttagat tgatttaaaa cttcattttt aatttaaaag 13800
gatctaggtg aagatccttt ttgataatct catgaccaaa atcccttaac gtgagtttc 13860
gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag atccttttt 13920
tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg tggtttgttt 13980
gccggatcaa gagctaccaa ctcttttcc gaaggtaact ggcttcagca gagcgcagat 14040
accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga actctgtagc 14100
accgcctaca tacctcgctc tgctaatcct gttaccagtg gctgctgcca gtggcgataa 14160
gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc agcggtcggg 14220
ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca ccgaactgag 14280
atacctacag cgtgagcatt gagaaagcgc cacgcttccc gaagggagaa aggcggacag 14340
gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc cagggggaaa 14400
cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc gtcgattttt 14460
gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg cctttttacg 14520
gttcctggcc ttttgctggc cttttgctca catgttcttt cctgcgttat cccctgattc 14580
tgtggataac cgtattaccg cctttgagtg agctgatacc gctcgccgca gccgaacgac 14640
cgagcgcagc gagtcagtga gcgaggaagc ggaagagcgc ccaatacgca aaccgcctct 14700
ccccgcgcgt tggccgattc attaatgcag ctggcacgac aggtttcccg actggaaagc 14760
gggcagtgag cgcaacgcaa ttaatgtgag ttagctcact cattaggcac cccaggcttt 14820
acactttatg cttccggctc gtatgttgtg tggaattgtg agcggataac aatttcacac 14880
aggaaacagc tatgaccatg attacgaatt tcgacctgca ggcatgcaag cttgcatgcc 14940
tgcaggtcga cgctcgcgcg acttggtttg ccattcttta gcgcgcgtcg cgtcacacag 15000
cttggccaca atgtggtttt tgtcaaacga agattctatg acgtgtttaa agtttaggtc 15060
gagtaaagcg caaatctttt ttaaccctag aaagatagtc tgcgtaaaat tgacgcatgc 15120
attcttgaaa tattgctctc tctttctaaa tagcgcgaat ccgtcgctgt gcatttagga 15180
catctcagtc gccgcttgga gctcccgtga ggcgtgcttg tcaatgcggt aagtgtcact 15240
gattttgaac tataacgacc gcgtgagtca aaatgacgca tgattatctt ttacgtgact 15300
tttaagattt aactcatacg ataattatat tgttatttca tgttctactt acgtgataac 15360
ttattatata tatattttct tgttatagat atcgtgacta atatataata aaatgggtag 15420
ttctttagac gatgagcata tcctctctgc tcttctgcaa agcgatgacg agcttgttgg 15480
tgaggattct gacagtgaaa tatcagatca cgtaagtgaa gatgacgtcc agagcgatac 15540
agaagaagcg tttatagatg aggtacatga agtgcagcca acgtcaagcg gtagtgaaat 15600
attagacgaa caaaatgtta ttgaacaacc aggttcttca ttggcttcta acagaatctt 15660
gaccttgcca cagaggacta ttagaggtaa gaataaacat tgttggtcaa cttcaaagtc 15720
cacgaggcgt agccgagtct ctgcactgaa cattgtcaga tcggcccgct cgcccgggga 15780
actagttcaa ttagagacta attcaattag agctaattca attaggatcc aagcttatcg 15840
atttcgaacc ctcgaccgcc ggagtataaa tagaggcgct tcgtctacgg agcgacaatt 15900
caattcaaac aagcaaagtg aacacgtcgc taagcgaaag ctaagcaaat aaacaagcgc 15960
agctgaacaa gctaaacaat cggggtaccg ctagagtcga tcccacccca cccaagaaga 16020
agcgcaaacc ggtcgccacc atggccctgt ccaacaagtt catcggcgac gacatgaaga 16080
tgacctacca catggacggc tgcgtgaacg gccactactt caccgtgaag ggcgagggca 16140
gcggcaagcc ctacgagggc acccagacct ccaccttcaa ggtgaccatg gccaacggcg 16200
gccccctggc cttctccttc gacatcctgt ccaccgtgtt catgtacggc aaccgctgct 16260
tcaccgccta ccccaccagc atgcccgact acttcaagca ggccttcccc gacggcatgt 16320
cctacgagag aaccttcacc tacgaggacg gcggcgtggc caccgccagc tgggagatca 16380
gcctgaaggg caactgcttc gagcacaagt ccaccttcca cggcgtgaac ttccccgccg 16440
acggcccccgt gatggccaag aagaccaccg gctgggaccc ctccttcgag aagatgaccg 16500
tgtgcgacgg catcttgaag ggcgacgtga ccgccttcct gatgctgcag ggcggcggca 16560

actacagatg ccagttccac acctcctaca agaccaagaa gcccgtgacc atgccccca 16620
accacgtggt ggagcaccgc atcgccagaa ccgacctgga caagggcggc aacagcgtgc 16680
agctgaccga gcacgccgtg gcccacatca cctccgtggt gcccttctcc ggactcagat 16740
cataatcagc cataccacat ttgtagaggt tttacttgct ttaaaaaacc tcccacacct 16800
ccccctgaac ctgaaacata aaatgaatgc aattgttgtt gttaacttgt ttattgcagc 16860
ttataatggt tacaaataaa gcaatagcat cacaaatttc acaaataaag catttttttc 16920
actgcattct agttgtggtt tgtccaaact catcaatgta tcttaccgcg gagtggacac 16980
gctagaccaa atgtgttctg tgatgacctg cagtaggaag acgaataggt ggcctatggc 17040
attattgtac ggaatgataa acattgcctg cataaattct tttattatat acagccataa 17100
tgtcagtagc aagggagaaa aggtccaaag tcgcaaaaaa tttatgagaa acctttacat 17160
gagcctgacg tcatcgttta tgcgtaagcg tttagaagct cctactttga agagatattt 17220
gcgcgataat atctctaata ttttgccaaa tgaagtgcct ggtacatcag atgacagtac 17280
tgaagagcca gtaatgaaaa aacgtactta ctgtacttac tgcccctcta aaataaggcg 17340
aaaggcaaat gcatcgtgca aaaaatgcaa aaaagttatt tgtcgagagc ataatattga 17400
tatgtgccaa agttgtttct gactgactaa taagtataat ttgtttctat tatgtataag 17460
ttaagctaat tacttatttt ataatacaac atgactgttt ttaaagtaca aaataagttt 17520
atttttgtaa aagagagaat gtttaaaagt tttgttactt tatagaagaa attttgagtt 17580
tttgtttttt tttaataaat aaataaacat aaataaattg tttgttgaat ttattattag 17640
tatgtaagtg taaatataat aaaacttaat atctattcaa attaataaat aaacctcgat 17700
atacagaccg ataaaacaca tgcgtcaatt ttacgcatga ttatctttaa cgtacgtcac 17760
aatatgatta tctttctagg g 17781

<210> 162
<211> 15482
<212> ADN
<213> Nhân tạo
<220>
<223> Trình tự plasmit LA3570

<220>
<221> Dấu hiệu hỗn tạp
<222> (1875)..(1875)
<223> n là a, c, g hoặc t
<400> 162
gggcggccgt ttttcttgaa atattgctct ctctttctaa atagcgcgaa tccgtcgctg 60
tgcatttagg acatctcagt cgccgcttgg agctcccaaa cgcgccagtg gtagtacaca 120
gtactgtggg tgttcagttt gaaatcctct tgcttctcca ttgtctcggt tacctttggt 180
caaatccatg ggttctattg cctatatact cttgcgatta ccagtgattg cgctattagc 240
tattagatgg attgttggcc aaacttgtcg cttaagtggc tgggaattgt aaccgtaggc 300
ccgagtgtaa tgatccccca taaaaagttt tcgcaatgcc tttatttttt gttgcaaatc 360
tctctttatt ctgcggtatt cttcattatt gcggggatgg ggaaagtgtt tatatagaag 420
caacttacga ttgaacccaa atgcacctga caagcaaggt caaagggcca gatttttaaa 480
tatattattt agtcttagga ctctctattt gcaattaaat tactttgcta cctgagggtt 540
aaatcttccc cattgataat aataattcca ctatatgttc aattgggttt caccgcgctt 600
agttacatga cgagccctaa tgagccgtcg gtggtctata aactgtgcct tacaaatact 660
tgcaactctt ctcgttttga agtcagcaga gttattgcta attgctaatt gctaattgct 720
tttaactgat ttcttcgaaa ttggtgctat gtttatggcg ctattaacaa gtatgaatgt 780
caggtttaac caggggatgc ttaattgtgt tctcaacttc aaaggcagaa atgtttactc 840
ttgaccatgg gtttaggtat aatgttatca agctcctcga gttaacgtta cgttaacgtt 900

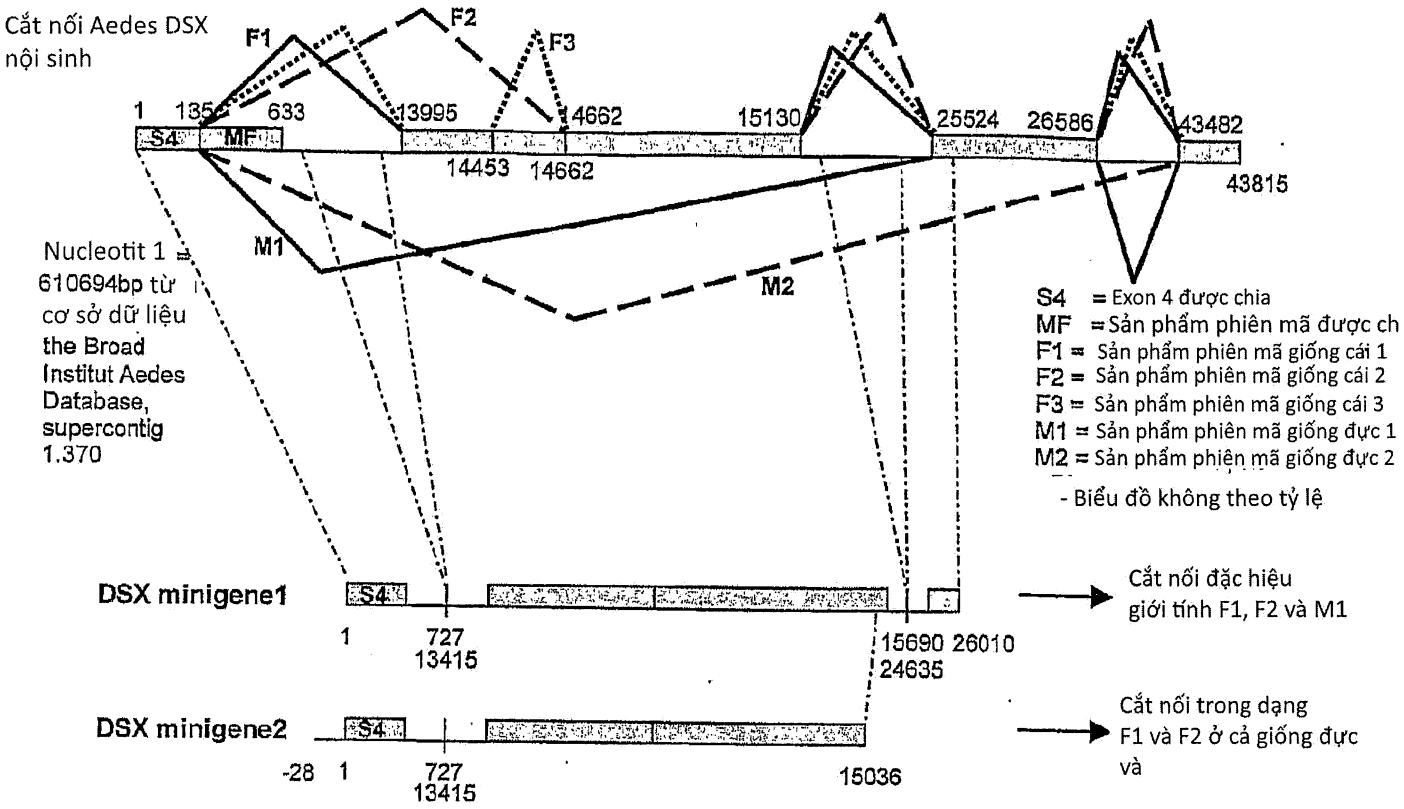
aacgttcgag gtcgactcta gggcctctct agatttacag gtctattttg agctctttgt 960
cagacactgt ttgcttgaaa ttcaagtctg tcagcacctt aaaaccaaaa ataaaaagaa 1020
taataaatga aatagtactt acttcccgcg gcgcaggttc gcatcgctac aagtgcgcgg 1080
gcggcgggga tgatctctgc gtggtaagcg gcagaggcaa caggtgtggc gcgtactcgg 1140
gcgcgatgac gtagcggggt gagcagcaca ccgagtacgt ccctttgcgc gcttgcagct 1200
ccaggagcga gcacagcgac cgctcgtaca tccgccactg gtggaccacc caatgtgcac 1260
ccaatgtgct gcaaggaagg cggggttaag tcgtcgagaa gtgatacaag aaatcggtct 1320
ttaaagtcgt aaggtccatt acctttaaaa atcgaaaacc cttaaactac tgtgtctaga 1380
aatctggacc ttacgaggtt aagtcgttag agaattgaaa gaaagcataa agaaactaga 1440
ccatatcatc gccttgtagc gaaaccacgt aagcgttttt ttgaaaatca aattaaaaac 1500
attctgatac gattttcttc aacaaaattt cattacaggt aaaaattaag accacraatt 1560
attgcctggg ttgaattgaa acaagcttgt ctattgtgtg gttttattaa caaaaatcac 1620
atccgaaggc gcttrtgtgg gtttcattat aaagccacga tatacagtct atacatttag 1680
ctgttcaagt tacaggtgaa tcgcaacctc caaggttaca agcggtataa aattwatatt 1740
gttaataatg tcaaatgtac caactatagt tttacattgg tcaaatgagc aatgtacggc 1800
cgtaaaatgg ccagtcgcag tgccagtaat gtagtttttt aaatccgtaa aaattaagtg 1860
ccatacyttt tttanctacc ttaaaataca aaaatattgg gaacmcacga acaccccaat 1920
aatagtgttt aaacagtcgt tgtcataaaa cgatatcaat aatctttgat gttataaaaa 1980
tatatgtttt tctttatttt aattgcccgg tagtcatgtt gtatacgagt attgtataaa 2040
gcaatcgttc tacaaatgac tcgttacgat gttcctgaga ttcctccata gcagtgagta 2100
gtaataaaaa gtcaattgta ccgcgatgaa atagataaaa tattattcta ccactcaccg 2160
aatagtccag cgtggcgacg acacaatagt ccttccatcg caaacagcaa cgcacagcaa 2220
aagtccacac aacacacagc acatacaaaa caaagtatca ttcgcaaaat aacttcatgg 2280
acgacgatag tacaccactt atattattaa tttcgctcag cattttccac cggtgttagc 2340
cgccgtactc atcgatgccc agggcgtcgg tgaacatctg ctcgaactcg aaatcggcca 2400
tatccagggc gccgtagggg gcgctatcgt gcggggtgaa tcccggtccc gggctatcgc 2460
catcgcccag catgtccagg tcgaagtcgt ccagggcatc ggcgtgggcc atcgccacat 2520
cctcgccatc caggtgcagc tcatcgccca ggctcacgtc ggtcggcggg gcggtcgaca 2580
ggcggcgggt gtgtccggcc ggcaggaagc tcaggcgcgg ggcggccagg cccgcctcct 2640
ccggggcatc atcatccggc agatccagca ggccctcgat ggtgctgccg tagttgttct 2700
tggtgcgggc gcggctgtag gcggggcccg agcccgactc gcatttcagt tgcttttcca 2760
atccgcagat aatcagctcc aagccgaaca ggaatgccgg ctcggctcct tgatgatcga 2820
acagctcgat tgcctgacgc agcagtgggg gcatcgaatc ggttgttggg gtctcgcgct 2880
cctcttttgc gacttgatgc tcttggtcct ccagcacgca gcccagggta aagtgaccga 2940
cggcgctcag agcgtagaga gcattttcca ggctgaagcc ttgctggcac aggaacgcga 3000
gctggttctc cagtgtctcg tattgctttt cggtcgggcg cgtgccgaga tggactttgg 3060
caccgtctcg gtgggacagc agagcgcagc ggaacgactt ggcgttattg cggaggaagt 3120
cctgccagga ctcgccttcc aacgggcaaa aatgcgtgtg gtggcggtcg agcatctcga 3180
tggccagggc atccagcagc gcccgcttat tcttcacgtg ccagtagagg gtgggctgct 3240
ccacgcccag cttctgcgcc aacttgcggg tcgtcagtcc ctcaatgcca acttcgttca 3300
acagctccaa cgcggagttg atgactttgg acttatccag gcggctgccc atggtggttt 3360
cggtccgtta gcgagtcgag ttcctcagct cgtggccatc gaagatgttc agattgtgct 3420
tcctcgcgta ctcgttgatg atcatcttcc ctggaaacat atgacgctag ctttacattc 3480
gcacagcggg gtatgaggaa ctgcatttat tacaatttat tatactatta ttataattcc 3540
cgtcgtcata attgtcgtcg gtcatgtcgt atcaggaggt gaaggatttg gtaggaagaa 3600
gagaggaatg gcgattactc caccgacaag agcgcagctc ttaaaaaaaa agagagataa 3660
ttcccgtgac cttaatataa gcatcatggc ttcataacct cgtgagaaaa cgcacataat 3720
ttcccgagaa atgcgtttcg gaggtgacct aaccagccca atacctgtgt tgtttgcctt 3780
cgggttggaa ggtcagatag gcattcaatt ctgtaatgaa ccggacctgt caaatcttca 3840

ggctaagtac agaaattata ccatcaaata aggtaacata attttgatca gatttcttta 3900
ttatttattt atctttagaa gacagagaga tgaggaggaa gggtgcagac aacattgcat 3960
cctacgtgca ctcaagaaca agtagaatgt ctacttgtat ttactaccta aaatacattt 4020
tattggacct cctagattta attacagttt tgaaatctct aacatctaaa ataatagccc 4080
cgggccttca attattgtaa aaggggaatg aatcttatgt tactataggt agtttcgcct 4140
cgagaggcat tcgcaacttg accgaacaaa cggtttcttc ctttagcgaa tgtattataa 4200
ttatccaaca cacaagactg cacgcagtac aagtaggtaa taatgcaata gattgacata 4260
aacggcaatt aacgaacgac agacgtacct accgcggtgt agagttgtag acctatgatt 4320
attcttcacg gagtttttta ttacaaactg tggtaaaacc tttataaacc accgtaatat 4380
acaagaataa agaacgaaac taattatgta taaacaactt atataaatac cactgctgga 4440
cgcagacgtc ccctcaatca actggacagg gaagatcgta ctccaccacg ctgcttcgtt 4500
acgggttggt agagaattaa ataaatgaat tgtatgaaaa aaaaaacgta agtaaacata 4560
taaaaaatgt aagttttcta tcaaaaactt cacctcgtat tcaaagaacg caaagaactt 4620
gtaatcaatc agtaattatc gtaccttcat caatttttcc agaagcctcg tcgaggccta 4680
gggcagattg tttagcttgt tcagctgcgc ttgtttattt gcttagcttt cgcttagcga 4740
cgtgttcact ttgcttgttt gaattgaatt gtcgctccgt agacgaagcg cctctattta 4800
tactccggcg ctcgttttcg agtttaccac tccctatcag tgatagagaa aagtgaaagt 4860
cgagtttacc actccctatc agtgatagag aaaagtgaaa gtcgagttta ccactcccta 4920
tcagtgatag agaaaagtga aagtcgagtt taccactccc tatcagtgat agagaaaagt 4980
gaaagtcgag tttaccactc cctatcagtg atagagaaaa gtgaaagtcg agtttaccac 5040
tccctatcag tgatagagaa aagtgaaagt cgagtttacc actccctatc agtgatagag 5100
aaaagtgaaa gtcgaaacct ggcgcgcccc ggccatcgag aaagagagag agaagagaag 5160
agagagaaca ttcgagaaag agagagagaa gagaagagag agaacatact ccctatcagt 5220
gatagagaag tccctatcag tgatagagat gtccctatca gtgatagaga gttccctatc 5280
agtgatagag acgtccctat cagtgataga gaagtcccta tcagtgatag agagatccct 5340
atcagtgata gagatttccc tatcagtgat agagaggtcc ctatcagtga tagagacttc 5400
cctatcagtg atagagaaat ccctatcagt gatagagaca tccctatcag tgatagagaa 5460
ctccctatca gtgatagaga cctccctatc agtgatagag atcgatgcgg ccgcgagcgc 5520
cggagtataa atagaggcgc ttcgtctacg gagcgacaat tcaattcaaa caagcaaagt 5580
gaacacgtcg ctaagcgaaa gctaagcaaa taaacaagcg cagctgaaca agctaaacaa 5640
tctgcaggta ccctggcggt aagttgatca aaggaaacgc aaagttttca agaaaaaaca 5700
aaactaattt gatttataac acctttagaa agcggggcta gccaccatgg gcagcgccta 5760
cagccgcgcc cgtaccaaga acaactatgg cagcaccatc gagggactgc tggacctgcc 5820
ggatgacgat gccccggagg aagccggcct ggccgccccc cgcctgagct tcctgcccgc 5880
cggacacacg cgccgcctga gcaccgcccc gccgaccgat gtgagcctgg gcgacgagct 5940
gcacctggat ggagaggatg tggcaatggc ccacgccgac gccctggacg atttcgacct 6000
ggatatgctg ggcgatggag atagccccgg accgggcttc acgccccacg atagcgcccc 6060
gtacggcgcc ctggacatgg ccgacttcga gttcgagcaa atgttcaccg acgcgctggg 6120
catcgatgag tatggcgggt aggtttaaac tcgcgttaag atacattgat gagtttggac 6180
aaaccacaac tagaatgcag tgaaaaaaat gctttatttg tgaaatttgt gatgctattg 6240
ctttatttgt aaccattata agctgcaata aacaagttaa caacaacaat tgcattcatt 6300
ttatgtttca ggttcagggg gaggtgtggg aggtttttta aagcaagtaa aacctctaca 6360
aatgtggtat ggctgattat gatcagttat ctagatccgg tggatcttac gggtcctcca 6420
ccttccgctt tttcttgggt cgagatctca ggaacaggtg gtggcggccc tcggtgcgct 6480
cgtactgctc cacgatggtg tagtcctcgt tgtgggaggt gatgtccagc ttggcgtcca 6540
cgtagtagta gccgggcagc tgcacgggct tcttggccat gtagatggac ttgaactcca 6600
ccaggtagtg gccgccgtcc ttcagcttca gggccttgtg ggtctcgccc ttcagcacgc 6660
cgtcgcgggg gtacaggcgc tcggtggagg cctcccagcc catggtcttc ttctgcatca 6720
cggggccgtc ggaggggaag ttcacgccga tgaacttcac cttgtagatg aagcagccgt 6780

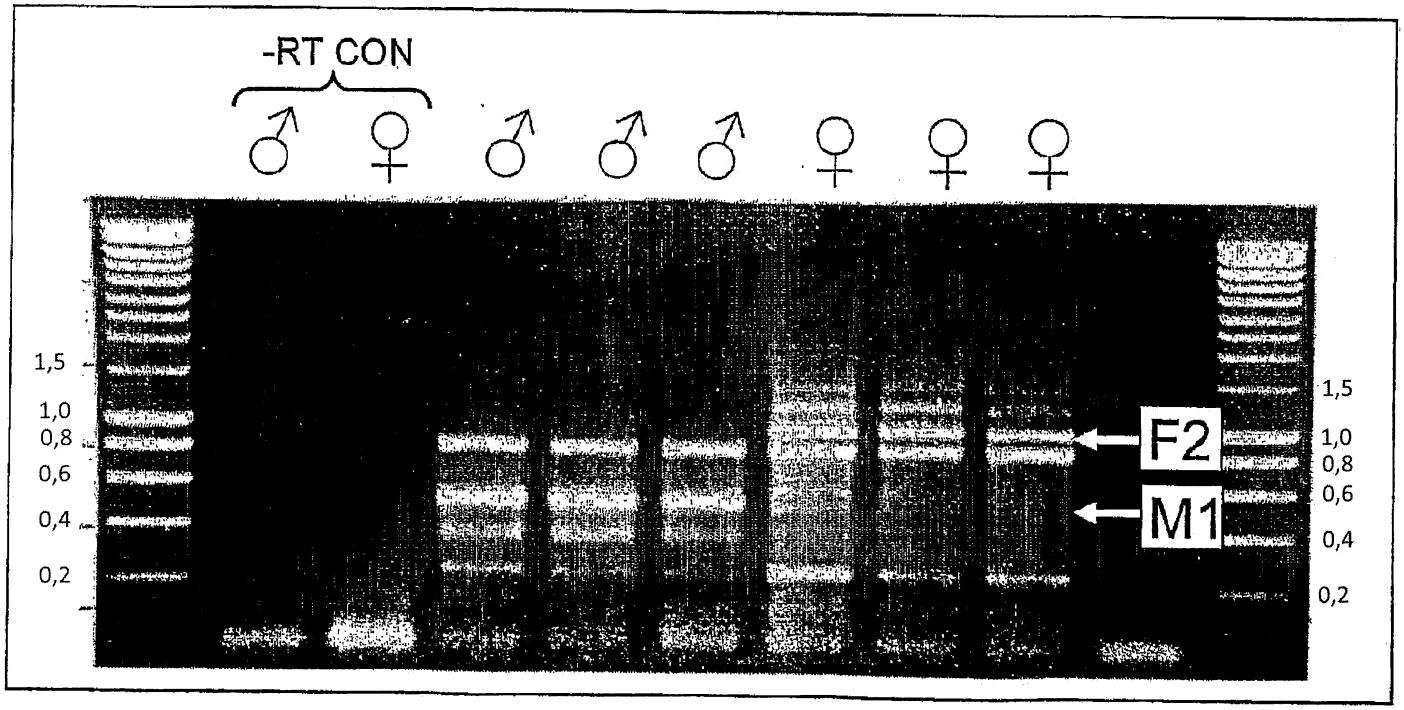
cctgcaggga ggagtcctgg gtcacggtcg ccacgccgcc gtcctcgaag ttcatcacgc 6840
gctcccactt gaagccctcg gggaaggaca gcttcttgta gtcgggggatg tcggcggggt 6900
gcttcacgta caccttggag ccgtactgga actgggggga caggatgtcc caggcgaagg 6960
gcaggggggcc gcccttggtc accttcagct tcacggtgtt gtggccctcg taggggcggc 7020
cctcgccctc gccctcgatc tcgaactcgt ggccgttcac ggtgccctcc atgcgcacct 7080
tgaagcgcat gaactcggtg atgacgttct cggaggaggc catggtggcg accggtttgc 7140
gcttcttctt gggtggggtg ggatctccca tggtggcctg aatctcaact tgcacctgaa 7200
ggtagtgcag caaggatgag caaaagggaa gaacccagaa aagaacggga aaacttaccc 7260
caattagaat tgcttgtcgc cgccagtgtc aacttgcaac tgaaacaata tccaacatga 7320
acgtcaattt atactgccct aatggcgaac acgataacaa tatttctttt attatgccct 7380
ctaaaaccaa cgcggttatc gtttatttat tcaaattaga tatagaacat ccgccgacat 7440
acaatgttaa tgcaaaaacg cgtttggtga gcggatacga aaacagtcgg ccgataaaca 7500
ttaatctgag gtcgataaca ccgtccttga acggaacacg aggagcgtac gtgatcagct 7560
gcattcgcgc gccgcgcctt tatcgagatt tatttgcata caacaagtac actgcgccgt 7620
tgggatttgt ggtaacgcgc acacatgcag agctgcaagt gtggcacatt ttgtctgtgc 7680
gcaaaacctt tgaagccaaa agtacgaggt ccgttacggg catgctacta gcgcacacgg 7740
acaatggacc cgacaaattc tacgccaagg atttaatgat aatgtcgggc aacgtatccg 7800
ttcattttat caataaccta caaaaatgtc gcgcgcatca caaagacatc gatatattta 7860
aacatttatg tcccgaactg caaatcgata atagtgttgt gcaacctcga gcgtccgttt 7920
gatttaacgt atagcttgca aatgaattat ttaattatca atcatgtttt acgcgtagaa 7980
ttctacccgt aaagcgagtt tagttatgag ccatgtgcaa aacatgacat cagcttttat 8040
ttttataaca aatgacatca tttcttgatt gtgttttaca cgtagaattc tactcgtaaa 8100
gcgagttcag ttttgaaaaa caaatgacat catcttttttg attgtgcttt acaagtagaa 8160
ttctacccgt aaatcaagtt cggttttgaa aaacaaatga gtcatattgt atgatatcat 8220
attgcaaaac aaatgactca tcaatcgatc gtgcgttaca cgtagaattc tactcgtaaa 8280
gcgagtttat gagccgtgtg caaaacatga catcatctcg atttgaaaaa caaatgacat 8340
catccactga tcgtgcatta caagtagaat tctactcgta aagccagttc ggttatgagc 8400
cgtgtacaaa acatgacatc agattatgac tcatacttga ttgtgtttta cgcgtagaat 8460
tctactcgta aagccagttc aattttaaaa acaaatgaca tcatccaaat taataaatga 8520
caagcaatgg gtaccatgcg gcctggcctc gcgctcgcgc gactgacggt cgtaagcacc 8580
cgcgtacgtg tccaccccgg tcacaacccc ttgtgtcatg tcggcgaccc tacgcccccа 8640
actgagagaa ctcaaaggtt accccagttg gggcactact cccgaaaacc gcttctgacc 8700
tgggaaaacg tgaagccccg gggcatccgc tgagggttgc cgccggggct tcggtgtgtc 8760
cgtcagtact taattaacac cgaaatcgta attcacggca tcattacaaa atattttgac 8820
gttttggacc tcgtccctaa tgacaccata acggtggcct tgaagtatat ttaaccctag 8880
aaagatagtc tgcgtaaaat tgacgcatgc attcttgaaa tattgctctc tctttctaaa 8940
tagcgcgaat ccgtcgctgt gcatttagga catctcagtc gccgcttgga gctcccgtga 9000
ggcgtgcttg tcaatgcggt aagtgtcact gattttgaac tataacgacc gcgtgagtca 9060
aaatgacgca tgattatctt ttacgtgact tttaagattt aactcatacg ataattatat 9120
tgttatttca tgttctactt acgtgataac ttattatata tatattttct tgttatagat 9180
atcgtgacta atatataata aaatgggtag ttctttagac gatgagcata tcctctctgc 9240
tcttctgcaa agcgatgacg agcttgttgg tgaggattct gacagtgaaa tatcagatca 9300
cgtaagtgaa gatgacgtcc aggaaatctg gccggccgca accattgtgg gaaccgtgcg 9360
atcaaacaaa cgcgagatac cggaagtact gaaaaacagt cgctccaggc cagtgggaac 9420
atcgatgttt tgttttgacg gaccccttac tctcgtctca tataaaccga agccagctaa 9480
gatggtatac ttattatcat cttgtgatga ggatgcttct atcaacgaaa gtaccggtaa 9540
accgcaaatg gttatgtatt ataatcaaac taaaggcgga gtggacacgc tagaccaaat 9600
gtgttctgtg atgacctgca gtaggaagac gaataggtgg cctatggcat tattgtacgg 9660
aatgataaac attgcctgca taaattcttt tattatatac agccataatg tcagtagcaa 9720

gggagaaaag gtccaaagtc gcaaaaaatt tatgagaaac ctttacatga gcctgacgtc 9780
atcgtttatg cgtaagcgtt tagaagctcc tactttgaag agatatttgc gcgataatat 9840
ctctaatatt ttgccaaatg aagtgcctgg tacatcagat gacagtactg aagagccagt 9900
aatgaaaaaa cgtacttact gtacttactg cccctctaaa ataaggcgaa aggcaaatgc 9960
atcgtgcaaa aaatgcaaaa aagttatttg tcgagagcat aatattgata tgtgccaaag 10020
ttgtttctga ctgactaata agtataattt gtttctatta tgtataagtt aagctaatta 10080
cttattttat aatacaacat gactgttttt aaagtacaaa ataagtttat ttttgtaaaa 10140
gagagaatgt ttaaaagttt tgttacttta tagaagaaat tttgagtttt tgtttttttt 10200
taataaataa ataaacataa ataaattgtt tgttgaattt attattagta tgtaagtgta 10260
aatataataa aacttaatat ctattcaaat taataaataa acctcgatat acagaccgat 10320
aaaacacatg cgtcaatttt acgcatgatt atctttaacg tacgtcacaa tatgattatc 10380
tttctagggt taaataatag tttctaattt ttttattatt cagcctgctg tcgtgaatac 10440
cgtatatctc aacgctgtct gtgagattgt cgtattctag ccttttttagt ttttcgctca 10500
tcgacttgat attgtccgac acattttcgt cgatttgcgt tttgatcaaa gacttgagca 10560
gagacacgtt aatcaactgt tcaaattgat ccatattaac gatatcaacc cgatgcgtat 10620
atggtgcgta aaatatattt tttaacccctc ttatactttg cactctgcgt taatacgcgt 10680
tcgtgtacag acgtaatcat gttttctttt ttggataaaa ctcctactga gtttgacctc 10740
atattagacc ctcacaagtt gcaaaacgtg gcatttttta ccaatgaaga atttaaagtt 10800
attttaaaaa atttcatcac agatttaaag aagaaccaaa aattaaatta tttcaacagt 10860
ttaatcgacc agttaatcaa cgtgtacaca gacgcgtcgg caaaaaacac gcagcccgac 10920
gtgttggcta aaattattaa atcaacttgt gttatagtca cggatttgcc gtccaacgtg 10980
ttcctcaaaa agttgaagac caacaagttt acggacacta ttaattattt gattttgccc 11040
cacttcattt tgtgggatca caattttgtt atattttaaa caaagcttgg cactggccgt 11100
cgttttacaa cgtcgtgact gggaaaaccc tggcgttacc caacttaatc gccttgcagc 11160
acatcccccc ttcgccagct ggcgtaatag cgaagaggcc cgcaccgatc gcccttccca 11220
acagttgcgc agcctgaatg gcgaatggcg cctgatgcgg tattttctcc ttacgcatct 11280
gtgcggtatt tcacaccgca tatggtgcac tctcagtaca atctgctctg atgccgcata 11340
gttaagccag ccccgacacc cgccaacacc cgctgacgcg ccctgacggg cttgtctgct 11400
cccggcatcc gcttacagac aagctgtgac cgtctccggg agctgcatgt gtcagaggtt 11460
ttcaccgtca tcaccgaaac gcgcgagacg aaagggcctc gtgatacgcc tatttttata 11520
ggttaatgtc atgataataa tggtttctta gacgtcaggt ggcacttttc ggggaaatgt 11580
gcgcggaacc cctatttgtt tatttttcta aatacattca aatatgtatc cgctcatgag 11640
acaataaccc tgataaatgc ttcaataata ttgaaaaagg aagagtatga gtattcaaca 11700
tttccgtgtc gcccttattc cctttttttgc ggcattttgc cttcctgttt ttgctcaccc 11760
agaaacgctg gtgaaagtaa aagatgctga agatcagttg ggtgcacgag tgggttacat 11820
cgaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag aacgttttcc 11880
aatgatgagc acttttaaag ttctgctatg tggcgcggta ttatcccgta ttgacgccgg 11940
gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttggttg agtactcacc 12000
agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca gtgctgccat 12060
aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag gaccgaagga 12120
gctaaccgct tttttgcaca acatgggggga tcatgtaact cgccttgatc gttgggaacc 12180
ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg tagcaatggc 12240
aacaacgttg cgcaaactat taactggcga actacttact ctagcttccc ggcaacaatt 12300
aatagactgg atggaggcgg ataaagttgc aggaccactt ctgcgctcgg cccttccggc 12360
tggctggttt attgctgata aatctggagc cggtgagcgt gggtctcgcg gtatcattgc 12420
agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga cggggagtca 12480
ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac tgattaagca 12540
ttggtaactg tcagaccaag tttactcata tatactttag attgatttaa aacttcattt 12600
ttaatttaaa aggatctagg tgaagatcct ttttgataat ctcatgacca aaatccctta 12660

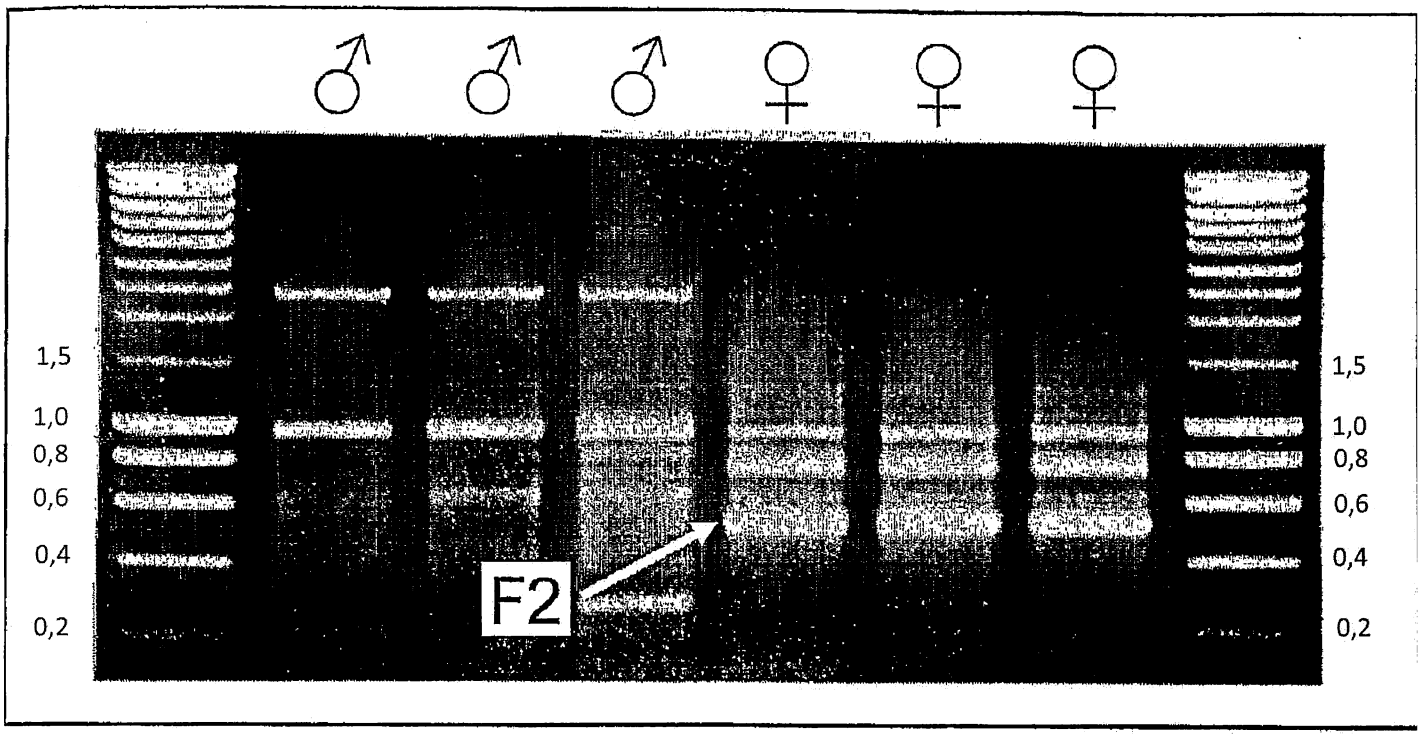
acgtgagttt tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg 12720
agatcctttt tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac cgctaccagc 12780
ggtggtttgt ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag 12840
cagagcgcag ataccaaata ctgtccttct agtgtagccg tagttaggcc accacttcaa 12900
gaactctgta gcaccgccta catacctcgc tctgctaatc ctgttaccag tggctgctgc 12960
cagtggcgat aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggc 13020
gcagcggtcg ggctgaacgg ggggttcgtg cacacagccc agcttggagc gaacgaccta 13080
caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgcttc ccgaagggag 13140
aaaggcggac aggtatccgg taagcggcag ggtcggaaca ggagagcgca cgagggagct 13200
tccagggga aacgcctggt atctttatag tcctgtcggg tttcgccacc tctgacttga 13260
gcgtcgattt ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc 13320
ggcctttta cggttcctgg ccttttgctg gccttttgct cacatgttct ttcctgcgtt 13380
atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata ccgctcgccg 13440
cagccgaacg accgagcgca gcgagtcagt gagcgaggaa gcggaagagc gcccaatacg 13500
caaaccgcct ctccccgcgc gttggccgat tcattaatgc agctggcacg acaggtttcc 13560
cgactggaaa gcgggcagtg agcgcaacgc aattaatgtg agttagctca ctcattaggc 13620
accccaggct ttacacttta tgcttccggc tcgtatgttg tgtggaattg tgagcggata 13680
acaatttcac acaggaaaca gctatgacca tgattacgaa tttcgacctg caggcatgca 13740
agcttgcatg cctgcaggtc gacgctcgcg cgacttggtt tgccattctt tagcgcgcgt 13800
cgcgtcacac agcttggcca caatgtggtt tttgtcaaac gaagattcta tgacgtgttt 13860
aaagtttagg tcgagtaaag cgcaaatctt ttttaaccct agaaagatag tctgcgtaaa 13920
attgacgcat gcattcttga aatattgctc tctctttcta aatagcgcga atccgtcgct 13980
gtgcatttag gacatctcag tcgccgcttg gagctcccgt gaggcgtgct tgtcaatgcg 14040
gtaagtgtca ctgattttga actataacga ccgcgtgagt caaaatgacg catgattatc 14100
ttttacgtga cttttaagat ttaactcata cgataattat attgttattt catgttctac 14160
ttacgtgata acttattata tatatatttt cttgttatag atatcgtgac taatatataa 14220
taaaatgggt agttctttag acgatgagca tatcctctct gctcttctgc aaagcgatga 14280
cgagcttgtt ggtgaggatt ctgacagtga aatatcagat cacgtaagtg aagatgacgt 14340
ccagagcgat acagaagaag cgtttataga tgaggtacat gaagtgcagc caacgtcaag 14400
cggtagtgaa atattagacg aacaaaatgt tattgaacaa ccaggttctt cattggcttc 14460
taacagaatc ttgaccttgc cacagaggac tattagaggt aagaataaac attgttggtc 14520
aacttcaaag tccacgaggc gtagccgagt ctctgcactg aacattgtca gatcggcccg 14580
gcggagtgga cacgctagac caaatgtgtt ctgtgatgac ctgcagtagg aagacgaata 14640
ggtggcctat ggcattattg tacggaatga taaacattgc ctgcataaat tcttttatta 14700
tatacagcca taatgtcagt agcaagggag aaaaggtcca aagtcgcaaa aaatttatga 14760
gaaaccttta catgagcctg acgtcatcgt ttatgcgtaa gcgtttagaa gctcctactt 14820
tgaagagata tttgcgcgat aatatctcta atattttgcc aaatgaagtg cctggtacat 14880
cagatgacag tactgaagag ccagtaatga aaaaacgtac ttactgtact tactgccccc 14940
ctaaaataag gcgaaaggca aatgcatcgt gcaaaaaatg caaaaaagtt atttgtcgag 15000
agcataatat tgatatgtgc caaagttgtt tctgactgac taataagtat aatttgtttc 15060
tattatgtat aagttaagct aattacttat tttataatac aacatgactg tttttaaagt 15120
acaaaataag tttatttttg taaaagagag aatgtttaaa agttttgtta ctttatagaa 15180
gaaattttga gtttttgttt ttttttaata aataaataaa cataaataaa ttgtttgttg 15240
aatttattat tagtatgtaa gtgtaaatat aataaaactt aatatctatt caaattaata 15300
aataaacctc gatatacaga ccgataaaac acatgcgtca attttacgca tgattatctt 15360
taacgtacgt cacaatatga ttatctttct agggttaaaa tgaatgtaag cactttatta 15420
acgaaatctt tgggaatatt tcgctcatca gcattttatt tgagcaggag tccgagatgc 15480
cc 15482


Cắt nối Aedes DSX nội sinh
F1
F2
F3
1
135
633
13995
14662
15130
25524
26586
43482
S4
MF
14453
14662
43815
M1
M2
Nucleotit 1 = 610694bp từ cơ sở dữ liệu the Broad Institut Aedes Database, supercontig 1.370
S4 = Exon 4 được chia
MF = Sản phẩm phiên mã được ch
F1 = Sản phẩm phiên mã giống cái 1
F2 = Sản phẩm phiên mã giống cái 2
F3 = Sản phẩm phiên mã giống cái 3
M1 = Sản phẩm phiên mã giống đực 1
M2 = Sản phẩm phiên mã giống đực 2
- Biểu đồ không theo tỷ lệ
DSX minigene1
S4
1
727
13415
15690
24635
26010
Cắt nối đặc hiệu giới tính F1, F2 và M1
DSX minigene2
S4
-28
1
727
13415
15036
Cắt nối trong dạng F1 và F2 ở cả giống đực và

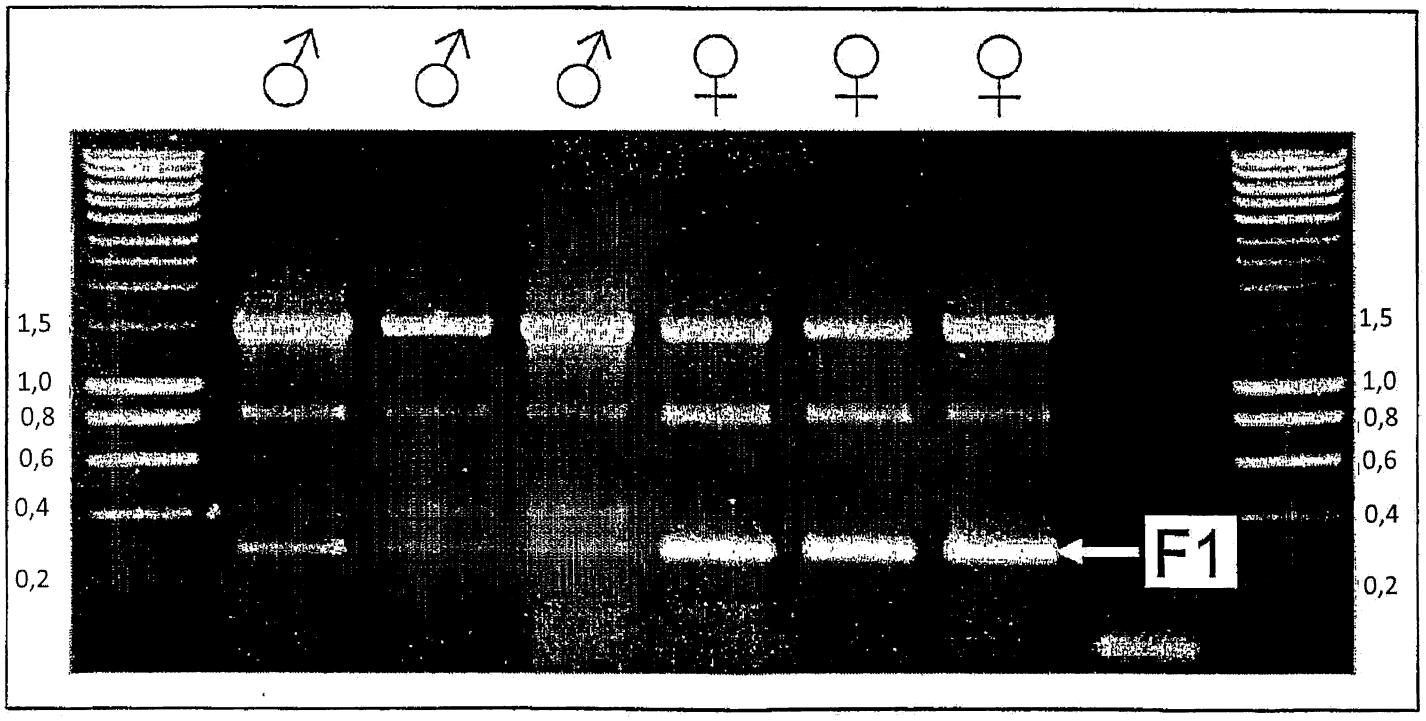
Hình 1


-RT CON
♂
♀
♂
♂
♂
♀
♀
♀
1,5
1,0
0,8
0,6
0,4
0,2
F2
M1
1,5
1,0
0,8
0,6
0,4
0,2

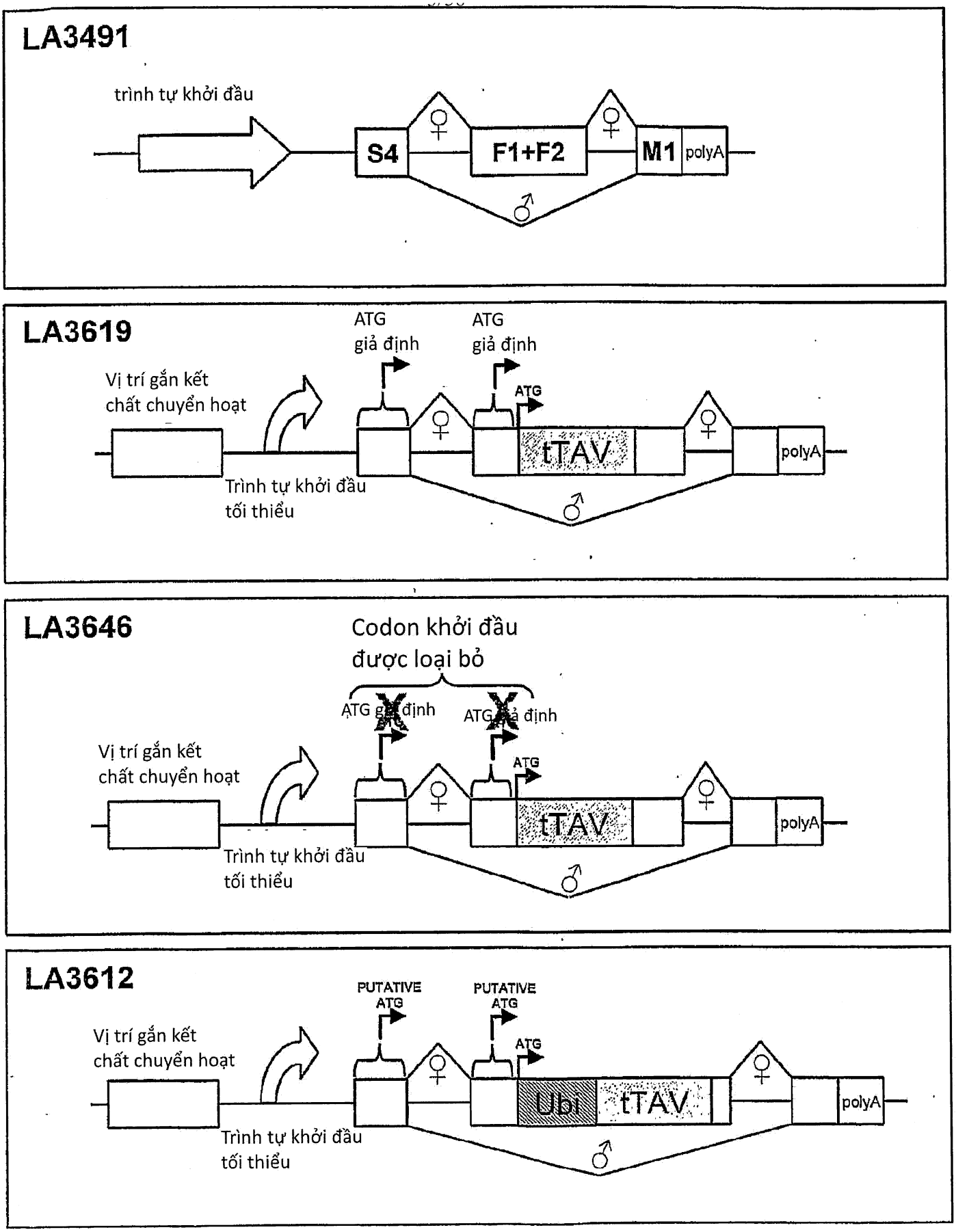
Hình 2


♂
♂
♂
♀
♀
♀
1,5
1,0
0,8
0,6
0,4
0,2
F2
1,5
1,0
0,8
0,6
0,4
0,2

Hình 3


♂
♂
♂
♀
♀
♀
1,5
1,0
0,8
0,6
0,4
0,2
F1
1,5
1,0
0,8
0,6
0,4
0,2

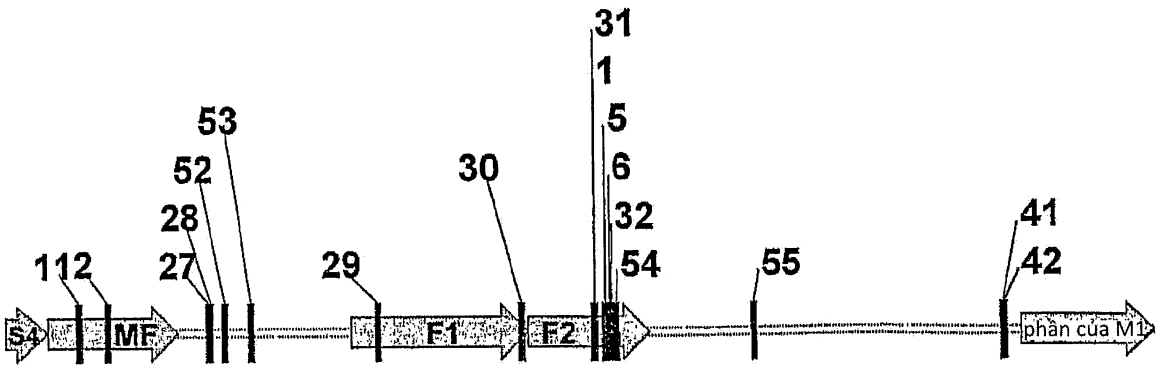
Hình 4


LA3491
trình tự khởi đầu
S4
F1+F2
M1
polyA
LA3619
ATG
giả định
ATG
giả định
Vị trí gắn kết
chất chuyển hoạt
ATG
tTAV
polyA
Trình tự khởi đầu
tối thiểu
LA3646
Codon khởi đầu
được loại bỏ
ATG giả định
ATG giả định
Vị trí gắn kết
chất chuyển hoạt
ATG
tTAV
polyA
Trình tự khởi đầu
tối thiểu
LA3612
PUTATIVE
ATG
PUTATIVE
ATG
Vị trí gắn kết
chất chuyển hoạt
ATG
Ubi
tTAV
polyA
Trình tự khởi đầu
tối thiểu

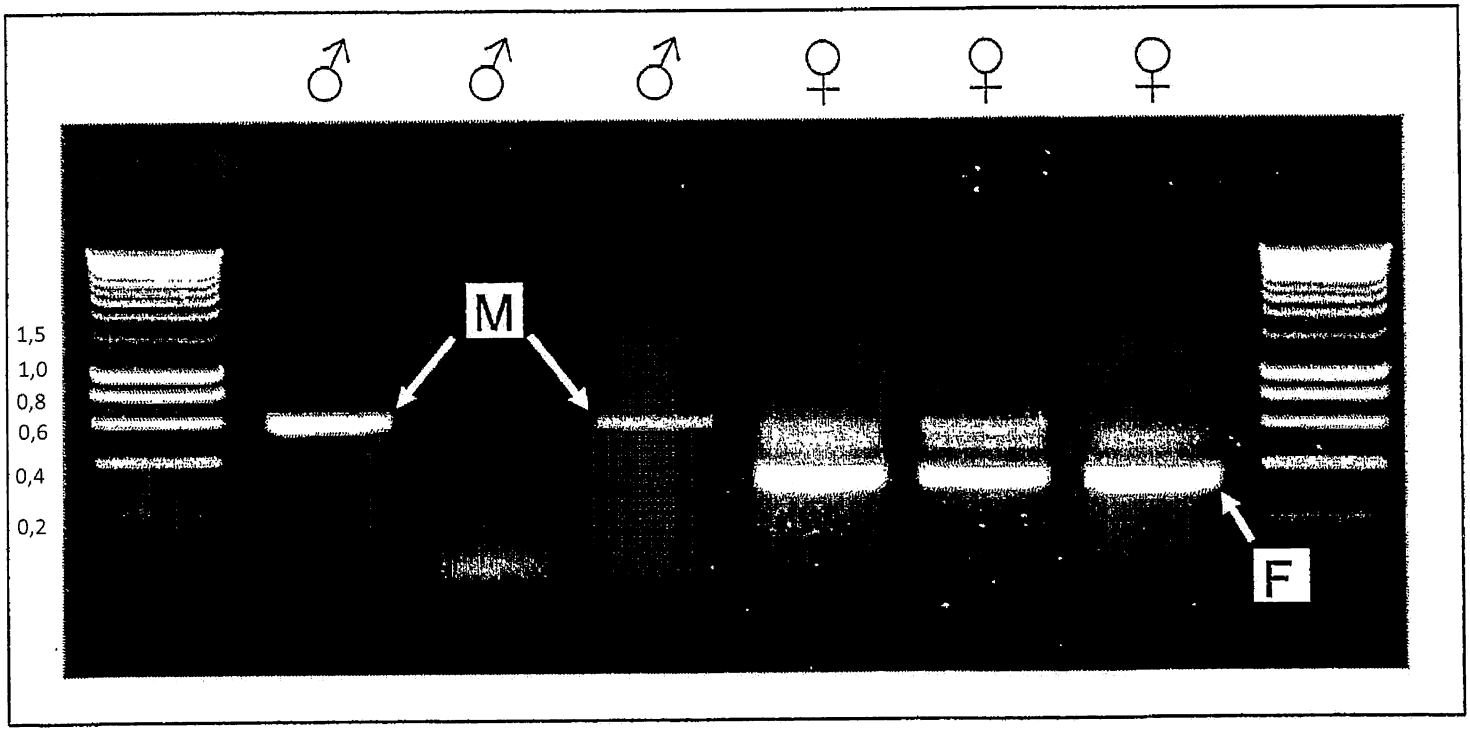
Hình 5

```
-dsx        TGTGCGTTGCTGTT--TGCGATGGAAGGACTAT-TGTGTCGTCGCCACGCTGGACTATTC 4287
byx-dsx     TGTGCGATGCTGTG--CGAATTTCAACGGAAATATTTGTTGTCGTAACATTGGATCTATG 1575
ling-dsx    TGACTGTTCCTGTAAATAAGCTTCGTTGGACAT-TGTGTC-TCAC-ACATCGGATCTCAT 3428
            **    * * ****          *      *    ** * ***  **    **    ***

-dsx        GGTGAGTGG------TAGAATAATA-TTTTATCTA-------------TTTCATCGCGGT 4327
byx-dsx     GGTAAGTT-------TAGTATAATAACTTTACTCT-------------GTTCACATTAGT 1615
ling-dsx    GGTAAGTGCTAGTGCTAGCATYRMAACTTAACTCTCTGAGCGAATTCCTTTGACTCTAAA 3488
            *** ***         *** **    *   ** *                   ** *

-dsx        ACAATTGACTTTTTATTACTACTCACTGCTATGGAGGAATCTCAGGAACAT----CGTAA 4383
byx-dsx     GAAACATACATTTG--TAAAATTTG-TGTTTT--ACTAATGTGAAATTTAT----TTTTG 1666
ling-dsx    GTCACACGRACAGCCATACAATCAA-AGCTACGCTCTAATTTTAAGATGACAWTCTGTAA 3547
               *              **   *      * *        *** * *     *       *
```

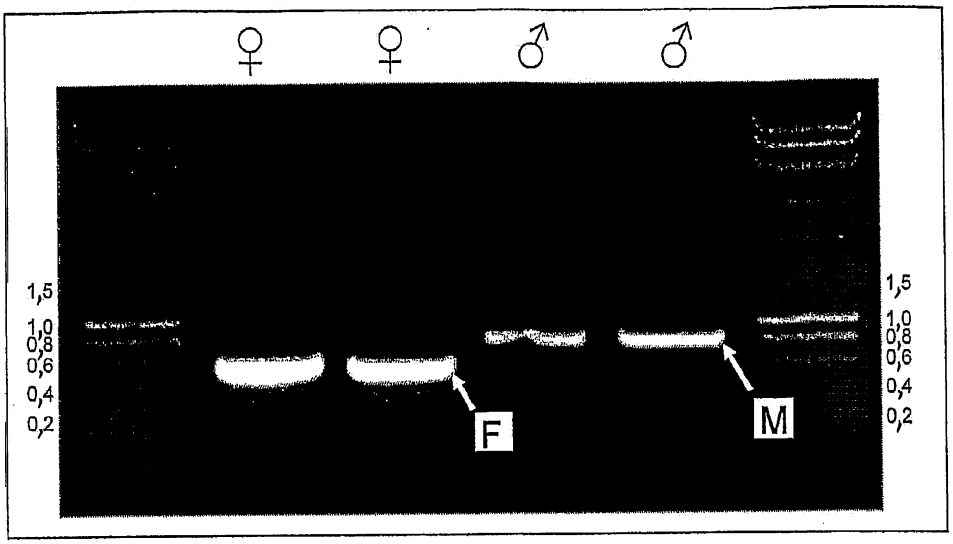
exon đặc hiệu giống cái thứ hai được đánh dấu bởi nucleotit in đậm. Các đoạn lặp bảo toàn AGTGAC/T được gạch chân


31
1
5
6
32
54
53
52
28
27
30
29
112
55
41
42
S4
MF
F1
F2
phần của M1

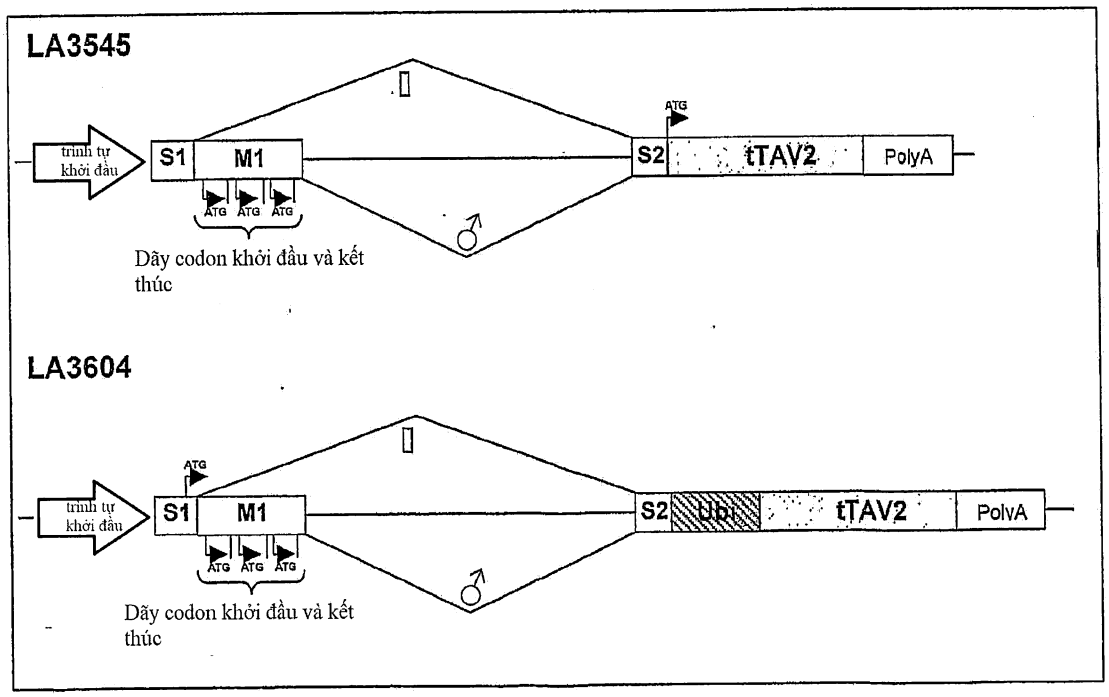
Hình 7


♂
♂
♂
♀
♀
♀
M
F
1,5
1,0
0,8
0,6
0,4
0,2

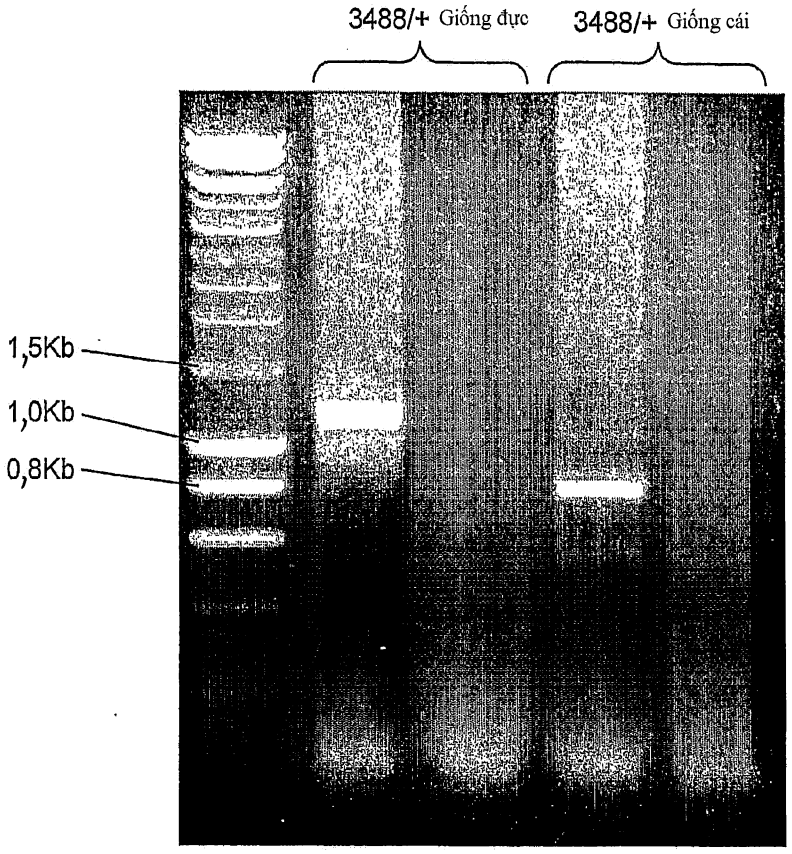
Hình 8



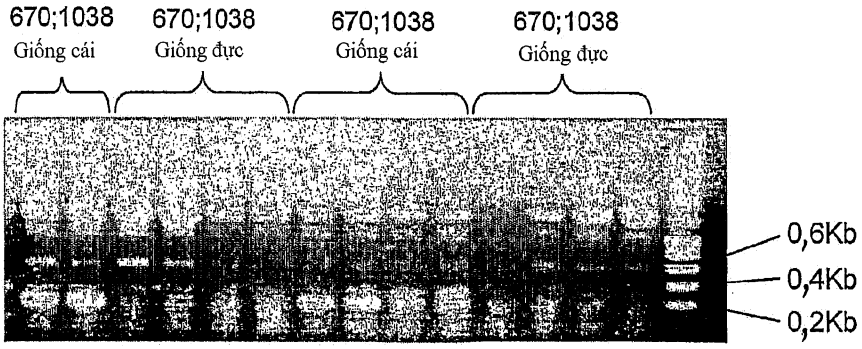
Hình 9



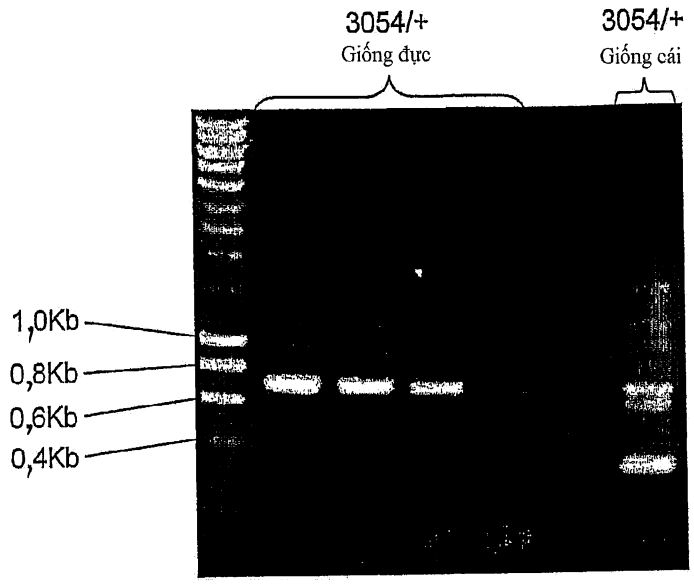
Hình 10



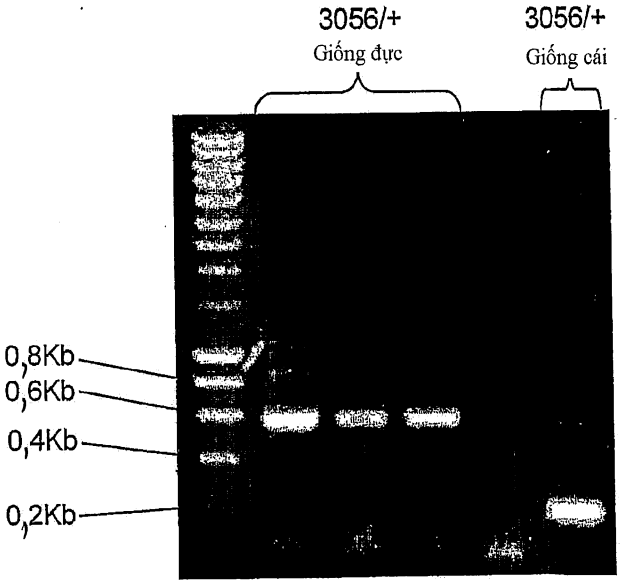
Hình 11



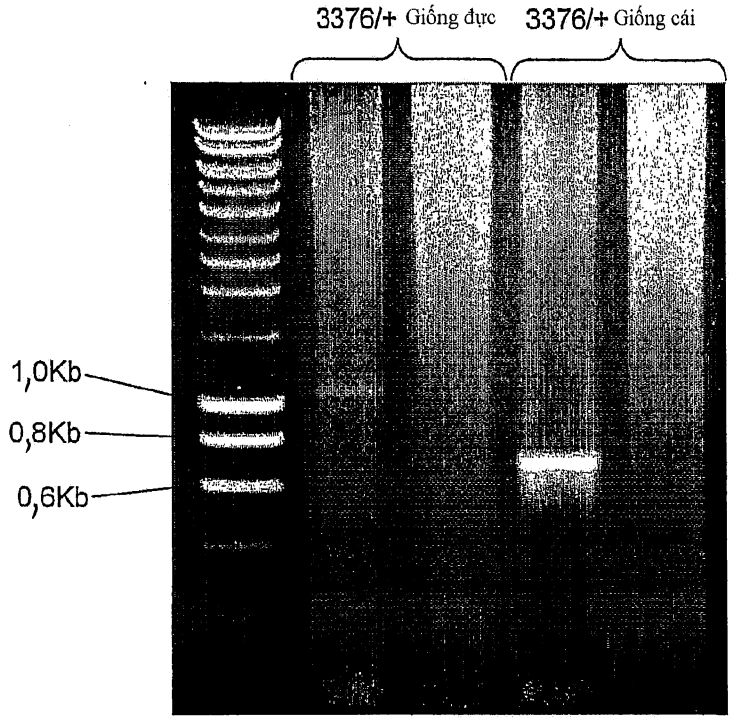
Hình 12



Hình 13



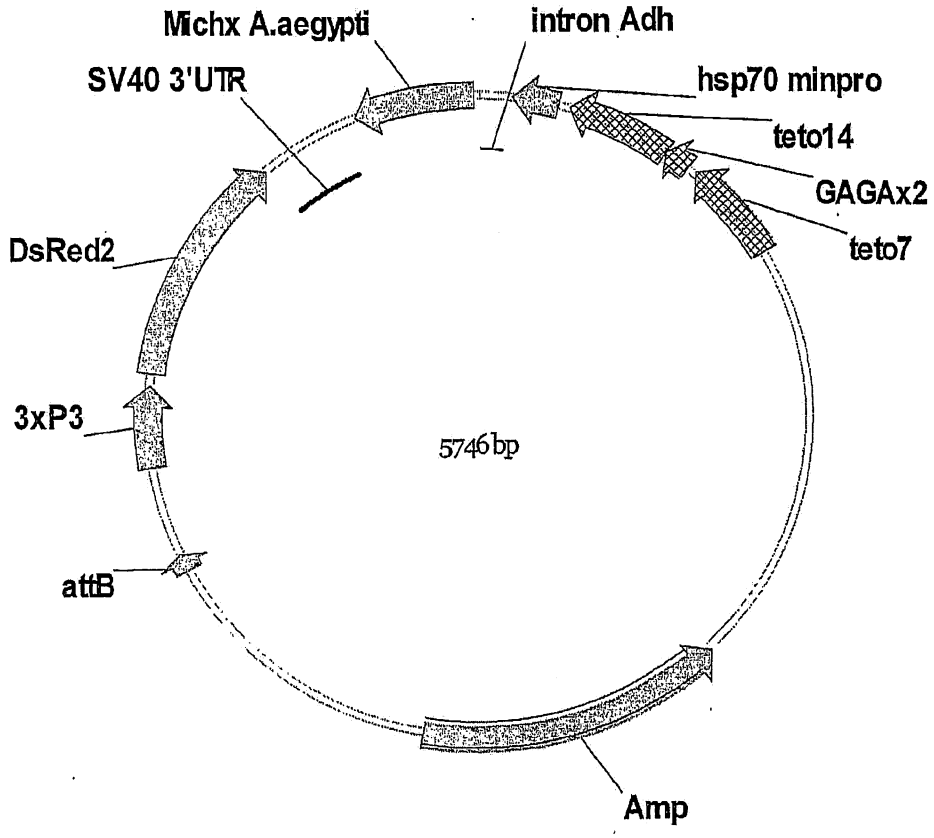
Hình 14



Hình 15

Hình 16

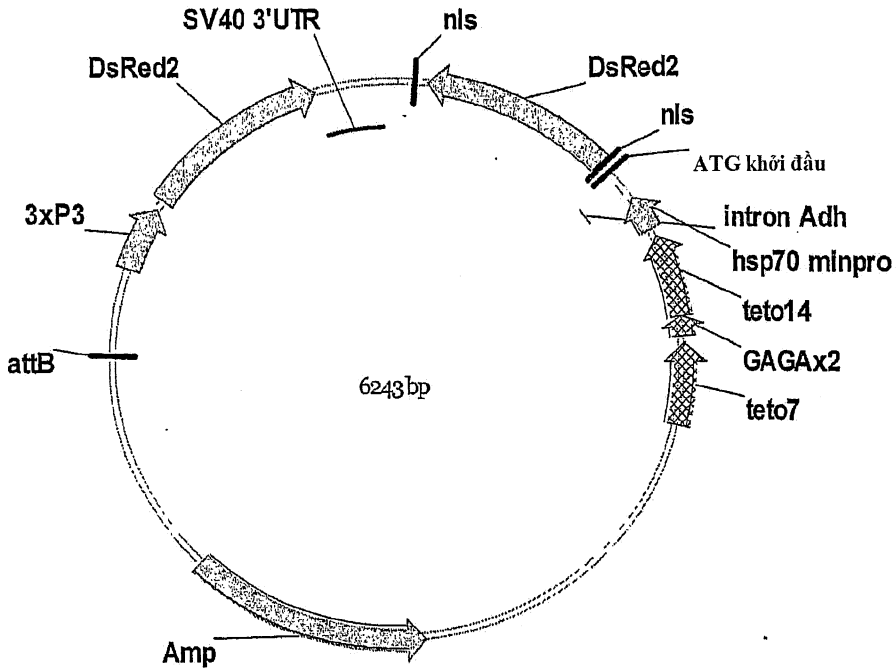
plasmid LA3582



#3581,2 AttB-3xP3DsRed2-teto21-hsp-adh-michxc

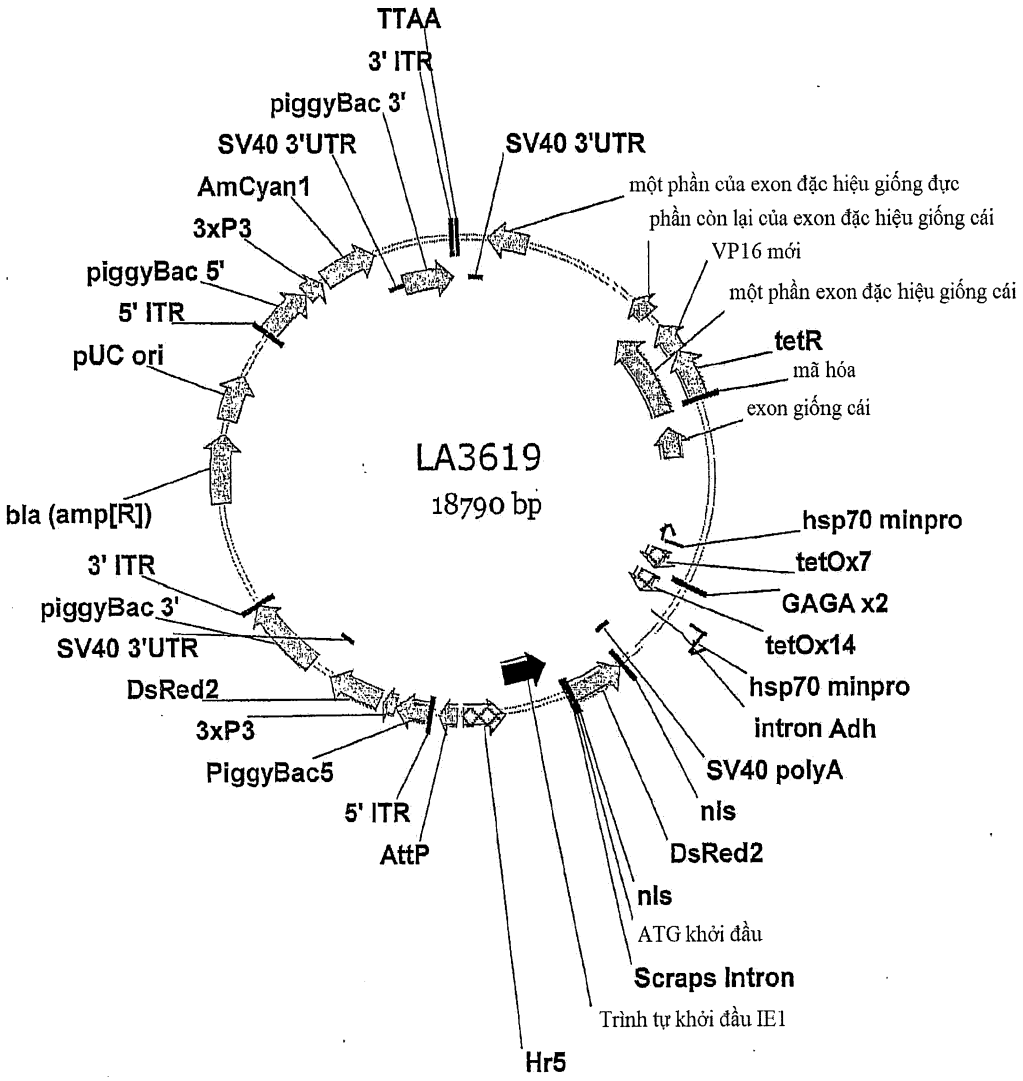
Hình 17

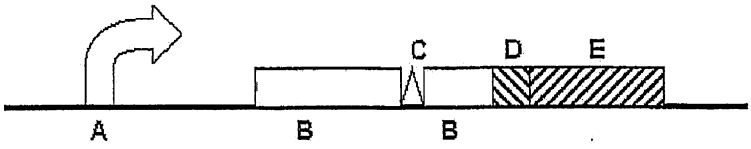
plasmid LA3576



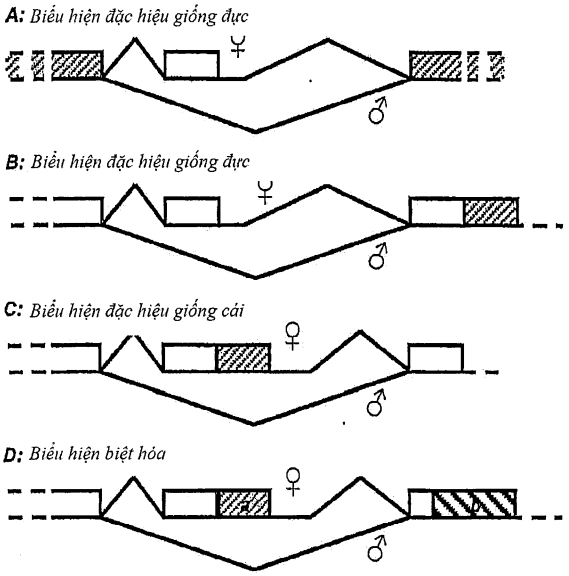
#3575,6 AttB-3xP3DsRed2-teto21-hsp-adh-dsred

Hình 18 - Sơ đồ plasmit LA3619

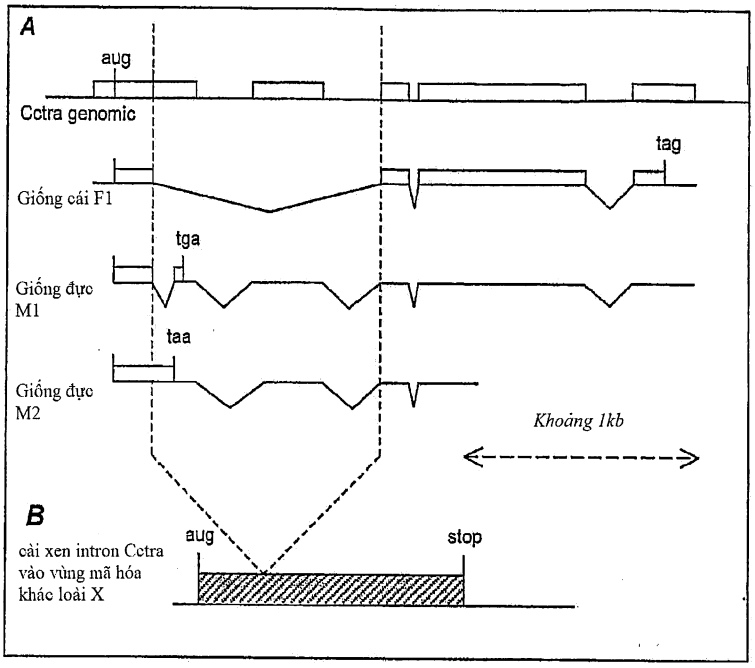




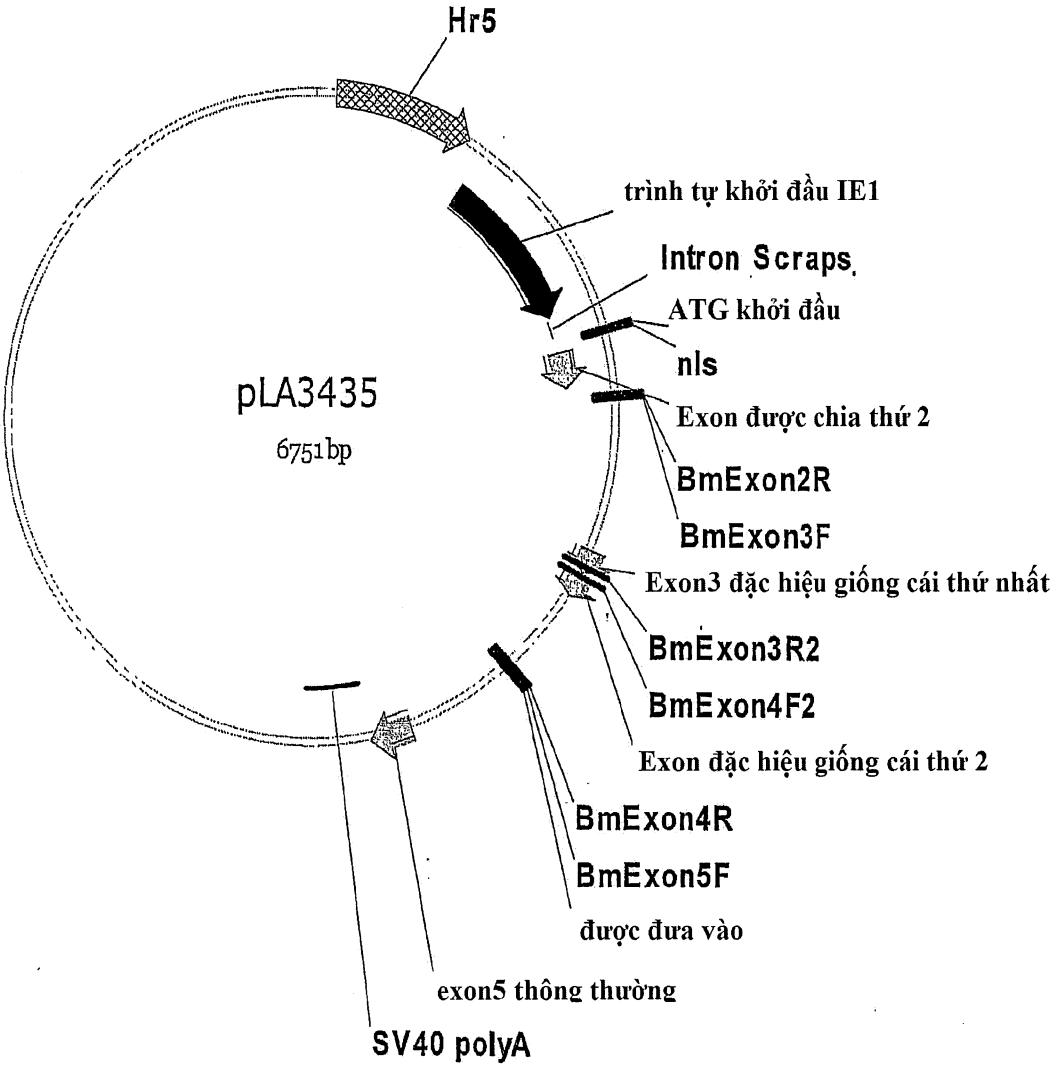
Hình 19


A: Biểu hiện đặc hiệu giống đực
♀
♂
B: Biểu hiện đặc hiệu giống đực
♀
♂
C: Biểu hiện đặc hiệu giống cái
♀
♂
D: Biểu hiện biệt hóa
♀
♂

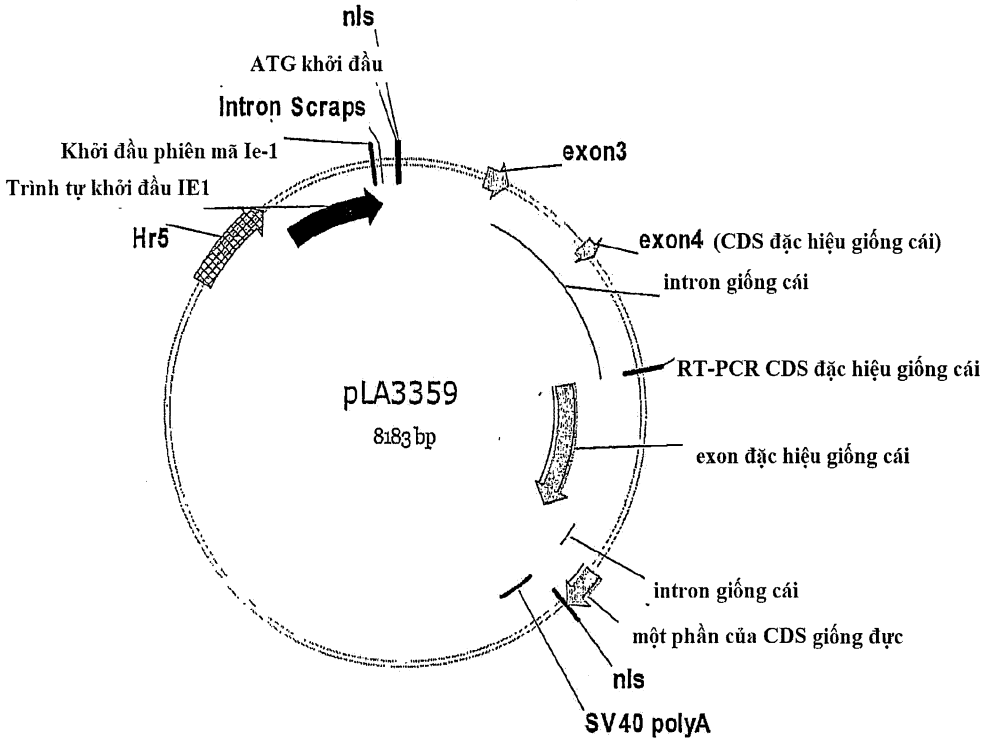
Hình 20


A
aug
Cctra genomic
tag
Giống cái F1
tga
Giống đực
M1
taa
Giống đực
M2
Khoảng 1kb
B
cài xen intron Cctra
vào vùng mã hóa
khác loài X
aug
stop

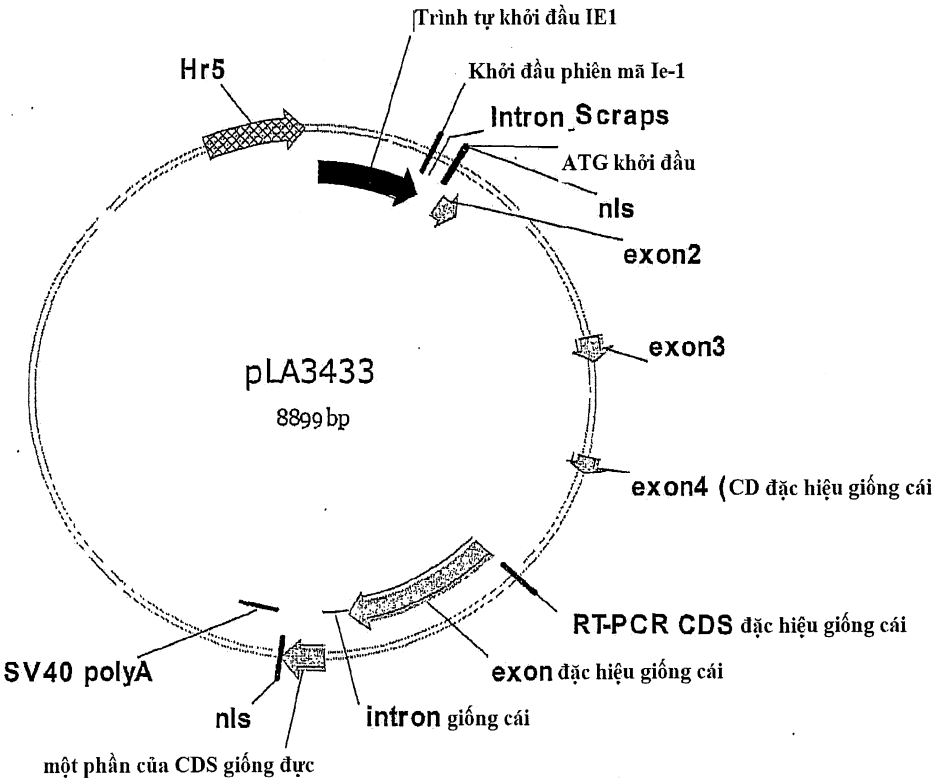
Hình 21


Hr5
trình tự khởi đầu IE1
Intron Scraps
ATG khởi đầu
nls
pLA3435
6751bp
Exon được chia thứ 2
BmExon2R
BmExon3F
Exon3 đặc hiệu giống cái thứ nhất
BmExon3R2
BmExon4F2
Exon đặc hiệu giống cái thứ 2
BmExon4R
BmExon5F
được đưa vào
exon5 thông thường
SV40 polyA

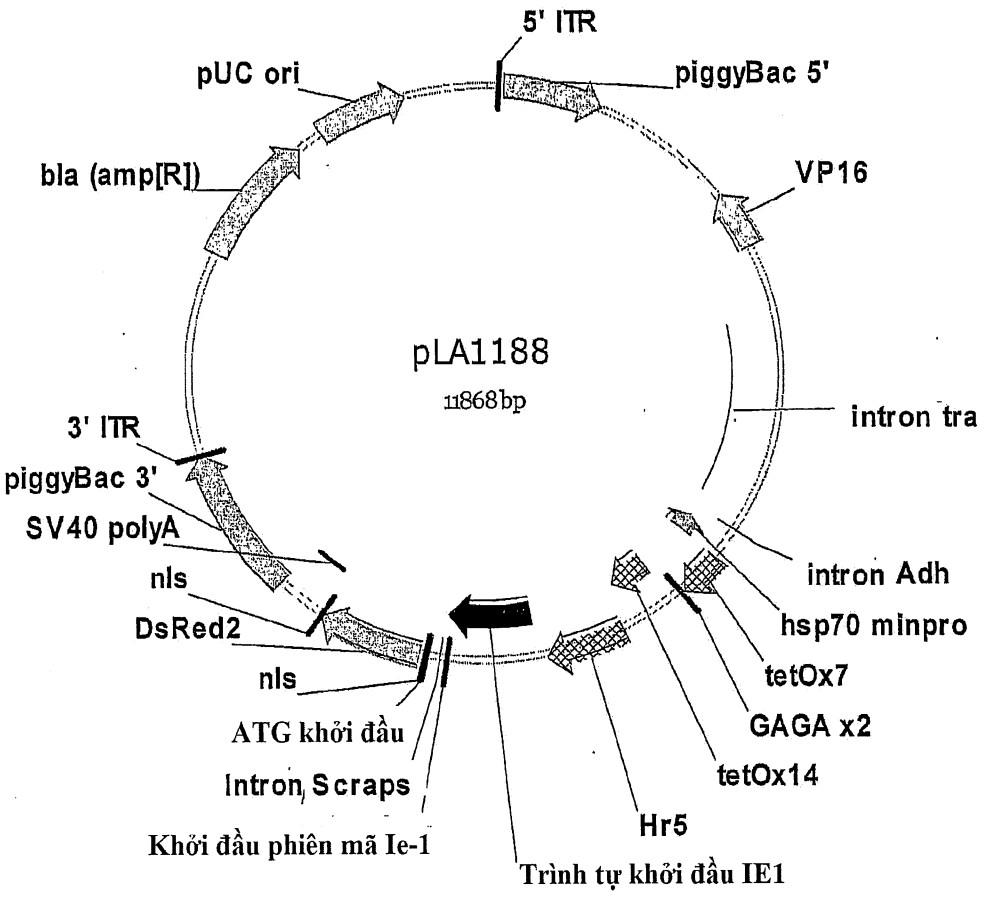
Hình 22


nls
ATG khởi đầu
Intron Scraps
Khởi đầu phiên mã Ie-1
Trình tự khởi đầu IE1
Hr5
exon3
exon4 (CDS đặc hiệu giống cái)
intron giống cái
RT-PCR CDS đặc hiệu giống cái
pLA3359
8183 bp
exon đặc hiệu giống cái
intron giống cái
một phần của CDS giống đực
nls
SV40 polyA

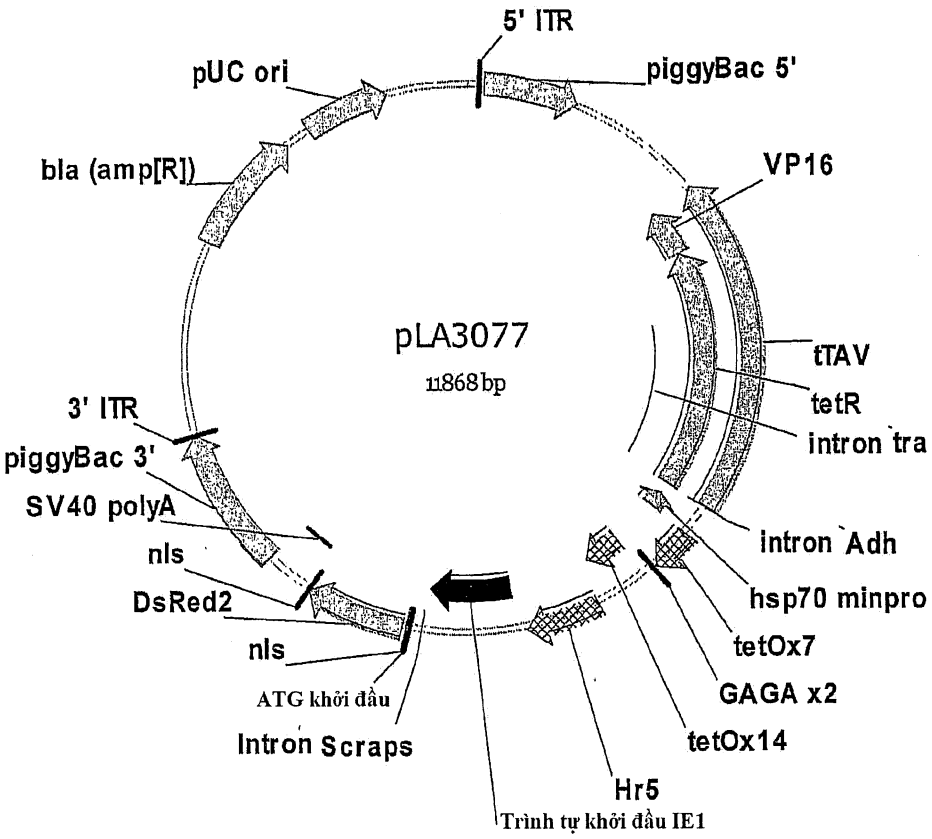
Hình 23


Trình tự khởi đầu IE1
Hr5
Khởi đầu phiên mã Ie-1
Intron Scraps
ATG khởi đầu
nls
exon2
exon3
pLA3433
8899 bp
exon4 (CD đặc hiệu giống cái
RT-PCR CDS đặc hiệu giống cái
exon đặc hiệu giống cái
SV40 polyA
nls
intron giống cái
một phần của CDS giống đực

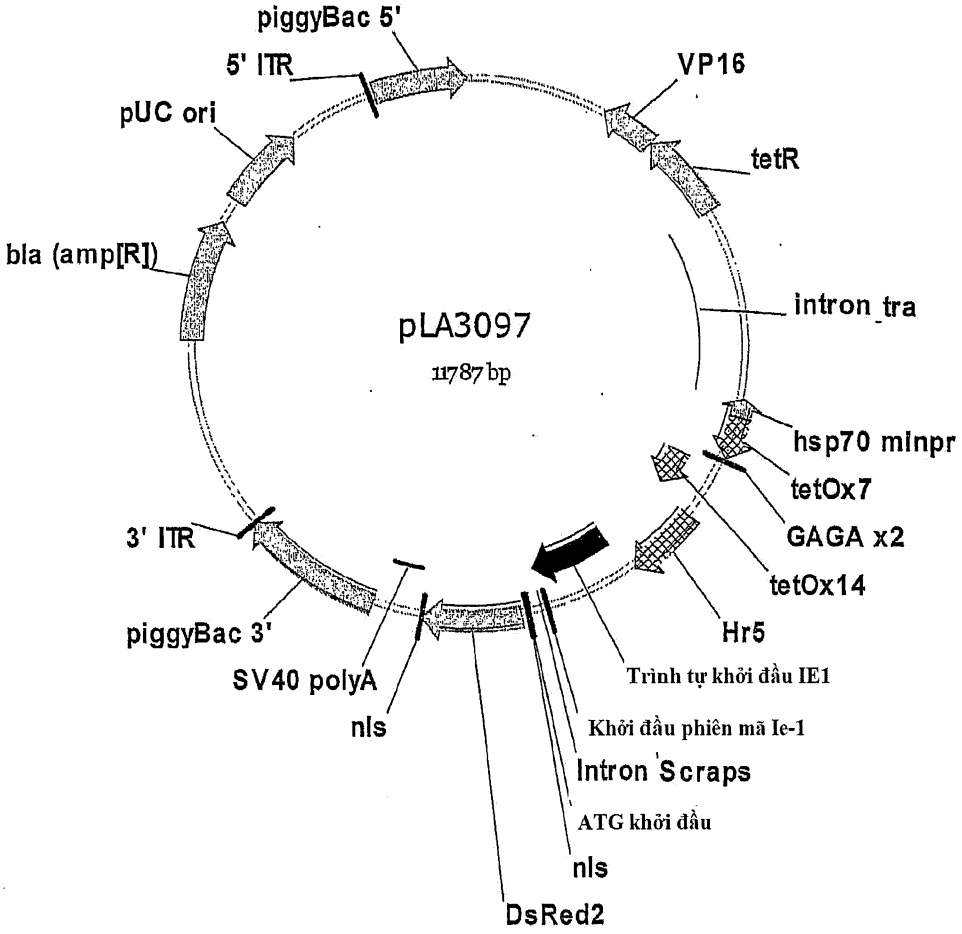
Hình 24


5' ITR
pUC ori
piggyBac 5'
bla (amp[R])
VP16
pLA1188
11868bp
intron tra
3' ITR
piggyBac 3'
SV40 polyA
intron Adh
nls
DsRed2
hsp70 minpro
tetOx7
nls
GAGA x2
ATG khởi đầu
tetOx14
Intron, Scraps
Hr5
Khởi đầu phiên mã Ie-1
Trình tự khởi đầu IE1

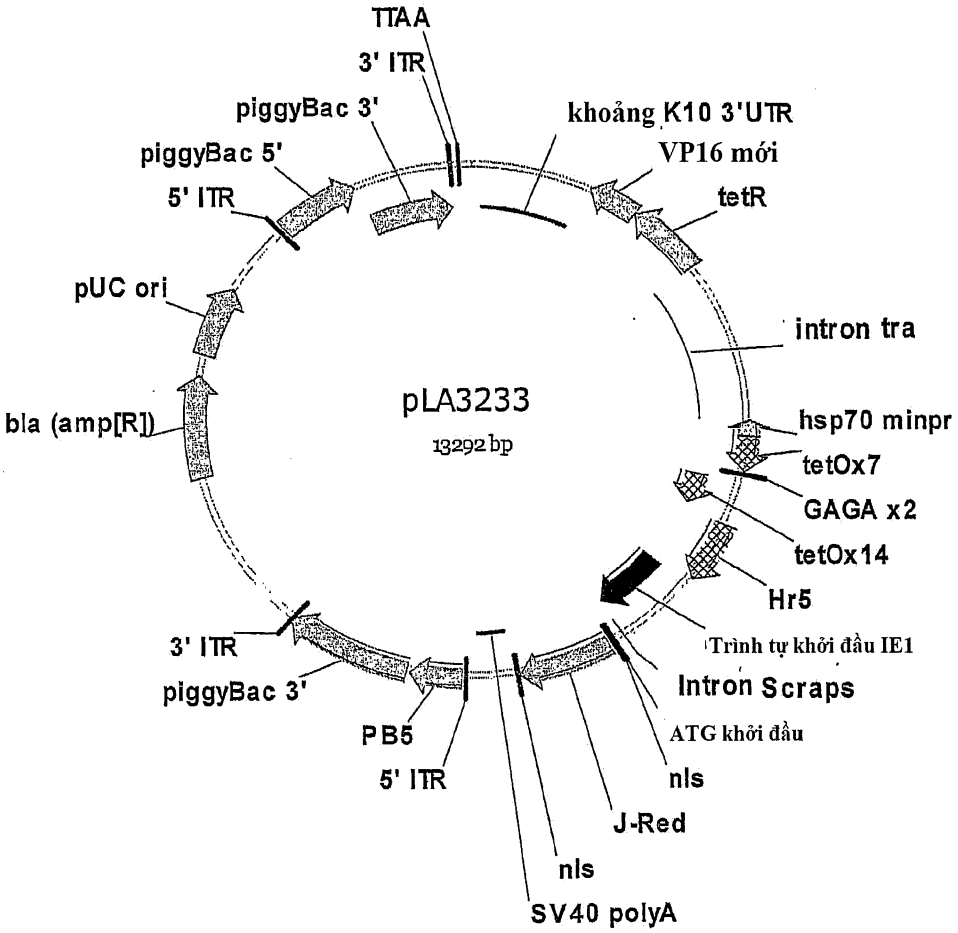
Hình 25


5' ITR
pUC ori
piggyBac 5'
bla (amp[R])
VP16
pLA3077
11868bp
tTAV
tetR
intron tra
3' ITR
piggyBac 3'
SV40 polyA
intron Adh
nls
hsp70 minpro
DsRed2
tetOx7
nls
GAGA x2
ATG khởi đầu
tetOx14
Intron Scraps
Hr5
Trình tự khởi đầu IE1

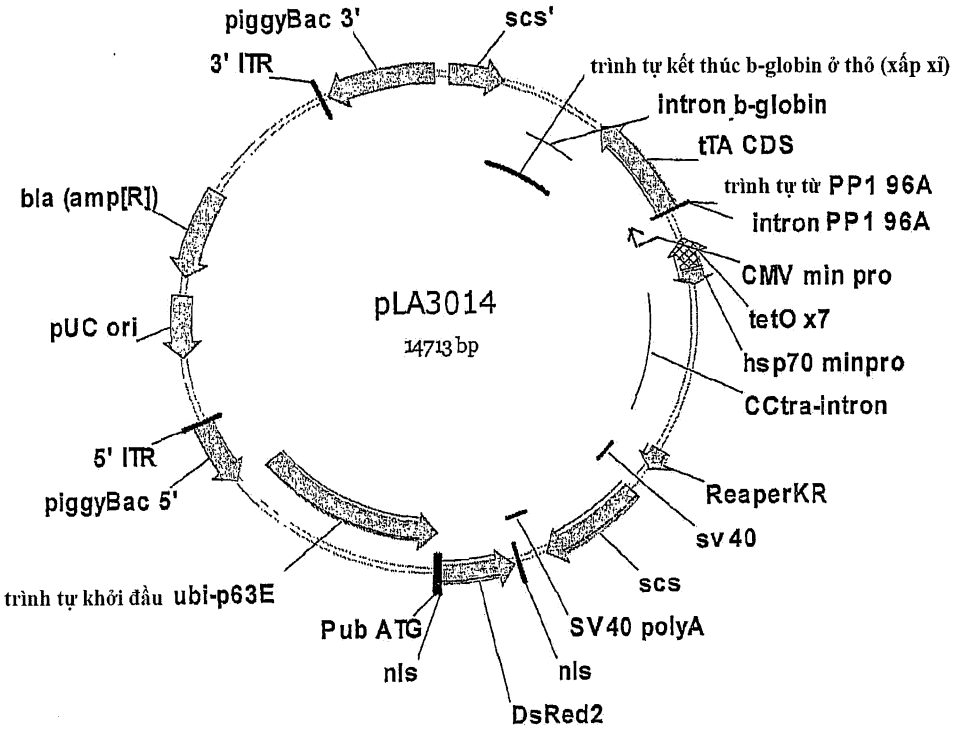
Hình 26


piggyBac 5'
5' ITR
pUC ori
bla (amp[R])
3' ITR
piggyBac 3'
SV40 polyA
nls
DsRed2
nls
ATG khởi đầu
Intron Scraps
Khởi đầu phiên mã Ie-1
Trình tự khởi đầu IE1
Hr5
tetOx14
GAGA x2
tetOx7
hsp70 minpr
intron_tra
pLA3097
11787 bp
tetR
VP16

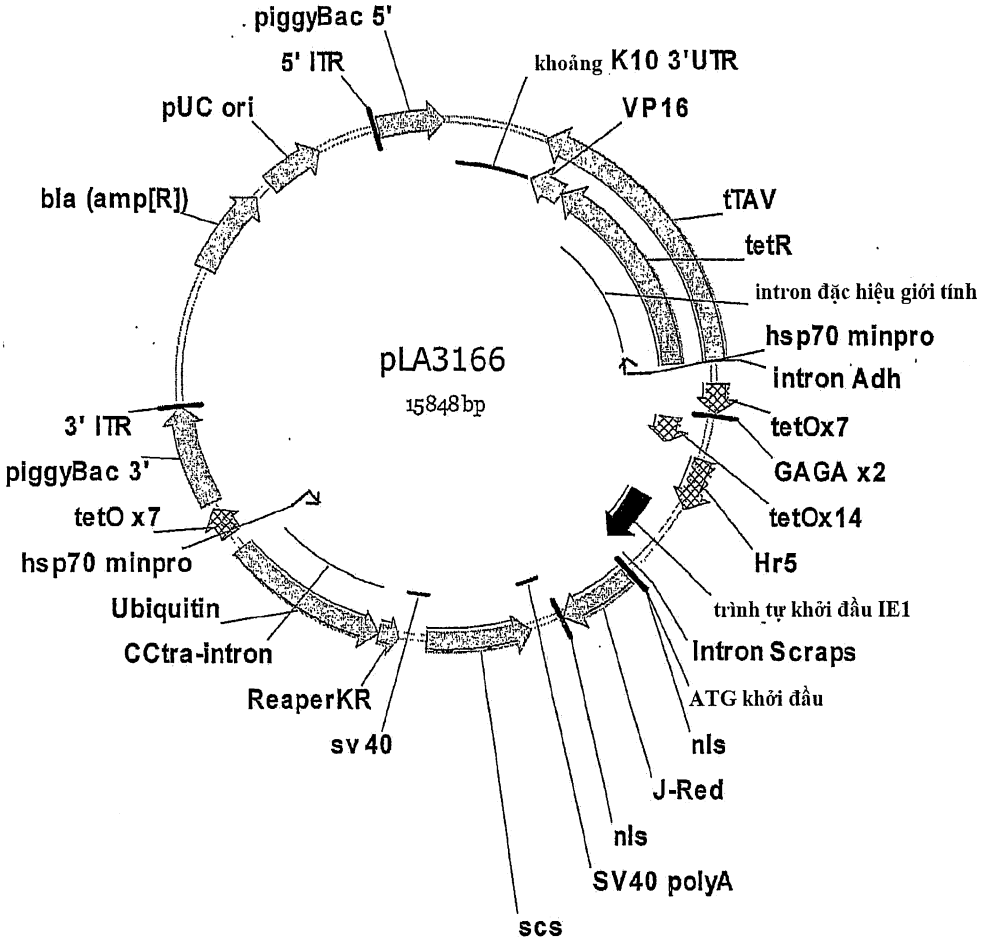
Hình 27


TTAA
3' ITR
piggyBac 3'
piggyBac 5'
5' ITR
pUC ori
bla (amp[R])
3' ITR
piggyBac 3'
PB5
5' ITR
nls
SV40 polyA
nls
J-Red
ATG khởi đầu
Intron Scraps
Trình tự khởi đầu IE1
Hr5
tetOx14
GAGA x2
tetOx7
hsp70 minpr
intron tra
tetR
VP16 mới
khoảng K10 3'UTR
pLA3233
13292 bp

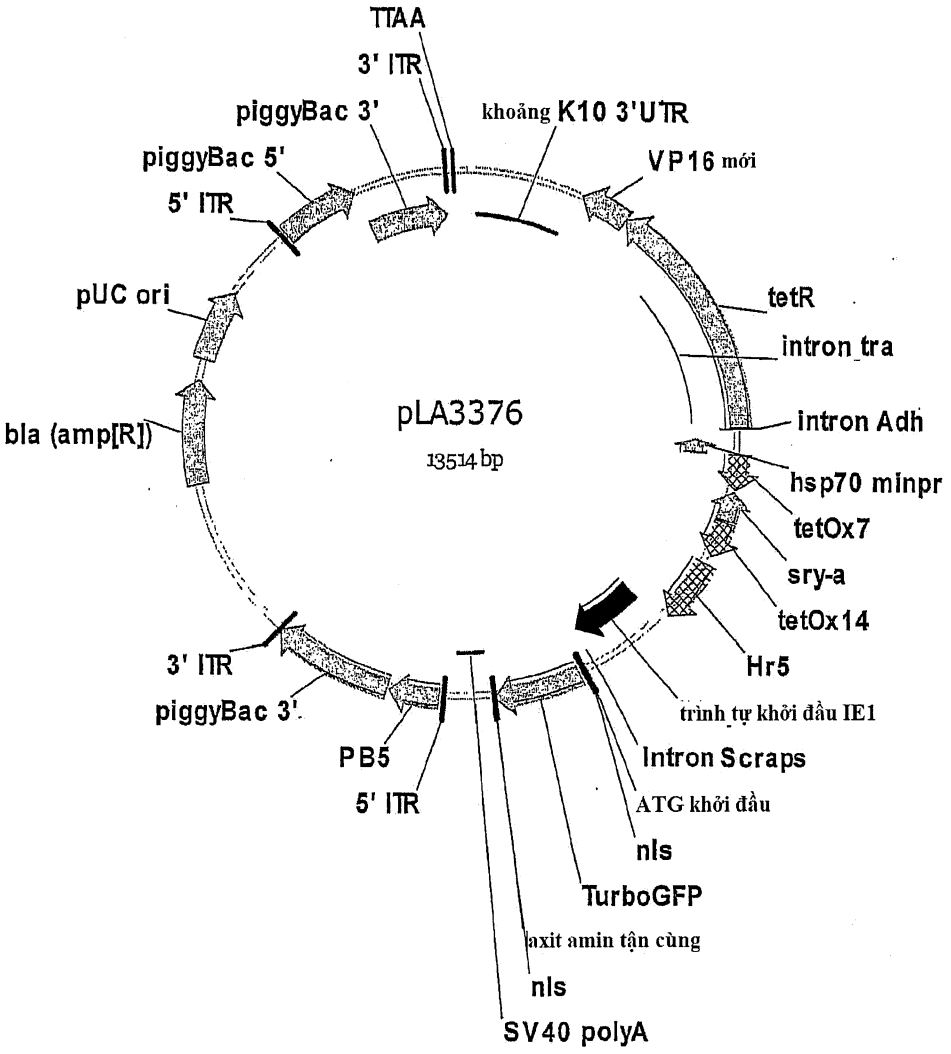
Hình 28


piggyBac 3'
scs'
3' ITR
trình tự kết thúc b-globin ở thỏ (xấp xỉ)
intron b-globin
tTA CDS
trình tự từ PP1 96A
intron PP1 96A
CMV min pro
tetO x7
hsp70 minpro
CCtra-intron
bla (amp[R])
pUC ori
pLA3014
14713 bp
5' ITR
piggyBac 5'
ReaperKR
sv40
scs
trình tự khởi đầu ubi-p63E
Pub ATG
SV40 polyA
nls
nls
DsRed2

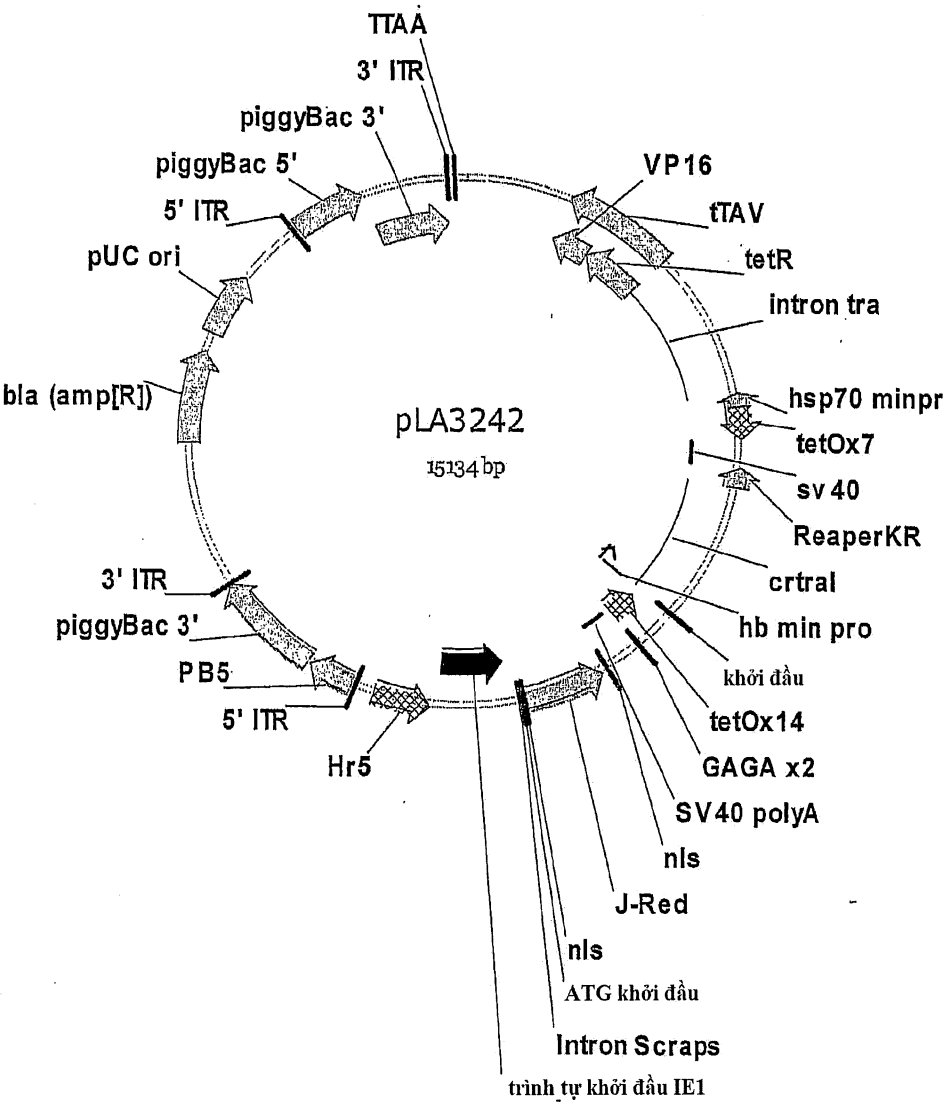
Hình 39


piggyBac 5'
5' ITR
khoảng
K10 3'UTR
VP16
pUC ori
bla (amp[R])
tTAV
tetR
intron đặc hiệu giới tính
hsp70 minpro
pLA3166
15848bp
intron Adh
3' ITR
tetOx7
piggyBac 3'
GAGA x2
tetO x7
tetOx14
hsp70 minpro
Hr5
Ubiquitin
trình tự khởi đầu IE1
CCtra-intron
Intron Scraps
ReaperKR
ATG khởi đầu
sv40
nls
J-Red
nls
SV40 polyA
scs

Hình 30


TTAA
3' ITR
piggyBac 3'
khoảng
K10 3'UTR
VP16 mới
piggyBac 5'
5' ITR
pUC ori
tetR
intron tra
pLA3376
13514 bp
intron Adh
bla (amp[R])
hsp70 minpr
tetOx7
sry-a
tetOx14
Hr5
3' ITR
piggyBac 3'
trình tự khởi đầu IE1
Intron Scraps
PB5
5' ITR
ATG khởi đầu
nls
TurboGFP
axit amin tận cùng
nls
SV40 polyA

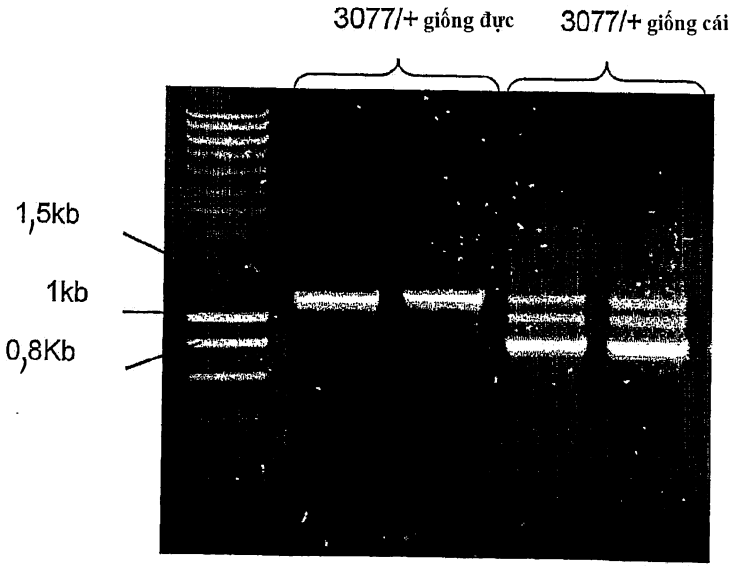
Hình 31


TTAA
3' ITR
piggyBac 3'
piggyBac 5'
5' ITR
pUC ori
bla (amp[R])
3' ITR
piggyBac 3'
PB5
5' ITR
Hr5
VP16
tTAV
tetR
intron tra
hsp70 minpr
tetOx7
sv 40
ReaperKR
crtral
hb min pro
khởi đầu
tetOx14
GAGA x2
SV40 polyA
nls
J-Red
nls
ATG khởi đầu
Intron Scraps
trình tự khởi đầu IE1
pLA3242
15134 bp

Hình 32

```
Native: CGTAGATTTG|GT…intron…AG|GTGAAGGCTC
LA1188: CTACTG|GCACGT…intron…AG|GTGAAGAATA
LA3077: AACGAAGTTG|GT…intron…AG|GTATTGAGGG
LA3097: AGCCACCATG|GT…intron…AG|GTCAGCCGCC
```

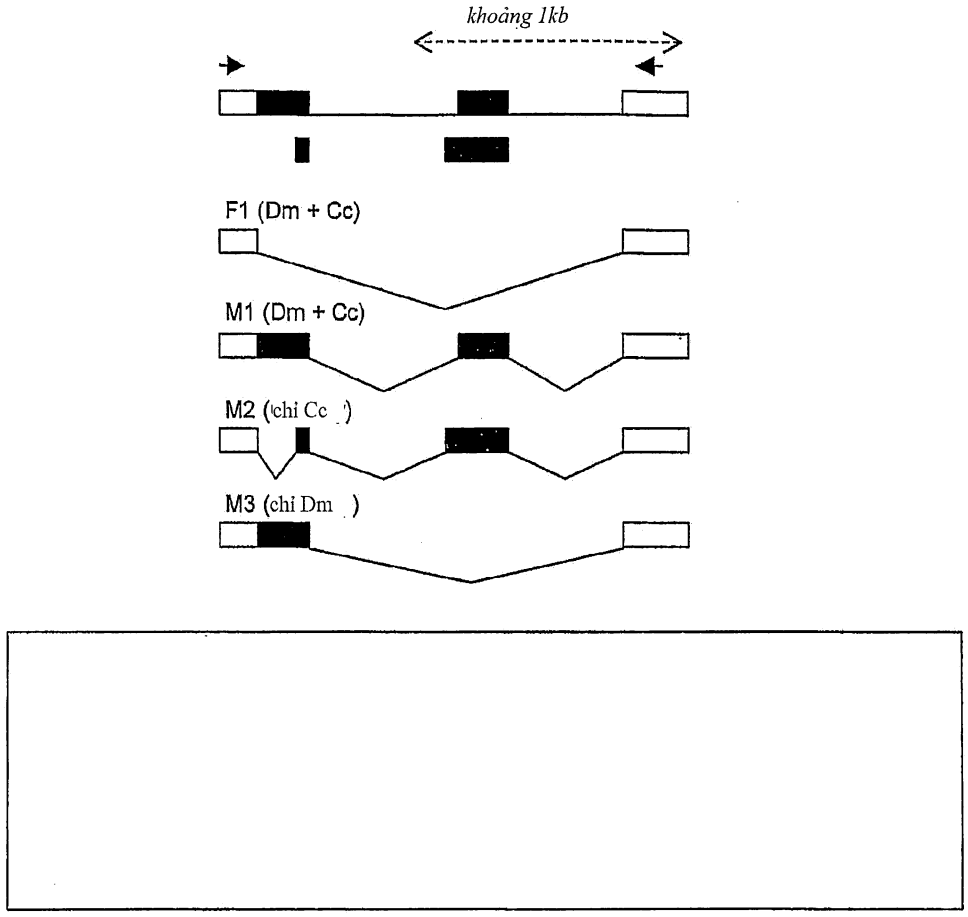
Hình 33



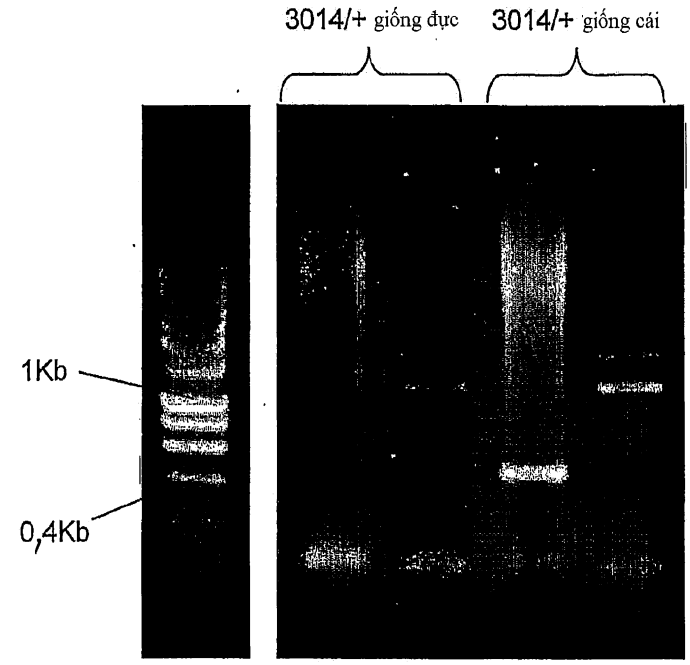
Hình 34

3077A	111	32	73	44
3077B	314	157	132	121
3077C	161	116	60	84
3077D	445	85	194	190
3097A	179	5	89	90
3097B	440	0	59	27
3097C	172	0	46	44
3233A	457	1	79	58
3233B	171	0	14	13
3014;1217	136	0	48	10
3166;1217	64	0	5	7

Hình 35



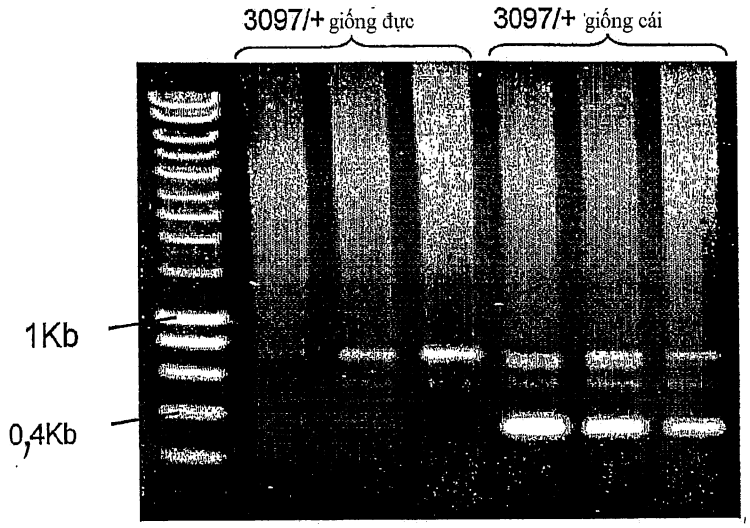
Hình 36



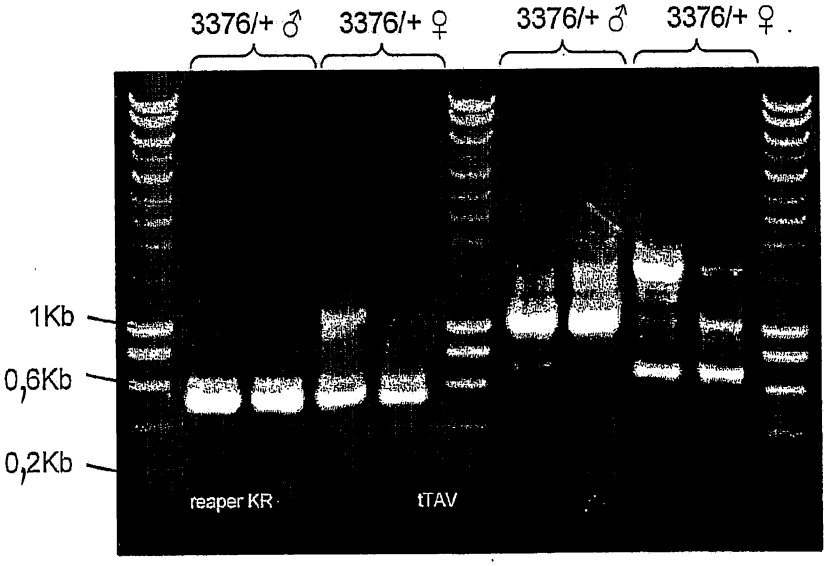
Hình 37

	NT giống đực	NT giống cái	TET giống đực	TET giống cái
3097A	136	0	21	19
3097B	295	11	14	11
3097C	96	12	22	21
3097D	103	15	82	67
3233A	78	6	32	5

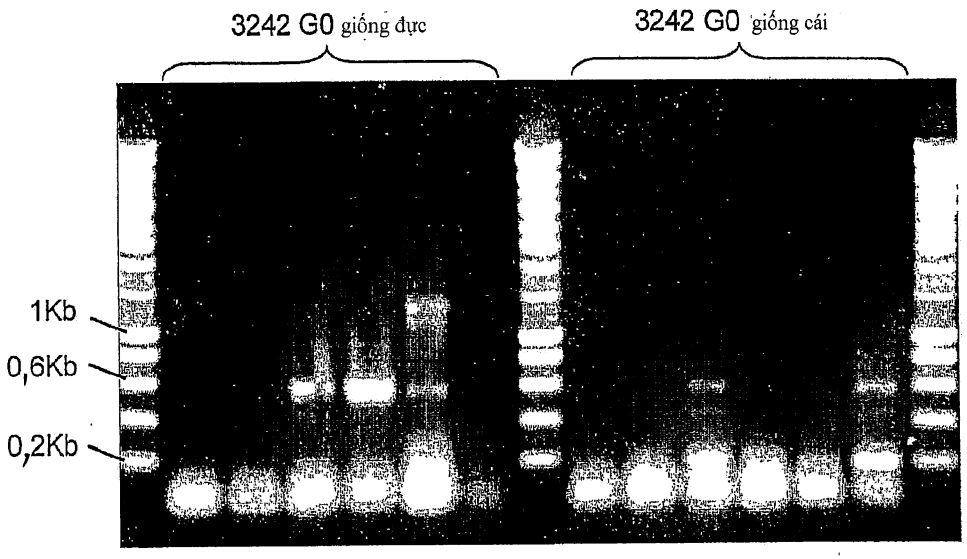
Hình 38



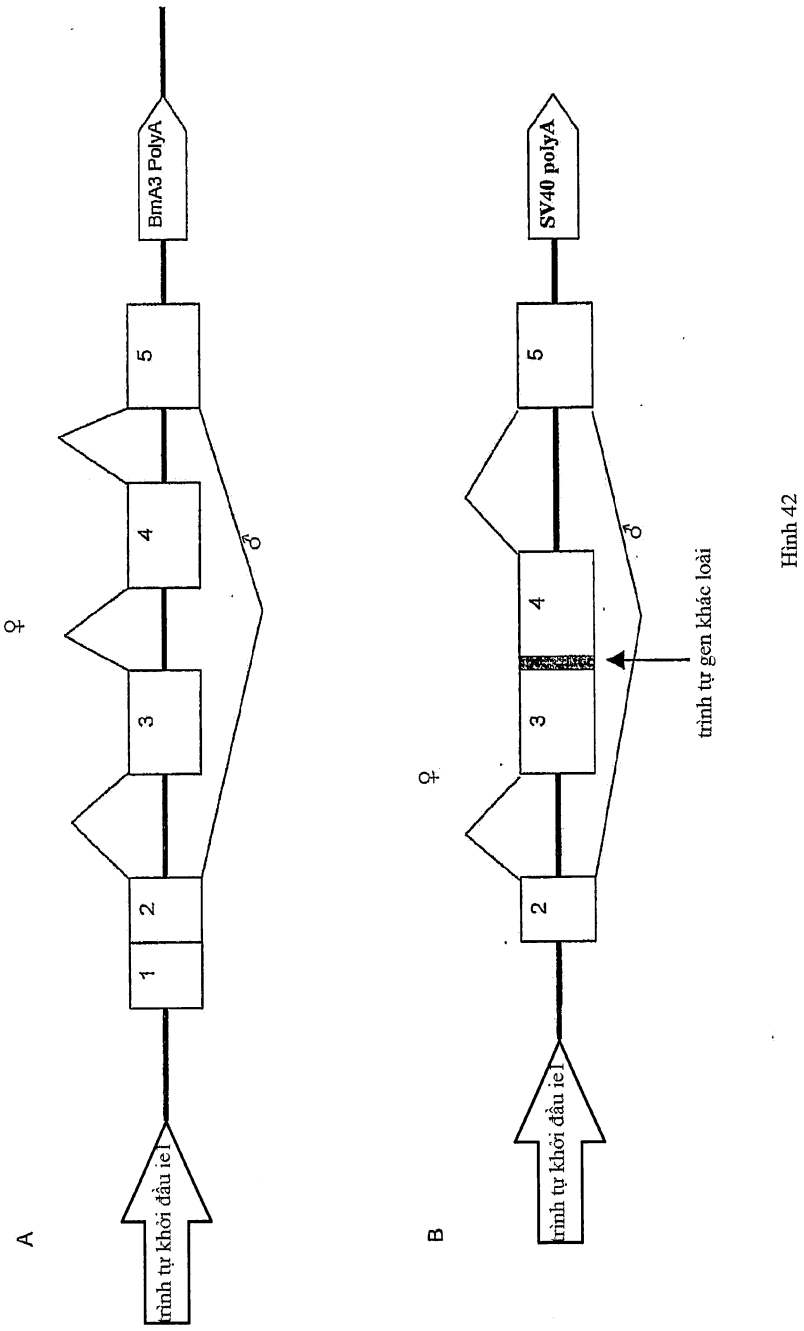
Hình 39



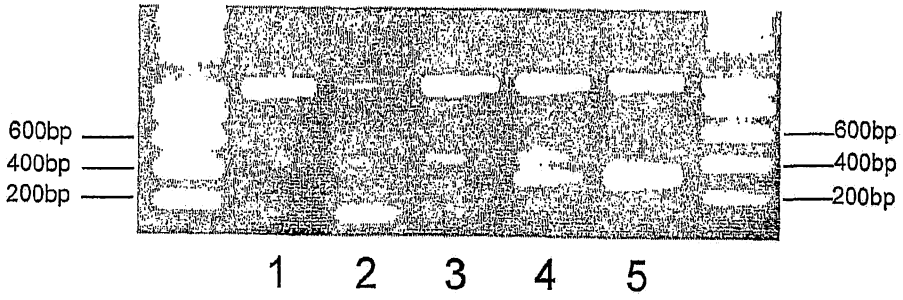
Hình 40



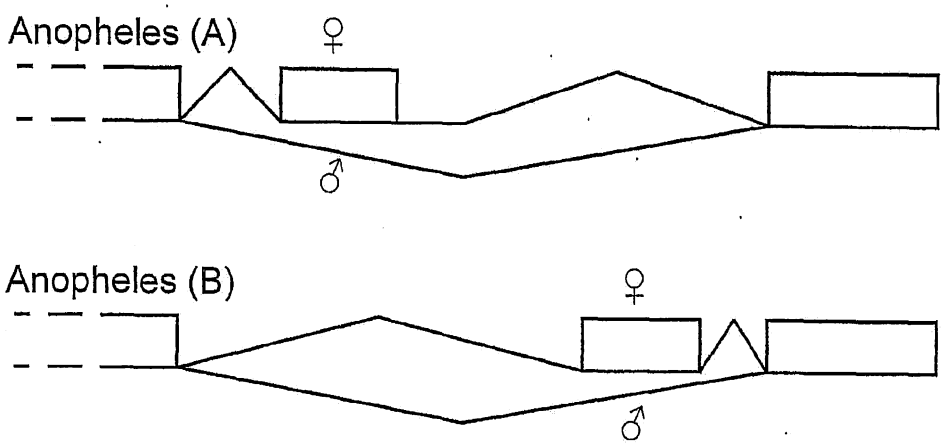
Hình 41


A
trình tự khởi đầu ie1
1
2
3
4
5
♀
♂
BmA3 PolyA
B
trình tự khởi đầu ie1
2
3
4
5
♀
♂
trình tự gen khác loài
SV40 polyA

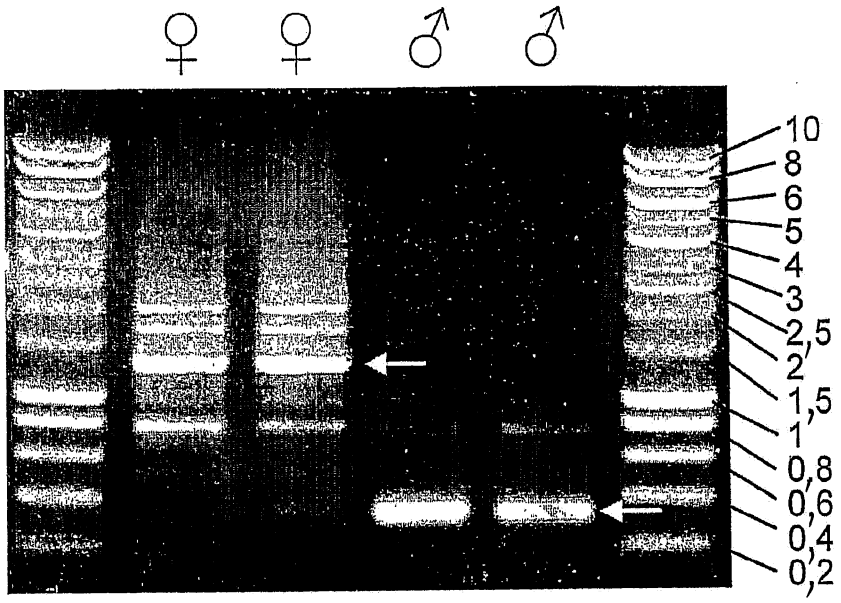
Hình 42



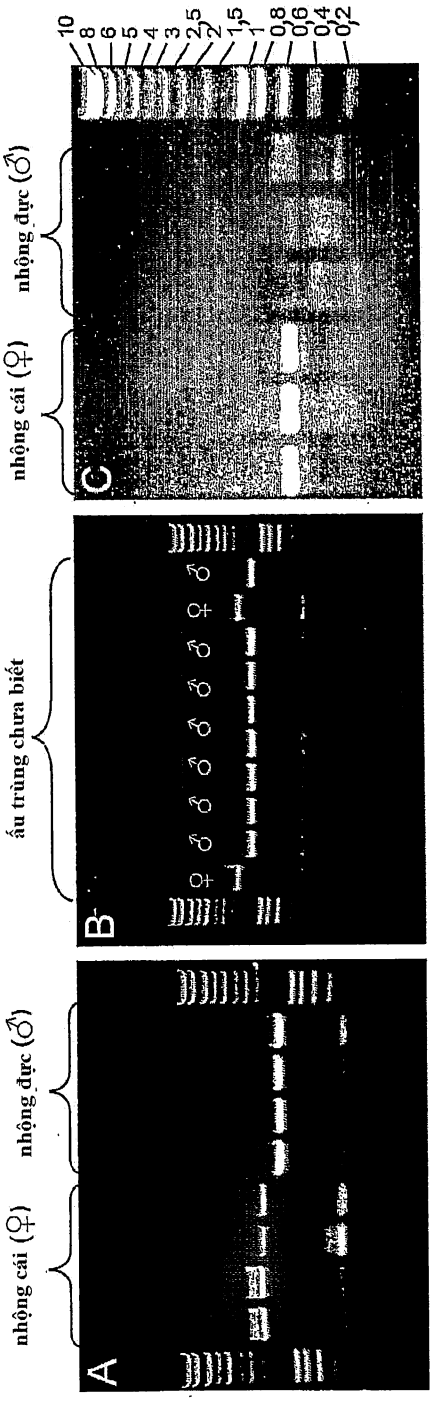
Hình 43



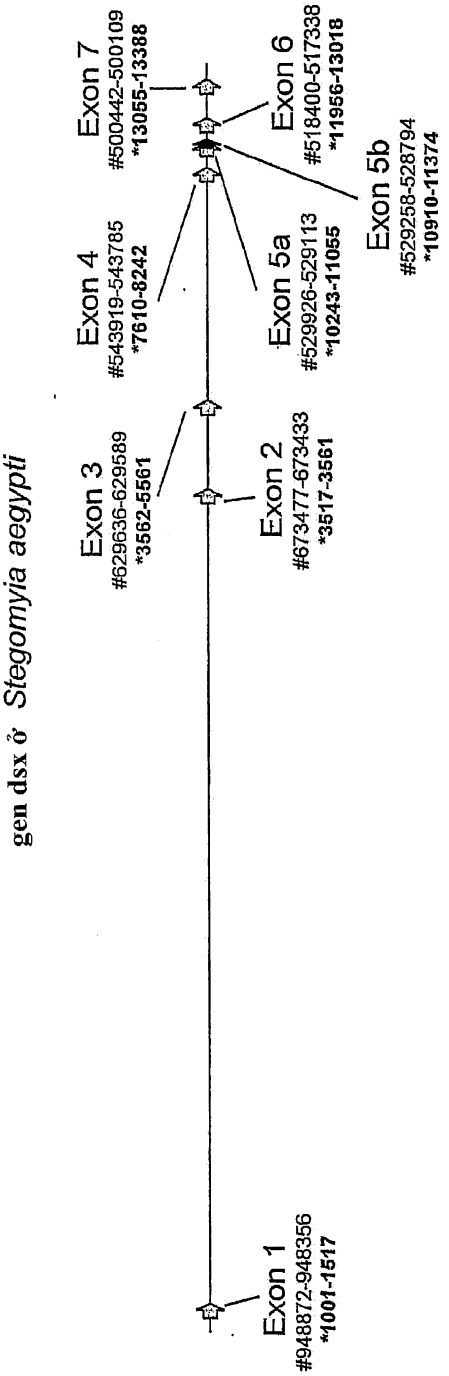
Hình 44



Hình 45

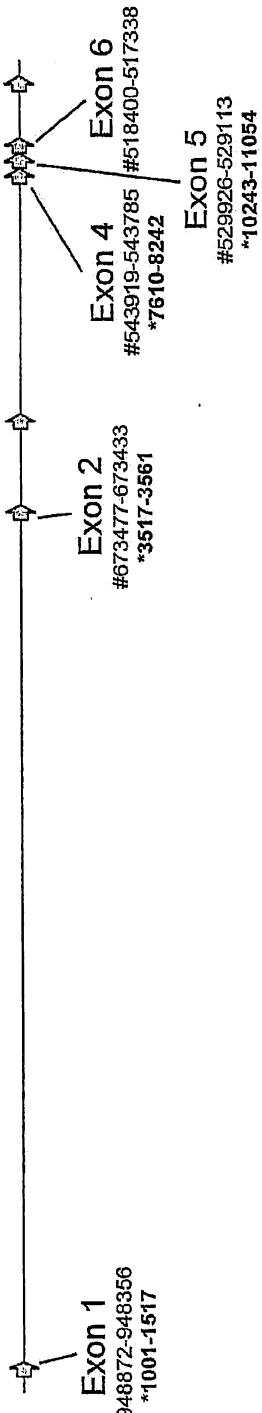

A
nhộng cái (♀)
nhộng đực (♂)
B
ấu trùng chưa biết
C
nhộng cái (♀)
nhộng đực (♂)
10
8
6
5
4
3
2,5
2
1,5
1
0,8
0,6
0,4
0,2

Hình 46

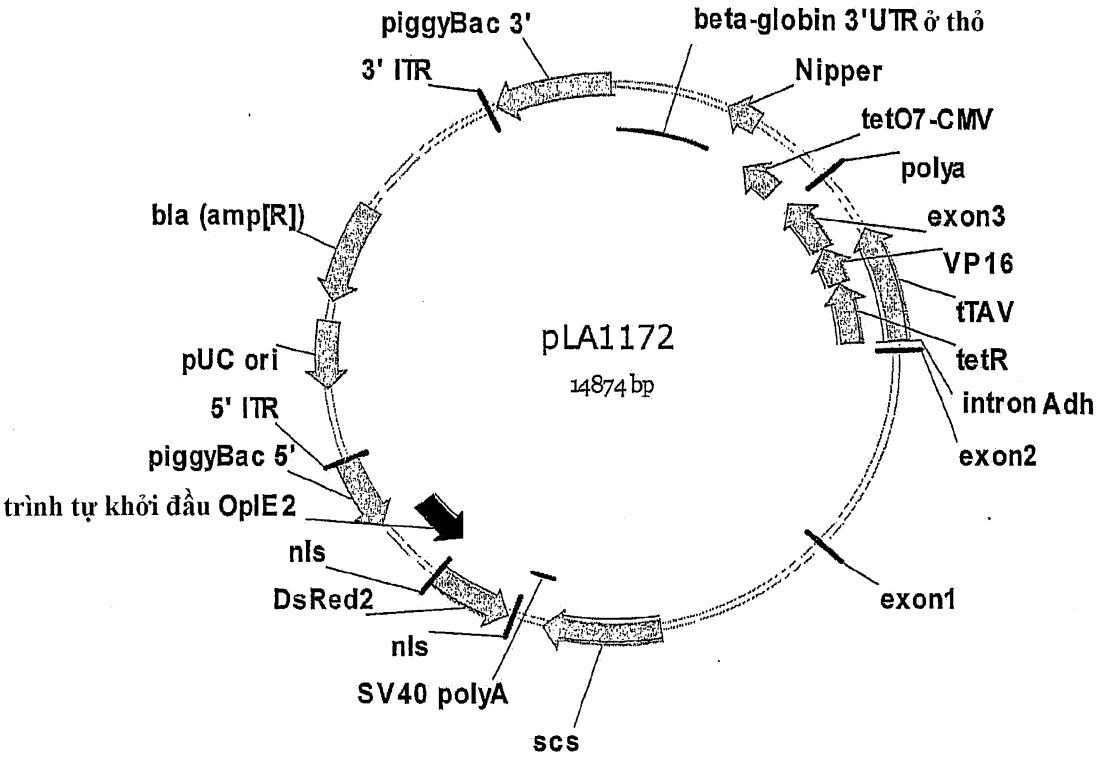

gen dsx ở Stegomyia aegypti
Exon 1
#948872-948356
*1001-1517
Exon 2
#673477-673433
*3517-3561
Exon 3
#629636-629589
*3562-5561
Exon 4
#543919-543785
*7610-8242
Exon 5a
#529926-529113
*10243-11055
Exon 5b
#529258-528794
*10910-11374
Exon 6
#518400-517338
*11956-13018
Exon 7
#500442-500109
*13055-13388

Hình 47

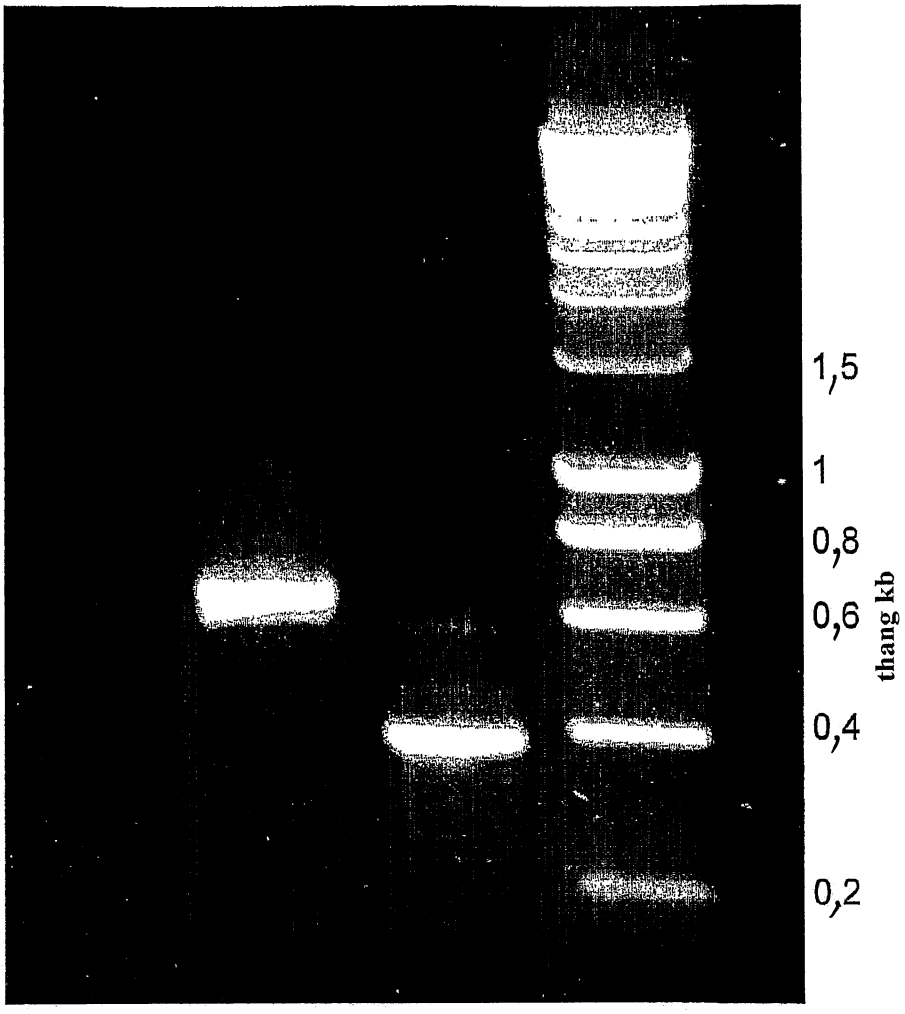
gen dsx ở *Stegomyia aegypti*



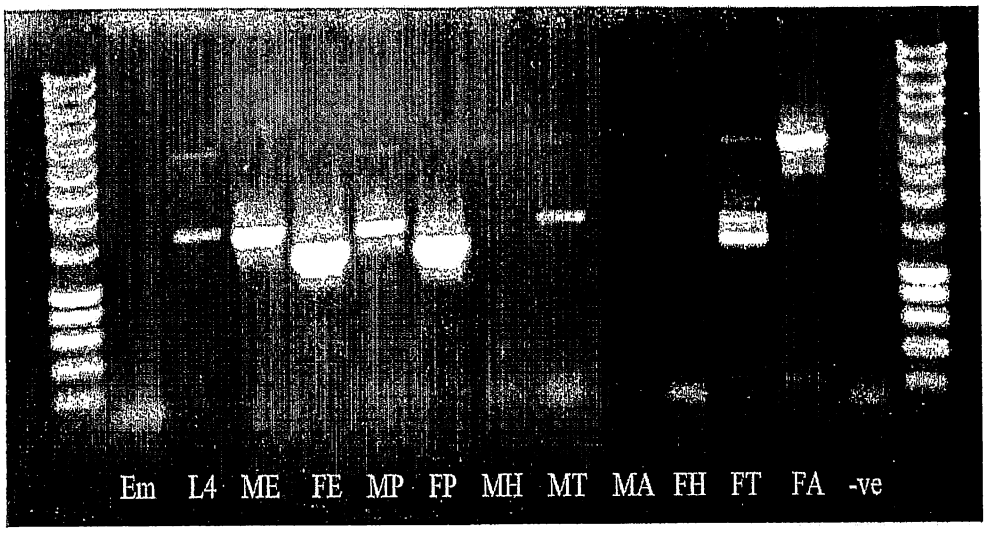
Hình 48


piggyBac 3'
3' ITR
beta-globin 3'UTR ở thỏ
Nipper
tetO7-CMV
polya
exon3
VP16
tTAV
tetR
intron Adh
exon2
exon1
bla (amp[R])
pUC ori
5' ITR
piggyBac 5'
trình tự khởi đầu OpIE2
nls
DsRed2
nls
SV40 polyA
scs
pLA1172
14874 bp

Hình 49

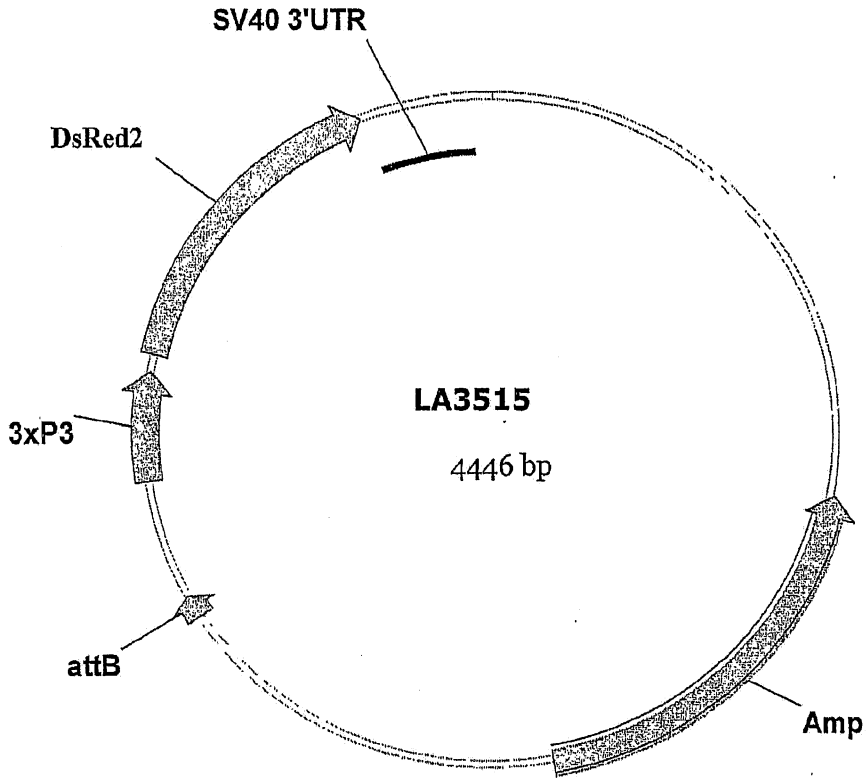


Hình 50

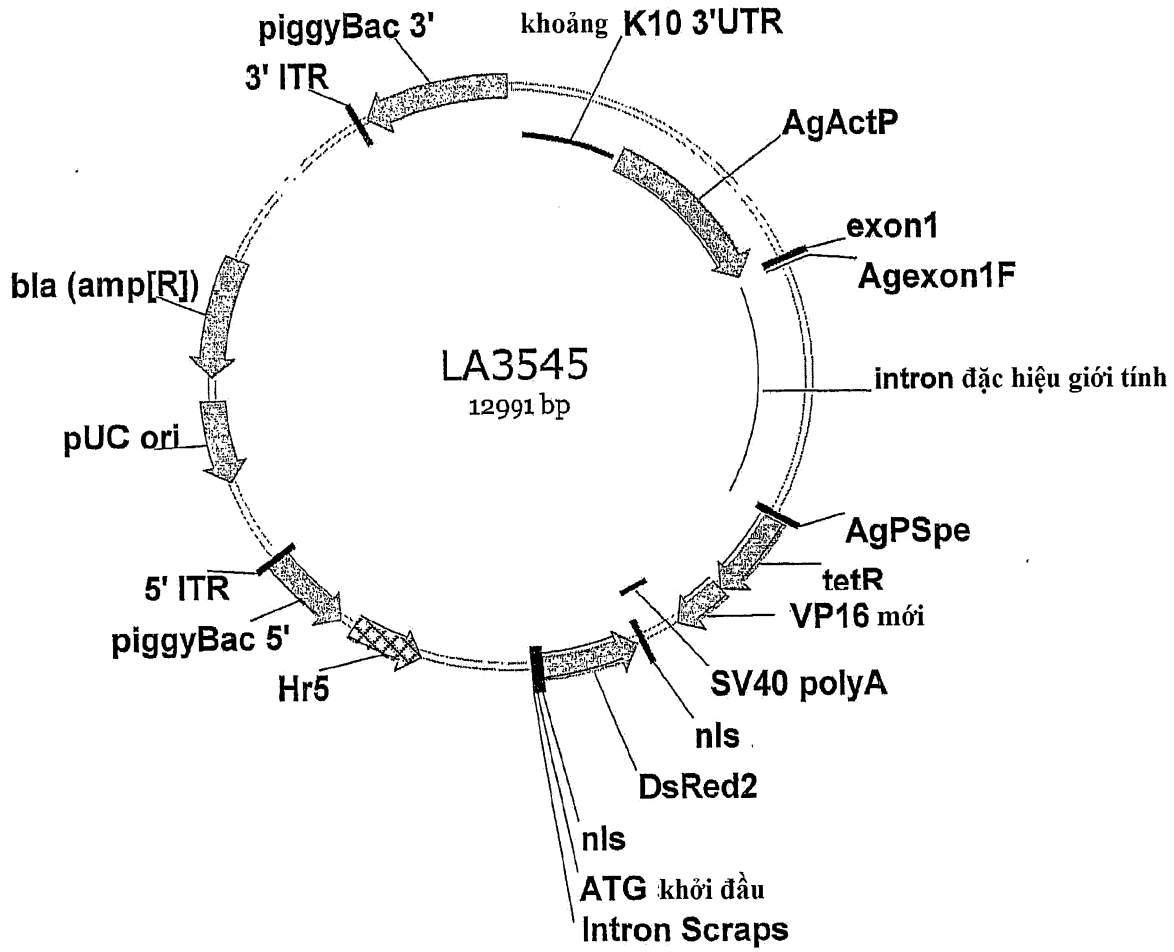


Hình 51

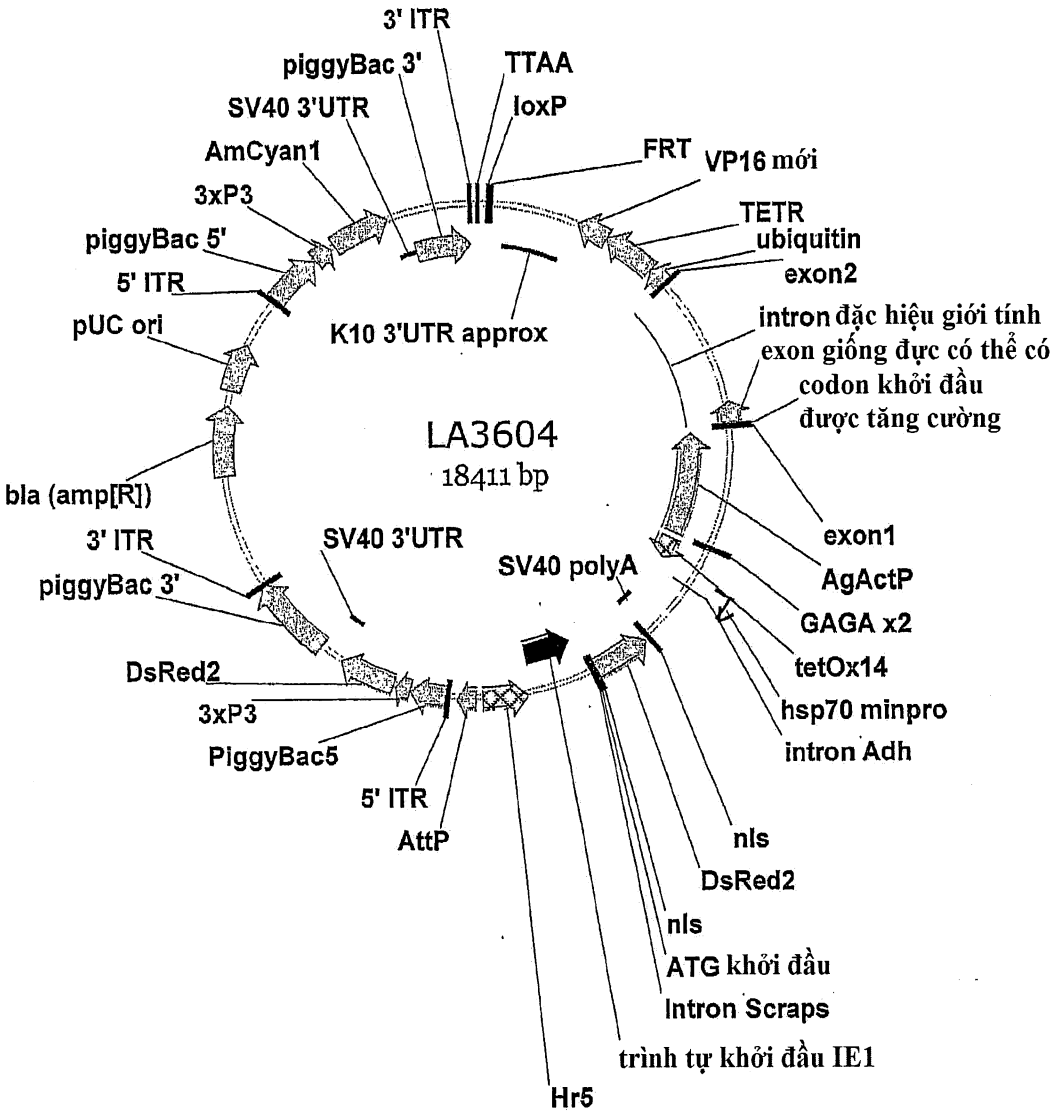
Hình 52- Sơ đồ plasmit LA3515



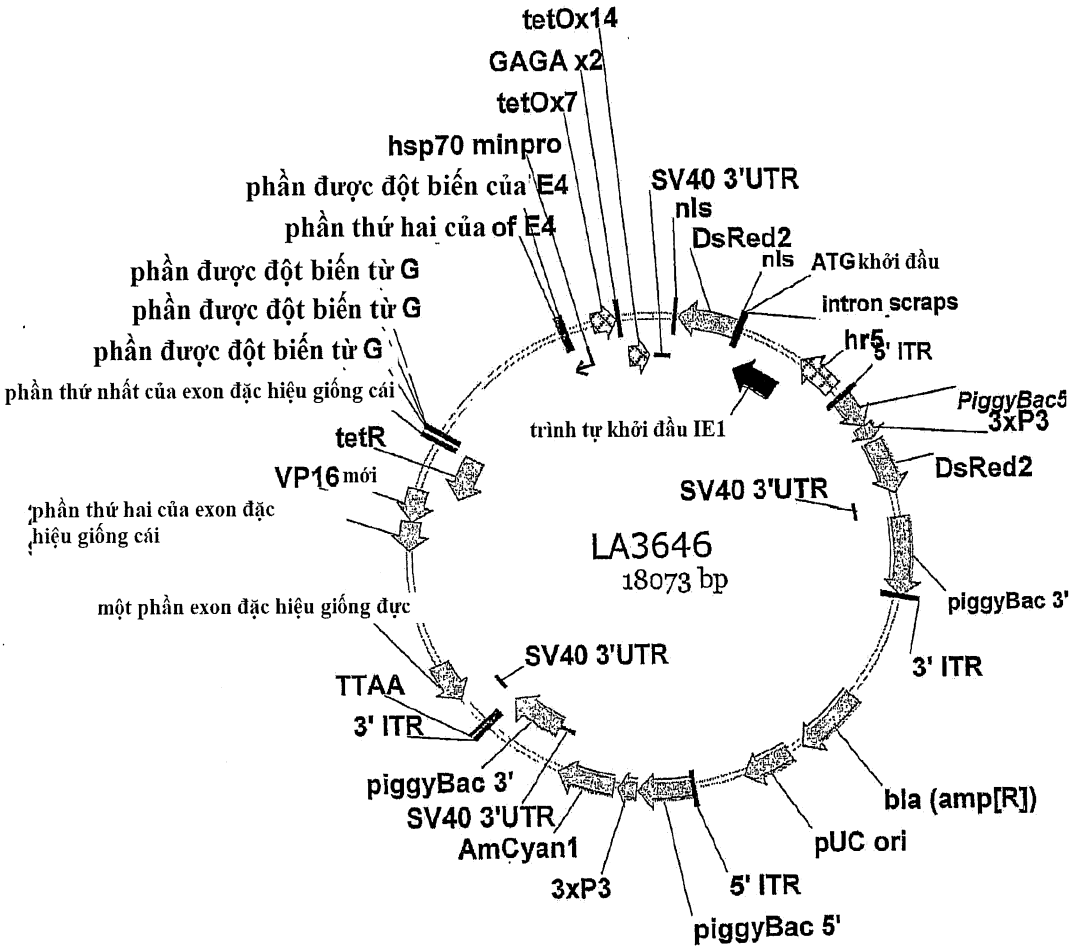
Hình 53 sơ đồ plasmit LA3545



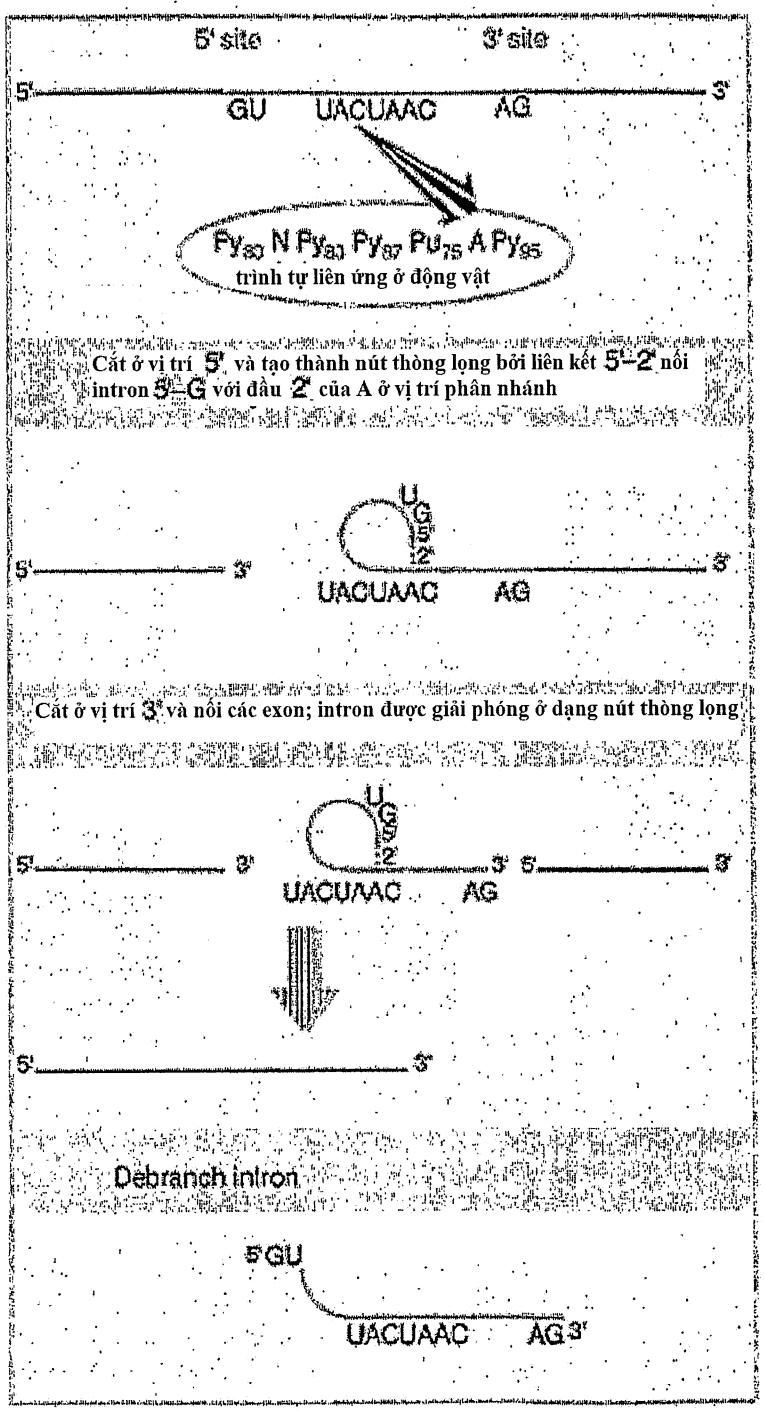
Hình 54 sơ đồ plasmit LA3604



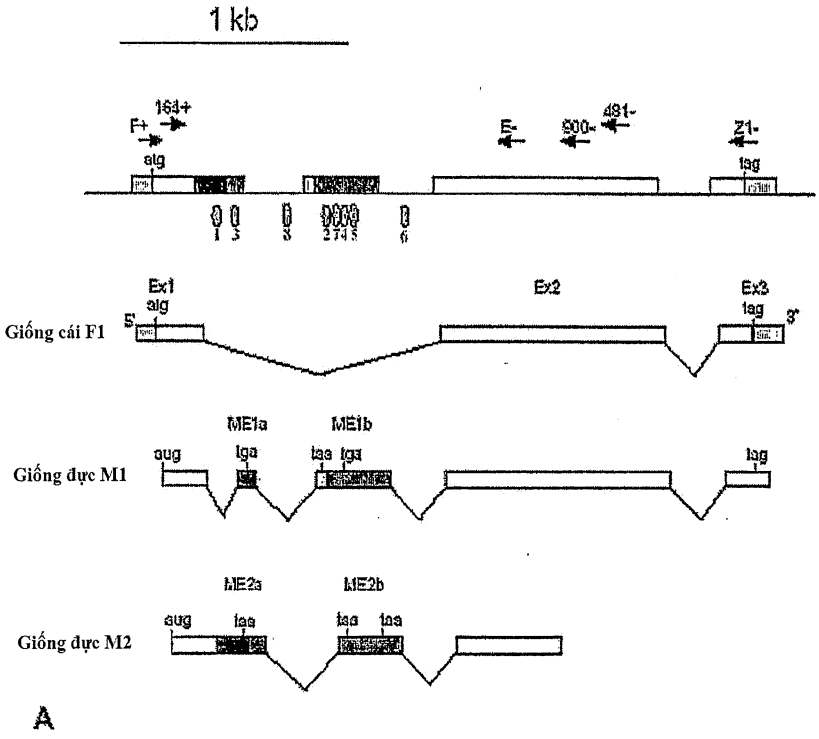
Hình 55 sơ đồ plasmit LA3646



Hình 56

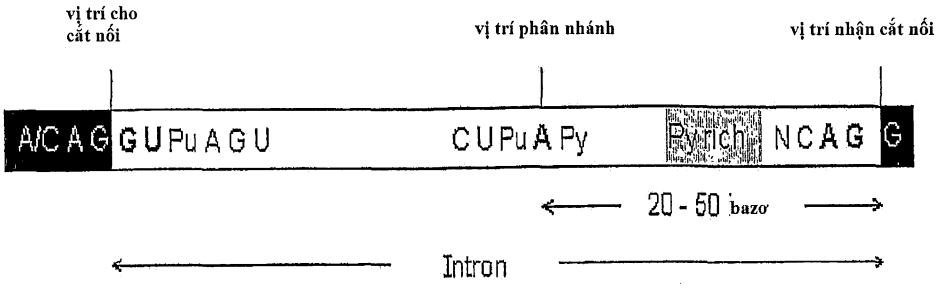

5' site
3' site
5'
3'
GU
UACUAAC
AG
Py80 N Py81 Py87 Pu75 A Py95
trình tự liên ứng ở động vật
Cắt ở vị trí 5' và tạo thành nút thòng lọng bởi liên kết 5'-2' nối intron 5'-G với đầu 2' của A ở vị trí phân nhánh
U
G
5'
2'
5'
3'
UACUAAC
AG
Cắt ở vị trí 3' và nối các exon; intron được giải phóng ở dạng nút thòng lọng
U
G
2'
5'
3'
3' 5'
3'
UACUAAC
AG
5'
3'
Debranch intron
5'GU
UACUAAC
AG 3'

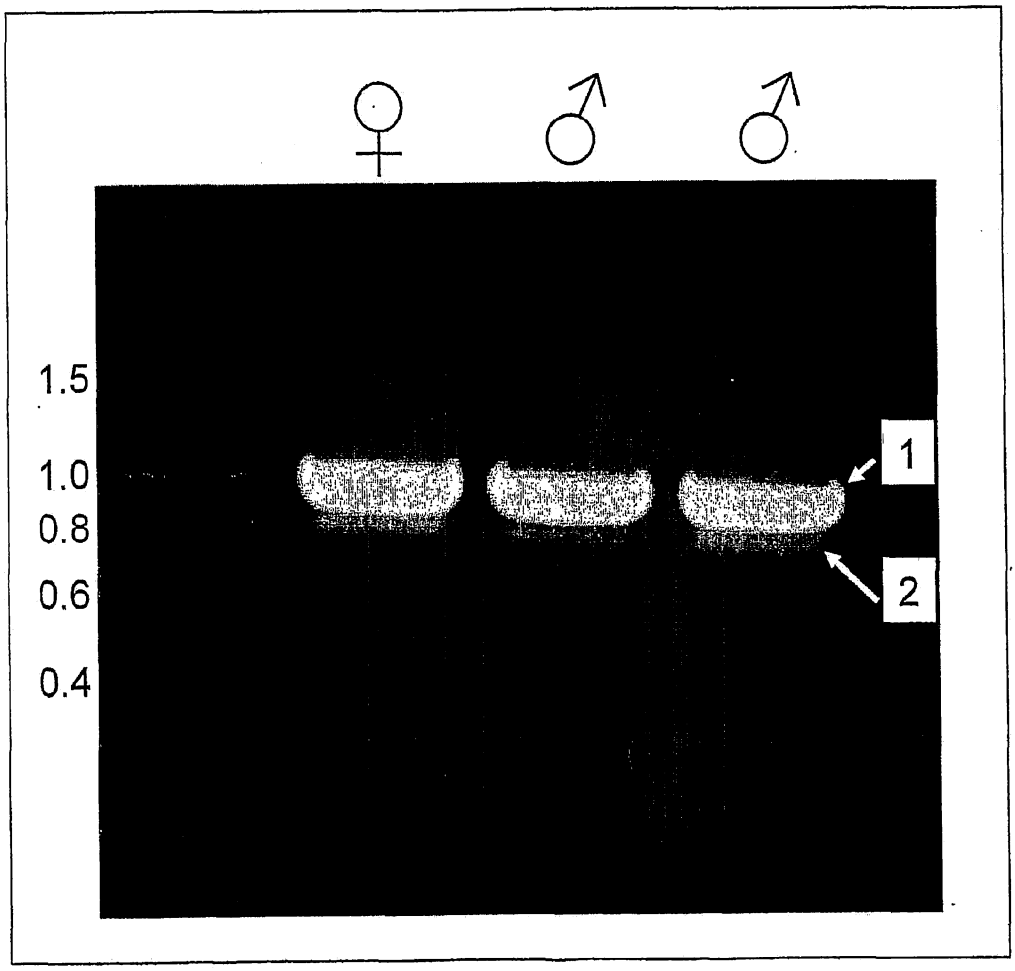
Hình 57


1 kb
164+
F+
atg
E-
900-
481-
Z1-
tag
1 3
3
2 7 4 5
6
Ex1
atg
5'
Ex2
Ex3
tag
3'
Giống cái F1
ME1a
ME1b
aug
tga
taa tga
tag
Giống đực M1
ME2a
ME2b
aug
taa
taa
taa
Giống đực M2

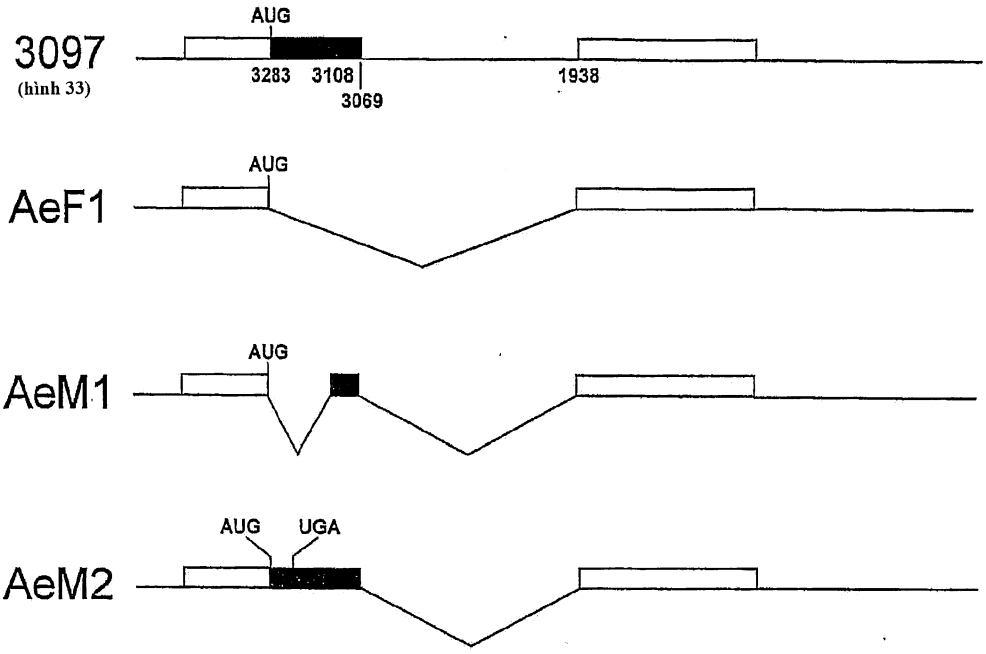
A

Hình 58

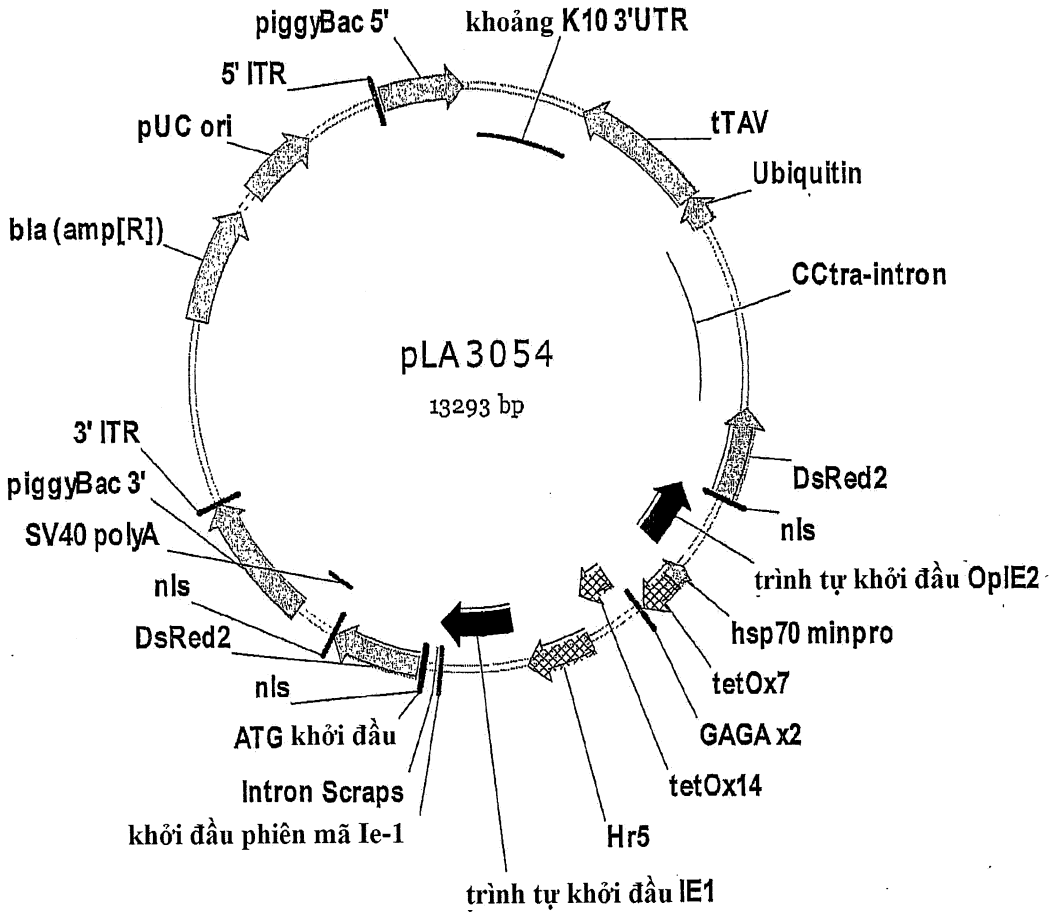




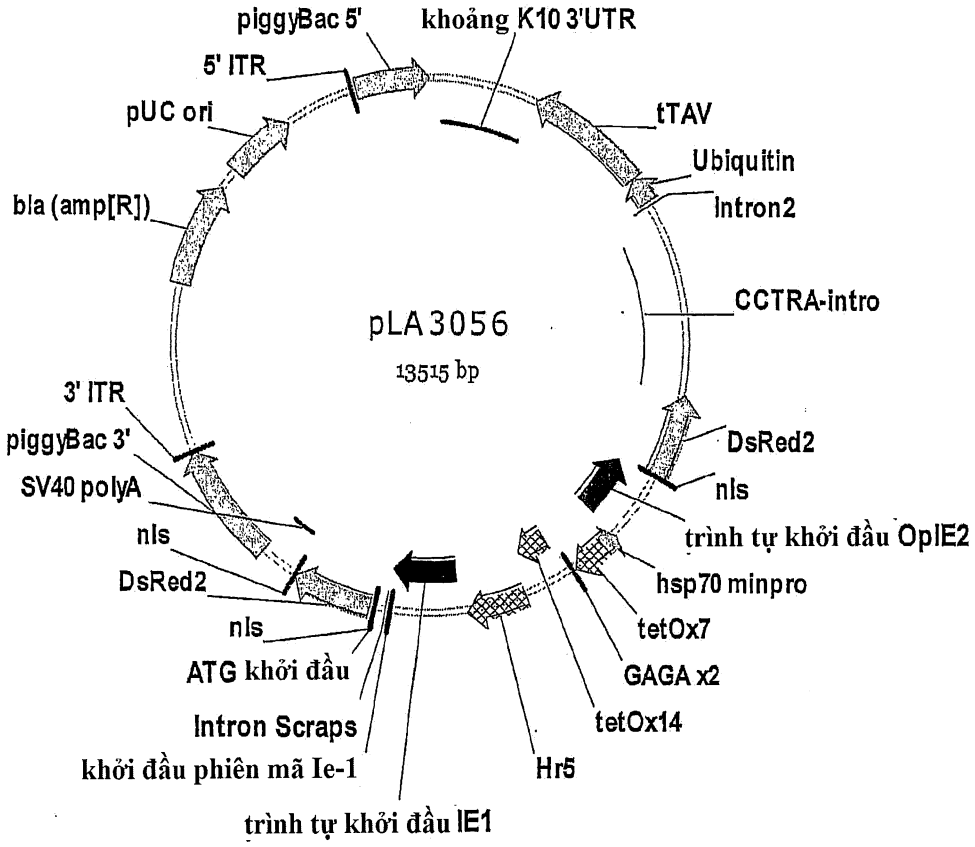
Hình 59


3097
(hình 33)
AUG
3283
3108
3069
1938
AeF1
AUG
AeM1
AUG
AeM2
AUG
UGA

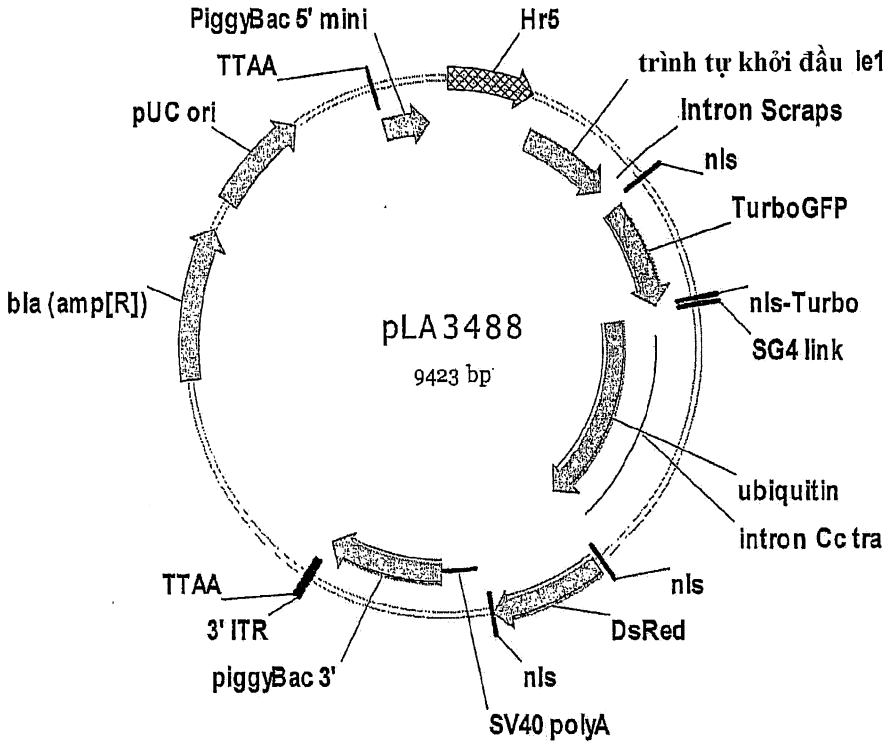
Hình 60


piggyBac 5'
khoảng K10 3'UTR
5' ITR
pUC ori
tTAV
Ubiquitin
bla (amp[R])
CCtra-intron
pLA3054
13293 bp
3' ITR
DsRed2
piggyBac 3'
nls
SV40 polyA
trình tự khởi đầu OpIE2
nls
hsp70 minpro
DsRed2
tetOx7
nls
GAGA x2
ATG khởi đầu
tetOx14
Intron Scraps
khởi đầu phiên mã Ie-1
Hr5
trình tự khởi đầu IE1

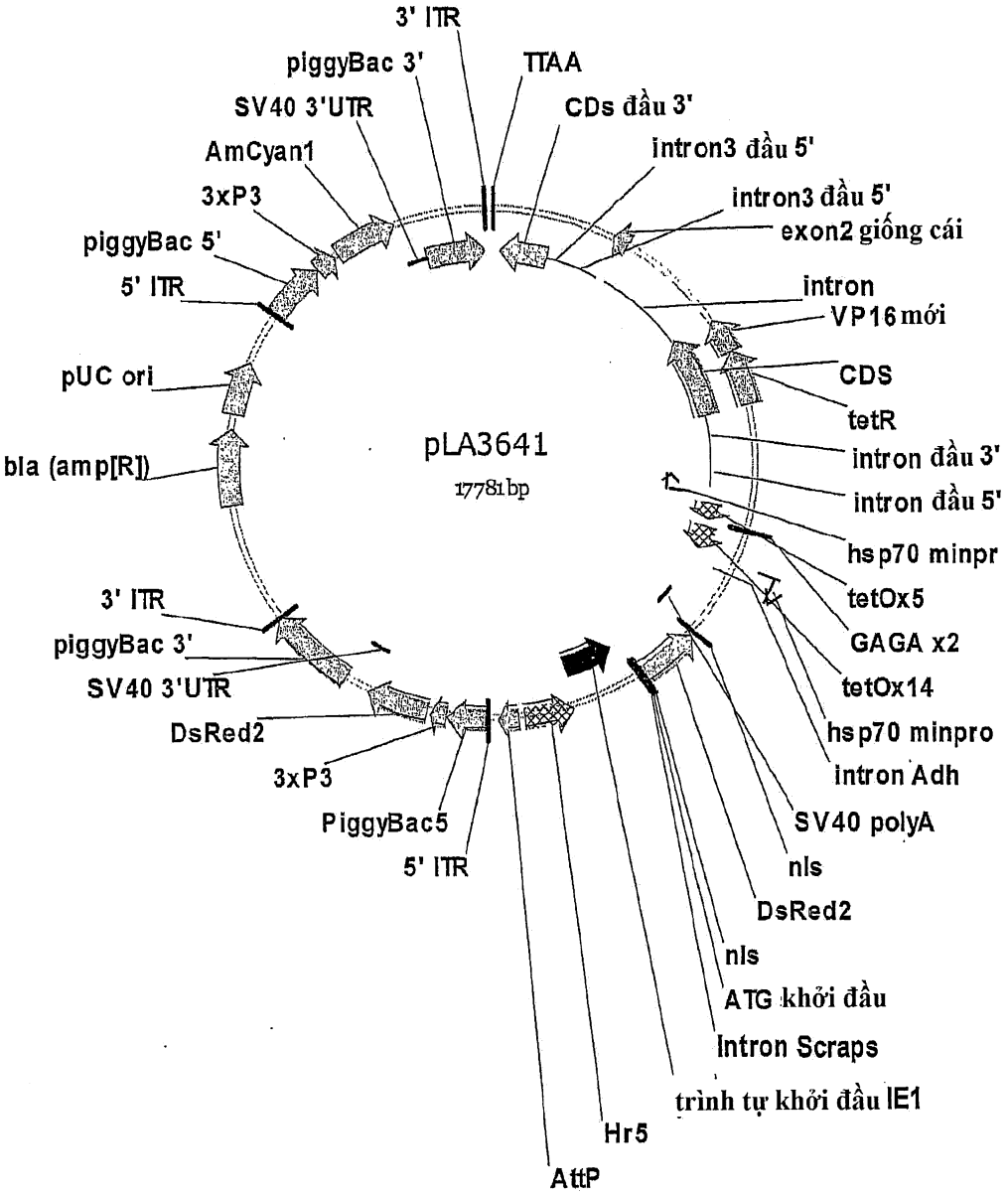
Hình 61


piggyBac 5'
khoảng K10 3'UTR
5' ITR
pUC ori
tTAV
Ubiquitin
bla (amp[R])
Intron2
CCTRA-intro
pLA3056
13515 bp
3' ITR
DsRed2
piggyBac 3'
nls
SV40 polyA
trình tự khởi đầu OpIE2
nls
hsp70 minpro
DsRed2
tetOx7
nls
GAGA x2
ATG khởi đầu
tetOx14
Intron Scraps
khởi đầu phiên mã Ie-1
Hr5
trình tự khởi đầu IE1

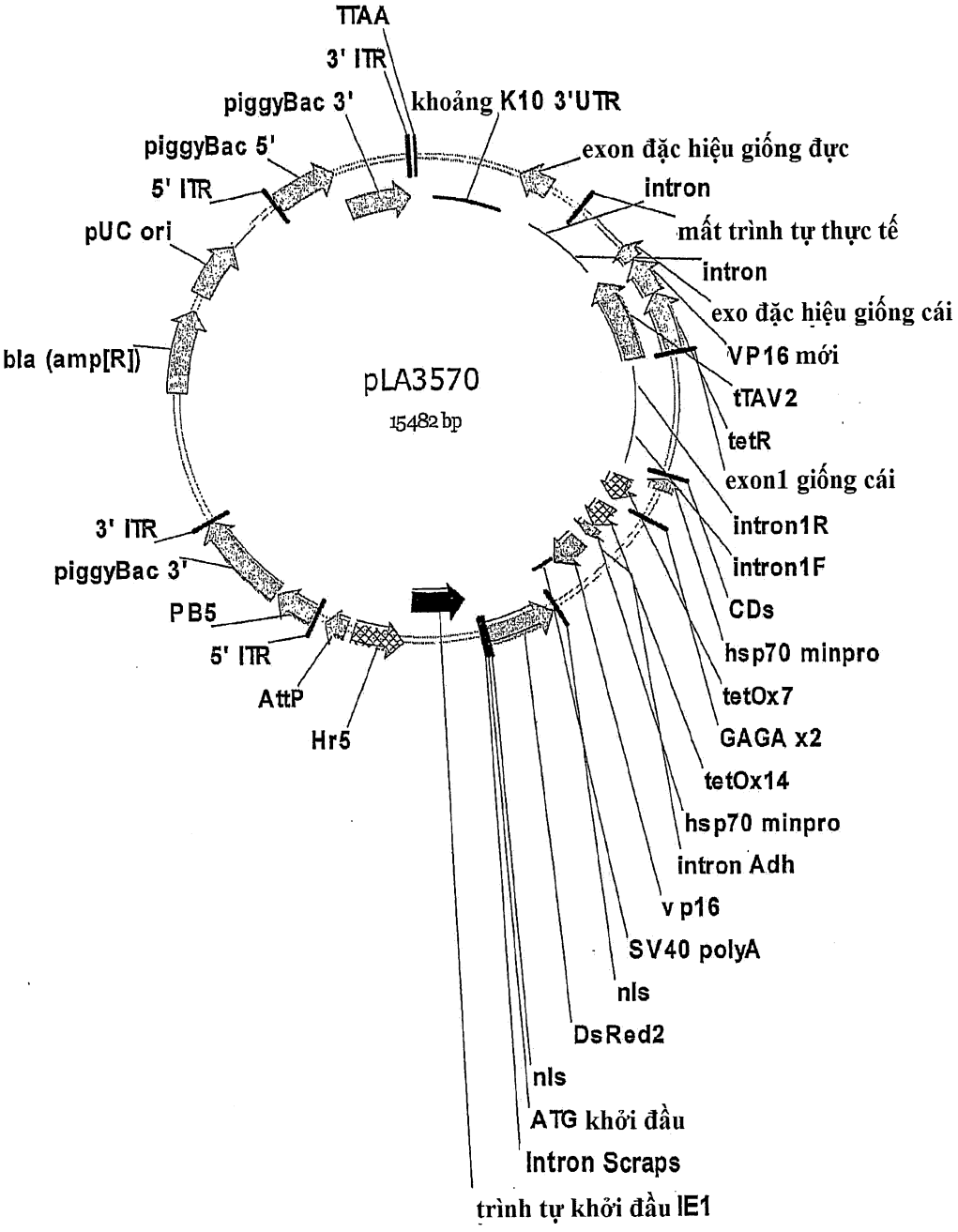
Hình 62


PiggyBac 5' mini
TTAA
pUC ori
bla (amp[R])
TTAA
3' ITR
piggyBac 3'
Hr5
trình tự khởi đầu le1
Intron Scraps
nls
TurboGFP
nls-Turbo
SG4 link
pLA3488
9423 bp
ubiquitin
intron Cc tra
nls
DsRed
nls
SV40 polyA

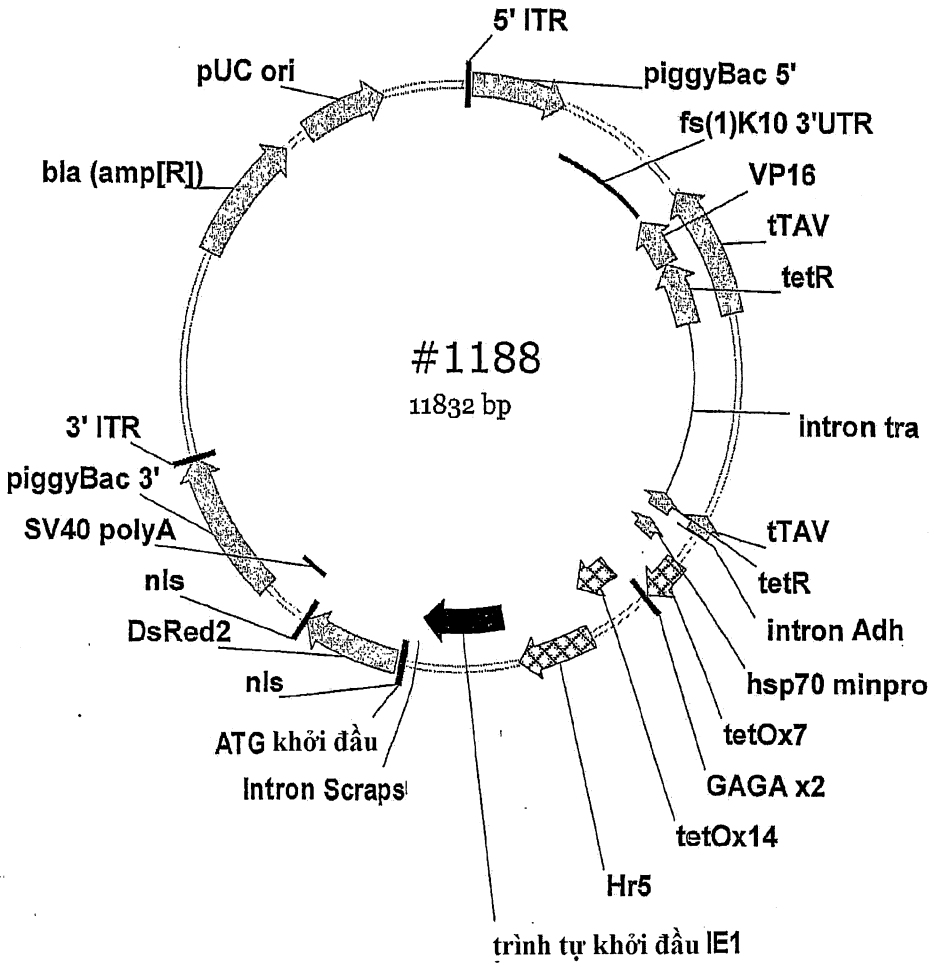
Hình 63


3' ITR
piggyBac 3'
TTAA
SV40 3'UTR
CDs đầu 3'
AmCyan1
intron3 đầu 5'
3xP3
intron3 đầu 5'
piggyBac 5'
exon2 giống cái
5' ITR
intron
VP16 mới
pUC ori
CDS
tetR
pLA3641
17781bp
bla (amp[R])
intron đầu 3'
intron đầu 5'
hsp70 minpr
3' ITR
tetOx5
piggyBac 3'
GAGA x2
SV40 3'UTR
tetOx14
DsRed2
hsp70 minpro
3xP3
intron Adh
PiggyBac5
SV40 polyA
5' ITR
nls
DsRed2
nls
ATG khởi đầu
Intron Scraps
trình tự khởi đầu IE1
Hr5
AttP

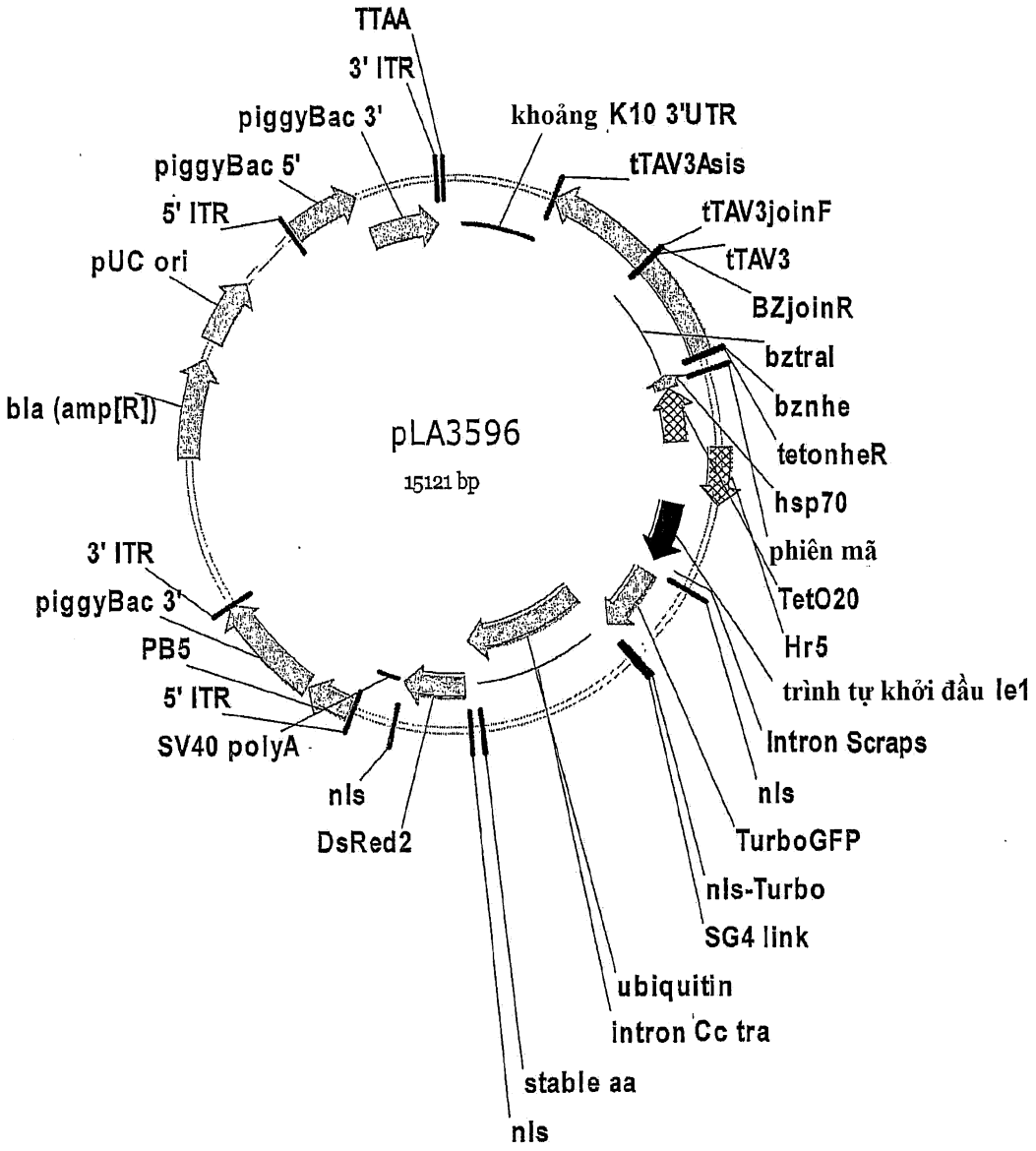
Hình 64


TTAA
3' ITR
piggyBac 3'
khoảng K10 3'UTR
piggyBac 5'
exon đặc hiệu giống đực
5' ITR
intron
pUC ori
mất trình tự thực tế
intron
exo đặc hiệu giống cái
bla (amp[R])
pLA3570
15482 bp
VP16 mới
tTAV2
tetR
exon1 giống cái
3' ITR
intron1R
piggyBac 3'
intron1F
PB5
CDs
5' ITR
hsp70 minpro
AttP
tetOx7
Hr5
GAGA x2
tetOx14
hsp70 minpro
intron Adh
v p16
SV40 polyA
nls
DsRed2
nls
ATG khởi đầu
Intron Scraps
trình tự khởi đầu IE1

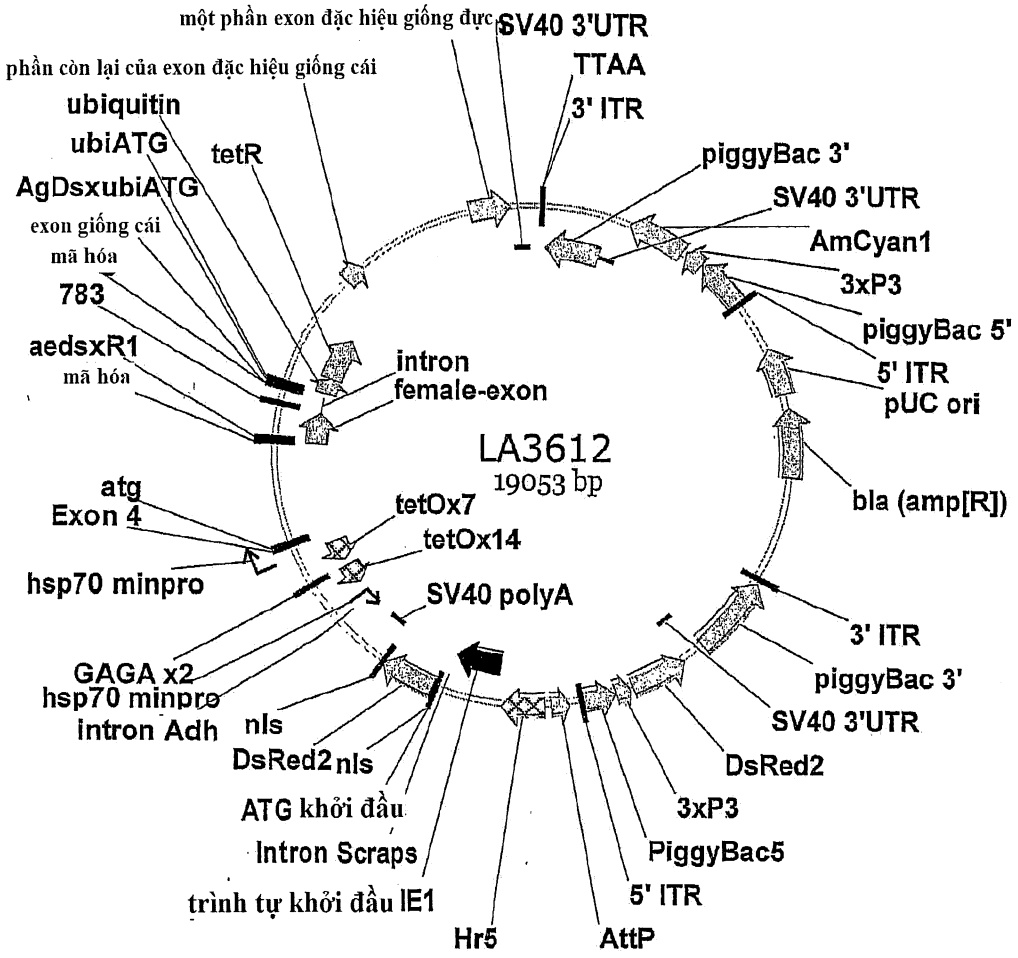
Hình 65


5' ITR
pUC ori
piggyBac 5'
fs(1)K10 3'UTR
bla (amp[R])
VP16
tTAV
tetR
#1188
11832 bp
intron tra
3' ITR
piggyBac 3'
SV40 polyA
tTAV
nls
tetR
DsRed2
intron Adh
nls
hsp70 minpro
ATG khởi đầu
tetOx7
Intron Scraps
GAGA x2
tetOx14
Hr5
trình tự khởi đầu IE1

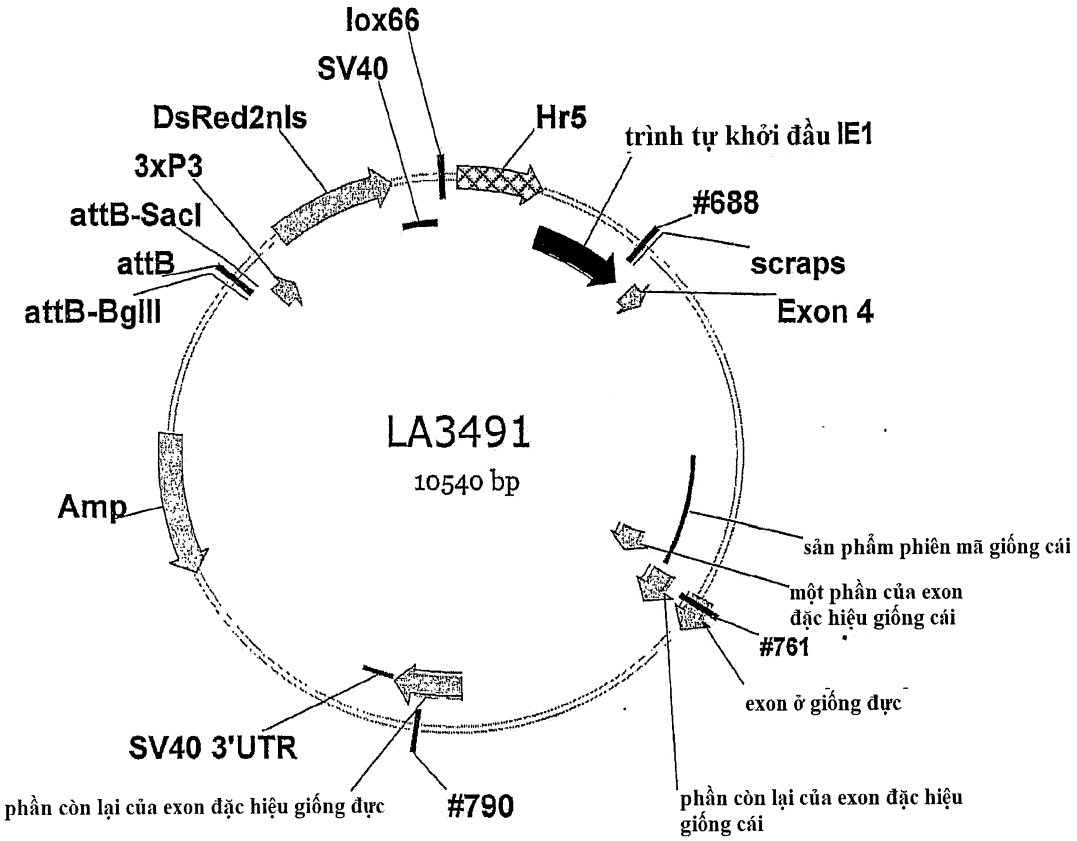
Hình 66


TTAA
3' ITR
piggyBac 3'
piggyBac 5'
5' ITR
pUC ori
bla (amp[R])
3' ITR
piggyBac 3'
PB5
5' ITR
SV40 polyA
nls
DsRed2
stable aa
nls
ubiquitin
intron Cc tra
nls-Turbo
SG4 link
TurboGFP
nls
Intron Scraps
trình tự khởi đầu le1
Hr5
TetO20
phiên mã
hsp70
tetonheR
bznhe
bztral
BZjoinR
tTAV3
tTAV3joinF
tTAV3Asis
khoảng K10 3'UTR
pLA3596
15121 bp

Hình 67


một phần exon đặc hiệu giống đực
SV40 3'UTR
TTAA
3' ITR
phần còn lại của exon đặc hiệu giống cái
ubiquitin
ubiATG
tetR
AgDsxubiATG
exon giống cái
mã hóa
783
aedsxR1
mã hóa
intron
female-exon
piggyBac 3'
SV40 3'UTR
AmCyan1
3xP3
piggyBac 5'
5' ITR
pUC ori
LA3612
19053 bp
bla (amp[R])
atg
Exon 4
tetOx7
tetOx14
hsp70 minpro
SV40 polyA
3' ITR
piggyBac 3'
SV40 3'UTR
DsRed2
3xP3
PiggyBac5
5' ITR
AttP
GAGA x2
hsp70 minpro
intron Adh
nls
DsRed2 nls
ATG khởi đầu
Intron Scraps
trình tự khởi đầu IE1
Hr5

Hình 68


lox66
SV40
DsRed2nls
Hr5
trình tự khởi đầu IE1
3xP3
attB-SacI
#688
attB
scraps
attB-BglII
Exon 4
LA3491
10540 bp
Amp
sản phẩm phiên mã giống cái
một phần của exon
đặc hiệu giống cái
#761
exon ở giống đực
SV40 3'UTR
phần còn lại của exon đặc hiệu giống đực
#790
phần còn lại của exon đặc hiệu
giống cái

Hình 69