(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) **1-0028549**
(51)[7] **C12N 9/88**
(13) **B**

(21) 1-2015-00872
(22) 15/08/2013
(86) PCT/US2013/055198 15/08/2013
(87) WO 2014/028772 A2 20/02/2014
(30) 61/683,914 16/08/2012 US
(45) 25/06/2021 399
(43) 25/08/2015 329A
(73) BANGLADESH JUTE RESEARCH INSTITUTE (BD)
Manik Mia Avenue, Dhaka, 1207 (BD) Bangladesh, People's Republic
(72) ISLAM, Mohammed, Shahidul (BD); HOSSEN, Mohammed, Mosaddeque (BD); HAQUE, Mohammed, Samiul (BD); ALAM, Mohammed, Monjurul (BD).
(74) Công ty TNHH Trần Hữu Nam và Đồng sự (TRAN H.N & ASS.)

(54) POLYNUCLEOTIT ĐƯỢC PHÂN LẬP MÃ HÓA PECTAT LYAZA TỪ NẤM MACROPHOMINA PHASEOLINA, PHÂN TỬ AXIT NUCLEIC, VẬT TRUYỀN VÀ CẤU TRÚC BIỂU HIỆN CHỨA POLYNUCLEOTIT NÀY

(57) Sáng chế đề cập đến polynucleotit được phân lập mã hóa cho enzym, có nguồn gốc từ nấm *Macrophomina phaseolina ("M. phaseolina")*, chịu trách nhiệm về việc làm thoái hóa, và nó có chứa và/hoặc gồm có trình tự nucleotit nêu trong SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61, hoặc trình tự bổ sung của các trình tự này. Sáng chế cũng đề cập đến polypeptit được phân lập được mã hóa bởi trình tự polynucleotit nêu trong SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63; cấu trúc gen tái tổ hợp có chứa polynucleotit; nấm biến nạp và chuyển gen có chứa cấu trúc gen tái tổ hợp, với hoặc có sự sản xuất tăng cường của enzym làm thoái biến pectin. Polypeptit theo sáng chế có thể được sử dụng, trong số những thứ khác, để sản xuất nước ép trái cây, sản phẩm dệt, bột giấy và giấy, cà phê, trà và chất chiết dầu và xử lý nước thải sản sinh ra pectic.

Lĩnh vực kỹ thuật được đề cập

Sáng chế đề cập đến polypeptit được phân lập từ *M. phaseolina* có hoạt tính pectat lyaza và polynucleotit được phân lập mã hóa cho polypeptit này, và phương pháp sản xuất và sử dụng các polynucleotit và polypeptit này. Sáng chế cũng đề cập đến cấu trúc axit nucleic, vật truyền, và tế bào chủ có chứa polynucleotit cũng như là phương pháp sản xuất và sử dụng polypeptit này. Polypeptit theo sáng chế có thể được sử dụng, cùng với những thứ khác, trong công nghiệp dệt để giảm và khử keo sợi và vải; trong công nghiệp thực phẩm để chiết nước ép trái cây và rau; trong công nghiệp giấy để sản xuất giấy có chất lượng tốt; và cũng để lên men cà phê và trà, chiết dầu và xử lý nước thải sản sinh ra pectic.

Tình trạng kỹ thuật của sáng chế

Pectin là nhóm dị polyme phức hợp, có mặt ở lớp mỏng ở giữa của thành tế bào thực vật sơ cấp và đóng vai trò làm chất gắn gian bào. Chúng cũng có chức năng quan trọng trong việc điều hòa nước của thực vật do bản chất keo của chúng. Lớp mỏng ở giữa mà nằm ở giữa các thành tế bào chủ yếu được tạo thành từ protopectin (dạng không tan của pectin).

Pectin được tạo thành bởi hai vùng xác định khác nhau: vùng "nhẵn" và vùng "có lông". Vùng "nhẵn" gồm có khung chính của axit D-galacturonic được liên kết α-1,4, tạo thành axit polygalactouronic có một số nhóm carboxyl được este hóa bằng metanol (Rouse A. Pectin: distribution, significance. In: Nagy S, Shaw P, Veldhuis Meds. Citrus Science and Technology, Vol I. Westport CT: AVI publishing Inc. 1977). Trong vùng "có lông", khung chính axit galacturonic bị làm đứt bằng đơn vị L-rhamnoza được liên kết α-1, 2.

Pectin/pectat lyaza giải trùng hợp pectin trong vùng nhẵn, mà phân giải liên kết glycosiđic thông qua sự tách β để tạo ra oligome mà là đầu không khử chưa no 4, 5 (Lombard V, Bernard T, Rancurel C, Brumer H, Coutinho PM, Henrissat B. A hierarchical classification of polysaccharide lyases for glycogenomics. Biochem J.

2010;432(3):437–444). Trong khi đó rhamnogalacturonaza chia ra trong các vùng có lông của pectin (Jensen MH, Otten H, Christensen U, Borchert TV, Christensen LL, Larsen S, Leggio LL. Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from *Aspergillus aculeatus*. J Mol Biol. 2010;404(1):100–111).

Pectin esteraza thủy phân pectin thành metanol và axit polygalacturonic. Enzym này có thể được tạo ra bởi một số nấm bao gồm *Aspergillus* sp, *Botrytis cinerea, Fusarium monilforme, Rhizopus stolonifer, Trichoderma* sp., v.v., (Polizeli ML, Jorge JA, Terenzi HF. Pectinase production by *Neurospora crassa*: purification and biochemical characterization of extracellular polygalacturonase activity, J. Gen. Microbiol. 1991;137: 1815-1823). Nhưng *Aspergillus* là nguồn chính để sản xuất thương mại pectin esteraza (Torres EF, Aguilar C, Esquivel JCC, Gonzales GV. Pectinase. In: Enzyme Technology, Pandey, A., Webb, C., Soccol, C.R., Larroche, C (Eds), Asiatech Publishiers Inc., New Delhi, India, 2005. pp. 273-296).

Enzym làm thoái biến pectin là công cụ quan trọng trong ngành công nghiệp thực phẩm, trước hết để xử lý trái cây và rau chẳng hạn như sự sản xuất nước ép trái cây hoặc sản xuất rượu. Các lĩnh vực ứng dụng khác bao gồm công nghiệp bột giấy và giấy (Reid I, Ricard M. Pectinase in papermaking: solving retention problems in mechanical pulps bleached with hydrogen peroxide. Enz. Microbiol. Technol. 2000;26:115-123), thức ăn gia súc (Barreto de Menezes TJ, Salva JG, Baldini VL, Papini RS, Sales AM. Protein enrichment of citrus wastes by solid substrate fermentation. Proc. Biochem. 1989;23:167-171), giầm sợi lanh và các loại sợi rau khác (Hoondal GS, Tiwari RP, Tiwari R, Dahiya N, Beg QK. Microbial alkaline pectinases and their applications: a review. Appl. Microbiol. Biotechnol. 2000;9:409–418), lên men cà phê và trà (Gar JG. Tea, coffee and cocoa. In: wood BJB, editor. Microbiology of fermented foods. 1985;vol 2. London: Elsevier Sci. Ltd. pp:133-154), chiết dầu (Scott D. Enzymes, industrial. In: Encyclopedia of Chemical Technology. Grayson M, Ekarth D and Othmer K (eds), 1978; Wiley, NY. pp: 173–224), làm sạch bằng phương pháp sinh học sợi bông (Singh R, Saxena S, Gupta R. Microbial pectinolytic enzymes: A review. Proc. Biochem. 2005;40:2931-2944), khử keo sợi vỏ thực vật (Kapoor M, Beg QK, Bhushan B, Singh K, Dadich KS, Hoondal GS. Application of alkaline and

thermostable polygalacturonase from *Bacillus* sp. MGcp-2 in degumming of ramie (*Boehmeria nivea*) and sunn hemp (*Crotolaria juncia*) bast fibers. Proc. Biochem. 2001;36:803–817), công nghiệp dệt (Karmakar SR. Chemical technology in the pretreatment processes of textiles. In: Textile science and technology series1st ed., Amsterdam: Elsevier Science B.V. 1999; p. 12) và quản lý chất thải (Kashyap DR, Vohra PK, Chopra S, Tewari R. Applications of pectinases in the commercial sector: a review. Biores. Technol. 2001;77:215–227). WO 98/45393 bộc lộ hợp phần tẩy rửa có chứa protopectinaza có khả năng tẩy rửa khác thường đối với đất bùn.

Các enzym này được sử dụng ở dạng riêng rẽ hoặc pha trộn (hỗn hợp) trong công nghiệp. Ví dụ, trong ngành công nghiệp thực phẩm nó được sử dụng dưới dạng enzym đơn lẻ chẳng hạn như pectin esteraza cần cho sự gel hóa cũng như là dạng kết hợp của các enzym khác nhau chẳng hạn như pectin esteraza với polygalacturonaza cần cho sự hóa lỏng của nguyên liệu thực vật (Heldt-Hansen HP, Kofod LV, Budolfsen G, Nielsen PM, Hüttel S, Bladt T. Application of tailor made pectinases. In: Visser J and Voragen AJG (eds), Pectin and Peactnases. Progress in Biotechnology. 1996;14:463-474).

Sự tách dòng và sự biểu hiện của một số enzym thu được từ *Aspergillus niger* đã được báo cáo. EP0278355 mô tả sự tách dòng của gen pectin lyaza, trình tự và sự biểu hiện của chúng. EP0353188 bổ sung một số pectin lyaza khác. EO0421919 bộc lộ hai polygalacturonaza và endo-polygalacturonaza khác đã được bộc lộ trong EP0388593. Cả hai đơn sáng chế này đều sử dụng *Aspergillus niger* làm nguồn gen. WO 94/14952 mô tả ba enzym có hoạt tính endo-polygalacturonaza thu được từ *Aspergillus aculeatus.* Tuy nhiên, không có công bố nào được báo cáo về việc tách dòng của gen mã hóa cho enzym làm thoái biến pectin từ *M. phaseolina.*

Sáng chế thực hiện nghiên cứu hệ gen để bộc lộ gen và protein được mã hóa của chúng của enzym làm thoái biến pectin có nguồn gốc từ *M. phaseolina* mà có thể được sử dụng, ngoài những cái khác, trong các quy trình công nghiệp hoặc nhằm mục đích công nghiệp.

Bản chất kỹ thuật của sáng chế

Ngoài những cái khác, sáng chế đề cập đến enzym làm thoái biến pectin mà có thể thu được từ *M. phaseolina*. Sáng chế cũng đề cập đến việc sử dụng nấm *M. phaseolina* trong việc làm thoái biến pectin.

Mục đích chính theo sáng chế là bộc lộ bộ trình tự nucleotit mã hóa cho pectat lyaza (SEQ ID NO. 1, và 2, của nấm *M. phaseolina*). Đối với mỗi gen theo sáng chế, trình tự khung đọc mở (ORF) được tạo dẫn xuất bằng tay từ trình tự hệ gen tương ứng bằng cách bỏ đi các trình tự intron dự đoán và nối các trình tự exon cùng nhau. Vật truyền, vật truyền biểu hiện, và tế bào chủ có chứa gen enzym cũng được bao gồm.

Ngoài ra, sáng chế đề xuất trình tự polypeptit suy ra từ trình tự ORF của gen. Trình tự polypeptit theo sáng chế tương ứng với trình tự polypeptit của pectat lyaza (SEQ ID NO. 3). Sáng chế cũng đề cập đến polynucleotit được phân lập có chứa phần bổ sung của trình tự nucleotit được mô tả ở trên.

Mục đích khác của sáng chế là đề xuất thông tin sinh học phân tử và di truyền của gen và enzym nêu trong mục đích chính để được khai thác/sử dụng để điều hòa, chuyển hóa sự thoái biến pectin để sản xuất sản phẩm công nghiệp có giá trị.

Ngoài ra, mục đích của sáng chế là làm thuận lợi cho sự sản xuất *in vitro* của polypeptit làm thoái biến pectin, và sáng chế cũng bao gồm cấu trúc biểu hiện có khả năng biểu hiện polypeptit có chứa ít nhất 80% axit amin liên tiếp như nêu trong SEQ ID NO. 3. Tốt hơn là, cấu trúc biểu hiện đã cài xen ADN hoặc cADN bằng các nucleotit liên tiếp như nêu trong SEQ ID NO. 2.

Mục đích khác của sáng chế bộc lộ cấu trúc gen tái tổ hợp có chứa khuôn polynucleotit có trình tự nucleotit nêu trong SEQ ID NO. 2, trong đó khuôn polynucleotit có thể biểu hiện ở tế bào chủ để tạo ra enzym mà làm thoái biến pectin. Tốt hơn là, cấu trúc gen tái tổ hợp còn bao gồm vùng trình tự khởi đầu được liên kết có điều khiển để tăng cường sự biểu hiện của khuôn polynucleotit.

Một trong số các phương án được ưu tiên của sáng chế là chủng ms6 của *M. phaseolina* được phân lập từ cây đay bị nhiễm. Polypeptit được phân lập cũng tốt hơn là có nguồn gốc từ chủng này.

Ngoài ra, mục đích của sáng chế là đề xuất phương pháp tiềm năng khả thi về mặt thương mại để phân lập enzym làm thoái biến pectin từ *M. phaseolina* để bắt kịp

nhu cầu ngày càng tăng đối với, trong số những thứ khác, việc sản xuất nước ép trái cây, sản phẩm dệt, bột giấy và giấy, cà phê, trà và chất chiết dầu.

Mục đích khác của sáng chế là sử dụng chất làm thoái biến pectin trong việc xử lý thức ăn gia súc, nước ép trái cây, sản phẩm dệt, bột giấy và giấy, cà phê, trà và chất chiết dầu, giầm/khử keo của sợi vỏ.

Ứng dụng bất kỳ hoặc tất cả các ứng dụng có khả năng được phát triển thành bộ kit để thương mại hóa dưới dạng sản phẩm nghiên cứu hoặc dưới hàng cung cấp cho công nghiệp và các ứng dụng khác. Bộ kit có thể có chứa polynucleotit và/hoặc polypeptit tương ứng với một hoặc nhiều gen *M. phaseolina* theo sáng chế, kháng thể, và/hoặc chất phản ứng khác.

Theo khía cạnh thứ nhất, mục đích của sáng chế là cung cấp trình tự polynucleotit được phân lập có độ tương đồng ít nhất là 80% so với trình tự nucleotit được nêu trong SEQ ID NO. 1 hoặc 2, hoặc dạng kết hợp bất kỳ của chúng hoặc trình tự bổ sung của chúng, trong đó polynucleotit nói trên mã hóa pectat lyaza.

Theo khía cạnh thứ hai, mục đích của sáng chế là cung cấp đoạn dò axit nucleic được phân lập mà sự lai hóa dưới điều kiện nghiêm ngặt trung bình với trình tự polynucleotit được phân lập bao gồm trình tự nuleotit được nêu trong SEQ ID NO. 1 hoặc 2, hoặc dạng kết hợp bất kỳ của chúng hoặc trình tự bổ sung của chúng, trong đó polynucleotit nói trên mã hóa pectat lyaza, và trong đó điều kiện nghiêm ngặt trung bình bao gồm sự lai hóa với ADN liên kết với bộ lọc ở 6x natri clorua/natri xitrat (SSC) ở khoảng 45°C sau đó là một hoặc nhiều lần rửa trong 0,2xSSC/0,1% SDS ở khoảng từ 50 đến 65°C.

Theo khía cạnh thứ ba, mục đích của sáng chế là cung cấp phân tử nucleic chứa trình tự polynucleotit theo khía cạnh thứ nhất liên tục bằng cách dừng codon trong một trình tự mã hóa mà mã hóa protein hoặc peptit khác loại.

Theo khía cạnh thứ tư, mục đích của sáng chế là cung cấp vật truyền tái tổ hợp chứa trình tự polynucleotit được phân lập theo khía cạnh thứ nhất.

Theo khía cạnh thứ năm, mục đích của sáng chế này là cung cấp cấu trúc biểu hiện chứa trình tự polynucleotit được phân lập theo khía cạnh thứ nhất, trong đó trình tự polynucleotit kết hợp có điều khiển với một trình tự nucleotit điều hòa chứa tín hiệu

điều hòa phiên mã hoặc dịch mã, hoặc cả hai, điều khiển sự biểu hiện của trình tự nucleotit trong tế bào chủ.

Theo khía cạnh thứ sáu, mục đích của sáng chế là cung cấp tế bào chủ được xử lý di truyền chứa trình tự polynucleotit được phân lập theo khía cạnh thứ nhất.

Theo khía cạnh thứ bảy, mục đích của sáng chế này là cung cấp phương pháp tạo polynucleotit gồm có các bước:

i) nuôi cấy tế bào được biến nạp với một cấu trúc biểu hiện theo khía cạnh thứ năm ở điều kiện có hiệu quả để tạo ra polypeptit; và

ii) phân lập polypeptit.

Theo khía cạnh thứ tám, mục đích của sáng chế này là cung cấp polypeptit được phân lập gồm một trình tự axit amin có độ tương đồng ít nhất là 80% so với trình tự axit amin được chọn từ nhóm SEQ ID NO. 3 hoặc dạng kết hợp bất kỳ của chúng, và thể hiện hoạt tính xúc tác của pectat lyaza.

Theo khía cạnh thứ chín, mục đích của sáng chế là cung cấp protein khảm gồm có polypeptit theo khía cạnh thứ tám được dung hợp qua liên kết cộng hóa trị với trình tự axit amin của polypeptit thứ hai.

Theo khía cạnh thứ mười, mục đích của sáng chế là cung cấp một tế bào chủ, bao gồm cấu trúc biểu hiện theo khía cạnh thứ năm.

Theo khía cạnh mười một, mục đích của sáng chế là cung cấp nấm chuyển gen của *M. phaseolina*.

Người có hiểu biết trung bình trong lĩnh vực sẽ dễ dàng nhận ra rằng sáng chế được làm cho thích hợp để thực hiện các mục đích và đạt được các kết quả và ích lợi đã nêu, cũng như là các mục đích vốn có trong đó. Các phương án được mô tả trong bản mô tả này không nhằm làm giới hạn hoặc hạt chế phạm vi của sáng chế.

Các đặc điểm, khía cạnh và ưu điểm này và các đặc điểm, khía cạnh và ưu điểm theo sáng chế sẽ trở nên dễ hiểu hơn khi tham khảo phần mô tả và các yêu cầu bảo hộ dưới đây.

Mô tả vắn tắt các hình vẽ

Các hình vẽ kèm theo, mà được kết hợp vào và tạo thành một phần của bản mô tả, minh họa các phương án của sáng chế và cùng với phần mô tả, để giải thích nguyên lý của sáng chế.

Fig. 1 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 1 và đường M là thang khối lượng phân tử ADN.

Fig. 2 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 4 và đường M là thang khối lượng phân tử ADN.

Fig. 3 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 7 và đường M là thang khối lượng phân tử ADN.

Fig. 4 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 10 và đường M là thang khối lượng phân tử ADN.

Fig. 5 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 13 và đường M là thang khối lượng phân tử ADN.

Fig. 6 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 16 và đường M là thang khối lượng phân tử ADN.

Fig. 7 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 19 và đường M là thang khối lượng phân tử ADN.

Fig. 8 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza của SEQ ID NO. 22 và đường M là thang khối lượng phân tử ADN.

Fig. 9 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza C của SEQ ID NO. 28 và đường M là thang khối lượng phân tử ADN.

Fig. 10 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza C của SEQ ID NO. 31 và đường M là thang khối lượng phân tử ADN.

Fig. 11 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza C của SEQ ID NO. 34 và đường M là thang khối lượng phân tử ADN.

Fig. 12 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza C của SEQ ID NO. 37 và đường M là thang khối lượng phân tử ADN.

Fig. 13 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza C của SEQ ID NO. 40 và đường M là thang khối lượng phân tử ADN.

Fig. 14 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectat lyaza C của SEQ ID NO. 43 và đường M là thang khối lượng phân tử ADN.

Fig. 15 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectinesteraza của SEQ ID NO. 46 và đường M là thang khối lượng phân tử ADN.

Fig. 16 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectinesteraza của SEQ ID NO. 49 và đường M là thang khối lượng phân tử ADN.

Fig. 17 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của pectinesteraza của SEQ ID NO. 55 và đường M là thang khối lượng phân tử ADN.

Fig. 18 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của rhamnogalacturonaza của SEQ ID NO. 58 và đường M là thang khối lượng phân tử ADN.

Fig. 19 là ảnh điện di trên gel agaroza thể hiện kết quả khuếch đại PCR trong đó đường 1 là polynucleotit của rhamnogalacturonaza của SEQ ID NO. 61 và đường M là thang khối lượng phân tử ADN.

Mô tả chi tiết sáng chế

Các định nghĩa và/hoặc phương pháp được nêu trong bản mô tả này xác định sáng chế và hướng dẫn người có hiểu biết trung bình trong lĩnh vực thực hiện sáng chế. Trừ khi có chỉ dẫn khác, các thuật ngữ cần được hiểu theo cách dùng thông thường của người có hiểu biết trung bình trong lĩnh vực thích hợp. Trong trường hợp mà định nghĩa và/hoặc phương pháp bất kỳ được phát hiện là không thống nhất với định nghĩa và/hoặc phương pháp bất kỳ nêu trong tài liệu tham khảo sáng chế hoặc không phải sáng chế bất kỳ được viện dẫn trong bản mô tả này hoặc trong tài liệu tham khảo bất kỳ được tìm thấy ở nơi khác, cần hiểu rằng định nghĩa và/hoặc phương pháp đã được đề cập mà được cung cấp/được thừa nhận rõ ràng trong đơn này sẽ được sử dụng trong bản mô tả này. Các thuật ngữ số ít bao gồm cả nghĩa số nhiều trừ khi nội dung chỉ ra rõ ràng theo cách khác. Tương tự, từ "hoặc" được dự định là bao gồm cả "và" trừ khi nội dung chỉ ra rõ ràng theo cách khác. Do đó "gồm có A hoặc B" nghĩa là bao gồm A, hoặc B, hoặc A và B. Cần hiểu thêm rằng tất cả các giá trị kích thước bazơ hoặc kích thước axit amin, và tất cả trọng lượng phân tử hoặc khối lượng phân tử, được nêu đối với axit nucleic hoặc polypeptit là xấp xỉ, và được nêu để mô tả. Mặc dù phương pháp và nguyên liệu tương tự hoặc tương đương với phương pháp và nguyên liệu được mô tả trong bản mô tả này có thể được sử dụng trong việc thực hiện hoặc thử nghiệm sáng chế, phương pháp và nguyên liệu thích hợp được mô tả dưới đây.

Sáng chế đề xuất trình tự nucleotit của gen *M. phaseolina* tham gia vào sự thoái biến pectin. Gen mã hóa cho protein có hoạt tính enzym làm thoái biến pectin mà để sử dụng trong ngành công nghiệp bao gồm thực phẩm, thức ăn gia súc, giấy, chất chiết dầu, dệt, nước thải hoặc được quan tâm trong công nghiệp. Được mô tả trong bản mô tả này dưới đây là gen theo sáng chế và gen khác của việc bộc lộ, việc xác định, xác

định đặc điểm, cải biến, và phương pháp sử dụng chúng trong các quy trình công nghiệp khác nhau.

Trình tự nucleotit của ADN hệ gen *M. phaseolina* được thu lấy bằng cách đọc trình tự ADN bằng súng bắn gen ngẫu nhiên hệ gen hoàn chỉnh. ADN hệ gen được điều chế từ sự phân lập *M. phaseolina* ms6 được phân lập từ cây đay bị nhiễm (*Corchorus* spp.). Trình tự nucleotit đã được tạo ra được lắp ráp để tạo thành giá treo và giá đỡ bằng thiết bị lắp ráp Newbler. Trình tự nucleotit được chú thích ban đầu bằng chương trình phần mềm, chẳng hạn như Augustus, Glimmer M (The Institute of Genome Research, Rockville, Md.) và Evidence Modeler (EVM), mà có thể xác định các vùng mã hóa giả định, các intron, và chỗ ghép nối. Sự lựa chọn tự động và bằng tay khác nữa của trình tự nucleotit được thực hiện để tinh chế và thiết lập đặc điểm chính xác của các vùng mã hóa và các đặc điểm gen khác.

Trình tự hệ gen theo sáng chế mà mã hóa cho enzym làm thoái biến pectin được xác định trước hết bằng cách so sánh trình tự nucleotit của ADN hệ gen *M. phaseolina* và trình tự nucleotit của gen enzym đã biết của vi sinh vật khác. Trình tự nucleotit của các gen *M. phaseolina* này, khung đọc mở, vị trí của các exon và các intron, cấu trúc của enzym, và sự hữu ích tiềm năng của chúng trong các ngành công nghiệp khác nhau, chẳng hạn như các enzym tham gia vào việc sản xuất thực phẩm và thức ăn chăn nuôi, đồ uống, vải dệt và thuốc tẩy, chưa được biết đến.

Trên 14000 cADN từ *M. phaseolina* được đọc trình tự một phần hoặc toàn bộ. Trong số chúng hai mươi mốt cADN mã hóa cho enzym mới có vai trò giả định trong sự thoái biến pectin được phát hiện.

Khung đọc mở (ORF) được phân tích bằng cách đọc trình tự hoàn toàn hoặc một phần của các dòng của thư viện cADN có nguồn gốc từ mARN *M. phaseolina* và được phân tích thêm bằng cách sử dụng phần mềm phân tích trình tự, và bằng cách xác định độ tương đồng với các trình tự đã biết trong cơ sở dữ liệu (công cộng/tư nhân).

Trong bản mô tả này, một số thuật ngữ được sử dụng trong toàn bộ bản mô tả có nghĩa như được chỉ ra trừ khi được chỉ ra rõ ràng là có nghĩa khác.

Thuật ngữ "gen", như được dùng trong bản mô tả này, được xác định là trình tự hệ gen của nấm *M. phaseolina*, cụ thể là trình tự polynucleotit mã hóa cho polypeptit

của một loạt các enzym. Thuật ngữ này có thể còn bao gồm phân tử axit nucleic có chứa trình tự nucleotit ngược chiều, xuôi chiều, và/hoặc intron.

Thuật ngữ "khung đọc mở (ORF)" nghĩa là một loạt các bộ ba nucleotit mã hóa cho axit amin mà không có codon kết thúc bất kỳ và trình tự bộ ba này có thể dịch mã được thành protein bằng cách sử dụng thông tin sử dụng codon thích hợp đối với sinh vật cụ thể.

"Trình tự mã hóa" hoặc "vùng mã hóa" dùng để chỉ phân tử axit nucleic có thông tin trình tự cần để tạo ra sản phẩm gen, chẳng hạn như axit amin hoặc polypeptit, khi trình tự này được biểu hiện. Trình tự mã hóa có thể bao gồm các trình tự không được dịch mã (ví dụ, các intron hoặc các vùng không được dịch mã 5' hoặc 3') ở trong các vùng được dịch mã, hoặc có thể thiếu các trình tự không được dịch mã xen kẽ này (ví dụ, như trong cADN).

Như được dùng trong bản mô tả này, "polynucleotit" là trình tự nucleotit chẳng hạn như mảnh axit nucleic. Polynucleotit có thể là polyme của ARN hoặc ADN tức là mạch đơn hoặc mạch kép, mà tùy ý có chứa bazơ nucleotit tổng hợp, không tự nhiên hoặc được thay đổi. Polynucleotit ở dạng polyme của ADN có thể được bao gồm một hoặc nhiều đoạn của cADN, ADN hệ gen, ADN tổng hợp, hoặc hỗn hợp của chúng. Polynucleotit được phân lập theo sáng chế cũng có thể là có nguồn gốc từ SEQ ID NO. 1, hoặc phần bổ sung của các trình tự này. Polynucleotit được phân lập theo việc bộc lộ cũng có thể có nguồn gốc từ SEQ ID NO. 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58, 61, 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59 và 62, hoặc phần bổ sung của các trình tự này.

"Được phân lập" nghĩa là được làm thay đổi "bởi bàn tay con người" từ tình trạng tự nhiên. Nếu hợp phần hoặc chất có trong tự nhiên, nó "được phân lập" nếu nó đã được thay đổi hoặc được lấy ra từ môi trường ban đầu của nó, hoặc cả hai. Ví dụ, polynucleotit hoặc polypeptit có mặt tự nhiên ở thực vật hoặc động vật sống không "được phân lập" nhưng cùng polynucleotit hoặc polypeptit được tách ra từ nguyên liệu đồng tồn tại của tình trạng tự nhiên của nó "được phân lập", là thuật ngữ được sử dụng trong bản mô tả này.

Thuật ngữ "tái tổ hợp", khi dùng trong bản mô tả này dùng để chỉ polypeptit hoặc protein, có nghĩa là polypeptit hoặc protein đó có nguồn gốc từ hệ thống biểu hiện tái tổ hợp (ví dụ, vi khuẩn hoặc động vật có vú). Polypeptit hoặc protein "vi khuẩn" dùng để chỉ polypeptit hoặc protein tái tổ hợp được tạo ra ở hệ thống biểu hiện vi khuẩn hoặc nấm. Polypeptit hoặc protein được biểu hiện ở hầu hết các hệ thống vi khuẩn, ví dụ, *Escherichia coli,* sẽ không có các cải biến glycosyl hóa; polypeptit hoặc protein được biểu hiện ở nấm sẽ được glycosyl hóa.

"Vật truyền" dùng để chỉ plasmit, phagơ, cosmit, nấm men, hoặc virut, trình tự sao chép nhân tạo (ARS) hoặc nhiễm sắc thể nhân tạo để biểu hiện polypeptit từ trình tự nucleotit. Thuật ngữ "vật truyền" cũng nhằm dùng để chỉ phân tử axit nucleic có khả năng vận chuyển axit nucleic khác vào nơi mà nó được liên kết. Một dạng của vật truyền là "plasmit," mà dùng để chỉ thòng lọng ADN mạch kép dạng vòng mà đoạn ADN bổ sung có thể được nối vào đó. Dạng khác của vật truyền là vật truyền virut, trong đó đoạn ADN bổ sung có thể được nối vào hệ gen virut. Các vật truyền nhất định có khả năng tự sao chép ở tế bào chủ mà chúng được đưa vào (ví dụ, vật truyền vi khuẩn có vị trí khởi đầu sao chép vi khuẩn và vật truyền động vật có vú thể kết hợp). Vật truyền khác có thể được hợp nhất vào hệ gen của tế bào chủ khi đưa vào trong tế bào chủ, và bằng cách đó được sao chép cùng với hệ gen vật chủ. Ngoài ra, vật truyền nhất định có khả năng định hướng sự biểu hiện của gen mà chúng được liên kết có điều khiển với. Vật truyền này trong bản mô tả này được đề cập đến là "vật truyền biểu hiện tái tổ hợp" (hoặc đơn giản là, "vật truyền biểu hiện"). Nhìn chung, vật truyền biểu hiện để sử dụng trong kỹ thuật ADN tái tổ hợp thường ở dạng plasmit. Trong bản mô tả này, "plasmit" và "vật truyền" có thể được dùng thay thế lẫn nhau vì plasmit là dạng thường dùng nhất của vật truyền. Tuy nhiên, sáng chế được dự định là bao gồm các dạng khác của vật truyền biểu hiện, chẳng hạn như vật truyền virut (ví dụ, thiếu khả năng tự sao chép retrovirut, adenovirut và virut được kết hợp ađeno), mà có chức năng tương đương.

Thuật ngữ "vật truyền biểu hiện" được xác định trong bản mô tả này là phân tử ADN dạng thẳng hoặc dạng vòng mà bao gồm polynucleotit mã hóa cho polypeptit theo sáng chế, mà được liên kết có điều khiển với nucleotit khác mà mang lại sự biểu hiện.

Thuật ngữ "cấu trúc biểu hiện" có thể bao gồm tổ hợp của (các) yếu tố di truyền có vai trò điều khiển trong sự biểu hiện gen, ví dụ, trình tự khởi đầu hoặc trình tự tăng cường, hoặc trình tự mã hóa mà được phiên mã thành ARN, mARN và được dịch mã thành protein, và mà được liên kết có điều khiển với trình tự khởi đầu hoặc các trình tự mở đầu và kết thúc phiên mã thích hợp.

Thuật ngữ "được liên kết có điều khiển" trong bản mô tả này có nghĩa là cấu hình trong đó trình tự điều khiển được đặt ở vị trí thích hợp so với trình tự mã hóa của trình tự polynucleotit sao cho trình tự điều khiển điều khiển sự biểu hiện của trình tự mã hóa của polypeptit.

Thuật ngữ "tế bào chủ", như được dùng trong bản mô tả này, bao gồm dạng tế bào bất kỳ mà dễ bị biến nạp, chuyển nạp, tải nạp, và dạng tương tự bằng cấu trúc axit nucleic hoặc vật truyền biểu hiện có chứa polynucleotit theo sáng chế.

Thuật ngữ "tế bào chủ tái tổ hợp" thường có nghĩa là tế bào được nuôi cấy mà có chứa đơn vị phiên mã tái tổ hợp, và sẽ biểu hiện polypeptit hoặc protein khác loại, và ARN được mã hóa bởi đoạn ADN hoặc gen tổng hợp trong đơn vị phiên mã tái tổ hợp. Tế bào có thể là tế bào nhân sơ hoặc tế bào nhân chuẩn.

"Polypeptit" như được dùng trong bản mô tả này, là mạch đơn thẳng của axit amin được liên kết cùng nhau bằng liên kết peptit, và có trình tự có chiều dài lớn hơn 100 axit amin.

Thuật ngữ "trình tự khởi đầu" như được dùng trong bản mô tả này, dùng để chỉ trình tự axit nucleic mà có chức năng điều khiển sự phiên mã của gen xuôi chiều. Trình tự khởi đầu thường thích hợp đối với tế bào chủ trong đó gen đích sẽ được biểu hiện. Trình tự khởi đầu cùng với trình tự axit nucleic điều hòa phiên mã và dịch mã khác (cũng được gọi là "trình tự điều khiển") là cần thiết để biểu hiện gen nhất định. Nhìn chung, trình tự điều hòa phiên mã và dịch mã bao gồm, nhưng không giới hạn ở, trình tự khởi đầu, vị trí liên kết ribosom, trình tự mở đầu và kết thúc phiên mã, trình tự mở đầu và kết thúc dịch mã, và trình tự tăng cường hoặc trình tự kích hoạt.

Thuật ngữ "*in-vitro*" như được dùng trong bản mô tả này, thường dùng để chỉ phản ứng sinh học mà xảy ra trong môi trường nhân tạo bên ngoài sinh vật sống, mà thường được thực hiện trong phòng thí nghiệm bằng cách sử dụng các thành phần của

sinh vật mà đã được phân lập từ hoàn cảnh sinh học thông thường của chúng để cho phép thực hiên phân tích chi tiết hơn và thuận tiện hơn.

Thuật ngữ "độ tương đồng theo %" được sử dụng thay thế lẫn nhau trong bản mô tả này với thuật ngữ "độ đồng nhất theo %" trong bản mô tả này và thường dùng để chỉ mức độ của trình tự axit nucleic hoặc axit amin độ đồng nhất giữa trình tự axit nucleic mà mã hóa cho bất kỳ một trong số polypeptit theo sáng chế hoặc trình tự axit amin của polypeptit theo sáng chế, khi được sắp thẳng hàng bằng cách sử dụng chương trình sắp thẳng hàng trình tự.

Ví dụ, như được dùng trong bản mô tả này, độ tương đồng 80% có nghĩa giống như là độ đồng nhất trình tự 80% được xác định bằng thuật toán xác định, và theo đó thể tương đồng của trình tự nhất định có độ đồng nhất trình tự lớn hơn 80% trên chiều dài của trình tự nhất định. Mức độ ví dụ của độ đồng nhất trình tự bao gồm, nhưng không giới hạn ở, độ đồng nhất trình tự ở 80, 85, 90, 95, 98% hoặc lớn hơn so với trình tự nhất định, *ví dụ*, trình tự mã hóa đối với polypeptit bất kỳ trong số các polypeptit theo sáng chế, như được mô tả trong bản mô tả này.

Chương trình máy tính ví dụ mà có thể được sử dụng để xác định độ đồng nhất giữa hai trình tự bao gồm, nhưng không giới hạn ở, bộ chương trình BLAST, *ví dụ*, BLASTN, BLASTX, và TBLASTX, BLASTP và TBLASTN, có thể truy cập công cộng tại www.ncbi.nlm.nih.gov/BLAST.

Tra cứu trình tự thường được thực hiện bằng cách sử dụng chương trình BLASTN khi đánh giá trình tự axit nucleic nhất định so với trình tự axit nucleic trong trình tự ADN ngân hàng gen và cơ sở dữ liệu công cộng khác. Chương trình BLASTX được ưu tiên để tra cứu trình tự axit nucleic mà đã được dịch mã trong tất cả các khung đọc mở đối với trình tự axit amin trong trình tự protein ngân hàng gen và cơ sở dữ liệu công cộng khác.

Sự sắp thẳng hàng được ưu tiên của trình tự được chọn để xác định "độ đồng nhất theo %" giữa hai hoặc nhiều trình tự được tạo thành bằng cách sử dụng, ví dụ, chương trình CLUSTAL-W.

Thuật ngữ “đoạn mồi” như được dùng trong bản mô tả này, là oligonucleotit có khả năng liên kết với trình tự axit nucleic đích và làm mồi cho sự tổng hợp axit

nucleic. Oligonucleotit khuếch đại như được xác định trong bản mô tả này tốt hơn là có chiều dài từ 10 đến 50, tốt nhất là từ 15 đến 25 nucletit. Oligonucleotit khuếch đại theo sáng chế có thể được tổng hợp hóa học.

Chữ viết tắt trong toàn bộ bản mô tả dùng để chỉ axit nucleic có chứa trình tự nucleotit là chữ viết tắt một ký tự thông thường. Do đó, khi được bao gồm trong axit nucleic, nucleotit mã hóa có trong tự nhiên được viết tắt như sau: ađenin (A), guanin (G), xytosin (C), tymin (T) và uraxin (U). Ngoài ra, trừ khi được xác định theo cách khác, trình tự axit nucleic được thể hiện trong bản mô tả này là theo chiều 5'→3'.

Như được dùng trong bản mô tả này, thuật ngữ "bổ sung" và từ phát sinh của nó được sử dụng để chỉ sự bắt cặp của axit nucleic theo quy tắc đã biết là A bắt cặp với T hoặc U và C bắt cặp với G. Sự bổ sung có thể là "một phần" hoặc "hoàn toàn". Trong sự bổ sung một phần, chỉ một số bazơ axit nucleic được bắt cặp theo quy tắc bắt cặp bazơ; trong khi đó trong sự bổ sung hoàn toàn hoặc toàn bộ, tất cả các bazơ được bắt cặp theo quy tắc bắt cặp. Mức độ bổ sung giữa các sợi axit nucleic có thể có ảnh hưởng đáng kể lên hiệu quả và độ mạnh của sự lai hóa giữa các sợi axit nucleic như đã biết trong lĩnh vực. Điều này có thể là cụ thể trong phương pháp phát hiện mà phụ thuộc vào sự liên kết giữa các axit nucleic.

Trình tự ADN theo sáng chế được tạo ra bằng phản ứng đọc trình tự và có thể có chứa sai sót nhỏ mà có thể tồn tại dưới dạng nucleotit được nhận dạng nhầm, sự cài xen, và/hoặc sự làm khuyết. Tuy nhiên, sai sót nhỏ này, nếu có, sẽ không làm nhiễu loạn việc xác định trình tự là gen của *M. phaseolina* mà mã hóa cho enzym được quan tâm ở quy mô công nghiệp, và được bao gồm cụ thể trong phạm vi của sáng chế.

Sáng chế bao gồm trình tự nucleotit hệ gen và trình tự mã hóa của gen mà mã hóa cho enzym của *M. phaseolina* được quan tâm ở quy mô công nghiệp. Theo đó, theo một phương án, sáng chế đề xuất SEQ ID NO. 2 xác định trình tự nucleotit của khung đọc mở (ORF) của gen được xác định. Theo một phương án khác, trình tự hệ gen của gen được xác định bằng SEQ ID NO. 1 được đề xuất. Theo việc bộc lộ, SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59 và 62, và/hoặc hỗn hợp/dạng kết hợp bất kỳ của chúng, mỗi SEQ ID này xác định trình tự nucleotit của khung đọc mở (ORF) của gen được xác định. Sáng chế đề xuất trình tự

hệ gen của gen được xác định bởi SEQ ID NO. 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 55, 58 và 61, và/hoặc hỗn hợp bất kỳ của chúng.

Như được dùng trong bản mô tả này, "gen" dùng để chỉ (i) gen có chứa và/hoặc gồm có ít nhất một trong số các trình tự nucleotit và/hoặc mảnh của chúng mà được nêu trong SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61; (ii) trình tự nucleotit bất kỳ hoặc mảnh của chúng mà mã hóa cho trình tự axit amin mà được nêu trong SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63; (iii) trình tự nucleotit bất kỳ mà lai hóa với phần bổ sung của trình tự nucleotit nêu trong SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61 trong điều kiện nghiêm ngặt trung bình, ví dụ, sự lai hóa với ADN liên kết với bộ lọc ở lượng có hiệu quả của 6x natri clorua/natri xitrat (SSC) ở nhiệt độ có hiệu quả, chẳng hạn như 45°C (xấp xỉ) sau đó là một hoặc nhiều lần rửa ở lượng có hiệu quả của SDS, chẳng hạn như 0,2x SSC/0,1% SDS, ở nhiệt độ có hiệu quả, chẳng hạn như từ 50 đến 65°C (xấp xỉ), hoặc trong điều kiện nghiêm ngặt cao, ví dụ, sự lai hóa với axit nucleic liên kết với bộ lọc ở lượng có hiệu quả của SSC, chẳng hạn như 6x SSC, ở nhiệt độ có hiệu quả, chẳng hạn như 45°C (xấp xỉ), sau đó là một hoặc nhiều lần rửa ở lượng có hiệu quả của SDS, chẳng hạn như 0,1x SSC/0,2% SDS, ở nhiệt độ có hiệu quả, chẳng hạn như 68°C (xấp xỉ), hoặc trong điều kiện lai hóa khác mà rõ ràng đối với người có hiểu biết trung bình trong lĩnh vực (xem tài liệu, ví dụ, Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. Current Protocols in Molecular Biology, 1994; Green Publishing Associates, Inc. and John Wiley & Sons, Inc., New York). Tốt hơn là, polynucleotit lai hóa với phần bổ sung của trình tự ADN được bộc lộ trong bản mô tả này mã hóa cho sản phẩm gen, ví dụ, sản phẩm gen mà tương đương về mặt chức năng với sản phẩm gen được mã hóa bởi một trong số các gen enzym hoặc mảnh của chúng.

Như được mô tả ở trên, trình tự gen không chỉ bao gồm trình tự nucleotit thoái hóa mà mã hóa cho trình tự axit amin của SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63, mà còn bao gồm trình tự nucleotit thoái hóa mà khi được dịch mã ở sinh vật không phải là *M. phaseolina,* sẽ tạo ra polypeptit

gồm có một trong số các trình tự axit amin của SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63, hoặc mảnh của chúng. Người có hiểu biết trung bình trong lĩnh vực sẽ biết làm thế nào để chọn lọc codon thích hợp hoặc cải biến trình tự nucleotit của SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61, khi sử dụng trình tự gen ở *M. phaseolina hoặc* ở sinh vật khác. Ví dụ, ở *Candida albicans,* codon CTG mã hóa cho gốc serin thay vì gốc leuxin.

Trình tự nucleotit theo sáng chế có thể được sử dụng làm gen đánh dấu di truyền và/hoặc gen đánh dấu trình tự để hỗ trợ cho sự phát triển của bản đồ di truyền, vật chất, hoặc trình tự của hệ gen *M. phaseolina*. Trình tự nucleotit và sản phẩm gen tương ứng theo sáng chế cũng có thể được dùng để phát hiện sự có mặt của *M. phaseolina.* Phương pháp dựa trên sự lai hóa và kháng thể đã biết trong lĩnh vực có thể được sử dụng để xác định sự có mặt và nồng độ của trình tự nucleotit và sản phẩm gen tương ứng theo sáng chế.

Trình tự nucleotit cũng có thể được dùng để xác định chất ức chế của enzym mà có thể có tác dụng trị liệu, trong bối cảnh enzym có thể đóng vai trò trong sự xâm nhập của vật chủ trong quá trình nhiễm.

Theo phương án khác, ngoài trình tự nucleotit của *M. phaseolina* được mô tả ở trên, thể tương đồng hoặc dạng trực giao của gen theo sáng chế mà có thể có mặt ở *M. phaseolina* và các loài nấm khác cũng được bao gồm. Được đặc biệt ưu tiên là thể tương đồng hoặc dạng trực giao của nấm sợi. Gen enzym có thể được xác định và được phân lập bằng các kỹ thuật sinh học phân tử đã biết trong lĩnh vực.

Thuật ngữ "nấm" như được dùng trong bản mô tả này bao gồm các ngành Ascomycota, Basidiomycota, Chytridiomycota, và Zygomycota (Hawksworth DL, Kirk PM, Sutton BC, Pegler DN. Ainsworth and Bisby's Dictionary of the Fungi (8th Ed.). 1995; CAB International, Wallingford, United Kingdom. 616p) và nấm men. Các nhóm điển hình của Ascomycota bao gồm, ví dụ, *Neurospora, Penicillium, Aspergillus*. Các nhóm điển hình của Basidiomycota bao gồm nấm mốc, nấm gỉ sắt, và nấm muội than. Các nhóm điển hình của Chytridiomycota bao gồm Allomyces, Blastocladiella, Coelomomyces. Các nhóm điển hình của Zygomycota bao gồm, ví dụ, Rhizopus và Mucor.

Thuật ngữ "nấm sợi" bao gồm tất cả các dạng sợi của nấm. Nấm sợi đặc trưng ở thể sợi nấm dinh dưỡng tạo thành bởi kitin, xenluloza, glucan, chitosan, mannan, và các polysacarit phức hợp khác. Sự sinh trưởng dinh dưỡng được thực hiện bằng cách kéo dài sợi nấm và sự dị hóa cacbon là hiếu khí bắt buộc.

Theo đó, sáng chế cũng đề xuất trình tự nucleotit của nấm mà có thể lai hóa với polynucleotit của gen. Sáng chế bao gồm axit nucleic được phân lập có chứa và/hoặc gồm có trình tự nucleotit mà có độ tương đồng ít nhất là 50% so với trình tự nucleotit được chọn từ nhóm có chứa và/hoặc gồm có: SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61, và/hoặc hỗn hợp bất kỳ của chúng.

Sáng chế cũng bao gồm axit nucleic được phân lập có chứa trình tự nucleotit của nấm mà lai hóa trong điều kiện nghiêm ngặt trung bình với axit nucleic thứ hai mà có chứa và/hoặc gồm có trình tự nucleotit được chọn từ nhóm có chứa và/hoặc gồm có SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61, và/hoặc hỗn hợp bất kỳ của chúng.

Sáng chế cũng bao gồm axit nucleic được phân lập có chứa trình tự nucleotit của nấm mà mã hóa cho polypeptit mà trình tự axit amin của nó có độ tương đồng ít nhất là 50% so với trình tự axit amin được chọn từ nhóm có chứa và/hoặc gồm có SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63.

Trình tự nucleotit theo sáng chế còn bao gồm trình tự nucleotit của nấm mà có độ tương đồng ít nhất là 40% so với trình tự nucleotit nêu trong SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61.

Để phân lập gen tương đồng, trình tự gen *M. phaseolina* được mô tả ở trên có thể được gắn nhãn và được sử dụng để sàng lọc thư viện cADN được tạo cấu trúc từ mARN thu được từ sinh vật quan tâm, bao gồm nhưng không giới hạn ở *M. phaseolina.* Theo đó, sáng chế đề cập đến đoạn dò axit nucleic, tốt hơn được gắn nhãn có thể phát hiện được, gồm có trình tự nucleotit bất kỳ trong số các trình tự nucleotit của SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56,

59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61 được đưa vào sáng chế. Điều kiện lai hóa có thể tốt hơn là có độ nghiêm ngặt thấp hơn khi thư viện cADN có nguồn gốc từ sinh vật khác với loại sinh vật mà trình tự đã được gắn nhãn có nguồn gốc từ đó. Việc sàng lọc cADN cũng có thể xác định dòng có nguồn gốc từ các bản phiên mã được nối thay đổi nhau ở cùng loài. Theo cách khác, đoạn dò đã được gắn nhãn có thể được sử dụng để sàng lọc thư viện hệ gen có nguồn gốc từ sinh vật quan tâm, một lần nữa, bằng cách sử dụng điều kiện nghiêm ngặt thích hợp. Điều kiện nghiêm ngặt thấp là đã biết đối với người có hiểu biết trung bình trong lĩnh vực, và sẽ thay đổi theo cách có thể dự đoán trước tùy thuộc vào sinh vật cụ thể mà thư viện và trình tự đã được gắn nhãn có nguồn gốc từ đó. (Chi tiết xem trong tài liệu Sambrook J, Russell DW. Molecular Cloning, A Laboratory Manual, Third edition, 2001, Cold Spring Harbor Press, N.Y.; và Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K. Current Protocols in Molecular Biology, 1994;Green Publishing Associates, Inc. and John Wiley & Sons, Inc., New York).

Ngoài ra, trình tự gen tương đồng có thể được phân lập bằng cách thực hiện phản ứng chuỗi polymeraza (PCR) bằng cách sử dụng hai tổ hợp đoạn mồi oligonucleotit thoái hóa được thiết kế trên cơ sở trình tự axit amin ở trong gen quan tâm. Khuôn cho phản ứng có thể là cADN thu được bằng cách phiên mã ngược mARN được điều chế từ sinh vật quan tâm. Sản phẩm PCR có thể được tách dòng phụ và được đọc trình tự để đảm bảo rằng trình tự được khuếch đại đại diện cho các trình tự của trình tự gen enzym tương đồng.

Sau đó, mảnh PCR có thể được dùng để phân lập dòng cADN nguyên chiều dài bằng nhiều phương pháp đã biết đối với người có hiểu biết trung bình trong lĩnh vực. Theo cách khác, mảnh đã được gắn nhãn có thể được sử dụng để sàng lọc thư viện hệ gen.

Theo phương án khác theo sáng chế, trình tự gen *M. phaseolina* có thể được sử dụng trong việc phát triển enzym được cải biến hoặc mới mà thể hiện tính chất hóa học và/hoặc vật lý đặc biệt mong muốn. Vì sự tương tự rõ ràng của các trình tự axit amin trong số các enzym của *M. phaseolina* và nấm sợi khác, cấu trúc của enzym của nấm khác có thể được sử dụng để dự đoán cấu trúc của enzym *M. phaseolina*, và hỗ trợ

trong sự cải biến hợp lý của enzym *M. phaseolina* về các tính chất hữu dụng và chất lượng cao. Trình tự được đề xuất bởi sáng chế cũng có thể được dùng làm nguyên liệu điều chế cho sự cải biến hợp lý hoặc thiết kế của enzym mới có các tính chất làm cho enzym có thể thực hiện tốt hơn trong các quy trình đòi hỏi khắt khe.

Trình tự nucleotit của gen có thể được làm thay đổi bằng kỹ thuật gây đột biến ngẫu nhiên và định hướng vị trí hoặc kỹ thuật tiến hóa phân tử có định hướng, chẳng hạn như nhưng không giới hạn ở phương pháp được mô tả trong tài liệu (Arnold FH. Protein engineering for unusal environments. Curr. Opinion Biotechnol. 1993;4:450-455), sự gây đột biến định hướng oligonucleotit (Reidhaar-Olson JF, Sauer RT. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. Science, 1988; 241:53-57), sự gây đột biến hóa học (Eckert KA, Drinkwater NR. recA-dependent and recA-independent N-ethyl-N-nitrosourea mutagenesis at a plasmid-encoded herpes simplex virus thymidine kinase gene in *E. coli*. Mutat Res. 1987;178:1–10), sự gây đột biến định hướng vị trí (Kunkel TA. Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA, 1985;82:488-492; Oliphant A, Nussbaum AL, Struhl K. Cloning of random-sequence oligodeoxynucleotides. Gene 1986;44 177-183), PCR dễ có lỗi (Cadwell RC, Joyce GF. Randomization of genes by PCR mutagenesis. PCR Methods Appl. 1992;2:28–33), sự gây đột biến cát sét (Stauss Hj, Davies H, Sadovnikova E, Chain B, Horowitz N, Sinclair C. Induction of cytotoxic T lymphocytes with peptides in vitro: identification of candidate T-cell epitopes in human papilloma virus. PNAS 1992; 89(17): 7871-7875), phương pháp tái sắp xếp ADN như được mô tả trong tài liệu S Stemmer WP. DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. PNAS 1994;91(22):10747-10751 và trong các patent Mỹ số 5,605,793; 6,117,679; và 6,132,970, và phương pháp như được mô tả trong các patent Mỹ số 5,939,250, 5,965,408, 6,171,820. Đột biến ở trình tự nucleotit có thể được xác định bằng cách đọc trình tự gen trong các dòng này.

Theo một phương án, polynucleotit dài 747bp được minh họa trong SEQ ID NO. 2 là dòng cADN nguyên chiều dài mã hóa cho protein pectat lyaza biểu thị khung đọc mở mã hóa cho polypeptit có 248 axit amin, như trong SEQ ID NO. 3, có khối lượng phân tử tính được là khoảng 26kD. Thông qua phân tích SMART của SEQ ID

NO. 2, phát hiện ra sự có mặt của miền pectat lyaza Pfam trong trình tự này. Pectat lyaza là enzym ngoại bào mà xúc tác cho sự tách loại bỏ của pectat để tạo ra oligosacarit có các nhóm 4-deoxy-alpha-D-gluc-4-enuronosyl ở các đầu không khử của chúng. Enzym này tham gia vào việc làm thoái biến pectin.

Sáng chế cũng đề cập đến (a) vật truyền axit nucleic mà có chứa và/hoặc gồm có trình tự nucleotit có chứa trình tự bất kỳ trong số các trình tự nêu trên của gen và/hoặc trình tự bổ sung của chúng; (b) cấu trúc biểu hiện mà có chứa trình tự nucleotit chứa và/hoặc gồm có trình tự bất kỳ trong số các trình tự nêu trên mã hóa của gen được liên kết có điều khiển với yếu tố điều hòa mà điều khiển sự biểu hiện của trình tự mã hóa; và (c) tế bào chủ tái tổ hợp mà có chứa và/hoặc gồm có trình tự bất kỳ trong số các trình tự nêu trên của gen, bao gồm các vùng mã hóa được liên kết có điều khiển với yếu tố điều hòa mà điều khiển sự biểu hiện của trình tự mã hóa ở tế bào chủ.

Các kỹ thuật để cải biến trình tự polynucleotit bằng cách sử dụng phương pháp ADN tái tổ hợp là đã biết trong lĩnh vực. Các trình tự khác nhau có thể được nối bằng kỹ thuật đã biết, chẳng hạn như kỹ thuật giới hạn, ghép các vị trí giới hạn bổ sung và nối, tạo đầu cắt cụt bằng cách làm đầy phần nhô ra và nối đầu cắt cụt hoặc dạng tương tự. Đa cầu nối và phần nối có thể được sử dụng, khi thích hợp, và được đưa vào hoặc được loại bỏ bằng kỹ thuật đã biết để cho phép dễ dàng tổ hợp vật truyền và cấu trúc biểu hiện ADN. Số lượng lớn của các vật truyền có sẵn để tách dòng và thao tác di truyền. Thông thường, việc tách dòng có thể được thực hiện ở *E. coli.*

Theo phương án khác của sáng chế, vật truyền mà có chứa trình tự gen enzym theo sáng chế có thể còn bao gồm chức năng sao chép mà giúp cho vật truyền có thể truyền, duy trì và nhân lên ở một hoặc nhiều loài của tế bào chủ, bao gồm nhưng không giới hạn ở tế bào *E. coli*, tế bào nấm sợi, tế bào nấm men, và tế bào Bacillus. Vật truyền có thể có chứa phương tiện bất kỳ để đảm bảo sự tự sao chép. Theo cách khác, vật truyền có thể là loại mà, khi được đưa vào tế bào chủ, được hợp nhất vào hệ gen và được sao chép cùng với (các) nhiễm sắc thể mà nó đã được hợp nhất vào. Việc lựa chọn vật truyền thường phụ thuộc vào khả năng tương thích của vật truyền với tế bào chủ mà vật truyền được đưa vào.

Cấu trúc biểu hiện theo sáng chế có chứa và/hoặc gồm có trình tự khởi đầu, trình tự nucleotit mã hóa cho trình tự gen theo sáng chế, trình tự kết thúc phiên mã và

trình tự đánh dấu chọn lọc (tùy ý). Phương pháp bất kỳ đã biết trong lĩnh vực để đưa cấu trúc biểu hiện này vào tế bào chủ có thể được sử dụng. Tế bào nấm có thể được biến nạp bằng quy trình bao gồm bước tạo thành thể nguyên sinh, biến nạp thể nguyên sinh, và phục hồi thành tế bào theo cách đã biết về bản chất. Quy trình thích hợp để biến nạp tế bào chủ *Aspergillus* và *Trichoderma* được mô tả trong tài liệu EP 238 023 và Yelton MM, Hames JE, Timberlake WE. Sự biến nạp của *Aspergillus nidulans* bằng cách sử dụng plasmit trpC. PNAS,1984; 81(5):1470-1474. Phương pháp thích hợp để biến nạp các loài *Fusarium* được mô tả bởi Malardier L, Daboussi MJ, Julien J, Roussel F, Scazzocchio C, Brygoo Y. Cloning of the nitrate reductase gene (niaD) of *Aspergillus nidulans* and its use for transformation of *Fusarium oxysporum*. Gene 1989; 78:147-156, và WO 96/00787. Nấm men có thể được biến nạp bằng cách sử dụng quy trình được mô tả bởi Becker DM, Guarente L. High-efficiency transformation of *yeast* by electroporation. In: Abelson JN and Simon MI (eds), Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology,1991;194:182-187, Academic Press, Inc., New York; Ito H, Fukuda Y, Murata K, Kimura A. Transformation of intact yeast cells treated with alkali cations. Journal of Bacteriology, 1983;153: 163-168 and Hinnen A, Hicks JB, Fink GR. Transformation of yeast. PNAS, 1978; 75 (4):1929-1933.

Để ứng dụng công nghiệp, enzym theo sáng chế được sản xuất bằng tế bào nấm. Tốt hơn là, tế bào chủ biểu hiện là tế bào nấm sợi mà đã được sử dụng trong lên men công nghiệp quy mô lớn. Dòng tế bào hoặc hệ vật chủ thích hợp có thể được lựa chọn để đảm bảo sự cải biến chính xác và xử lý protein lạ được biểu hiện. Tốt hơn là, vật chủ biểu hiện được chọn mà có khả năng tiết hiệu quả protein nội sinh của chúng. Tế bào chủ cũng có thể được lựa chọn về sự thiếu hụt hoạt tính proteaza ngoại bào vì enzym được tiết ra có thể bị thoái biến trong môi trường nuôi cấy.

Sáng chế cũng đề cập đến phương pháp sản xuất polypeptit theo sáng chế, bao gồm các bước: (i) nuôi cấy tế bào, mà ở dạng dại của nó có khả năng sản xuất polypeptit, trong điều kiện có lợi cho sự sản xuất của polypeptit; và (ii) thu hồi polypeptit. Theo khía cạnh được ưu tiên, tế bào là *M. phaseolina*.

Sáng chế cũng đề cập đến phương pháp sản xuất polypeptit , bao gồm các bước: (i) nuôi cấy tế bào chủ trong điều kiện có lợi cho sự sản xuất của polypeptit, trong đó

tế bào chủ có chứa trình tự nucleotit của SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59, 62, 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 hoặc 61, trong đó trình tự nucleotit mã hóa cho polypeptit mà bao gồm hoặc gồm có hệ polypeptit trưởng thành của SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 hoặc 63, và (ii) thu hồi polypeptit.

Tế bào chủ biểu hiện hoặc thể biến nạp có thể được nuôi cấy trong môi trường dinh dưỡng thích hợp để sinh trưởng và biểu hiện protein bằng cách sử dụng phương pháp đã biết trong lĩnh vực. Ví dụ, tế bào có thể được nuôi cấy bằng cách nuôi cấy trong bình thót cổ có lắc, và lên men quy mô nhỏ hoặc quy mô lớn (bao gồm lên men liên tục, theo mẻ, theo mẻ có bổ sung cơ chất, hoặc trạng thái rắn) trong thiết bị lên men thí nghiệm hoặc công nghiệp được thực hiện trong môi trường thích hợp và trong điều kiện cho phép polypeptit được biểu hiện và/hoặc được phân lập. Sự nuôi cấy diễn ra trong môi trường dinh dưỡng thích hợp có chứa nguồn cacbon và nitơ và muối vô cơ, bằng cách sử dụng quy trình đã biết trong lĩnh vực (xem tài liệu: More Gene Manipulations in Fungi. Bennett J W, Lasure L., (eds). 1991; Academic Press, San Diego, CA). Nếu polypeptit được tiết vào môi trường dinh dưỡng, polypeptit có thể được thu hồi trực tiếp từ môi trường. Nếu polypeptit không được tiết vào môi trường, nó có thể được thu hồi từ dịch tan tế bào.

Polypeptit có thể được phát hiện bằng cách sử dụng phương pháp đã biết trong lĩnh vực mà đặc hiệu đối với polypeptit. Các phương pháp phát hiện này có thể bao gồm việc sử dụng kháng thể đặc hiệu, sự tạo thành sản phẩm enzym, hoặc sự biến mất của cơ chất enzym. Thử nghiệm enzym có thể được dùng để xác định hoạt tính của polypeptit.

Polypeptit được tạo ra có thể được thu hồi bằng cách sử dụng phương pháp đã biết trong lĩnh vực. Polypeptit có thể được thu hồi bằng các phương pháp khác nhau từ môi trường dinh dưỡng bằng quy trình thông thường bao gồm, nhưng không giới hạn ở, lọc, ly tâm, chiết, phun khô, làm bay hơi và kết tủa hoặc dạng kết hợp của chúng.

Polypeptit theo sáng chế có thể được tinh chế bằng nhiều quy trình đã biết trong lĩnh vực bao gồm, nhưng không giới hạn ở, phương pháp sắc ký (chẳng hạn như trao đổi ion, ái lực, kỵ nước, theo điểm đẳng điện, và loại trừ kích thước), quy trình điện di

(chẳng hạn như điện di điểm đẳng điện), khả năng hòa tan vi sai (chẳng hạn như kết tủa amoni sulfat), SDS-PAGE, hoặc chiết để thu được polypeptit về cơ bản tinh khiết (xem chi tiết trong tài liệu Protein Purification, Principles, High Resolution Methods and Applications. Janson JC, Rydén L. (eds(). 1989; VCH Publishers Inc., New York).

Sáng chế cũng đề cập tới sản phẩm gen (ví dụ, ARN hoặc protein) mà được mã hóa bởi trình tự gen nêu trong SEQ ID NO. 2, 5, 8, 11, 14, 17, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, 53, 56, 59 và 62. Sản phẩm gen enzym theo sáng chế cũng bao gồm ARN hoặc protein mà được mã hóa bởi trình tự hệ gen của gen như nêu trong SEQ ID NO. 1, 4, 7, 10, 13, 16, 19, 22, 25, 28, 31, 34, 37, 40, 43, 46, 49, 52, 55, 58 và 61. Enzym theo sáng chế bao gồm trình tự axit amin được chọn từ nhóm có chứa và/hoặc gồm có SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63.

Enzym theo sáng chế thể hiện ít nhất một trong số các hoạt tính của enzym được chọn từ nhóm có chứa và/hoặc gồm có pectat lyaza, pectat lyaza C, pectinesteraza và rhamnogalacturonaza. Sản phẩm gen enzym theo sáng chế có thể dễ dàng được tạo ra, ví dụ, bằng kỹ thuật tổng hợp hoặc bằng phương pháp theo công nghệ ADN tái tổ hợp bằng cách sử dụng các kỹ thuật đã biết trong lĩnh vực (xem tài liệu, Creighton TE. Proteins: Structures and Molecular Principles, 1983; W. H. Freeman and Co., N.Y.).

Phương pháp và hợp phần theo sáng chế cũng bao gồm protein và polypeptit mà tiêu biểu cho sản phẩm gen tương đương về mặt chức năng. Sản phẩm gen tương đương về mặt chức năng này bao gồm, nhưng không giới hạn ở, các biến thể tự nhiên của polypeptit có trình tự axit amin nêu trong SEQ ID NO. 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60 và 63.

Sản phẩm gen tương đương này có thể có chứa, ví dụ sự làm khuyết, sự thêm hoặc sự thế của các gốc axit amin ở trong trình tự axit amin được mã hóa bởi trình tự gen enzym được mô tả ở trên, nhưng dẫn đến sự thay đổi câm, do đó tạo ra sản phẩm tương đương về mặt chức năng. Sự thế axit amin có thể được tạo ra trên cơ sở sự tương tự về độ phân cực, điện tích, khả năng hòa tan, tính kỵ nước, tính ưa nước và/hoặc bản chất lưỡng tính của các gốc tham gia.

Các cải biến khác của sản phẩm trình tự gen mã hóa được mô tả ở trên có thể được tạo ra để tạo ra polypeptit mà phù hợp hơn, ví dụ, đối với sự biểu hiện, đối với quy mô lớn, v.v., trong tế bào chủ được chọn. Ví dụ, các gốc xystein có thể được làm khuyết hoặc được thế bằng axit amin khác để loại trừ cầu đisulfua.

Phương án khác của sáng chế còn bao gồm enzym theo sáng chế ở dạng rắn. Enzym ở dạng rắn hoặc hạt enzym có thể được sử dụng, ví dụ, trong chất tẩy rửa rắn và trong thức ăn gia súc. Phương pháp sản xuất dạng rắn của enzym là đã biết trong lĩnh vực, chẳng hạn như nhưng không giới hạn ở sự tạo hạt (làm lạnh bằng cách phun trong nguyên liệu sáp), ép đùn, kết tụ, hoặc kết viên (pha loãng bằng nguyên liệu trơ và chất gắn). Hợp phần enzym rắn có chứa dạng rắn của enzym theo sáng chế, ở dạng bột hỗn hợp, viên nén, và dạng tương tự, được định hình dạng.

Bản mô tả này bao gồm nội dung trong các yêu cầu bảo hộ kèm theo, cũng như là nội dung của phần mô tả trên đây. Mặc dù sáng chế này được mô tả ở dạng được ưu tiên với mức độ phân biệt nhất định, cần hiểu rằng việc mô tả dạng được ưu tiên này chỉ làm ví dụ và nhiều thay đổi về chi tiết của cấu trúc và sự kết hợp và sự sắp xếp của các phần có thể được áp dụng.

Ví dụ thực hiện sáng chế

Các ví dụ sau đây nhằm minh họa thêm cho sáng chế, không nhằm làm giới hạn sáng chế ở các phương án cụ thể được mô tả trong đó.

Ví dụ 1: Phân lập ADN hệ gen từ *M. phaseolina*

ADN hệ gen được phân lập từ chủng *M. phaseolina* ms6 bằng cách sử dụng quy trình được mô tả bởi Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. Practical *Streptomyces* Genetics, 2000; John Innes Foundation, Norwich, UK, pp. 162–208. Một cách vắn tắt, hệ sợi được cạo ra từ đĩa Petri dự trữ bằng cách sử dụng móc cấy và được cấy vào môi trường đextroza khoai tây lỏng trong bình thót cổ hình nón. Bình thót cổ hình nón được ủ không lắc trong thời gian khoảng 72 giờ ở nhiệt độ 30°C. Hệ sợi phát triển trên bề mặt ở mặt phân cách không khí-môi trường. Hệ sợi được thu hoạch bằng cách sử dụng tăm tiệt trùng và được đặt giữa khăn giấy tiệt trùng và được rửa một vài lần bằng dung dịch sinh lý (Đệm natri phosphat có độ pH 7,0). Hệ sợi được ép để loại bỏ chất lỏng dư và hệ sợi thu được được để cho khô trong không

khí trong thời gian 30 phút. Hệ sợi khô chưa hoàn toàn được đặt trong nitơ lỏng và được nghiền để tạo ra bột mịn và cuối cùng phân lập ADN theo quy trình được mô tả trong tài liệu Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. Practical *Streptomyces* Genetics, 2000; John Innes Foundation, Norwich, UK, pp. 162–208.

Ví dụ 2: Thiết kế và tổng hợp đoạn mồi

Đoạn mồi được sử dụng trong nghiên cứu này được thiết kế từ hệ phiên mã được chọn lọc bằng tay và "mô hình gen" được dự đoán từ trình tự hệ gen của *M. phaseolina* ms6, bằng cách lựa chọn các trình tự bằng tay với các ORF hoàn chỉnh hoặc bằng cách sử dụng cơ sở dữ liệu trong đó gen tương tự đã được phân lập thành công từ thực vật khác. Phân tích thông tin sinh học so sánh của các trình tự nucleotit thu được từ hệ phiên mã được thực hiện bằng cách sử dụng NCBI BLAST, BLASTP, RPS-BLAST, BLASTX và PSI-BLAST để xác định dạng tương đồng của gen liên quan và để xác định chính xác gen. Sự sắp thẳng hàng trình tự nucleotit được thực hiện bằng cách sử dụng clustalW phiên bản 1.82 bất cứ khi nào tìm thấy nhiều trình tự từ "tổ hợp gen". Sau đó, sự sắp thẳng hàng này được hiệu chỉnh. Đoạn mồi đặc hiệu gen (cả xuôi và ngược) được chọn lọc bằng tay hoặc thông qua Đoạn mồi 3 cộng với công cụ và đoạn mồi được tổng hợp tùy chỉnh.

Tất cả các oligonucleotit được sử dụng trong nghiên cứu này được tổng hợp và được tinh chế bằng HPLC bởi nhà cung cấp và mua của hãng Integrated DNA Technologies (IDT). Dung dịch dự trữ 100 pmol được điều chế trong ddH_2O đã thanh trùng và được lưu trữ ở nhiệt độ -20°C, trong các phần phân ước để sử dụng.

Trình tự oligonucleotit được sử dụng làm đoạn mồi cho PCR

Tên gen	SEQ ID NO.	Nhận dạng trình tự	Trình tự đoạn mồi	Sản phẩm khuếch đại (bp)
Pectat lyaza	1	64	Xuôi TCATCACCGACCCACAATCGC	859
		65	Ngược CTTAGCAAGCGGGAGCCGTC	
Pectat lyaza	4	66	Xuôi CAGACAACGGCGTCGACAGT	1029

		67	Ngược ACGGACCACATCCAACGCGA	
Pectat lyaza	7	68	Xuôi ATGCCCTGGTACCCTGCCCC	1010
		69	Ngược GAAAGGTCCCGGGCTCACGC	
Pectat lyaza	10	70	Xuôi CCGCTCCTCTGTGCGTTGTTGAA	1036
		71	Ngược ACAGCGATAGAGCCGCAACACG	
Pectat lyaza	13	72	Xuôi TTTCCTCACCACCACTTCCCCTCT	982
		73	Ngược CGAAAGCCAGCCCAAGCGAC	
Pectat lyaza	16	74	Xuôi GGCAAGCATCCTCTCTCCGGC	921
		75	Ngược CCTGAGGCCCATCGTCCGAGT	
Pectat lyaza	19	76	Xuôi TCCCTGCCTCTCCATCCTACCCT	1083
		77	Ngược CCAACCGTGTGGCCGTCGAA	
Pectat lyaza	22	78	Xuôi ACATGCAGTTCAAGTACGCCGCT	875
		79	Ngược GCAGGGGCCCACACCTCTTG	
Pectat lyaza	25	80	Xuôi TCCAAGGCAGAGCCTCCGAAACA	2382
		81	Ngược GCTCCCTGCGGGCGTGTTTTA	
Pectat lyaza C	28	82	Xuôi ACAGACCCCATCACATCCGCCA	1790
		83	Ngược TTGCATTGAGCCCTCTTGGGCTGT	
Pectat lyaza C	31	84	Xuôi CTGTCCTCCAGCAGCAGCCTC	1133

		85	Ngược CGAGCAACCCGTCGAGGTCAA	
Tên gen	SEQ ID NO.		Trình tự đoạn mồi	Sản phẩm khuếch đại (bp)
Pectat lyaza C	34	86	Xuôi CATGAAGGCCACCACCCTCGC	1080
		87	Ngược CCGAACCCTGGCTCGGGCAT	
Pectat lyaza C	37	88	Xuôi ACCCCCACGCCGAACAATACC	1534
		89	Ngược CCGCCAAGTCTATCCGCTCGC	
Pectat lyaza C	40	90	Xuôi GGCTTCGGTGGACCGTCCTAT	1642
		91	Ngược CCACCCGCTCCGCCCTTAAA	
Pectat lyaza C	43	92	Xuôi AGGAGATGGGCTGCCGTCCT	1051
		93	Ngược GCTCAGCAAGCGTCCAACCCA	
Pectinesteraza	46	94	Xuôi TCCCGTGCCATGGTAGCCTTT	1352
		95	Ngược GCTGGCCACCACAATCCACA	
Pectinesteraza	49	96	Xuôi TGGCCTCTTGATCAGGCTCGT	830
		97	Ngược GCGGTGACCCATCCCTGCTT	
Pectinesteraza	52	98	Xuôi TCGTCCACCGGTACCACGTT	7503
		99	Ngược TGCTACGCAAGTGTGCAAAGTGT	
Pectinesteraza	55	100	Xuôi ATCACAGCATGCCTCGCCTCG	1063
		101	Ngược CAGTCCTGCCGGTTCCTTGCAT	

Rhamnogalacturonaza	58	102	Xuôi CGGAGGATTGCTGCGGGACTT	1844
		103	Ngược CCGCGAAACCATCCACAACACG	
Rhamnogalacturonaza	61	104	Xuôi ACGGGCCAAGGTGCCAGAAC	1901
		105	Ngược CACCCTTCTCCTGACCCTCGCT	

Ví dụ 3: Khuếch đại, tách dòng và đọc trình tự pectat lyaza, pectat lyaza C, pectinesteraza và rhamnogalacturonaza từ *M. phaseolina* ms6

ARN tổng số được phân lập từ hệ sợi ba ngày tuổi phát triển trên môi trường lỏng như đã được mô tả từ trước bởi Chomczynski P and Sacchi N, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction (Anal Biochem 1987, 162: 156-159). Chất lượng hoặc tính toàn vẹn của ARN được kiểm tra bằng cách điện di trên gel agaroza và được định lượng bằng cách sử dụng Thermo Scientific Nano Drop 2000 theo quy trình tiêu chuẩn. Sợi cADN thứ nhất được tổng hợp bằng cách sử dụng transcriptaza ngược SuperScript III (Invitrogen) theo hướng dẫn của nhà sản xuất. Gen được khuếch đại từ cADN bằng PCR bằng cách sử dụng đoạn mồi đặc hiệu gen. Phản ứng PCR (50µl) có chứa 1µl cADN, 20pmol của mỗi đoạn mồi, 5µl 10X đệm PCR, 5µl hỗn hợp dNTP 2,5mM và 1,0 đơn vị ADN polymeraza PfuTaq. PCR được thực hiện trong Thermal Cycler (Applied Biosystems) bằng cách sử dụng các điều kiện sau đây: làm biến tính ban đầu trong thời gian 5 phút ("min") ở nhiệt độ 95°C sau đó là 35 chu kỳ biến tính ở nhiệt độ 95°C trong thời gian 30 giây ("sec"), ủ ở nhiệt độ 59-61°C trong thời gian 30 giây và kéo dài ở nhiệt độ 72°C trong thời gian từ 1 đến 2,0 phút tùy thuộc vào chiều dài của gen đích, với sự kéo dài cuối cùng ở nhiệt độ 72°C trong thời gian 7 phút. Sản phẩm PCR được phân tích bằng gel agaroza 1% bằng cách sử dụng đệm 1XTAE và sản phẩm khuếch đại được rửa giải từ gel bằng cách sử dụng bộ kit chiết gel QIAGEN theo hướng dẫn của nhà sản xuất. Sản phẩm PCR đã tinh sạch được nối vào bộ kit tách dòng pCR®8/GW/TOPO® TA (Invitrogen) và được biến nạp vào tế bào *E. coli* đủ khả năng (Invitrogen). Plasmit được phân lập từ các khuẩn lạc giả định bằng cách sử dụng

bộ kit QIAprip Spin Miniprep (QIAGEN) theo hướng dẫn của nhà sản xuất. Sự có mặt của vật cài xen được kiểm tra bằng cách sử dụng đoạn mồi đặc hiệu gen và plasmit dương tính được đưa đi đọc trình tự.

Ví dụ 4: Phân tích trình tự

Trình tự nucleotit và trình tự axit amin lần lượt được phân tích bằng chương trình BLASTN và chương trình BLASTP. Các trình tự được báo cáo từ thực vật khác được sắp thẳng hàng với ClustalW. Phân tích phát sinh loài được thực hiện bằng cách sử dụng Neighbour Joining (NJ).

SEQ ID NO : 1
CHIỀU DÀI : 1047 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

GAAGAGTATATAGAGGTGGCCACGACTGCACTGTTGAACTCATCACCGACCCACAATCGCTTTCTTTCTGCCCAC
TCTCGTTCACAACACAGATCTCTGATCTACCACATTCGTTCTTACCGAGGATCAAAAATTCCTCCCATCTCCACA
ATGCAGTACAAGTACACCGTCCTTGCCGCCACTATGGCCGCCGTTGCCTTCGCCCAGCAGAACTTGATTGGCATT
CCTACCGGTACTAAAGCTAAGCCTTTCACCCAGCCCGAGACCATTACCTTCTTCGACGGTAAAATGATGGAATAC
GGCCGCGGCAAGCCCTGTGGCACCGACGATGACAAGGGTGACCTTGAGGCCGTCTTCATCATCAAGCCCGGCGGC
ACGCTGCAGAACGCCATCATCGGCGCCGACTCCCTTGAGGGCGTGCACTGCGAGGGCGCCTGCACCATCAAGAAC
GTCTGGTTCAAGGACGTGTGCGAAGACGCCATCACCCTCAAGGGCAACGGCCCGTACCTGATCACCGGTGGCGGT
GCGCAGCACGCCAAGGACAAGGTCGTCCAGCACAACGGCAAGGGCACCGTGACCATCTCCGATTACAAGATCGGC
GCCGTCGGTAAGCTGTACCGCAGCTGTGGCAACTGCAGCAACAACGGCGGCCCCCGCAACGTCGTCCTCGACCGC
ATCTCTTCCTTCGGCCCCGGCACCACGTCCGACCTTGTCGGCATCAACTCCAACTACAACGACGTCGCCACCATC
AGTGGCGTCTGCGGCCCTGTCAAGAACATGTGTCAGGAGTTCACCGGCATCGAGAAGGACGGCAACAAGGAGAGC
CCCCACAGGGAGCCCCCGATTGGTGCTTGCAAGGGCCCCCAGGGCCAGCTCAAGACGGCTCCCGCTTGCTAAGCG
CCTCACGGAATTATCATGGGTGTGCAGTTACGGGAAAGGAAATCGATATCGACCTCCCTATTGGTAACATCAATC
TGGCTCCAAATTGCATCATCCGCCACTGGTCGGCCATGACAGGACAGAGCGAAGTCAAAAGTTTCGGTGTAC

SEQ ID NO : 2
CHIỀU DÀI : 747
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(747)

```
atgcagtacaagtacaccgtccttgccgccactatggccgccgttgccttcgcccagcag
 M  Q  Y  K  Y  T  V  L  A  A  T  M  A  A  V  A  F  A  Q  Q
aacttgattggcattcctaccggtactaaagctaagcctttcacccagcccgagaccatt
 N  L  I  G  I  P  T  G  T  K  A  K  P  F  T  Q  P  E  T  I
accttcttcgacggtaaaatgatggaatacggccgcggcaagccctgtggcaccgacgat
 T  F  F  D  G  K  M  M  E  Y  G  R  G  K  P  C  G  T  D  D
gacaagggtgaccttgaggccgtcttcatcatcaagcccggcggcacgctgcagaacgcc
 D  K  G  D  L  E  A  V  F  I  I  K  P  G  G  T  L  Q  N  A
atcatcggcgccgactcccttgagggcgtgcactgcgagggcgcctgcaccatcaagaac
 I  I  G  A  D  S  L  E  G  V  H  C  E  G  A  C  T  I  K  N
gtctggttcaaggacgtgtgcgaagacgccatcaccctcaagggcaacggcccgtacctg
 V  W  F  K  D  V  C  E  D  A  I  T  L  K  G  N  G  P  Y  L
atcaccggtggcggtgcgcagcacgccaaggacaaggtcgtccagcacaacggcaagggc
```

I T G G G A Q H A K D K V V Q H N G K G
accgtgaccatctccgattacaagatcggcgccgtcggtaagctgtaccgcagctgtggc
T V T I S D Y K I G A V G K L Y R S C G
aactgcagcaacaacggcggcccccgcaacgtcgtcctcgaccgcatctcttccttcggc
N C S N N G G P R N V V L D R I S S F G
cccggcaccacgtccgaccttgtcggcatcaactccaactacaacgacgtcgccaccatc
P G T T S D L V G I N S N Y N D V A T I
agtggcgtctgcggccctgtcaagaacatgtgtcaggagttcaccggcatcgagaaggac
S G V C G P V K N M C Q E F T G I E K D
ggcaacaaggagagcccccacagggagcccccgattggtgcttgcaagggcccccagggc
G N K E S P H R E P P I G A C K G P Q G
cagctcaagacggctcccgcttgctaa
Q L K T A P A C -

SEQ ID NO : 3
CHIỀU DÀI : 248
LOẠI : PRT
SINH VẬT : *M. phaseolina*
MQYKYTVLAATMAAVAFAQQNLIGIPTGTKAKPFTQPETITFFDGKMMEYGRGKPCGTDDDKGDLEAVFIIKPGG
TLQNAIIGADSLEGVHCEGACTIKNVWFKDVCEDAITLKGNGPYLITGGGAQHAKDKVVQHNGKGTVTISDYKIG
AVGKLYRSCGNCSNNGGPRNVVLDRISSFGPGTTSDLVGINSNYNDVATISGVCGPVKNMCQEFTGIEKDGNKES
PHREPPIGACKGPQGQLKTAPAC*

SEQ ID NO : 4
CHIỀU DÀI : 1167 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*
CTACCACGAATCGGAATTAAACGACGAATCTCCTGGTGAAGACTATAAAAGGCAGACAACGGCGTCGACAGTAAA
GACCATCGCAATCCAACAGTCCAGTTCCTCCTCAAGTCTCTCACTCCCATCCTCCACTTTCCAATCACTTCCAAC
ATGAAGTTCATCTACGCCGCTATCACCGCCACCGTGGCTGGCCTTGCCTCTGCGCAGTCCCTGACCATCCCCACC
CGCTCCGGCAGCAAGGTCGTCCTCTCCGCGCCCAGTACCATCTCCGGCTCGCAAGACTTTGGTAACAAGGAGTTC
GACCGCGGAATCCCCTGCGACTCGGACGATGACACTGGCAGCTCCAGCGCCGTCTTCATCCTCAAGAACGGCGCC
AGCATCTCCAACGTCATCATTGGTACCGACGCGCTCGAGGGCGTCCACTGCGAGGGTGCCTGCACCCTGACCAAC
GTCTGGTTCCGTGACGTCTGCGAGGGTGAGCTAAATCCACAACATACCATCAAACACTCAAGTAGTGCACACGAA
TGCTAATGGGAAAATCCAGACGCCATCTCCGCGCTCGGCACCGGCAATGTCCTCATCCAGGGCGGCGGTGCGCAG
AACGCCAAGGACAAGGTCGTCCAGCACAACGGCCGCGGCACCGTCACCATCAAGAACTACACCGTCGTCAACGCC
GGCAAGCTGTACCGCAGCTGCGGCGACTGCACCAACAACGGCGGCCCCCGCAACGTCGTCGTGGACAACGTCCGC
GTCAACGGCATGACCTCTGACCTCGTCGGCATCAACTCCAACTACGGTGACAGTAAGTCGGCAAAAGATACTTCA
AATCGTTGGAGGAACACTCAATACTGACTGTCGATTCCAGCTGCTACCATCAGCAACTCTTGCGGCGTCTCCAAG
AAGGTCTGCCAGGAGTACAAGGGCGTCACCAAGGGCAACGGCGACAGCGAGAAGGTCTCCACCACTGCCAACTGC
AAGGGTGCTCAGGGTACTCTGTCTAAGCTTCCCACTTGCTAGACAACCTAAAAGCAAAACCAAACTAGAGCTTGG
GCACGGATCACTCGCGTTGGATGTGGTCCGTGGAATCCTTGGTCAGTGCTCAATCGCCTGCTTCGAGTGGGAAAC
GGCCGTCTCAATTGATACCGGGCGAGTGTCTGCGCATGGATG

SEQ ID NO : 5
CHIỀU DÀI : 735
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(735)
atgaagttcatctacgccgctatcaccgccaccgtggctggccttgcctctgcgcagtcc
M K F I Y A A I T A T V A G L A S A Q S
ctgaccatccccacccgctccggcagcaaggtcgtcctctccgcgcccagtaccatctcc

```
 L  T  I  P  T  R  S  G  S  K  V  V  L  S  A  P  S  T  I  S
ggctcgcaagactttggtaacaaggagttcgaccgcggaatcccctgcgactcggacgat
 G  S  Q  D  F  G  N  K  E  F  D  R  G  I  P  C  D  S  D  D
gacactggcagctccagcgccgtcttcatcctcaagaacggcgccagcatctccaacgtc
 D  T  G  S  S  S  A  V  F  I  L  K  N  G  A  S  I  S  N  V
atcattggtaccgacgcgctcgagggcgtccactgcgagggtgcctgcaccctgaccaac
 I  I  G  T  D  A  L  E  G  V  H  C  E  G  A  C  T  L  T  N
gtctggttccgtgacgtctgcgaggacgccatctccgcgctcggcaccggcaatgtcctc
 V  W  F  R  D  V  C  E  D  A  I  S  A  L  G  T  G  N  V  L
atccagggcggcggtgcgcagaacgccaaggacaaggtcgtccagcacaacggccgcggc
 I  Q  G  G  G  A  Q  N  A  K  D  K  V  V  Q  H  N  G  R  G
accgtcaccatcaagaactacaccgtcgtcaacgccggcaagctgtaccgcagctgcggc
 T  V  T  I  K  N  Y  T  V  V  N  A  G  K  L  Y  R  S  C  G
gactgcaccaacaacggcggcccccgcaacgtcgtcgtggacaacgtccgcgtcaacggc
 D  C  T  N  N  G  G  P  R  N  V  V  V  D  N  V  R  V  N  G
atgacctctgacctcgtcggcatcaactccaactacggtgacactgctaccatcagcaac
 M  T  S  D  L  V  G  I  N  S  N  Y  G  D  T  A  T  I  S  N
tcttgcggcgtctccaagaaggtctgccaggagtacaagggcgtcaccaagggcaacggc
 S  C  G  V  S  K  K  V  C  Q  E  Y  K  G  V  T  K  G  N  G
gacagcgagaaggtctccaccactgccaactgcaagggtgctcagggtactctgtctaag
 D  S  E  K  V  S  T  T  A  N  C  K  G  A  Q  G  T  L  S  K
cttcccacttgctag
 L  P  T  C  -
```

SEQ ID NO : 6
CHIỀU DÀI : 244
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MKFIYAAITATVAGLASAQSLTIPTRSGSKVVLSAPSTISGSQDFGNKEFDRGIPCDSDDDTGSSSAVFILKNGA
SISNVIIGTDALEGVHCEGACTLTNVWFRDVCEDAISALGTGNVLIQGGGAQNAKDKVVQHNGRGTVTIKNYTVV
NAGKLYRSCGDCTNNGGPRNVVVDNVRVNGMTSDLVGINSNYGDTATISNSCGVSKKVCQEYKGVTKGNGDSEKV
STTANCKGAQGTLSKLPTC*
```

SEQ ID NO : 7
CHIỀU DÀI : 1110 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
CAGCGGTATATAAATGCCCTGGTACCCTGCCCCAAAATGCCTTCTTCACCACCACAGTCCTTCGTTTCAATCGTT
CAGCACATTCACTCTTTCATTCACACTCACTAACCGATCTGCGATCGTGTTCTTTGCCTTCCACTACCAGACACA
ATGTTCTCCAAGCTTTTCCTGCTTCCCCTCCTGGCGGCTTCCGCCCTGGCTGCTCCTGCCGACGACACTTTCGGC
TACGAGCTCGTTCGCCGCGCGAACTTCCCCATTCCTGCCTCCAAGGGCAACGTCAAGCTCAGCGCTGCCAAGTCC
GTGTCCGGCACCTTCGATGGTGGCATGAAGACCTACGGCCGTGGTGTCAAGTGCACCGGTCAGGCCGAGGGTGGT
GACAAGGACGCCGTCTTCATCCTCGAGAACGGTGCTACCCTCAAGAACGCCATCATTGGTACCGACCAGATCGAG
GGCGTCCACTGCAAGGGCTCCTGCACCATCGAAAACGTCTGGTGGGCCGGTGTCTGCGAGGACGCGCTCTCCCTG
AAGGGTGACGGTTCCGCCAAGGTCATCGGCGGCGGCGCTACTGGCGCTGAAGACAAGGTACGTACGCCTCTGATA
CGCCTTCGCTCAGATTGCTGACCAGATCTGGTACAGGTCATCCAGCACAACGGCGTTGGCTCCGTCTCGATTGAT
GGCTTCACTGTTGCCGACTTCGGCAAGCTCTACCGCTCGTGCGGAAACTGCAAGAAGATGGGCAAGAGAACCGTC
ACCATCAAGAACGTGAAGGCTAGCAACGGAAAGCTTTTGGCTGGCATCAACTCCAACTACGGCGACACTGCTACC
ATCACCGGCACCTGCGCTACCTCCGTCAAGAAGGTCTGCACCGAGTTCAAGGGCAACAACAGCGGCAAGGAGCCC
ACCGAGATCAGCTCCGGCCCCAGCAACGCCTGCAAGTACTCCTCCCTCAAGGCCTGCTAGGGCCTGCTAGACGGT
TGCTCAACAAACAAACATGGCATTTGCCGCGTGAGCCCGGGACCTTTCCGGGCTCGGGAGAAGAAAGTGGATGGC
ATTCTATTCTTGTATCGACTTCTAGTACTAGAGCATCTTCTATACCTCTAATTTACTGTG
```

SEQ ID NO : 8
CHIỀU DÀI : 756
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(756)

```
atgttctccaagcttttcctgcttcccctcctggcggcttccgccctggctgctcctgcc
 M  F  S  K  L  F  L  L  P  L  L  A  A  S  A  L  A  A  P  A
gacgacactttcggctacgagctcgttcgccgcgcgaacttccccattcctgcctccaag
 D  D  T  F  G  Y  E  L  V  R  R  A  N  F  P  I  P  A  S  K
ggcaacgtcaagctcagcgctgccaagtccgtgtccggcaccttcgatggtggcatgaag
 G  N  V  K  L  S  A  A  K  S  V  S  G  T  F  D  G  G  M  K
acctacggccgtggtgtcaagtgcaccggtcaggccgagggtggtgacaaggacgccgtc
 T  Y  G  R  G  V  K  C  T  G  Q  A  E  G  G  D  K  D  A  V
ttcatcctcgagaacggtgctaccctcaagaacgccatcattggtaccgaccagatcgag
 F  I  L  E  N  G  A  T  L  K  N  A  I  I  G  T  D  Q  I  E
ggcgtccactgcaagggctcctgcaccatcgaaaacgtctggtgggccggtgtctgcgag
 G  V  H  C  K  G  S  C  T  I  E  N  V  W  W  A  G  V  C  E
gacgcgctctccctgaagggtgacggttccgccaaggtcatcggcggcggcgctactggc
 D  A  L  S  L  K  G  D  G  S  A  K  V  I  G  G  G  A  T  G
gctgaagacaaggtcatccagcacaacggcgttggctccgtctcgattgatggcttcact
 A  E  D  K  V  I  Q  H  N  G  V  G  S  V  S  I  D  G  F  T
gttgccgacttcggcaagctctaccgctcgtgcggaaactgcaagaagatgggcaagaga
 V  A  D  F  G  K  L  Y  R  S  C  G  N  C  K  K  M  G  K  R
accgtcaccatcaagaacgtgaaggctagcaacggaaagcttttggctggcatcaactcc
 T  V  T  I  K  N  V  K  A  S  N  G  K  L  L  A  G  I  N  S
aactacggcgacactgctaccatcaccggcacctgcgctacctccgtcaagaaggtctgc
 N  Y  G  D  T  A  T  I  T  G  T  C  A  T  S  V  K  K  V  C
accgagttcaagggcaacaacagcggcaaggagcccaccgagatcagctccggccccagc
 T  E  F  K  G  N  N  S  G  K  E  P  T  E  I  S  S  G  P  S
aacgcctgcaagtactcctccctcaaggcctgctag
 N  A  C  K  Y  S  S  L  K  A  C  -
```

SEQ ID NO : 9
CHIỀU DÀI : 251
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MFSKLFLLPLLAASALAAPADDTFGYELVRRANFPIPASKGNVKLSAAKSVSGTFDGGMKTYGRGVKCTGQAEGG
DKDAVFILENGATLKNAIIGTDQIEGVHCKGSCTIENVWWAGVCEDALSLKGDGSAKVIGGGATGAEDKVIQHNG
VGSVSIDGFTVADFGKLYRSCGNCKKMGKRTVTIKNVKASNGKLLAGINSNYGDTATITGTCATSVKKVCTEFKG
NNSGKEPTEISSGPSNACKYSSLKAC*
```

SEQ ID NO : 10
CHIỀU DÀI : 1127 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
CCGCTCCTCTGTGCGTTGTTGAACCTTGCTGTTTGTAGGTATAAAGCCCCCCCGTCATTGCTGAGGACGTGTATG
GACAGGATCCCTCAAACAGATTGTCTTCTACATCTCAGCGCATTAAGAACCCCATAAATACCAAGAGTTACAAAC
ATGCGTTCTACAAATCTGCTCGTCCTCCTCGCAACCTCCCTCAGCCTGGTCGCAGCGAGCCCGTTGGATGCGCCG
ACCAAAGTGATGGGCAAGCGGTTCGCCACCAATGTTCTACCCGCATCATCTGGACATGTTGTCCTCCCCTCAGCA
ACCACTGTGAGCACTTTTGACGGCGGAATGAAGAAGTACGACCGTGGTACCTCTTGCACTGGTCAGTCCGAAGGG
GGGGATGCGGACGCTGTCTTCCTCGTTCAGTCCGGAGGCACCCTCAAGAATGTCATCATTGGTGCGGACCAATCT
```

GAAGGTGTCCACTGTCTCGGCCCGTGCACGATCCAGAACGTGTGGTGGGAGGCCGTCTGCGAGGGTGAGAGCCAG
ACGCTTCGCGTTGTAAAGGGGCGAAATGCTGACCTCAAGACCCAGACGCCTTGACCATCAAACAAACTTCTGGGA
CCTCCTACGTCGTTGGCGGCGGTGCTTTTGGCGCGTCAGACAAAATTATCCAGCACAACGGAGGCGGTACCGTCT
CGATCAAGGACTTCTACGCCCAAGATTTCGGCAAAGTTTACCGCAGCTGCGGCAATTGCGGCACTCAGTACAAAC
GCACCGTCACCATGTCTGGGATTTGGGCCGTTAATGGTGATCTCCTTGCCGGCGTTAACTCCAATTACGGTGATA
CCGCGACCATTTCCGGCACTTGTGCAGACAACGTGGACAACATCTGCGCCTGGTACGAAGGAAACGATGATGGCG
ATGAGCCCACCAAGTTGGGCACTGGCATCAGCTCTTACTGTGTCTATACCGCCAATGGTGTCGATGACTGCCCTT
GAGTTGAACATTGGCTTCTATTTGCCACGACACTATGATCGTGTTGCGGCTCTATCGCTGTTGTGTCCGGTTGGT
GAGGGCACCCGTACATATGTGGAATGGAGGTCACGTGGTGGATAGCATGTACATATTCTTGCCTGTTATGCCTAC
AC

SEQ ID NO : 11
CHIỀU DÀI : 771
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(771)

```
atgcgttctacaaatctgctcgtcctcctcgcaacctccctcagcctggtcgcagcgagc
 M  R  S  T  N  L  L  V  L  L  A  T  S  L  S  L  V  A  A  S
ccgttggatgcgccgaccaaagtgatgggcaagcggttcgccaccaatgttctacccgca
 P  L  D  A  P  T  K  V  M  G  K  R  F  A  T  N  V  L  P  A
tcatctggacatgttgtcctcccctcagcaaccactgtgagcacttttgacggcggaatg
 S  S  G  H  V  V  L  P  S  A  T  T  V  S  T  F  D  G  G  M
aagaagtacgaccgtggtacctcttgcactggtcagtccgaaggggggggatgcggacgct
 K  K  Y  D  R  G  T  S  C  T  G  Q  S  E  G  G  D  A  D  A
gtcttcctcgttcagtccggaggcaccctcaagaatgtcatcattggtgcggaccaatct
 V  F  L  V  Q  S  G  G  T  L  K  N  V  I  I  G  A  D  Q  S
gaaggtgtccactgtctcggcccgtgcacgatccagaacgtgtggtgggaggccgtctgc
 E  G  V  H  C  L  G  P  C  T  I  Q  N  V  W  W  E  A  V  C
gaggacgccttgaccatcaaacaaacttctgggacctcctacgtcgttggcggcggtgct
 E  D  A  L  T  I  K  Q  T  S  G  T  S  Y  V  V  G  G  G  A
tttggcgcgtcagacaaaattatccagcacaacggaggcggtaccgtctcgatcaaggac
 F  G  A  S  D  K  I  I  Q  H  N  G  G  G  T  V  S  I  K  D
ttctacgcccaagatttcggcaaagtttaccgcagctgcggcaattgcggcactcagtac
 F  Y  A  Q  D  F  G  K  V  Y  R  S  C  G  N  C  G  T  Q  Y
aaacgcaccgtcaccatgtctgggatttgggccgttaatggtgatctccttgccggcgtt
 K  R  T  V  T  M  S  G  I  W  A  V  N  G  D  L  L  A  G  V
aactccaattacggtgataccgcgaccatttccggcacttgtgcagacaacgtggacaac
 N  S  N  Y  G  D  T  A  T  I  S  G  T  C  A  D  N  V  D  N
atctgcgcctggtacgaaggaaacgatgatggcgatgagcccaccaagttgggcactggc
 I  C  A  W  Y  E  G  N  D  D  G  D  E  P  T  K  L  G  T  G
atcagctcttactgtgtctataccgccaatggtgtcgatgactgcccttga
 I  S  S  Y  C  V  Y  T  A  N  G  V  D  D  C  P  -
```

SEQ ID NO : 12
CHIỀU DÀI : 256
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MRSTNLLVLLATSLSLVAASPLDAPTKVMGKRFATNVLPASSGHVVLPSATTVSTFDGGMKKYDRGTSCTGQSEG
GDADAVFLVQSGGTLKNVIIGADQSEGVHCLGPCTIQNVWWEAVCEDALTIKQTSGTSYVVGGGAFGASDKIIQH
NGGGTVSIKDFYAQDFGKVYRSCGNCGTQYKRTVTMSGIWAVNGDLLAGVNSNYGDTATISGTCADNVDNICAWY
EGNDDGDEPTKLGTGISSYCVYTANGVDDCP*

SEQ ID NO : 13
CHIỀU DÀI : 1164 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

AGTGACATATAAGTCCTCTGTTCTGCCATCGGAAAATGCTTCTCATCACCAACATTCTCTTCCTTCTTTCCTCAC
CACCACTTCCCCTCTTTCAAACAATTACCGGATAGATCTGTGATCATCCCCTACTTTCGTTTAACAACATACAAA
ATGCAGTCCAAGATCGTTCTCGGCCTTTCTCTCCTCGCCGCCACCGGCATGGCGGCTCCCTCCGAGCCCCTCCGC
GTCCGCGCCACCCTCCCCATCCCCTCCTCCAAGGGCTCTGTCACCTACGATGAGGTCAAGTCCATCACCGGCACT
TTCGACGGTGGCATGAAGACCTACGGCCGTGGTGTTTCCTGCACTGGCCAGAAGGAGGGTGGCAACAAGGACGCC
GTCTTCCAGCTTGAGGACGGTGCTACCCTCAAGAACGTCATCATTGGCGAGGACCAGATCGAGGGTATCCACTGC
ATGGGCAGCTGCACCCTGGAGAACGTCTGGTGGAGCGCCGTCTGCGAGGGTTTGTCTTCTTCCATCCACTGAAGC
ACTTCACAGCTAACGTCATCTCAGACGCTCTTACCTTCAAGGGTGACGGTGACGCCAAGGTCATCGGCGGCGGTG
CCCAGGGCGCCGACGACAAGGTCCTCCAGCACAACGGTGTCGGCGATGTGACCATCGACGGCTTCACCGTTGTCG
ATTTCGGCAAGCTGTATCGCTCTTGCGGCAGTAAGTTCCCCCCGGCTTCTGCCCACCTCCATCCCTCATTCACCT
GCTCACACACGTGTTCCAGACTGCAAGCAGAACGGCGGCACCCGCAACGTCAACATCTCCAACGTCAAGGCCTAC
AACGGCAAGGTCCTCACCGGCATCAACTCCAACTACGGCGACGTCGCCACCTTCAAGGACACCTGCGCCTCTTCC
GTCAAGGACATCTGCGTCGAGTACAAGGGTACCAACAACAACGACGAGGAGCCCTCCAAGATCGGCTCCGGTCCT
TCCGACAACTGCGTCTACTCCGACATCCCCTCTTGCTAGACGGCTCGCCAACGGTCGCTTGGGCTGGCTTTCGCG
GATGCCCATTGTGAATAATCCGCTTGTTTCTTGTATGTAATTATTTTAAAGGCAAGAAGACCATTTAAATCAATA
AATAGAGACGGGGCTAGACCTCGCTTGTACTACGCACAC

SEQ ID NO : 14
CHIỀU DÀI : 750
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(750)

```
atgcagtccaagatcgttctcggcctttctctcctcgccgccaccggcatggcggctccc
 M  Q  S  K  I  V  L  G  L  S  L  L  A  A  T  G  M  A  A  P
tccgagcccctccgcgtccgcgccaccctccccatcccctcctccaagggctctgtcacc
 S  E  P  L  R  V  R  A  T  L  P  I  P  S  S  K  G  S  V  T
tacgatgaggtcaagtccatcaccggcactttcgacggtggcatgaagacctacggccgt
 Y  D  E  V  K  S  I  T  G  T  F  D  G  G  M  K  T  Y  G  R
ggtgtttcctgcactggccagaaggagggtggcaacaaggacgccgtcttccagcttgag
 G  V  S  C  T  G  Q  K  E  G  G  N  K  D  A  V  F  Q  L  E
gacggtgctaccctcaagaacgtcatcattggcgaggaccagatcgagggtatccactgc
 D  G  A  T  L  K  N  V  I  I  G  E  D  Q  I  E  G  I  H  C
atgggcagctgcaccctggagaacgtctggtggagcgccgtctgcgaggacgctcttacc
 M  G  S  C  T  L  E  N  V  W  W  S  A  V  C  E  D  A  L  T
ttcaagggtgacggtgacgccaaggtcatcggcggcggtgcccagggcgccgacgacaag
 F  K  G  D  G  D  A  K  V  I  G  G  G  A  Q  G  A  D  D  K
gtcctccagcacaacggtgtcggcgatgtgaccatcgacggcttcaccgttgtcgatttc
 V  L  Q  H  N  G  V  G  D  V  T  I  D  G  F  T  V  V  D  F
ggcaagctgtatcgctcttgcggcaactgcaagcagaacggcggcacccgcaacgtcaac
 G  K  L  Y  R  S  C  G  N  C  K  Q  N  G  G  T  R  N  V  N
atctccaacgtcaaggcctacaacggcaaggtcctcaccggcatcaactccaactacggc
 I  S  N  V  K  A  Y  N  G  K  V  L  T  G  I  N  S  N  Y  G
gacgtcgccaccttcaaggacacctgcgcctcttccgtcaaggacatctgcgtcgagtac
 D  V  A  T  F  K  D  T  C  A  S  S  V  K  D  I  C  V  E  Y
aagggtaccaacaacaacgacgaggagccctccaagatcggctccggtccttccgacaac
 K  G  T  N  N  N  D  E  E  P  S  K  I  G  S  G  P  S  D  N
```

tgcgtctactccgacatcccctcttgctag
C V Y S D I P S C -

SEQ ID NO : 15
CHIỀU DÀI : 249
LOẠI : PRT
SINH VẬT : *M. phaseolina*
MQSKIVLGLSLLAATGMAAPSEPLRVRATLPIPSSKGSVTYDEVKSITGTFDGGMKTYGRGVSCTGQKEGGNKDA
VFQLEDGATLKNVIIGEDQIEGIHCMGSCTLENVWWSAVCEDALTFKGDGDAKVIGGGAQGADDKVLQHNGVGDV
TIDGFTVVDFGKLYRSCGNCKQNGGTRNVNISNVKAYNGKVLTGINSNYGDVATFKDTCASSVKDICVEYKGTNN
NDEEPSKIGSGPSDNCVYSDIPSC*

SEQ ID NO : 16
CHIỀU DÀI : 1238 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*
CTATAAAGATGGAGACCAGCCAGCTTCTAGAGGCTTTTCCCTCCAGGCAAGCATCCTCTCTCCGGCCATCACTTT
CATTCTTTCGCATTCTTCCACTCTTTTGTCCAACTCCGACCCGAACAGATCTGTGATCGTTCACTACACTACACA
ATGTTGACCAAGACTCTTCTTCCCCTCCTCGCCATCTCTGGCGCCGTCTCTGCCGTCAAGACCAAGACCTTCGGC
AAGAACCTCGTCCGCCGTGCGGACTTCCCTATCCCTGAATCCCAGGGTTCCGTCACCTTTGACGCCGCCGAGGAG
GTCGATGGCGAGTACGACGGTGGCTACAAGACGTACGGCCGTGGTGTTTCTTGCACTGGCCAGGAGGAGGGTGGC
GACTCTGACGCCGTCTTCATCCTGAAGGACGGTGCTTCCCTCAAGAACGCCATCATCGGCTCTGACCAGATTGAG
GGTGTCCACTGCGAGGGCTCCTGCACTCTTGAGAACGTTTGGTGGGAGGCCGTCTGCGAGGACGCTCTGTCCTTC
AAGGGTGACGGTGATGCCACCGTCACTGGCGGTGGTGCAACTGGTGCTGAGGACAAAGTCCTCCAGCAAAACGGT
ATTGGCTCCATCACCGTCGATGGCTTCACCGTTGTGGACTTCGGCAAGCTGTACCGCTCTTGCGGTAACTGCGAC
GAGATGGGCCAGCGCACTGTACGTCAATGCTATCGCTCATTCTGTAGGGCAGAGAGTGCTGACATGGCTGTGCGA
CTTAGATCACCCTCAAGAACGTCAAGGCCTACTCCGGCAAGAAGCTTGTCGGCATCAACTCCAACTACGGTGACA
GTGCGTCATGTTCCTTTAGGCTTTCAACCCTTGAGCCATGTGCTAATAGCATCCTCAGCTGCCACCATCACCGAC
ACTTGCGCGACCGACGTCTCTGACATTTGCACGGAGTACGTAGTCACTCGGACGATGGGCCTCAGGAACGAAGAC
ACTAACATAATTATCTTTACAGGTACGAAGGAAACGATACCGGAGACGAACCTGAAGAGATCAGCTCCGGCCCCT
CTGATGCCTGCATTTACAGCGACGTTCCGGAGTGCTAGACGCACAGCTCGAACATGCCGCCTCAAAGCAAGCATT
CTGATGGAGGGCCAGGGGGTGGGAAGGGATTTAGAGCCTCCGAAAAGGTGGAAGTGGGTTTAGGACTTCGTTGTA
TAATATAAACTTGTTTCCGAAAATAAGAACCATCCTGC

SEQ ID NO : 17
CHIỀU DÀI : 756
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(756)
atgttgaccaagactcttcttcccctcctcgccatctctggcgccgtctctgccgtcaag
M L T K T L L P L L A I S G A V S A V K
accaagaccttcggcaagaacctcgtccgccgtgcggacttccctatccctgaatcccag
T K T F G K N L V R R A D F P I P E S Q
ggttccgtcacctttgacgccgccgaggaggtcgatggcgagtacgacggtggctacaag
G S V T F D A A E E V D G E Y D G G Y K
acgtacggccgtggtgtttcttgcactggccaggaggagggtggcgactctgacgccgtc
T Y G R G V S C T G Q E E G G D S D A V
ttcatcctgaaggacggtgcttccctcaagaacgccatcatcggctctgaccagattgag
F I L K D G A S L K N A I I G S D Q I E
ggtgtccactgcgagggctcctgcactcttgagaacgtttggtgggaggccgtctgcgag
G V H C E G S C T L E N V W W E A V C E
gacgctctgtccttcaagggtgacggtgatgccaccgtcactggcggtggtgcaactggt

```
 D  A  L  S  F  K  G  D  G  D  A  T  V  T  G  G  G  A  T  G
gctgaggacaaagtcctccagcaaaacggtattggctccatcaccgtcgatggcttcacc
 A  E  D  K  V  L  Q  Q  N  G  I  G  S  I  T  V  D  G  F  T
gttgtggacttcggcaagctgtaccgctcttgcggtaactgcgacgagatgggccagcgc
 V  V  D  F  G  K  L  Y  R  S  C  G  N  C  D  E  M  G  Q  R
actatcaccctcaagaacgtcaaggcctactccggcaagaagcttgtcggcatcaactcc
 T  I  T  L  K  N  V  K  A  Y  S  G  K  K  L  V  G  I  N  S
aactacggtgacactgccaccatcaccgacacttgcgcgaccgacgtctctgacatttgc
 N  Y  G  D  T  A  T  I  T  D  T  C  A  T  D  V  S  D  I  C
acggagtacgaaggaaacgataccggagacgaacctgaagagatcagctccggcccctct
 T  E  Y  E  G  N  D  T  G  D  E  P  E  E  I  S  S  G  P  S
gatgcctgcatttacagcgacgttccggagtgctag
 D  A  C  I  Y  S  D  V  P  E  C  -
```

SEQ ID NO : 18
CHIỀU DÀI : 251
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MLTKTLLPLLAISGAVSAVKTKTFGKNLVRRADFPIPESQGSVTFDAAEEVDGEYDGGYKTYGRGVSCTGQEEGG
DSDAVFILKDGASLKNAIIGSDQIEGVHCEGSCTLENVWWEAVCEDALSFKGDGDATVTGGGATGAEDKVLQQNG
IGSITVDGFTVVDFGKLYRSCGNCDEMGQRTITLKNVKAYSGKKLVGINSNYGDTATITDTCATDVSDICTEYEG
NDTGDEPEEISSGPSDACIYSDVPEC*
```

SEQ ID NO : 19
CHIỀU DÀI : 1245 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
GGTTCACCTCATCAACTCGCTTCTTCATTCGTTCCTTGCTGTGCAAGACACCATCCAGACGGACATTCCATTACA
ACGGTCTTTGATAGCCAATCCAACTTTGTCTACTCCCTGCCTCTCCATCCTACCCTACCAATAAGCCCATCCAAA
ATGCAGTACAGCATCCGCTCGCTCGCCTTCGGCCTCGCCGCCCTGGCCGTCTCCACCGACGCTTTCATGGTTCCC
CGTGACCCTAGCAAGGGCTGGGGCTTCCCCAACGCCAACAACGCTCGCGTTGCTGCTGCCGCCGCCTCCACTCCT
GTCCAGACCCCCGGCAACACCCCTACCACTCCCGTGGCCACCCCTGGTGTCACTGGCGGCGCCACTGCCGCCGCT
GGCTTCCCCGCCTCGTCCGGCACCAGCCAGCTCAGCGCCCCCATGACCGTCACCGGCAGCTTCGACGGTGGCATG
AAGGCCTTCGGCCGCGGTGTCTCCTGCACTGGCCAGGCTGAGGGTGGCGACAGCGACGCTGTCTTCATGATCGAG
GAGGGCGGTACTCTCTCCAACGTCATCATCGCCGCCGATCAGATCGAGGGTGTCCACTGCTTCGGTTCCTGCACT
CTCAAGAACGTCTGGTGGGTCGCCGTCTGCGAGGACGCTTTCACCATCAAGGAGCAGGGCGCCTCTGGCACTACC
CACATCATCGGCGGTGGTGCTCAGGGCGCCGAAGACAAGGTTCTCCAGCACAACGGCGGCGGCACTCTCGCTGTC
TCTGGCTTCCTCGCCAAGGACTTCGGTAAGCTCTACCGCAGCTGCGGCAACTGCGACGAGATGCCCGAGCGCCAC
GTTACCATCGACGGCGTTACCGCTGAGTCTGGTGACATCCTGGTCGGCATCAACTCCAACATGGGCGACAGTAAG
TAACTCCTGATTTTGCAAGCGATGAGGCAATGCTAACAATGCGCAGCTGCTACCATCACCAACACCTGCGCCACC
GGCGTCGAGACCATCTGCCAGGAGTTCAACGGTGTCACGGACGGCAGCGAGCCCGAGGCTGCCGGCAGCGGCCCT
AGCACTGCTTGCAAGTACACTGCCGCTGACGCCCAGGCCTGCTAAGCACATCCCTTTCTTGATTGACCAGTAAAA
GTTCGACGGCCACACGGTTGGGTAATGCAGGGATTGGCAAAATGTCCGTCCGGCATAATTAGCACTTCTTTGATG
TAGATATTGGGCCGGGTGTCCACAATGTATAACTCTGACATGTCA
```

SEQ ID NO : 20
CHIỀU DÀI : 894
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(894)

```
atgcagtacagcatccgctcgctcgccttcggcctcgccgccctggccgtctccaccgac
 M  Q  Y  S  I  R  S  L  A  F  G  L  A  A  L  A  V  S  T  D
gctttcatggttccccgtgaccctagcaagggctggggcttccccaacgccaacaacgct
 A  F  M  V  P  R  D  P  S  K  G  W  G  F  P  N  A  N  N  A
cgcgttgctgctgccgccgcctccactcctgtccagacccccggcaacacccctaccact
 R  V  A  A  A  A  A  S  T  P  V  Q  T  P  G  N  T  P  T  T
cccgtggccacccctggtgtcactggcggcgccactgccgccgctggcttccccgcctcg
 P  V  A  T  P  G  V  T  G  G  A  T  A  A  A  G  F  P  A  S
tccggcaccagccagctcagcgcccccatgaccgtcaccggcagcttcgacggtggcatg
 S  G  T  S  Q  L  S  A  P  M  T  V  T  G  S  F  D  G  G  M
aaggccttcggccgcggtgtctcctgcactggccaggctgagggtggcgacagcgacgct
 K  A  F  G  R  G  V  S  C  T  G  Q  A  E  G  G  D  S  D  A
gtcttcatgatcgaggagggcggtactctctccaacgtcatcatcgccgccgatcagatc
 V  F  M  I  E  E  G  G  T  L  S  N  V  I  I  A  A  D  Q  I
gagggtgtccactgcttcggttcctgcactctcaagaacgtctggtgggtcgccgtctgc
 E  G  V  H  C  F  G  S  C  T  L  K  N  V  W  W  V  A  V  C
gaggacgctttcaccatcaaggagcagggcgcctctggcactacccacatcatcggcggt
 E  D  A  F  T  I  K  E  Q  G  A  S  G  T  T  H  I  I  G  G
ggtgctcagggcgccgaagacaaggttctccagcacaacggcggcggcactctcgctgtc
 G  A  Q  G  A  E  D  K  V  L  Q  H  N  G  G  G  T  L  A  V
tctggcttcctcgccaaggacttcggtaagctctaccgcagctgcggcaactgcgacgag
 S  G  F  L  A  K  D  F  G  K  L  Y  R  S  C  G  N  C  D  E
atgcccgagcgccacgttaccatcgacggcgttaccgctgagtctggtgacatcctggtc
 M  P  E  R  H  V  T  I  D  G  V  T  A  E  S  G  D  I  L  V
ggcatcaactccaacatgggcgacactgctaccatcaccaacacctgcgccaccggcgtc
 G  I  N  S  N  M  G  D  T  A  T  I  T  N  T  C  A  T  G  V
gagaccatctgccaggagttcaacggtgtcacggacggcagcgagcccgaggctgccggc
 E  T  I  C  Q  E  F  N  G  V  T  D  G  S  E  P  E  A  A  G
agcggccctagcactgcttgcaagtacactgccgctgacgcccaggcctgctaa
 S  G  P  S  T  A  C  K  Y  T  A  A  D  A  Q  A  C  -
```

SEQ ID NO : 21
CHIỀU DÀI : 297
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MQYSIRSLAFGLAALAVSTDAFMVPRDPSKGWGFPNANNARVAAAAASTPVQTPGNTPTTPVATPGVTGGATAAA
GFPASSGTSQLSAPMTVTGSFDGGMKAFGRGVSCTGQAEGGDSDAVFMIEEGGTLSNVIIAADQIEGVHCFGSCT
LKNVWWVAVCEDAFTIKEQGASGTTHIIGGGAQGAEDKVLQHNGGGTLAVSGFLAKDFGKLYRSCGNCDEMPERH
VTIDGVTAESGDILVGINSNMGDTATITNTCATGVETICQEFNGVTDGSEPEAAGSGPSTACKYTAADAQAC*
```

SEQ ID NO : 22
CHIỀU DÀI : 1047 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
GAATGGTATAAAGAGCGGTGGGATATCCATCATTGAGCTCATCACCAACCCACAATTACTTTCGTTCTACACTCC
TCTCGTTTACACAACGGATCACTGATCTACCACATTCGCTCTTACCAAAGATCCTGTCTACCAACCATCTACAAC
ATGCAGTTCAAGTACGCCGCTCTCGCTGCCGCCCTGGCCGCCGTCGGCTCCACGCAGAGCCTCCCCATCCCCGCC
TCCCGCGGAGTTGATCCCAACAAGGGCCAAAGGATCGTCAAGCCCGGCAATCCTTTCGACGGCGGGATGAAGGAG
TTTGGCCGTGGTGTCGTCTGCAACGACAAGGCCGACACTGGCGTCAGGAACGCCGTCTTCGTCATCGAGGACGGC
GGTGTTCTGCGCAACGCCATCATCGGCGCCGACGCCGTCGAGGGCATCCACTGCGAGGGCAAGTGCACCATCGAG
AACGTCTGGTTCCGCGACGTGTGCGAGGACGCCATCACGCTGAAGGGCAACGGCCCTTACACCATCACTGGCGGC
```

GGCGCCCAGAACGCCGGCGACAAGGTCATCCAGCACAACGGCAAGGGCGAGCTGCGCATCTCAAACTACCAGGTC
AACAACGTCGGCAAGCTGTTCCGCACCTGCGGCAACTGCAGCAACAACGGCGGCCCGCGCAGCATCGTCGCCACC
GGCATCAGGGCCTTCGGCGTCACCAGCGACCTCATCGGCATCAACTCCAACTACGGCGACAAGGCCTCCATCACC
GGCTCCTGCGGCAACACCAAGACGGTCTGCCAGGAATATGTCGGCATCGAGAAGGGCGCCAACGGCGGGAAGGAC
AGCGAGAAGAGGGTCCCGCCGGTCGGCGCTTGCACCGGCCAGGGCGGTCTCGCCAGGCTCCCTTCCTGCTAGACG
CCTCGCGGAATCTTGCGCTGAGTATGGGATGGGAGGAATGGATTGACCGCCTCAGTTTGGGATCCGTGCGAACAG
TCGGGAACTTCGCAACTGCCTCATCCGTCAAGAGGTGTGGGCCCCTGCCGGACAGGATGAAACGCGTTTATG

SEQ ID NO : 23
CHIỀU DÀI : 747
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(747)

```
atgcagttcaagtacgccgctctcgctgccgccctggccgccgtcggctccacgcagagc
 M  Q  F  K  Y  A  A  L  A  A  A  L  A  A  V  G  S  T  Q  S
ctccccatccccgcctcccgcggagttgatcccaacaagggccaaaggatcgtcaagccc
 L  P  I  P  A  S  R  G  V  D  P  N  K  G  Q  R  I  V  K  P
ggcaatcctttcgacggcgggatgaaggagtttggccgtggtgtcgtctgcaacgacaag
 G  N  P  F  D  G  G  M  K  E  F  G  R  G  V  V  C  N  D  K
gccgacactggcgtcaggaacgccgtcttcgtcatcgaggacggcggtgttctgcgcaac
 A  D  T  G  V  R  N  A  V  F  V  I  E  D  G  G  V  L  R  N
gccatcatcggcgccgacgccgtcgagggcatccactgcgagggcaagtgcaccatcgag
 A  I  I  G  A  D  A  V  E  G  I  H  C  E  G  K  C  T  I  E
aacgtctggttccgcgacgtgtgcgaggacgccatcacgctgaagggcaacggcccttac
 N  V  W  F  R  D  V  C  E  D  A  I  T  L  K  G  N  G  P  Y
accatcactggcggcggcgcccagaacgccggcgacaaggtcatccagcacaacggcaag
 T  I  T  G  G  G  A  Q  N  A  G  D  K  V  I  Q  H  N  G  K
ggcgagctgcgcatctcaaactaccaggtcaacaacgtcggcaagctgttccgcacctgc
 G  E  L  R  I  S  N  Y  Q  V  N  N  V  G  K  L  F  R  T  C
ggcaactgcagcaacaacggcggcccgcgcagcatcgtcgccaccggcatcagggccttc
 G  N  C  S  N  N  G  G  P  R  S  I  V  A  T  G  I  R  A  F
ggcgtcaccagcgacctcatcggcatcaactccaactacggcgacaaggcctccatcacc
 G  V  T  S  D  L  I  G  I  N  S  N  Y  G  D  K  A  S  I  T
ggctcctgcggcaacaccaagacggtctgccaggaatatgtcggcatcgagaagggcgcc
 G  S  C  G  N  T  K  T  V  C  Q  E  Y  V  G  I  E  K  G  A
aacggcgggaaggacagcgagaagagggtcccgccggtcggcgcttgcaccggccagggc
 N  G  G  K  D  S  E  K  R  V  P  P  V  G  A  C  T  G  Q  G
ggtctcgccaggctcccttcctgctag
 G  L  A  R  L  P  S  C  -
```

SEQ ID NO : 24
CHIỀU DÀI : 248
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MQFKYAALAAALAAVGSTQSLPIPASRGVDPNKGQRIVKPGNPFDGGMKEFGRGVVCNDKADTGVRNAVFVIEDG
GVLRNAIIGADAVEGIHCEGKCTIENVWFRDVCEDAITLKGNGPYTITGGGAQNAGDKVIQHNGKGELRISNYQV
NNVGKLFRTCGNCSNNGGPRSIVATGIRAFGVTSDLIGINSNYGDKASITGSCGNTKTVCQEYVGIEKGANGGKD
SEKRVPPVGACTGQGGLARLPSC*

SEQ ID NO : 25
CHIỀU DÀI : 2495 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)

LOẠI : ADN
SINH VẬT : *M. phaseolina*
GTTTGGTGCGTGGGAACGCATTTAACGGTGGATGAATCGATCCAAGGCTTCCAAGGCAGAGCCTCCGAAACAGTC
AATATTCCATCCAAACCAACGCCTATTGGATTCAAGATTTGGGCACTTTGTAACGAGGCTTTTTTGGTTGATTGG
ATGTTCCATTCAAATGGCGAAGGACCGTGGGATCTTGACGAGTATTACACCGACGAACAGTACGAAATTAACCTC
ACACAGACGGCTGCAGTGGTGCCCGATTTGGTGGCTCAGTTATGCCCACATGAGCTGCCATTTGACAGCCCAAAA
TACATCGTTTGGATAGACAATCTGCTTACTTCTGCGCGGCTTATGACAACGCTCCGGAACGAGAATATTGGCGCT
GCAGGGACTGTTCGAATGGGCAAGACACAGCGTGAGAAGAATGAAGAGAAGGCAGCTAGCAAAAAGAAGCAAGCC
ACCAAAGAAGATAACCGTGGTTTACACCAAAGGCTTCGAGATTTGAGGTCAAAAAATGAAGGACAGATCGAATGG
GGCACGCAGTATTGCTGTCTATCACAGGACGAGCAGTCAATGCAATTTGGCTGGCAGGATGCAAGAATTGTGCTA
TTCATGAGCACTGTTCATGACGGTAAACAATGGGTGATTCGACGACGACGGCGGCCCACAAAGACCTCCACCAAT
TCGAAGATCACTCAGAAGCCATTTGGTGATAACGTGGAACAAGATATGGAGATTCCGAGGTGGGCAGACGAGTAC
AATCACAACAAGAATGCGGTTGACCGTTTTGACCAATTCAAAGCCGAACGCCAAATCAATCGACTTTGTTATCGA
ACATGGAAGCCTCTTTGGAACTTCCTCTTCCAATCATCGATCATCAGTGCATATCTGCTCACTTGTAGAGGCTTG
GATGACGATGAGAAACCTCCATTCGCGTCCCTGAACGCATTCCGCCAGAGGCTTTGGGAGCAGCTCTTTGAACGC
TCAGAACGCGTTGACGGACGCCGTGGACGAGCAGAGCCTCGTTATCGTGAGATTCCACGCGAAGAACATCAATAT
GAGAGGCTTAAGGTCCAAGGGAACTGCAGTGAATGTATGGCTAACATGCGCAAATCAACGCGTATTCGAAGGCCA
CTTGCAGAGCGTTCCACCAACATTCGGCCACCAAGGCCAAGGTGGGGCTGTGGGCTATGTAATCTGAATCTTTGT
GAAGGCATTTGCTTCAAAAATCACTTGAATCGGGTGATAGGAAGCAGTTGAATTTGATGTGGAAGTTGGGCTCGA
TTTGGTCGCGAATTCGTCGCGATTTAGATCTAAAACACCAAGATTTGGGCCGTAAATCGAATATATATAAATGCA
AATAGACTCATTTGAATTTGACAGCTTATGCGACCTCAAAACACTACGTTTTAGTTGAATTATTGCGATTATTGG
GCTGCGAAAAGCCCCTGATGACTTTGGTGTTCTGACCCCTGTCACCACCCCTGTCGCTCTCAAGCTCTCCATAGA
TTGAGAGCATATCCAGCCTACGAAGGGATACCACTGCCCCCTCCTGGCGGCTTCCGCCCTGGCTGCTCCTGCCGA
CGACACTTTCGGCTACGAGCTCGTTCGCCGTGCGAACTTCCCTATTCCTGCCTCCAAGGGCACCGTCAAGTACAG
CGCTGCCAAGTCCATCTCCGGCACTTTCGATGGTGGCCTGAAGACCCTACAGCCGTGGTGTCAAGTGCATCGGTC
AGGCTGAGGGCGGTGACAAAAAAGCCGTCTTCATCCCTGAGGACGGTGCTAGCCTCAAGAGCGCCATCACTGGCA
CCGAGCAGATCGAGGGCGTCCACTGCAAGGGGCTCCTGCACTATCGAGAACGTCTGGTGGGCTGGTGTCTGCGAG
GATGCTCTCTCCCTCAAGGGCAACGGCAACGCCAAGATCATTGGTGGCGGTGCCACTGGTGCCGAAGACAAGGGT
AAGCTCTCTTCTGCAAGAGCGTTGCGATCGGATTTCTAACACTCTCTGGTGCAGGTCATCCAGCACCGGTTTTGG
ATCCGTCTCTATTGACGGCTTCACTGTTGCCGATTTCGGCAAGCTCTACCGCTCTTGCGGAAACTGCAAGAAGCA
GGGCAAGAGAACTGTCACCATCAAGAATGTCAAGGCATCATCCGGCAAGCTGCTCGCTGGTATCAACTCCAACTA
CGGCGACACTGCTACCATCACCGGCACCTGCGCTGCCTCCGTCAAGAAGATCTGCACCGAGTTCATGGGCAACAA
CAGCGGCAAGGAGCCCAGTGTGGTTTTCAATCTCGATCTTATCAATCTTATCAATCGATCCCTATCAATCGATAT
CAAGTCCATCATGGAATTGAATTGACGATTGATAAGGATCGACAGCCTCCTGGTTGCCCTGCCAAACGCCACCAA
ACAAACGCCATAAAACACGCCCGCAGGGAGCCGCAGGACCCGAAATGCATCAAACAGTCCTCTTCTATTGCATTA
TCAATTATCGGGCATGGATG

SEQ ID NO : 26
CHIỀU DÀI : 1707
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1707)
atgttccattcaaatggcgaaggaccgtgggatcttgacgagtattacaccgacgaacag
M F H S N G E G P W D L D E Y Y T D E Q
tacgaaattaacctcacacagacggctgcagtggtgcccgatttggtggctcagttatgc
Y E I N L T Q T A A V V P D L V A Q L C
ccacatgagctgccatttgacagcccaaaatacatcgtttggatagacaatctgcttact
P H E L P F D S P K Y I V W I D N L L T
tctgcgcggcttatgacaacgctccggaacgagaatattggcgctgcagggactgttcga
S A R L M T T L R N E N I G A A G T V R
atgggcaagacacagcgtgagaagaatgaagagaaggcagctagcaaaaagaagcaagcc

```
 M  G  K  T  Q  R  E  K  N  E  E  K  A  A  S  K  K  K  Q  A
accaaagaagataaccgtggtttacaccaaaggcttcgagatttgaggtcaaaaaatgaa
 T  K  E  D  N  R  G  L  H  Q  R  L  R  D  L  R  S  K  N  E
ggacagatcgaatggggcacgcagtattgctgtctatcacaggacgagcagtcaatgcaa
 G  Q  I  E  W  G  T  Q  Y  C  C  L  S  Q  D  E  Q  S  M  Q
tttggctggcaggatgcaagaattgtgctattcatgagcactgttcatgacggtaaacaa
 F  G  W  Q  D  A  R  I  V  L  F  M  S  T  V  H  D  G  K  Q
tgggtgattcgacgacgacggcggcccacaaagacctccaccaattcgaagatcactcag
 W  V  I  R  R  R  R  R  P  T  K  T  S  T  N  S  K  I  T  Q
aagccatttggtgataacgtggaacaagatatggagattccgaggtgggcagacgagtac
 K  P  F  G  D  N  V  E  Q  D  M  E  I  P  R  W  A  D  E  Y
aatcacaacaagaatgcggttgaccgttttgaccaattcaaagccgaacgccaaatcaat
 N  H  N  K  N  A  V  D  R  F  D  Q  F  K  A  E  R  Q  I  N
cgactttgttatcgaacatggaagcctctttggaacttcctcttccaatcatcgatcatc
 R  L  C  Y  R  T  W  K  P  L  W  N  F  L  F  Q  S  S  I  I
agtgcatatctgctcacttgtagaggcttggatgacgatgagaaacctccattcgcgtcc
 S  A  Y  L  L  T  C  R  G  L  D  D  D  E  K  P  P  F  A  S
ctgaacgcattccgccagaggctttgggagcagctctttgaacgctcagaacgcgttgac
 L  N  A  F  R  Q  R  L  W  E  Q  L  F  E  R  S  E  R  V  D
ggacgccgtggacgagcagagcctcgttatcgtgagattccacgcgaagaacatcaatat
 G  R  R  G  R  A  E  P  R  Y  R  E  I  P  R  E  E  H  Q  Y
gagaggcttaaggtccaagggaactgcagtgaatcctacgaagggataccactgccccct
 E  R  L  K  V  Q  G  N  C  S  E  S  Y  E  G  I  P  L  P  P
cctggcggcttccgccctggctgctcctgccgacgacactttcggctacgagctcgttcg
 P  G  G  F  R  P  G  C  S  C  R  R  H  F  R  L  R  A  R  S
ccgtgcgaacttccctattcctgcctccaagggcaccgtcaagtacagcgctgccaagtc
 P  C  E  L  P  Y  S  C  L  Q  G  H  R  Q  V  Q  R  C  Q  V
catctccggcactttcgatggtggcctgaagaccctacagccgtggtgtcaagtgcatcg
 H  L  R  H  F  R  W  W  P  E  D  P  T  A  V  V  S  S  A  S
gtcaggctgagggcggtgacaaaaaagccgtcttcatccctgaggacggtgctagcctca
 V  R  L  R  A  V  T  K  K  P  S  S  S  L  R  T  V  L  A  S
agagcgccatcactggcaccgagcagatcgagggcgtccactgcaaggggctcctgcact
 R  A  P  S  L  A  P  S  R  S  R  A  S  T  A  R  G  S  C  T
atcgagaacgtctggtgggctggtgtctgcgaggatgctctctccctcaagggcaacggc
 I  E  N  V  W  W  A  G  V  C  E  D  A  L  S  L  K  G  N  G
aacgccaagatcattggtggcggtgccactggtgccgaagacaagggtcatccagcaccg
 N  A  K  I  I  G  G  G  A  T  G  A  E  D  K  G  H  P  A  P
gttttggatccgtctctattgacggcttcactgttgccgatttcggcaagctctaccgct
 V  L  D  P  S  L  L  T  A  S  L  L  P  I  S  A  S  S  T  A
cttgcggaaactgcaagaagcagggcaagagaactgtcaccatcaagaatgtcaaggcat
 L  A  E  T  A  R  S  R  A  R  E  L  S  P  S  R  M  S  R  H
catccggcaagctgctcgctggtatcaactccaactacggcgacactgctaccatcaccg
 H  P  A  S  C  S  L  V  S  T  P  T  T  A  T  L  L  P  S  P
gcacctgcgctgcctccgtcaagaagatctgcaccgagttcatgggcaacaacagcggca
 A  P  A  L  P  P  S  R  R  S  A  P  S  S  W  A  T  T  A  A
aggagcccagtgtggttttcaatctcgatcttatcaatcttatcaatcgatccctatcaa
 R  S  P  V  W  F  S  I  S  I  L  S  I  L  S  I  D  P  Y  Q
tcgatatcaagtccatcatggaattga
 S  I  S  S  P  S  W  N  -
```

SEQ ID NO : 27
CHIỀU DÀI : 568

LOẠI : PRT
SINH VẬT : *M. phaseolina*

MFHSNGEGPWDLDEYYTDEQYEINLTQTAAVVPDLVAQLCPHELPFDSPKYIVWIDNLLTSARLMTTLRNENIGA
AGTVRMGKTQREKNEEKAASKKKQATKEDNRGLHQRLRDLRSKNEGQIEWGTQYCCLSQDEQSMQFGWQDARIVL
FMSTVHDGKQWVIRRRRRPTKTSTNSKITQKPFGDNVEQDMEIPRWADEYNHNKNAVDRFDQFKAERQINRLCYR
TWKPLWNFLFQSSIISAYLLTCRGLDDDEKPPFASLNAFRQRLWEQLFERSERVDGRRGRAEPRYREIPREEHQY
ERLKVQGNCSESYEGIPLPPPGGFRPGCSCRRHFRLRARSPCELPYSCLQGHRQVQRCQVHLRHFRWWPEDPTAV
VSSASVRLRAVTKKPSSSLRTVLASRAPSLAPSRSRASTARGSCTIENVWWAGVCEDALSLKGNGNAKIIGGGAT
GAEDKGHPAPVLDPSLLTASLLPISASSTALAETARSRARELSPSRMSRHHPASCSLVSTPTTATLLPSPAPALP
PSRRSAPSSWATTAARSPVWFSISILSILSIDPYQSISSPSWN*

SEQ ID NO : 28
CHIỀU DÀI : 1922 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

CAAGATGTGCTTTTGCTCTCCAGGACATTCCTCTCCATCCCTCGTTTCCAATCACTTTCAGTCTTTGCCACTCCT
TCGCTGTCCACTGTAGTCCGGAACTACATCAGTAAACCACCTTCTTTTCCAACAGACCCCATCACATCCGCCAAA
ATGGTCAACATTGCTGCCGTCTGCCGCGCTGCGGCTTTCTTTCTTCCTGTCCTCGCCGCCGCCCAGAGCGGTGTC
GTTGGTACCCCTTCCGGTTTCGCCGCCGGCACCACTGGTGGTGGTGACGCTGCCGCCGCCGCTCCCTCCGACGTT
GCTGAGCTCATCTCTTGGCTTGAGGATGAGACTCCCCGTGTCATCCTCATCGACAAGGAGTTCAACTTCCTCGGC
ACGGAGGGCACCAAGAGCGATCAGGGCTGCCGTCCCGACAGCAACAAGTGCCCTGGCAAGGGTGGCCAGGATGCC
ATCAACCAGGCCAACTGGTGCGACAACGGCAACGCTGGCGAGGGTGTCTCGGCCGTCACCGTCAAGTACGACCAG
GCTGCGCTCAAGGGCATCGCTGTCAAGAGCAACAAGTCCATCGTCGGTGTTGGCGACAAGGGTGTCCTCAAGGGC
AAGGGCCTCAGCCTGACTGGTGGCGTTGAGAACGTTATCATCCAGAACATCCACATCACCGACCTGAACCCCGAG
TACATCTGGGGTGGTGATGCCATCACGCTCGCTGGTTCCGACAAGGTCTGGAGTAAGTACTTCCGAACTGTTTCG
ATCCCATACGGTTGGCCTTGAAACCCAAATCGGGAAACGCACTATACCCAATATGGCAATTTTCTACGAACATCG
AGACACCGCAGCAGAACACGAGTTTCACGCCATTGCTTGTTTCTGTGGCTGTCAATGCTGTCTCGAAGTCCCTGT
ATGCTGTTTGGCTTGATGCGCTAGTGCGTTTCCTGAGAGAGGTGCCCACCACACCATCATGCCCCAAGCTCCCTT
ACCCCTTTCCCAGCATCCCACATCATGCTATATCCTTTACTATGCTTTCAACAGTAAAACTAACGCCCGAAACGT
CGATCATGTCAAAATTTCTCTCGTCGGTCGCCAGATGATCGTTACCGGCTACGACAAGGCCGGTAGGGTCACGAT
CTCCAACTCCGAGTTCGACGGCCAGACCGACTGGTCCGCCTCCTGCAACGGCGAACACTACTGGACCGTCCTCGT
GTACGGTGCCGACGACCAGATCACCTTCGCAAACAACTACATGCACGACACCTCGGGCCGTTCGCCCAAGATCGG
TGGTGCCGACGGCGCTGGCATCACCTTCCACGCGGTCAACAACGTCTTCCAGACCAACAAGGGCCATAACTTCGA
CATCGGCTCCGGTGCCGAGGTTTTGCTCGAGGGTAACACCTTCAACGAGTGCAACCAGCCCATCACCTCCAAGTC
GGCCAACGAGGGCGGCGCCATCTTCAACACCCCCTCCGGCTCCGAGAGCTCCTGCTCCAGCTACCTCAAGCGCGC
CTGCCAGGCCAACACTCTCACCAGCTCCGGTGACTTCGGCACTTACAAGGACACCTCGGTTCTCGAGGGCTTCAA
CGGCGCCACCTCCATCTGGGAGGCCGTCAGCGCTGAGGAGGCTGCTGCGTCCGTCATCGCCAACGCCGGTATTGG
CAAGCTCCAGGGCGGTGCTAGCAACTCCACTGCCAAGTTCAGTATGTCTTCTCCCTTCAACCCTGTATCACAGGA
GACATGTACTAATAATTCTCTCCTCTGCAGGGCGCTTCCAATCCTAGATTAATAAGCACGATCGACCGAATACTA
GCATCAGTACGGTTTTTAGATAAAAAGGGATCTGTCTGGTTTACTGTAGATATTTAGCCTTCCTTTCTTTGGCTT
CACATATTTTTAACTTTACAGCCCAAGAGGGCTCAATGCAAATACTT

SEQ ID NO : 29
CHIỀU DÀI : 1314
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1314)

atggtcaacattgctgccgtctgccgcgctgcggctttctttcttcctgtcctcgccgcc
M V N I A A V C R A A A F F L P V L A A
gcccagagcggtgtcgttggtacccccttccggtttcgccgccggcaccactggtggtggt

```
 A  Q  S  G  V  V  G  T  P  S  G  F  A  A  G  T  T  G  G  G
gacgctgccgccgccgctccctccgacgttgctgagctcatctcttggcttgaggatgag
 D  A  A  A  A  A  P  S  D  V  A  E  L  I  S  W  L  E  D  E
actccccgtgtcatcctcatcgacaaggagttcaacttcctcggcacggagggcaccaag
 T  P  R  V  I  L  I  D  K  E  F  N  F  L  G  T  E  G  T  K
agcgatcagggctgccgtcccgacagcaacaagtgccctggcaagggtggccaggatgcc
 S  D  Q  G  C  R  P  D  S  N  K  C  P  G  K  G  G  Q  D  A
atcaaccaggccaactggtgcgacaacggcaacgctggcgagggtgtctcggccgtcacc
 I  N  Q  A  N  W  C  D  N  G  N  A  G  E  G  V  S  A  V  T
gtcaagtacgaccaggctgcgctcaagggcatcgctgtcaagagcaacaagtccatcgtc
 V  K  Y  D  Q  A  A  L  K  G  I  A  V  K  S  N  K  S  I  V
ggtgttggcgacaagggtgtcctcaagggcaagggcctcagcctgactggtggcgttgag
 G  V  G  D  K  G  V  L  K  G  K  G  L  S  L  T  G  G  V  E
aacgttatcatccagaacatccacatcaccgacctgaaccccgagtacatctggggtggt
 N  V  I  I  Q  N  I  H  I  T  D  L  N  P  E  Y  I  W  G  G
gatgccatcacgctcgctggttccgacaaggtctggatgcgtttcctgagagagcatccc
 D  A  I  T  L  A  G  S  D  K  V  W  M  R  F  L  R  E  H  P
acatcatgctatatcctttactatgctttcaacagtaaaactaacgcccgaaacgtcgat
 T  S  C  Y  I  L  Y  Y  A  F  N  S  K  T  N  A  R  N  V  D
catgtcaaaatttctctcgtcggtcgccagatgatcgttaccggctacgacaaggccggt
 H  V  K  I  S  L  V  G  R  Q  M  I  V  T  G  Y  D  K  A  G
agggtcacgatctccaactccgagttcgacggccagaccgactggtccgcctcctgcaac
 R  V  T  I  S  N  S  E  F  D  G  Q  T  D  W  S  A  S  C  N
ggcgaacactactggaccgtcctcgtgtacggtgccgacgaccagatcaccttcgcaaac
 G  E  H  Y  W  T  V  L  V  Y  G  A  D  D  Q  I  T  F  A  N
aactacatgcacgacacctcgggccgttcgcccaagatcggtggtgccgacggcgctggc
 N  Y  M  H  D  T  S  G  R  S  P  K  I  G  G  A  D  G  A  G
atcaccttccacgcggtcaacaacgtcttccagaccaacaagggccataacttcgacatc
 I  T  F  H  A  V  N  N  V  F  Q  T  N  K  G  H  N  F  D  I
ggctccggtgccgaggttttgctcgagggtaacaccttcaacgagtgcaaccagcccatc
 G  S  G  A  E  V  L  L  E  G  N  T  F  N  E  C  N  Q  P  I
acctccaagtcggccaacgagggcggcgccatcttcaacacccctccggctccgagagc
 T  S  K  S  A  N  E  G  G  A  I  F  N  T  P  S  G  S  E  S
tcctgctccagctacctcaagcgcgcctgccaggccaacactctcaccagctccggtgac
 S  C  S  S  Y  L  K  R  A  C  Q  A  N  T  L  T  S  S  G  D
ttcggcacttacaaggacacctcggttctcgagggcttcaacggcgccacctccatctgg
 F  G  T  Y  K  D  T  S  V  L  E  G  F  N  G  A  T  S  I  W
gaggccgtcagcgctgaggaggctgctgcgtccgtcatcgccaacgccggtattggcaag
 E  A  V  S  A  E  E  A  A  A  S  V  I  A  N  A  G  I  G  K
ctccagggcggtgctagcaactccactgccaagttcaggcgcttccaatcctag
 L  Q  G  G  A  S  N  S  T  A  K  F  R  R  F  Q  S  -
```

SEQ ID NO : 30
CHIỀU DÀI : 437
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MVNIAAVCRAAAFFLPVLAAAQSGVVGTPSGFAAGTTGGGDAAAAAPSDVAELISWLEDETPRVILIDKEFNFLG
TEGTKSDQGCRPDSNKCPGKGGQDAINQANWCDNGNAGEGVSAVTVKYDQAALKGIAVKSNKSIVGVGDKGVLKG
KGLSLTGGVENVIIQNIHITDLNPEYIWGGDAITLAGSDKVWMRFLREHPTSCYILYYAFNSKTNARNVDHVKIS
LVGRQMIVTGYDKAGRVTISNSEFDGQTDWSASCNGEHYWTVLVYGADDQITFANNYMHDTSGRSPKIGGADGAG
ITFHAVNNVFQTNKGHNFDIGSGAEVLLEGNTFNECNQPITSKSANEGGAIFNTPSGSESSCSSYLKRACQANTL
TSSGDFGTYKDTSVLEGFNGATSIWEAVSAEEAAASVIANAGIGKLQGGASNSTAKFRRFQS*

SEQ ID NO : 31
CHIỀU DÀI : 1326 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

ACCAATCGTCACAGATCCGCTGGACACGATGCAGCGTGCACATATAGATCGGCTCCGCAGCGTTAATGCTGTCCT
CCAGCAGCAGCCTCAACTTTCTCTTCACCTTTCTCTTCACCTTTCTCTTCACCACCAGCCCAGCCCACCCTCACC
ATGAAGTTTCTTTCCACTTTGGCCCTGGCGGCCCTCGCGCACCTGGCGTGCGCCGTCCCCATGGCCGAGAAGCGT
GCCCCCGCGGCCACGCCCGTCGGCTACGCCTCGCAGAACGGCGGCGTGACCGGCGGTCAGGGCGGCACCACCACC
ACCGTCTCCTCGCTGCCCGAGATGACGGCTGCCCTTAAGAAGGGGGACACGGAGAAGAAGGTCGTCTACGTCAAG
GGCAAGATCTCCGGCAAGGCGAAGATCTACGTTGGCTCGAACAAGGTGCGGCCTCACCCCTTCACTCCATGACAT
CCGGGTAGCTAACGCAGCCTGCAGTCCATCCTCGGCGTTGACTCCAGCTCCGGCCTCGAGGGCATCGGACTCCTC
GTCCGTGACGCCAAGAACGTCATCATCCGCAACCTGGCCATCTCCAAGGTCGAGGCCGACACGGGCGGCGACGCC
ATCGCCATCGACGGCTCCACCAACGTCTGGGTCGACCACTGCGACCTGTCCAGCGACCTGGCCGCCGACAAGGAC
TTCTACGACGGCCTGCTCGACATCTCCCACGGCGCCGACTACGTCACCGTCTCGAACGTCTACTTCCACGACCAC
CACAAGAACTCGCTCGTCGGCCACTCCGACTCCAACGCCGGCGAGGACACGGGCAAGCTGCACGTCACCTACGCC
AACAACTACTGGAGCAACGTCGGCTCGCGCTGCCCGCTCGTGCGCTTCGGCACCGTCCACATCGTCAACAACTAC
TTCGAGGACGTCTCCGTCTCCGGCATCAACACCCGCATGGGCGCCCAGGTCCTCGTCGAGAACAACGTCTTCAAC
AACGTCGTGCAGGCCCTCGTCTCGATCGACTCCAAGGAGCTCGGCTACGCCGTCGCGCGAGGCAACGACTGGGGC
ACCTCCAAGAACGAGGCCCCCGAGGGTACCCTTACCAAGGTGCCTTACACCTACACCGCTGTTGAAGCGAGCGCG
GTCAAGGCTGCTGTTGTTGGCAGCGCGGGCAACACCCTCTCTGGCTTGTAAACATTTGACCTCGACGGGTTGCTC
GAGAGAAGCCTATCTACATAGCCTGTGTTTACAAGCAATTCTCTGTAGATATTTTTATACAATCTCTAATTGTAA
ATACGAAATCAAATCCGAATTAGTGGAATGCTTGCCTTAAGCTCTTCGGAG

SEQ ID NO : 32
CHIỀU DÀI : 972
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(972)

atgaagtttctttccactttggccctggcggccctcgcgcacctggcgtgcgccgtcccc
M K F L S T L A L A A L A H L A C A V P
atggccgagaagcgtgcccccgcggccacgcccgtcggctacgcctcgcagaacggcggc
M A E K R A P A A T P V G Y A S Q N G G
gtgaccggcggtcagggcggcaccaccaccaccgtctcctcgctgcccgagatgacggct
V T G G Q G G T T T T V S S L P E M T A
gcccttaagaaggggacacggagaagaaggtcgtctacgtcaagggcaagatctccggc
A L K K G D T E K K V V Y V K G K I S G
aaggcgaagatctacgttggctcgaacaagtccatcctcggcgttgactccagctccggc
K A K I Y V G S N K S I L G V D S S S G
ctcgagggcatcggactcctcgtccgtgacgccaagaacgtcatcatccgcaacctggcc
L E G I G L L V R D A K N V I I R N L A

atctccaaggtcgaggccgacacgggcggcgacgccatcgccatcgacggctccaccaac
I S K V E A D T G G D A I A I D G S T N
gtctgggtcgaccactgcgacctgtccagcgacctggccgccgacaaggacttctacgac
V W V D H C D L S S D L A A D K D F Y D
ggcctgctcgacatctcccacggcgccgactacgtcaccgtctcgaacgtctacttccac
G L L D I S H G A D Y V T V S N V Y F H
gaccaccacaagaactcgctcgtcggccactccgactccaacgccggcgaggacacgggc
D H H K N S L V G H S D S N A G E D T G
aagctgcacgtcacctacgccaacaactactggagcaacgtcggctcgcgctgcccgctc

```
 K  L  H  V  T  Y  A  N  N  Y  W  S  N  V  G  S  R  C  P  L
gtgcgcttcggcaccgtccacatcgtcaacaactacttcgaggacgtctccgtctccggc
 V  R  F  G  T  V  H  I  V  N  N  Y  F  E  D  V  S  V  S  G
atcaacacccgcatgggcgcccaggtcctcgtcgagaacaacgtcttcaacaacgtcgtg
 I  N  T  R  M  G  A  Q  V  L  V  E  N  N  V  F  N  N  V  V
caggccctcgtctcgatcgactccaaggagctcggctacgccgtcgcgcgaggcaacgac
 Q  A  L  V  S  I  D  S  K  E  L  G  Y  A  V  A  R  G  N  D
tggggcacctccaagaacgaggcccccgagggtacccttaccaaggtgccttacacctac
 W  G  T  S  K  N  E  A  P  E  G  T  L  T  K  V  P  Y  T  Y
accgctgttgaagcgagcgcggtcaaggctgctgttgttggcagcgcgggcaacaccctc
 T  A  V  E  A  S  A  V  K  A  A  V  V  G  S  A  G  N  T  L
tctggcttgtaa
 S  G  L  -
```

SEQ ID NO : 33
CHIỀU DÀI : 323
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MKFLSTLALAALAHLACAVPMAEKRAPAATPVGYASQNGGVTGGQGGTTTTVSSLPEMTAALKKGDTEKKVVYVK
GKISGKAKIYVGSNKSILGVDSSSGLEGIGLLVRDAKNVIIRNLAISKVEADTGGDAIAIDGSTNVWVDHCDLSS
DLAADKDFYDGLLDISHGADYVTVSNVYFHDHHKNSLVGHSDSNAGEDTGKLHVTYANNYWSNVGSRCPLVRFGT
VHIVNNYFEDVSVSGINTRMGAQVLVENNVFNNVVQALVSIDSKELGYAVARGNDWGTSKNEAPEGTLTKVPYTY
TAVEASAVKAAVVGSAGNTLSGL*

SEQ ID NO : 34
CHIỀU DÀI : 1320 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

TCTCCAAGTTCATCTTCATCTTCAGTCTTTCTCTATTAAACGCGTTGACTCGCGCCATTCTTTGACAGCCATCAG
CCTGCTGAGCACGCTTGACCACATATTCCTTCCCTTGAAAGTTCCCGTTTGCTCAACAACACTTCCACCGCAACC
ATGAAGGCCACCACCCTCGCTACTTTTGCTCTTTCCGCCGTGAGCGCTCTGGCAGCGCCCACCAAGACCTTCGCC
AAGCGTGCCGCCATCACTGACGGCCCCGACGTCGGCTATGCCACCCTGAACGGTGGCACCACTGGCGGTGCTGGC
GGTTCCACCACCACCGTCTCCAGCCTGGCCCAGTTCACCGCTGCCGCCGAGGCTGATGGCGCCGCCGTCATTGTC
GTCTCCGGCAACATCTCTGGCGCCGCCAAGGTCCGCGTCGGCAGCGACAAGACCATCATCGGCAAGGACTCCAGC
GCCGTGCTCGAGGGCGTCGGTCTCTACATCAACAAGCAGAAGAACGTCATCGTCCGCAACCTGTCCATCAAGAAT
GTGCTTGCGGAGAATGGTGACGCCATCGGCATCCAGGCTTCCCAGAACGTCTGGATCGACCACTGCGACCTGTCC
TCGGACCGTGACCACGACAAGGACTACTACGACGGTCTCCTCGACGTGACCCACGCCTCTGACTACATCACCCTG
TCCAACAACTACCTCCATGACCACTGGAAGGCCTCGCTCGTCGGCCACTCCGACTCCAACGGCTCTGAGGACAAG
GGCCACTTGACCGTCACCTACTACCAGAACCACTTCGAGAACCTGAACTCGCGTGGTCCTTCTTTCCGCTTCGGC
ACCGGTCACATTGTATGTTGCCCACCAACCTGACACCACAGTAATGATTCTAACGGCAAATGCTAGGTCAACAAC
CTCTACACCAGCGTCAGCGACGGTATCAACGCTCGCCAGGGCGCCCAGCTCCTTGTTGAGGGCAACGTCTTCACT
GGCAGCAAGAAGCCGCTGTACAGCACCGATGCCGGCTACGCCGCCGTCAACGACAACGACTTCGGCGGCGCTGAG
AACACTGCTGAGGCTGGCACCCTCACTGCCTCTGACCTCGGCTACAAGTATACCGCTCTGAAGTCTTCCGAGGTC
TCTGCCGCCGTCTCCAAGTCGGCTGGCGCCACGCTCACTTTCTAAACGAAGTCGCGACGCTCTCTCGTTTTGGAA
CTCGCCGCAATGCCCGAGCCAGGGTTCGGATGCATGGAAGGAATGCATTGTGATGGATGAAGAGCAACATACTTT
TTTGGGGGGGAAAAAGCACAGGGCGTGCTACTTTTTCATGTACAT

SEQ ID NO : 35
CHIỀU DÀI : 966
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS

VỊ TRÍ : (1)……(966)

```
atgaaggccaccaccctcgctacttttgctctttccgccgtgagcgctctggcagcgccc
 M  K  A  T  T  L  A  T  F  A  L  S  A  V  S  A  L  A  A  P
accaagaccttcgccaagcgtgccgccatcactgacggccccgacgtcggctatgccacc
 T  K  T  F  A  K  R  A  A  I  T  D  G  P  D  V  G  Y  A  T
ctgaacggtggcaccactggcggtgctggcggttccaccaccaccgtctccagcctggcc
 L  N  G  G  T  T  G  G  A  G  G  S  T  T  T  V  S  S  L  A
cagttcaccgctgccgccgaggctgatggcgccgccgtcattgtcgtctccggcaacatc
 Q  F  T  A  A  A  E  A  D  G  A  A  V  I  V  V  S  G  N  I
tctggcgccgccaaggtccgcgtcggcagcgacaagaccatcatcggcaaggactccagc
 S  G  A  A  K  V  R  V  G  S  D  K  T  I  I  G  K  D  S  S
gccgtgctcgagggcgtcggtctctacatcaacaagcagaagaacgtcatcgtccgcaac
 A  V  L  E  G  V  G  L  Y  I  N  K  Q  K  N  V  I  V  R  N
ctgtccatcaagaatgtgcttgcggagaatggtgacgccatcggcatccaggcttcccag
 L  S  I  K  N  V  L  A  E  N  G  D  A  I  G  I  Q  A  S  Q
aacgtctggatcgaccactgcgacctgtcctcggaccgtgaccacgacaaggactactac
 N  V  W  I  D  H  C  D  L  S  S  D  R  D  H  D  K  D  Y  Y
gacggtctcctcgacgtgacccacgcctctgactacatcaccctgtccaacaactacctc
 D  G  L  L  D  V  T  H  A  S  D  Y  I  T  L  S  N  N  Y  L
catgaccactggaaggcctcgctcgtcggccactccgactccaacggctctgaggacaag
 H  D  H  W  K  A  S  L  V  G  H  S  D  S  N  G  S  E  D  K
ggccacttgaccgtcacctactaccagaaccacttcgagaacctgaactcgcgtggtcct
 G  H  L  T  V  T  Y  Y  Q  N  H  F  E  N  L  N  S  R  G  P
tctttccgcttcggcaccggtcacattgtcaacaacctctacaccagcgtcagcgacggt
 S  F  R  F  G  T  G  H  I  V  N  N  L  Y  T  S  V  S  D  G
atcaacgctcgccagggcgcccagctccttgttgagggcaacgtcttcactggcagcaag
 I  N  A  R  Q  G  A  Q  L  L  V  E  G  N  V  F  T  G  S  K
aagccgctgtacagcaccgatgccggctacgccgccgtcaacgacaacgacttcggcggc
 K  P  L  Y  S  T  D  A  G  Y  A  A  V  N  D  N  D  F  G  G
gctgagaacactgctgaggctggcaccctcactgcctctgacctcggctacaagtatacc
 A  E  N  T  A  E  A  G  T  L  T  A  S  D  L  G  Y  K  Y  T
gctctgaagtcttccgaggtctctgccgccgtctccaagtcggctggcgccacgctcact
 A  L  K  S  S  E  V  S  A  A  V  S  K  S  A  G  A  T  L  T
ttctaa
 F  -
```

SEQ ID NO : 36
CHIỀU DÀI : 321
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MKATTLATFALSAVSALAAPTKTFAKRAAITDGPDVGYATLNGGTTGGAGGSTTTVSSLAQFTAAAEADGAAVIV
VSGNISGAAKVRVGSDKTIIGKDSSAVLEGVGLYINKQKNVIVRNLSIKNVLAENGDAIGIQASQNVWIDHCDLS
SDRDHDKDYYDGLLDVTHASDYITLSNNYLHDHWKASLVGHSDSNGSEDKGHLTVTYYQNHFENLNSRGPSFRFG
TGHIVNNLYTSVSDGINARQGAQLLVEGNVFTGSKKPLYSTDAGYAAVNDNDFGGAENTAEAGTLTASDLGYKYT
ALKSSEVSAAVSKSAGATLTF*
```

SEQ ID NO : 37
CHIỀU DÀI : 1896 (bao gồm 150 bp 5 ′ UTR và 150 bp 3 ′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

CAGCCAATACGGTGGCCTGAAACATACTCATATTCCTATATCCAATTCCACCCAGAAACATATTTACACGGCCCG
TCCCTTAGCCCAATTGACCCCGGAAGCGTCTCTCTTTTTCCCTTACGCTCGTTTCTTTTTCTCCCTGTCATGCTG
ATGGCGTCCTGTCTTCCCCTGGCCCTAGTGGCCTTTTCTTCACTTTCCTATGCTCTTCAGCCCTACAGTGGCCCG
GCGCCAGACCTTCCTGATCGTAAGCCATTCGGCTTCGGTGCCGCCGCCACAGGTGGCGGTACCCCCACGCCGAAC
AATACCTACCTGGTCGACAACATGCCTGATCTGCGGACCGTTCTAAAGATGGAGACTCCCCGCACAGTTTACGTG
AAGGGCGAGATCAAAGGCAACGAAATCAACGAGACCACGACCGGCAATTGCCAATTCTACATTGACAGCAGCAAC
AACTCCAAGTTCAACTTTACCCTCTACATCCAGAGCTACAACGAGACCTACATGGGCCAGGTTAAGGCTGCCTCC
GAAGCGGGCCAGCTATTCGACGGACAGAACGCCACCGAGCTCCTCAATCTACTCGGCCGCCAGAACGTGAGCTTC
CATGTGCCCGGAAAGGTTTCAAAATAACAGCAACGCTAACCCACCTCGCAGGGCTGGCGCGGCCAAGTGCAAAAC
GTGCAAAAGTCCTACGAAGCCATTGACGTGGCCAGCAACCTTACGCTTATCGGCTGGGACTCGTCCGCGTACCTG
AATGGCGTCAATCTCGGGTTCAACTCGCGCTCCAACATCATCATGCGCAACCTGCGCATTTCGTCGCCGCGTGAC
TGCTTCCCGTCGCCGGAAACCTACCCCAGCAGCTGGAACGCGCGTTACGACGCCGTGTCCATGGTGTCGACGACG
ACCGCGTGGCTCGACGGCAACATCTTCGAGGACGGCCCCGTCGCCGTCGCGCCCGACGACTTCGCGGGCGGGTGG
AAGGTCGACCGGTACGACGGCTTGTTTGACGCCGAGGACGGCAGCGACGACATCACCTTCTCGCACAACATCGTC
ACCAACCACCACAAGAGCCTGCTGTGGGGCGGCGGCAACAAGGAGGCGGACCGCGACATCGGCAAGATGAAGTTC
ACCGTCTTCGGCAACCGCTTCGTCGACAGCTTGTCGCGCAACCCGCTGATGCGCTTCGGCACCTTCTACATCGTC
GCCAACGTCTTCGAGAACTACGCCGAGCGCGCGCCGCTGTTCGAGGACGACAGCGTCGCCGCGAGCGCGGCCGCC
TCGCGGCTCGTTCGACGCGCGGACAATGCCACCACCTACAGGCCGGACTTCCAGTACAACATGGGCATCTACAAC
GCCAGCACCGTGTACGTCGCGGCGAATGCGTTTGTCCAGACGGGCACCTATGCGGACGATTCGTCGCGCGTCTTC
TCCTTCTCGGACCTTGCGACGCCGGGCCTGCCCGCGACGCTGTGCTCGCCGGCTGATGGCGCGAACGGGACGGCG
TCGGCGTCGTCCCTGCCGAAGAGCGTTTTCAATGGCAGGCCTATCGACCTTGCGAAGAATGCCAAGAACACGTGG
GCGTATTTCCTGGAGAGCAAGAATGAGGACGGCGAGTTCGCCGAGGGTGGGTTCGTGATTGGATGTGACGGGCTG
GAGGAGCAGGAGACGCCGGTTGCGTTTAAGGATGGTAACGAGGTCGATGCATATGTGCGGAAGAATGCAGGGCAG
GTAGGGAGGGCGACGCCGTAGGGGTGCAAGTGGATGCCAGGCTGATGGTTGGTGGTGGAATGTTTGTGGAACCGC
GAGCGGATAGACTTGGCGGGAAGTGCCTCCTCTTGGCGGAAGGTATGCATGTACAATGGGTCAACGGTAGAGATC
TGCGGCTTTTCCGGCGGCGGG

SEQ ID NO : 38
CHIỀU DÀI : 1536
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1536)

```
atggcgtcctgtcttcccctggccctagtggccttttcttcactttcctatgctcttcag
 M  A  S  C  L  P  L  A  L  V  A  F  S  S  L  S  Y  A  L  Q
ccctacagtggcccggcgccagaccttcctgatcgtaagccattcggcttcggtgccgcc
 P  Y  S  G  P  A  P  D  L  P  D  R  K  P  F  G  F  G  A  A
gccacaggtggcggtacccccacgccgaacaatacctacctggtcgacaacatgcctgat
 A  T  G  G  G  T  P  T  P  N  N  T  Y  L  V  D  N  M  P  D
ctgcggaccgttctaaagatggagactccccgcacagtttacgtgaagggcgagatcaaa
 L  R  T  V  L  K  M  E  T  P  R  T  V  Y  V  K  G  E  I  K
ggcaacgaaatcaacgagaccacgaccggcaattgccaattctacattgacagcagcaac
 G  N  E  I  N  E  T  T  T  G  N  C  Q  F  Y  I  D  S  S  N
aactccaagttcaactttaccctctacatccagagctacaacgagacctacatgggccag
 N  S  K  F  N  F  T  L  Y  I  Q  S  Y  N  E  T  Y  M  G  Q
gttaaggctgcctccgaagcgggccagctattcgacggacagaacgccaccgagctcctc
 V  K  A  A  S  E  A  G  Q  L  F  D  G  Q  N  A  T  E  L  L
aatctactcggccgccagaacggctggcgcggccaagtgcaaaacgtgcaaaagtcctac
 N  L  L  G  R  Q  N  G  W  R  G  Q  V  Q  N  V  Q  K  S  Y
gaagccattgacgtggccagcaaccttacgcttatcggctgggactcgtccgcgtacctg
 E  A  I  D  V  A  S  N  L  T  L  I  G  W  D  S  S  A  Y  L
aatggcgtcaatctcgggttcaactcgcgctccaacatcatcatgcgcaacctgcgcatt
```

```
 N  G  V  N  L  G  F  N  S  R  S  N  I  I  M  R  N  L  R  I
tcgtcgccgcgtgactgcttcccgtcgccggaaacctaccccagcagctggaacgcgcgt
 S  S  P  R  D  C  F  P  S  P  E  T  Y  P  S  S  W  N  A  R
tacgacgccgtgtccatggtgtcgacgacgaccgcgtggctcgacggcaacatcttcgag
 Y  D  A  V  S  M  V  S  T  T  T  A  W  L  D  G  N  I  F  E
gacggccccgtcgccgtcgcgcccgacgacttcgcgggcgggtggaaggtcgaccggtac
 D  G  P  V  A  V  A  P  D  D  F  A  G  G  W  K  V  D  R  Y
gacggcttgtttgacgccgaggacggcagcgacgacatcaccttctcgcacaacatcgtc
 D  G  L  F  D  A  E  D  G  S  D  D  I  T  F  S  H  N  I  V
accaaccaccacaagagcctgctgtggggcggcggcaacaaggaggcggaccgcgacatc
 T  N  H  H  K  S  L  L  W  G  G  G  N  K  E  A  D  R  D  I
ggcaagatgaagttcaccgtcttcggcaaccgcttcgtcgacagcttgtcgcgcaacccg
 G  K  M  K  F  T  V  F  G  N  R  F  V  D  S  L  S  R  N  P
ctgatgcgcttcggcaccttctacatcgtcgccaacgtcttcgagaactacgccgagcgc
 L  M  R  F  G  T  F  Y  I  V  A  N  V  F  E  N  Y  A  E  R
gcgccgctgttcgaggacgacagcgtcgccgcgagcgcggccgcctcgcggctcgttcga
 A  P  L  F  E  D  D  S  V  A  A  S  A  A  A  S  R  L  V  R
cgcgcggacaatgccaccacctacaggccggacttccagtacaacatgggcatctacaac
 R  A  D  N  A  T  T  Y  R  P  D  F  Q  Y  N  M  G  I  Y  N
gccagcaccgtgtacgtcgcggcgaatgcgtttgtccagacgggcacctatgcggacgat
 A  S  T  V  Y  V  A  A  N  A  F  V  Q  T  G  T  Y  A  D  D
tcgtcgcgcgtcttctccttctcggaccttgcgacgccgggcctgcccgcgacgctgtgc
 S  S  R  V  F  S  F  S  D  L  A  T  P  G  L  P  A  T  L  C
tcgccggctgatggcgcgaacgggacggcgtcggcgtcgtccctgccgaagagcgttttc
 S  P  A  D  G  A  N  G  T  A  S  A  S  S  L  P  K  S  V  F
aatggcaggcctatcgaccttgcgaagaatgccaagaacacgtgggcgtatttcctggag
 N  G  R  P  I  D  L  A  K  N  A  K  N  T  W  A  Y  F  L  E
agcaagaatgaggacggcgagttcgccgagggtgggttcgtgattggatgtgacgggctg
 S  K  N  E  D  G  E  F  A  E  G  G  F  V  I  G  C  D  G  L
gaggagcaggagacgccggttgcgtttaaggatggtaacgaggtcgatgcatatgtgcgg
 E  E  Q  E  T  P  V  A  F  K  D  G  N  E  V  D  A  Y  V  R
aagaatgcagggcaggtagggagggcgacgccgtag
 K  N  A  G  Q  V  G  R  A  T  P  -
```

SEQ ID NO : 39
CHIỀU DÀI : 511
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MASCLPLALVAFSSLSYALQPYSGPAPDLPDRKPFGFGAAATGGGTPTPNNTYLVDNMPDLRTVLKMETPRTVYV
KGEIKGNEINETTTGNCQFYIDSSNNSKFNFTLYIQSYNETYMGQVKAASEAGQLFDGQNATELLNLLGRQNGWR
GQVQNVQKSYEAIDVASNLTLIGWDSSAYLNGVNLGFNSRSNIIMRNLRISSPRDCFPSPETYPSSWNARYDAVS
MVSTTTAWLDGNIFEDGPVAVAPDDFAGGWKVDRYDGLFDAEDGSDDITFSHNIVTNHHKSLLWGGGNKEADRDI
GKMKFTVFGNRFVDSLSRNPLMRFGTFYIVANVFENYAERAPLFEDDSVAASAAASRLVRRAADNTTYRPDFQYN
MGIYNASTVYVAANAFVQTGTYADDSSRVFSFSDLATPGLPATLCSPADGANGTASASSLPKSVFNGRPIDLAKN
AKNTWAYFLESKNEDGEFAEGGFVIGCDGLEEQETPVAFKDGNEVDAYVRKNAGQVGRATP*
```

SEQ ID NO : 40
CHIỀU DÀI : 1798 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

CGGGTTTTGGGAATATCTCCGGGCTTCGGTGGACCGTCCTATCCGTCTGAGCATGCCAATATGTTTTCTTGCCTT
TCTGTCCGTATATAAAATGCCAGTCAAGAAGAATTCTCGTGTTTCACACCTTTCTTCATCCGATCTCGCCTCAGT
ATGCGTGCATCAGTCTCAATTACGGTCTCGCTACTGGCATTTGCCCTGAGGGCGTGCGCTGTTGGTGTGGTTGGG
AGTCCGGAGGGATTTGGCTCCGGGACAACAGGTGGAGGTAGTGCCACGCCGGTTTATCCTTCGACCACGAGCGAA
TTGCAGTCCTACTTGGAAGACGATGTTGCCCGAGTTATCGTCCTGAACAAGGAGTATGCAACCAAACAATTGTTT
AATAAGCCTAAAACTGACCATGTCCAGGTTCAACTTCAAGGGCACCACCACCACTGCGACGGGCTGCAGACCCAC
ATCCAACACTTGCCCCGGTAATGGTGGCCAAGATGCAATTAATGCCAACGGTTGGTTAGTATCTGCTGGTTGCTC
TCATCTGGTTCTTTGTGACTAATGTTCAACGCAGGTGCGGAAGTGCTCCGGCGGTCTCCGTGAAGTACGACACGG
CCGCTATTACGCCCATAAATCTCGGCTCTAACAAGTCTCTTATCGGCATTGGCACATCTGGTGTCATCCGCGGAA
AGGGTCTCCGTATCGTCAGTAAAACTAATTCTTGCTATCAGTTGACGAAGAAACTTACAGTTGATGAAATCTCTT
CAGGCCAACGGAGCTAAAAATGTCATCATCCAAAATATTCATATTACCGATCTCAACCCGCAGTACATCTGGGGT
GGAGATGCCATTACCCTGGATGGTGCTGATCTGGTTTGGGTGGACCACGTTAAGGGACGTATACTCCCGAACCCT
CGCTCGAGCATTATTCTGACCTTGCCAACTACAGATCTCTCTCATCGGCCGCCAGATGTTTGTCGCGGGCTACGG
GGCATCTAACAGAGTAACTATCAGCAACACGGAGTTCGACGGCTCAACCTCCTGGAGCGCAACTTGCGACGGGCA
TGTACGTGCCGCCTCCCAGCCCACCACCACCTCCTTTCTTTCACTGGAGTAGCTTAGATGCTAATATGCCTTATC
TTTCCTTCTAGCACTACTGGACTATTTTGCTTTTGGGCTCAAACGACTTGATCACAATGAAAGGCAACTATATCC
ACCACACCAGCGGCCGTGGACCCAAGATCGGAGGCAACACCCTGCTCCACGCCGTAAGTTCCGCCCCACAAACAC
CCTTGCCTCACCAATCGCCTGCTGAAACATGCTATTTCCACAGGTCAATAATGAGTGGTACGCCGTTTCGGGCCA
CGCCTTCGACCTGGGCGAGTCGGGCGCCATGGCCGTTGCCGAGGGCAACGTTTTCCAGAACGTCGTCACGCCGCT
GCTTTCGCCCGTCGAGGGCCGCATCTTCACGGCGCCCAGCACTAGCGCCAACACCGCCTGCGCGACCTACCTTGG
CCGCAACTGCGTCGTCAACGCCTTCGGCAGCAGCGGCACCTTCTCCGGAACCGACACCAGCTTCTTCTCCAACTT
CAACGGCAAGACGATCGCGAGTGCAGCCGCCGCCGCATACAAGGATTCGACAGCGGGTGTGGGGAAGATTTAAGG
GCGGAGCGGGTGGTTTTGCATTTAGGGGCTGTGATATTGCGTTGGATGTTGGAATGAGGTCAGCAAAAGAAAATC
TAAAATCAACGATCATGAATGCTGGCTGGAAGGAAATTGTTTGAGAACTACTTGTGCTTTTTTTTtATTACTG

SEQ ID NO : 41
CHIỀU DÀI : 1137
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1137)

```
atgcgtgcatcagtctcaattacggtctcgctactggcatttgccctgagggcgtgcgct
 M  R  A  S  V  S  I  T  V  S  L  L  A  F  A  L  R  A  C  A
gttggtgtggttgggagtccggagggatttggctccgggacaacaggtggaggtagtgcc
 V  G  V  V  G  S  P  E  G  F  G  S  G  T  T  G  G  G  S  A
acgccggtttatccttcgaccacgagcgaattgcagtcctacttggaagacgatgttgcc
 T  P  V  Y  P  S  T  T  S  E  L  Q  S  Y  L  E  D  D  V  A
cgagttatcgtcctgaacaaggagttcaacttcaagggcaccaccaccactgcgacgggc
 R  V  I  V  L  N  K  E  F  N  F  K  G  T  T  T  T  A  T  G
tgcagacccacatccaacacttgccccggtaatggtggccaagatgcaattaatgccaac
 C  R  P  T  S  N  T  C  P  G  N  G  G  Q  D  A  I  N  A  N
ggttggtgcggaagtgctccggcggtctccgtgaagtacgacacggccgctattacgccc
 G  W  C  G  S  A  P  A  V  S  V  K  Y  D  T  A  A  I  T  P
ataaatctcggctctaacaagtctcttatcggcattggcacatctggtgtcatccgcgga
 I  N  L  G  S  N  K  S  L  I  G  I  G  T  S  G  V  I  R  G
aagggtctccgtatcttgatgaaatctcttcaggccaacggagctaaaaatgtcatcatc
 K  G  L  R  I  L  M  K  S  L  Q  A  N  G  A  K  N  V  I  I
caaaatattcatattaccgatctcaacccgcagtacatctggggtggagatgccattacc
 Q  N  I  H  I  T  D  L  N  P  Q  Y  I  W  G  G  D  A  I  T
ctggatggtgctgatctggtttgggtggaccacatctctctcatcggccgccagatgttt
 L  D  G  A  D  L  V  W  V  D  H  I  S  L  I  G  R  Q  M  F
gtcgcgggctacggggcatctaacagagtaactatcagcaacacggagttcgacggctca
```

```
 V  A  G  Y  G  A  S  N  R  V  T  I  S  N  T  E  F  D  G  S
acctcctggagcgcaacttgcgacgggcatcactactggactattttgcttttgggctca
 T  S  W  S  A  T  C  D  G  H  H  Y  W  T  I  L  L  L  G  S
aacgacttgatcacaatgaaaggcaactatatccaccacaccagcggccgtggacccaag
 N  D  L  I  T  M  K  G  N  Y  I  H  H  T  S  G  R  G  P  K
atcggaggcaacaccctgctccacgccgtcaataatgagtggtacgccgtttcgggccac
 I  G  G  N  T  L  L  H  A  V  N  N  E  W  Y  A  V  S  G  H
gccttcgacctgggcgagtcgggcgccatggccgttgccgagggcaacgttttccagaac
 A  F  D  L  G  E  S  G  A  M  A  V  A  E  G  N  V  F  Q  N
gtcgtcacgccgctgctttcgcccgtcgagggccgcatcttcacggcgcccagcactagc
 V  V  T  P  L  L  S  P  V  E  G  R  I  F  T  A  P  S  T  S
gccaacaccgcctgcgcgacctaccttggccgcaactgcgtcgtcaacgccttcggcagc
 A  N  T  A  C  A  T  Y  L  G  R  N  C  V  V  N  A  F  G  S
agcggcaccttctccggaaccgacaccagcttcttctccaacttcaacggcaagacgatc
 S  G  T  F  S  G  T  D  T  S  F  F  S  N  F  N  G  K  T  I
gcgagtgcagccgccgccgcatacaaggattcgacagcgggtgtggggaagatttaa
 A  S  A  A  A  A  A  Y  K  D  S  T  A  G  V  G  K  I  -
```

SEQ ID NO : 42
CHIỀU DÀI : 378
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MRASVSITVSLLAFALRACAVGVVGSPEGFGSGTTGGGSATPVYPSTTSELQSYLEDDVARVIVLNKEFNFKGTT
TTATGCRPTSNTCPGNGGQDAINANGWCGSAPAVSVKYDTAAITPINLGSNKSLIGIGTSGVIRGKGLRILMKSL
QANGAKNVIIQNIHITDLNPQYIWGGDAITLDGADLVWVDHISLIGRQMFVAGYGASNRVTISNTEFDGSTSWSA
TCDGHHYWTILLLGSNDLITMKGNYIHHTSGRGPKIGGNTLLHAVNNEWYAVSGHAFDLGESGAMAVAEGNVFQN
VVTPLLSPVEGRIFTAPSTSANTACATYLGRNCVVNAFGSSGTFSGTDTSFFSNFNGKTIASAAAAAAYKDSTAGV
GKI*
```

SEQ ID NO : 43
CHIỀU DÀI : 1631 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
GGTGGTGGTCTTGCACCTCTTCCGGTCCACTACTCACATTAACCCGTTCAATCGTTCCCACACATTCACTTATAT
TCCAACTCTCACACATTCGTTCTGTATCACGCAATCAATCTTTTTGTTTTTGTTTTTTGATCTATTTTAGCAAAC
ATGAAGCTCACTCTCTTGTCTATTGTTCTCGCTCTTGCCCACCAGACTCGGGCCACCAGGGTTACGGGCGCTGCC
GAAGGCTTCGCCAAAGGCGTCACAGGGGGTGGCTCTGCTGCCCCTGTCACCCCCAAAAATATTGTTGAGCTGACC
AAATACCTTGCCGACGCTCAGCCCCGTGTCATCATCTTGGACAAGGAGTAAGTCCGCTATCACGGGAGGGCTTCA
ACCGCTCTATCTGGGGGATTTTCTCCAGCCGAACGAACGCTGACATCTATCTTCCTCAGATTCAACTTCAAGGGC
TCAGAGGGAACCGTCAAGGAGATGGGCTGCCGTCCTCGTACCAAATGCTCCGAGACTGGCGGTGGCCAAGATGCC
ATCAACCACGCTAGCTGGTGCACAAACGGCAACGCTGGTGCCGGCTCCAAAGCCGTTTCTGTCACCTATGATAGA
GCGGGAACCCTCGCCTTAAAGGTAAACTCCAACAAATCCATTGTCGGTGTCGGCAACAAGGGTGTCATCCGCGGC
AAGGGTCTCCGTCTCGCCGGCGCTAAGAACGTCATCATCCAAAACATCCACATTACCGAGCTTAACCCTCAGTAC
ATCTGGGGCGGTGACGGCATTCAGATCGACGGCGCCGACATGGTCTGGATCGACAGGTCCAAGATCTCTCTCGTG
GGCCGGCAGATGCTTGTCCTTGGCAACGGCGCGTCCGGCAGGGTCTCGGTCACCAACAACGAATTCGACGGCTCT
ACTAGTATGCTGATCTCCATAGCCCCCTTGATCACCAAACGGCTAATCGCTCTTTCCAGGCTGGTCGGCTACCTG
TGACGGTCACCACTACTGGGCCCTCTACTTCACCGGTTCCAACGACATGGTCACGTTCAAGGGTAACTACATCCA
CCACACTTCTGGCCGGAGCCCCAAGATCGGCGGTACCACGCTCCTCCACGCTGTAAGATCCCCAATGATCCGTGT
CTCGTGACTCGCGCTGATTCCACGCCCAGGTCAACAACGTCTGGTATGCCAACTCCGGCCACGCTTTCGAGATCC
TCGCGCCCAAGTCGAACGCCCTCATCGAGGGCAACATCTTCCAGAACGTCGTCCGACCCATGAAGGCCAGCACGG
GCAAGGTCTTTGCTCCCTCTTCGGTCCAGGCCGTCTGCAATCGCGCTCTTGGCCGCGCTTGCCAGGCCAACTCTT
```

TCGGCTCCTCTGGCCCCTTGACCGGCACCGACAGCAGCTTCCTCGGTCTGTTCAAGGGCAAGAACCCCGCTACTG
CTTCCGCCGCCAGCCCCGCCATTTCTGGCCGCGCTGGTGTCGGGAAGATTGCCTAAATGCTGGCTATGTTCTGGG
TTGGACGCTTGCTGAGCAAGCTGTCTTGGCTGGTTGGGTGCATGCACTCAGTTGATAGGCACAACCAACAATCTC
TGGACGTGTTATCTCACTTGTCTGGCACTGAGTGGGAATGATTTATGACAGTTTTT

SEQ ID NO : 44
CHIỀU DÀI : 1137
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1137)

```
atgaagctcactctcttgtctattgttctcgctcttgcccaccagactcgggccaccagg
 M  K  L  T  L  L  S  I  V  L  A  L  A  H  Q  T  R  A  T  R
gttacgggcgctgccgaaggcttcgccaaaggcgtcacaggggtggctctgctgcccct
 V  T  G  A  A  E  G  F  A  K  G  V  T  G  G  G  S  A  A  P
gtcacccccaaaaatattgttgagctgaccaaataccttgccgacgctcagccccgtgtc
 V  T  P  K  N  I  V  E  L  T  K  Y  L  A  D  A  Q  P  R  V
atcatcttggacaaggaattcaacttcaagggctcagagggaaccgtcaaggagatgggc
 I  I  L  D  K  E  F  N  F  K  G  S  E  G  T  V  K  E  M  G
tgccgtcctcgtaccaaatgctccgagactggcggtggccaagatgccatcaaccacgct
 C  R  P  R  T  K  C  S  E  T  G  G  G  Q  D  A  I  N  H  A
agctggtgcacaaacggcaacgctggtgccggctccaaagccgtttctgtcacctatgat
 S  W  C  T  N  G  N  A  G  A  G  S  K  A  V  S  V  T  Y  D
agagcgggaaccctcgccttaaaggtaaactccaacaaatccattgtcggtgtcggcaac
 R  A  G  T  L  A  L  K  V  N  S  N  K  S  I  V  G  V  G  N
aagggtgtcatccgcggcaagggtctccgtctcgccggcgctaagaacgtcatcatccaa
 K  G  V  I  R  G  K  G  L  R  L  A  G  A  K  N  V  I  I  Q
aacatccacattaccgagcttaaccctcagtacatctggggcggtgacggcattcagatc
 N  I  H  I  T  E  L  N  P  Q  Y  I  W  G  G  D  G  I  Q  I
gacggcgccgacatggtctggatcgacaggtccaagatctctctcgtgggccggcagatg
 D  G  A  D  M  V  W  I  D  R  S  K  I  S  L  V  G  R  Q  M
cttgtccttggcaacggcgcgtccggcagggtctcggtcaccaacaacgaattcgacggc
 L  V  L  G  N  G  A  S  G  R  V  S  V  T  N  N  E  F  D  G
tctactagctggtcggctacctgtgacggtcaccactactgggccctctacttcaccggt
 S  T  S  W  S  A  T  C  D  G  H  H  Y  W  A  L  Y  F  T  G
tccaacgacatggtcacgttcaagggtaactacatccaccacacttctggccggagcccc
 S  N  D  M  V  T  F  K  G  N  Y  I  H  H  T  S  G  R  S  P
aagatcggcggtaccacgctcctccacgctgtcaacaacgtctggtatgccaactccggc
 K  I  G  G  T  T  L  L  H  A  V  N  N  V  W  Y  A  N  S  G
cacgctttcgagatcctcgcgcccaagtcgaacgccctcatcgagggcaacatcttccag
 H  A  F  E  I  L  A  P  K  S  N  A  L  I  E  G  N  I  F  Q
aacgtcgtccgacccatgaaggccagcacgggcaaggtctttgctccctcttcggtccag
 N  V  V  R  P  M  K  A  S  T  G  K  V  F  A  P  S  S  V  Q
gccgtctgcaatcgcgctcttggccgcgcttgccaggccaactctttcggctcctctggc
 A  V  C  N  R  A  L  G  R  A  C  Q  A  N  S  F  G  S  S  G
cccttgaccggcaccgacagcagcttcctcggtctgttcaagggcaagaaccccgctact
 P  L  T  G  T  D  S  S  F  L  G  L  F  K  G  K  N  P  A  T
gcttccgccgccagccccgccatttctggccgcgctggtgtcgggaagattgcctaa
 A  S  A  A  S  P  A  I  S  G  R  A  G  V  G  K  I  A  -
```

SEQ ID NO : 45
CHIỀU DÀI : 378

LOẠI : PRT
SINH VẬT : *M. phaseolina*

MKLTLLSIVLALAHQTRATRVTGAAEGFAKGVTGGGSAAPVTPKNIVELTKYLADAQPRVIILDKEFNFKGSEGT
VKEMGCRPRTKCSETGGGQDAINHASWCTNGNAGAGSKAVSVTYDRAGTLALKVNSNKSIVGVGNKGVIRGKGLR
LAGAKNVIIQNIHITELNPQYIWGGDGIQIDGADMVWIDRSKISLVGRQMLVLGNGASGRVSVTNNEFDGSTSWS
ATCDGHHYWALYFTGSNDMVTFKGNYIHHTSGRSPKIGGTTLLHAVNNVWYANSGHAFEILAPKSNALIEGNIFQ
NVVRPMKASTGKVFAPSSVQAVCNRALGRACQANSFGSSGPLTGTDSSFLGLFKGKNPATASAASPAISGRAGVG
KIA*

SEQ ID NO : 46
CHIỀU DÀI : 1573 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*
CGGATGCCGTGCAGCTGTCAAAGTCATGATAAAAGTCCCCACTCCCAGGTACATAATGTGATCGGTAGTTCTGCG
CTCTCCTTTCCAACGGTTGTCCCTTGAGCTTCCCGTGCCATGGTAGCCTTTTCTTCCCTTACGTCTGGCAATGCC
ATGGCGTCCCTCCTGTCTCTATCCGTCCTTTCAGCTTTTCTTGCTTCACCGACACTTGCATACAGTCGTGCTGAA
TGCCAGGCTCCATCTGCCGACCCCCTGGCGGGCTGCGTACCCGGAACGCTGCTCGTCAGTCCAAACAGCACCGTG
ACGGCTAACAACAGCACCTCCTTTACCTCTGTCCAATCAGCTATCCTTTCCCTCGGCAACACCACCACACCCGCC
ACCATCCTCATTCTTCCCGGCACCTATGTTGAGCAAGTCAATATCACCCGCCCGGGACCAGTGACGCTCCTCGGG
CAGACTTCATCCCCGAACAATTCCTCCACCAACGGCGTTAAGATCCTCTGGCGTCAGGCCACTGGCAACTCGGTC
AACACCTTCGACAACGCCTATACCTCTGTCCTCACTGTCGCGCCGACCCTCGAGAGCAGCCTAACAGGCAGCGGG
CCCACCGGATACGCCGTCCCCGCCGGCACGCCCTTCGGCAACGAGGACTTCCGCGCATACAACGTCGACTTCGTC
AACGACTACGCGCCCTACTCGGCCGGCCCGGCACTCGCCATTAGCATCAGCTACGCCAACGCGGGCTTCTACTTC
TGCGGCTTCTACAGCTATCAGGACACCGTAAGTGTCATCTCCACTCGTCCACTACACGCAACAAACAAACATGCA
TCCGAGAATCCCTTTTCCCCACCAACTAACGCCAAACCGCAAGGTCTACATCGGCAAGCTCGGCAACGCCTACTT
CTACGACTCCATCATCGCCGGCCAAACCGACTTCCTCTACGGCTTCGGCACCGCCTGGATCCAGTCATCGCAGCT
CTCTCTCCGCTCCTGCGGCGGCGGCATCACCGCCTGGAAGGGCACCAATACGTCCTTCCCCAACGCCTACGGCGT
CTACATCCACGACAGCGTCGTCGAGAAGGCCAACGCCTCGCTCTCCATCGCCGGGCTCTGCGCCCTCGGCAGGCC
CTGGAATGCGCAGCATCGCTCCATCTTCGCGAACACGTGGCTCGATGACAGCATCAAGCCGAGCGGGTACATCAT
CTGGGGGAGCACGGATCCGAGGACCAACAATTACACGTTTATGGCGGAGTATGAGGACTTTGGGCCGGGTTGGAA
TGAGACCGGGAGGAGGGCGGCGAATATTACGAAGGTGTTGACGGAGGCCGAGTATGAGCCGTATGACAGTCTGGA
GAAGGTGTTCCAGTATCCGTTCAGTGGCGAGTTCGGGAACGTGGGGTGGATCGATGAGAGTCCGGAGGCTTGAGC
CGGTGGGGATAGTGTGGATTGTGGTGGCCAGCGGCGTAGACGAGGCGGATCTGGTAAGAACTGCGTAGCATGAAT
GAAATTCGTTTGTACATACGTCTTTATATCGAGCAGCGCTTATGCAAGTTAAATGACGCCGTATTTGCATGCG

SEQ ID NO : 47
CHIỀU DÀI : 1182
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1182)
atggcgtccctcctgtctctatccgtcctttcagcttttcttgcttcaccgacacttgca
M A S L L S L S V L S A F L A S P T L A
tacagtcgtgctgaatgccaggctccatctgccgacccccctggcgggctgcgtacccgga
Y S R A E C Q A P S A D P L A G C V P G
acgctgctcgtcagtccaaacagcaccgtgacggctaacaacagcacctcctttacctct
T L L V S P N S T V T A N N S T S F T S
gtccaatcagctatcctttccctcggcaacaccaccacacccgccaccatcctcattctt
V Q S A I L S L G N T T T P A T I L I L
cccggcacctatgttgagcaagtcaatatcacccgcccgggaccagtgacgctcctcggg

```
 P  G  T  Y  V  E  Q  V  N  I  T  R  P  G  P  V  T  L  L  G
cagacttcatccccgaacaattcctccaccaacggcgttaagatcctctggcgtcaggcc
 Q  T  S  S  P  N  N  S  S  T  N  G  V  K  I  L  W  R  Q  A
actggcaactcggtcaacaccttcgacaacgcctatacctctgtcctcactgtcgcgccg
 T  G  N  S  V  N  T  F  D  N  A  Y  T  S  V  L  T  V  A  P
accctcgagagcagcctaacaggcagcgggcccaccggatacgccgtccccgccggcacg
 T  L  E  S  S  L  T  G  S  G  P  T  G  Y  A  V  P  A  G  T
cccttcggcaacgaggacttccgcgcatacaacgtcgacttcgtcaacgactacgcgccc
 P  F  G  N  E  D  F  R  A  Y  N  V  D  F  V  N  D  Y  A  P
tactcggccggcccggcactcgccattagcatcagctacgccaacgcgggcttctacttc
 Y  S  A  G  P  A  L  A  I  S  I  S  Y  A  N  A  G  F  Y  F
tgcggcttctacagctatcaggacaccgtctacatcggcaagctcggcaacgcctacttc
 C  G  F  Y  S  Y  Q  D  T  V  Y  I  G  K  L  G  N  A  Y  F
tacgactccatcatcgccggccaaaccgacttcctctacggcttcggcaccgcctggatc
 Y  D  S  I  I  A  G  Q  T  D  F  L  Y  G  F  G  T  A  W  I
cagtcatcgcagctctctctccgctcctgcggcggcggcatcaccgcctggaagggcacc
 Q  S  S  Q  L  S  L  R  S  C  G  G  G  I  T  A  W  K  G  T
aatacgtccttccccaacgcctacggcgtctacatccacgacagcgtcgtcgagaaggcc
 N  T  S  F  P  N  A  Y  G  V  Y  I  H  D  S  V  V  E  K  A
aacgcctcgctctccatcgccgggctctgcgccctcggcaggccctggaatgcgcagcat
 N  A  S  L  S  I  A  G  L  C  A  L  G  R  P  W  N  A  Q  H
cgctccatcttcgcgaacacgtggctcgatgacagcatcaagccgagcgggtacatcatc
 R  S  I  F  A  N  T  W  L  D  D  S  I  K  P  S  G  Y  I  I
tgggggagcacggatccgaggaccaacaattacacgtttatggcggagtatgaggacttt
 W  G  S  T  D  P  R  T  N  N  Y  T  F  M  A  E  Y  E  D  F
gggccgggttggaatgagaccgggaggagggcggcgaatattacgaaggtgttgacggag
 G  P  G  W  N  E  T  G  R  R  A  A  N  I  T  K  V  L  T  E
gccgagtatgagccgtatgacagtctggagaaggtgttccagtatccgttcagtggcgag
 A  E  Y  E  P  Y  D  S  L  E  K  V  F  Q  Y  P  F  S  G  E
ttcgggaacgtggggtggatcgatgagagtccggaggcttga
 F  G  N  V  G  W  I  D  E  S  P  E  A  -
```

SEQ ID NO : 48
CHIỀU DÀI : 393
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MASLLSLSVLSAFLASPTLAYSRAECQAPSADPLAGCVPGTLLVSPNSTVTANNSTSFTSVQSAILSLGNTTTPA
TILILPGTYVEQVNITRPGPVTLLGQTSSPNNSSTNGVKILWRQATGNSVNTFADNYTSVLTVAPTLESSLTGSG
PTGYAVPAGTPFGNEDFRAYNVDFVNDYAPYSAGPALAISISYANAGFYFCGFYSYQDTVYIGKLGNAYFYDSII
AGQTDFLYGFGTAWIQSSQLSLRSCGGGITAWKGTNTSFPNAYGVYIHDSVVEKANASLSIAGLCALGRPWNAQH
RSIFANTWLDDSIKPSGYIIWGSTDPRTNNYTFMAEYEDFGPGWNETGRRAANITKVLTEAEYEPYDSLEKVFQY
PFSGEFGNVGWIDESPEA*
```

SEQ ID NO : 49
CHIỀU DÀI : 1330 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
TCAGGCTCGTATGCCTTGTCCTTCATTTGCTTGCACTCTAGGGCTGGCTAAGACCTCCTTACTTGAATAATGTAT
CTCTCTACGTCCGCCTTTATAGTCTGCCTGGCGGCTCCAAAACCCCTGAGCCTCACAACATTCATTCGCTTAACC
ATGCATAGTTCTTCCATCCTTTTCAGCCTTTTCGGCTCCGCACTCGCTCTGACATCTCCACCGCCAGGAGCGCTG
ACTGTTGGAGGCTCTGAGGGAAAATTCTCCACCGTTCAAGCTGCGGTTGACGCTCTCCAGAACACCACTGCTCGG
CAATCCATATTCATCTACTCCGGCACTTACGAGGAGCAGGTGTATATCGCCAAACACAATGGCCCCATTTCCATT
```

TATGGACAAGCCTCGGATGACAGCTCATACCACACCAATACTGTCACTCTCTCTTTCGGCCTCAGCCAGGCCTTC
AACCTTAGCAACGACCTCACGGCCACCCTCCGCGCTCACTCTCCCGATTTCAACTTGTACAACGTGAATGTGGAG
AACACCTACGGCAAAGGGAGCCAGGCCGTCGCTGTCTCGGCCTACGGTACCGAGCAGGCATACTACGGCTGCAAG
CTCACGGGCTTCCAAGACACCCTCCTCACCCAGAGAGGCAGGCACTACTTCGTAAACACCTACATTGAAGGTAAG
GAGGTCTTTCTCAAGCTGCGGAAAGCGCGGTCTCTGACGGACATGCACAGGCGCCACCGACTTCATTTTCGGCCA
GTATTCCGCGGCCTGGTTTGAGAACTGCGACCTGCGCGTCCCCGCGGCCAAGCAGGGATGGGTCACCGCGAGCGG
CCGCAACACGACCAACGACGGCTGGTATGTCATCAACGGAGGCAGCGTGCAAGCGGCACCCGGCCAGAACGTAAC
TGCCGGCGCGTACTTCCTGGGCCGGCCGTGGCGCAGCTTCGCGCGCGCCGTCTTCCAGAACGTCTACTTGAGCGA
GGTCATCAACCCTGCGGGCTGGGCCATCTGGTCCACGAGCACGCCCAACACGGGCAACGTCACGTTCGCGGAGTA
CAACAACACCGGACCTGGTGCTTCCAACGGGACGCGGGCCAGCTTTTCGCAGCAGCTGGCGGAGCCGGTGGACAT
CACGGAGGTGCTAGGATCGAACTACACACTCTGGATTGATCCCAAGTTCTTGTGAATAGCAAGCGGGTAGAAAGC
CGCGGTACGAACGCCTTTGCCGTAATTAGATAATCCCAGGAAGACGGAGAGAATGATTATCTTCTCCTCTAATGC
GCCGTGGGACGACGGATCCGTGACTGAATAGTGAGAGCAAAAGTGCGCCTGCATG

SEQ ID NO : 50
CHIỀU DÀI : 975
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(975)

atgcatagttcttccatccttttcagccttttcggctccgcactcgctctgacatctcca
 M H S S S I L F S L F G S A L A L T S P
ccgccaggagcgctgactgttggaggctctgagggaaaattctccaccgttcaagctgcg
 P P G A L T V G G S E G K F S T V Q A A
gttgacgctctccagaacaccactgctcggcaatccatattcatctactccggcacttac
 V D A L Q N T T A R Q S I F I Y S G T Y
gaggagcaggtgtatatcgccaaacacaatggccccatttccatttatggacaagcctcg
 E E Q V Y I A K H N G P I S I Y G Q A S
gatgacagctcataccacaccaatactgtcactctctctttcggcctcagccaggccttc
 D D S S Y H T N T V T L S F G L S Q A F
aaccttagcaacgacctcacggccaccctccgcgctcactctcccgatttcaacttgtac
 N L S N D L T A T L R A H S P D F N L Y
aacgtgaatgtggagaacacctacggcaaagggagccaggccgtcgctgtctcggcctac
 N V N V E N T Y G K G S Q A V A V S A Y
ggtaccgagcaggcatactacggctgcaagctcacgggcttccaagacaccctcctcacc
 G T E Q A Y Y G C K L T G F Q D T L L T
cagagaggcaggcactacttcgtaaacacctacattgaaggcgccaccgacttcattttc
 Q R G R H Y F V N T Y I E G A T D F I F
ggccagtattccgcggcctggtttgagaactgcgacctgcgcgtccccgcggccaagcag
 G Q Y S A A W F E N C D L R V P A A K Q
ggatgggtcaccgcgagcggccgcaacacgaccaacgacggctggtatgtcatcaacgga
 G W V T A S G R N T T N D G W Y V I N G
ggcagcgtgcaagcggcacccggccagaacgtaactgccggcgcgtacttcctgggccgg
 G S V Q A A P G Q N V T A G A Y F L G R
ccgtggcgcagcttcgcgcgcgccgtcttccagaacgtctacttgagcgaggtcatcaac
 P W R S F A R A V F Q N V Y L S E V I N
cctgcgggctgggccatctggtccacgagcacgcccaacacgggcaacgtcacgttcgcg
 P A G W A I W S T S T P N T G N V T F A
gagtacaacaacaccggacctggtgcttccaacgggacgcgggccagcttttcgcagcag
 E Y N N T G P G A S N G T R A S F S Q Q
ctggcggagccggtggacatcacggaggtgctaggatcgaactacacactctggattgat
 L A E P V D I T E V L G S N Y T L W I D

cccaagttcttgtga
P K F L -

SEQ ID NO : 51
CHIỀU DÀI : 324
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MHSSSILFSLFGSALALTSPPPGALTVGGSEGKFSTVQAAVDALQNTTARQSIFIYSGTYEEQVYIAKHNGPISI
YGQASDDSSYHTNTVTLSFGLSQAFNLSNDLTATLRAHSPDFNLYNVNVENTYGKGSQAVAVSAYGTEQAYYGCK
LTGFQDTLLTQRGRHYFVNTYIEGATDFIFGQYSAAWFENCDLRVPAAKQGWVTASGRNTTNDGWYVINGGSVQA
APGQNVTAGAYFLGRPWRSFARAVFQNVYLSEVINPAGWAIWSTSTPNTGNVTFAEYNNTGPGASNGTRASFSQQ
LAEPVDITEVLGSNYTLWIDPKFL*

SEQ ID NO : 52
CHIỀU DÀI : 7706 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

GCTGCTTCTCTGAGCGGCTCCAAGAACGCCTTTTACAACTGCCAATTCATCTCCGCCGGCGATTTAGGAATATCC
TCGTCCACCGGTACCACGTTCGTCTACAATTCGTACATCGAAGCGCGTTCTAAGCTCATCTACAATACGCCTAAC
ATGTACATTTACGGCTCAACTCTCACGCCTACGGCGAGCAACGCACTCTTGGTTTATAACAAGGGCGTGTCTTCT
GGCACGGTGGTGACCACCAACTCCACCGTTGTCTTCGACTCCTGCACCGTCACTCCGAAGCCGGGCACCACTGTG
ACCGGTATTTCCCTTGCCGCAGCGAACGGTGTTGGGTCAATTGTGGTCTATAGAAACTCCGTCCTGCCAGGCTTC
ATTCTCAGCACCGGTATTCACGTTGACGCCAAGACCCAGGCAGCACAGAACTTCTACGGCGAGTTCGGTAACACA
GGTGCCGGATCCTACTCTTCCAACTCCGCCGCGAGGGCCTCCTACGTGAAGAGCCTGAGCGTCGATCAGCTTTCT
CAATTCTCCGTTGATCAAGTGTTCGGCACCGCAAACACTAACTGGATCGACTCTTCAGTCATCTCCCTCGTCCAG
GATTCGGATGCATCTCAAGCGGCAAAGGCGACCACGGCAGGTGTGGTGTCATACACTACTACGTCTTCCTCTTCA
AGTACTACAAGCTCTGTACTGTCGAGCAGCTTGGTCACCGCATCGGGCAGCTCCGTGTCCACATCATCAAGCTCG
GTGTCGTCTACGTCTTCAACCTCCGCCTCTGCCAGTGACGCGCCTTCGCTTTCCAGCAGTGCTGCTCTCGCTTCA
TCAGTACCTGGATCTTCCTCCAACACTACAGTGACCACCAGCGGCACAGCAAGCTCAAGCTCTGTCACTGGTTCC
TCGTTCGCACTGTCCACTAGCTCGATCTCCTCCAGCGTGAGCACTCCGGGGGACACAGCGAGCTCCACCGCCGCT
GCGTCTAATTCGACTTCGAGTTGCGCTCTTCCTTCCTCAGTGCCCACCACTGCTCGCGTCGTGGGTCCGGCTGGG
TCCTGCGCCAACTACACCAGCATCGCCGATGCCGTCAAAGACCTTTCTACGGATCAATCTAAGACGGAATATGTG
TACATTCTCGCTGGCACCTATACCGAGCAGATCATTTTCAGCCGTGTAGGTCCCACCGTCTTCCGCGGCGAGACA
TCAAGCGAGCTCGATCAGTCCAGCAACAAAGTGACTATCAAGTCTTCTACTGGTGTGCCCTCCAGTAGCGGGGGA
TCTTCCAGTACCGCCCCGTTCCAGGCGACTCAGTACTACAGCAAGAGTATCAGTTTCTATAACATCAATTTCGAG
AACACCTACGCTGCAGCTACGGGTTACAATGCGGTTGCATTGTCCAGCAAAGCTTTGAAGGCCTATTATTACAAC
TGCGGGATCACTTCCTCTCAAGGAGCTCTGCTGCTCAACTTCGGCGCCCACTTCTTCTCGGGCTGCAAGATTACC
GGTACCACCGACATCGTCTGGGGCCAAGGTGGCGCTTACATCTACAATTCGAAGATCGTTAGCACCGGAACCACG
ACTGGTCAGTCTCTCTCCGCACAGTCGTACCAGTCACAATACAATCCTTCGCAGTTTGTCTTCGACACTTGCGCT
TTCGTGCCGAACGACAGCACCGTGCCGAAGGCGAGCACGTACTTGGGCCGCGACTATACAGCTTCGGCACGTGTT
GCGGTGATCAACTCTTACCTGGACGCTCATATTACTCCTGTTGGATGGTTGATTGCGTCCAAGACTACGAACGTG
ACGTTTGTTGAGGCCAGCAACTCCGGGCCTGGAGCGTCGACCGCATCTAGAGTTTCCCAGATTGTGACCGATACG
TCTGCCTACGCCGCCAACAACGTGCTCGGTTCTCTTTCGATCGACACCGCCGCTGTTGCTCCTGTTGCTGCCTTC
CCCGACAGCGTCTACGGTAGTCCTCTTTCCCCCAGCTCTTCTACTGTCTTGGCCAACTCAACGACAGCGACTTCC
GTGTCTGCTTCGACATCCAGCACTTCTGCTGCTGCAACTGCGGCCAACACCCTCATTGTCTCAACCACGCCTGCC
TCTGGCGAATACGGCAACGTCACAGCAGCCATCGCCGCCCTTCCCAACGATAGCAAGGAGTACACCATCTACATC
CGCGCCGGCACCTACCAGGAGCAGTTCACCATTGTGCGGAATGGCAAGGTCACACTGCGTGGCGAGACGGCGTTC
CCCAACGACTTCTCGCAGAACCAGGTCACCATCCAGTTCTCCTATGGTGTTCTCACCAGCGCTGGCCAGAACGAG
TTGACGCCCGTTATCAACGCCAAGAAGAACGATGGCTCCGGCCTCGCGTTGTACAACATCAACTTCATCAACACG
TACCCGCAGACGAAGAACACGGCTGCGCTTGCGGCTGACTTCTACGGTACCAATATGGCTGCCTATGGCTGCAAG

TTCGTCGGCTACCAGGACACCCTCCTCGCCAACAAGGGCACCCAAGTCTTTTCCAACAGCTACATTGAAGGCTCT
ATTGACTATATCTGGGGTTTCTCGACGGCCTACTTCCACCAGTGCTACATCGCATCCAACACAGCTGGTGGCTAC
ATTTCTGCCATGAGCCGCGCTTCGGCTTCGGCTACTGGAGGTTATGTCTTTGACAGCTGCTATGTCACCTACACG
AGCACCTACGGATCGACCTTTGGCACCAGCTACCTCGGTCGCCCGTACTCGAGCTACAGCATCGCCGTCTACATG
AACTCGTTCATTGACAAGCACATTAGCCCCGCCGGTTGGGCTGTGTGGCAAACGAGTAACCCGCAGACGGACAAC
GTCCTGTTTGGAGAGTTCAACAACACCGGCCCCGGAAGCTGGTCCAGCTCTCGCGCGTCTTTCGCGACGAACTTG
ACCGAGTCCCAGGCGGATGCTTACAAGCTGTCTACTTGGATCGGCAGCACGTCGTGGCTTGACATGGATGCTTAC
AACTACGTGCCTTCGTACGATATTTCCGGCGCAGCTGGTGCATCAACCACGACCCCCTCGGCGTCAGCTTCCTCT
ACCACTGCCACAGCCACTGCCACCTGGGCGCATCCCTCGAGTGGTGCTACCCCACCTATCGGTGCCGTCCTGGTG
TCCGTTGGAGGCTCCGTCAACGGGTCATACAGCAATCTTACCGCCGCCCTCGCGTCTCTGCCATCGGATAGCTCG
ACACAGGTCATCTTCATGTACCCCGGAACCTACAACGAGCAGCCCCCGGCCGTAAACAGGCCCGGCCCCATCCAG
ATCATCGGTGCTCAAGATGGCAATCCAGGCCAGAGCTACAAGACCAATAAAGTCATTCTGACCCAGTCTCGCGGC
TTGTCGGTCTCTCCGTTGCCGACCGGACACTCTGACGCTGAAACCGCAACATTCTCCACTACCAGCAACAAAATT
GCCATGTACAACATCGACATCATCAACTCCGATAATCTTGACGGTTCGCTTTCCTCTTACGTGACGCTGGCCGGC
TCAATTTATGGCTCGCGCATTGCGTTCTATGGGTGCAGCTTCATTGGCTGGCAGGATACGCTGTTGACTGGAAGC
ACCAGTGGCTACCAGTACTACGAATCCTGCTACATTGATGGCGCCATCGATTTCATCTGGGGCTACTCTAAGGCT
TACTTCAAAGGCTGCACCATCGGCGCCAAAAGGCAAAAGTCGGCCATCACGGCTCACAGCCGTGCTTCCTCCAGC
GCAGTCGGAGGCTACATCTTCGACCAGTGCCTGTTCACTGCTGCCTCGTCTGCGACGGTTGACCTTACGCAGTCT
GTTTACCTCGGCCGCCCTTACAGCAAGTACGCCCTCGTGGTTGTCAAGAACTCGTACCTGGACAAGACGATCCAG
CCTGCTGGTTGGAAGATCTGGTCTGCCACCGACCCGCGTACCGACTACGTTACCTTCGCCGAGTTCAACAACTCC
GGACCTGGTAACTGGGAGAACAACGCCGCGGCGCGTACCGCTTTCGGCTACTGCACCCTGTTGACGTCGGACACG
TACTCCCTATCTGCCGTCATGGACTCACCCTCTGACTGGATTGACATGACGTACTGGGACTCGATCACCACACCG
ACAGTGGCCGCTGTTGCTACGGGAAACACCACTACGGCGGTCAACGGTACTTCGGTCTATGATGGCACCACCCCA
CCTGCCGGCGCATTGATTGTCTCCAAGACGGCCATTGAGGGCGTGACGACGTACGATACCATCCAGAGCGCTCTG
AATGCGCTGCCTACTTCGAGCAGCAAGACCGGCACCATCTTCATCTACCCCGGTGTGTACTCGGAGCAGCTTGTG
CTGAGCAAGTCTGGTACCACCGTGTTCATCGGCTACTCAAACTCCACCGACGATTACCTGCAGAACCAGGTGACG
ATCGACTTCAACAAGGGAATCGATACGCAAGCAGATGCTTCCAACTCTGACAGCGCTACTGTGTATGCCACAGGC
AACTACTTCCAGGCATACAACATCAACTTCAAGAACTCTTTCGGCACAACCGAGGACTACGCCTCGCTCGGCTTC
GGTGTCAAGTCCAGCAAGTACGCATCGCTGTATGGCTGCCAGGTCTGGGGCAATCAGGACTCGCTACTCATCAAC
GGCTACTTCTTCGCCTTCAACTCGCTGATTGTCGGTAACATCGACATGATCTGGGGCTCCGGCGCGGGCTATTTC
CTCAGCTCAACCATCTCACCCAACACCGACGACGTCAGCCTGACCGCCAGCAAGCGCGCGACCAACACTACAGCT
GCCGGTTTTGTCTTCGACCAGTGCACCGTTAAGCCCGCCCCCGGTACCGGCCCCTTCACCGAGATCAGCCTCGGA
CGTCCGTGGAACAACCTTGCCCGTGTCGCGTATATCGAGACGTATCTCGATTCCAGCGTTGAGGCTGCAGGCTGG
AGCCAGTGGTCCAAGTCCAATCCACAGACCGAAGGAGTGACTTTTGCCGAGTACGGGAACTATGGACCGGGTGCA
AGCACTTCGGGCCGTGCGAAATTTTCCACGCAGCTCTCGGCTGCCGATGCCGCCCAGTTTCAGCTCGCCAACTTC
TTCGCCGTCACCTCCTGGATCAACTTCACGCGGATCGATGTCCAACCCTTCGTGGCCAGCGAGGTTGTTGTTCCA
ACCTCGGCGGTCACTTCGTCTGTCCTTCTCTCTTCCACCCTTAGCACTCCGATCTCGTCTTCAACGCTTATACTG
AGCACCCTCTTCCTCACTAAAGTCACTACCGACAAGGAGACTCTCTTCACAACCGTCACTGGCGCTGTTCCGACC
CTCACTTCAACGCAGACTATCACGCTGGACATGGGTGCCACCGTTACCCCCGACCCAGTGTACAAGACCAGCACC
GTGAAGAGTACGACAACGATCATTGAGACCGTATCGCAGCCCGATGTTACGCAGACGTCTACTGTGGTCGTCACC
AGCGATATCGGAACCACAATCACACCGGAGCCGAGCACGATCACCACTGTGCTCAAGCAGACGACTACTGTTTTC
GCCACCTCGACCAAGGCGCCCCAGACGATTACTGAGAAGAGTACCATTACTTCGACGAGTTTGGCTACCCGGACG
CTCGATCCCATTACGTCAACACTCAGCCTCGGATCTACCGTTTATGTGACCAGCGTTTTCACCCCGAAGGCCGCG
CGGGTCACCTCTAGTCTAACCATCACTACGGGCACTGGCGGCACGTCAACTAAGACGACCAAAGCAACCACGACT
TACGTCACTGTCACTTCTGTGAAAATGACAACGAAGAAATCCACAACAACCCTCTCCTGTGTGCCCACCGACAGT
CTCCAAAGGCGCTCGTACCCCTTAGGTCCCCGCGCCGCCGGTGACAGCGTCACCACATCCACCATTTTCTCCACC
CTCCTCACCTACGTCAAGACCTCTACAGCAACCTTAGCCCCCGGCTCGACCGCCACGACCCTTGTGTCGATCACC
AAGACCGTCGGCTCCACCACCACCCTCAAGCCCACCACCATCACGGTCTCCACTGTCTCCATCGCCACCAAGTCC
TCCACCCTCACCCACCCGGCATCCACCTTCTTCACAACCGTAACCTCGACCAAACTCACCGGCAAAACCACCTCG
CTGAAAGCCTCAACACTGACGGTGACATCCACCTCCCTTGCAGCCACCCGCACCTCCGTCACCACCCTCTCCGCT
GCGGGCGCCGCTACAACGACTTCTCTCGCCACCGTGACGCGGAAATCGACGATAACGCTCCCCGCGAGCACCGTC
GTCAAAACATCTTCGCGGGACGCCACGAGCAGTGTAGTCATCACGGCGGCGGCGAGCACGAGTACGGTGTGGAAG

ACGGCCACGATCACGCTGGCACCGAGCCAGACAGTGGTGTTGCAGAGCACAGTCGTCAAAGAAACGACGGCAAAG
GTAACAAGCACGGTTCTTAGTACGATAACAAAGACGGCGAAAGGGGCGGCGGCGTGTACAGAGGTATAAGTAAAG
ATTGGAAGAGAAGAAGGAGGGGAATAGAGTTGAAAATAAATAAAGAGAATAAATAGTAGCGGGGGTTTTATATAT
TTATAATATTTTAAGTAACTAGTTACTTAATTATTAAATTTCCTAAAGGCCGAATTCTACACCTGCTCACTGCTG
ATGGACGATACGGGCTGTGTTCCACGATCTATAGTCACATGGAAAAGCAGAAGTTGCAAAATTATATTATACTTA
CACGAAATATTTCCCGGAAAATCAGACACCTTAGTGTATTACAGTTGTGCCAAGGTTCTATCTAAGTTGAAACTC
CTTTATCTACTGTGAGAAAGAGACTTGCATAGACTTAGGTATATTTCACGAGAATAACAGCAATCAACATATTTG
TAGCAGTAGAGCATTGCTTGCACAGGATAGGTTCAATTGCCTTTGAAGTTTCTAAGTCACTAGAGCTGCCTCGTT
CAGTCAGAGTCGTTCTTTCTGGCGTTCAGCGAATGACTGCTGGCATCCAAAATCGAGAATGAAGTGATTCCATAT
GTGAGGATGCAATGCGACCATTCAATCGAACAGCGCCTCCGCACTCGCAGCCTGAGCGAGTTCCCCTCTCAATCG
TGTGGATATAATCCCTTGTCGACTGAATCAAGTAAGGGGCATAAGCCTGAGTAGCTCTCTTGTGTCCAGCCACAA
CAATGTGCGGGTTCAAAGCCTCGATGCGATCAATAGCATCAAGCCGCTCCCTGCGTTTGGCCGGGGCGTTTGCCT
CCCCGAAATGTGGATGACAATCGCCGTAGATGATATCGCCAGAGACAACGAGACGGACGGAAGGGACGTGGGGGA
AGCTGGAAAACTCGTTGTCGGTGTGCCCAACTTCAATTCCGTGCCATTTATGGCCGTTCGACGGAGAACTCCCCC
GACGCTGGAAGCGCCTTGGGCACTGGCTTGGCAGTATCGAGCTGGCCGGGGAACATGCCAGGCCAGATGTGCTCG
CAACTCGTATTGCATACGGGATAATACAACGTGCACACTCTTGGCGTGCAATATAACACTTTGCACACTTGCGTA
GCAAAATACACACCTGTTTTACTAAGAAATACGATCTTTTCTTAAATATCTCTACTACAACGCAGTAATATATAA
ATAGCTAACTTAAATATAGAAATTTAGGGTCTTCTTTAGGATTAGGAGGGTTAGGG

SEQ ID NO : 53
CHIỀU DÀI : 5895
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(5895)

```
atgtacatttacggctcaactctcacgcctacggcgagcaacgcactcttggtttataac
 M  Y  I  Y  G  S  T  L  T  P  T  A  S  N  A  L  L  V  Y  N
aagggcgtgtcttctggcacggtggtgaccaccaactccaccgttgtcttcgactcctgc
 K  G  V  S  S  G  T  V  V  T  T  N  S  T  V  V  F  D  S  C
accgtcactccgaagccgggcaccactgtgaccggtatttcccttgccgcagcgaacggt
 T  V  T  P  K  P  G  T  T  V  T  G  I  S  L  A  A  A  N  G
gttgggtcaattgtggtctatagaaactccgtcctgccaggcttcattctcagcaccggt
 V  G  S  I  V  V  Y  R  N  S  V  L  P  G  F  I  L  S  T  G
attcacgttgacgccaagacccaggcagcacagaacttctacggcgagttcggtaacaca
 I  H  V  D  A  K  T  Q  A  A  Q  N  F  Y  G  E  F  G  N  T
ggtgccggatcctactcttccaactccgccgcgagggcctcctacgtgaagagcctgagc
 G  A  G  S  Y  S  S  N  S  A  A  R  A  S  Y  V  K  S  L  S
gtcgatcagctttctcaattctccgttgatcaagtgttcggcaccgcaaacactaactgg
 V  D  Q  L  S  Q  F  S  V  D  Q  V  F  G  T  A  N  T  N  W
atcgactcttcagtcatctccctcgtccaggattcggatgcatctcaagcggcaaaggcg
 I  D  S  S  V  I  S  L  V  Q  D  S  D  A  S  Q  A  A  K  A
accacggcaggtgtggtgtcatacactactacgtcttcctcttcaagtactacaagctct
 T  T  A  G  V  V  S  Y  T  T  T  S  S  S  S  S  T  T  S  S
gtactgtcgagcagcttggtcaccgcatcgggcagctccgtgtccacatcatcaagctcg
 V  L  S  S  S  L  V  T  A  S  G  S  S  V  S  T  S  S  S  S
gtgtcgtctacgtcttcaacctccgcctctgccagtgacgcgccttcgctttccagcagt
 V  S  S  T  S  S  T  S  A  S  A  S  D  A  P  S  L  S  S  S
gctgctctcgcttcatcagtacctggatcttcctccaacactacagtgaccaccagcggc
 A  A  L  A  S  S  V  P  G  S  S  S  N  T  T  V  T  T  S  G
acagcaagctcaagctctgtcactggttcctcgttcgcactgtccactagctcgatctcc
 T  A  S  S  S  S  V  T  G  S  S  F  A  L  S  T  S  S  I  S
tccagcgtgagcactccgggggacacagcgagctccaccgccgctgcgtctaattcgact
```

S S V S T P G D T A S S T A A A S N S T
tcgagttgcgctcttccttcctcagtgcccaccactgctcgcgtcgtgggtccggctggg
S S C A L P S S V P T T A R V V G P A G
tcctgcgccaactacaccagcatcgccgatgccgtcaaagacctttctacggatcaatct
S C A N Y T S I A D A V K D L S T D Q S
aagacggaatatgtgtacattctcgctggcacctataccgagcagatcattttcagccgt
K T E Y V Y I L A G T Y T E Q I I F S R
gtaggtcccaccgtcttccgcggcgagacatcaagcgagctcgatcagtccagcaacaaa
V G P T V F R G E T S S E L D Q S S N K
gtgactatcaagtcttctactggtgtgccctccagtagcgggggatcttccagtaccgcc
V T I K S S T G V P S S S G G S S S T A
ccgttccaggcgactcagtactacagcaagagtatcagtttctataacatcaatttcgag
P F Q A T Q Y Y S K S I S F Y N I N F E
aacacctacgctgcagctacgggttacaatgcggttgcattgtccagcaaagctttgaag
N T Y A A A T G Y N A V A L S S K A L K
gcctattattacaactgcgggatcacttcctctcaaggagctctgctgctcaacttcggc
A Y Y Y N C G I T S S Q G A L L L N F G
gcccacttcttctcgggctgcaagattaccggtaccaccgacatcgtctggggccaaggt
A H F F S G C K I T G T T D I V W G Q G
ggcgcttacatctacaattcgaagatcgttagcaccggaaccacgactggtcagtctctc
G A Y I Y N S K I V S T G T T T G Q S L
tccgcacagtcgtaccagtcacaatacaatccttcgcagtttgtcttcgacacttgcgct
S A Q S Y Q S Q Y N P S Q F V F D T C A
ttcgtgccgaacgacagcaccgtgccgaaggcgagcacgtacttgggccgcgactataca
F V P N D S T V P K A S T Y L G R D Y T
gcttcggcacgtgttgcggtgatcaactcttacctggacgctcatattactcctgttgga
A S A R V A V I N S Y L D A H I T P V G
tggttgattgcgtccaagactacgaacgtgacgtttgttgaggccagcaactccgggcct
W L I A S K T T N V T F V E A S N S G P
ggagcgtcgaccgcatctagagtttcccagattgtgaccgatacgtctgcctacgccgcc
G A S T A S R V S Q I V T D T S A Y A A
aacaacgtgctcggttctctttcgatcgacaccgccgctgttgctcctgttgctgccttc
N N V L G S L S I D T A A V A P V A A F
cccgacagcgtctacggtagtcctctttcccccagctcttctactgtcttggccaactca
P D S V Y G S P L S P S S S T V L A N S
acgacagcgacttccgtgtctgcttcgacatccagcacttctgctgctgcaactgcggcc
T T A T S V S A S T S S T S A A A T A A
aacaccctcattgtctcaaccacgcctgcctctggcgaatacggcaacgtcacagcagcc
N T L I V S T T P A S G E Y G N V T A A
atcgccgcccttcccaacgatagcaaggagtacaccatctacatccgcgccggcacctac
I A A L P N D S K E Y T I Y I R A G T Y
caggagcagttcaccattgtgcggaatggcaaggtcacactgcgtggcgagacggcgttc
Q E Q F T I V R N G K V T L R G E T A F
cccaacgacttctcgcagaaccaggtcaccatccagttctcctatggtgttctcaccagc
P N D F S Q N Q V T I Q F S Y G V L T S
gctggccagaacgagttgacgcccgttatcaacgccaagaagaacgatggctccggcctc
A G Q N E L T P V I N A K K N D G S G L
gcgttgtacaacatcaacttcatcaacacgtacccgcagacgaagaacacggctgcgctt
A L Y N I N F I N T Y P Q T K N T A A L
gcggctgacttctacggtaccaatatggctgcctatggctgcaagttcgtcggctaccag
A A D F Y G T N M A A Y G C K F V G Y Q
gacaccctcctcgccaacaagggcacccaagtcttttccaacagctacattgaaggctct

D T L L A N K G T Q V F S N S Y I E G S
attgactatatctggggtttctcgacggcctacttccaccagtgctacatcgcatccaac
I D Y I W G F S T A Y F H Q C Y I A S N
acagctggtggctacatttctgccatgagccgcgcttcggcttcggctactggaggttat
T A G G Y I S A M S R A S A S A T G G Y
gtctttgacagctgctatgtcacctacacgagcacctacggatcgacctttggcaccagc
V F D S C Y V T Y T S T Y G S T F G T S
tacctcggtcgcccgtactcgagctacagcatcgccgtctacatgaactcgttcattgac
Y L G R P Y S S Y S I A V Y M N S F I D
aagcacattagccccgccggttgggctgtgtggcaaacgagtaacccgcagacggacaac
K H I S P A G W A V W Q T S N P Q T D N
gtcctgtttggagagttcaacaacaccggccccggaagctggtccagctctcgcgcgtct
V L F G E F N N T G P G S W S S S R A S
ttcgcgacgaacttgaccgagtcccaggcggatgcttacaagctgtctacttggatcggc
F A T N L T E S Q A D A Y K L S T W I G
agcacgtcgtggcttgacatggatgcttacaactacgtgccttcgtacgatatttccggc
S T S W L D M D A Y N Y V P S Y D I S G
gcagctggtgcatcaaccacgacccctcggcgtcagcttcctctaccactgccacagcc
A A G A S T T T P S A S A S S T T A T A
actgccacctgggcgcatccctcgagtggtgctaccccacctatcggtgccgtcctggtg
T A T W A H P S S G A T P P I G A V L V
tccgttggaggctccgtcaacgggtcatacagcaatcttaccgccgccctcgcgtctctg
S V G G S V N G S Y S N L T A A L A S L
ccatcggatagctcgacacaggtcatcttcatgtaccccggaacctacaacgagcagccc
P S D S S T Q V I F M Y P G T Y N E Q P
ccggccgtaaacaggcccggccccatccagatcatcggtgctcaagatggcaatccaggc
P A V N R P G P I Q I I G A Q D G N P G
cagagctacaagaccaataaagtcattctgacccagtctcgcggcttgtcggtctctccg
Q S Y K T N K V I L T Q S R G L S V S P
ttgccgaccggacactctgacgctgaaaccgcaacattctccactaccagcaacaaaatt
L P T G H S D A E T A T F S T T S N K I
gccatgtacaacatcgacatcatcaactccgataatcttgacggttcgctttcctcttac
A M Y N I D I I N S D N L D G S L S S Y
gtgacgctggccggctcaatttatggctcgcgcattgcgttctatgggtgcagcttcatt
V T L A G S I Y G S R I A F Y G C S F I
ggctggcaggatacgctgttgactggaagcaccagtggctaccagtactacgaatcctgc
G W Q D T L L T G S T S G Y Q Y Y E S C
tacattgatggcgccatcgatttcatctggggctactctaaggcttacttcaaaggctgc
Y I D G A I D F I W G Y S K A Y F K G C
accatcggcgccaaaaggcaaaagtcggccatcacggctcacagccgtgcttcctccagc
T I G A K R Q K S A I T A H S R A S S S
gcagtcggaggctacatcttcgaccagtgcctgttcactgctgcctcgtctgcgacggtt
A V G G Y I F D Q C L F T A A S S A T V
gaccttacgcagtctgtttacctcggccgcccttacagcaagtacgccctcgtggttgtc
D L T Q S V Y L G R P Y S K Y A L V V V
aagaactcgtacctggacaagacgatccagcctgctggttggaagatctggtctgccacc
K N S Y L D K T I Q P A G W K I W S A T
gacccgcgtaccgactacgttaccttcgccgagttcaacaactccggacctggtaactgg
D P R T D Y V T F A E F N N S G P G N W
gagaacaacgccgcggcgcgtaccgctttcggctactgcaccctgttgacgtcggacacg
E N N A A A R T A F G Y C T L L T S D T
tactccctatctgccgtcatggactcaccctctgactggattgacatgacgtactgggac

```
 Y  S  L  S  A  V  M  D  S  P  S  D  W  I  D  M  T  Y  W  D
tcgatcaccacaccgacagtggccgctgttgctacgggaaacaccactacggcggtcaac
 S  I  T  T  P  T  V  A  A  V  A  T  G  N  T  T  T  A  V  N
ggtacttcggtctatgatggcaccaccccacctgccggcgcattgattgtctccaagacg
 G  T  S  V  Y  D  G  T  T  P  P  A  G  A  L  I  V  S  K  T
gccattgagggcgtgacgacgtacgataccatccagagcgctctgaatgcgctgcctact
 A  I  E  G  V  T  T  Y  D  T  I  Q  S  A  L  N  A  L  P  T
tcgagcagcaagaccggcaccatcttcatctaccccggtgtgtactcggagcagcttgtg
 S  S  S  K  T  G  T  I  F  I  Y  P  G  V  Y  S  E  Q  L  V
ctgagcaagtctggtaccaccgtgttcatcggctactcaaactccaccgacgattacctg
 L  S  K  S  G  T  T  V  F  I  G  Y  S  N  S  T  D  D  Y  L
cagaaccaggtgacgatcgacttcaacaagggaatcgatacgcaagcagatgcttccaac
 Q  N  Q  V  T  I  D  F  N  K  G  I  D  T  Q  A  D  A  S  N
tctgacagcgctactgtgtatgccacaggcaactacttccaggcatacaacatcaacttc
 S  D  S  A  T  V  Y  A  T  G  N  Y  F  Q  A  Y  N  I  N  F
aagaactctttcggcacaaccgaggactacgcctcgctcggcttcggtgtcaagtccagc
 K  N  S  F  G  T  T  E  D  Y  A  S  L  G  F  G  V  K  S  S
aagtacgcatcgctgtatggctgccaggtctggggcaatcaggactcgctactcatcaac
 K  Y  A  S  L  Y  G  C  Q  V  W  G  N  Q  D  S  L  L  I  N

ggctacttcttcgccttcaactcgctgattgtcggtaacatcgacatgatctggggctcc
 G  Y  F  F  A  F  N  S  L  I  V  G  N  I  D  M  I  W  G  S
ggcgcgggctatttcctcagctcaaccatctcacccaacaccgacgacgtcagcctgacc
 G  A  G  Y  F  L  S  S  T  I  S  P  N  T  D  D  V  S  L  T
gccagcaagcgcgcgaccaacactacagctgccggttttgtcttcgaccagtgcaccgtt
 A  S  K  R  A  T  N  T  T  A  A  G  F  V  F  D  Q  C  T  V
aagcccgcccccggtaccggccccttcaccgagatcagcctcggacgtccgtggaacaac
 K  P  A  P  G  T  G  P  F  T  E  I  S  L  G  R  P  W  N  N
cttgcccgtgtcgcgtatatcgagacgtatctcgattccagcgttgaggctgcaggctgg
 L  A  R  V  A  Y  I  E  T  Y  L  D  S  S  V  E  A  A  G  W
agccagtggtccaagtccaatccacagaccgaaggagtgacttttgccgagtacgggaac
 S  Q  W  S  K  S  N  P  Q  T  E  G  V  T  F  A  E  Y  G  N
tatggaccgggtgcaagcacttcgggccgtgcgaaattttccacgcagctctcggctgcc
 Y  G  P  G  A  S  T  S  G  R  A  K  F  S  T  Q  L  S  A  A
gatgccgcccagtttcagctcgccaacttcttcgccgtcacctcctggatcaacttcacg
 D  A  A  Q  F  Q  L  A  N  F  F  A  V  T  S  W  I  N  F  T
cggatcgatgtccaacccttcgtggccagcgaggttgttgttccaacctcggcggtcact
 R  I  D  V  Q  P  F  V  A  S  E  V  V  V  P  T  S  A  V  T
tcgtctgtccttctctcttccacccttagcactccgatctcgtcttcaacgcttatactg
 S  S  V  L  L  S  S  T  L  S  T  P  I  S  S  S  T  L  I  L
agcaccctcttcctcactaaagtcactaccgacaaggagactctcttcacaaccgtcact
 S  T  L  F  L  T  K  V  T  T  D  K  E  T  L  F  T  T  V  T
ggcgctgttccgaccctcacttcaacgcagactatcacgctggacatgggtgccaccgtt
 G  A  V  P  T  L  T  S  T  Q  T  I  T  L  D  M  G  A  T  V
acccccgacccagtgtacaagaccagcaccgtgaagagtacgacaacgatcattgagacc
 T  P  D  P  V  Y  K  T  S  T  V  K  S  T  T  T  I  I  E  T
gtatcgcagcccgatgttacgcagacgtctactgtggtcgtcaccagcgatatcggaacc
 V  S  Q  P  D  V  T  Q  T  S  T  V  V  V  T  S  D  I  G  T
acaatcacaccggagccgagcacgatcaccactgtgctcaagcagacgactactgttttc
 T  I  T  P  E  P  S  T  I  T  T  V  L  K  Q  T  T  T  V  F
gccacctcgaccaaggcgccccagacgattactgagaagagtaccattacttcgacgagt
 A  T  S  T  K  A  P  Q  T  I  T  E  K  S  T  I  T  S  T  S
```

```
ttggctacccggacgctcgatcccattacgtcaacactcagcctcggatctaccgtttat
 L  A  T  R  T  L  D  P  I  T  S  T  L  S  L  G  S  T  V  Y
gtgaccagcgttttcaccccgaaggccgcgcgggtcacctctagtctaaccatcactacg
 V  T  S  V  F  T  P  K  A  A  R  V  T  S  S  L  T  I  T  T
ggcactggcggcacgtcaactaagacgaccaaagcaaccacgacttacgtcactgtcact
 G  T  G  G  T  S  T  K  T  T  K  A  T  T  T  Y  V  T  V  T
tctatgatatcgccagagacaacgagacggacggaagggacgtgggggaagctggaaaac
 S  M  I  S  P  E  T  T  R  R  T  E  G  T  W  G  K  L  E  N
tcgttgtcggtgtgcccaacttcaattccgtgccatttatggccgttcgacggagaactc
 S  L  S  V  C  P  T  S  I  P  C  H  L  W  P  F  D  G  E  L
ccccgacgctggaagcgccttgggcactggcttggcagtatcgagctggccggggaacat
 P  R  R  W  K  R  L  G  H  W  L  G  S  I  E  L  A  G  E  H
gccaggccagatgtgctcgcaactcgtattgcatacgggataatacaacgtgcacactct
 A  R  P  D  V  L  A  T  R  I  A  Y  G  I  I  Q  R  A  H  S
tggcgtgcaatataa
 W  R  A  I  -
```

SEQ ID NO : 54
CHIỀU DÀI : 1964
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MYIYGSTLTPTASNALLVYNKGVSSGTVVTTNSTVVFDSCTVTPKPGTTVTGISLAAANGVGSIVVYRNSVLPGF
ILSTGIHVDAKTQAAQNFYGEFGNTGAGSYSSNSAARASYVKSLSVDQLSQFSVDQVFGTANTNWIDSSVISLVQ
DSDASQAAKATTAGVVSYTTTSSSSSTTSSVLSSSLVTASGSSVSTSSSSVSSTSSTSASASDAPSLSSSAALAS
SVPGSSSNTTVTTSGTASSSSVTGSSFALSTSSISSSVSTPGDTASSTAAASNSTSSCALPSSVPTTARVVGPAG
SCANYTSIADAVKDLSTDQSKTEYVYILAGTYTEQIIFSRVGPTVFRGETSSELDQSSNKVTIKSSTGVPSSSGG
SSSTAPFQATQYYSKSISFYNINFENTYAAATGYNAVALSSKALKAYYYNCGITSSQGALLLNFGAHFFSGCKIT
GTTDIVWGQGGAYIYNSKIVSTGTTTGQSLSAQSYQSQYNPSQFVFDTCAFVPNDSTVPKASTYLGRDYTASARV
AVINSYLDAHITPVGWLIASKTTNVTFVEASNSGPGASTASRVSQIVTDTSAYAANNVLGSLSIDTAAVAPVAAF
PDSVYGSPLSPSSSTVLANSTTATSVSASTSSTSAAATAANTLIVSTTPASGEYGNVTAAIAALPNDSKEYTIYI
RAGTYQEQFTIVRNGKVTLRGETAFPNDFSQNQVTIQFSYGVLTSAGQNELTPVINAKKNDGSGLALYNINFINT
YPQTKNTAALAADFYGTNMAAYGCKFVGYQDTLLANKGTQVFSNSYIEGSIDYIWGFSTAYFHQCYIASNTAGGY
ISAMSRASASATGGYVFDSCYVTYTSTYGSTFGTSYLGRPYSSYSIAVYMNSFIDKHISPAGWAVWQTSNPQTDN
VLFGEFNNTGPGSWSSSRASFATNLTESQADAYKLSTWIGSTSWLDMDAYNYVPSYDISGAAGASTTTPSASASS
TTATATATWAHPSSGATPPIGAVLVSVGGSVNGSYSNLTAALASLPSDSSTQVIFMYPGTYNEQPPAVNRPGPIQ
IIGAQDGNPGQSYKTNKVILTQSRGLSVSPLPTGHSDAETATFSTTSNKIAMYNIDIINSDNLDGSLSSYVTLAG
SIYGSRIAFYGCSFIGWQDTLLTGSTSGYQYYESCYIDGAIDFIWGYSKAYFKGCTIGAKRQKSAITAHSRASSS
AVGGYIFDQCLFTAASSATVDLTQSVYLGRPYSKYALVVVKNSYLDKTIQPAGWKIWSATDPRTDYVTFAEFNNS
GPGNWENNAAARTAFGYCTLLTSDTYSLSAVMDSPSDWIDMTYWDSITTPTVAAVATGNTTTAVNGTSVYDGTTP
PAGALIVSKTAIEGVTTYDTIQSALNALPTSSSKTGTIFIYPGVYSEQLVLSKSGTTVFIGYSNSTDDYLQNQVT
IDFNKGIDTQADASNSDSATVYATGNYFQAYNINFKNSFGTTEDYASLGFGVKSSKYASLYGCQVWGNQDSLLIN
GYFFAFNSLIVGNIDMIWGSGAGYFLSSTISPNTDDVSLTASKRATNTTAAGFVFDQCTVKPAPGTGPFTEISLG
RPWNNLARVAYIETYLDSSVEAAGWSQWSKSNPQTEGVTFAEYGNYGPGASTSGRAKFSTQLSAADAAQFQLANF
FAVTSWINFTRIDVQPFVASEVVVPTSAVTSSVLLSSTLSTPISSSTLILSTLFLTKVTTDKETLFTTVTGAVPT
LTSTQTITLDMGATVTPDPVYKTSTVKSTTTIIETVSQPDVTQTSTVVVTSDIGTTITPEPSTITTVLKQTTTVF
ATSTKAPQTITEKSTITSTSLATRTLDPITSTLSLGSTVYVTSVFTPKAARVTSSLTITTGTGGTSTKTTKATTT
YVTVTSMISPETTRRTEGTWGKLENSLSVCPTSIPCHLWPFDGELPRRWKRLGHWLGSIELAGEHARPDVLATRI
AYGIIQRAHSWRAI*
```

SEQ ID NO : 55
CHIỀU DÀI : 1281 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)

LOẠI : ADN
SINH VẬT : *M. phaseolina*

CACCGGTTGTTGATCATTTACACCCTCAGCGAGCATTCCCGGCCAGGCGCCTCTTGTTGAACACTGTCCAGTCCT
TCACTCCTCTCCTTCCACCCGCTACCCGGTCTTTTTCCTTTTGCCATTCCAATTGCGTGAACCGTCCATCACAGC
ATGCCTCGCCTCGCCAGCCTCCTCGCCCTTGCAAGCCCAGTGCTTGCCCTGACCTCCCCTCCGTCAGGCGCCCTC
ACCGTCGGCTCCAGCGGCAAGTACTCGACCATCCAGGACGCCGTCGACGCGCTCAGCACCAGCTCCTCGTCGGCC
CAGACCATCTTCATCTACCAGGGCACCTACAAGGAACAAGTCGTGATCCCGAAGCTATCGGGCGCCCTCACCATC
TACGGCTACTCGAAGGATGGATCATCGTACAGCGGCAACACCGTGACCATCAGCGCGGGCAAGTCGCAGGCCGAC
GGCCTCTCCAACGACGGCACCGCCACCCTCGCCGTCCACACGGGCAACATCAAGGTCTACAACATCAACGTCGCC
AACACGTACGGCTCCGGCTCGCAGGCCGTCGCGCTGTCGGCCTACGCCAGCGGCAACCACGGCTACTACGGCGTC
AAGCTCACGGGCTTCCAAGACACGCTCCTGGCGCAGGAGGGCAAACAAGTCTACGCCAACTCGTACATCGAGGGC
GCGACCGATTTCATTTTCGGCCAGAAGGCCGTCGCGTGGTTTGAGAACTGCGACCTGCGCATCGCCTCGGCCTCG
CTCGGCTACGTGACGGCCAACGGCCGCGACAGCAGCTCCAACCCATCCTACTACGTCATCAACAACTCGACCGTC
GCCGCCGCCGACGGCGCCACCGTCAAGAGCGGCGGCATCTACCTCGGCCGCCCCTGGCGCAACTACGCCCGCGTC
GTCTTCCAGGAGACCAGCCTGTCCAACATCATCAACAGCGCCGGCTGGGTCCAATGGGGCAGCAGCGACCCACGC
ACCGACAATGTCAACTTCGCCGAGTACAAAAACTCGGGCGCCGGCGCCTCTACCAGCGGCCGTGCTAGCTTTTCC
AAGCAGTTGAGCAGCCCCGTCAGCATCTCGGAAGTCTTGGGCAGCAACTACGCCGACTGGATCGACACCAGCTAC
TTCTAAACGCGTGAGGCTTCGCTAATGTCGGTTGGATCGGATGGAGGCGGGATGTAGCATGCAAGGAACCGGCAG
GACTGTTTTAACTTTCTAATGTGGATATTtttGTCTCTCGGACAAATATAAAAATAAGCTTTCTGGAGTGACTTT
TGATGA

SEQ ID NO : 56
CHIỀU DÀI : 981
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(981)

```
atgcctcgcctcgccagcctcctcgcccttgcaagcccagtgcttgccctgacctcccct
 M  P  R  L  A  S  L  L  A  L  A  S  P  V  L  A  L  T  S  P
ccgtcaggcgccctcaccgtcggctccagcggcaagtactcgaccatccaggacgccgtc
 P  S  G  A  L  T  V  G  S  S  G  K  Y  S  T  I  Q  D  A  V
gacgcgctcagcaccagctcctcgtcggcccagaccatcttcatctaccagggcacctac
 D  A  L  S  T  S  S  S  S  A  Q  T  I  F  I  Y  Q  G  T  Y
aaggaacaagtcgtgatcccgaagctatcgggcgccctcaccatctacggctactcgaag
 K  E  Q  V  V  I  P  K  L  S  G  A  L  T  I  Y  G  Y  S  K
gatggatcatcgtacagcggcaacaccgtgaccatcagcgcgggcaagtcgcaggccgac
 D  G  S  S  Y  S  G  N  T  V  T  I  S  A  G  K  S  Q  A  D
ggcctctccaacgacggcaccgccaccctcgccgtccacacgggcaacatcaaggtctac
 G  L  S  N  D  G  T  A  T  L  A  V  H  T  G  N  I  K  V  Y
aacatcaacgtcgccaacacgtacggctccggctcgcaggccgtcgcgctgtcggcctac
 N  I  N  V  A  N  T  Y  G  S  G  S  Q  A  V  A  L  S  A  Y
gccagcggcaaccacggctactacggcgtcaagctcacgggcttccaagacacgctcctg
 A  S  G  N  H  G  Y  Y  G  V  K  L  T  G  F  Q  D  T  L  L
gcgcaggagggcaaacaagtctacgccaactcgtacatcgagggcgcgaccgatttcatt
 A  Q  E  G  K  Q  V  Y  A  N  S  Y  I  E  G  A  T  D  F  I
ttcggccagaaggccgtcgcgtggtttgagaactgcgacctgcgcatcgcctcggcctcg
 F  G  Q  K  A  V  A  W  F  E  N  C  D  L  R  I  A  S  A  S
ctcggctacgtgacggccaacggccgcgacagcagctccaacccatcctactacgtcatc
 L  G  Y  V  T  A  N  G  R  D  S  S  S  N  P  S  Y  Y  V  I
aacaactcgaccgtcgccgccgccgacggcgccaccgtcaagagcggcggcatctacctc
 N  N  S  T  V  A  A  A  D  G  A  T  V  K  S  G  G  I  Y  L
ggccgcccctggcgcaactacgcccgcgtcgtcttccaggagaccagcctgtccaacatc
```

G R P W R N Y A R V V F Q E T S L S N I
atcaacagcgccggctgggtccaatggggcagcagcgacccacgcaccgacaatgtcaac
I N S A G W V Q W G S S D P R T D N V N
ttcgccgagtacaaaaactcgggcgccggcgcctctaccagcggccgtgctagcttttcc
F A E Y K N S G A G A S T S G R A S F S
aagcagttgagcagccccgtcagcatctcggaagtcttgggcagcaactacgccgactgg
K Q L S S P V S I S E V L G S N Y A D W
atcgacaccagctacttctaa
I D T S Y F -

SEQ ID NO : 57
CHIỀU DÀI : 326
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MPRLASLLALASPVLALTSPPSGALTVGSSGKYSTIQDAVDALSTSSSSAQTIFIYQGTYKEQVVIPKLSGALTI
YGYSKDGSSYSGNTVTISAGKSQADGLSNDGTATLAVHTGNIKVYNINVANTYGSGSQAVALSAYASGNHGYYGV
KLTGFQDTLLAQEGKQVYANSYIEGATDFIFGQKAVAWFENCDLRIASASLGYVTANGRDSSSNPSYYVINNSTV
AAADGATVKSGGIYLGRPWRNYARVVFQETSLSNIINSAGWVQWGSSDPRTDNVNFAEYKNSGAGASTSGRASFS
KQLSSPVSISEVLGSNYADWIDTSYF*

SEQ ID NO : 58
CHIỀU DÀI : 1943 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

CTATATTTCCATTTTTGGGAACGTACGCGTACGCCTTCTGTGGGGCCCCTTACTCCCGGAGGATTGCTGCGGGAC
TTGGAGAATGTATAAAGTACAGCAATCGAGTTGTCTGCGAGCTAGAGCGTTCTCCAGGCAGTGTCTAAACGTTTC
ATGATGTTGTCTAACGTCATCGTCGGCCTTCTAGCTGCCGCCTATGGAAAGCCTGTGCTGGCAGCGTTTGGGCTC
ACAGACAATGGCGACAGTTACACCATTGATGCTGGCTCCACCAATCCCCTGGTAATCACTGTCGACAGTTCCAAC
TGCGACATCACCTCTTTCGTCTACCGGAGCCAGAAATACCAGTATGCCTCGAAAGGATCCCACATTAGCTCAGGC
CTTGGCTCGGCTGATGTGACGGCAGAGACTATTGATGGTAAAGCCTGAAGACTGTGATGCAACTTCCCATCCGCT
GACCACCCGCAGATGTAATCAAGGTCACTTGCACAACCGACACACTGATCCACTACTATGTCGTCAAATCCGGAG
AGAGCAACGTCTACATGGCTACGTACACTTCAGCTGAGCCTTCAGTGGGAGAGCTGCGCTACATCGCACGACTTG
ATCCTTCGCTGCTTCCATATGAGTACCCCTTCGGCGACGTCTCCACCACCGCCGACAGCACTTCGACAGTGGAAG
GTTCTGACGTCTTCATCGTCGACGGAGAGACTCGCAGCAAATTCTACTCCAGCCAGCGCTTCATCGACGACGACG
TCCACTGCGTCTACAGCGACGACGTCCACGCGTGCATGCTGATGCCCCAGTACGAGACCTCGTCCGGCGGGCCGT
TCCACCGAGACATCAACACCAACAACGCCGGCGACTCGACCAACCTGTACTGGTACATGAACAGCGGACACGTGC
AGACCGAGGCCTTCCGGATGGGCCTCCACGGGCCCTACGCCCTCACGTGGTCGCGCTCTGGCGTGCCCAGTCTCA
GCGGCCTCGACTTCTCCTTCATGGCGGACCTCGACCTGGAAGGCTACGTCGCCGACAGCGGGCGCGGCACCGTCT
CGGGCACAGCATCCGGCGTCTCCAGCGACTATGAGGTCGTGGTGCACTGGTACAACGACGACGCGCAGTACTGGA
CGTACGCCTCTTCGTCCGGCGCCTTCACGTCGCCCGCCATGAAGCCGGGGACCTACACCATGGCCCTGTACCAGA
CCGAACTCAAAGTGGCCACGTCCTCGGTGACGGTCAAGGCCGGATCGGCCACCAGCGCCAACATCGCCAGCACCT
GGAGCTCCAATACCACGCTCTTCCAGATCGGCGACTGGGACGGGCAGCCGACCGGCTTCCTCAACGCGGCCAGCC
AGCTGCGCATGCACCCGTCCGACTCGCGCATGGCCGACTGGGGCCCCGTCACCTACACCGTCGGCTCGTCCTCCG
ACAGCTCCGTCCCGATGGCGCTGTTCAAGGACGTCAACGACCCGCTGACCATCAGCTTCACGCTGGCGTCTTCGC
AGGCCAGCGGCGCGGCGACGCTGCGCGTCGGCACCACGCTGTCGTTCGCGGGCGGGCGGCCGAGCGTCACCGTCA
ACGACTACAGCGACTCCGCCGCCGCGCCGACCAAGATCGACTCCCGGGGCGTGACGCGCGGCGCGTATCGCGGGT
ACGGCGAGATCTACGACTTCGAGCTGCCGGACGGGACGCTGTCGACGAGCAATACGATCACCATTTCGGTGATTT
CGGGGAGCAGTGGGGCGGATTTCCTGGAGCCGAATTTTGTGAGTTCTCGCATTCCTCTTGGCCACTAGCATGTTT
GAGAGGAATGTGACGCTGACGGTGTGACAGATCCTTGATGCTATTGAGCTGTGGAGATAGGGTTGATAGGTCTGC
GGACGTGTTGTGGATGGTTTCGCGGGTCTGGGATCTTGATTGGTTGTTGGGATGGTATGCAGCAAAGA

SEQ ID NO : 59
CHIỀU DÀI : 1593
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1593)

```
atgatgttgtctaacgtcatcgtcggccttctagctgccgcctatggaaagcctgtgctg
 M  M  L  S  N  V  I  V  G  L  L  A  A  A  Y  G  K  P  V  L
gcagcgtttgggctcacagacaatggcgacagttacaccattgatgctggctccaccaat
 A  A  F  G  L  T  D  N  G  D  S  Y  T  I  D  A  G  S  T  N
cccctggtaatcactgtcgacagttccaactgcgacatcacctctttcgtctaccggagc
 P  L  V  I  T  V  D  S  S  N  C  D  I  T  S  F  V  Y  R  S
cagaaataccagtatgcctcgaaaggatcccacattagctcaggccttggctcggctgat
 Q  K  Y  Q  Y  A  S  K  G  S  H  I  S  S  G  L  G  S  A  D
gtgacggcagagactattgatgatgtaatcaaggtcacttgcacaaccgacacactgatc
 V  T  A  E  T  I  D  D  V  I  K  V  T  C  T  T  D  T  L  I
cactactatgtcgtcaaatccggagagagcaacgtctacatggctacgtacacttcagct
 H  Y  Y  V  V  K  S  G  E  S  N  V  Y  M  A  T  Y  T  S  A
gagccttcagtgggagagctgcgctacatcgcacgacttgatccttcgctgcttccatat
 E  P  S  V  G  E  L  R  Y  I  A  R  L  D  P  S  L  L  P  Y
gagtacccccttcggcgacgtctccaccaccgccgacagcacttcgacagtggaaggttct
 E  Y  P  F  G  D  V  S  T  T  A  D  S  T  S  T  V  E  G  S
gacgtcttcatcgtcgacggagagactcgcagcaaattctactccagccagcgcttcatc
 D  V  F  I  V  D  G  E  T  R  S  K  F  Y  S  S  Q  R  F  I
gacgacgacgtccactgcgtctacagcgacgacgtccacgcgtgcatgctgatgccccag
 D  D  D  V  H  C  V  Y  S  D  D  V  H  A  C  M  L  M  P  Q
tacgagacctcgtccggcgggccgttccaccgagacatcaacaccaacaacgccggcgac
 Y  E  T  S  S  G  G  P  F  H  R  D  I  N  T  N  N  A  G  D
tcgaccaacctgtactggtacatgaacagcggacacgtgcagaccgaggccttccggatg
 S  T  N  L  Y  W  Y  M  N  S  G  H  V  Q  T  E  A  F  R  M
ggcctccacgggccctacgccctcacgtggtcgcgctctggcgtgcccagtctcagcggc
 G  L  H  G  P  Y  A  L  T  W  S  R  S  G  V  P  S  L  S  G
ctcgacttctccttcatggcggacctcgacctggaaggctacgtcgccgacagcgggcgc
 L  D  F  S  F  M  A  D  L  D  L  E  G  Y  V  A  D  S  G  R
ggcaccgtctcgggcacagcatccggcgtctccagcgactatgaggtcgtggtgcactgg
 G  T  V  S  G  T  A  S  G  V  S  S  D  Y  E  V  V  V  H  W
tacaacgacgacgcgcagtactggacgtacgcctcttcgtccggcgccttcacgtcgccc
 Y  N  D  D  A  Q  Y  W  T  Y  A  S  S  S  G  A  F  T  S  P
gccatgaagccggggacctacaccatggccctgtaccagaccgaactcaaagtggccacg
 A  M  K  P  G  T  Y  T  M  A  L  Y  Q  T  E  L  K  V  A  T
tcctcggtgacggtcaaggccggatcggccaccagcgccaacatcgccagcacctggagc
 S  S  V  T  V  K  A  G  S  A  T  S  A  N  I  A  S  T  W  S
tccaataccacgctcttccagatcggcgactgggacgggcagccgaccggcttcctcaac
 S  N  T  T  L  F  Q  I  G  D  W  D  G  Q  P  T  G  F  L  N
gcggccagccagctgcgcatgcacccgtccgactcgcgcatggccgactggggccccgtc
 A  A  S  Q  L  R  M  H  P  S  D  S  R  M  A  D  W  G  P  V
acctacaccgtcggctcgtcctccgacagctccgtcccgatggcgctgttcaaggacgtc
 T  Y  T  V  G  S  S  S  D  S  S  V  P  M  A  L  F  K  D  V
aacgacccgctgaccatcagcttcacgctggcgtcttcgcaggccagcggcgcggcgacg
 N  D  P  L  T  I  S  F  T  L  A  S  S  Q  A  S  G  A  A  T
ctgcgcgtcggcaccacgctgtcgttcgcgggcgggcggccgagcgtcaccgtcaacgac
 L  R  V  G  T  T  L  S  F  A  G  G  R  P  S  V  T  V  N  D
```

```
tacagcgactccgccgccgcgccgaccaagatcgactcccggggcgtgacgcgcggcgcg
 Y  S  D  S  A  A  A  P  T  K  I  D  S  R  G  V  T  R  G  A
tatcgcgggtacggcgagatctacgacttcgagctgccggacgggacgctgtcgacgagc
 Y  R  G  Y  G  E  I  Y  D  F  E  L  P  D  G  T  L  S  T  S
aatacgatcaccatttcggtgatttcggggagcagtggggcggatttcctggagccgaat
 N  T  I  T  I  S  V  I  S  G  S  S  G  A  D  F  L  E  P  N
tttgtgagttctcgcattcctcttggccactag
 F  V  S  S  R  I  P  L  G  H  -
```

SEQ ID NO : 60
CHIỀU DÀI : 530
LOẠI : PRT
SINH VẬT : *M. phaseolina*

```
MMLSNVIVGLLAAAYGKPVLAAFGLTDNGDSYTIDAGSTNPLVITVDSSNCDITSFVYRSQKYQYASKGSHISSG
LGSADVTAETIDDVIKVTCTTDTLIHYYVVKSGESNVYMATYTSAEPSVGELRYIARLDPSLLPYEYPFGDVSTT
ADSTSTVEGSDVFIVDGETRSKFYSSQRFIDDDVHCVYSDDVHACMLMPQYETSSGGPFHRDINTNNAGDSTNLY
WYMNSGHVQTEAFRMGLHGPYALTWSRSGVPSLSGLDFSFMADLDLEGYVADSGRGTVSGTASGVSSDYEVVVHW
YNDDAQYWTYASSSGAFTSPAMKPGTYTMALYQTELKVATSSVTVKAGSATSANIASTWSSNTTLFQIGDWDGQP
TGFLNAASQLRMHPSDSRMADWGPVTYTVGSSSDSSVPMALFKDVNDPLTISFTLASSQASGAATLRVGTTLSFA
GGRPSVTVNDYSDSAAAPTKIDSRGVTRGAYRGYGEIYDFELPDGTLSTSNTITISVISGSSGADFLEPNFVSSR
IPLGH*
```

SEQ ID NO : 61
CHIỀU DÀI : 2009 (bao gồm 150 bp 5′ UTR và 150 bp 3′ UTR)
LOẠI : ADN
SINH VẬT : *M. phaseolina*

```
TCGCTGATCGTCTTGGATATGCGCCTGGCAGTGCATTCGCGCCGGGGACTACAAAAGCAGGGGACGCATGCCACA
ACAGAATTGGCTTTTCTTGTCACTGCACTTCCTGCCGCTCCAAGTCCCCCCCTCGCGTTTGAGGGAACCGCCACA
ATGGCTCTCTTTGCTCTGTTTCTGACCTTTATCTCTCTTTTCGCACCGTCGGTGCTGGCCGCCTTTGGAGTGACT
ACATCCTCTTCCAGCTATGTTGTCGATGCCGGCTCATCCAACCCCTTCGTTGTTACCATTTCGCGCAGCAGTTGC
GACATTACCTCGATCAAGTACCGTGGAGAGGAATTCCAGTACTCCGGCAAGGGTTCGCACATCTCGTCCGGCCTT
GGATCTGCGACCGTAACGTCAGAGATCGTCAGCAGTGAGCGTCTCTTGCAACCCGTCATACGCCCCTCAGGCTAA
CAATCGACAGGCACTTATGCCAAGATCACCTGCACGGCTGGCAGCTTGACCCATTACATTATTGTCAAGTCCGGA
GAGAGCTCTTTGTACATGGGCACTTACTTCACTGAGGAGCCCTCGATTGGTGAAGCCCGTTTCATCGCTCGCTTG
GACCCGGCCAAGCTCCCTCTCGAATACCCCTACGGCACTGCCTCCACTACTGCTGGGAGCAGCAGCACCGTCGAG
GGATCCGACGTCTTTGTTGTGAATGGCCAGACCCGCAGCAAGTTCTACTCTAGCCAACGTTTCATCGACGACAAG
GTGCAGTGCGTGTACCGCGATGATGATGCTGTCCACGCTTGCATGATCTTGCAGCCTCTCTCCTACGAGGGATCC
AGCGGCGGTCCCTTCTTCAGGGACATCAACACCAACAACGCCGGTGACTCGACCAACCTGTACTTCTACATGAAC
TCCAACCATGCTCAGACTGAGAGTTACCGTATGGGCTTCCACGGCCCGTACCAGTTGCAGTTCAGCCGCTCTGGT
ATTCCCAACAGCTTTGACGCTTCGTTCTTCGCCGACCTCAAGCTCTCCGGCTACGTTGCTGAGTCTGCTCGTGGC
TACGTCAAGGGTACTGCTTCTGGAGTTGGTAGCTCTTACCAGAAAGTCCTTCATTGGTACAACAGCAACGCGCAG
TACTGGGTCTATGCCTCGTCTAACGGTGCCTTTACTTCTCCTGCCATGAAGCCTGGCACCTACACCCAGGTCCTC
TACCAGGACGAACTCAAGGTTGCCACCGACTCGGTCACTGTCTCTGCTGGATCGACGGTCACCAAGAACATTGCA
TCTACCTTCTCTTTCCCTTCGACCATCTGGACAATTGGCGACTGGGACGGCCAGCCATTCGGCTTCCGCAACGCC
GACAAGATCGAGCGCATGCACCCGTCCGACAGCCGCATGAGCAGCTGGGGGCCGTTGACCTACACGGTCGGCTCC
AGCGCACTGACTGACGTTCCCATGGCGCTCTTCAAGGGCGTCAACACGCCTTTCACCATCAAGTTCACGCTCTCC
TCGTCGCAGACAGGTGCGGCGGTCCTCCGGATCGGCACGACTCTGGCCTTCGCCAGCGGCAGGCCGCAGCCCAAG
ATCAACAGCTACAGCCCTTCGGCGCCGGCGGCACCGACGAAGATCGATTCGCGCGGCGTCACACGTGGCACTTAC
CGCGGTTTGGGTGAGATCTACACCTTCGACATTCCCGCGGGCACTCTCGTGAGTGGCTCGAACACGATCACCATC
GACTGCATCTCGGGAAGCTCGGGCGACACCTACCTGTCTCCCAACTTTGTAAGTTTCGCTCTGCATACCTCCGCC
ATGCGCCGTGCTAACGCAAGAAACAGATCCTTGACGCCATTGACCTGTACCTCAAGTGAGCGAGGGTCAGGAGAA
```

GGGTGCTAGCTGAGGTTCGCTTCTGTAAATAAGTTTACGTCCTACATTGAACATACATTTTGGCTTTGATTGCCG
GTCTGCATTACTCGTCAATACGAAGCTTCGCGAGCGCGCAGCTCCACCGAATGCGTGGC

SEQ ID NO : 62

CHIỀU DÀI : 1605
LOẠI : ADN
SINH VẬT : *M. phaseolina*
TÊN/KHÓA DẤU HIỆU : CDS
VỊ TRÍ : (1)……(1605)

```
atggctctctttgctctgtttctgacctttatctctcttttcgcaccgtcggtgctggcc
 M  A  L  F  A  L  F  L  T  F  I  S  L  F  A  P  S  V  L  A
gcctttggagtgactacatcctcttccagctatgttgtcgatgccggctcatccaacccc
 A  F  G  V  T  T  S  S  S  S  Y  V  V  D  A  G  S  S  N  P

ttcgttgttaccatttcgcgcagcagttgcgacattacctcgatcaagtaccgtggagag
 F  V  V  T  I  S  R  S  S  C  D  I  T  S  I  K  Y  R  G  E
gaattccagtactccggcaagggttcgcacatctcgtccggccttggatctgcgaccgta
 E  F  Q  Y  S  G  K  G  S  H  I  S  S  G  L  G  S  A  T  V
acgtcagagatcgtcagcagcacttatgccaagatcacctgcacggctggcagcttgacc
 T  S  E  I  V  S  S  T  Y  A  K  I  T  C  T  A  G  S  L  T
cattacattattgtcaagtccggagagagctctttgtacatgggcacttacttcactgag
 H  Y  I  I  V  K  S  G  E  S  S  L  Y  M  G  T  Y  F  T  E
gagccctcgattggtgaagcccgtttcatcgctcgcttggacccggccaagctccctctc
 E  P  S  I  G  E  A  R  F  I  A  R  L  D  P  A  K  L  P  L
gaatacccctacggcactgcctccactactgctgggagcagcagcaccgtcgagggatcc
 E  Y  P  Y  G  T  A  S  T  T  A  G  S  S  S  T  V  E  G  S
gacgtctttgttgtgaatggccagacccgcagcaagttctactctagccaacgtttcatc
 D  V  F  V  V  N  G  Q  T  R  S  K  F  Y  S  S  Q  R  F  I
gacgacaaggtgcagtgcgtgtaccgcgatgatgatgctgtccacgcttgcatgatcttg
 D  D  K  V  Q  C  V  Y  R  D  D  D  A  V  H  A  C  M  I  L
cagcctctctcctacgagggatccagcggcggtcccttcttcagggacatcaacaccaac
 Q  P  L  S  Y  E  G  S  S  G  G  P  F  F  R  D  I  N  T  N
aacgccggtgactcgaccaacctgtacttctacatgaactccaaccatgctcagactgag
 N  A  G  D  S  T  N  L  Y  F  Y  M  N  S  N  H  A  Q  T  E
agttaccgtatgggcttccacggcccgtaccagttgcagttcagccgctctggtattccc
 S  Y  R  M  G  F  H  G  P  Y  Q  L  Q  F  S  R  S  G  I  P
aacagctttgacgcttcgttcttcgccgacctcaagctctccggctacgttgctgagtct
 N  S  F  D  A  S  F  F  A  D  L  K  L  S  G  Y  V  A  E  S
gctcgtggctacgtcaagggtactgcttctggagttggtagctcttaccagaaagtcctt
 A  R  G  Y  V  K  G  T  A  S  G  V  G  S  S  Y  Q  K  V  L
cattggtacaacagcaacgcgcagtactgggtctatgcctcgtctaacggtgcctttact
 H  W  Y  N  S  N  A  Q  Y  W  V  Y  A  S  S  N  G  A  F  T
tctcctgccatgaagcctggcacctacacccaggtcctctaccaggacgaactcaaggtt
 S  P  A  M  K  P  G  T  Y  T  Q  V  L  Y  Q  D  E  L  K  V
gccaccgactcggtcactgtctctgctggatcgacggtcaccaagaacattgcatctacc
 A  T  D  S  V  T  V  S  A  G  S  T  V  T  K  N  I  A  S  T
ttctctttcccttcgaccatctggacaattggcgactgggacggccagccattcggcttc
 F  S  F  P  S  T  I  W  T  I  G  D  W  D  G  Q  P  F  G  F
cgcaacgccgacaagatcgagcgcatgcacccgtccgacagccgcatgagcagctggggg
 R  N  A  D  K  I  E  R  M  H  P  S  D  S  R  M  S  S  W  G
ccgttgacctacacggtcggctccagcgcactgactgacgttcccatggcgctcttcaag
 P  L  T  Y  T  V  G  S  S  A  L  T  D  V  P  M  A  L  F  K
ggcgtcaacacgcctttcaccatcaagttcacgctctcctcgtcgcagacaggtgcggcg
 G  V  N  T  P  F  T  I  K  F  T  L  S  S  S  Q  T  G  A  A
gtcctccggatcggcacgactctggccttcgccagcggcaggccgcagcccaagatcaac
```

```
 V  L  R  I  G  T  T  L  A  F  A  S  G  R  P  Q  P  K  I  N
agctacagcccttcggcgccggcggcaccgacgaagatcgattcgcgcggcgtcacacgt
 S  Y  S  P  S  A  P  A  A  P  T  K  I  D  S  R  G  V  T  R
ggcacttaccgcggtttgggtgagatctacaccttcgacattcccgcgggcactctcgtg
 G  T  Y  R  G  L  G  E  I  Y  T  F  D  I  P  A  G  T  L  V
agtggctcgaacacgatcaccatcgactgcatctcgggaagctcgggcgacacctacctg
 S  G  S  N  T  I  T  I  D  C  I  S  G  S  S  G  D  T  Y  L
tctcccaactttatccttgacgccattgacctgtacctcaagtga
 S  P  N  F  I  L  D  A  I  D  L  Y  L  K  -
```

SEQ ID NO : 63
CHIỀU DÀI : 534
LOẠI : PRT
SINH VẬT : *M. phaseolina*

MALFALFLTFISLFAPSVLAAFGVTTSSSSYVVDAGSSNPFVVTISRSSCDITSIKYRGEEFQYSGKGSHISSGL
GSATVTSEIVSSTYAKITCTAGSLTHYIIVKSGESSLYMGTYFTEEPSIGEARFIARLDPAKLPLEYPYGTASTT
AGSSSTVEGSDVFVVNGQTRSKFYSSQRFIDDKVQCVYRDDDAVHACMILQPLSYEGSSGGPFFRDINTNNAGDS
TNLYFYMNSNHAQTESYRMGFHGPYQLQFSRSGIPNSFDASFFADLKLSGYVAESARGYVKGTASGVGSSYQKVL
HWYNSNAQYWVYASSNGAFTSPAMKPGTYTQVLYQDELKVATDSVTVSAGSTVTKNIASTFSFPSTIWTIGDWDG
QPFGFRNADKIERMHPSDSRMSSWGPLTYTVGSSALTDVPMALFKGVNTPFTIKFTLSSSQTGAAVLRIGTTLAF
ASGRPQPKINSYSPSAPAAPTKIDSRGVTRGTYRGLGEIYTFDIPAGTLVSGSNTITIDCISGSSGDTYLSPNFI
LDAIDLYLK*

YÊU CẦU BẢO HỘ

1. Trình tự polynucleotit được phân lập chứa trình tự nucleotit có độ tương đồng ít nhất là 80% so với trình tự nucleotit như nêu trong SEQ ID NO. 1 hoặc 2 hoặc trình tự bổ sung của chúng, trong đó polynucleotit nói trên mã hóa pectat lyaza.

2. Đoạn dò axit nucleic được phân lập lai hóa dưới điều kiện nghiêm ngặt trung bình với trình tự polynucleotit được phân lập bao gồm trình tự nuleotit được nêu trong SEQ ID NO. 1 hoặc 2, hoặc trình tự bổ sung của chúng, trong đó polynucleotit nói trên mã hóa pectat lyaza, và trong đó điều kiện nghiêm ngặt trung bình bao gồm sự lai hóa với ADN liên kết với bộ lọc ở 6x natri clorua/natri xitrat (SSC) ở khoảng 45°C sau đó là một hoặc nhiều lần rửa trong 0,2xSSC/0,1% SDS ở khoảng từ 50 đến 65°C.

3. Trình tự polynucleotit được phân lập theo điểm 1 hoặc đoạn dò axit nucleic được phân lập theo điểm 2, trong đó trình tự này là ADN hệ gen.

4. Trình tự polynucleotit được phân lập theo điểm 1 hoặc đoạn dò axit nucleic được phân lập theo điểm 2, trong đó trình tự này là cADN.

5. Trình tự polynucleotit được phân lập theo điểm 1 hoặc đoạn dò axit nucleic được phân lập theo điểm 2, trong đó trình tự này là ARN.

6. Trình tự polynucleotit được phân lập theo điểm 1 hoặc đoạn dò axit nucleic được phân lập theo điểm 2, trong đó trình tự này là sợi đơn.

7. Phân tử axit nucleic chứa trình tự polynucleotit được phân lập theo điểm 1, không bị gián đoạn bởi codon kết thúc ở trong trình tự mã hóa mà mã hóa cho protein hoặc peptit khác loại.

8. Vật truyền tái tổ hợp gồm trình tự polynucleotit được phân lập theo điểm 1.

9. Cấu trúc biểu hiện chứa trình tự polynucleotit được phân lập theo điểm 1, trong đó trình tự polynucleotit được kết hợp có điều khiển với trình tự nucleotit điều hòa có chứa tín hiệu điều hòa phiên mã hoặc dịch mã, hoặc cả hai, mà kiểm soát sự biểu hiện của trình tự nucleotit ở tế bào chủ.

10. Tế bào chủ được xử lý di truyền chứa trình tự polynucleotit được phân lập theo điểm 1.

11. Phương pháp sản xuất polypeptit bao gồm các bước:

i. nuôi cấy tế bào được biến nạp cấu trúc biểu hiện theo điểm 9 trong điều kiện có hiệu quả để tạo ra polypeptit, và

ii. phân lập polypeptit.

12. Polypeptit được phân lập chứa trình tự axit amin có độ tương đồng ít nhất là 80% so với trình tự axit amin được nêu trong SEQ ID NO. 3 và thể hiện hoạt tính xúc tác của pectat lyaza.

13. Protein khảm chứa polypeptit theo điểm 12 được dung hợp qua liên kết cộng hóa trị với trình tự axit amin của polypeptit thứ hai.

14. Tế bào chủ chứa cấu trúc biểu hiện theo điểm 9.

15. Nấm chuyển gen của *M. phaseolina* bao gồm cấu trúc biểu hiện theo điểm 9.

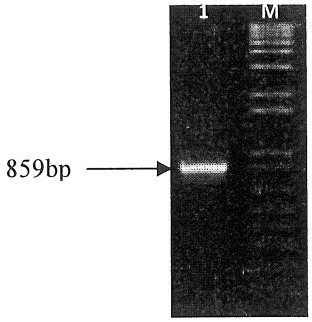

Fig. 1

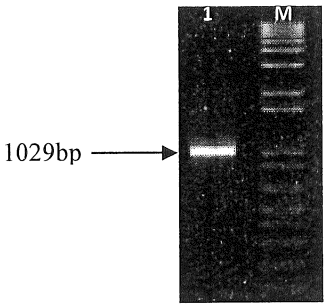

Fig. 2

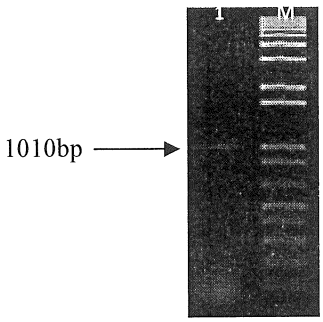

Fig. 3

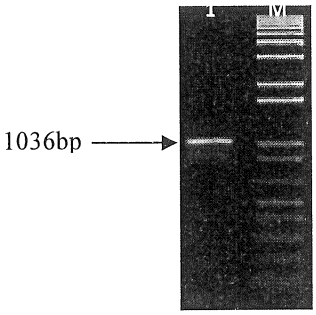

Fig. 4

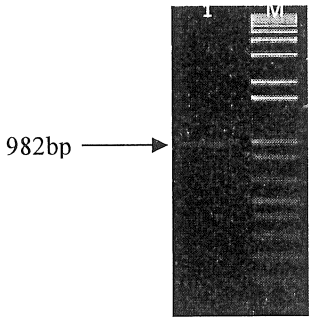

Fig. 5

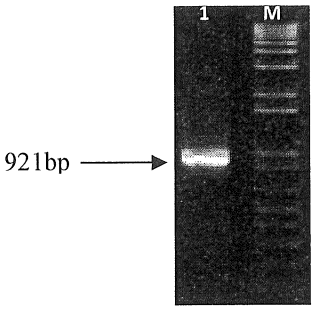

Fig. 6

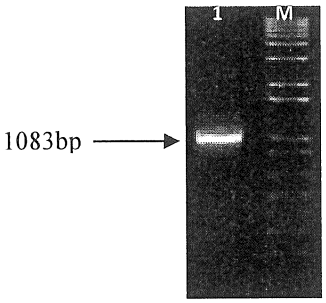

Fig. 7

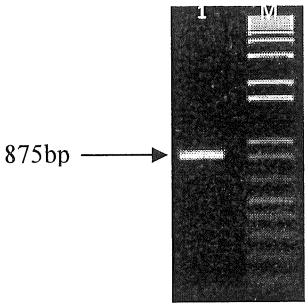

Fig. 8

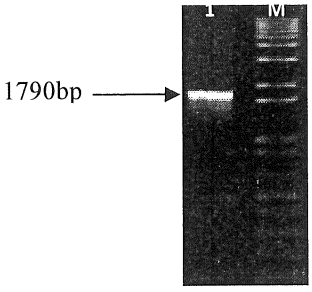

Fig. 9

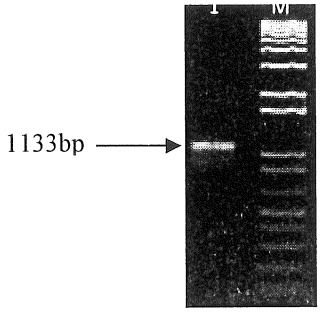

Fig. 10

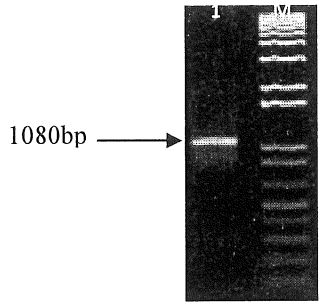

Fig. 11

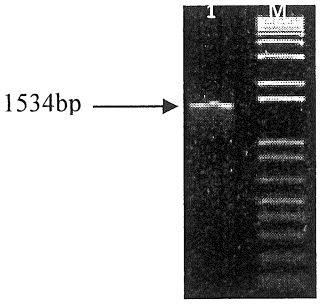

Fig. 12

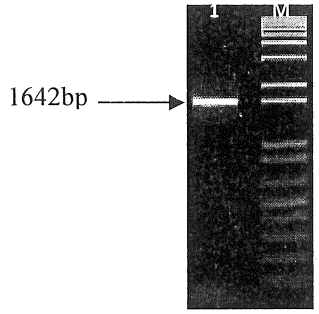

Fig. 13

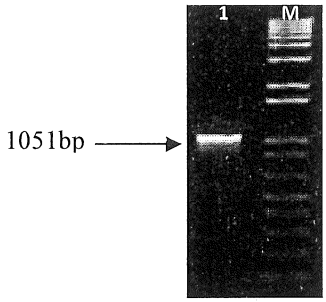

Fig. 14

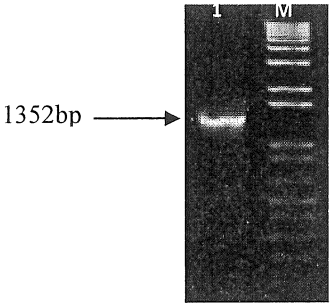

Fig. 15

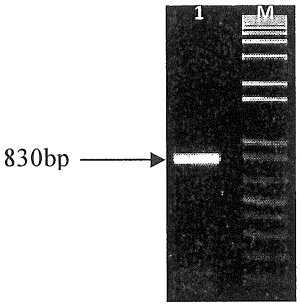

Fig. 16

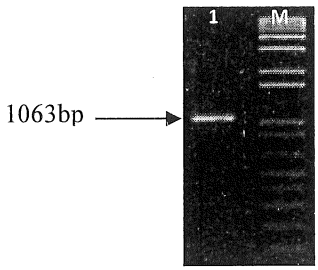

Fig. 17

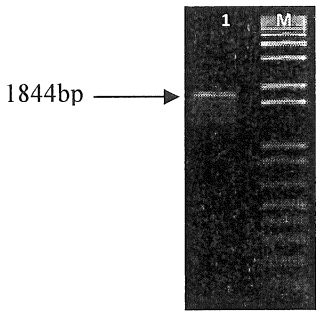

Fig. 18

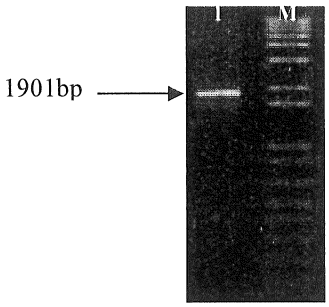

Fig. 19