

## ĐẶC ĐIỂM CỦA ĐOẠN GEN MÃ HÓA COAT PROTEIN PHÂN LẬP TỪ SOYBEAN MOSAIC VIRUS

Lò Thị Mai Thu<sup>1</sup>, Hoàng Hà<sup>2</sup>, Lê Văn Sơn<sup>2</sup>, Chu Hoàng Hà<sup>2</sup>, Chu Hoàng Mậu<sup>3\*</sup>

<sup>1</sup>Trường Đại học Tây Bắc

<sup>2</sup>Viện Công nghệ sinh học, Viện Hàn lâm KH & CN Việt Nam

<sup>3</sup>Đại học Thái Nguyên, \*chuoangmau@tnu.edu.vn

**TÓM TẮT:** Bệnh khảm lá đậu tương do Soybean mosaic virus (SMV) thuộc họ Potyviridae gây ra, là một trong những nguyên nhân chính làm giảm năng suất và chất lượng đậu tương. Để phát hiện và xác định chính xác nguyên nhân gây bệnh khảm trên đậu tương tại một số vùng thuộc tỉnh Sơn La, chúng tôi sử dụng kỹ thuật RT-PCR với cặp mồi thiết kế đặc hiệu cho họ Potyviridae và kết quả đã nhân bản được đoạn gen (cDNA) mã hóa protein vỏ (coat protein-CP) đặc hiệu duy nhất. Kết quả tách dòng và so sánh trình tự nucleotide cho thấy trình tự đoạn gen CP của SMV ở hai dòng SL1 và SL2 có độ tương đồng từ 99,3%-100% so với đoạn gen mã hóa CP của SMV được công bố trên Ngân hàng Gen quốc tế với mã số X63771. SMV dòng SL1, SL2-Việt Nam có quan hệ gần và phân bố cùng nhóm với hai dòng Trung Quốc có mã số X63771 và U25673 trên Ngân hàng Gen quốc tế. Trình tự đoạn gen CP của dòng SL1, SL2 -Việt Nam được sử dụng làm nguyên liệu và thông tin để thiết kế vector mang cấu trúc RNAi phục vụ chuyển gen nhằm cải thiện khả năng kháng SMV của cây đậu tương Việt Nam.

*Từ khóa:* Bệnh khảm, gen CP, *Glycine max*, protein vỏ, Soybean mosaic virus.

### MỞ ĐẦU

Bệnh khảm lá đậu tương do soybean mosaic virus (SMV) thuộc họ Potyviridae gây ra. SMV có thể gây ra thiệt hại đáng kể về sản lượng, làm suy giảm năng suất đậu tương tới 40% khi các cây bị nhiễm trước khi ra hoa, 91% hạt đậu có vết lốm đốm và trong một số trường hợp có thể gây thiệt hại lên tới 94% tổng sản lượng đậu tương [10].

Hạt SMV có dạng hình que, dài 720-850 nm và rộng 11-12 nm, được tạo thành khoảng 2000 tiểu đơn vị tạo nên lớp vỏ protein (coat protein-CP) sắp xếp theo kiểu xoắn ốc xung quanh hệ gen RNA virus [4, 17, 19]. Hệ gen của SMV là RNA sợi đơn dương, có đuôi polyA ở đầu 3' và một protein virion liên kết tại đầu 5'. Đây là loại protein duy nhất bên cạnh CP được phát hiện trong các hạt virus. Hệ gen của SMV chứa vùng bảo thủ (NTRs) theo chiều 5'→3' [17], chiều dài của NTR 5' là 131-205 nucleotide, trong khi NTR 3' của potyviruses khác nhau là hoàn toàn không đồng nhất về kích thước và trình tự. Hệ gen RNA của hai chủng G2 và G7 (thuộc SMV) đã được sắp xếp theo một trật tự [11] và tổng số nucleotide của mỗi chủng là 9588 nucleotide (không tính đuôi poly A), mã hóa cho các phân tử protein có 3066 amino acid. SMV là tác nhân

gây bệnh khảm lá ở đậu tương, lan truyền do rệp làm môi giới. Sự lan truyền cây bệnh sang cây khỏe do rệp ở ngoài đồng vẫn là chủ yếu, bệnh còn truyền qua hạt. Thực tế hiện nay cho thấy, trong sản xuất đậu tương mới dừng lại ở biện pháp phòng là chính mà chưa có thuốc trị SMV, chính vì vậy, nghiên cứu tạo cây đậu tương kháng virus phục vụ công tác chọn tạo giống đậu tương sạch bệnh rất cần thiết. Trong bài báo này chúng tôi trình bày kết quả phân lập trình tự đoạn gen mã hóa protein vỏ của virus gây bệnh khảm lá ở đậu tương trồng ở Sơn La, tạo nguyên liệu để thiết kế vector phục vụ chuyển gen tạo cây đậu tương kháng SMV theo nguyên lý RNAi.

### VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Sử dụng các mẫu lá đậu tương có biểu hiện nhiễm bệnh khảm lá do SMV thu tại huyện Mai Sơn (SL1) và huyện Phù Yên (SL2) tỉnh Sơn La.

Sử dụng bộ kit GeneJET Plant RNA Purification Mini Kit (Hãng Thermo Scientific) để tách RNA tổng số từ mẫu lá đậu tương nhiễm bệnh SMV. Phản ứng RT-PCR được tiến hành trong hai bước: (1). cDNA được tạo ra từ mRNA bằng phiên mã ngược trong 20 µl dung

dịch có chứa 5 µg RNA tổng số, 200 ng mỗi ngẫu nhiên, 2 µl 10 mM dNTPs; bổ sung nước khử ion tới thể tích 13 µl. Dung dịch phản ứng này được ủ ở nhiệt độ 65°C trong 5 phút, sau đó giữ ở nhiệt độ 4°C. Tiếp theo bổ sung 4 µl 5X đệm RT tổng hợp cDNA, 1 µl 0,1 M DTT, 1 µl chất ức chế ribonuclease tái tổ hợp RNase OUTTM (40 đơn vị/µl), 1 µl cloned AMV RT (15 đơn vị/µl) vào dung dịch trên trước khi thực hiện chu kỳ nhiệt như sau: 1 chu kỳ 25°C trong 10 phút, 1 chu kỳ 45°C trong 50 phút, 1 chu kỳ 85°C trong 5 phút và giữ ở 4°C; (2). Phản ứng PCR nhân gen đặc hiệu được thực hiện trong dung dịch bao gồm 28,6 µl nước, 5 µl đệm PCR 10X, 5 µl MgCl<sub>2</sub> 25 mM, 4 µl dNTPs 2,5 mM, 2 µl 10 pmol/µl mỗi loại mỗi SMVcp-F và SMVcp-R, 0,4 µl Taq polymerase 5 u/µl, 4 µl cDNA và theo chu kỳ nhiệt như sau: 01 chu kỳ 94°C trong 3 phút, 35 chu kỳ 94°C trong 1 phút, 57°C trong 45 giây, 72°C trong 45 giây, 01 chu kỳ 72°C trong 5 phút và giữ ở 4°C cho đến khi phân tích.

Sản phẩm PCR được điện di trên gel agarose 1% trong đệm 1X TAE. Gel được nhuộm trong dung dịch ethidium bromide với nồng độ 0,1 µg/ml.

Tách dòng gen được tiến hành theo Sambrook et al. (1989) [18]. Sản phẩm PCR được tách khỏi gel, tinh sạch bằng bộ kit của Qiagen và gắn trực tiếp vào vector tách dòng pBT, sau đó được biến nạp vào tế bào *E. coli* DH5α. Chọn lọc các khuẩn lạc dương tính (khuẩn lạc màu trắng) trên môi trường LB có bổ sung 50 mg/l ampicillin, 30 mg/l X-gal và 0,1 mM IPTG. Sự có mặt của DNA tái tổ hợp được

kiểm tra bằng phản ứng colony-PCR, sử dụng trực tiếp các khuẩn lạc dương tính và cặp mồi đặc hiệu.

DNA plasmid được tách chiết theo kit QIAprep Spin Miniprep. Chọn các mẫu DNA plasmid tái tổ hợp có kích thước như mong muốn để tinh sạch và đem xác định trình tự nucleotide. Phân tích, so sánh trình tự nucleotide của đoạn gen bằng phần mềm BioEdit.

## KẾT QUẢ VÀ THẢO LUẬN

### Khuếch đại đoạn gen mã hoá protein vỏ (coat protein-CP) từ hệ gen của SMV

Tách chiết RNA tổng số từ lá đậu tương nghi nhiễm SMV từ hai mẫu SL1, SL2 thu ở Sơn La, sau đó điện di kiểm tra trên gel agarose 1%, kết quả cho thấy sản phẩm tách chiết RNA tổng số đảm bảo chất lượng để tạo cDNA và tiến hành phản ứng PCR.

Từ những thông tin về trình tự đoạn gen CP của SMV ở đậu tương đã công bố tại Ngân hàng Gen quốc tế, sử dụng BLAST trong NCBI so sánh 96 trình tự gen CP của SMV cho thấy các trình tự gen CP trên GenBank có độ tương đồng từ 94% đến 100% và kích thước vùng tương đồng được xác định có số lượng hơn 700 nucleotide. Từ kết quả so sánh và xác định vùng tương đồng của các trình tự gen CP và dựa vào trình tự gen CP có mã số X63771, chúng tôi đã thiết kế cặp mồi đặc hiệu (SMVcp-F/SMVcp-R) để nhân bản đoạn gen CP với kích thước dự tính khoảng 0,7 kb. Trình tự cặp mồi PCR được thiết kế tại các vùng gen có độ bảo thủ cao nhất và được trình bày trong bảng 1.

Bảng 1. Trình tự cặp mồi thiết kế và sử dụng trong nghiên cứu

Mồi	Trình tự
SMVcp-F	5'-TTACCATCAGGCAAGGAGCAGGAAG-3'
SMVcp-R	5'-TGGACTTGATGGGAACATCTCAACT-3'

Phản ứng PCR nhân bản đoạn để khuếch đại đoạn cDNA được thực hiện với cặp mồi đã thiết kế SMVcp-F/SMVcp-R, nhiệt độ gắn mồi khả dụng cho phản ứng PCR là 57°C, sau 35 chu kỳ phản ứng. Kết quả kiểm tra sản phẩm RT-PCR bằng điện di trên gel agarose 1% ở hình 1 cho

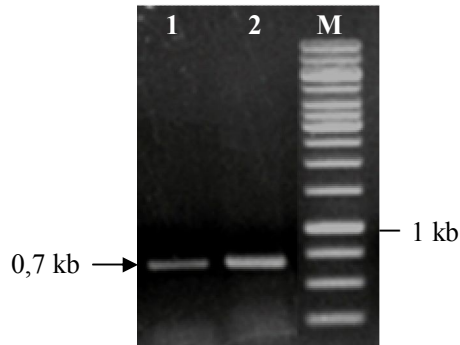
thấy có một băng DNA đặc hiệu với kích thước khoảng 0,7 kb, tương ứng với kích thước dự tính khi thiết kế cặp mồi PCR.

### Tách dòng và xác định trình tự đoạn gen CP của SMV dòng SL1 và dòng SL2

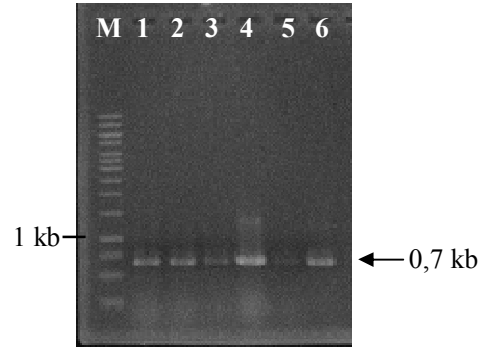
Sản phẩm PCR được tách khỏi bản gel, tinh

sạch bằng bộ kit của Qiagen và gắn trực tiếp vào vector tách dòng pBT, sau đó được biến nạp vào tế bào *E. coli* DH5 $\alpha$ . Tiến hành chọn dòng bằng phản ứng colony-PCR trực tiếp từ khuẩn

lạc với cặp mồi đặc hiệu. Kết quả điện di kiểm tra sản phẩm colony-PCR trên gel agarose 1% cho thấy xuất hiện một băng DNA có kích thước khoảng 0,7 kb (hình 2).



Hình 1. Hình ảnh điện di sản phẩm RT-PCR khuếch đại đoạn gen CP (cDNA) của SMV  
1, 2. SL1, SL2; M. Thang DNA chuẩn 1 kb.



Hình 2. Hình ảnh điện di sản phẩm colony-PCR từ khuẩn lạc  
M. thang DNA chuẩn 1 kb; 1-6. gen CP khuếch đại từ khuẩn lạc

```

10      20      30      40      50      60      70
X63771  TTACCATCAG GCAAGGAGCA GGAAGGAGG ATGGATACAG GTAAGGATCC AAAGAAGAGC ACCAGCAGTA
SL1     .....
SL2     ..... T.....G.....

80      90      100     110     120     130     140
X63771  GCAAGGAGGC TGGTACAGC AGCAAGAGTG TAAATGTTGG ATCAAAGGGA AAGGTGGTGC CGCGTTTGCA
SL1     .....
SL2     .....

150     160     170     180     190     200     210
X63771  GAAGATTACA AGAAGATGA ATCTTCCAAT GGTGAAGGGG AAGATCATTC TCAGTTTGGG CCACTTGCTT
SL1     .....
SL2     .....

220     230     240     250     260     270     280
X63771  GAGTATAAAG CTAATCAGGT TGAATTTATTC ANTACTCGAG CAAGCAAGAGC AGAATTCGAG CGGTGGTACA
SL1     .....
SL2     .....

290     300     310     320     330     340     350
X63771  ATGCAGTTAA AGATGAATAT GAGCTTGATG ATGAACAGAT GGGTGTGCTT ATGAATGGTT TCATGGTATG
SL1     .....
SL2     .....

360     370     380     390     400     410     420
X63771  GTGCATTGAC AATGGTACAT CTCAGATGCG CAATGGCGTG TGGGTGATGA TGCATGGAGA AGAACAGATT
SL1     .....
SL2     .....

430     440     450     460     470     480     490
X63771  GCATATCCGC TGA AACCCAT TGTGAGAAAT GCAAAACCAA CTCTGAGACA ACTCATGCAC CACTTCTCAG
SL1     .....
SL2     .....

500     510     520     530     540     550     560
X63771  ATGCAGGAGA AGCTTACATT GAGATGAGAA ATTCTGAAG TCCGTATATG CCTAGATATG GACTACTGAG
SL1     .....
SL2     .....

570     580     590     600     610     620     630
X63771  GAATTGAGA  GATAGAGAGC TAGCAAGCTA TGCTTTTGAT TTCTATGAGG TFACTTCCAA AACACCAAAC
SL1     .....
SL2     .....

640     650     660     670     680     690     700
X63771  AGGGCAAGGG AAGCAATAGC GCAGATGAG GCTGCAGCTC TTTCGGGAGT TAACAACAAG TTGTTTGGAC
SL1     .....
SL2     ..... G..T..A..

710     720
X63771  TTGATGGGAA CACTCTCACT
SL1     .....
SL2     .....
    
```

Hình 3. Trình tự đoạn gen CP (cDNA) của SMV phân lập tại tỉnh Sơn La và trình tự đoạn gen CP của dòng virus SMV được đăng ký GenBank có mã số X63771 X63771. Trình tự gen CP của SMV trên GenBank; SL1, SL2. SMV dòng SL1, SL2-Việt Nam.

Những dòng khuẩn lạc dương tính với PCR đã được chọn để tách plasmid tái tổ hợp và đem giải trình tự nucleotide. Kết quả xác định trình tự nucleotide cho thấy, đoạn gen *CP* phân lập từ hệ gen SMV dòng SL1, SL2 gây bệnh khảm lá trên cây đậu tương thu tại Sơn La có 720 nucleotide (hình 3).

Kết quả so sánh với trình tự của đoạn gen *CP* trên Ngân hàng Gen quốc tế có mã số

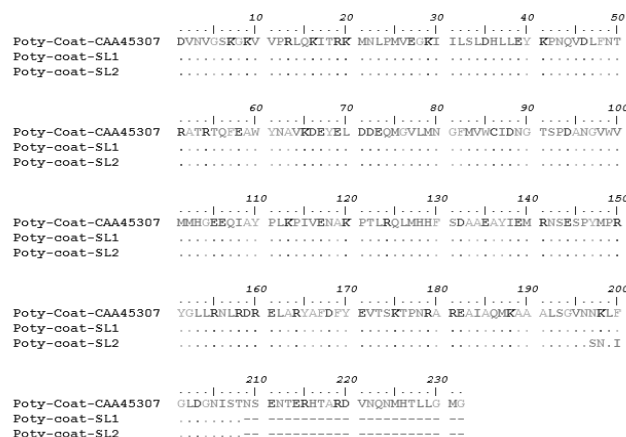
X63771 (trình tự sử dụng để thiết kế cặp mồi nhân đoạn gen *CP*) bằng BLAST trong NCBI cho thấy đoạn gen *CP* (cDNA) mà chúng tôi phân lập được từ SMV dòng SL1 có độ tương đồng là 100%, còn trình tự đoạn gen *CP* phân lập từ dòng SL2 có độ tương đồng là 99,3%. Hình 3 cho thấy, trình tự đoạn gen *CP* của dòng virus SL2 có sự sai khác so với trình tự gen *CP* có mã số X63771 và dòng SL1 ở 5 vị trí nucleotide: 31, 38, 686, 690, 694 (bảng 2).

**Bảng 2.** Các vị trí sai khác giữa trình tự đoạn gen *CP* của SMV dòng SL2 so với dòng SL1 và trình tự có mã số X63771

Vị trí nucleotide	X63771	SL1	SL2
31	A	A	T
38	C	C	G
686	A	A	G
690	G	G	T
694	T	T	A

**Bảng 3.** Các vị trí sai khác giữa trình tự amino acid của protein suy diễn dòng SL2 so với dòng SL1 và trình tự có mã số CAA45307

Vị trí amino acid	CAA45350	SL1	SL2
11	M	M	L
13	T	T	R
220	N	N	S
230	K	K	N
232	F	F	I



**Hình 4.** So sánh trình tự amino acid của vùng Poty-coat của dòng SL1, SL2 và của protein có mã số CAA45307 trên GenBank

Poty-coat- CAA45307: Trình tự amino acid của vùng poty-coat của *CP* dòng SMV có mã số CAA45307 trên GenBank; Poty-coat-SL1: Trình tự amino acid của protein suy diễn từ đoạn gen *CP* của SMV dòng SL1; Poty-coat-SL2: Trình tự amino acid của protein suy diễn từ đoạn gen *CP* của SMV dòng SL2

Protein *CP* có mã số CAA45307 trên GenBank được suy diễn từ trình tự gen *CP* của SMV, mã số là X63771, có 267 amino acid; còn protein suy diễn từ trình tự đoạn gen *CP* của SMV dòng SL1 và SL2 mà chúng tôi phân lập có 240 amino acid, kết quả so sánh vùng tương đồng của ba trình tự amino acid cho thấy, *CP* của dòng SMV, mã số CAA45307, có hệ số tương đồng so với protein suy diễn của dòng SL1 là 100% và so với dòng SL2 là 97,9%. Bảng 3 thể hiện 5 vị trí sai khác (11, 13, 220, 230, 232) về trình tự amino acid của protein suy diễn dòng SL2 so với protein CAA45307 và dòng SL1.

Vùng Poty-coat của *CP* ở SMV, mã số CAA45307, có 232 amino acid, vùng Poty-coat của *CP* dòng SL1 và dòng SL2 có 208 amino acid. Kết quả so sánh vùng tương đồng về trình tự amino acid của vùng Poty-coat của *CP* dòng SL1 và dòng SL2 so với vùng Poty-coat của *CP* ở SMV có mã số CAA45307 trên GenBank cho thấy dòng SL2 có sự sai khác ở 3 vị trí amino acid

so với CAA45307 và dòng SL1 (hình 4).

#### Phân tích sự đa dạng của trình tự đoạn gen *CP* của các dòng SMV

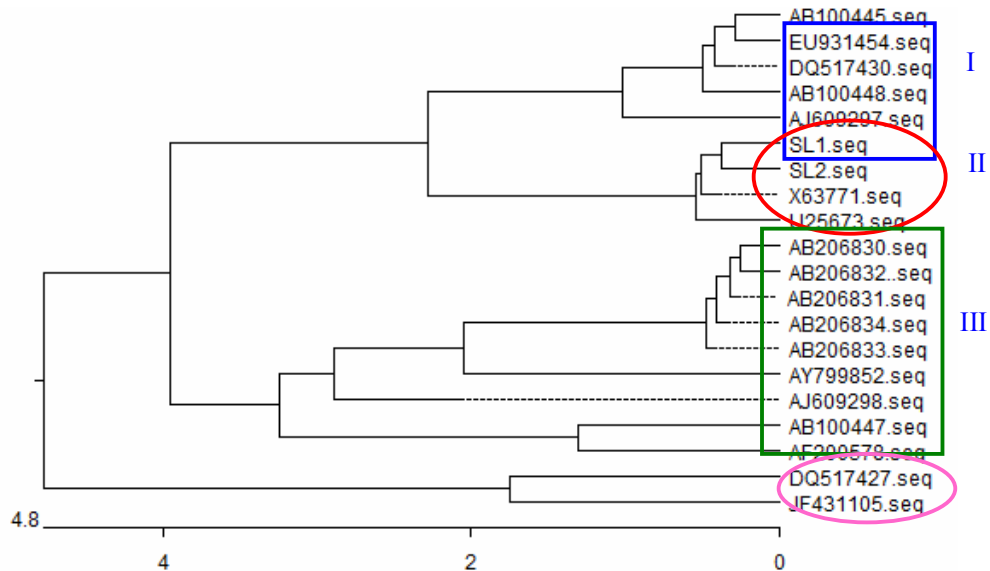
Chúng tôi tiến hành so sánh trình tự nucleotide của đoạn gen *CP* của SMV dòng SL1 và SL2 phân lập từ Sơn La, Việt Nam với 18 trình tự gen *CP* đã công bố trên GenBank (bảng 4) để xác định hệ số tương đồng và hệ số sai khác của các trình tự gen *CP*, đồng thời thiết lập sơ đồ hình cây để phân tích sự đa dạng của các dòng virus thông qua trình tự gen *CP* của SMV. Trong các trình tự của đoạn gen *CP* thuộc 20 dòng SMV được sử dụng so sánh có 2 dòng Việt Nam, 8 dòng Nhật Bản, 2 dòng Hàn Quốc, 2 dòng Mỹ, 4 dòng Trung Quốc, 1 dòng Canada, 1 dòng Ucraina, kết quả so sánh cho thấy, hệ số tương đồng giữa các cặp so sánh giao động từ 89,8% đến 99,7%; còn hệ số sai khác từ 0,3% đến 12,3%. Mối quan hệ giữa các dòng SMV dựa trên cơ sở so sánh trình tự nucleotide của đoạn gen *CP* được thể hiện ở hình 5.

**Bảng 4.** Các trình tự đoạn gen *CP* của hai dòng virus ở Việt Nam và các trình tự có mã số trên GenBank được sử dụng trong phân tích so sánh [23]

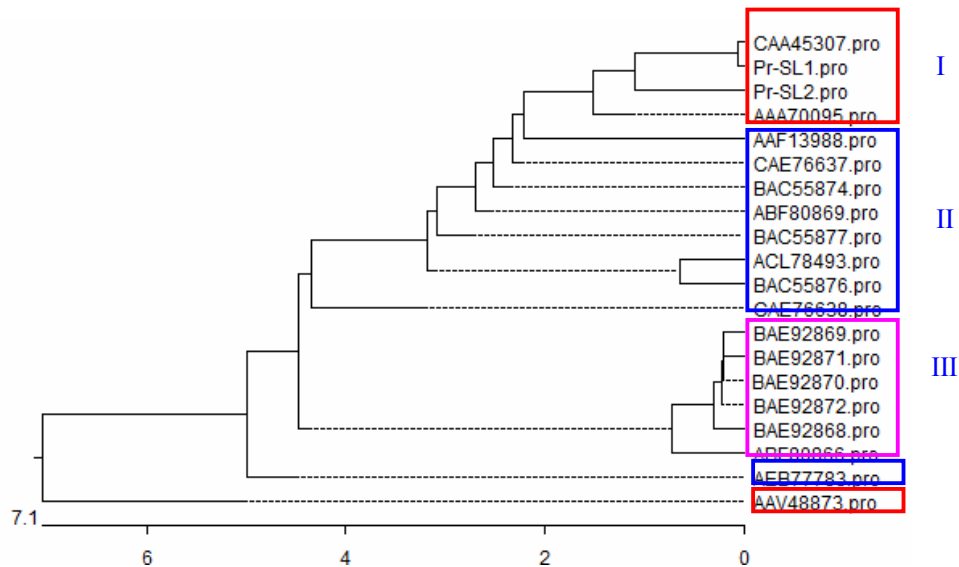
STT	Dòng virus/ Mã số trên GenBank	Vùng phân lập	Năm công bố	Tác giả
1	HG965102 (SL1)	Sơn La, Việt Nam	2013	Lo et al.
2	HG965103 (SL2)	Sơn La, Việt Nam	2013	Lo et al.
3	X63771	Trung Quốc	1992	Chu
4	U25673	Trung Quốc	1995	Chu et al.
5	AB100445	Nhật Bản	2003	Kanematsu et al.
6	AB100447	Nhật Bản	2003	Kanematsu et al.
7	AB100448	Nhật Bản	2003	Kanematsu et al.
8	AB206830	Nhật Bản	2008	Saruta et al.
9	AB206831	Nhật Bản	2008	Saruta et al.
10	AB206832	Nhật Bản	2008	Saruta et al.
11	AB206833	Nhật Bản	2008	Saruta et al.
12	AB206834	Nhật Bản	2008	Saruta et al.
13	AF200578	Hoa Kỳ	1999	Latorre
14	AJ609297	Hàn Quốc	2003	Choi
15	AJ609298	Hàn Quốc	2003	Choi
16	AY799852	Hoa Kỳ	2004	Fayad và Tolin
17	DQ517427	Trung Quốc	2006	Wang
18	DQ517430	Trung Quốc	2006	Wang
19	EU931454	Canada	2009	Stromvik et al.
20	JF431105	Ukraine	2012	Sherepitko et al.

Sơ đồ hình cây ở hình 5 dựa trên so sánh trình tự nucleotide của đoạn gen CP cho thấy 20 dòng virus phân bố ở 4 nhóm (I, II, III, IV) với khoảng cách di truyền 4,8%; dòng SL1, SL2

(Việt Nam) và hai dòng có mã số X63771, U25673 (Trung Quốc) phân bố trong cùng một nhóm (Nhóm II) với hệ số tương đồng khoảng 99%.



Hình 5. Sơ đồ hình cây về mối quan hệ di truyền giữa các dòng SMV thiết lập dựa trên trình tự đoạn gen CP phân lập từ 20 dòng SMV từ Việt Nam, Nhật Bản, Hàn Quốc, Mỹ, Trung Quốc, Canada, Ucraina SL1- dòng SL1; SL2- dòng SL1; 18 dòng SMV có mã số trên GenBank.



Hình 6. Sơ đồ hình cây thiết lập dựa trên trình tự amino acid suy diễn từ gen CP phân lập từ 20 dòng SMV từ Việt Nam, Nhật Bản, Hàn Quốc, Mỹ, Trung Quốc, Canada, Ucraina SL1- dòng SL1; SL2- dòng SL2; 19 dòng SMV mang mã số trên GenBank.

Tiếp tục phân tích mối quan hệ của 20 dòng SMV dựa trên trình tự amino acid suy diễn, hình 6 cho thấy 20 trình tự amino acid chia làm hai nhánh, nhánh thứ nhất chỉ có một dòng, nhóm V (AAV48873) và nhánh thứ hai có 19 dòng còn lại (Nhóm I, II, III, IV) với khoảng cách di truyền là 7,1%. Hai dòng SL1, SL2 (Việt Nam) và hai dòng Trung Quốc có quan hệ gần nhau, phân bố trong nhóm I.

Khoảng cách di truyền được tính trên cơ sở so sánh trình tự nucleotide của đoạn gen *CP* giữa 20 dòng SMV là 4,8% (hình 5), còn dựa trên trình tự amino acid, khoảng cách di truyền giữa 20 dòng SMV lại lớn hơn rất nhiều (7,1%) (hình 6). Kết quả phân tích phù hợp với nhận xét cho rằng các potyvirus lây nhiễm trên cây họ đậu và các loài thực vật khác có sự biểu hiện sự đa dạng về đặc tính sinh học hơn là sự đa dạng về trình tự nucleotide trong nucleic acid [1]. Để phân biệt các loài, các chủng virus có thể tiến hành so sánh trình tự nucleotide, đặc biệt là vùng đầu 3' không mã hóa của hệ gen virus và trình tự gen *CP* [1], bởi vì thành phần và trình tự amino acid của *CP* có tính đặc trưng cho loài virus và cho từng cá thể. Trong thực tế rất ít có tương đồng về trình tự *CP* giữa các chi virus thực vật khác nhau. Từ cơ sở này mà người ta tiến hành phân loại các thành viên của nhóm potyvirus có thể dựa trên trình tự gen *CP* và trình tự amino acid của *CP* [1, 20]. Như vậy có thể nói, cách tiếp cận phân tích trình tự amino acid của *CP* và trình tự nucleotide của gen *CP* từ các potyvirus đã được coi là công cụ mạnh, không chỉ để nghiên cứu phân loại các loài virus thuộc chi, mà còn đối với các chủng trong potyvirus, nhận xét này cũng tìm thấy trong báo cáo của McKern et al. (1993) [15]. Phân tích sự phát sinh loài dựa trên trình tự nucleotide của Potyvirus cho thấy, các thành viên có trình tự với hệ số tương đồng là 38-71% được nhận dạng là các loài virus khác nhau; còn các thành viên có trình tự với hệ số tương đồng ít nhất là 90% được nhận dạng là cùng chủng virus [20]. Các chủng virus riêng biệt cũng đã được phân biệt bởi sự thay đổi trong hệ gen của chúng. Gillings et al. (2009) đã sử dụng enzyme giới hạn *RsaI* và *HinfI* cắt sản phẩm PCR khuếch đại gen *CP* từ *Citrus tristeza virus* nhận thấy đã có sự thay đổi về trình tự gen *CP* [6].

Khi nghiên cứu so sánh các trình tự amino acid của *CP* của các dòng virus, Gough & Shukla (1992) [7] cho rằng có sự khác biệt lớn về kích thước và trình tự amino acid của *CP* ở vùng đầu N, nhưng lại có sự tương đồng cao về trình tự ở vùng đầu C của *CP*. Như vậy, trình tự vùng nucleotide của gen *CP* mã hóa cho vùng có độ tương đồng cao ở đầu C của *CP* có thể sử dụng theo hai cách tiếp cận, đó là (1) khuếch đại đoạn bảo thủ để tạo đoạn gen CPi (đoạn gen *CP* interference) và (2) sử dụng như nguồn thông tin để thiết kế đoạn gen bảo thủ CPi phục vụ tạo vector mang cấu trúc RNAi. Đối với virus gây bệnh khảm trên cây đậu tương, kết quả tách dòng và xác định trình tự đoạn gen *CP* từ SMV dòng SL1, SL2 của chúng tôi cũng nhằm phục vụ phân tích sự đa dạng của các dòng SMV trên cây đậu tương Việt Nam và xác định vùng tương đồng để cung cấp thông tin cho việc thiết kế đoạn gen CPi của SMV trong kỹ thuật RNAi. Hướng nghiên cứu này cũng đã được thực hiện đối với virus gây bệnh đốm vòng đu đủ, virus gây bệnh khảm dưa chuột, virus gây bệnh xoắn vàng lá cà chua [8, 9, 21, 22].

Berger et al. (1997) đã phân tích trình tự hệ gen của một số potyvirus ở cây họ đậu và cây phát sinh loài đã được thiết lập [1]. Cách tiếp cận phân tích sự phát sinh loài cũng được một số tác giả sử dụng để phân biệt các potyvirus khác nhau [2]. Ở Việt Nam một số tác giả cũng đã sử dụng gen *CP* để phân tích sự đa dạng của các loài virus gây bệnh thực vật [9, 21]. Thiết lập cây phát sinh của các dòng virus gây bệnh xoắn lá cà chua của Lê Thị Hồng Ngọc và nnk. (2005) cho thấy các dòng virus ở Việt Nam gần với các dòng virus của Trung Quốc, Đài Loan [16]. Trong nghiên cứu của chúng tôi, trình tự đoạn gen *CP* phân lập từ SMV dòng SL1 và SL2, Việt Nam có kích thước 720 nucleotide, mã hoá 240 amino acid và vùng Poty-coat của coat protein có 208 amino acid có quan hệ gần và phân bố cùng nhóm với hai dòng SMV của Trung Quốc có mã số X63771 và U25673 trên Ngân hàng Gen quốc tế. Nhận xét này phù hợp với báo cáo của Kirthi et al. (2004) [12] cho rằng, các chủng virus phân lập từ những vùng địa lý càng gần nhau thì có độ tương đồng di truyền càng cao. Tuy nhiên, một số nghiên cứu lại cho rằng do có sự tái tổ hợp thường xuyên

trong hệ gen mà ngày nay các loài virus ngày càng có hệ gen đa dạng hơn [5, 13, 14].

## KẾT LUẬN

Trình tự đoạn gen *CP* phân lập từ SMV dòng SL1 và dòng SL2 có kích thước 720 nucleotide, mã hoá 240 amino acid và vùng Poty-coat của coat protein có 208 amino acid. SMV dòng SL1 và dòng SL2-Việt Nam có quan hệ gần và phân bố cùng nhóm với hai dòng SMV Trung Quốc có mã số X63771 và U25673 trên Ngân hàng Gen quốc tế. Trình tự đoạn gen *CP* của dòng SL1 và SL2-Việt Nam đã đăng ký trên Ngân hàng gen quốc tế với mã số HG965102, HG965103 và được sử dụng làm nguyên liệu và cung cấp thông tin để thiết kế vector mang cấu trúc RNAi phục vụ chuyển gen nhằm cải thiện khả năng kháng virus SMV của cây đậu tương Việt Nam.

**Lời cảm ơn:** Công trình được hoàn thành có sự dụng thiết bị Phòng thí nghiệm trọng điểm về công nghệ gen, Viện Công nghệ sinh học và được sự tài trợ của đề tài Khoa học-Công nghệ cấp Bộ Giáo dục & Đào tạo, mã số B2013-TN04-05.

## TÀI LIỆU THAM KHẢO

- Berger P. H., Wyatt S., Shiel P. J., Silbernagel M. J., Druffel K., Mink G. I., 1997. Phylogenetic analysis of the Potyviridae with emphasis on legume-infecting potyviruses. *Arch. Virol.*, 142: 1979-1999.
- Bousalem M., Douzery E. J. P., Fargette D., 2000. High genetic diversity, distant phylogenetic relationships and intraspecies recombination events among natural populations of Yam mosaic virus: a contribution to understanding potyvirus evolution. *J. Gen. Virol.*, 81: 243-255.
- Chen K. C., Lin C. Y., Kuan C. C., Sung H. Y., Chen C. S., 2002. A novel defensin encoded by a mungbean cDNA exhibits insecticidal activity against bruchid. *Journal of Agricultural and Food Chemistry*, 50(25): 7258-7263.
- Dougherty W. G., Carrington J. C., 1988. Expression and function of potyviral gene products. *Annu. Rev. Phytopath.*, 26: 123-143.
- Fondong V. N., Pita J. S., Rey M. E. C., de Kochko A., Beachy R. N., Fauquet C. M., 2000. Evidence of synergism between African cassava mosaic virus and a new double-recombinant geminivirus infecting cassava in Cameroon. *J. Gen. Virol.*, 81: 287-297.
- Gillings M., Broadbent P., Indsto J., Lee R., 1993. Characterisation of isolates and strains of citrus tristeza closterovirus using restriction analysis of the coat protein gene amplified by the polymerase chain reaction. *J. Virol. Meth.*, 44: 305-317.
- Gough K. H., Shukla D. D., 1992. Major sequence variations in the N-terminal region of the capsid protein of a severe strain of passion fruit woodiness potyvirus. *Arch. Virol.*, 124: 389-396.
- Chu Hoàng Hà, Đỗ Xuân Đồng, Phạm Bích Ngọc, Lâm Đại Nhân, Lê Văn Sơn, Lê Trần Bình, 2011. Nghiên cứu tạo giống đu đủ kháng bệnh đốm vòng ứng dụng cơ chế RNAi. *Hội thảo Quốc gia bệnh hại thực vật Việt Nam*, 316-326.
- Chu Hoàng Hà, Nguyễn Minh Hùng, Bùi Chi Lăng, Lê Trần Bình, 2004. Đánh giá tính đa dạng của các dòng virus gây bệnh đốm vòng đu đủ của Việt Nam thông qua tách dòng, xác định và so sánh trình tự gen mã hoá protein vỏ (CP). *Tạp chí Công nghệ sinh học*, 2(4):451-459.
- Hill J., 1999. *Soybean mosaic*. In Hartman G. L., Sinclair, J. B., Rupe J. C (eds.) *Compendium of soybean diseases*. 4th edition. APS Press. St. Paul, MN, 70-71.
- Jayaram C., Hill J. H., Miller W. A., 1992. Complete nucleotide sequences of two soybean mosaic virus strains differentiated by response of soybean containing the Rsv resistance gene. *J. Gen. Virol.*, 73: 2067-2077.
- Kirithi N., Priyadarshini C. G., Sharma P., Maiya S. P., Hemalatha V., Sivaraman P.,



- Dhawan P., Rishi N., Savithri H. S., 2004. Genetic variability of begomoviruses associated with cotton leaf curl disease originating from India. *Arch. Virol.*, 149 (10): 2047-2057.
13. Kirthi N., Maiya S. P., Murthy M. R., Savithri H. S., 2002. Evidence for recombination among the tomato leaf curl virus strains/species from Bangalore, India. *Arch. Virol.*, 147(2): 255-272.
14. Maruthi M. N., Seal S., Colvin J., Briddon R. W., Bull S. E., 2004. East African cassava mosaic Zanzibar virus - a recombinant begomovirus species with a mild phenotype. *Arch. Virol.*, 149: 2365-2377.
15. McKern N. M., Barnett O. W., Whittaker L. A., Mishra A., Strike P. M., Xiao X. W., Ward C. W., Shukla D. D., 1993. Sequence relationship among the coat proteins of strains of peamosaic, white lupin mosaic and bean yellow mosaic potyviruses. *Phytopathology*, 83: 255-361.
16. Lê Thị Hồng Ngọc, Chu Hoàng Hà, Hà Thị Thanh Bình, Nguyễn Thị Loan, Nguyễn Quốc Thông, Lê Trần Bình, 2005. Giải mã đoạn gen protein vỏ của virus gây bệnh xoắn lá ở vùng chuyên canh cà chua thuộc tỉnh Hưng Yên. *Tạp chí công nghệ Sinh học*, 3: 223-230.
17. Riechmann J. L., Lain S., Garcia J. A., 1992. Highlights and prospects of potyvirus molecular biology. *J. Gen. Virol.*, 73:1-16.
18. Sambrook J., Fritschi E. F., Maniatis T., 1989. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.
19. Shukla D. D., Ward C. W., Brunt A. A., 1994. *The Potyviridae*, CAB International, Wallingford, UK, 516p.
20. Shukla D. D., Ward C. W., 1988. Amino acid sequence homology of coat proteins as a basis for identification and classification of the potyvirus group. *J. Gen. Virol.*, 69: 2703-2710.
21. Nguyễn Thị Tâm, Nguyễn Vũ Thanh Thanh, Chu Hoàng Mậu, 2010. Phân lập và xác định trình tự gen mã hoá protein vỏ của virus Y ở khoai tây trồng tại Thái Nguyên. *Tạp chí Sinh học* 32(1): 81-87.
22. Nguyễn Thị Hải Yến, Phạm Thị Vân, Chu Hoàng Hà, Chu Hoàng Mậu, Lê Trần Bình, 2008. Phân lập gen mã hóa protein vỏ của virus gây bệnh xoắn vàng lá cà chua thu thập trên cây cà chua dại tại tỉnh Thái Nguyên. *Tạp chí Công nghệ sinh học*, 6(4): 467-474.
23. [Http://www.ncbi.nlm.nih.gov/nuccore/](http://www.ncbi.nlm.nih.gov/nuccore/)

## CHARACTERISTICS OF GENE ENCODING COAT PROTEIN ISOLATED FROM SOYBEAN MOSAIC VIRUS

Lo Thi Mai Thu<sup>1</sup>, Hoang Ha<sup>2</sup>, Le Van Son<sup>2</sup>, Chu Hoang Ha<sup>2</sup>, Chu Hoang Mau<sup>3\*</sup>

<sup>1</sup>Tay Bac University

<sup>2</sup>Institute of Biotechnology, VAST

<sup>3</sup>Thai Nguyen University of Education

### SUMMARY

Leaf mosaic disease of soybean is caused by soybean mosaic virus (SMV) from the family Potyviridae and is one of the main causes of reduced productivity and quality of soybean. To detect and determine the exact cause of mosaic disease on soybean leaves in some areas of Son La province, we used RT-PCR technique with specifically designed primers for Potyviridae; consequently, *CP* gene (cDNA) encoding the coat protein (CP) has been amplified. Results of cloning and comparison of nucleotide sequences showed the sequences of *CP* gene of SMV lines SL1 and SL2 had the similarity of 99.3-100% compared to that published

in the GenBank with code number X63771. SMV lines SL1 and SL2 (Vietnam) have close relations and the same group of allocation with two Chinese lines, X63771 and U25673 in the GenBank. The nucleotide sequence of the *CP* gene of SMV line SL-Vietnam will be used as material in constructing RNAi vectors for gene transfer to enhance SMV-antiviral activity of soybean in Vietnam.

*Keywords:* *Glycine max*, coat protein, *CP* gene, mosaic disease, soybean mosaic virus.

*Ngày nhận bài:* 15-9-2013