

NGHIÊN CỨU ĐẶC ĐIỂM TRÌNH TỰ GIEN CYSTATIN CỦA MỘT SỐ DÒNG LẠC CÓ NGUỒN GỐC TỪ MÔ SẸO CHỊU CHIẾU XẠ VÀ XỬ LÝ MẤT NƯỚC

VŨ THỊ THU THỦY, NGUYỄN THỊ TÂM

Trường đại học Sư phạm, Đại học Thái Nguyên

CHU HOÀNG MẬU

Đại học Thái Nguyên

NGUYỄN VŨ THANH THANH

Trường đại học Khoa học, Đại học Thái Nguyên

Lạc là cây công nghiệp ngắn ngày có giá trị kinh tế về nhiều mặt và được trồng ở nhiều nơi trên thế giới. Với mục đích nghiên cứu nâng cao khả năng chịu hạn của cây lạc, bài báo này trình bày kết quả phân tích trình tự gen cystatin- một gen được bàn luận nhiều trong những năm gần đây bởi mối liên quan của nó với tính chống chịu của thực vật [1, 2, 9, 10, 12, 17]....

Cystatin là một loại protein phổ biến trong cơ thể động vật, thực vật.... Đó là tên gọi chung cho các chất có bản chất protein ức chế hoạt động của cystein protease. Các chất ức chế được chia ít nhất làm 3 họ nhỏ do có sự sai khác nhau chủ yếu về trình tự tương đồng của các amino acid, số liên kết disunfit và khối lượng phân tử của protein. Trong đó, cystatin họ 1 có tên gọi là stefin là nhóm có cấu tạo khoảng 100 amino acid, khối lượng khoảng 11 kDa, không có cầu disunfit và các nhóm cacbonhydrat trong thành phần cấu tạo; cystatin họ 2 gọi tên cystatin là nhóm chứa khoảng 120 - 126 amino acid, khối lượng đạt 13,4-14,4 kDa và hầu hết không được bổ sung hydrat cacbon; nhóm cystatin thứ 3 có tên là kininogen là nhóm lớn nhất gồm những glycoprotein có khối lượng khoảng 60-120 kD [14]. Những chất ức chế cystatin protease của thực vật được mã hóa bởi 1 họ gen, tuy nhiên những hiểu biết về những gen này còn rất hạn chế [7, 8, 14].... Sự biểu hiện của gen ức chế protease thường bị giới hạn bởi các tổ chức đặc biệt hoặc không giống nhau trong những giai đoạn phát triển khác nhau của cây như: giai đoạn hạt nảy mầm, quá trình rụng lá sớm [9, 12]... ở thực vật cystatin được gọi là phytocystatin (phycys) và tạo thành một phân họ

độc lập trên cây phát sinh loài của gia đình cystatin. Phytocystatin gồm có nhiều loại, tuy nhiên điểm chung của chúng là tất cả đều chứa thêm so với ở động vật một motif LARFAV bảo thủ ở đoạn xoắn anpha vùng cuối N [12]. Các loại phytocystatin khác nhau có kích thước giao động từ 12 đến 85 kDa, cũng chia làm 3 nhóm: Nhóm 1 và nhóm 2 có kích thước 12-16 kDa, không có cầu disunfide và vùng glycosilation giả định, gen mã hoá chúng không có intron hoặc có 1 intron ở đoạn giữa LARFAV và QxVxG; nhóm thứ 3 có 3 intron trên gen mã hoá với kích thước 23 kDa, có một phần mở rộng đầu C [7, 8]. Phytocystatin cũng thực hiện nhiều chức năng sinh lý quan trọng như: điều chỉnh quá trình phân giải protein trong suốt giai đoạn hạt tiềm sinh hoặc giai đoạn hạt nảy mầm [12]; góp phần bảo vệ thực vật bằng cách ngăn ngừa sự phân giải các protease ngoại sinh là các loại côn trùng gây hại, giun tròn... [10]. Cystatin phân lập từ *Arabidopsis thaliana* cho thấy trong điều kiện stress (muối, hạn, lạnh...) các chất ức chế đều biểu hiện nhiều hơn, đồng thời khuyến cáo có thể sử dụng cystatin để cải tạo khả năng chống stress ở thực vật [17]. Đặc tính này cũng phát hiện thấy trên cây ngô, cây rau dền (*Armanthus hypochondriacus*) [4, 12]. Gần đây một số tác giả đã công bố những công trình nghiên cứu về gen cystatin và bàn luận về mối liên quan của cystatin với khả năng chống chịu của thực vật như khả năng chịu hạn, chịu lạnh, chịu mặn, khả năng chống vi sinh vật gây bệnh [4, 5, 9, 10, 12, 17]... Cystatin của cây lạc được tách chiết từ mRNA là một đoạn nucleic acid dài 297 nucleotide mã hoá phân tử protein dài

98 amino acid (AY722693) [16]. Năm 2009, chúng tôi đã phân lập và xác định trình tự gen này từ hệ gen của giống lạc L18. Kết quả cho thấy gen cystatin của giống lạc L18 dài 461 nucleotide, trong đó có 2 exon và 1 intron; đoạn mã hoá dài 297 nucleotide và mã hoá sản phẩm dài 98 amino acid. Đoạn mã hoá của gen cystatin từ giống lạc L18 và đoạn mã hoá của gen cystatin với mã số AY722693 có mức tương đồng 100% [3, 15].

Giống lạc L18 được đánh giá là giống có khả năng chịu hạn kém và kém hơn so với giống lạc L23; Các dòng lạc có nguồn gốc từ giống L18 được tạo ra bằng kỹ thuật nuôi cấy *in vitro* kết hợp với kỹ thuật gây đột biến thực nghiệm liệu có làm thay đổi trong cấu trúc gen cystatin của những dòng này là lý do để chúng tôi tiếp tục nghiên cứu gen này. Gen cystatin được phân lập từ lá non của một số dòng lạc được tạo ra và chọn lọc bằng công nghệ nuôi cấy mô tế bào thực vật, trên cơ sở đó tìm hiểu đặc tính chịu hạn của cây lạc và làm cơ sở cho các nghiên cứu tiếp theo.

I. PHƯƠNG PHÁP NGHIÊN CỨU

1. Vật liệu

Giống lạc L18 (ĐC) có nguồn gốc từ Trung Quốc, năng suất đạt 55-70 tạ/ha, thời gian sinh trưởng vụ Xuân 120-130 ngày; có khả năng cho năng suất cao và kháng được bệnh lá nhưng khả năng chịu hạn kém do Viện Cây lương thực và cây thực phẩm Việt Nam cung cấp.

Dòng lạc R4.6 - có nguồn gốc từ mô sẹo chịu mất nước giống lạc L18 do thối khô 9 giờ liên tục.

Dòng M4.8 - có nguồn gốc từ mô sẹo của giống lạc L18 chịu ảnh hưởng của chiếu xạ liều 2 krad sau đó gây mất nước bằng thối khô 9 giờ liên tục.

Giống lạc L23 có nguồn gốc từ Trung Quốc, năng suất đạt 45-55 tạ/ha, thời gian sinh trưởng vụ Xuân 120-135 ngày, có nhiều đặc điểm nông học tốt, có khả năng kháng bệnh lá và khả năng chịu hạn khá do Viện Cây lương thực và cây thực phẩm Việt Nam cung cấp.

2. Phương pháp

Tách chiết DNA tổng số theo phương pháp của Gewel và cs., 1991 [6].

Nhân gel cystatin thực hiện bằng kỹ thuật PCR với cặp mồi được thiết kế dựa trên trình tự nucleotide của gen cystatin đã công bố (2004) tại ngân hàng gen quốc tế với mã số AY722693 [16]: Cys-AraF: 5'-ATGGCAGCAGTGGGTGC A-3'; Cys-AraR: 5'-TTAAGCATTGGAGCCAT CAC-3'.

Phản ứng được tiến hành trong hỗn hợp gồm: DNA khuôn (100 ng/ μ l) 1 μ l; mồi (10 pmol/ μ l) 2,0 μ l, dNTPs (2,5 mM) 2,0 μ l; Mg²⁺(25 nM) 2,5 μ l; Taq polymerase (5U/ μ l) 0,5 μ l; buffer PCR (10X) 2,5 μ l; Nước khử ion vừa đủ 25 μ l. Hỗn hợp được thực hiện với 30 chu kỳ phản ứng, các giai đoạn gồm: 94°C/3 phút; 94°C/30 giây; 56°C/1 phút; 72°C/1 phút; 72°C/10 phút và lưu giữ sản phẩm 4°C.

Tách dòng gen được tiến hành theo Sambrook và Russell (2001) [11].

Xác định trình tự DNA theo phương pháp Sanger thực hiện trên máy xác định trình tự DNA tự động ABI PRISM 3100 Avant Genetic Analyzer.

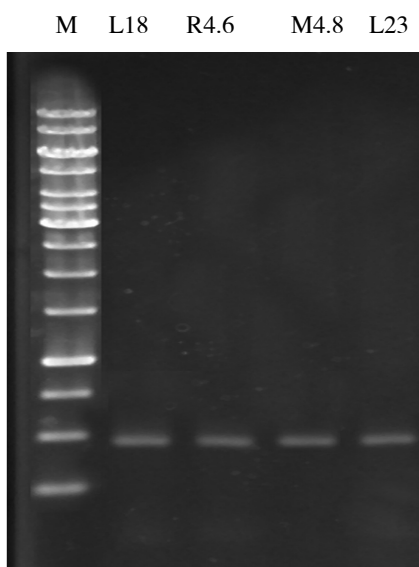
Kết quả đọc trình tự được xử lý trên máy vi tính bằng phần mềm BioEdit và DNASTAR.

II. KẾT QUẢ VÀ THẢO LUẬN

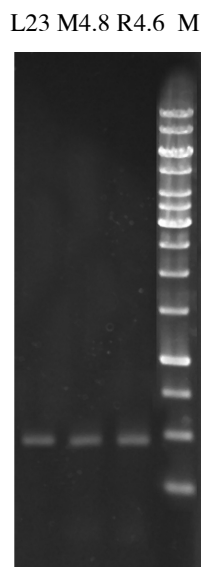
1. Kết quả nhân gen cystatin

Sau khi tách chiết DNA tổng số, tinh sạch và pha về nồng độ thích hợp, chúng tôi đã tiến hành nhân gen cystatin bằng phương pháp PCR với DNA của hai giống lạc (L18, L23) và 2 dòng chọn lọc từ giống L18. Kết quả nhân gen được kiểm tra trên gel agarose 1% cùng với marker và sản phẩm nhân gen cystatin của giống lạc L18 đã công bố (hình 1).

Sản phẩm của phản ứng PCR thu được ở hình 1 cho thấy, chỉ có 1 băng vạch, kích thước khoảng gần 500 bp tương đương với kích thước gen cystatin của giống lạc L18 mà chúng tôi đã công bố [3]. Do đó chúng tôi cho rằng, kết quả phản ứng PCR với mồi cystatin được nhân lên từ DNA các giống và dòng lạc có kích thước gần 500 bp là sự có mặt của gen cystatin trong hệ gen cây lạc.



Hình 1. Hình ảnh điện di sản phẩm nhân gen cystatin lạc



Hình 2. Hình ảnh điện di sản phẩm colony-PCR

2. Kết quả tách dòng và xác định trình tự gen cystatin

Chúng tôi tiến hành tách dòng sản phẩm thu được để nghiên cứu sâu hơn về gen cystatin. Mục này trình bày kết quả tách dòng gen được thực hiện từ sản phẩm nhân gen của 2 dòng chọn lọc là R4.6; M4.8 và giống L23.

Gen cystatin vừa nhân ở trên được làm sạch bằng bộ kit của Fermentas, sau đó được gắn vào vector pBT để tạo vector tái tổ hợp. Vector tái tổ hợp được biến nạp vào tế bào *E. coli* DH5 α bằng phương pháp sốc nhiệt. Nuôi khuẩn *E. coli* biến nạp trên môi trường LB có bổ sung thạch agar 1,6%, ampicilin 100 mg/ml, X-gal 40 mg/ml và IPTG 100 mM, ở 37°C qua đêm (16 giờ). Sau 16 giờ nuôi ổn định ở 37°C, tiến hành chọn khuẩn lạc có màu trắng, hình tròn đều trên môi trường cấy chuyển sang môi trường LB lỏng (10 g peptone + 5 g yeast extract + 10 g NaCl + nước cất) bổ sung ampicilin 100 mg/ml, lắc 200 vòng/phút ở 37°C qua đêm. Các lọ khuẩn được kiểm tra bằng phản ứng colony-PCR với

chính cặp mỗi nhân gen. Kết quả kiểm tra (hình 2) chứng tỏ việc ghép nối sản phẩm PCR vào vector tách dòng, chọn dòng và phản ứng colony-PCR.

Trình tự nucleotide của gen cystatin được xác định trên máy tự động ABI PRISM® 3100 Avant Genetic Analyzer. Sử dụng phần mềm DNASTar xử lý kết quả cho thấy, trình tự gen cystatin của giống lạc L23, dòng R4.6 và M4.8 có chiều dài 461 nucleotide. So sánh các trình tự này với trình tự nucleotide phân lập từ cDNA trên ngân hàng gen quốc tế mã số AY722693, chúng tôi nhận thấy đoạn gen cystatin mà chúng tôi phân lập gồm 2 exon và 1 intron; đoạn exon 1 gồm 102 nucleotide bắt đầu từ vị trí 1 đến 102 và đoạn exon 2 có 195 nucleotide, từ vị trí 267 đến 461; đoạn intron ở giữa có 164 nucleotide, từ vị trí 103 đến 266. Chúng tôi tiến hành so sánh trình tự nucleotide của gen cystatin phân lập được với một số trình tự nucleotide của gen cystatin cây lạc đã công bố trên Ngân hàng gen, kết quả thể hiện ở hình 3.

	10	20	30	
	-----+-----+-----+			
1	A T G G C A G C A G T G G G T G C A C C T C G C G A A G T A			L23.seq
1	A T G G C A G C A G T G G G T G C A C C T C G C G A A G T A			L18.seq
1	A T G G C A G C A G T G G G T G C A C C T C G C G A A G T A			R4.6.seq
1	A T G G C A G C A G T G G G T G C A C C T C G C G A A G T A			M4.8.seq
1	A T G G C A G C A G T G G G T G C A C C T C G C G A A G T A			AU 723567.seq
1	A T G G C A G C A G T G G G T G C A C C T C G C G A A G T A			AY722693.seq

```

                                     40          50          60
-----+-----+-----+
31  G C T G G A A A C G A G A A C A G C C T C G A G A T C G A T L23.seq
31  G C C G G A A A C G A G A A C A G C C T C G A G A T C G A T L18.seq
31  G C C G G A A A C G A G A A C A G C C T C G A G A T C G A T R4.6.seq
31  G C C G G A A A C G A G A A C A G C C T C G A G A T C G A T M4.8.seq
31  G C C G G A A A C G A G A A C A G C C T C G A G A T C G A T AU 723567.seq
31  G C C G G A A A C G A G A A C A G C C T C G A G A T C G A T AY722693.seq
                                     70          80          90
-----+-----+-----+
61  A G T C T T G C T C G C T T T G C T G T T G A T G A A C A C L23.seq
61  A G T C T T G C T C G C T T T G C T G T C G A T G A A C A C L18.seq
61  A G T C T T G C T C G C T T T G C T G T C G A T G A A C A C R4.6.seq
61  A G T C T T G C T C G C T T T G C T G T T G A T G A A C A C M4.8.seq
61  A G T C T T G C T C G C T T T G C T G T C G A T G A A C A C AU 723567.seq
61  A G T C T T G C T C G C T T T G C T G T C G A T G A A C A C AY722693.seq
                                     100         110         120
-----+-----+-----+
91  A A C A A G A A A C A G G G T T T C T T T C T C T C T C T T G L23.seq
91  A A C A A G A A A C A G G G T T T C T T T C T C T C T C T A C L18.seq
91  A A C A A G A A A C A G G G T T T C T T T C T C T C T C T A C R4.6.seq
91  A A C A A G A A A C A G G G T T T C T T T C T C T C T C T A G M4.8.seq
91  A A C A A G A A A C A G - - - - - - - - - - - - - - - - AU 723567.seq
91  A A C A A G A A A C A G - - - - - - - - - - - - - - - - AY722693.seq
                                     130         140         150
-----+-----+-----+
121 C G A A G A T C G C T T T C G C T T T G G T T T T G G T T C L23.seq
121 C G A T C A T C G C T T T G G C T C T G G T T T T G G T T C L18.seq
121 C G A T C A T C G C T T T G G C T C T G G T T T T G G T T C R4.6.seq
121 C G A A G A T C G C T T T G G C T C T G G T T T T G G T T C M4.8.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AU 723567.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AY722693.seq
                                     160         170         180
-----+-----+-----+
151 G T T T T G T G T G A T T G A T C A T G A T T A G T G T T C L23.seq
151 G T T T T G T A T G A T T G A T C A T G A T T A G T G T T C L18.seq
151 G T T T T G T A T G A T T G A T C A T G A T T A G T G T T C R4.6.seq
151 G T T T T G T A T G A T T G A T C A T G A T T A G T G T T C M4.8.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AU 723567.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AY722693.seq
                                     190         200         210
-----+-----+-----+
181 G A T T T T G A T T T A A T T T T A A A A T A A T T T G G G L23.seq
181 G A T T T T G A T T T A A T T T T A A A A T A A T T T G G G L18.seq
181 G A T T T T G A T T T A A T T T T A A A A T A A T T T G G G R4.6.seq
181 G A T T T T G A T T T A A T T T T A A A A T A A T T T G G G M4.8.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AU 723567.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AY722693.seq
                                     220         230         240
-----+-----+-----+
211 T G G T T G T T T T G G A T T T T A T C T A A T G T G A C L23.seq
211 T G G T T G T T T T G G A T T T T A T C T A A T G T G A C L18.seq
211 T G G T T G T T T T G G A T T T T A T C T A A T G T G A C R4.6.seq
211 T T G T T G T C T T G G A T T T T A T C T A A T G T A A C M4.8.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AU 723567.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AY722693.seq
                                     250         260         270
-----+-----+-----+
241 G A A T T A A A T T G G G A A T G A T C A T G C A G A A T G L23.seq
241 G A A T T A A A T T G G G A A T G A T G A T G C A G A A T G L18.seq
241 G A A T T A A A T T G G G A A T G A T G A T G C A G A A T G R4.6.seq
241 G A A T T A A A T T G G G A A T G A T C A T G C A G A A T G M4.8.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AU 723567.seq
103 - - - - - - - - - - - - - - - - - - - - - - - - - - - - AY722693.seq

```


Bảng 1

**Sự sai khác về nucleotide tại một số vị trí trên đoạn
gen cystatin của giống và dòng lạc nghiên cứu**

STT	Vị trí	L23	L18	R4.6	M4.8	AU723567	AY722693	Vùng sai khác
1	33	T	C	C	C	C	C	EXON
2	81	T	C	C	T	C	C	EXON
3	87	A	A	A	C	A	A	EXON
4	88	C	C	C	A	C	C	EXON
5	89	A	A	A	C	A	A	EXON
6	90	C	C	C	A	C	C	EXON
7	92	A	A	A	C	A	A	EXON
8	93	C	C	C	A	C	C	EXON
9	95	A	A	A	G	A	A	EXON
10	96	G	G	G	A	G	G	EXON
11	99	A	A	A	C	A	A	EXON
12	100	C	C	C	A	C	C	EXON
13	101	A	A	A	G	A	A	EXON
14	119	T	A	A	A	-	-	INTRON
15	120	G	C	C	C	-	-	INTRON
16	124	A	T	T	A	-	-	INTRON
17	125	G	C	C	G	-	-	INTRON
18	134	C	G	G	G	-	-	INTRON
19	138	T	C	C	C	-	-	INTRON
20	158	G	A	A	A	-	-	INTRON
21	212	G	G	G	T	-	-	INTRON
22	218	T	T	T	C	-	-	INTRON
23	238	G	G	G	A	-	-	INTRON
24	260	C	G	G	C	-	-	INTRON
25	271	C	G	G	C	G	G	EXON
26	278	T	C	C	C	C	C	EXON
27	326	C	T	T	T	T	T	EXON
28	389	T	G	G	G	G	G	EXON

Bảng 2

Độ tương đồng và độ sai khác đoạn mã hoá các trình tự gen cystatin

	Độ tương đồng (%)							
		L23	L18	R4.6	M4.8	AY722693	AU723567	
Độ sai khác (%)	L23		97,9	97,9	94,8	97,9	97,9	L23
	L18	2,1		100,0	95,5	100,0	100,0	L18
	R4.6	2,1	0,0		95,5	100,0	100,0	R4.6
	M4.8	5,2	4,5	4,5		95,5	95,5	M4.8
	AY722693	2,1	0,0	0,0	4,5		100,0	AY722693
	AU723567	2,1	0,0	0,0	4,5	0,0		AU723567
		L23	L18	R4.6	M4.8	AY722693	AU723567	

Xác định mức tương đồng và mức sai khác về trình tự nucleotide của đoạn mã hóa protein trên gen cystatin các dòng lạc nghiên cứu, kết

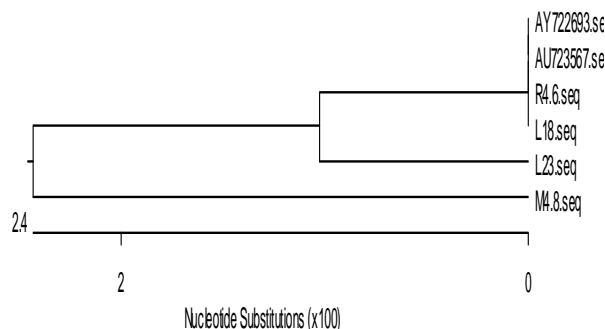
quả nhận được mức tương đồng từ 94,8% đến 100% và sự sai khác từ 0,0% đến 5,2%.

Biểu đồ hình cây ở hình 4 thể hiện mối quan

hệ di truyền của 6 dòng lạc trên cơ sở phân tích trình tự đoạn mã hóa gen cystatin, kết quả trên sơ đồ cho thấy, 6 dòng lạc được phân thành 2 nhóm chính: nhóm I phân thành 2 nhóm phụ: nhóm phụ 1 gồm giống lạc L18, dòng R4.6 (có nguồn gốc tái sinh từ mô sẹo chịu mất nước của giống L18) AY722693, AU723567 và nhóm phụ 2 là giống L23; nhóm II là dòng M4.8 là dòng tái sinh từ mô sẹo chịu ảnh hưởng của chiếu xạ và thối khô của giống L18. Như vậy, trong 6 trình tự nucleotide so sánh trên phương diện cấu trúc đoạn mã hoá gen cystatin thì dòng R4.6 có độ tương đồng 100% so với các trình tự đã công bố trên Ngân hàng gen quốc tế; giống L23 có độ sai khác 4,5% với các trình tự đã công bố và khác với dòng M4.8 là 5,2%.

So sánh trình tự amino acid: trình tự nucleotide đoạn mã hoá gen cystatin của giống lạc L23 và các dòng lạc nghiên cứu mà chúng tôi phân lập được dài 297 nucleotide mã hoá phân tử protein gồm 98 amino acid, có chứa 2

vùng bảo thủ của nhóm là L²²A²³R²⁴F²⁵A²⁶V²⁷ và Q⁴⁹V⁵⁰V⁵¹A⁵²G⁵³, vùng intron nằm xen kẽ giữa 2 amino acid vị trí số 34 và 35; tức là giữa 2 đoạn bảo thủ LARFAV và QxVxG, tương ứng với nhóm phân loại 1 và 2 của Marcia và cs. (2008) và Martinez và cs. (2007) [7, 8].



Hình 4. Sơ đồ so sánh mối tương đồng nucleotide đoạn mã hóa gen cystatin của các mẫu nghiên cứu với 2 trình tự công bố trên Ngân hàng gen quốc tế (NCBI)

		10		20		30																									
1	M	A	A	V	G	A	P	R	E	V	A	G	N	E	N	S	L	E	I	D	S	L	A	R	F	A	V	D	E	H	L23.pro
1	M	A	A	V	G	A	P	R	E	V	A	G	N	E	N	S	L	E	I	D	S	L	A	R	F	A	V	D	E	H	L18.pro
1	M	A	A	V	G	A	P	R	E	V	A	G	N	E	N	S	L	E	I	D	S	L	A	R	F	A	V	D	E	H	R4.6.pro
1	M	A	A	V	G	A	P	R	E	V	A	G	N	E	N	S	L	E	I	D	S	L	A	R	F	A	V	D	D	T	M4.8.pro
1	M	A	A	V	G	A	P	R	E	V	A	G	N	E	N	S	L	E	I	D	S	L	A	R	F	A	V	D	E	H	AY722693.pro
1	M	A	A	V	G	A	P	R	E	V	A	G	N	E	N	S	L	E	I	D	S	L	A	R	F	A	V	D	E	H	AU723567.pro
					40																										
31	N	K	K	Q	N	A	L	L	E	F	K	R	V	I	S	A	K	Q	Q	V	V	A	G	T	L	H	H	I	T	L	L23.pro
31	N	K	K	Q	N	G	L	L	E	F	K	R	V	I	S	A	K	Q	Q	V	V	A	G	T	L	H	H	I	T	L	L18.pro
31	N	K	K	Q	N	G	L	L	E	F	K	R	V	I	S	A	K	Q	Q	V	V	A	G	T	L	H	H	I	T	L	R4.6.pro
31	T	R	N	R	N	A	L	L	E	F	K	R	V	I	S	A	K	Q	Q	V	V	A	G	T	L	H	H	I	T	L	M4.8.pro
31	N	K	K	Q	N	G	L	L	E	F	K	R	V	I	S	A	K	Q	Q	V	V	A	G	T	L	H	H	I	T	L	AY722693.pro
31	N	K	K	Q	N	G	L	L	E	F	K	R	V	I	S	A	K	Q	Q	V	V	A	G	T	L	H	H	I	T	L	AU723567.pro
					70																										
61	E	A	A	S	G	D	S	K	N	V	Y	E	A	K	V	W	E	K	P	W	M	N	F	K	E	V	Q	E	F	K	L23.pro
61	E	A	A	S	G	D	S	K	N	V	Y	E	A	K	V	W	E	K	P	W	M	N	F	K	E	V	Q	E	F	K	L18.pro
61	E	A	A	S	G	D	S	K	N	V	Y	E	A	K	V	W	E	K	P	W	M	N	F	K	E	V	Q	E	F	K	R4.6.pro
61	E	A	A	S	G	D	S	K	N	V	Y	E	A	K	V	W	E	K	P	W	M	N	F	K	E	V	Q	E	F	K	M4.8.pro
61	E	A	A	S	G	D	S	K	N	V	Y	E	A	K	V	W	E	K	P	W	M	N	F	K	E	V	Q	E	F	K	AY722693.pro
61	E	A	A	S	G	D	S	K	N	V	Y	E	A	K	V	W	E	K	P	W	M	N	F	K	E	V	Q	E	F	K	AU723567.pro
91	L	A	G	D	G	S	N	A	.																					L23.pro	
91	L	A	G	D	G	S	N	A	.																					L18.pro	
91	L	A	G	D	G	S	N	A	.																					R4.6.pro	
91	L	A	G	D	G	S	N	A	.																					M4.8.pro	
91	L	A	G	D	G	S	N	A	.																					AY722693.pro	
91	L	A	G	D	G	S	N	A	.																					AU723567.pro	

Hình 5. So sánh trình tự aminoacid của các của các mẫu nghiên cứu với 2 trình tự công bố trên Ngân hàng gen

**Sự sai khác về amino acid của các mẫu nghiên cứu
với 2 trình tự công bố trên Ngân hàng gen**

Vị trí	L23	L18	R4.6	M4.8	AU723567	AY722693
29	E	E	E	D	E	E
30	H	H	H	T	H	H
31	N	N	N	T	N	N
32	K	K	K	R	K	K
33	K	K	K	N	K	K
34	Q	Q	Q	P	Q	Q
36	A	G	A	A	A	A

So sánh trình tự amino acid của giống lạc L23 và 2 dòng R4.6; M4.8 với các trình tự đã công bố có 7 vị trí sai khác trên chuỗi amino acid tại vị trí số 29, 30, 31, 32, 33, 34 và 36 của các dòng và giống nghiên cứu, kết quả cụ thể trình bày ở bảng 3.

Nghiên cứu sự thay đổi các amino acid trên phân tử protein của giống L23, giống L18 và 2 dòng chọn lọc, chúng tôi thấy, 6 vị trí E²⁹, H³⁰, N³¹, K³², K³³ và Q³⁴ của giống L23, L18 và R4.6 là giống nhau và giống với trình tự đã công bố trên NCBI, đoạn trình tự này bị thay thế ở dòng M4.8 là D²⁹, T³⁰, T³¹, R³², N³³, P³⁴. Vị trí amino acid số 36 là G của giống L18 thay thế vị trí là A của giống L23, dòng R4.6, dòng M4.8 và 2 trình tự trên NCBI. Tuy nhiên, khả năng chịu hạn của thực vật là tính trạng đa gen và bản thân gen cystatin là gen liên quan đến nhiều khả năng chống chịu khác của thực vật, kết quả nghiên cứu là căn cứ để nghiên cứu sâu hơn về gen cystatin ở lạc.

III. KẾT LUẬN

Đã phân lập và xác định được trình tự gen cystatin trên giống lạc L23 và các dòng R4.6; M4.8 có nguồn gốc từ mô sẹo chịu chiếu xạ và thối khô của giống lạc L18.

Gen cystatin trên lạc có chiều dài 461 nucleotide, chứa một intron ở giữa 2 vùng bảo thủ là L²²A²³R²⁴F²⁵A²⁶V²⁷ và Q⁴⁹V⁵⁰V⁵¹A⁵²G⁵³.

Đã phát hiện 28 vị trí nucleotide sai khác trên gen cystatin và 7 amino acid trên phân tử protein của giống L23, giống L18, dòng R4.6 và dòng M4.8 với một số trình tự đã công bố trên lạc; đoạn mã hoá phân tử protein phân lập được mức độ sai khác nhau từ 2,1% đến 5,2%.

Lời cảm ơn: Công trình được thực hiện với sự hỗ trợ của đề tài Khoa học - Công nghệ cấp Bộ B2009-TN04-04 và giúp đỡ của Phòng thí nghiệm Công nghệ tế bào thực vật, Viện Công nghệ Sinh học.

TÀI LIỆU THAM KHẢO

1. **Vũ Công Hậu, Ngô Thế Dân, Trần Thị Dung**, 1994: Cây lạc. Nxb. Nông nghiệp, Hà Nội: 11-53.
2. **Chu Thị Thơm, Phan Thị Lại, Nguyễn Văn Tó**, 2006: Kỹ thuật trồng và chăm sóc cây lạc. Nxb. Lao động Xã hội, 2-86.
3. **Vũ Thị Thu Thủy, Nguyễn Vũ Thanh Thanh, Chu Hoàng Mậu, Nguyễn Thị Tâm**, 2009: Phân tích trình tự gen Cystatin của giống lạc L18 (*Arachis hypogaea* L.): 397-400. Báo cáo hội nghị Sinh học toàn quốc 2009, Nxb. Đại học Thái Nguyên.
4. **Massonneau A., Condamine P., Wisniewski J. P., Zivy M., Rogowsky P. M.**, 2005: Maize cystatins respond to developmental cues, cold stress and drought. *Biochimica et Biophysica Acta*, 1729: 186-199.
5. **Diop N. N., Kidric., Repellin A., Gareil M., Lameta A., Pham A. T. T., Zuil-Fodil Y.**, 2004: A multicystatin is induced by drought-stress in cowpea [*Vigna unguiculata* (L.) Walp.] leaves. *FEBS Lett.*, 577(3): 545-550.
6. **Gawel N. J., Jarrnet R. T.**, 1991: Genomic DNA isolation. [Http://www.weiheinstephan.de/pbpz/bambara/htm/dna.htm](http://www.weiheinstephan.de/pbpz/bambara/htm/dna.htm).

7. **Marcia M. P., Andreia C. T. Z., Guilherme L., Giancarlo P., Rogerio M.**, 2008: Molecular evolution and diversification of plant cysteine proteinase inhibitors: New insights after the poplar genome. *Molecular Phylogenetics and Evolution*, 49: 355-349.
8. **Martinez M., Diaz-Mendoza M., Carrillo L., Diaz I.**, 2007: Cacoxy terminal extended phytocystatins are bifunctional inhibitors of papain and legumain cystein proteinase. *FEBS Lett.*, 581: 2914-2918.
9. **Massonneau A., Condamine P., Wisniewski J. P., Zivy M., Rogowsky P. M.**, 2005: Maize cystatins respond to developmental cues, cold stress and drought. *Biochim Biophys Acta.*, 1729(3): 186-199.
10. **Paulraj K. Lawrence, Kripa Ram Koundal**, 2002: Plant protease inhibitors in control of phytophagous insects *Electron. J. Biotechnol*, 5(1).
11. **Sambrook J., Russell D. W.**, 2001: *Molecular cloning a laboratory manual* 3rd edition, Cold Spring Haror Laboratory Press.
12. **Silvia V. R., Armando G. R., Claudia M. V., Alicia C. L., Francisco D. V., Norma M. G., Carla S. H., John D. F.**, 2007: Cloning of a cDNA encoding a cystatin from grain amaranth (*Amaranthus hypochondriacus*) showing a tissue-specific expression that is modified by germination and abiotic stress. *Plant Physiology and Biochemistry*, 45: 790-798.
13. **Srinivasan T., Raja Rajesh Kumar K. and Kirti P. B.**, 2008: *Arachis diogoi* cystatin mRNA, complete cds., EMBL GenBank, Accession EU723567.
14. **Turk V., Bode W.**, 1991: The cystatins: protein inhibitors of cysteine proteinases. *FEBS Lett.*, 285: 213-219.
15. **Vu T. T. T., Nguyen T. V. T., Chu M. H. and Nguyen T. T.**, 2010: Accession FN811133, *Arachis hypogaea* cystatin gene 1, exons 1-2. Genbank.
16. **Yan Y., Wang. L., Huang S.**, 2004: *Arachis hypogaea* cystein proteinase inhibitor mRNA, complete cds, EMBL GenBank, Accession AY722693.
17. **Zhang X., Liu S., Takano T.**, 2008: Two cysteine proteinase inhibitors from *Arabidopsis thaliana*, AtCYSa and AtCYSb, increasing the salt, drought, oxidation and cold tolerance. *Plant Mol. Biol.*, 68(1-2): 131-143.

CHARACTERISTICS OF NUCLEOTIDE SEQUENCE OF CYSTATIN GENE OF SOME PEANUT LINES DERIVED FROM IRRADIATED AND DEHYDRATED

VU THI THU THUY, NGUYEN THI TAM,
CHU HOANG MAU, NGUYEN VU THANH THANH

SUMMARY

Cystatin are super family of cystein protease inhibitor protein which is found in plants, animals and micro-organism. They are divided into three inhibitory families based on the difference of amino acid sequence, disunity bond and molecular weight including: the first is stefines, the second is cystatins and the third is kininogens. In plant cystatin is name phytocystatin (physic), it is also an independent brand of cystatin families on phylogenetic. Physic has very importance physiological functions such as regulation of activity of endogenous cystein protease during seed development, germination and programmed cell death. Recently, there are many researches on relationship between cystatin and resistance in plant (drought tolerance, cold tolerance, salt tolerance and ability of against pathogenic microorganisms). Analyzing and comparing cytatin gene sequences not only can provide valuable knowledge for the conservation and evaluation of the evolution of cytatin genes, but also shed light on issues related to function of cytatin protein.

Isolation and sequencing cystatin gene of the L23 peanut cultivar and two lines: R4.6, M4.8 derived from callus treated with irradiation and dehydration of L18 peanut cultivar. The results showed that cystatin genes had 461 nucleotides and contained one intron between two conservative regions L²²A²³R²⁴F²⁵A²⁶V²⁷ and Q⁴⁹V⁵⁰V⁵¹A⁵²G⁵³. There are 28 different nucleotide positions on cystatin genes and seven amino acids on protein molecules of L23 peanut varieties and the M4.8, R4.6 lines in comparison with published sequences of peanut cystatin gene. The gene sequence encoded protein of peanut lines in this research have different levels from 2.1% to 5.2% and could be used for further research on the relationship between cystatin gene structure related to drought tolerance of peanut.

Ngày nhận bài: 15-11-2010