

ĐẶC ĐIỂM VÀ SỰ BIỂU HIỆN CỦA HAI HỌ PROTEIN LATE EMBRYOGENESIS ABUNDANT (LEA) Ở ĐẬU TƯƠNG (*Glycine max*)

Cao Phi Bằng

Trường Đại học Hùng Vương, Phú Thọ, phibang.cao@hvu.edu.vn

TÓM TẮT: Các protein Late Embryogenesis Abundant (LEA) có mặt với số lượng lớn trong phôi và giữ vai trò quan trọng đối với sự phát triển phôi. Nhiều họ LEA được cho là giữ vai trò quan trọng đối với tính chịu một số stress vô sinh ở thực vật. LEA5 và protein hạt chín (Seed Maturation Protein, SMP) là hai họ của siêu họ LEA theo cách phân loại hiện hành. Trong nghiên cứu này, chúng tôi xác định được năm gen mã hóa cho mỗi họ LEA5 và SMP trong hệ gen của cây đậu tương (*Glycine max*). Đối với các LEA5 của cây đậu tương, nhìn chung mỗi gen có một intron, ngoại trừ *GmLEA5-3*, gen này không được chú giải trong hệ gen và không có intron. Các protein suy diễn có kích thước nhỏ, có tính axit yếu, ưa nước mạnh. Các gen SMP có một hoặc nhiều intron. Protein suy diễn có tính axit yếu, ái lực với nước yếu. Tất cả các gen LEA5 và SMP không biểu hiện ở các mô sinh dưỡng của cây đậu tương nhưng biểu hiện đặc hiệu ở hạt. Cường độ biểu hiện tăng dần theo sự phát triển của hạt, mạnh nhất ở thời điểm 42 ngày sau thụ phấn. Sự biểu hiện đặc hiệu mô của các gen này gợi ý về vai trò quan trọng của chúng đối với sự phát triển của hạt đậu tương. Lần đầu tiên, các đặc điểm và sự biểu hiện của các LEA5 và các SMP của cây đậu tương được báo cáo trong công trình này, mở đường cho các nghiên cứu chức năng của các gen của hai họ này ở các loài cây họ Đậu.

Từ khóa: Biểu hiện gen, cây di truyền, đặc điểm của gen, đậu tương, LEA5, SMP.

MỞ ĐẦU

Các điều kiện cực hạn có tác động lên sinh trưởng, phát triển và sản lượng của thực vật. Đồng thời, thực vật cũng phát triển nhiều đáp ứng thích nghi với điều kiện môi trường [4]. Các stress vô sinh như hạn, mặn, áp suất thẩm thấu, lạnh và băng giá dẫn tới sự tích lũy một nhóm các protein ưa nước gọi là LEA (LEA = Late Embryogenesis Abundant) [15]. Sự tích lũy các protein này trong phôi của hạt ngủ và hạt phấn chín, cả hai cấu trúc chịu khô có khả năng chống lại sự mất nước trong những thời gian dài, dẫn tới giả thuyết về vai trò của các protein này đối với khả năng chịu mất nước của thực vật [15]. LEA lần đầu được phát hiện ở hạt cây bông (*Gossypium hirsutum*), tích lũy ở pha phát sinh phôi muộn [9]. Phù hợp với sự có mặt của các motif hoặc thành phần amino acid khác nhau, các LEA của thực vật được phân chia vào nhiều nhóm khác nhau, tương ứng với các họ LEA1, LEA2, LEA3, LEA4, LEA5, LEA6, SMP và dehydrin (Pfam) [15]. Trong các họ LEA ở thực vật, những hiểu biết về cấu trúc và chức năng của họ dehydrin được nghiên cứu nhiều nhất [13]. Ngược lại những nghiên cứu về

các họ LEA5 và SMP ở thực vật còn rất ít.

Các protein họ LEA5 điển hình hiện biết ở thực vật là AtEM1 và AtEM6 của cây *Arabidopsis* [17], hai protein Lea B19 của cây lúa mạch [14, 21]. Hai gen *AtEM1* và *AtEM6* biểu hiện khác nhau về không gian và thời gian trong quá trình chín của phôi. *AtEM1* biểu hiện mạnh nhất ở đỉnh rễ trong khi *AtEM6* biểu hiện ở phôi, mô phân sinh đỉnh và các mô tiền mạch [23; 24]. *Lea B19* biểu hiện ở hạt [21].

SMP điển hình được biết là Rab28 của cây ngô [18] và Rab28 của cây *Arabidopsis* [2]. Ở *Arabidopsis*, *AtRab28* biểu hiện trong quá trình phát sinh phôi nhưng không bị cảm ứng bởi abscisic acid hoặc điều kiện mất nước [2]. Trong khi đó, gen *ZmRab28* của cây ngô cũng biểu hiện trong quá trình phát sinh phôi, đồng thời bị cảm ứng bởi abscisic acid [18]. Hơn nữa, cây ngô chuyển gen *Rab18* có khả năng chịu mất nước cao hơn so với đối chứng [1].

Ở các cây họ đậu, đã có nhiều nghiên cứu được tiến hành đối với các gen thuộc một số họ LEA khác nhau (LEA1, LEA2, LEA3, LEA4, LEA6, dehydrin). Đặc điểm cấu trúc, sự phân loại và sự biểu hiện của các gen thuộc các họ

LEA khác nhau đã được nghiên cứu ở nhiều cây họ đậu như đậu tương, đậu cove, cây *Medicago truncatula*, cây đậu triều (*Cajanus cajan*), đặc biệt khi hệ gen của các loài này đã được giải trình tự [5]. Tuy nhiên, khác với các họ LEA khác, họ LEA5 và SMP chưa hề được nghiên cứu ở cây họ đậu, ngoại trừ gen *GmPM11* là một gen thuộc họ LEA5 đã được đặc trưng hóa [21].

Trong công trình này, chúng tôi trình bày các kết quả nghiên cứu về hai họ gen LEA có liên quan đến sự phát triển của hạt ở cây đậu tương. Trong đó, số lượng gen, các đặc trưng của gen, kết quả phân tích cây di truyền và sự biểu hiện của các gen trong mỗi họ sẽ được giới thiệu. Các kết quả nghiên cứu thu được có ý nghĩa lớn, bổ sung những hiểu biết còn thiếu về hai họ LEA này ở cây họ đậu cũng như ở thực vật nói chung.

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Cơ sở dữ liệu về hệ gen và RNAseq của cây đậu tương

Chúng tôi sử dụng hệ gen của cây đậu tương đã giải trình tự bởi Schmutz et al. (2010) [19]. Tập hợp dữ liệu RNAseq được lấy từ Severin et al. (2010) [20].

Xác định các gen thuộc họ LEA5 và SMP ở cây đậu tương

Chúng tôi sử dụng các protein LEA5 và SMP của cây *Arabidopsis* [15] làm khuôn dò để tìm kiếm các gen tương đồng trong hệ gen của cây đậu tương nhờ chương trình TBLASTN theo phương pháp nghiên cứu đã được sử dụng bởi Cao Phi Bằng và Trần Thị Thanh Huyền (2015) [7]. Sau đó chúng tôi dùng Pfam (cơ sở dữ liệu các họ protein [10] để xác nhận sự hiện diện của các vùng bảo thủ trong trình tự protein.

Phân tích tin sinh học về các họ gen được nghiên cứu

Các đặc điểm vật lí, hóa học của các gen/protein nghiên cứu được phân tích nhờ các công cụ của ExPASy [11]. Cấu trúc exon/intron được xây dựng nhờ GSDS 2.0 [12]. Các phương pháp nghiên cứu này được miêu tả trong công trình được thực hiện bởi Cao Phi Bằng và Trần Thị Thanh Huyền (2015) [7]. Các motif bảo thủ được phát hiện nhờ MEME (Multiple EM for

Motif Elicitation) [3].

Nghiên cứu sự biểu hiện gen

Sự biểu hiện của các gen được phân tích qua kết quả RNAseq (giải trình tự tập hợp ARN thông tin) của cây đậu tương [20].

KẾT QUẢ VÀ THẢO LUẬN

Xác định họ gen LEA5 và SMP ở cây đậu tương

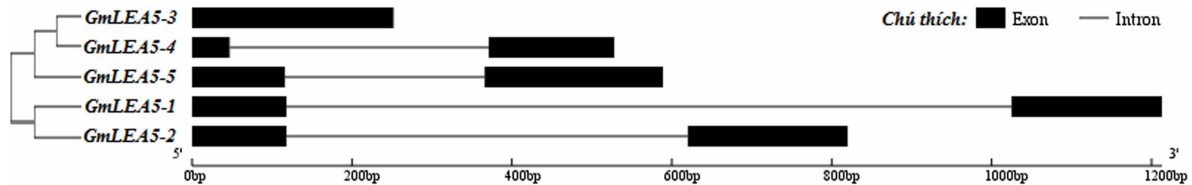
Nhờ sử dụng chương trình TBLASTN, với các khuôn dò là các SMP và LEA5 của cây *A. thaliana*, chúng tôi đã xác định được 5 gen mã hóa cho các LEA5 và 5 gen SMP trong hệ gen của cây đậu tương. Phân tích pfam để kiểm tra các trình tự protein suy diễn, các LEA5 của cây đậu tương đều có đoạn trình tự amino acid bảo thủ của họ LEA_5 (PF00477), đồng thời các SMP của cây đậu tương cũng có ít nhất hai vùng bảo thủ SMP (PF04927). Khi so sánh với các loài thực vật có hệ gen đã giải trình tự khác, số lượng gen LEA5 của cây đậu tương chỉ ít hơn ở loài dương xỉ *Selaginella moellendorffii*, bằng với cây bạch đàn (*Eucalyptus grandis*) và lớn hơn ở các loài khác. Trong khi đó, số lượng gen SMP của cây đậu tương ít hơn của cây dương xỉ, *A. thaliana* và bạch đàn, bằng với của cây lúa và nhiều hơn ở một số loài (bảng 1). Sự sai khác về số lượng gen mã hóa LEA5 và SMP giữa các loài gợi ý rằng số lượng gen LEA5 hoặc SMP trong hệ gen của mỗi loài cây có thể không liên quan đến sự tiến hóa ở thực vật.

Đặc điểm họ gen LEA5 ở cây đậu tương

Các gen LEA5 của cây đậu tương có kích thước nhỏ, chỉ dài từ 252 tới 1213 nucleotide, có chứa một intron hoặc mã hóa liên tục (hình 1). Đặc điểm này tương đồng với các LEA5 của cây bạch đàn, loài cây có cả 5 gen mã hóa LEA5 không liên tục với một intron [16]. Các gen này quy định các phân tử protein nhỏ, từ 77 tới 112 amino acid, tương ứng với khối lượng từ 8,83 kDa tới 12,25 kDa. Các protein LEA5 của cây đậu tương có tính axit hoặc axit yếu (pI dao động trong khoảng 5,22 tới 6,59). Các LEA5 này rất ưa nước, giá trị GRAVY dao động từ -1,617 tới -1,414 (bảng 2). Các đặc điểm của LEA5 của cây đậu tương tương đồng với nhiều LEA5 của các loài khác như cây mơ [8], nhưng hai trong ba LEA5 của cây mơ có tính kiềm mạnh.

Để tìm các motif bảo thủ, đặc trưng cho họ LEA5 của thực vật, chúng tôi sử dụng công cụ MEME để tìm kiếm từ tập hợp 33 protein LEA5 của các loài cây được giới thiệu trong bảng 1. Kết quả chúng tôi tìm thấy 5 motif bảo thủ, đặc trưng cho các LEA5 (hình 2). Trong đó, motif 2 tương

đồng với motif bảo thủ đặc trưng cho các LEA5 đã được báo cáo bởi Hundertmark và Hincha [15]. Các motif này có sự xuất hiện của một số amino acid như Glycine (G) và Glutamic acid (E). Riêng motif 4, Glutamic acid ít xuất hiện trong khi các Arginine (R) xuất hiện nhiều hơn.



Hình 1. Cấu trúc exon/intron của các LEA5 của cây đậu tương (CDS: trình tự mã hóa protein, intron: trình tự không mã hóa protein)

Bảng 1. Số lượng gen LEA5 và SMP ở một số thực vật

Loài	Số lượng gen họ LEA5	Số lượng gen họ SMP
Đậu tương (<i>G. max</i>)	5	5
Rêu (<i>P. patens</i>)	3	3
Dương xỉ (<i>S. moellendorffii</i>)	6	8
Thông (<i>P. taeda</i>)	3	3
<i>A. thaliana</i>	2	6
Khoai tây (<i>S. tuberosum</i>)	2	2
Bạch đàn (<i>E. grandis</i>)	5	6
Dương (<i>P. trichocarpa</i>)	1	2
Mơ (<i>P. mume</i>)	4	3
Lúa (<i>O. sativa</i>)	2	5

Đặc điểm họ gen SMP ở cây đậu tương

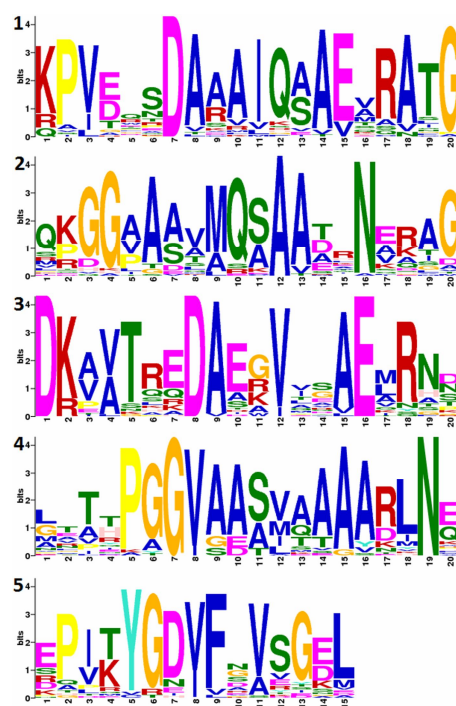
Năm gen mã hóa SMP của cây đậu tương có kích thước tương đối khác nhau, gen ngắn nhất chỉ có 979 nucleotide trong khi gen dài nhất có kích thước 2374 nucleotide. Ba trong năm gen SMP có một intron, hai gen còn lại lần lượt có hai (*GmSMP1*) và ba (*GmSMP2*) intron (hình 3). Như vậy số lượng intron của các gen không giống nhau, đặc điểm này khác với các SMP của cây bạch đàn, cả sáu SMP của cây bạch đàn đều có hai intron [16]. Trình tự protein suy diễn SMP của cây đậu tương có từ 176 amino acid đến 284 amino acid. Cả năm SMP của cây đậu tương đều có tính axit hoặc axit yếu (pI nằm trong khoảng 4,27 đến 6,48). Các protein này có ái lực với nước không cao, giá trị GRAVY chỉ nằm trong khoảng -0,536 tới -0,182 (bảng 3).

Tính axit và tính ưa nước yếu của các SMP của cây đậu tương giống với các SMP của cây mơ [8] và của cây *A. thaliana* [6].

Các motif bảo thủ của các SMP cũng được phát hiện nhờ MEME từ tập hợp 43 protein SMP của các loài thực vật từ rêu, dương xỉ, thông tới các cây một lá mầm và hai lá mầm (hình 4). Glycine vẫn có mặt ở cả 5 motif bảo thủ, nhưng với tỷ lệ không lớn. Các amino acid có mặt nhiều nhất trong các motif bảo thủ của SMP là các amino acid có chuỗi bên kỵ nước như các Alanine (A), Valine (V), Leucine (L)... Đặc điểm này góp phần giải thích tại sao các SMP của cây đậu tương lại có ái lực với nước yếu. Ngoài ra, các Aspartic acid và Glutamic acid cũng có mặt ở cả 5 motif bảo thủ của các SMP.



Hình 2. Các motif bảo thủ được tìm thấy trong các trình tự protein LEA5 được phát hiện nhờ chương trình MEME [3].



Hình 4. Các motif bảo thủ được tìm thấy trong các trình tự protein SMP được phát hiện nhờ chương trình MEME [3].



Hình 3. Cấu trúc exon/intron của các SMP của cây đậu tương (CDS: trình tự mã hóa protein, intron: trình tự không mã hóa protein)

Bảng 2. Các gen LEA5 ở cây đậu tương

Gen	Tên Locus	Kích thước gen/CDS (bp)	Chiều dài protein (aa)	Khối lượng phân tử protein (kDa)	pI	GRAVY	Số lượng intron
<i>GmLEA5-1</i>	Glyma01g29480	1213/306	101	11,14	6,31	-1,568	1
<i>GmLEA5-2</i>	Glyma03g07470	820/318	105	11,51	5,53	-1,617	1
<i>GmLEA5-3</i>	-	252/252	83	9,37	6,59	-1,418	0
<i>GmLEA5-4</i>	Glyma16g19405	528/204	77	8,83	5,22	-1,497	1
<i>GmLEA5-5</i>	Glyma18g43320	589/339	112	12,25	5,33	-1,414	1

Bảng 3. Các gen SMP ở cây đậu tương

Gen	Tên Locus	Kích thước gen/CDS (bp)	Chiều dài protein (aa)	Khối lượng phân tử protein (kDa)	pI	GRAVY	Số lượng intron
<i>GmSMP1</i>	Glyma10g03310	1201/789	262	27,46	5,16	-0,455	2
<i>GmSMP2</i>	Glyma10g30090	1442/855	284	29,45	6,48	-0,536	3
<i>GmSMP3</i>	Glyma10g39270	979/771	256	26,24	4,75	-0,322	1
<i>GmSMP4</i>	Glyma11g16090	2374/540	179	18,24	4,27	-0,182	1
<i>GmSMP5</i>	Glyma20g28550	1025/771	256	26,06	4,90	-0,238	1

Bảng 4. Sự biểu hiện của các gen *LEA5* và *SMP* ở cây đậu tương (đơn vị tính RPKM, số bản mã phiên (ARN)/1000 nucleotide của mã phiên/một triệu bản mã phiên)

Gen	Mô														Chú thích
	Lá non	Rễ	Nốt sần	Hoa	Quả nhỏ (7 DAF)	Quả trung bình (10DAF)	Quả lớn (14DAF)	Hạt (10DAF)	Hạt (14DAF)	Hạt (21DAF)	Hạt (25DAF)	Hạt (28DAF)	Hạt (35DAF)	Hạt (42DAF)	
<i>GmLEA5-1</i>	0	0	0	0	0	0	0	0	0	0	0	0	1	4	120
<i>GmLEA5-2</i>	0	0	0	0	0	0	0	0	0	0	0	0	10	54	100
<i>GmLEA5-3</i>	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	80
<i>GmLEA5-4</i>	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	60
<i>GmLEA5-5</i>	0	0	0	0	0	0	0	1	3	12	117	92	323	353	40
<i>GmSMP1</i>	0	0	0	1	2	1	1	0	1	1	5	5	11	26	20
<i>GmSMP2</i>	0	0	0	0	0	0	0	0	1	0	1	0	2	2	0
<i>GmSMP3</i>	0	0	0	0	0	0	0	0	0	1	3	1	8	18	<i>DAF</i>
<i>GmSMP4</i>	0	0	0	0	0	0	0	2	13	19	54	44	91	94	= Số
<i>GmSMP5</i>	0	0	0	0	0	1	2	0	0	1	12	7	26	82	ngày
															sau
															thụ
															phần

Sự biểu hiện đặc hiệu theo mô của các *LEA5* và *SMP* của cây đậu tương

Sự biểu hiện của các *LEA5* và *SMP* của cây đậu tương được nghiên cứu ở cả các mô sinh dưỡng (lá, rễ, nốt sần) và các mô sinh sản (hoa, quả và hạt) thông qua phân tích dữ liệu RNAseq của các mô trên [20] (bảng 4). Chỉ hai gen *GmLEA5-3* và *GmLEA5-4* không biểu hiện ở bất cứ mô nào. Các gen *LEA5* còn lại và tất cả các *SMP* đều biểu hiện ở hạt, đặc biệt trong giai đoạn phát triển muộn. Cường độ biểu hiện của

các gen này cao nhất ở trong hạt thời điểm 42 ngày sau thụ phấn. Ở giai đoạn phát triển sớm của hạt (10 ngày sau thụ phấn), chỉ có hai gen *GmLEA5-5* và *GmSMP3* biểu hiện, nhưng với cường độ yếu. Hai gen này là hai gen biểu hiện ở tất cả các pha phát triển của hạt. Đến giai đoạn 14 ngày sau thụ phấn, có thêm hai gen *GmSMP1* và *GmSMP2* biểu hiện. Gen *GmSMP5* chỉ bắt đầu biểu hiện ở hạt thời điểm 21 ngày sau thụ phấn. Nhìn chung, các gen này biểu hiện với cường độ tăng dần theo sự phát

triển của hạt. Thời điểm 25 ngày sau thụ phấn có thể là thời điểm quan trọng trong sự phát triển của hạt đậu tương khi sự biểu hiện của các gen *LEA5* và *SMP* (chỉ xét các gen có biểu hiện) tạo thành một đỉnh tạm thời, cao hơn cả thời điểm 28 ngày sau thụ phấn. Họ *LEA5* không biểu hiện ở mô sinh sản khác, chỉ hai gen thuộc họ *SMP* biểu hiện ở mô quả (*GmSMP1* và *GmSMP5*), chỉ duy nhất *GmSMP5* biểu hiện ở hoa. Tất cả các *LEA5* và *SMP* của cây đậu tương không biểu hiện ở các mô sinh dưỡng. Những kết quả này cho phép đặt giả thiết về sự biểu hiện đặc hiệu ở mô hạt của các *LEA5* và *SMP* của cây đậu tương, đồng thời gợi ý rằng các gen này có vai trò quan trọng trong sự phát triển của hạt ở loài cây này. Giả thiết này phù hợp với các báo cáo về sự biểu hiện cũng như vai trò của các *LEA5* và *SMP* ở một số loài thực vật khác.

Trong nghiên cứu của Du et al. (2013), các *LEA5* và *SMP* của cây mơ không biểu hiện ở các mô sinh dưỡng, chỉ hai *SMP* (*PmLEA5* và *PmLEA26*) biểu hiện yếu ở hoa [8]. Ở *A. thaliana*, tất cả các *LEA5* và *SMP* đều có biểu hiện ở hạt. Ngoài hạt, chỉ duy nhất một gen *LEA5* (*AtEM6*) biểu hiện ở nhiều mô và một gen *SMP* biểu hiện ở chồi [15]. Sự biểu hiện của các *LEA5* và *SMP* của cây *A. thaliana* được khẳng định trong công trình của Bies-Ethève et al. (2008) [6]. Một số ít các nghiên cứu sớm hơn cũng chỉ ra sự biểu hiện đặc hiệu của các *LEA5* và *SMP* trong hạt của các loài *A. thaliana* [2; 23; 24], ngô [18] và lúa mạch [14]. Đặc biệt, Manfre et al. (2006) đã chỉ ra sự cần thiết của gen *AtEM6* (họ *LEA5*) đối với sự phát triển bình thường của hạt cây *A. thaliana* [17].

KẾT LUẬN

Sử dụng các protein *LEA5* và *SMP* của cây *A. thaliana* làm khuôn, chúng tôi đã xác định được 5 gen tương đồng của mỗi họ trên trong hệ gen của cây đậu tương (*Glycine max*). Chúng tôi đã phân tích các đặc điểm lý - hóa của các gen cũng như các protein suy diễn của các gen này. Các *LEA5* và *SMP* có nhiều motif bảo thủ, các motif bảo thủ này có chứa nhiều Glycine (họ *LEA5*) hoặc các amino acid có chuỗi bên không phân cực (họ *SMP*).

Các *LEA5* và *SMP* của cây đậu tương biểu hiện đặc hiệu trong mô hạt, đặc biệt ở giai đoạn phát triển muộn của hạt. Cường độ biểu hiện của các gen cao nhất ở thời điểm 42 ngày sau thụ phấn. Các gen này không biểu hiện ở các mô sinh dưỡng của cây đậu tương. Sự biểu hiện đặc hiệu mô này gợi ý về vai trò quan trọng của các *LEA5* và *SMP* đối với sự phát triển của hạt đậu tương. Công trình này lần đầu tiên làm sáng tỏ các đặc điểm, sự biểu hiện của các *LEA5* và các *SMP* của cây đậu tương, mở đường cho các nghiên cứu chức năng của các gen của hai họ này ở các loài cây họ Đậu.

Lời cảm ơn: Công trình này được hoàn thành với sự hỗ trợ kinh phí từ chương trình nghiên cứu khoa học cơ bản của Trường Đại học Hùng Vương, tỉnh Phú Thọ.

TÀI LIỆU THAM KHẢO

1. Amara I., Capellades M., Ludevid M. D., Pages M., Goday A., 2013. Enhanced water stress tolerance of transgenic maize plants over-expressing LEA Rab28 gene. *J. Plant Physiol.*, 170(9): 864-873.
2. Arenas-Mena C., Raynal M., Borrell A., Varoquaux F., Cutanda M. C., Stacy R. A., Pagès M., Delseny M., Culiáñez-Macià F. A., 1999. Expression and cellular localization of Atrab28 during *Arabidopsis* embryogenesis. *Plant. Mol. Biol.*, 40(2): 355-363.
3. Bailey T. L., Boden M., Buske F. A., Frith M., Grant C. E., Clementi L., Ren J., Li W. W., Noble W. S., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, 37: W202-W208.
4. Bartels D., Sunkar R., 2005. Drought and Salt tolerance in plants. *Crit. Rev. Plant Sci.*, 24(1): 23-58.
5. Battaglia M., Covarrubias A. A., 2013. Late embryogenesis abundant (LEA) proteins in legumes. *Front. Plant Sci.*, 4. DOI: 10.3389/fpls.2013.00190
6. Bies-Ethève N., Gaubier-Comella P., Debures A., Lasserre E., Jobet E., Raynal M., Cooke R., Delseny M., 2008. Inventory, evolution and expression profiling diversity

- of the LEA (late embryogenesis abundant) protein gene family in *Arabidopsis thaliana*. *Plant Mol. Biol.*, 67(1-2): 107-124.
7. Cao Phi Bằng, Trần Thị Thanh Huyền, 2015. Phân tích họ gen β -amylase ở cây đậu tương (*Glycine max*). *Tạp chí Sinh học*, 37(1se): 165-176. DOI: 10.15625/0866-7160/v37n1se.6106.
 8. Du D., Zhang Q., Cheng T., Pan H., Yang W., Sun L., 2013. Genome-wide identification and analysis of late embryogenesis abundant (LEA) genes in *Prunus mume*. *Mol. Biol. Rep*, 40(2): 1937-1946.
 9. Dure L., Greenway S. C., Galau G. A., 1981. Developmental biochemistry of cottonseed embryogenesis and germination: changing messenger ribonucleic acid populations as shown by in vitro and in vivo protein synthesis. *Biochem.*, 20(14): 4162-4168.
 10. Finn R. D., Bateman A., Clements J., Coghill P., Eberhardt R. Y., Eddy S. R., Heger A., Hetherington K., Holm L., Mistry J., Sonnhammer E. L., Tate J., Punta M., 2014. Pfam: the protein families database. *Nucleic Acids Res.*, 42: D222-D230.
 11. Gasteiger E., Hoogland C., Gattiker A., Wilkins M. R., Appel R. D., Bairoch A., 2005. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, Springer: 571-607.
 12. Guo A. Y., Zhu Q. H., Chen X., Luo J. C., 2007. GSDS: a gene structure display server. *Yi Chuan*, 29:1023-1026.
 13. Hanin M., Brini F., Ebel C., Toda Y., Takeda S., Masmoudi K., 2011. Plant dehydrins and stress tolerance: versatile proteins for complex mechanisms. *Plant Signal. Behav.*, 6(10): 1503-1509.
 14. Hollung K., Espelund M., Jakobsen K. S., 1994. Another lea B19 gene (Group1 lea) from barley containing a single 20 amino acid hydrophilic motif. *Plant Mol. Biol.*, 25(3): 559-564.
 15. Hundertmark M., Hinch D. K., 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC genomics*, 9(1): 118. DOI: 10.1186/1471-2164-9-118
 16. Li Q., Yu H., Cao P. B., Fawal N., Mathé C., Azar S., Cassan-Wang H., Myburg AA2, Grima-Pettenati J1, Marque C., Teulières C., Dunand C., 2015. Explosive tandem and segmental duplications of multigenic families in *Eucalyptus grandis*. *Genome Biol. Evol.*, 7(4):1068-81.
 17. Manfre A. J., Lanni L. M., Marcotte W. R., 2006. The *Arabidopsis* group 1 late embryogenesis abundant protein ATEM6 is required for normal seed development *Plant Physiol.*, 140: 140-149.
 18. Niogret M. F., Cullanez-Macia F. A., Goday A., Mar Alba M., Pages M., 1996. Expression and cellular localization of rab28 mRNA and Rab28 protein during maize embryogenesis. *Plant J.*, 9(4): 549-557.
 19. Schmutz J., Cannon S. B., Schlueter J., Ma J., Mitros T., Nelson W., Hyten D. L., Song Q., Thelen J. J., Cheng J., Xu D., Hellsten U., May G. D., Yu Y., Sakurai T., Umezawa T., Bhattacharyya M. K., Sandhu D., Valliyodan B., Lindquist E., Peto M., Grant D., Shu S., Goodstein D., Barry K., Futrell-Griggs M., Abernathy B., Jianchang Du J., Tian Z., Zhu L., Gill N., Joshi T., Libault M., Sethuraman A., Zhang X. C., Shinozaki K., T. Nguyen H. T., Wing R. A., Cregan P., Specht J., Grimwood J., Rokhsar D., Stacey G., Shoemaker R. C., Jackson S. A., 2010. Genome sequence of the palaeopolyploid soybean. *Nature*, 463(7278): 178-183.
 20. Severin A. J., Woody J. L., Bolon Y. T., Joseph B., Diers B. W., Farmer A. D., Muehlbauer G. J., Nelson R. T., Grant D., Specht J. E., Graham M. A., Cannon S. B., May G. D., Vance C. P., Shoemaker R. C., 2010. RNA-Seq Atlas of *Glycine max*: a guide to the soybean transcriptome. *BMC Plant Biol.*, 10, 160. DOI: 10.1186/1471-2229-10-160

21. Shih M. D., Hsieh T. Y., Jian W. T., Wu M. T., Yang S. J., Hoekstra F. A., Hsing Y. I., 2012. Functional studies of soybean (*Glycine max* L.) seed LEA proteins GmPM6, GmPM11, and GmPM30 by CD and FTIR spectroscopy. *Plant Sci.*, 196: 152-159.
22. Stacy R. A., Espelund M., Saeboe-Larssen S., Hollung K., Helliesen E., Jakobsen K. S., 1995. Evolution of the Group 1 late embryogenesis abundant (Lea) genes: analysis of the Lea B19 gene family in barley. *Plant Mol. Biol.*, 28(6): 1039-1054.
23. Vicent C. M., Delseny M., 1999. Isolation of total RNA from *Arabidopsis thaliana* seeds. *Anal. Biochem.*, 268: 412-413.
24. Vicent C. M., Hull G., Guillemot J., Devic M., Delseny M., 2000. Differential expression of the *Arabidopsis* genes coding for Em-like proteins. *J. Exp. Bot.*, 51(348): 1211-1220.

CHARACTERISATION AND EXPRESSION OF TWO LATE EMBRYOGENESIS ABUNDANT (LEA) PROTEIN FAMILIES IN SOYBEAN (*Glycine max*)

Cao Phi Bang

Hung Vuong University, Phu Tho, phibang.cao@hvu.edu.vn

SUMMARY

Late Embryogenesis Abundant proteins (LEA) are abundantly present in late embryogenesis and play an important role in the development of embryos. Many LEA subfamilies were reported to have a major role in abiotic stress tolerance in plants. According to current classification, LEA5 family and SMP family belong to the LEA superfamily. In this study, we identified five coding genes for each of these families in the genome of the soybean (*Glycine max*). In the case of soybean LEA5 genes, there is generally only one intron per gene, except for GmLEA5-3. This gene was not annotated in the soybean genome and had no intron. All of the soybean LEA5 predicted proteins were small in size, weakly acidic but highly hydrophilic. The SMP genes have one or more introns interspersed within the coding region. Their predicted proteins were weakly acidic and hydrophilic. All of the soybean *LEA5* and *SMP* genes did not express in vegetative tissues while most of them were tissue-specifically expressed in seeds. The expression level increased in accordance with seed development, and arrived at the maximum level 42 days after flowering. The tissue-specific expression suggested that these genes played an important role in the development of soybean seeds. In brief, for the first time, the characters and the expression of the soybean *LEA5* and *SMP* genes were reported, paving the way for future gene function research about these two LEA families in Leguminosae plants.

Keywords: Gene expression, phylogenetic tree, gene characterization, LEA5, SMP, soybean.

Ngày nhận bài: 8-1-2015