

## XÁC ĐỊNH VÀ PHÂN TÍCH CÁC GENE *HSP90* Ở CÂY ĐU ĐỦ (*Carica papaya* L.) BẰNG PHƯƠNG PHÁP TIN SINH HỌC

Lê Thị Mận<sup>1</sup>, Nguyễn Quốc Nam<sup>2</sup>, Trần Thị Thanh Huyền<sup>3</sup>, La Việt Hồng<sup>4</sup>  
và Cao Phi Bằng<sup>1,\*</sup>

<sup>1</sup>Khoa Khoa học Tự nhiên, Trường Đại học Hùng Vương

<sup>2</sup>Trường Trung học Cơ sở Kim Đức

<sup>3</sup>Khoa Sinh học, Trường Đại học Sư phạm Hà Nội

<sup>4</sup>Khoa Sinh - Kỹ thuật Nông nghiệp, Trường Đại học Sư phạm Hà Nội 2

**Tóm tắt.** Họ gene *HSP90* đã được chứng minh có vai trò quan trọng đối với khả năng chống chịu cũng như sự phát triển của thực vật. Đu đủ là cây ăn quả có giá trị dinh dưỡng cao, nguồn gốc nhiệt đới nhưng hiện nay được trồng rộng rãi ở nhiều vùng cận nhiệt đới trên thế giới. Vì vậy, cây đu đủ phải đối mặt với nhiều tác nhân của môi trường trong quá trình sống. Nghiên cứu này nhằm xác định và phân tích họ gene *HSP90* ở cây đu đủ bằng phương pháp tin sinh học. Bảy gene *HSP90* đã được xác định ở trong hệ gene của cây đu đủ (*Carica papaya* L.) nhờ sử dụng các phương pháp nghiên cứu tin sinh học. Các gen *HSP90* của cây đu đủ có kích thước từ 2650 tới 8136 nucleotide, mã hóa không liên tục, số lượng intron từ 2 đến 19. Các protein suy diễn có kích thước từ 348 tới 796 amino acid, khối lượng phân tử nằm trong khoảng 39,92 kDa tới 90,61 kDa. Trong số bảy *HSP90*, *CpHSP90-1* và *CpHSP90-4* được coi là các gene giả (pseudogene) do kích thước của chúng bé hơn so với các *HSP90* thông thường và protein thiếu vùng bảo tồn HATPase\_c. Các protein suy diễn có tính acid với pI dao động từ 4,69 đến 5,42, ngoại trừ *CpHSP90-1* (pI 7,03). Căn cứ vào kết quả phân tích cấu trúc, vị trí khu trú dưới tế bào và cây phả hệ, các *HSP90* của cây đu đủ được phân chia thành hai nhóm, nhóm I (*HSP90* tế bào chất gồm 4 protein) và nhóm II (*HSP90* bào quan gồm 2 protein). Phân tích dữ liệu RNA-seq cho thấy rằng các *HSP90* của cây đu đủ có sự biểu hiện khác nhau ở các loại mô/cơ quan khác nhau ở các giai đoạn phát triển khác nhau. Trong đó, hầu hết *HSP90* của cây đu đủ biểu hiện mạnh ở chồi hoa hoặc quả ở giai đoạn 2 hoặc 3. Trong đó, gene *CpHSP90-2* là gene có mức độ biểu hiện mạnh nhất, kế tiếp là *CpHSP90-5*. Ngược lại, *CpHSP90-1* hầu như không biểu hiện hoặc biểu hiện rất thấp ở các mô/cơ quan này. Tất cả các *HSP90* của cây đu đủ đều bị cảm ứng bởi điều kiện đông lạnh, trong đó *CpHSP90-1* có mức độ cảm ứng mạnh nhất (12,13 lần), tuy nhiên gene này là một gene giả và có mức độ biểu hiện cơ sở rất thấp. *CpHSP90-2* có mức độ cảm ứng mạnh (2,81 lần) đồng thời cũng có mức độ biểu hiện cơ sở lớn. Kết quả nghiên cứu này có ý nghĩa lớn, là cơ sở cho các nghiên cứu về tách dòng gene, phân tích chức năng của các gene trong họ *HSP90* và chọn giống ở cây đu đủ trong đáp ứng với các điều kiện stress vô sinh của môi trường cũng như sự phát triển của cây đu đủ.

**Từ khóa:** biểu hiện gene, cây đu đủ (*Carica papaya* L.), cây phả hệ, in silico, protein sốc nhiệt 90 (*HSP90*).

## 1. Mở đầu

Cây đu đủ có tên khoa học là *Carica papaya* L. Đây là loại cây ăn quả quan trọng được trồng nhiều ở vùng nhiệt đới và cận nhiệt đới. Quả đu đủ chín chứa lượng lớn tiền vitamin A, vitamin C và các chất chống oxy hóa, quả có vị ngọt, mềm [1]. Trong môi trường sống, cây đu đủ luôn chịu tác động của các yếu tố vô sinh như ánh sáng, nhiệt độ, chế độ nước, gió cũng như nhiều tác nhân gây bệnh [2].

Các HSP90 là tập hợp nhóm protein sốc nhiệt với khối lượng phân tử xấp xỉ 90 kDa. Sự điều hòa biểu hiện các gene mã hóa HSP90 chịu tác động của nhiệt độ cũng như một số nhân tố khác [3]. HSP90 cũng được chỉ ra có chức năng của các phân tử chaperon [4]. Vì vậy, HSP90 có vai trò quan trọng trong nhiều quá trình sinh lý quan trọng cũng như truyền tín hiệu stress [3]. Về cấu trúc, các protein HSP90 có chứa các vùng bảo tồn ở đầu N và đầu C, giữa hai vùng bảo tồn này là vùng liên kết với độ dài thay đổi [5]. Các gene mã hóa HSP90 đã được phân tích trong hệ gene của các cây *Arabidopsis thaliana* (7 gene) [6], nho (*Vitis vinifera*) (7 gene) [7], dưa chuột (*Cucumis sativus*) (6 gene) [8] và cây thuốc lá (*Nicotiana tabacum*) (21 gene) [9].

Nghiên cứu này hướng tới mục tiêu xác định và phân tích đặc điểm của các gene HSP90 bằng các phương pháp tin sinh học ở cây đu đủ, loại cây ăn quả nhiệt đới quan trọng. Kết quả nghiên cứu là tiền đề cho các nghiên cứu về chọn tạo giống loại cây ăn quả có giá trị này cũng như tiền đề cho các nghiên cứu sâu hơn về chức năng của họ HSP90. Hướng nghiên cứu các họ gene liên quan đến khả năng chống chịu của cây đu đủ bằng phương pháp tin sinh học vốn đã được thực hiện trong nhóm nghiên cứu [10].

## 2. Nội dung nghiên cứu

### 2.1. Cơ sở dữ liệu và phương pháp nghiên cứu

#### 2.1.1. Cơ sở dữ liệu

Hệ gene của cây đu đủ (*Carica papaya* L.) đã được giải trình tự [11] và được đặt trên server phytozome (<https://phytozome-next.jgi.doe.gov/>). Dữ liệu RNA-seq được khai thác dữ liệu là cơ sở dữ liệu Gene Expression Omnibus, NCBI (<https://www.ncbi.nlm.nih.gov/geo/>) của chồi hoa ở hai giai đoạn phát triển trước (pre-M) và sau (post-M) nguyên phân (GSE137547) [12], lá và quả ở các giai đoạn phát triển (GSE116581) [13] và lá cây non được xử lý hay không xử lý bởi đông lạnh (GSE130188).

#### 2.1.2. Phương pháp nghiên cứu

\* *Xác định các gene HSP90 ở cây đu đủ*

Chương trình TBLASTN [14] được sử dụng để tìm kiếm các gene tương đồng trên toàn hệ gene của cây đu đủ, sử dụng bảy protein HSP90 của cây *A. thaliana* [6] làm trình tự khuôn dò.

\* *Phân tích in silico các HSP90 của cây đu đủ*

Các đặc điểm vật lý, hóa học của các CpHSP90 được phân tích bằng công cụ ProtParam trên server ExPASy (Expert Protein Analysis System) [15]. Vị trí khu trú dưới tế bào được phân tích nhờ ProtComp 9.0.

\* *Xây dựng cây phả hệ*

Các protein HSP90 của cây đu đủ, *A. thaliana*, cây nho và dưa chuột được sắp dây bằng MAFFT [16], cây phả hệ được xây dựng nhờ phần mềm MEGA X [17].

\* *Phân tích sự biểu hiện gene*

Sự biểu hiện của các gene được xác định qua phân tích dữ liệu RNA-seq của cây đu đủ trong ngân hàng Gene Expression Omnibus của chồi hoa ở hai giai đoạn phát triển trước (pre-M) và sau (post-M) nguyên phân (GSE137547) [12], lá và quả ở các giai đoạn phát triển

(GSE116581) [13], lá cây non được xử lí hay không xử lí bởi đông lạnh (GSE130188). Mức độ biểu hiện tương đối của các gene *CpHSP90* được tính dựa trên giá trị biểu hiện của gene (FPKM) theo công thức:

$$A = \frac{B / C}{D / E}$$

Trong đó: A: Mức độ biểu hiện tương đối của *CpHSP90-x*, B: Mức độ biểu hiện của gene *CpHSP90-x* ở điều kiện đông lạnh (FPKM), C: Mức độ biểu hiện của gene *CpHSP90-x* ở điều kiện thường (FPKM), D: Mức độ biểu hiện của gene *elF4E* ở điều kiện đông lạnh (FPKM), E: Mức độ biểu hiện của gene *elF4E* ở điều kiện thường (FPKM)

Gene *elF4E* được chọn làm gene chuẩn vì gene này đã được chứng minh biểu hiện ổn định trong các điều kiện khác nhau ở cây đu đủ và phù hợp để chuẩn hóa trong các nghiên cứu biểu hiện gene ở các điều kiện khác nhau của cây đu đủ [18].

## 2.2. Kết quả nghiên cứu và thảo luận

### 2.2.1. Xác định và phân tích họ gene *HSP90* ở cây đu đủ

**Bảng 1. Các trình tự *HSP90* ở cây đu đủ**

Gene	Tên locus	Nhóm	GS (bp)	PL (aa)	MW (kD)	pI	GRAVY	In	SCL
<i>CpHSP90-1</i>	evm.TU.supercontig_10.56	I	6029	348	39,92	7,03	-0,832	3	Cyto
<i>CpHSP90-2</i>	evm.TU.supercontig_12.106	I	2989	699	80,03	5,00	-0,579	2	Cyto
<i>CpHSP90-3</i>	evm.TU.supercontig_29.53	I	2650	703	80,80	4,99	-0,604	3	Cyto
<i>CpHSP90-4</i>	evm.TU.supercontig_64.164	II	4808	423	48,39	4,69	-0,628	10	ER
<i>CpHSP90-5</i>	evm.TU.supercontig_107.112	I	2728	703	80,94	4,96	-0,616	3	Cyto
<i>CpHSP90-6</i>	evm.TU.supercontig_136.24	II	6030	792	90,01	5,42	-0,525	19	Mito
<i>CpHSP90-7</i>	evm.TU.supercontig_998.4	II	8136	796	90,61	4,90	-0,585	18	Chlo

Chú thích: GS = Kích thước gen, PL = Chiều dài phân tử protein, MW = Khối lượng phân tử protein, pI = điểm đẳng điện, In = Số lượng intron, SCL = Vị trí khu trú dưới tế bào của protein, Cyto: tế bào chất, Chlo: lục lạp, Mito: ti thể, ER: lưới nội chất

Tổng số 7 gene có thể mã hóa cho các *HSP90* được xác định trong hệ gene của cây đu đủ (Bảng 1). Các *HSP90* điển hình được biết có chứa các vùng bảo tồn đặc trưng *HSP90* (PF00183) và *HATPase\_c* (PF02518) [19]. Kiểm tra Pfam [20] các protein suy diễn chỉ ra tất cả các *CpHSP90* có chứa vùng bảo tồn *HSP90* (PF00183) trong khi chỉ năm trong bảy *CpHSP90* có chứa vùng bảo tồn *HATPase\_c* (PF02518), ngoại trừ *CpHSP90-1* và *CpHSP90-4*. Hai gene này với kích thước ngắn và thiếu vùng bảo tồn nên được xác định là gene giả (pseudogene). Với bảy gene, họ gene *HSP90* của cây đu đủ tương đương so với ở các loài *A. thaliana* (7 gene) [6], nho (7 gene) [7], lớn hơn so với dưa chuột (6 gene) [8] nhưng bé hơn so với cây thuốc lá (21 gene) [9].

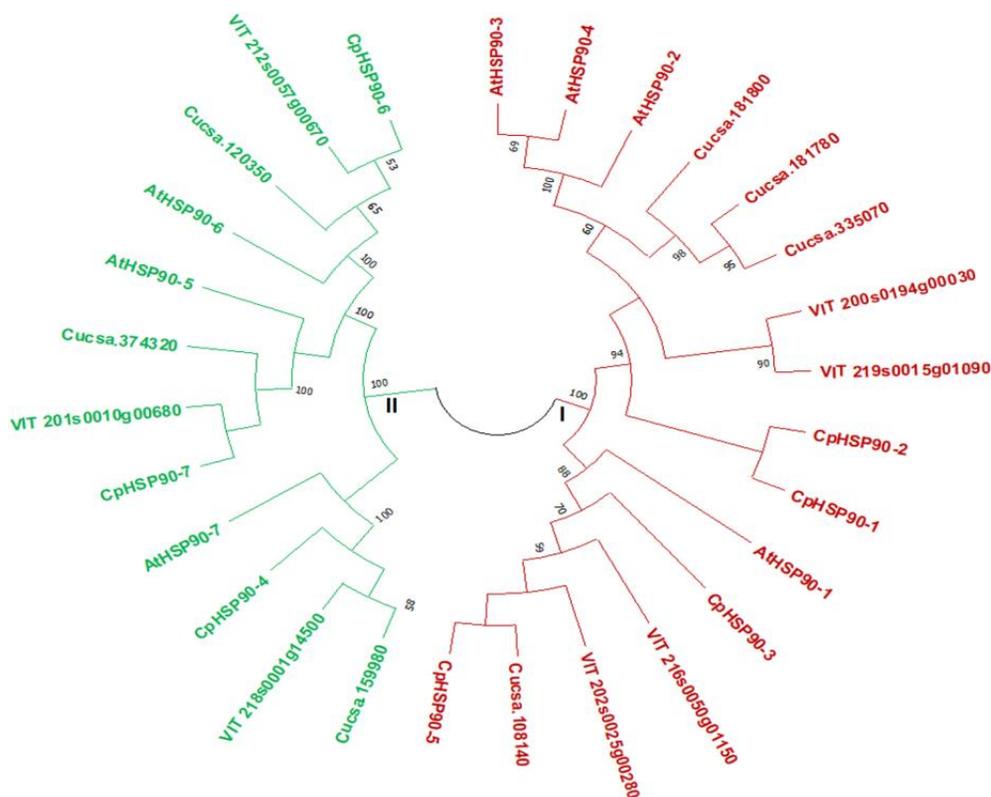
Các gene *HSP90* của cây đu đủ có độ dài trình tự nucleotide khác nhau, từ 2650 tới 8136 nucleotide (Bảng 1). Tất cả các gene đều mã hóa không liên tục, số lượng intron ở các gene *CpHSP90-1* đến *CpHSP90-7* lần lượt bằng 3, 2, 3, 10, 3 19 và 18. Ở cây *A. thaliana*, các gene có từ 2 đến 19 intron [6]. Tương tự, ở cây thuốc lá số intron dao động từ 3 đến 19 [9]. Các protein suy diễn có độ dài từ 348 tới 796 amino acid, khối lượng phân tử nằm trong khoảng 39,92 kDa tới 90,61 kDa, trong đó, hai pseudogene mã hóa các protein *CpHSP90-1* và *CpHSP90-4* có lần lượt 348 và 423 amino acid, trong khi các protein còn lại có từ 699 đến 796 amino acid, khối lượng dao động từ 80,03 đến 90,61 kDa. Ngoại trừ *CpHSP90-1*, các protein

suy diễn có tính acid, giá trị điểm đẳng điện (pI) lí thuyết của các cpHSP90 dao động từ 4,69 tới 5,42 (Bảng 1). Điều này chứng tỏ các CpHSP90 giàu các amino acid có tính acid. Các đặc điểm này khá tương đồng với đặc điểm của các HSP90 của cây *A. thaliana* [6] và cây thuốc lá [9].

Kết quả phân tích vị trí khu trú dưới tế bào cho thấy, có 4 trong tổng số 7 HSP90 của cây đu đủ khu trú ở trong tế bào chất (CpHSP90-1, CpHSP90-2, CpHSP90-3 và CpHSP90-5). Ba protein còn lại nằm ở các bào quan như lục lạp (CpHSP90-7), ti thể (CpHSP90-6) và lưới nội chất (CpHSP90-4). Kết quả nghiên cứu này tương đồng với ở cây *A. thaliana*, khi có các HSP90 nằm ở tế bào chất (AtHSP90-1 đến AtHSP90-4), lục lạp (AtHSP90-5), ti thể (AtHSP90-6), lưới nội chất (AtHSP90-7) [6].

### 2.2.2. Phân tích đặc điểm tiến hóa các HSP90 ở cây đu đủ

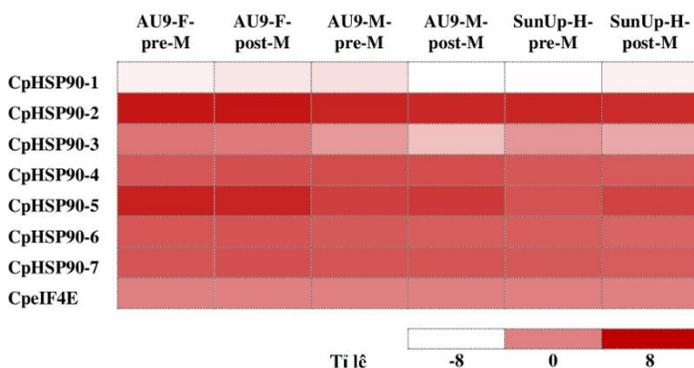
Cây phả hệ (Hình 1) cho thấy các HSP90 ở cây đu đủ được xếp vào hai nhóm chính, nhóm I (gồm 4 thành viên, CpHSP90-1 đến CpHSP90-3, CpHSP90-5), và nhóm II chỉ gồm 3 thành viên còn lại, CpHSP90-4, CpHSP90-6 và CpHSP90-7. Phù hợp với phân tích vị trí khu trú dưới tế bào, các HSP90 nhóm I khu trú trong tế bào chất, nhóm II chỉ khu trú ở các bào quan (Bảng 1). Một cách phân chia khác đã được miêu tả, các HSP90 được chia thành 5 nhóm, trong đó các HSP90 tế bào chất được chia thành hai nhóm trong khi mỗi HSP90 bào quan là một nhóm khác nhau [6, 7]. Trong khi đó, trong nghiên cứu về HSP90 của cây thuốc lá cho thấy các HSP90 có thể được xếp vào 10 nhóm khác nhau, nhưng không phải tất cả các nhóm đều có đại diện của *A. thaliana* [9]. Cây phả hệ cũng cho thấy rằng có một hiện tượng nhân gene sau quá trình phát sinh loài ở cây đu đủ (so với *A. thaliana*, dưa chuột và nho), hình thành nên hai gene *CpHSP90-1* và *CpHSP90-2*.



**Hình 1.** Cây phả hệ Maximum-Likelihood được thiết lập từ các HSP90 ở các loài đu đủ (Cp), *A. thaliana* (At) và nho (VIT) nhờ phần mềm MEGAX [18] với 1000 bootstrap lặp lại  
Giá trị bootstraps được thể hiện ở gốc của nhánh

### 2.2.3. Phân tích *in silico* sự biểu hiện gene *HSP90* ở cây đu đủ

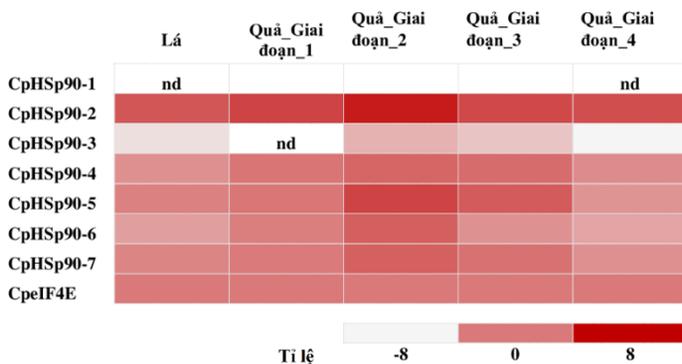
Kết quả phân tích sự biểu hiện gene *HSP90* của cây đu đủ ở các loại mô khác nhau cũng như điều kiện tác động khác nhau được trình bày trong các Hình 2, 3 và 4.



**Hình 2. Bản đồ nhiệt thể hiện sự biểu hiện tương đối của các gene *HSP90* ở các mô chồi hoa của cây đu đủ (giá trị log<sub>2</sub> của tỉ lệ tương đối giữa mức độ biểu hiện của gene nghiên cứu với gene *CpeIF4E*)**

AU9: giống đu đủ AU9, SunUp: giống đu đủ SunUp, F: hoa cái, M: hoa đực, H: hoa lưỡng tính, Pre-M: trước nguyên phân, Post-M: sau nguyên phân

Ngoại trừ *CpHSP90-1*, các gene còn lại đều biểu hiện ở chồi hoa đực, cái và cả lưỡng tính (Hình 2). Trong đó, *CpHSP90-2* biểu hiện mạnh nhất ở tất cả các loại chồi hoa. Khi phân tích sâu hơn, các gene *CpHSP90-2*, *CpHSP90-3*, *CpHSP90-5* và *CpHSP90-6* biểu hiện mạnh hơn ở chồi hoa cái so với chồi hoa đực và chồi hoa lưỡng tính. Trong khi đó, *CpHSP90-4* biểu hiện ở chồi hoa đực mạnh hơn ở chồi hoa cái và chồi hoa lưỡng tính. *CpHSP90-7* biểu hiện ở hai loại chồi hoa đực và cái tương đương và đều cao hơn so với chồi hoa lưỡng tính. Sự biểu hiện của các gene này cũng không giống nhau theo pha phát triển của chồi hoa. Các gene *CpHSP90-2*, *CpHSP90-4*, *CpHSP90-6* và *CpHSP90-7* biểu hiện ở hoa cái mạnh hơn trong pha sau nguyên phân so với trong pha trước giảm phân. Tương tự, *CpHSP90-5* biểu hiện ở chồi hoa đực và lưỡng tính mạnh hơn trong pha sau nguyên phân so với trong pha trước giảm phân. Ngược lại, gene này lại biểu hiện ở chồi hoa cái yếu hơn trong pha sau nguyên phân so với trong pha trước giảm phân. Tương tự, ba gene *CpHSP90-2*, *CpHSP90-4* và *CpHSP90-6* biểu hiện ở hoa cái và hoa lưỡng tính yếu hơn trong pha sau nguyên phân so với trong pha trước nguyên phân. Riêng *CpHSP90-3* luôn biểu hiện mạnh hơn trong pha trước nguyên phân so với trong pha sau nguyên phân ở tất cả các loại chồi hoa.

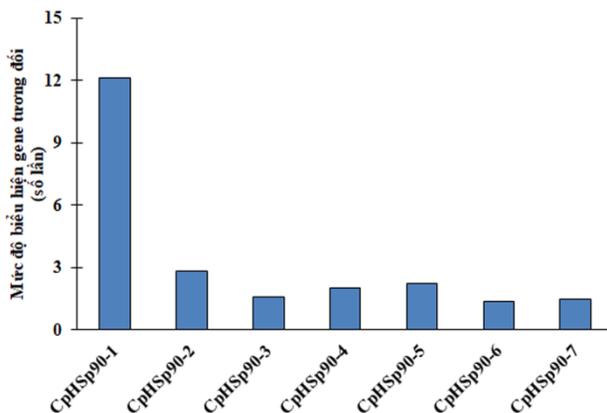


**Hình 3. Bản đồ nhiệt thể hiện sự biểu hiện tương đối của các gene *HSP90* ở các mô lá và quả của cây đu đủ (giá trị log<sub>2</sub> của tỉ lệ tương đối giữa mức độ biểu hiện của gene nghiên cứu với gene *CpeIF4E*)**

DPA: ngày sau nở hoa

Khi phân tích sự biểu hiện của các gene *CpHSP90* trong mô lá và quả ở bốn giai đoạn phát triển khác nhau, kết quả cho thấy gene *CpHSP90-1* hầu như không biểu hiện hoặc biểu hiện rất yếu ở các loại mô nghiên cứu (Hình 3). Sự biểu hiện của *CpHSP90-3* được phát hiện ở cả lá và các mô quả nhưng cũng ở mức yếu. Ngược lại, *CpHSP90-2* là gene biểu hiện mạnh nhất ở các mô nghiên cứu. Bốn gene *CpHSP90-2*, *CpHSP90-4*, *CpHSP90-5*, *CpHSP90-6* và *CpHSP90-7* biểu hiện mạnh nhất ở mô quả giai đoạn 2 và giai đoạn 3 so với các loại mô khác.

Sự biểu hiện khác nhau của các gene theo loại mô cũng như theo pha phát triển cho thấy vai trò đa dạng của các *CpHSP90* đối với sự phát triển của đu đủ, đặc biệt ở giai đoạn phát triển sinh sản. Kết quả này khẳng định các nghiên cứu trước đây khi chứng minh *CpHSP90* tham gia vào điều tiết sự phát triển thực vật khi biểu hiện khác nhau ở các loại mô khác nhau [7, 8].



**Hình 4. Mức độ biểu hiện tương đối của các gene HSP90 ở lá cây đu đủ không/có xử lý đông lạnh (giá trị tỉ lệ tương đối giữa mức độ biểu hiện của gene nghiên cứu ở điều kiện đông lạnh/điều kiện thường)**

Tất cả các gene *HSP90* của cây đu đủ đều biểu hiện ở điều kiện đông lạnh cao hơn so với ở điều kiện thường (Hình 4). Mức độ biểu hiện tương đối của các gene *CpHSP90-1-7* lần lượt đạt 12,13; 2,81; 1,56; 2,01; 2,25; 1,39 và 1,48. Như vậy, gene *CpHSP90-1* phản ứng với điều kiện đông lạnh mạnh nhất trong các gene nghiên cứu, kế tiếp là gene *CpHSP90-2*, *CpHSP90-5* và *CpHSP90-4*. Gene *CpHSP90-6* phản ứng với điều kiện đông lạnh thấp nhất (1,39 lần) trong họ gene *HSP90* ở cây đu đủ. Tuy nhiên, cần lưu ý rằng, gene *CpHSP90-2* và *CpHSP90-5* là hai gene có mức biểu hiện cơ sở lớn nhất trong họ *HSP90* ở điều kiện thường (lần lượt bằng 348,55 và 135,05 FPKM) (Bảng 2).

**Bảng 2. Mức độ biểu hiện của các gene HSP90 ở lá cây đu đủ trong điều kiện không/có xử lý đông lạnh**

Gene	Không xử lý đông lạnh (FPKM)	Có xử lý đông lạnh (FPKM)
<i>CpHSp90-1</i>	1,19	12,79
<i>CpHSp90-2</i>	348,55	868,72
<i>CpHSp90-3</i>	11,02	15,26
<i>CpHSp90-4</i>	17,3	30,8
<i>CpHSp90-5</i>	135,05	269,57
<i>CpHSp90-6</i>	88,65	109,39
<i>CpHSp90-7</i>	19,92	26,19
<i>eIF4E</i>	96,69	85,69

Nhiều nghiên cứu đã được chỉ ra các *HSP90* có đáp ứng với một số điều kiện stress vô sinh. Có 4 trong 5 gene *HSP90* ở cây nho đáp ứng với điều kiện nhiệt độ cao (37 và 45 °C) nho [7]. Tương tự, cả 6 gene *HSP90* ở cây dưa chuột đều đáp ứng với điều kiện nhiệt độ cao [8]. Ở cây thuốc lá, các gene *NtHSP90-4*, *NtHSP90-5* và *NtHSP90-9* phản ứng mạnh với nhiệt độ cao trong khi các gene khác như *NtHSP90-6* và *NtHSP90-7* không bị cảm ứng bởi nhiệt độ cao [9]. Sự biểu hiện của các gene này cũng có đáp ứng tương tự với các điều kiện khác như hạn, mặn, lạnh và với ABA [9]. Kết quả nghiên cứu biểu hiện gene gợi ý rằng các *HSP90* có thể là nhóm gene tiềm năng trong các nghiên cứu tiếp theo về phản ứng của cây đu đủ với các điều kiện stress vô sinh cũng như trong các nghiên cứu chọn giống, cải tạo giống cây đu đủ đáp ứng với các thay đổi của môi trường.

### 3. Kết luận

Trong hệ gene của cây đu đủ có tổng số 7 gene mã hóa *HSP90* được xác định, trong đó có hai gene giả (pseudogene). Cấu trúc gene và các đặc điểm lí - hóa của các protein *HSP90* của cây đu đủ cũng được phân tích. Ngoại trừ hai gene giả, các *HSP90* của cây đu đủ có chiều dài từ 699 đến 796 amino acid, khối lượng từ 80,03 đến 90,61 kDa. Các protein *CpHSP90* có tính acid và có ái lực với nước lớn. Các *HSP90* của cây đu đủ được xếp vào hai nhóm, nhóm I (các *HSP90* tế bào chất) và II (các *HSP90* bào quan). Sự biểu hiện gene của các gene *HSP90* ở lá cây đu đủ đã được phân tích từ các *RNA-seq* cho thấy sự biểu hiện của các gene khác nhau không giống nhau ở các mô chồi hoa cái, đực và lưỡng tính ở hai giai đoạn trước và sau nguyên phân. Tương tự, các gene này cũng biểu hiện khác nhau trong mô lá và mô quả ở các giai đoạn phát triển. Trong đó, *CpHSP90-2* là gene biểu hiện mạnh nhất trong họ *HSP90* của cây đu đủ. Khi khảo sát sự biểu hiện trong điều kiện thường và điều kiện đông lạnh thông qua ngân hàng RNA-Seq, tất cả các *CpHSP90* đều biểu hiện tăng cường trong điều kiện đông lạnh so với điều kiện thường, mạnh nhất là gene *CpHSP90-1*, nhưng gene này có mức độ biểu hiện cơ sở rất thấp. *CpHSP90-2* biểu hiện tăng cường trong điều kiện đông lạnh với mức độ gấp 2,81 lần và có mức độ biểu hiện cơ sở lớn nhất trong họ *HSP90* của cây đu đủ. Kết quả này có ý nghĩa lớn, mở đường cho việc tách dòng gene và phân tích chức năng của các gene trong họ *HSP90* ở cây đu đủ trong đáp ứng với các điều kiện stress vô sinh khác.

**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 (Đề tài mã số: 24/2020/HĐKH.HV20-24).

### TÀI LIỆU THAM KHẢO

- [1] R. Ming, Q. Yu, P. H. Moore, R. E. Paull, N. J. Chen, M-L. Wang, Y. J. Zhu, M. A. Schuler, J. Jiang and A. H. Paterson., 2012, Genome of papaya, a fast growing tropical fruit tree. *Tree Genetics & Genomes*, Vol. 8, No. 3, pp. 445-462, doi: 10.1007/s11295-012-0490-y.
- [2] E. Campostrini and D. M. Glenn, 2007, Ecophysiology of papaya: a review, *Brazilian Journal of Plant Physiology*, Vol. 19, No. 4, pp. 413-424.
- [3] Z. S. Xu, Z. Y. Li, Y. Chen, M. Chen, L. C. Li, and Y. Z. Ma, 2012 Heat shock protein 90 in plants: molecular mechanisms and roles in stress responses. *Int J Mol Sci*, Vol. 13, No. 12, pp. 15706-23, doi: 10.3390/ijms131215706.
- [4] S. A. Walford, J. A. Thomson, J. M. Farrant, and S. G. Mundree, 2003 The *HSP90* family of chaperones: a look at the structure, function and mode of action. *Current Topics in Plant Biology*, Vol 4, pp. 1-25.
- [5] L. H. Pearl and C. Prodromou, 2000. Structure and in vivo function of Hsp90. *Current opinion in Structural Biology*, Vol. 10, No. 1, pp. 46-51.

- [6] P. Krishna and G. Gloor, 2001 The Hsp90 family of proteins in *Arabidopsis thaliana*. *Cell Stress Chaperones*, Vol. 6, No. 3, pp. 238-46.
- [7] G. Banilas, E. Korkas, V. Englezos, A. A. Nisiotou, and P. Hatzopoulos, 2012. Genome-wide analysis of the heat shock protein 90 gene family in grapevine (*Vitis vinifera* L.), *Australian Journal of Grape and Wine Research*, Vol. 18, No. 1, pp. 29-38.
- [8] K. Zhang, S. He, Y. Sui, Q. Gao, S-H. Jia, X-M. Lu, and L. Jia, 2021. Genome-wide characterization of HSP90 gene family in cucumber and their potential roles in response to abiotic and biotic stresses. *Front Genet*, Vol. 12, p. 584886, doi: 10.3389/fgene.2021.584886.
- [9] Z. Song, F. Pan, C. Yang, H-F. Jia, H. Jiang, F. He, Na. Li, X. Lu, and H. Zhang, 2019. Genome-wide identification and expression analysis of HSP90 gene family in *Nicotiana tabacum*. *BMC Genetics*, Vol. 20, No. 1, p. 35, doi: 10.1186/s12863-019-0738-8.
- [10] Lê Thị Mận, Trần Thị Thanh Huyền, Lê Chí Toàn, Trần Thị Thanh Loan, Cao Phi Bằng, 2021. Xác định và phân tích các gene mã hóa protein liên kết stress (SAP) ở cây đu đủ (*Carica papaya* L.) bằng phương pháp *in silico*. *Tạp chí Khoa học Trường Đại học Sư phạm Hà Nội*, Vol. 66, No. 1, pp. 111-118, 2021, doi: 10.18173/2354-1059.2021-0014.
- [11] R. Ming, S. Hou, Y. Feng, Q. Yu, A. Dionne-Laporte, J. H. Saw, P. Senin, W. Wang, B. V. Ly, K. L. T. Lewis, S. L. Salzberg, L. Feng, M. R. Jones, R. L. Skelton, J. E. Murray, C. Chen, W. Qian, J. Shen, P. Du, M. Eustice, E. Tong, H. Tang, E. Lyons, R. E. Paull, T. P. Michael, K. Wall, D. W. Rice, H. Albert, M-L Wang, Y. J. Zhu, M. Schatz, N. Nagarajan, R. A. Acob, P. Guan, A. Blas, C. M. Wai, C. M. Ackerman, Y. Ren, C. Liu, J. Wang, J. Wang, J-K. Na, E. V. Shakirov, B. Haas, J. Thimmapuram, D. Nelson, X. Wang, J. E. Bowers, A. R. Gschwend, A. L. Delcher, R. Singh, J. Y. Suzuki, S. Tripathi, K. Neupane, H. Wei, B. Irikura, M. Paidi, N. Jiang, W. Zhang, G. Presting, A. Windsor, R. Navajas-Pérez, M. J. Torres, F. A. Feltus, B. Porter, Y. Li, A. M. Burroughs, M-C Luo, L. Liu, D. A. Christopher, S. M. Mount, P. H. Moore, T. Sugimura, J. Jiang, M. A. Schuler, V. Friedman, T. Mitchell-Olds, D. E. Shippen, C. W. dePamphilis, J. D. Palmer, M. Freeling, A. H. Paterson, D. Gonsalves, L. Wang and M. Alam, 2008, The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* L.). *Nature*, Vol. 452, pp. 991-996, doi: 10.1038/nature06856.
- [12] D. Zerpa-Catanho, J. Wai, M. L. Wang, L. Yu, J. Nguyen, and R. Ming, 2019. Differential gene expression among three sex types reveals a MALE STERILITY 1 (CpMS1) for sex differentiation in papaya. *BMC Plant Biol*, Vol. 19, No. 1, p. 545, doi: 10.1186/s12870-019-2169-0.
- [13] P. Lü, S. Yu, N. Zhu, Y-R. Chen, B. Zhou, Y. Pan, D. Tzeng, J. P. Fabi, J. Argyris, J. Garcia-Mas, N. Ye, J. Zhang, D. Grierson, J. Xiang, Z. Fei, J. Giovannoni, and S. Zhong, 2018. Genome encode analyses reveal the basis of convergent evolution of fleshy fruit ripening. *Nat Plants*, Vol. 4, No. 10, pp. 784-791, doi: 10.1038/s41477-018-0249-z.
- [14] E. M. Gertz, Y. K. Yu, R. Agarwala, A. A. Schaffer, and S. F. Altschul, 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol*, Vol. 4, p. 41, 2006, doi: 10.1186/1741-7007-4-41.
- [15] E. Gasteiger, C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, and A. Bairoch, 2005. Protein identification and analysis tools on the ExPASy server. *The proteomics protocols handbook*: Springer, pp. 571-607.
- [16] K. Katoh and D. M. Standley, 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, Vol. 30, No. 4, pp. 772-80, doi: 10.1093/molbev/mst010.

- [17] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*, Vol. 35, No. 6, pp. 1547-1549, doi: 10.1093/molbev/msy096.
- [18] X. Zhu, X. Li, W. Chen, J. Chen, W. Lu, L. Chen, D. Fu, 2012. Evaluation of new reference genes in papaya for accurate transcript normalization under different experimental conditions. *PLoS One*, Vol. 7, No. 8, p. e44405, doi: 10.1371/journal.pone.0044405.
- [19] B. Chen, D. Zhong, and A. Monteiro, 2006. Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms. *BMC Genomics*, Vol. 7, No. 1, p. 156, doi: 10.1186/1471-2164-7-156.
- [20] R. D. Finn, A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, L. Holm, J. Mistry, E. L. L. Sonnhammer, J. Tate, and M. Punta, 2014. Pfam: the protein families database. *Nucleic Acids Res*, Vol. 42, No. Database issue, pp. D222-30, doi: 10.1093/nar/gkt1223.

## ABSTRACT

### Identification and analysis of *HSP90* genes in papaya (*Carica papaya* L.) by using bioinformatics method

Le Thi Man<sup>1</sup>, Nguyen Quoc Nam<sup>2</sup>, Tran Thi Thanh Huyền<sup>3</sup>, La Viet Hong<sup>4</sup>, Cao Phi Bang<sup>1</sup>

<sup>1</sup>Faculty of Natural Sciences, Hung Vuong university

<sup>2</sup>Kim Duc Secondary School, Phu Tho

<sup>3</sup>Faculty of Biology, Hanoi National University of Education

<sup>4</sup>Faculty of Biology-Agriculture, Hanoi Pedagogical University N<sup>o</sup> 2

The *HSP90* gene family has been shown to play an important role in the tolerance and development of plants. Papaya, which is a fruit crop with high nutritional value, is native to the tropics but now is widely cultivated in many subtropical regions of the world. Therefore, papaya plants have to face many environmental factors during their life. This study aims to identify and analyze the *HSP90* gene family in papaya by bioinformatics method. A total of seven *HSP90* genes have been identified in the genome of papaya (*Carica papaya* L.) by using the bioinformatic methods. The full-length genomic sequence of papaya *HSP90* genes were ranging from 2650 to 8136 nucleotides, non continuous coding, with number of intro ranging from two to 19. The predicted protein sequences included from 348 to 796 amino acids, according to the molecular weight ranged from 39.92 to 90.61 kDa. Among seven *CpHSP90*, the two *CpHSP90-1* and *CpHSP90-4* were considered pseudogenes due to their small size. These proteins were acidic with a pI value ranging from 4.69 to 5.42, except *CpHSP90-1* (pI 7.03). Based on the protein structure, subcellular localization and the phylogenetic analysis, the papaya *HSP90* were divided into two groups, I (cytoplasmic *HSP90*, four members) and II (organelle *HSP90*, three members). Analysis of transcriptomes showed that the papaya *HSP90*s were differentially expressed in different tissues at different development stages. In which, most of the papaya *HSP90* is highly expressed in flower buds or in fruits at stage 2 or stage 3. *CpHSP90-2* had the highest level of expression, followed by *CpHSP90-5*. In contrast, *CpHSP90-1* was not expressed or very weakly expressed in these studied tissues. All of seven *HSP90* genes of papaya were induced by freeze-thaw awakening treatment (in comparison with control treatment), among them, *CpHSP90-1* was strongest induced by stress (12.13-folds), however, it was a pseudogene and had a very low level of basal expression. *CpHSP90-2* had a high induction level (2.81-folds), and also had a high basal expression level compared to other *HSP90* genes of papaya. The results of this work have an important significance and will serve as a base for the further research on gene cloning, functional analysis of *HSP90* genes and breeding of papaya in response to environmental abiotic stresses and the development of this fruit crop.

**Keywords:** gene expression, *in silico*, papaya (*Carica papaya* L.), phylogenetic tree, heat shock protein 90 (*HSP90*).