

ANALYSIS OF A CORE SET OF DROUGHT AND SALT THE DROUGHT AND SALT-RESPONSIVE GENES BY EXPLORING MICROARRAY DATASETS IN CHICKPEA (*Cicer arietinum*)

Nguyen Quoc Trung¹, Tong Van Hai¹, Trinh Thi Lam Hong¹, La Viet Hong², Phan Thi Thu Hien², Tran Van Tien⁴, Chu Duc Ha^{3*}

¹Vietnam National University of Agriculture, ²Hanoi Pedagogical University 2

³VNU - University of Engineering and Technology, ⁴National Academy of Public Administration

ARTICLE INFO	ABSTRACT
Received: 27/02/2022	The purpose of this research was to define a core set of differentially expressed genes (DEGs) in roots under the drought and salt stress in chickpea (<i>Cicer arietinum</i>). By exploring all potential microarray datasets related to drought and salt stress in chickpea, a total of 41 DEGs ($ \text{fold-change} \geq 15$), including 15 highly up-regulated ($\text{fold-change} \geq 15$) and 26 highly down-regulated ($\text{fold-change} \leq -15$) genes was screened in drought- and salt- treated roots. By annotating against the assemblies of chickpea, we found that most of 41 DEG-encoded proteins were annotated as functional and regulatory proteins. Next, our results indicated that 41 DEG-encoded proteins were highly variable in physic-chemical features, like sizes, molecular weights, isoelectric points, instability index and grand average of hydropathy. Furthermore, the prediction of subcellular localization suggested that 41 proteins were localized on many major organelles, especially in the nucleus, cytoplasm and plasma membrane. Taken together, our study could provide a significant core set of DEGs related to drought and salt stress in roots of chickpea for further functional characterization.
Revised: 18/4/2022	
Published: 18/4/2022	
KEYWORDS	
Chickpea	
Drought stress	
Salt stress	
Transcriptome	
Differentially expressed gene	
Bioinformatics	

PHÂN TÍCH NHÓM GEN CHÍNH ĐÁP ỨNG VỚI STRESS HẠN VÀ MẶN Ở CÂY ĐẬU GÀ (*Cicer arietinum*) BẰNG PHÂN TÍCH DỮ LIỆU GIẢI MÃ HỆ PHIÊN MÃ

Nguyễn Quốc Trung¹, Tống Văn Hải¹, Trịnh Thị Lam Hồng¹, La Việt Hồng², Phan Thị Thu Hiền², Trần Văn Tiên⁴, Chu Đức Hà^{3*}

¹Học viện Nông nghiệp Việt Nam, ²Trường Đại học Sư phạm Hà Nội 2

³Trường Đại học Công nghệ - ĐH Quốc gia Hà Nội, ⁴Học viện Hành chính Quốc gia

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 27/02/2022	Mục tiêu của nghiên cứu này nhằm xác định các nhóm gen chính có mức độ biểu hiện khác biệt (DEG) tại mẫu rễ trong điều kiện stress hạn và mặn ở cây đậu gà (<i>Cicer arietinum</i>). Bằng cách khai thác các bộ dữ liệu microarray liên quan đến stress hạn và mặn ở đậu gà, tổng cộng 41 DEG ($ \text{fold-change} \geq 15$), bao gồm 15 DEG tăng cường biểu hiện mạnh ($\text{fold-change} \geq 15$) và 26 DEG giảm biểu hiện mạnh ($\text{fold-change} \leq -15$) đã được sàng lọc ở mẫu rễ trong điều kiện stress hạn và mặn ở cây đậu gà. Dựa trên hệ tham chiếu của cây đậu gà, phần lớn các DEG được chú giải chức năng mã hóa cho các protein điều hòa và chức năng liên quan đến cơ chế đáp ứng và chống chịu stress thâm thấu. Phân tích đặc tính lý hóa cho thấy các protein đa dạng về kích thước, trọng lượng phân tử, điểm đẳng điện, độ bất ổn định và độ ưa nước trung bình. Bên cạnh đó, 41 protein được dự đoán cư trú tại nhiều bào quan chính trong tế bào, chủ yếu là ở nhân, tế bào chất và màng sinh chất. Tóm lại, kết quả của nghiên cứu đã cung cấp thông tin về nhóm DEG quan trọng liên quan đến stress hạn và mặn ở rễ cây đậu gà định hướng cho các phân tích chức năng gen.
Ngày hoàn thiện: 18/4/2022	
Ngày đăng: 18/4/2022	
TỪ KHÓA	
Đậu gà	
Hạn	
Mặn	
Hệ phiên mã	
Gen có biểu hiện đáp ứng	
Tin sinh học	

DOI: <https://doi.org/10.34238/tnu-jst.5595>

* Corresponding author. Email: cd.ha@vnu.edu.vn

1. Giới thiệu

Đậu gà (*Cicer arietinum*) là loài quan trọng thứ ba trong các cây họ đậu được trồng phổ biến hiện nay trên thế giới. Là một trong những loại cây có nhiều mục đích sử dụng khác nhau, đậu gà được canh tác ở nhiều vùng sinh thái trên thế giới. Hạt đậu gà, chứa nhiều chất dinh dưỡng, có thể được sử dụng làm thực phẩm có giá trị cho con người, nguyên liệu chế biến thức ăn chăn nuôi và sản xuất nhiên liệu sinh học [1]. Bên cạnh đó, trồng luân phiên đậu gà, tương tự như các cây họ đậu khác, có thể giúp cải tạo đất do có khả năng cố định nitơ [2]. Tuy nhiên, canh tác cây đậu gà trên thế giới hiện nay đang gặp nhiều khó khăn do ảnh hưởng của các điều kiện bất thuận, điển hình như stress hạn và mặn [3]. Tại Việt Nam, cây đậu gà chưa được trồng phổ biến, hầu hết các nghiên cứu vẫn chỉ dừng lại ở quy mô phòng thí nghiệm và nhà kính. Do đó, tìm hiểu về cơ chế đáp ứng stress hạn và mặn ở đậu gà đang được quan tâm hiện nay, nhất là khi thông tin di truyền của cây đậu gà đã được giải mã thành công gần đây, từ đó phục vụ công tác khảo nghiệm và chọn tạo giống đậu gà thích ứng với biến đổi khí hậu tại Việt Nam [3].

Trước đây, các nghiên cứu đã được tiến hành nhằm xác định mức độ biểu hiện của một số nhóm gen thiết yếu liên quan đến đáp ứng stress mặn và hạn ở cây đậu gà. Cụ thể, các nhóm gen mã hóa protein điều hòa (như nhân tố phiên mã NAC) [4] và protein chức năng (như protein vận chuyển đường sucrose SWEET) [5] đã được xác định và phân tích mức độ phiên mã ở một số cơ quan/bộ phận chính trong điều kiện xử lý stress hạn. Trước đó, một số nhóm protein tham gia vào điều hòa sự biểu hiện gen, điển hình như nhân tố phiên mã ARF cũng được phân tích trong điều kiện stress mặn ở cây đậu gà [6]. Hầu hết các báo cáo chỉ tập trung vào vai trò của một số nhóm gen cụ thể trong điều kiện stress hạn hoặc mặn, chưa có nhiều phân tích tiến hành với dữ liệu phiên mã toàn hệ gen. Gần đây, với thành tựu của công nghệ giải trình tự hệ phiên mã, mức độ biểu hiện toàn hệ gen ở cây đậu gà trong các điều kiện xử lý stress đã lần lượt được phân tích. Theo đó, dữ liệu biểu hiện của toàn hệ gen tại lá và rễ trong điều kiện xử lý stress hạn [7]-[9] và mặn [10], [11] đã được nghiên cứu một cách riêng lẻ, từ đó phân tích chức năng của những nhóm gen có biểu hiện khác biệt (differentially expressed gene, DEG) để đưa ra cơ chế đáp ứng với stress ở cấp độ phân tử. Tuy nhiên, chưa có nhiều nghiên cứu kết hợp các dữ liệu microarray trên GEO NCBI để tìm ra danh sách DEG trong điều kiện xử lý stress mặn và hạn trên cây đậu gà, từ đó đề xuất ra các gen ứng viên định hướng cho phân tích chức năng gen [3].

Nghiên cứu này được thực hiện khai thác tất cả các dữ liệu microarray liên quan đến xử lý stress hạn và mặn ở cây đậu gà. Theo đó, nghiên cứu đã tái phân tích dữ liệu nhằm đưa ra một danh sách DEG tiềm năng có mức độ phiên mã đáp ứng mạnh. Chức năng của các gen sau đó tiếp tục được chú giải dựa trên hệ tham chiếu của cây đậu gà. Cuối cùng, tính chất và vị trí cư trú nội bào của các protein được phân tích bằng công cụ tin sinh học.

2. Vật liệu và phương pháp nghiên cứu

2.1. Vật liệu nghiên cứu

Nghiên cứu này đã thu thập tất cả dữ liệu microarray liên quan đến xử lý stress hạn và mặn ở cây đậu gà trên cơ sở GEO NCBI [12] theo mô tả trong nghiên cứu trước đây. Bên cạnh đó, hệ tham chiếu, bao gồm hệ gen và hệ protein của cây đậu gà [13] cũng đã được khai thác trên cơ sở Phytozome [14] và NCBI.

2.2. Phương pháp nghiên cứu

- Phương pháp tìm kiếm dữ liệu microarray: Tất cả các dữ liệu RNA-Seq liên quan đến xử lý stress hạn (với từ khóa 'drought') và mặn (với từ khóa 'salt') trên cây đậu gà (với từ khóa 'chickpea') được khai thác trên cổng GEO NCBI [12]. Các thông tin về bố trí thí nghiệm, mẫu thu thập và dữ liệu biểu hiện được thu thập để sàng lọc. Theo đó, các dữ liệu microarray ở cùng mẫu cơ quan/bộ phận xử lý stress hạn và mặn được sử dụng để tái phân tích.

- Phương pháp phân tích dữ liệu microarray: Mức độ biểu hiện của toàn hệ gen trong các dữ liệu RNA-Seq được tái phân tích dựa theo mô tả trong các nghiên cứu trước đây [4], [5]. Trong đó, giá trị fold-change của gen được tính theo công thức (FPKM trong xử lý stress)/(FPKM trong điều kiện đối chứng). Trong nghiên cứu này, DEG được định nghĩa là gen có giá trị |fold-change| ≥ 15 , nghĩa là gen tăng cường biểu hiện (giá trị fold-change ≥ 15) hoặc gen giảm biểu hiện (giá trị fold-change ≤ -15). Các dữ liệu sau đó được mô hình hóa theo đồ thị nhiệt bằng ngôn ngữ Python [15].

- Phương pháp chú giải chức năng gen: Mã định danh của các DEG được đối chiếu vào bản chú giải hệ gen của cây đậu gà được mô tả trong nghiên cứu trước đây [13]. Cụ thể, công PCD (Pulse Crop Database), Phytozome [14] và NCBI cung cấp hệ tham chiếu của cây đậu gà [13] được sử dụng để tìm kiếm trình tự protein đầy đủ tương ứng với từng mã định danh của DEG. Sau đó, trình tự protein đầy đủ của từng DEG đã khai thác được truy vấn trên công Pfam [16] để kiểm chứng chức năng gen.

- Phương pháp phân tích tính chất protein: Trình tự protein đầy đủ của các DEG được sử dụng để truy vấn trên công cụ ExPasy ProtParam [17] theo mô tả trong các nghiên cứu trước đây [4], [5]. Theo đó, các tính chất của protein, bao gồm kích thước (số lượng gốc amino acid), trọng lượng phân tử (kilo Dalton), điểm đẳng điện (< 7 , $= 7$ và > 7 tương ứng độ acid, trung tính và độ base), độ bất ổn định (< 41 và ≥ 41 tương ứng ổn định và bất ổn định) và độ ưa nước trung bình (< 0 và > 0 tương ứng ưa nước và kỵ nước) được phân tích [17].

- Phương pháp dự đoán vị trí cư trú nội bào protein: Trình tự protein đầy đủ của các DEG được phân tích trên công cụ Yloc [18] theo nghiên cứu trước đây [5]. Theo đó, dựa trên các tín hiệu đặc trưng cho bào quan, các protein có thể được dự đoán nằm trên 9 bào quan chính ở tế bào thực vật, bao gồm nhân, tế bào chất, ty thể, màng sinh chất, hệ thống bao gói, peroxisome, thể Golgi, không bào và lục lạp [18].

3. Kết quả và bàn luận

3.1. Phân tích dữ liệu biểu hiện microarray liên quan đến stress hạn và mặn ở cây đậu gà

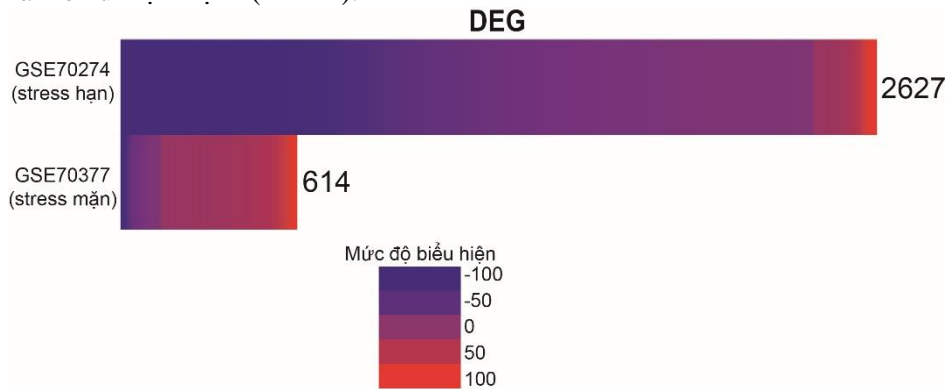
Để thu thập các dữ liệu microarray ứng viên cho phân tích, nghiên cứu này đã rà soát tất cả các dữ liệu liên quan đến stress hạn và mặn ở cây đậu gà trên công GEO NCBI [12]. Kết quả đã xác định được tổng số 10 dữ liệu, gồm 6 dữ liệu microarray giải mã hệ phiên mã ở mẫu lá, rễ, chồi và hoa xử lý hạn và 4 dữ liệu microarray phân tích ở mẫu rễ và chồi xử lý mặn (Bảng 1). Dựa trên tiêu chí chọn mẫu xử lý và phương pháp phân tích dữ liệu, nghiên cứu đã lựa chọn ra dữ liệu GSE70274 (đại diện cho stress hạn) và GSE70377 (đại diện cho stress mặn) để tái phân tích nhằm tìm ra nhóm DEG đáp ứng với cả hai điều kiện xử lý.

Bảng 1. Khai thác các dữ liệu microarray liên quan đến stress hạn và mặn ở cây đậu gà

STT	Mã dữ liệu	Điều kiện	Mẫu xử lý	Phương pháp
1	GSE104609	Hạn	Lá ở giai đoạn phân sinh đỉnh chồi	Illumina HiSeq 2000
2	GSE89228	Hạn	Lá ở giai đoạn sinh trưởng	Agilent ChickpeaGXP_8X60K chip
3	GSE70274	Hạn	Rễ ở giai đoạn sinh trưởng và sinh thực	Illumina HiSeq 2000
4	GSE12812	Hạn	Rễ ở cây 3 - 5 tuần tuổi	SAGE:26:NlaIII
5	GSE8554	Hạn	Chồi, hoa, mô rễ ở giai đoạn sinh thực	Pulse Chip
6	GSE7416	Hạn	Chồi lá, hoa và quả non ở giai đoạn sinh thực	Pulse Chip
7	GSE110127	Mặn	Rễ ở giai đoạn sinh dưỡng và sinh thực	Ion Torrent PGM
8	GSE70377	Mặn	Rễ ở giai đoạn sinh trưởng và sinh thực	Illumina HiSeq 2000
9	GSE26638	Mặn	Rễ ở giai đoạn sinh trưởng	454 GS
10	GSE7418	Mặn	Chồi và rễ ở giai đoạn sinh trưởng	Pulse Chip

Tiếp theo, phân tích các dữ liệu microarray cho thấy số lượng DEG đáp ứng ở mẫu rễ xử lý stress hạn và mặn tương đối khác biệt. Cụ thể, kết quả tái phân tích dữ liệu GSE70274 cho thấy số lượng DEG đáp ứng với điều kiện xử lý stress hạn ở mẫu rễ (giá trị |fold-change| ≥ 15) thu được là 2.627 gen, với 220 gen có mức độ biểu hiện tăng mạnh (fold-change ≥ 15) và 2.407 gen

có biểu hiện giảm mạnh (fold-change ≤ -15) (Hình 1). Kết quả tái phân tích dữ liệu GSE70377 đã chỉ ra 614 DEG ở mẫu rễ xử lý stress mặn, bao gồm 474 gen tăng cường biểu hiện mạnh và 140 gen kìm hãm biểu hiện mạnh (Hình 1).



Hình 1. Bản đồ nhiệt mô tả các DEG trong hai dữ liệu liên quan đến stress hạn và mặn ở cây đậu gà

Kết quả kết hợp 2 dữ liệu đã xác định được tổng số 41 DEG có xu hướng biểu hiện tương tự nhau trong xử lý stress hạn và mặn. Trong đó, nghiên cứu đã chỉ ra 15 DEG được tăng cường biểu hiện (fold-change ≥ 15) và 26 DEG kìm hãm biểu hiện (fold-change ≤ -15) trong điều kiện xử lý stress hạn và mặn (Bảng 1). Trong đó, *Ca_22375* được xác định là DEG có mức độ biểu hiện tăng mạnh nhất ở mẫu rễ trong stress hạn (73,96 lần), trong khi gen *Ca_23121* tăng cường phiên mã mạnh nhất ở rễ xử lý stress mặn (149,54 lần) (Bảng 2). Bên cạnh đó, *Ca_10731* và *Ca_22885* là 2 DEG kìm hãm biểu hiện mạnh nhất ở mẫu rễ trong stress hạn và mặn, với giá trị đạt lần lượt là -832,70 và -58,66 lần (Bảng 2).

Bảng 2. Mức độ biểu hiện và chú giải chức năng gen của 41 DEG đáp ứng stress hạn và mặn ở cây đậu gà

STT	Tên gen	Mức độ biểu hiện (fold)		Mã PFAM	Chú giải chức năng
		Stress hạn	Stress mặn		
1	<i>Ca_00495</i>	19,54	74,69	PF00657	GDSL-like lipase
2	<i>Ca_01016</i>	21,20	19,02	PF00067	Cytochrome P450
3	<i>Ca_02849</i>	21,59	27,20	PF01852	StAR-related lipid-transfer
4	<i>Ca_08179</i>	16,37	19,81	-	-
5	<i>Ca_09553</i>	18,43	23,18	PF08100	O-methyltransferase
6	<i>Ca_09735</i>	28,39	32,80	PF00069	Kinase
7	<i>Ca_13069</i>	19,70	124,95	PF16113	Enoyl-CoA hydratase/isomerase
8	<i>Ca_13252</i>	22,57	29,37	PF02984	Cyclin
9	<i>Ca_13953</i>	17,54	23,25	PF12638	Staygreen
10	<i>Ca_14015</i>	64,26	64,31	PF03031	NLI interacting factor-like phosphatase
11	<i>Ca_15711</i>	67,24	15,23	PF13507	CobB/CobQ-like glutamine amidotransferase
12	<i>Ca_22375</i>	73,96	31,25	PF03171	2OG-Fe(II) oxygenase
13	<i>Ca_23092</i>	16,79	16,75	PF02466	NADH dehydrogenase
14	<i>Ca_23121</i>	59,89	149,54	PF00056	Lactate dehydrogenase
15	<i>Ca_23668</i>	16,72	74,58	PF00588	SpoU rRNA methylase
16	<i>Ca_00759</i>	-22,96	-20,78	PF00808	Histone-like transcription factor
17	<i>Ca_02282</i>	-25,38	-18,56	PF05916	GINS complex protein
18	<i>Ca_02410</i>	-17,73	-35,64	PF07250	Glyoxal oxidase
19	<i>Ca_04882</i>	-140,55	-58,05	PF00078	Reverse transcriptase
20	<i>Ca_05326</i>	-115,61	-18,07	-	-
21	<i>Ca_05499</i>	-86,16	-18,75	PF03004	Transposase
22	<i>Ca_07290</i>	-63,76	-41,39	PF13923	Zinc finger, C3HC4 type
23	<i>Ca_07754</i>	-231,66	-32,89	-	-

STT	Tên gen	Mức độ biểu hiện (fold)		Mã PFAM	Chú giải chức năng
		Stress hạn	Stress mặn		
24	<i>Ca_09232</i>	-64,77	-15,77	PF00955	Bicarbonate transporter
25	<i>Ca_10731</i>	-832,70	-56,12	-	-
26	<i>Ca_11640</i>	-58,31	-29,33	PF08541	3-Oxoacyl-(acyl-carrier-protein) synthase III
27	<i>Ca_12146</i>	-30,75	-22,04	-	-
28	<i>Ca_12448</i>	-19,21	-25,70	PF01594	AI-2E family transporter
29	<i>Ca_13380</i>	-35,03	-41,81	PF01370	Epimerase
30	<i>Ca_13432</i>	-37,48	-17,41	PF00314	Thaumatococcus
31	<i>Ca_14927</i>	-88,59	-36,28	PF00909	Ammonia transporter
32	<i>Ca_16099</i>	-43,98	-36,44	PF01063	Amino-transferase nhóm IV
33	<i>Ca_16555</i>	-25,06	-18,04	PF00501	AMP-binding enzyme
34	<i>Ca_18704</i>	-37,61	-24,50	PF04564	U-box
35	<i>Ca_18859</i>	-58,09	-31,04	PF01012	Electron-transferring flavoprotein
36	<i>Ca_19509</i>	-193,35	-16,63	PF07731	Multicopper oxidase
37	<i>Ca_22396</i>	-67,61	-16,56	PF13041	PPR repeat family
38	<i>Ca_22885</i>	-25,23	-58,66	PF00098	Zinc finger
39	<i>Ca_23896</i>	-21,09	-33,68	PF03732	Retrotransposon capsid-like protein
40	<i>Ca_24594</i>	-264,49	-23,17	PF03732	Retrotransposon capsid-like protein
41	<i>Ca_26813</i>	-383,29	-29,13	PF03255	Acetyl co-enzyme A carboxylase carboxyltransferase alpha subunit

Chú thích: “-“: Chưa xác định

Trong nghiên cứu trước đây, xử lý stress hạn cũng đã gây ra sự thay đổi về mức độ biểu hiện của toàn hệ gen ở cây đậu gà một cách đa dạng. Cụ thể, tổng số 1562 DEG ($|\text{fold-change}| \geq 2$), gồm 693 gen tăng cường biểu hiện và 869 gen giảm biểu hiện đã được xác định ở lá xử lý stress hạn [7]. Kiểm tra mức độ phiên mã toàn hệ gen ở mẫu lá xử lý ngừng tưới đã chỉ ra tổng số 1078 DEG ($|\text{fold-change}| \geq 2$), gồm 403 gen tăng cường biểu hiện và 775 gen giảm biểu hiện [8]. Trong một dữ liệu về xử lý stress mặn ở mẫu rễ ở các giai đoạn, tổng số 761 DEG ($|\text{fold-change}| > 1$), bao gồm 510 gen tăng cường biểu hiện và 251 gen giảm biểu hiện và 1309 DEG (bao gồm 1000 gen tăng cường biểu hiện và 309 gen giảm biểu hiện) đã được xác định trong giai đoạn sinh trưởng sinh dưỡng và sinh thực [10].

3.2. Chú giải chức năng của các gen đáp ứng stress hạn và mặn ở cây đậu gà

Trong nghiên cứu này, chú giải chức năng của 41 DEG đáp ứng stress hạn và mặn ở cây đậu gà được tiến hành bằng cách phân tích trên hệ tham chiếu của cây đậu gà [13] khai thác trên PCD, Phytozome [14] và NCBI. Kết quả cho thấy, đa số các DEG đều mã hóa các protein điều hòa và protein chức năng quan trọng tham gia vào quá trình chống chịu stress ở cây trồng (Bảng 2). Cụ thể, 12 (trên tổng số 41) DEG được xác định mã hóa cho các nhóm protein chức năng, chủ yếu là các protein vận chuyển như *Ca_09232* mã hóa cho bicarbonate transporter (có vùng bảo thủ đặc trưng PF00955) [16], *Ca_14927* mã hóa cho ammonia transporter (có vùng bảo thủ đặc trưng PF00909) [16] (Bảng 2). Trong khi đó, phần lớn các DEG (24 trên tổng số 41) mã hóa protein điều hòa, chủ yếu là các enzyme và nhân tố phiên mã (Bảng 2). Ví dụ, *Ca_09553* và *Ca_23092* là 2 DEG lần lượt mã hóa cho hai enzyme, O-methyltransferase (với vùng bảo thủ đặc trưng PF08100) và NADH dehydrogenase (với vùng bảo thủ đặc trưng PF02466) [16]. Hai gen, *Ca_07290* và *Ca_00759* là 2 DEG được chú giải mã hóa cho các thành viên của 2 nhóm nhân tố phiên mã, bao gồm Zinc finger C3HC4 type (với vùng bảo thủ đặc trưng PF13923) và Histone-like transcription factor (với vùng bảo thủ đặc trưng PF00808) [16]. Bên cạnh đó, 5 (trên tổng số 41) DEG, bao gồm *Ca_08179*, *Ca_05326*, *Ca_07754*, *Ca_10731* và *Ca_12146* chưa được chú giải chức năng gen [16] trên hệ tham chiếu của cây đậu gà [13] (Bảng 2). Đáng chú ý, 4 gen (ngoại trừ *Ca_08179*) đều có xu hướng giảm biểu hiện ở mẫu rễ xử lý stress hạn và mặn ở

cây đậu gà (Bảng 1, 2). Như vậy, các gen này có thể được sử dụng làm nguồn vật liệu cho phân tích chức năng gen trong nghiên cứu tiếp theo.

Trước đây, các gen cảm ứng với bất lợi thâm thấu (bao gồm stress hạn và mặn) được ghi nhận là mã hóa cho 2 nhóm protein chính, bao gồm nhóm protein chức năng (protein đóng vai trò trực tiếp trong cơ chế kháng stress) và protein điều hòa (các yếu tố tham gia vào điều hòa chuyển hóa tín hiệu và biểu hiện gen chức năng trong cơ chế đáp ứng stress) [17], [18]. Trong cơ chế đáp ứng và kháng stress hạn ở cây đậu gà, hàng loạt các DEG đã được ghi nhận mã hóa cho protein liên quan đến điều hòa đóng/mở khí khổng, sửa đổi thành tế bào (như xyloglucan endotransglycosylase), vận chuyển xuyên màng và truyền tín hiệu, cân bằng nội môi, điều hòa phiên mã và sinh tổng hợp lipid [7], [8]. Ngoài ra, một số DEG đáp ứng stress hạn có thể mã hóa trong nhóm nhân tố phiên mã MYB, một số enzyme liên quan đến khử dạng ôxi nguyên tử hoạt động (như ascorbate peroxidase), một số enzyme tham gia vào cơ chế quang hợp (như carbonic anhydrase) [7], [8]. Tương tự, trong stress mặn, hầu hết các DEG đã được báo cáo mã hóa cho các protein liên quan đến cấu trúc thành tế bào, vận chuyển xuyên màng, dẫn truyền tín hiệu, đáp ứng stress ôxi hóa và điều hòa phiên mã [9]. Cụ thể, hầu hết các DEG tăng cường biểu hiện trong điều kiện stress mặn ở cây đậu gà mã hóa cho các protein vận chuyển lipid, nhân tố phiên mã NAC, protein vận chuyển đường, nhóm enzyme peroxidase, cytochrome P450 và LEA, trong khi các DEG giảm biểu hiện trong điều kiện stress mặn ở cây đậu gà mã hóa cho nhóm pectin esterase, cysteine protease, abscisic acid hydroxylase và thụ cảm glutamate [9].

3.3. Phân tích đặc tính và vị trí cư trú của protein liên quan đến stress hạn và mặn ở cây đậu gà

Để khai thác tính chất của các protein mã hóa bởi 41 DEG đáp ứng với stress hạn và mặn ở cây đậu gà, các đặc tính lý hóa và vị trí cư trú nội bào đã lần lượt được phân tích bằng các công cụ tin sinh học [17], [19] dựa theo các nghiên cứu trước đây [4], [5]. Theo đó, kết quả phân tích đặc tính lý hóa và dự đoán vị trí phân bố của protein được trình bày ở bảng 3. Các protein có kích thước dao động từ 57 (Ca_05326) đến 1002 gốc amino acid (Ca_09735 quy định enzyme kinase), tương ứng với trọng lượng phân tử dao động từ 6,25 đến 112,63 kDa (Bảng 3). Phần lớn các protein (25 trên tổng số 41) có điểm đẳng điện < 7, tương ứng với tính acid, trong khi 16 (trên tổng số 41) protein có tính base, với giá trị điểm đẳng điện > 7 (Bảng 3). Hầu hết các protein (31 trên tổng số 41) có độ ưa nước trung bình đạt giá trị âm, chứng tỏ các protein này có tính ưa nước, trong khi 10 (trên tổng số 41) protein có tính kỵ nước, với độ ưa nước trung bình có giá trị dương (Bảng 3). Ngoài ra, 21 và 20 (trên tổng số 41) protein lần lượt có độ bất ổn định < 41 và > 41, chứng tỏ các protein này lần lượt có tính ổn định và không có tính ổn định trong điều kiện ống nghiệm.

Bảng 3. Đặc tính của các protein mã hóa bởi 41 DEG đáp ứng stress hạn và mặn ở cây đậu gà

STT	Mã protein	L	mW	pI	GRAVY	II	Vị trí cư trú
1	Ca_00495	367	40,36	8,32	-0,16	28,93	Mạng lưới nội chất
2	Ca_01016	300	32,63	5,19	-0,07	33,98	Lục lạp
3	Ca_02849	265	29,92	9,02	-0,27	49,14	Tế bào chất
4	Ca_08179	308	32,78	5,58	-0,60	68,73	Nhân
5	Ca_09553	371	40,98	5,55	-0,13	28,46	Tế bào chất
6	Ca_09735	1002	112,63	9,05	-0,46	48,38	Tế bào chất
7	Ca_13069	407	46,07	5,60	-0,30	35,90	Ty thể
8	Ca_13252	361	41,42	7,94	-0,44	42,15	Nhân
9	Ca_13953	299	33,71	8,80	-0,11	27,20	Tế bào chất
10	Ca_14015	469	53,60	4,57	-0,44	51,12	Tế bào chất
11	Ca_15711	259	30,30	6,72	-0,52	45,73	Mạng lưới nội chất
12	Ca_22375	364	41,34	5,84	-0,53	42,69	Tế bào chất
13	Ca_23092	147	16,78	4,12	-0,60	39,19	Nhân
14	Ca_23121	630	69,44	6,07	-0,05	35,71	Tế bào chất
15	Ca_23668	77	8,46	7,82	-0,41	45,55	Nhân

STT	Mã protein	L	mW	pI	GRAVY	II	Vị trí cư trú
16	Ca_00759	157	17,42	6,43	-0,59	38,92	Nhân
17	Ca_02282	133	15,75	5,32	-0,56	54,89	Tế bào chất
18	Ca_02410	328	36,62	9,02	-0,66	47,36	Nhân
19	Ca_04882	525	59,90	5,92	-0,26	48,98	Tế bào chất
20	Ca_05326	57	6,25	6,37	0,21	14,84	Tế bào chất
21	Ca_05499	108	12,12	9,69	0,16	40,11	Màng sinh chất
22	Ca_07290	86	9,71	10,05	0,27	37,18	Không bào
23	Ca_07754	254	28,12	6,66	-0,27	52,92	Tế bào chất
24	Ca_09232	723	80,88	9,05	0,13	42,60	Màng sinh chất
25	Ca_10731	715	76,87	9,38	0,63	41,09	Lục lạp
26	Ca_11640	156	16,94	6,29	-0,17	21,02	Tế bào chất
27	Ca_12146	188	20,74	4,73	-0,50	48,92	Tế bào chất
28	Ca_12448	663	73,48	8,81	0,29	38,19	Tế bào chất
29	Ca_13380	344	39,04	9,16	-0,45	40,57	Nhân
30	Ca_13432	325	34,27	4,66	-0,12	50,92	Màng sinh chất
31	Ca_14927	460	49,91	5,76	0,52	27,05	Màng sinh chất
32	Ca_16099	75	8,44	4,98	-0,59	56,66	Nhân
33	Ca_16555	694	75,90	6,90	-0,03	34,22	Peroxisome
34	Ca_18704	415	45,62	7,56	0,10	47,42	Màng sinh chất
35	Ca_18859	358	37,78	6,66	0,07	36,19	Ty thể
36	Ca_19509	418	46,55	6,07	-0,10	35,99	Màng sinh chất
37	Ca_22396	384	43,20	5,02	0,05	33,18	Tế bào chất
38	Ca_22885	297	32,98	9,24	-0,96	42,63	Nhân
39	Ca_23896	218	25,44	5,41	-0,39	42,25	Thể Golgi
40	Ca_24594	967	109,11	6,03	-0,46	51,30	Nhân
41	Ca_26813	755	83,32	8,87	-0,52	40,34	Tế bào chất

Ghi chú: L - kích thước (gốc amino acid), mW - trọng lượng phân tử (kDa), pI - điểm đẳng điện, GRAVY - độ ưa nước trung bình, II - độ bất ổn định.

Tiếp theo, dự đoán vị trí cư trú nội bào đã chỉ ra rằng, các protein mã hóa bởi 41 DEG phân bố đa dạng ở các bào quan chính trong tế bào (Bảng 3). Cụ thể, 2 (trên tổng số 41) protein, Ca_00495 và Ca_15711 được dự đoán cư trú tại mạng lưới nội chất, trong khi 2 (bao gồm Ca_01016 và Ca_10731) và 2 (bao gồm Ca_13069 và Ca_18859) protein được xác định nằm lần lượt tại bào quan lục lạp và ty thể (Bảng 3). Trong khi đó, chỉ có 1 protein phân bố tại 3 bào quan, như thể Golgi (Ca_23896), peroxisome (Ca_16555) và không bào (Ca_07290) (Bảng 3). Đáng chú ý, các protein còn lại được dự đoán nằm rải rác tại màng sinh chất (6 trên tổng số 41), nhân (10 trên tổng số 41) và trong tế bào chất (16 trên tổng số 41) (Bảng 3).

4. Kết luận

Khai thác các dữ liệu microarray đã sàng lọc được 2.627 và 614 DEG đáp ứng với stress hạn và mặn (giá trị $|\text{fold-change}| \geq 15$) ở rễ của cây đậu gà. Trong đó, 15 và 26 DEG tăng cường và giảm biểu hiện trong cả stress hạn và mặn ở rễ đã được xác định.

Chú giải chức năng gen cho thấy 12 DEG mã hóa cho protein chức năng, 24 DEG mã hóa cho protein điều hòa và 5 DEG chưa được chú giải chức năng. Các protein này chủ yếu là các enzyme, nhân tố phiên mã và protein vận chuyển, tham gia trực tiếp vào cơ chế chống chịu stress thẩm thấu trong tế bào.

Đánh giá đặc tính cho thấy các protein có kích thước từ 57 (6,25 kDa) đến 1002 gốc amino acid (112,63 kDa). 25 protein có tính acid và 16 protein có tính base. 31 protein có tính ưa nước và 10 protein có tính kỵ nước. Trong đó, 16 protein được dự đoán cư trú tại tế bào chất, 10 protein nằm trong nhân tế bào và 6 protein nằm trên màng sinh chất, các protein còn lại nằm rải rác tại các bào quan chính trong tế bào.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] L. Yegrem, "Nutritional composition, antinutritional factors, and utilization trends of Ethiopian chickpea (*Cicer arietinum* L.)," *Int J Food Sci*, vol. 2021, p. 5570753, 2021.
- [2] N. Esfahani, S. Sulieman, J. Schulze, K. Yamaguchi-Shinozaki, K. Shinozaki, and L. S. Tran, "Mechanisms of physiological adjustment of N₂ fixation in *Cicer arietinum* L. (chickpea) during early stages of water deficit: single or multi-factor controls," *Plant J*, vol. 79, no. 6, pp. 964-980, 2014.
- [3] A. Rani, P. Devi, U. Jha, K. D. Sharma, K. Siddique, and H. Nayyar, "Developing climate-resilient chickpea involving physiological and molecular approaches with a focus on temperature and drought stresses," *Front Plant Sci*, vol. 10, p. 1759, 2020.
- [4] C. V. Ha, N. Esfahani, Y. Watanabe, U. T. Tran, and S. Sulieman, "Genome-wide identification and expression analysis of the *CaNAC* family members in chickpea during development, dehydration and ABA treatments," *Plos ONE*, vol. 9, no. 12, p. e114107, 2014.
- [5] V. H. La, D. H. Chu, D. C. Tran, H. K. Nguyen, T. Q. Le, M. C. Hoang, P. B. Cao, A. C. Pham, D. B. Nguyen, Q. T. Nguyen, V. L. Nguyen, V. C. Ha, T. H. Le, H. H. Le, D. T. Le, and L. S. Tran, "Insights into the gene and protein structures of the CaSWEET family members in chickpea (*Cicer arietinum*), and their gene expression patterns in different organs under various stress and abscisic acid treatments," *Gene*, vol. 819, p. 146210, 2022.
- [6] V. K. Singh, M. S. Rajkumar, and R. Garg, "Genome-wide identification and co-expression network analysis provide insights into the roles of auxin response factor gene family in chickpea," *Sci Rep*, vol. 7, p. 10895, 2017.
- [7] S. Badhan, P. Kole, A. Ball, and N. Mantri, "RNA sequencing of leaf tissues from two contrasting chickpea genotypes reveals mechanisms for drought tolerance," *Plant Physiol Biochem*, vol. 129, pp. 295-304, 2018.
- [8] R. Sinha, A. Gupta, and M. Senthil-Kumar, "Concurrent drought stress and vascular pathogen infection induce common and distinct transcriptomic responses in chickpea," *Front Plant Sci*, vol. 8, p. 333, 2017.
- [9] C. Molina, B. Rotter, and R. Horres, "SuperSAGE: the drought stress-responsive transcriptome of chickpea roots," *BMC Genomics*, vol. 9, p. 553, 2008.
- [10] M. Kaashyap, R. Ford, H. Kudapa, and M. Jain, "Differential regulation of genes involved in root morphogenesis and cell wall modification is associated with salinity tolerance in chickpea," *Sci Rep*, vol. 8, no. 1, p. 4855, 2018.
- [11] C. Molina, M. Zaman-Allah, F. Khan, and N. Fatnassi, "The salt-responsive transcriptome of chickpea roots and nodules via deepSuperSAGE," *BMC Plant Biol*, vol. 11, p. 31, 2011.
- [12] T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis, and A. Soboleva, "NCBI GEO: archive for functional genomics data sets – update," *Nucleic Acids Res*, vol. 41, pp. D991-D995, 2013.
- [13] M. Jain, G. Misra, R. K. Patel, P. Priya, S. Jhanwar, A. W. Khan, N. Shah, V. K. Singh, R. Garg, G. Jeena, M. Yadav, C. Kant, P. Sharma, G. Yadav, S. Bhatia, A. K. Tyagi, and D. Chattopadhyay, "A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.)," *Plant J*, vol. 74, no. 5, pp. 715-729, 2013.
- [14] D. M. Goodstein, S. Shu, R. Howson, R. Neupane, R. D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D. S. Rokhsar, "Phytozome: A comparative platform for green plant genomics," *Nucleic Acids Res*, vol. 40, pp. D1178-D1186, 2012.
- [15] Y. Liao, G. K. Smyth, and W. Shi, "The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads," *Nucleic Acids Res*, vol. 47, p. e47, 2019.
- [16] S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C. Tosatto, and R. D. Finn, "The Pfam protein families database in 2019," *Nucleic Acids Res*, vol. 47, p. gky995, 2018.
- [17] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, and A. Bairoch, "Protein identification and analysis tools on the ExPASy Server," In John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press, pp. 571-607, 2005.
- [18] K. Shinozaki and K. Yamaguchi-Shinozaki, "Gene networks involved in drought stress response and tolerance," *J Exp Bot*, vol. 58, no. 2, pp. 221-227, 2007.
- [19] S. Briesemeister, J. Rahnenführer, and O. Kohlbacher, "YLoc - an interpretable web server for predicting subcellular localization," *Nucleic Acids Res*, vol. 38, pp. W497-W502, 2010.