

NGHIÊN CỨU MỨC ĐỘ BIỂU HIỆN CỦA CÁC GEN LẬP TRONG HỌ NHÂN TỐ PHIÊN MÃ NAC LIÊN QUAN ĐẾN ĐÁP ỨNG HẠN Ở CÂY LẠC (*Arachis hypogaea*)

Trần Thị Thanh Huyền¹, Chu Đức Hà², Trần Đăng Khoa², La Việt Hồng³,
Hoàng Phương Loan⁴, Trần Văn Tiến⁵, Bùi Thị Thu Hương⁴ và Đồng Huy Giới⁴

¹Khoa Sinh học, Trường Đại học Sư phạm Hà Nội

²Khoa Công nghệ Nông nghiệp, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

³Viện Nghiên cứu Khoa học và Ứng dụng, Trường Đại học Sư phạm Hà Nội 2

⁴Khoa Công nghệ Sinh học, Học viện Nông nghiệp Việt Nam

⁵Học viện Hành chính Quốc gia

Tóm tắt. NAC là một trong những nhóm nhân tố phiên mã (transcription factor, TF) đặc trưng có số lượng thành viên lớn nhất ở thực vật. Ở loài lạc (*Arachis hypogaea*), 29 gen *AhNAC* đã được xác định và phân tích đặc tính cấu trúc. Tuy nhiên, chức năng của những gen này liên quan đến cơ chế chống bất lợi vẫn chưa được đề cập. Trong nghiên cứu này, tổng số bảy sự kiện lập gen, bắt nguồn từ hiện tượng lập đoạn trên các nhiễm sắc thể đã được dự đoán trong họ gen *AhNAC* ở lạc. Phân tích tỉ lệ Ka/Ks cho thấy chọn lọc tự nhiên đã kim hãm sự xuất hiện của các đột biến điểm ở phần lớn các gen lập, từ đó bảo tồn cấu trúc và chức năng của gen trong tiến hóa. Dựa trên sơ đồ hình cây giữa trình tự protein đầy đủ của NAC ở lạc và 23 NAC đáp ứng hạn ở đậu gà (*Cicer arietinum*) đã được xây dựng, 12 *AhNAC* đã được đề xuất chia sẻ chức năng tương tự như các TF NAC ở đậu gà. Hơn nữa, vùng promoter của các gen này được ghi nhận có nhiều nhóm yếu tố điều hòa *cis*- đáp ứng tín hiệu hormone và đáp ứng hạn. Khai thác dữ liệu RNA-Seq đã chứng minh đa số các gen đề xuất có biểu hiện đáp ứng trong điều kiện bất lợi thâm thấu. Kết quả của nghiên cứu này đã cung cấp những dẫn liệu quan trọng cho phân tích chức năng gen nhằm nâng cao tính chống chịu bất lợi ở cây lạc.

Từ khóa: lạc, tin sinh học, nhân tố phiên mã, NAC.

1. Mở đầu

Lạc (*Arachis hypogaea*) là một trong những cây họ đậu quan trọng nhất và được trồng phổ biến và rộng rãi ở hầu hết các vùng nhiệt đới và cận nhiệt đới. Đây là nguồn cung cấp protein, axit béo không bão hòa đơn, một số chất chống ôxi hóa và các hợp chất trao đổi thứ cấp khác [1, 2]. Hơn nữa, trồng lạc cũng được xem là biện pháp luân canh và xen canh hiệu quả nhằm sản xuất sinh khối làm vật liệu che phủ bề mặt đất, cải tạo đất nhờ cơ chế cố định đạm, đồng thời tăng thêm thu nhập cho nông hộ [3]. Tuy nhiên, canh tác lạc hiện nay đang chịu ảnh hưởng bất lợi từ các điều kiện ngoại cảnh bất thuận, đặc biệt là sự thiếu hụt về nguồn nước [4].

Đề đáp ứng với sự thay đổi của môi trường, cây trồng nói chung và loài *A. hypogaea* nói riêng đã thích nghi với cơ chế điều hòa sự biểu hiện của hàng loạt gen chức năng thông qua vai trò của các nhóm nhân tố phiên mã (TF) [5]. Trong đó, TF NAC là một trong những nhóm protein điều hòa đặc

Ngày nhận bài: 18/5/2021. Ngày sửa bài: 21/3/2022. Ngày nhận đăng: 28/3/2022.

Tác giả liên hệ: Trần Thị Thanh Huyền. Địa chỉ e-mail: tranthanhhuuyen@hnue.edu.vn

trong ở thực vật có số lượng gen thành viên nhiều nhất, được chứng minh là tham gia vào các quá trình sinh học quan trọng diễn ra ở cây trồng [6]. Nghiên cứu đã chứng minh rằng một số gen thành viên của nhóm TF NAC ở loài đậu gà (*Cicer arietinum*) có mức độ biểu hiện đáp ứng mạnh với xử lý hạn [6, 7]. Trong khi đó, nghiên cứu gần đây đã xác định được tổng số 29 gen mã hóa TF NAC (*AhNAC*) phân bố rải rác ở hệ gen của lạc [8]. Những kết quả này đã gợi mở ra hướng nghiên cứu nhằm làm rõ về cơ chế phát sinh của họ gen mã hóa TF NAC và vai trò của những gen lặp liên quan đến đáp ứng hạn ở cây lạc.

Nghiên cứu này được thực hiện nhằm chỉ ra những sự kiện lặp gen xảy ra ở họ gen mã hóa TF NAC ở lạc. Dựa trên các phân tích tin sinh học, giả thuyết về vai trò của chọn lọc tự nhiên đến cơ chế nhân rộng của họ gen mã hóa TF NAC ở lạc đã được làm rõ. Bên cạnh đó, chức năng của gen mã hóa TF NAC liên quan đến đáp ứng hạn đã được đề xuất dựa trên sơ đồ hình cây và đánh giá vùng promoter. Cuối cùng, dữ liệu biểu hiện phiên mã của các gen ứng viên đã được khai thác trong điều kiện xử lý hạn.

2. Nội dung nghiên cứu

2.1. Dữ liệu nghiên cứu

Hệ gen, hệ protein của giống lạc mô hình Tifrunner (mã số dự án: PRJNA419393) [9] được khai thác trên PeanutBase [10] và NCBI.

Trình tự protein đầy đủ của 23 TF NAC đáp ứng hạn ở loài *C. arietinum* được khai thác dựa theo nghiên cứu gần đây [7]. Thông tin chú giải, bao gồm mã định danh, trình tự gen (CDS), trình tự protein của 29 gen mã hóa TF NAC ở lạc được khai thác dựa trên nghiên cứu trước đây [8].

2.2. Phương pháp nghiên cứu

Phương pháp phân tích mức độ tương đồng: Các trình tự CDS của gen *AhNAC* (.fasta) được sử dụng để căn trình tự tương đồng trên ClustalX [11]. Kết quả được tiếp tục phân tích trên BioEDIT [12] để khai thác mức độ tương đồng (%) giữa các thành viên trong họ gen *AhNAC* ở lạc.

Phương pháp xác định sự kiện lặp gen: Một sự kiện lặp xảy ra trong họ gen *AhNAC* ở lạc được định nghĩa là một cặp gen có mức độ tương đồng trong cấu trúc lớn hơn 60%, tương tự như mô tả trong nghiên cứu trước đây [7, 13]. Trong đó, nếu các gen lặp phân bố trong cùng một vùng 20 kb trên nhiễm sắc thể được hiểu là lặp đoạn nhiễm sắc thể (tandem duplication), trong khi các gen lặp phân bố trên các nhiễm sắc thể khác nhau được hiểu là lặp đoạn trên các nhiễm sắc thể (segmental duplication) [13, 14].

Phương pháp tính toán giá trị thay thế đồng nghĩa và trái nghĩa: Trị số thay thế trái nghĩa K_a (non-synonymous substitutions per non-synonymous site) và thay thế đồng nghĩa K_s (synonymous substitutions per synonymous site) giữa cặp gen lặp được phân tích trên DNAsp [15] dựa trên trình tự CDS và protein đầy đủ theo mô tả trong nghiên cứu trước đây [13, 14].

Phương pháp xây dựng sơ đồ hình cây: Trình tự protein đầy đủ của *AhNAC* ở *A. hypogaea* và *CaNAC* ở *C. arietinum* [7] được sử dụng để thiết lập cây phân loại Neighbor-Joining với giá trị bootstrap 1000. Giá trị cut-off tại các nhánh được kiến nghị là 50% dựa theo nghiên cứu trước đây [13, 14].

Phương pháp dự đoán yếu tố điều hòa *cis*- trên vùng promoter: Đoạn 1000 bp (trước mã mở đầu 'ATG') trên vùng promoter của gen *AhNAC* được khai thác trong hệ gen của giống lạc Tifrunner [9] trên NCBI. Các yếu tố đáp ứng tín hiệu hormone, bao gồm yếu tố đáp ứng với axit abscisic (ABA) ('ACGTG'), yếu tố đáp ứng axit jasmonic (JA) (CGTCA-motif, 'CGTCA' và TGACG-motif, 'TGACG'), yếu tố đáp ứng gibberellin (GARE-motif, 'TCTGTTG' và hộp P, 'CCTTTTG'), yếu tố đáp ứng auxin (TGA-element, 'AACGAC' và AuxRR-core, 'GGTCCAT'),

yếu tố đáp ứng ethylene ('ATTTTAAA'), thu thập trong các nghiên cứu trước đây [13, 14, 16], được xác định trên vùng promoter bằng công cụ PlantCARE [17]. Sáu yếu tố đáp ứng hạn, bao gồm yếu tố đáp ứng mất nước (DRE, 'GCCGAC' hoặc 'ACCGAC'), vị trí nhận biết TF MYB (MYBRS, 'CAACTG' hoặc 'TAACTG'), vị trí nhận biết TF MYC (MYCRS, 'CACATG'), coupling element 3 (CE3, 'CACGCG'), hộp T/G ('CACGTT') và evening element (EE, 'AATATC') khai thác từ nghiên cứu trước đây [13, 14, 16] được sàng lọc trên toàn bộ promoter của gen bằng BioEDIT [11].

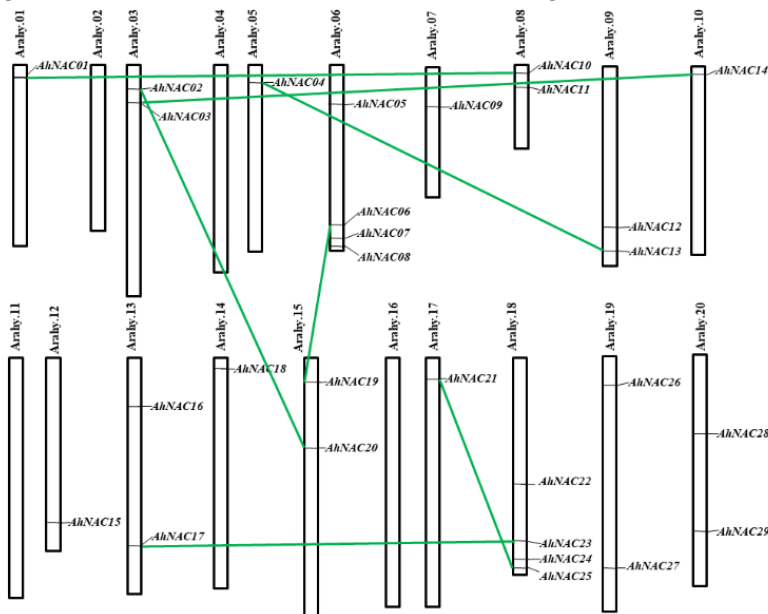
Phương pháp khai thác dữ liệu RNA-Seq: Dữ liệu xử lý cây lạc trong điều kiện bất lợi về áp suất thẩm thấu (mã số dự án: GSE56439) được khai thác trên GEO NCBI dựa theo nghiên cứu trước đây [18]. Cụ thể, cây lạc hai tuần tuổi được xử lý với nước bổ sung 30% polyethylene glycol (PEG) 600 trong 30 phút. Các mẫu được thu thập để tách ARN tổng số phục vụ phân tích hệ phiên mã bằng Illumina HiSeq2000 [18]. Số liệu được khai thác và phân tích bằng thuật toán geomean.

2.3. Kết quả và thảo luận

2.3.1. Phân tích hiện tượng lặp xảy ra ở họ gen mã hóa TF NAC ở lạc

Để làm rõ về cơ chế nhân rộng của họ mã hóa TF NAC ở lạc trong quá trình tiến hóa, các sự kiện lặp gen đã được dự đoán dựa trên mức độ tương đồng giữa các gen thành viên. Kết quả được minh họa ở Hình 1 và Bảng 1. Theo đó, tổng số bảy cặp gen lặp đã được dự đoán trong họ gen *AhNAC* ở lạc. Các cặp gen lặp chia sẻ mức độ tương đồng trong cấu trúc đoạn CDS từ 60,50 (*AhNAC21/25*) đến 70,70% (*AhNAC04/13*) (Bảng 1).

Đáng chú ý, các cặp gen lặp này đều được phân bố ở những nhiễm sắc thể khác nhau trong hệ gen của lạc (Hình 1, Bảng 1). Cụ thể, ba cặp gen lặp, *AhNAC01/10*, *02/20* và *03/14* lần lượt cư trú trên các cặp nhiễm sắc thể Arahy.01/08, 03/15 và 03/10, trong khi các cặp gen lặp còn lại, *AhNAC04/13*, *06/19*, *17/23* và *21/25* cũng lần lượt nằm trên các cặp nhiễm sắc thể khác nhau, Arahy.05/09, 06/15, 13/18 và 17/18 (Hình 1, Bảng 1). Kết quả này cho thấy tất cả các sự kiện lặp xảy ra trong họ gen *AhNAC* ở lạc đều xuất phát từ hiện tượng lặp đoạn trên các nhiễm sắc thể khác nhau. Nghiên cứu này cũng đồng thuận với các nghiên cứu trước đây khi xem xét về cơ chế lặp xảy ra ở họ gen mã hóa TF NAC ở loài *C. arietinum* [7] cũng như các loài thực vật khác [6].



Hình 1. Vị trí phân bố của các cặp gen lặp của họ *AhNAC* trên hệ gen của lạc

Bảng 1. Thông tin về các cặp gen lặp ở họ AhNAC ở lạc

Cặp gen lặp	Vị trí phân bố	Cơ chế lặp	Mức độ tương đồng (%)	Giá trị Ka	Giá trị Ks	Tỉ lệ Ka/Ks
AhNAC01/10	Arahy.01/08	S	61,40	0,4373	0,4955	0,88
AhNAC02/20	Arahy.03/15	S	66,70	0,3140	0,3984	0,79
AhNAC03/14	Arahy.03/10	S	63,40	0,3117	0,3321	0,94
AhNAC04/13	Arahy.05/09	S	70,70	0,2935	0,3435	0,85
AhNAC06/19	Arahy.06/15	S	69,50	0,3035	0,2708	1,12
AhNAC17/23	Arahy.13/18	S	67,20	0,3097	0,4236	0,73
AhNAC21/25	Arahy.17/18	S	60,50	0,4780	0,4494	1,06

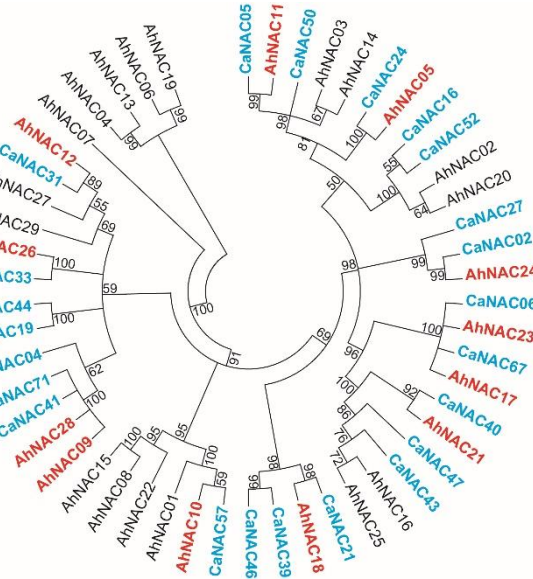
Ghi chú: S - Lặp đoạn trên các nhiễm sắc thể, Ka - Giá trị thay thế trái nghĩa, Ks - Giá trị thay thế đồng nghĩa

Tiếp theo, tỉ lệ Ka/Ks của các sự kiện lặp gen được tính toán nhằm đưa ra giả thuyết về vai trò của áp lực chọn lọc tự nhiên đến sự nhân rộng của họ gen AhNAC ở lạc trong quá trình tiến hóa. Kết quả cho thấy phần lớn các cặp gen lặp (5/7) có tỉ lệ Ka/Ks nhỏ hơn 1 (Bảng 1), cho thấy sự sai khác trong cấu trúc (thể hiện ở số lượng đột biến trái nghĩa) xảy ra ít hơn ở các cặp gen lặp. Trong khi đó, hai cặp gen lặp, AhNAC06/19 và 21/25 có tỉ lệ Ka/Ks lớn hơn 1 (Bảng 1), nghĩa là số lượng đột biến trái nghĩa xảy ra nhiều hơn, chứng tỏ các cặp gen lặp có sự sai khác đáng kể trong cấu trúc. Nghiên cứu trước đây cho thấy sự sai khác trong cấu trúc của các cặp gen lặp là động lực để hình thành chức năng gen mới, trong khi các cặp gen lặp có sự tương đồng cao trong cấu trúc được giữ toàn vẹn về mặt chức năng [14].

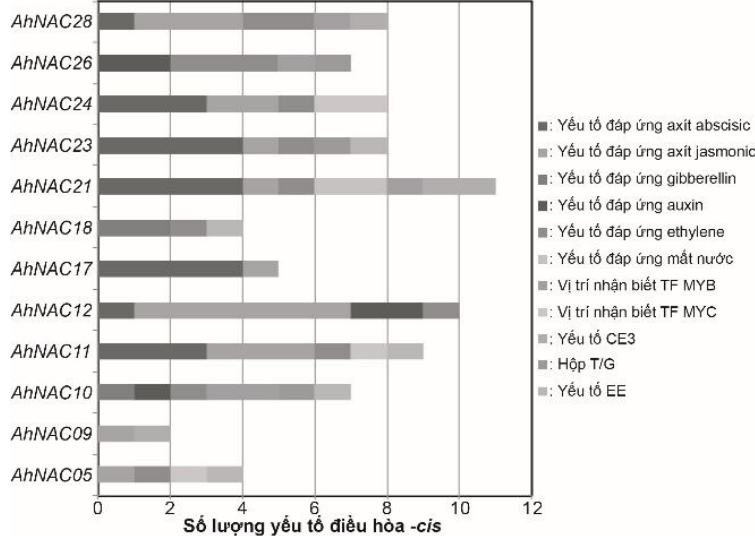
2.3.2. Phân tích mối tương quan giữa TF NAC ở lạc và các NAC đáp ứng hạn ở đậu gà và khai thác yếu tố điều hòa *cis*- trên vùng promoter của gen

Để củng cố giả thuyết về chức năng của các gen mã hóa TF NAC liên quan đến đáp ứng hạn ở lạc, sơ đồ hình cây đã được xây dựng dựa trên trình tự protein đầy đủ của 23 CaNAC đã biết chức năng ở *C. arietinum* [7] và 29 AhNAC ở lạc [8] (Hình 2). Dựa theo giá trị cut-off 50%, tổng số 12 AhNAC thành viên đã được xác định nằm cùng nhánh với CaNAC đáp ứng với hạn [7] (Hình 2). Cụ thể, giá trị bootstrap của các nhánh giữa AhNAC và CaNAC dao động từ 59 (CaNAC57 và AhNAC10) đến 100 (CaNAC33 và AhNAC26, CaNAC41/71 và AhNAC09/28, CaNAC06/67 và AhNAC17/23, CaNAC24 và AhNAC05) (Hình 2). Gần đây, phương pháp thiết lập sơ đồ hình cây giữa các protein đã biết chức năng với các protein mục tiêu theo giá trị bootstrap đã được ghi nhận trên nhóm TF NF-Y trên sắn [19] và đậu gà [14], từ đó cung cấp những dẫn liệu quan trọng cho phân tích thực nghiệm kiểm chứng lại các giả thuyết về vai trò của các gen mục tiêu trong đáp ứng bất lợi.

Để tìm hiểu về chức năng của 12 gen AhNAC này, các yếu tố điều hòa *cis*- đáp ứng hạn và tín hiệu hormone [13, 14, 16] đã được xác định trên đoạn trình tự 1000 bp thuộc vùng promoter. Theo đó, một yếu tố đáp ứng ABA, hai yếu tố đáp ứng JA, hai yếu tố đáp ứng gibberellin, hai yếu tố đáp ứng auxin, một yếu tố đáp ứng ethylene và tổng số sáu yếu tố điều hòa *cis*- liên quan đến đáp ứng hạn [13, 14, 16] đã được phân tích (Hình 3). Kết quả cho thấy, vùng promoter của tất cả 12 gen AhNAC đều chứa ít nhất một nhóm yếu tố đáp ứng với hormone (Hình 3). Trong đó, vùng promoter của gen AhNAC09 chỉ chứa một yếu tố đáp ứng với JA, trong khi có bốn nhóm yếu tố điều hòa, đáp ứng ABA, JA, auxin và ethylene (Hình 3). Đáng chú ý, các yếu tố đáp ứng ABA đã được tìm thấy trên vùng promoter của đa số các gen AhNAC, tỉ lệ trung bình đạt xấp xỉ 1,67 yếu tố/gen. Về mặt lí thuyết, ABA là hormone được chứng minh đóng vai trò chủ đạo trong cơ chế chống chịu và thích nghi với điều kiện hạn ở cây trồng [5, 6].



Hình 2. Sơ đồ hình cây giữa họ TF NAC ở lạc và 23 CaNAC đáp ứng hạn ở *C. arietinum*

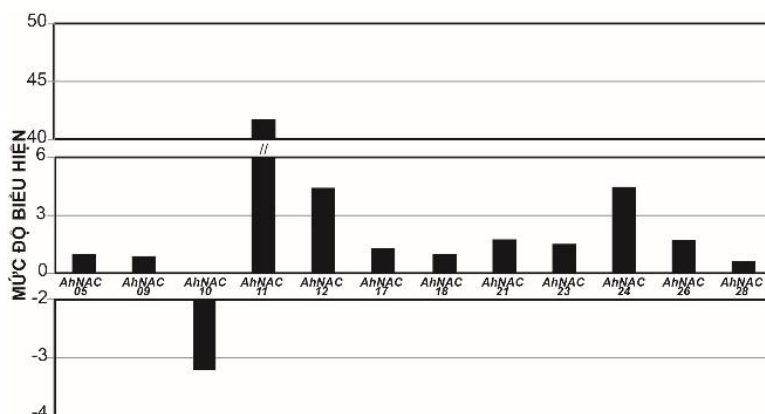


Hình 3. Sự phân bố của yếu tố điều hòa cis- trên vùng promoter của các gen AhNAC ở lạc

Trong nghiên cứu này, các yếu tố điều hòa cis- đáp ứng với hạn [13, 14, 16] đã được tập trung khai thác trên vùng promoter của 12 gen AhNAC (Hình 3). Kết quả cho thấy phần lớn các gen (¹⁰/₁₂) đều chứa ít nhất một yếu tố đáp ứng hạn, ngoại trừ hai gen, AhNAC12 và 17 không chứa yếu tố điều hòa cis- đáp ứng hạn (Hình 3). Vùng promoter của gen AhNAC21 và 10 chứa nhiều yếu tố điều hòa đáp ứng hạn nhất, lần lượt bao gồm hai DRE, một MYBRS và hai yếu tố CE3, và hai MYBRS, một hộp T/G và một yếu tố EE (Hình 3). Trong số các yếu tố đáp ứng hạn, MYBRS và yếu tố EE xuất hiện nhiều nhất (5) trên vùng promoter ở các gen, lần lượt tại AhNAC05 (1 EE), 10 (2 MYBRS và 1 EE), 11 (1 EE), 18 (1 EE), 21 (1 MYBRS), 23 (1 EE), 26 (1 MYBRS) và 28 (1 MYBRS) (Hình 3). Sự phân bố dày đặc của các yếu tố điều hòa cis- đáp ứng hạn này cũng đồng thuận với những phân tích vùng promoter của các nhóm gen mã hóa TF NF-Y liên quan đến đặc tính chống chịu bất lợi ở các cây trồng khác như mô tả trong nghiên cứu trước đây [14, 19].

2.3.3. Khai thác mức độ biểu hiện của các gen mã hóa TF NAC có tiềm năng đáp ứng hạn

Để nghiên cứu chức năng của 12 gen mã hóa TF NAC liên quan đến tính chống chịu ở lạc, dữ liệu RNA-Seq đã được khai thác nhằm phân tích mức độ biểu hiện của gen trong điều kiện xử lý PEG 600 [18]. Trong nghiên cứu này, một gen có mức độ biểu hiện thay đổi ít nhất 1,5-fold được xem là đáp ứng có ý nghĩa (tăng cường biểu hiện với giá trị fold-change $\geq 1,5$ và kim hãm biểu hiện với giá trị fold-change $\leq -1,5$) trong điều kiện xử lý. Kết quả được minh họa ở Hình 4.



Hình 4. Dữ liệu biểu hiện của 12 gen AhNAC trong điều kiện xử lý PEG 600 ở lạc

Kết quả cho thấy $6/12$ gen có đáp ứng tăng, $1/12$ gen có đáp ứng giảm, trong khi $5/12$ gen AhNAC có mức độ biểu hiện không thay đổi có ý nghĩa trong điều kiện xử lý PEG 600 (Hình 4). Trong đó, AhNAC11 là gen có mức độ biểu hiện tăng mạnh nhất (~41,73-fold), trong khi AhNAC10 bị kim hãm biểu hiện nhất (~3,22-fold) trong điều kiện xử lý PEG 600 (Hình 4). Các gen AhNAC24 và 12 có mức độ biểu hiện đáp ứng tăng, lần lượt đạt ~4,46 và 4,42-fold (Hình 4). Mức độ biểu hiện của ba gen, AhNAC21, 23 và 26 tăng nhẹ, được ghi nhận là ~1,75; 1,51 và 1,70-fold (Hình 4). Trong nghiên cứu trước đây, gen CaNAC05 ở *C. arrietinum* (mã hóa protein cùng nhánh với AhNAC11) (Hình 2) được ghi nhận biểu hiện mạnh ở lá xử lý mất nước [7].

3. Kết luận

Đã xác định được bảy cặp gen lặp trong họ gen mã hóa TF NAC ở lạc. Các cặp gen lặp này đều xuất phát từ hiện tượng lặp đoạn trên các nhiễm sắc thể. Trong đó, $5/7$ cặp gen lặp được bảo tồn cấu trúc trong quá trình tiến hóa.

Đã xác định được 12 gen AhNAC có tiềm năng đáp ứng với điều kiện hạn thông qua xây dựng sơ đồ nhóm với các TF NAC đã biết chức năng ở loài *C. arrietinum*. Vùng promoter của các gen AhNAC gồm nhiều nhóm yếu tố cis- đáp ứng tín hiệu hormone và đáp ứng hạn.

Đã chứng minh được $10/12$ gen AhNAC có mức độ biểu hiện đáp ứng với xử lý PEG 600 dựa trên khai thác dữ liệu RNA-Seq. Gen AhNAC11 được tăng cường mức độ biểu hiện mạnh nhất (~41,73-fold) trong điều kiện xử lý PEG 600.

TÀI LIỆU THAM KHẢO

- [1] O.T. Toomer, 2018. Nutritional chemistry of the peanut (*Arachis hypogaea*). *Crit. Rev. Food Sci. Nutr.*, Vol. 58, Iss. 17, pp. 3042-3053.
- [2] X. Zhao, J. Chen, F. Du, 2012. Potential use of peanut by-products in food processing: a review. *J. Food Sci. Technol.*, Vol. 49, Iss. 5, pp. 521-529.

- [3] J. Dai, W. Qiu, N. Wang, T. Wang, H. Nakanishi, Y. Zuo, 2019. From Leguminosae/Gramineae intercropping systems to see benefits of intercropping on iron nutrition. *Front. Plant Sci.*, Vol. 10, pp. 605.
- [4] S. Bhogireddy, A. Xavier, V. Garg, *et al.*, 2020. Genome-wide transcriptome and physiological analyses provide new insights into peanut drought response mechanisms. *Sci. Rep.*, Vol. 10, Iss. 1, pp. 4071.
- [5] X.L.T. Hoang, D.N.H. Nhi, N.B.A. Thu, N.P. Thao, L.P. Tran, 2017. Transcription factors and their roles in signal transduction in plants under abiotic stresses. *Curr. Genomics*, Vol. 18, Iss. 6, pp. 483-497.
- [6] L.S. Tran LS, R. Nishiyama, K. Yamaguchi-Shinozaki, K. Shinozaki, 2010. Potential utilization of NAC transcription factors to enhance abiotic stress tolerance in plants by biotechnological approach. *GM Crops*, Vol. 1, Iss. 1, pp. 32-39.
- [7] C.V. Ha, M.N. Esfahani, Y. Watanabe, U.T. Tran, S. Sulieman, K. Mochida, D.V. Nguyen, L.S. Tran, 2014. Genome-wide identification and expression analysis of the CaNAC family members in chickpea during development, dehydration and ABA treatments. *PLoS One*, Vol. 9, Iss. 12, pp. e114107.
- [8] C.D. Ha, P.T. Linh, N.T. Hao, N.Q. Trung, L.V. Hong, P.P. Th, L.T.N. Quynh, H.T. Quyen, P.M. Trien, 2021. Structural analysis of genes encoding NAC transcription factor in peanut (*Arachis hypogaea*) by the bioinformatics approach. *Proceeding of 2021 Vietnam National Conference on Biotechnology*, pp. 543-548 (in Vietnamese).
- [9] Bertoli D.J., Jenkins J., Clevenger J. *et al.*, 2019. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nat. Genet.*, Vol. 51, pp. 877-884.
- [10] D. Sudhansu, Ethalinda K., S. Cannon, Scott R. Kalberer, Andrew D., Farmer S., B. Cannon, 2016. PeanutBase and other bioinformatic resources for peanut. In *Peanut genetics, processing, and utilization*, edited by H. Thomas Stalker and Richard F. Wilson, AOCS Press, p. 241-252.
- [11] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics*, Vol. 23, pp. 2947-2948.
- [12] Hall T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, Vol. 41, pp. 95-98.
- [13] C.D. Ha, L.X. Dac, T.T.T. Huyen, P.T.L. Thu, 2017. Evolutionary analysis and expression profiling of the SWEET sugar transporter gene family in cassava (*Manihot esculenta* Crantz). *HNUE Journal of Science*, Vol. 62, Iss. 10, pp. 91-99.
- [14] Chu H.D., Nguyen K.H., Watanabe Y., *et al.*, 2018. Identification, structural characterization and gene expression analysis of members of the Nuclear factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments. *Int. J. Mol. Sci.*, Vol. 19, Iss. 11, pp. 3290.
- [15] P. Librado, J. Rozas, 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, Vol. 25, Iss. 11, pp. 1451-1452.
- [16] Chu Duc Ha, Nguyen Thi Kim Lien, Tran Thi Thanh Huyen, Pham Thi Ly Thu, Le Tien Dung, 2017. Computational analysis of the Methionine sulfoxide reductase gene family in soybean (*Glycine max*) and their response in abiotic stresses. *HNUE Journal of Science*, Vol. 62, Iss. 10, pp. 127-133.
- [17] M. Lescot, Patrice D., Gert T., K. Marchal, Y. Moreau, Y. Van de Peer, Pierre R., S. Rombauts, 2002. PlantCARE, a database of plant cis-acting regulatory elements and a

- portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.*, Vol. 30, Iss. 1, pp. 325-327.
- [18] Li X., Lu J., Liu S., Liu X. et al., 2004. Identification of rapidly induced genes in the response of peanut (*Arachis hypogaea*) to water deficit and abscisic acid. *BMC Biotechnol.*, Vol. 14, p. 58.
- [19] C.D. Ha, L.V. Hong, L.H.T. Phuong, L.T. Thao, H.T. Thao, P.T.L. Thu, 2018. Study on structure of genes encoding Nuclear factor-YB subunit associated with drought tolerance in cassava. *Journal of Vietnam Agricultural Science and Technology*, Vol. 5, Iss. 90, pp. 5-9 (in Vietnamese).

ABSTRACT

Expression analysis of the duplicated genes in the NAC transcription factor gene family related to the drought stress response in peanuts (*Arachis hypogaea*)

Tran Thi Thanh Huyen¹, Chu Duc Ha², Tran Dang Khoa², La Viet Hong³,
Hoang Phuong Loan⁴, Tran Van Tien⁵, Bui Thi Thu Huong⁴ and Dong Huy Gioi⁴

¹*Faculty of Biology, Hanoi National University of Education*

²*Faculty of Agricultural Technology, University of Engineering Technology, VNU*

³*Institute of Science and Application, Hanoi Pedagogical University 2*

⁴*Faculty of Biotechnology, Vietnam National University of Agriculture*

⁵*National Academy of Public Administration*

NAC has been considered one of the largest transcription factor families in plant species. In peanut (*Arachis hypogaea*), 29 members of the *AhNAC* gene family have been identified and characterized. However, the information of their functions related to the mechanism of the stress response has been reported. In this study, a total of seven duplication events, which occurred from the segmental duplication, has been predicted in the *AhNAC* gene family in peanuts. The Ka/Ks ratios revealed that natural selection prevented the presence of the point mutations in the most duplicated genes, consequently, maintaining their structure and function during evolution. Based on the unrooted phylogenetic tree between *AhNAC* proteins in peanuts and 23 drought-responsive CaNAC in chickpea (*Cicer arietinum*), we proposed 12 members of the *AhNAC* that potentially share similar functions with well-characterized NAC in chickpea. Furthermore, various hormone- and drought-responsive *cis*-regulatory elements have been found in the promoter regions of these candidate genes. Re-analysis of the previous RNA-Seq demonstrated that the majority of the candidate *AhNAC* genes were responsive to the osmotic stress. Taken together, our study could provide an important foundation for further functional characterization towards the improvement of stress tolerance in peanuts.

Keywords: peanut, bioinformatics, transcription factor, NAC.