

## NGHIÊN CỨU TƯƠNG TÁC VÀ KHAI THÁC DỮ LIỆU BIỂU HIỆN GIỮA CÁC TIỂU PHẦN CỦA YẾU TỐ NHÂN Y Ở LOÀI ĐẬU GÀ (*Cicer arietinum*)

Chu Đức Hà<sup>1</sup>, Lê Thị Ngọc Quỳnh<sup>2\*</sup>

<sup>1</sup>Viện Khoa học Nông nghiệp Việt Nam, <sup>2</sup>Đại học Thủy lợi

### TÓM TẮT

Nghiên cứu về tương tác protein và mức độ biểu hiện của gen mã hóa protein đích được xem là những công cụ hỗ trợ đắc lực cho việc đưa ra giả thuyết về chức năng của gen. Trong nghiên cứu này, khả năng tương tác giữa ba tiểu phần của nhân tố phiên mã Yếu tố nhân Y (NF-Y) và mức độ biểu hiện của các gen mã hóa đã được phân tích trên cây đậu gà (*Cicer arietinum*). Kết quả mô hình hóa interactome đã cho thấy tiểu phần CaNF-YB có thể kết hợp chặt chẽ với CaNF-YC thành dạng nhị hợp, sau đó tương tác với CaNF-YA tạo thành phức hợp CaNF-Y hoàn chỉnh, tương tự như ở các loài thực vật khác. Trong đó, bốn dạng kết hợp của ba tiểu phần đã được dự đoán là [(CaNF-YB19/-YC01)/-YA06], [(CaNF-YB19/-YC01)/-YA02], [(CaNF-YB18/-YC01)/-YA06] và [(CaNF-YB18/-YC01)/-YA02]. Tiếp theo, cấu trúc vùng bảo thủ của các tiểu phần CaNF-YA có tương tác đều chứa hai vùng riêng biệt, là đoạn nhận biết, bám DNA và đoạn tương tác với cấu trúc CaNF-YB/YC. Cấu trúc vùng bảo thủ của các tiểu phần CaNF-YB và -YC có tương tác được ghi nhận với hai vùng tương tác với CaNF-YA và với CaNF-YC hoặc CaNF-YB. Khai thác dữ liệu phiên mã trước đây đã chỉ ra rằng các gen mã hóa tiểu phần CaNF-Y có tương tác đều thể hiện mức độ biểu hiện mạnh ở ít nhất một vị trí trong cây. Kết quả của nghiên cứu này đã gợi mở những dẫn liệu quan trọng cho việc định hướng nghiên cứu chức năng của các gen mã hóa tiểu phần CaNF-Y quan tâm.

**Từ khóa:** tin sinh học, tương tác protein, cấu trúc 3D, nhân tố phiên mã, yếu tố nhân Y, đậu gà

**Ngày nhận bài:** 11/12/2019; **Ngày hoàn thiện:** 31/12/2019; **Ngày đăng:** 10/01/2020

## INTERACTION AND DATA MINING OF EXPRESSION OF THE NUCLEAR FACTOR-Y SUBUNITS IN CHICKPEA (*Cicer arietinum*)

Chu Duc Ha<sup>1</sup>, Le Thi Ngoc Quynh<sup>2\*</sup>

<sup>1</sup>Vietnam Academy of Agricultural Sciences, <sup>2</sup>Thuyloi University

### ABSTRACT

Protein - protein interaction and data mining of the expression patterns of genes were considered as the major reliable tools to construct the theory of the gene functions. Here, the interactions between Nuclear factor - Y (NF-Y) subunits and the expression profiles of corresponding genes found in chickpea (*Cicer arietinum*) were performed. Based on the interactome database, we predicted that members of CaNF-YB subunits can slightly interact with members of CaNF-YC subunit to construct the heterodimer, which then merges with CaNF-YA subunit to create the CaNF-Y complex as fully reported in other plant species. Among them, four formulas of protein-protein interaction, including [(CaNF-YB19/-YC01)/-YA06], [(CaNF-YB19/-YC01)/-YA02], [(CaNF-YB18/-YC01)/-YA06] and [(CaNF-YB18/-YC01)/-YA02] were predicted. Next, the conserved domain of interacted CaNF-YA subunit consisted of two distinct sub-domains, including the DNA binding sub-domain and the CaNF-YB/YC interaction domain. The conserved domains of interacted CaNF-YB and -YC subunits were reported to harbor two regions, CaNF-YA interaction sub-domain and CaNF-YC or -YB interaction sub-domain. Of our interest, data mining of the previous transcriptome atlas revealed that genes encoding interacted CaNF-Y subunits were highly expressed in at least one organ. Taken together, our study would provide the important information for the further characterization of genes encoding CaNF-Y subunits.

**Keywords:** bioinformatics, protein interaction, three-dimension structure, transcription factor, nuclear factor-Y, chickpea

**Received:** 11/12/2019; **Revised:** 31/12/2019; **Published:** 10/01/2020

\* Corresponding author. Email: [ngoc.quynh.19051990@gmail.com](mailto:ngoc.quynh.19051990@gmail.com)

## 1. Giới thiệu

Yếu tố nhân Y (Nuclear factor - Y, NF-Y) là một trong những nhóm nhân tố phiên mã (transcription factor, TF) đặc trưng ở thực vật [1]. Về mặt cấu trúc, phức hợp TF NF-Y được hình thành từ ba tiểu phần, bao gồm NF-YA, -YB và -YC, trong đó, NF-YA chứa vùng domain đặc trưng cho tương tác với NF-YB và NF-YC và vùng bám DNA, trong khi NF-YB chứa hai domain chức năng riêng biệt, tương tác với tiểu phần NF-YA và với NF-YC, và một domain cần thiết cho quá trình bám ADN [2]. Cấu trúc của tiểu phần NF-YC được ghi nhận gồm vùng đặc trưng gồm hai đoạn tương tác với NF-YA (NF-YA interaction) và một vùng có thể liên kết nhị trùng hóa với NF-YB [1, 2].

Đáng chú ý, rất nhiều nghiên cứu đã chỉ ra rằng, ba tiểu phần này được mã hóa bởi họ đa gen ở thực vật [1], như ở lúa gạo (*Oryza sativa*) [3], đậu tương (*Glycine max*) [4] và đậu gà (*Cicer arietinum*) [5]. Vì vậy, xác định tương tác giữa những thành viên trong ba họ tiểu phần protein để tạo thành TF NF-Y được xem là rất phức tạp, đặc biệt là liên quan đến khả năng đáp ứng với điều kiện ngoại cảnh bất lợi [6, 7].

Trong nghiên cứu này, dữ liệu từ họ gen mã hóa TF NF-Y ở *C. arietinum*, CaNF-Y, đã được khai thác [5] nhằm phân tích khả năng tương tác giữa các thành viên của họ CaNF-YA, CaNF-YB và CaNF-YC. Từ đó, mức độ phiên mã của các gen tiềm năng trong điều kiện bất lợi đã được phân tích dựa trên các cơ sở dữ liệu RNA-seq trước đây.

## 2. Dữ liệu và phương pháp nghiên cứu

### 2.1. Dữ liệu nghiên cứu

Toàn bộ dữ liệu về 8, hai mươi một và 11 thành viên của lần lượt ba họ CaNF-YA, -YB và -YC ở giống đậu gà kabuli đã được khai thác và sử dụng dựa trên mô tả gần đây [5].

### 2.2. Phương pháp nghiên cứu

#### 2.2.1. Phân tích tương tác protein - protein

Tương tác protein - protein (protein-protein interaction, PPI) của họ CaNF-Y khai thác từ nghiên cứu trước đây [5] được xây dựng dựa

trên cơ sở dữ liệu STRING [8] với hệ tham chiếu của giống đậu gà "kabuli" CDC Frontier [9]. Hệ số tương tác trong mạng lưới PIP được lựa chọn ở các giá trị 0,4 (trung bình), 0,7 (cao) và 0,9 (cao nhất) tương ứng với mức độ tin cậy của phép phân tích [8].

#### 2.2.2. Căn trình tự tương đồng

Trình tự protein của các CaNF-Y có tương tác được căn trình tự tương đồng bằng công cụ Clustal X [10] để xác định vùng bảo thủ đặc trưng cho tiểu phần NF-YA, -YB và -YC ở thực vật [1, 2] trên công cụ Pfam [11].

#### 2.2.3. Xác định mức độ phiên mã dựa vào RNA-seq

Hai dữ liệu biểu hiện của các gen mã hóa tiểu phần CaNF-Y có tương tác được khai thác trên CTDB (Chickpea Transcriptome Database) [12]. Trong đó, dữ liệu phiên mã ở ba cơ quan sinh sản, bao gồm mô phân sinh đỉnh chồi (SAM), lá non (YL) và cây non đang nảy mầm (GS), và 8 mẫu mô thu thập ở các giai đoạn phát triển của nụ hoa (FB1-FB4) và hoa (FL1-FL4) đã được thu thập [12]. Đồng thời, dữ liệu phiên mã ở 5 cơ quan, bao gồm thân (S), rễ (R), lá trưởng thành (ML), nụ hoa (FB) và quả non (YP) cũng đã được khai thác [12]. Dữ liệu được phân tích và mô phỏng bằng thuật toán R.

#### 2.2.4. Xây dựng mô hình cấu trúc 3D

Cấu trúc 3D của các phân tử protein CaNF-Y có tương tác được xây dựng bằng công cụ Pyhre2 dựa trên đoạn trình tự protein tương ứng. Độ tin cậy (confidence) và độ bao phủ (coverage) của thuật toán được chọn từ giá trị cao nhất để tăng mức độ chính xác. Cấu trúc  $\alpha$  helix và chuỗi  $\beta$  được xác định để xây dựng cấu trúc bậc 2 của protein. Mô hình được biểu diễn trên công cụ Illustrator.

## 3. Kết quả và bàn luận

### 3.1. Dự đoán khả năng tương tác giữa các tiểu phần CaNF-Y ở loài đậu gà

Tương tác PIP giữa các tiểu phần CaNF-Y được phân tích dựa trên các chức năng đã biết hay tương tác vật lý được hình thành giữa các phân tử protein trên STRING [8]. Kết quả cho thấy, mạng lưới PIP được hình thành ở 13

tiểu phần CaNF-Y (trên tổng số 40 thành viên của họ CaNF-Y đã được báo cáo gần đây [5]), bao gồm hai CaNF-YA, bốn CaNF-YB, sáu CaNF-YC và một phân tử tương đồng với Dr1 (Hình 1). Trong đó, 12 cạnh (edge) và 13 nút (node) ở các tương tác PIP được ghi nhận với độ tin cậy ở mức cao nhất (lớn hơn 0,9), giá trị p trong PPI nhỏ hơn  $1,0e-16$  (Hình 1). Bốn protein, bao gồm CaNF-YB19 (XP\_004507590.1), -YC04 (XP\_004494629.1), -YC05 (XP\_004495741.1) và -YC06 (XP\_004501704.1) đã được biết rõ cấu trúc 3D như mô tả trong nghiên cứu trước đây [13] (Hình 1, 4).

Đáng chú ý, các tiểu phần CaNF-YB và -YC có khả năng tương tác qua lại (tám trên 13 cặp liên kết) nhiều hơn so với CaNF-YA và -YB (Hình 1). Điều này đã ghi nhận tương tự ở các loài thực vật khác, tiểu phần NF-YA và -YB rất ít hoặc không tương tác với nhau, trong khi NF-YB và -YC có thể tương tác tạo nên nhiều cấu trúc dị nhị tụ (heterodimer) khác nhau [14, 15]. Các kết quả này cho thấy sự tương tác giữa hai tiểu phần NF-YB và -YC xảy ra phổ biến ở *C. arietinum* và trong thực vật nói chung. Như vậy, bốn phức hợp CaNF-Y hoàn chỉnh (tạo nên từ các PIP giữa CaNF-YA, -YB và -YC) đã được xác định, bao gồm [(CaNF-YB19/-YC01)/-YA06], [(CaNF-YB19/-YC01)/-YA02], [(CaNF-YB18/-YC01)/-YA06] và [(CaNF-YB18/-YC01)/-YA02] (Hình 1).

### 3.2. Phân tích cấu trúc giữa các tiểu phần CaNF-Y có tương tác ở đậu gà

Nhằm tìm hiểu rõ hơn về khả năng tương tác giữa bốn cặp tiểu phần CaNF-Y ở đậu gà, cấu trúc vùng bảo thủ đã được xem xét và phân tích trên Pfam [11]. Kết quả cho thấy hai tiểu phần CaNF-YA02 và -YA06 có cấu trúc rất bảo thủ (Hình 2A), đồng thuận với các báo cáo trước đây về OsNF-YA ở lúa gạo [3], GmNF-YA ở đậu tương [4], cũng như ở các loài thực vật khác [1, 6, 7]. Cụ thể, vùng bảo thủ của CaNF-YA02 và -YA06 chứa trình tự đặc hiệu nhận biết hộp CCAAT trên trình tự DNA đích và trình tự tương tác với NF-YB/C (Hình 2A).

Trong khi đó, CaNF-YB18 và -YB19 chứa vùng cuộn xoắn histone (Hình 3), hình thành nên liên kết với NF-YA và tiểu phần NF-YC. Đáng chú ý, vị trí amino acid số 1 - 63 của CaNF-YB19 mô tả trong Hình 2B cũng chính là vùng bảo thủ đã được tìm thấy trong histone của vi khuẩn cổ (archaeobacteria) và các TF giống histone của sinh vật nhân thực [16]. Tương tự, vùng bảo thủ của CaNF-YC01 được xác định gồm hai đoạn riêng biệt, tương tác với NF-YA và tương tác với NF-YB (Hình 2C). Bên cạnh đó, để có cái nhìn tổng quan về cấu trúc của các CaNF-Y có tương tác, mô hình 3D của các phân tử này đã được xây dựng trên công cụ Phyre 2 (Hình 3). Tóm lại, kết quả đã chứng minh rõ hơn tính bảo thủ của vùng trong trình tự của NF-Y và vai trò quan trọng trong thực hiện chức năng tương tác nhằm đảm bảo sự điều hòa biểu hiện gen.

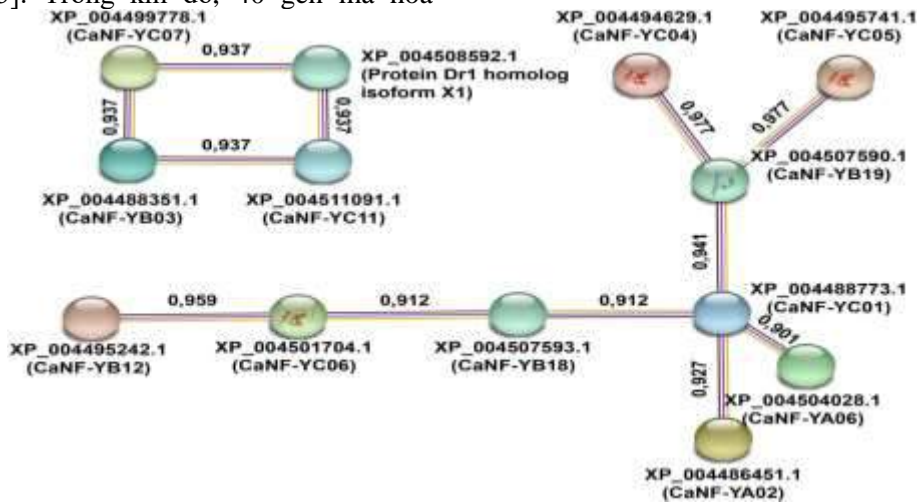
### 3.3. Đánh giá biểu hiện của tiểu phần CaNF-Y có tương tác ở các bộ phận trên cây đậu gà

Để đánh giá biểu hiện của các gen mã hóa tiểu phần CaNF-Y có tương tác, hai dữ liệu hệ phiên mã ở các cơ quan chính trên cây đậu gà đã được khai thác. Kết quả cho thấy hai gen CaNF-YB19 và -YC01 không có thông tin, trong khi gen CaNF-YA06 có mức độ biểu hiện dưới ngưỡng phát hiện (Hình 4). Xem xét trên dữ liệu phiên mã ở các mô sinh sản và ở các giai đoạn phát triển của hoa, hầu hết các gen mã hóa các tiểu phần tương tác đều có xu hướng biểu hiện mạnh ở ít nhất một bộ phận trong điều kiện thường (Hình 4). Năm gen CaNF-YA02, -B03, -C04, -C05 và -C06 có biểu hiện mạnh ở tất cả các mẫu và có xu hướng thể hiện đặc thù ở từng mẫu riêng biệt (Hình 4). Gen CaNF-YA02 và -YB03 lần lượt biểu hiện rất mạnh ở mô phân sinh đỉnh chồi (SAM) và hoa ở giai đoạn phát triển thứ 3 (FL3), trong khi CaNF-YC04 tăng cường biểu hiện ở nụ hoa giai đoạn 1 (FB1) và 2 (FB2) (Hình 4). Gen CaNF-YC05 tăng cường phiên mã ở FL1 và FL2, trong khi CaNF-YC06 có biểu hiện đặc thù ở mẫu lá non (YL) (Hình 4). Bên cạnh đó, xem xét trên hệ phiên mã ở năm mẫu mô chính cho thấy các gen mã hóa tiểu

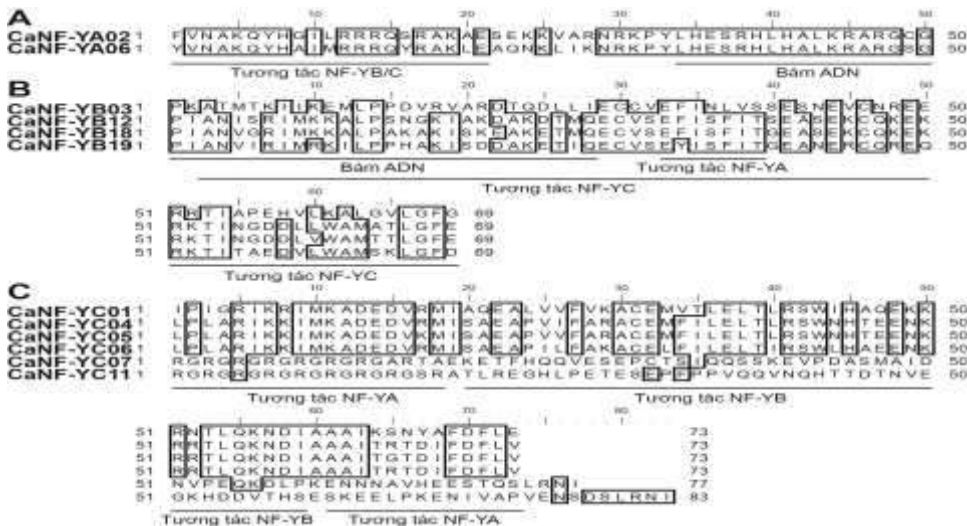
phân tương tác cũng có mức độ biểu hiện đa dạng. Cụ thể, khai thác dữ liệu cho thấy gen CaNF-YB03 và -YC11 có biểu hiện mạnh ở mô FB, trong khi gen CaNF-YC05 có xu hướng biểu hiện mạnh ở cả năm mẫu mô (Hình 4). Đáng chú ý, CaNF-YC04 biểu hiện đặc thù đồng thời ở R và FB, trong khi (Hình 4). Những kết quả này đã cung cấp những dẫn liệu quan trọng cho việc xây dựng giả thuyết về sự hình thành TF CaNF-Y ở các cơ quan trên cây đậu gà.

Ở lúa, 34 thành viên của họ OsNF-Y đã được phân lập và mạng lưới PIP giữa các tiểu phần OsNF-Y (OsNF-YA, -YB và -YC) cũng được đưa ra [3]. Trong khi đó, 40 gen mã hóa

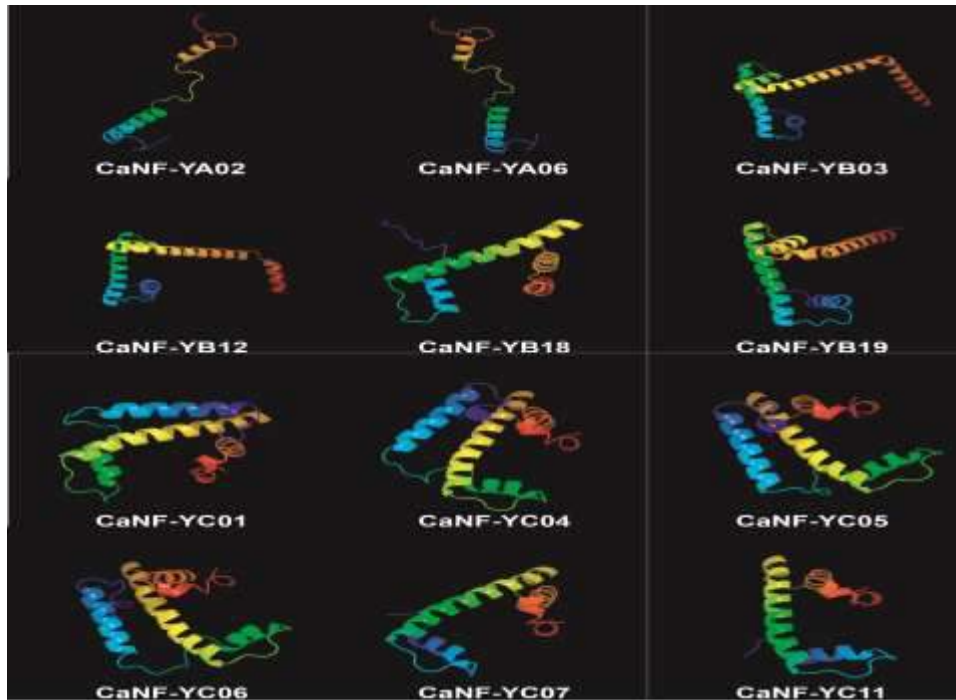
CaNF-Y (8 CaNF-YA, 21 -YB và 11 -YC) đã được tìm thấy [5]. Tuy nhiên chưa có công bố về các tương tác có thể xảy ra giữa các tiểu phần. Những kết quả thu được đã bổ sung cho nghiên cứu trước đây [5] để chỉ rõ ra bộ ba tương tác mạnh giữa từng tiểu phần CaNF-A, -YB và -YC để tạo nên phức hợp TF CaNF-Y hoàn chỉnh trong cây đậu gà. Bên cạnh đó, chúng tôi cũng đã gợi ý tương tác giữa các tiểu phần CaNF-YB và -YC là mạnh mẽ và phổ biến trong đậu gà nói riêng và có lẽ ở thực vật nói chung. Đây là những dẫn chứng quan trọng nhằm định hướng được thành viên đóng vai trò quan trọng trong họ CaNF-Y.



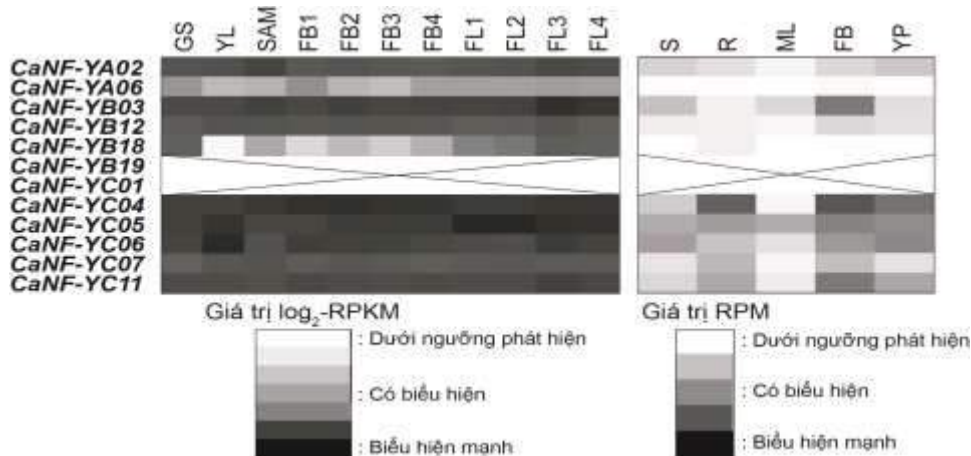
Hình 1. Mạng lưới tương tác protein - protein được hình thành từ họ CaNF-Y ở đậu gà



Hình 2. Trình tự amino acid những vùng bảo thủ trong các NF-Y được dự đoán tương tác protein-protein mạnh để tạo thành phức hợp NF-Y hoàn chỉnh. A. CaNF-YA02, CaNF-YA06; B. CaNF-YB18, CaNF-YB19; C. CaNF-YC01 và D. Mô hình phức hợp CaNF-Y hoàn chỉnh bám vào trình tự thông qua hộp CCAAT



**Hình 3.** Mô hình hóa cấu trúc 3D của các tiểu phần CaNF-Y có tương tác



**Hình 4.** Khai thác dữ liệu biểu hiện của các gen mã hóa tiểu phần CaNF-Y có tương tác ở các mẫu mô trên cây đậu gà

**4. Kết luận**

Các thành viên của hai tiểu phần CaNF-YB và -YC, (CaNF-YB19/-YC01, CaNF-YB18/-YC01) được dự đoán có tương tác với nhau. Phức hợp này sau đó kết hợp với tiểu phần CaNF-YA (CaNF-YA02, -YA06) để tạo thành cấu trúc TF NF-Y hoàn chỉnh ở đậu gà.

Cấu trúc của tiểu phần CaNF-YA có tương tác chứa vùng nhận biết DNA và vùng tương tác với cấu trúc NF-YB/YC. Cấu trúc của tiểu

phần CaNF-YB và CaNF-YC có tương tác được xác định gồm hai vùng riêng biệt nhằm tương tác với nhau và tương tác với NF-YA.

Khai thác dữ liệu biểu hiện cho thấy hầu hết các gen mã hóa tiểu phần CaNF-Y có tương tác đều có xu hướng biểu hiện mạnh ở ít nhất một cơ quan trong điều kiện thường.

**TÀI LIỆU THAM KHẢO/ REFERENCES**

[1]. M. E. Zanetti, C. Ripodas and A. Niebel, "Plant NF-Y transcription factors: Key players in

- plant-microbe interactions, root development and adaptation to stress," *Biochim Biophys Acta*, vol. 1860, no. 5, pp. 645-654, 2017.
- [2]. R. Mantovani, "The molecular biology of the CCAAT-binding factor NF-Y," *Gene*, vol. 239, no. 1, pp. 15-27, 1999.
- [3]. W. Yang, Z. Lu, Y. Xiong and J. Yao, "Genome-wide identification and co-expression network analysis of the *OsNF-Y* gene family in rice," *Crop J.*, vol. 5, no. 1, pp. 21-31, 2017.
- [4]. T. N. Quach, H. T. Nguyen, B. Valliyodan, T. Joshi, D. Xu and H. T. Nguyen, "Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response," *Mol Genet Genomics*, vol. 290, no. 3, pp. 1095-1115, 2015.
- [5]. H. D. Chu, K. H. Nguyen, Y. Watanabe, D. T. Le, T. L. T. Pham, K. Mochida and L. P. Tran, "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, no. 11, pp. 3290, 2018.
- [6]. S. L. Pereira, C. P. Martins, A. O. Sousa, L. R. Camillo, C. P. Araujo, G. M. Alcantara, D. S. Camargo, L. C. Cidade and A. F. Almeida, "Genome-wide characterization and expression analysis of citrus NUCLEAR FACTOR-Y (NF-Y) transcription factors identified a novel NF-YA gene involved in drought-stress response and tolerance," *PLoS One*, vol. 13, no. 6, pp. e0199187, 2018.
- [7]. S. A. Filichkin, M. Ansariola, V. N. Fraser and M. Megraw, "Identification of transcription factors from NF-Y, NAC, and SPL families responding to osmotic stress in multiple tomato varieties," *Plant Sci.*, vol. 274, no. pp. 441-450, 2018.
- [8]. D. Szklarczyk et al., "The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible," *Nucleic Acids Res.*, vol. 45, no. D1, pp. D362-D368, 2016.
- [9]. R. K. Varshney et al., "Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement," *Nat Biotech*, vol. 31, no. 3, pp. 240-246, 2013.
- [10]. M. A. Larkin et al., "Clustal W and Clustal X version 2.0," *Bioinformatics*, vol. 23, no. 21, pp. 2947-2948, 2007.
- [11]. S. El-Gebali et al., "The Pfam protein families database in 2019," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D427-D432, 2019.
- [12]. M. Verma, V. Kumar, R. K. Patel, R. Garg and M. Jain, "CTDB: An integrated chickpea transcriptome database for functional and applied genomics," *PLoS One*, vol. 10, no. 8, pp. e0136880, 2015.
- [13]. I. Letunic and P. Bork, "20 years of the SMART protein domain annotation resource," *Nucleic Acids Res.*, vol. 46, no. D1, pp. D493-D496, 2017.
- [14]. V. Calvenzani et al., "Interactions and CCAAT-binding of *Arabidopsis thaliana* NF-Y subunits," *PloS One*, vol. 7, no. 8, pp. e42902, 2012.
- [15]. D. Hackenberg, Y. Wu, A. Voigt, R. Adams, P. Schramm and B. Grimm, "Studies on differential nuclear translocation mechanism and assembly of the three subunits of the *Arabidopsis thaliana* transcription factor NF-Y," *Mol Plant*, vol. 5, no. 4, pp. 876-888, 2012.
- [16]. S. K. Burley, X. Xie, K. L. Clark and F. Shu, "Histone-like transcription factors in eukaryotes," *Curr. Opin Struct Biol.*, vol. 7, no. 1, pp. 94-102, 1997.