

Xác định các gen mã hóa Nuclear factor-YB trên sắn (*Manihot esculenta* Crantz) bằng công cụ tin sinh học

Chu Đức Hà¹, Lê Thị Thảo^{2,3,*}, Lê Quỳnh Mai³, Phạm Thị Lý Thu¹

¹Viện Di truyền Nông nghiệp, Phạm Văn Đồng, Bắc Từ Liêm, Hà Nội, Việt Nam

²Công ty CP bóng đèn phích nước Rạng Đông, Hạ Đình, Thanh Xuân, Hà Nội, Việt Nam

³Khoa Sinh, Trường Đại học Khoa học Tự nhiên, ĐHQGHN, 334 Nguyễn Trãi, Thanh Xuân, Hà Nội, Việt Nam

Nhận ngày 16 tháng 8 năm 2017

Chỉnh sửa ngày 14 tháng 9 năm 2017; Chấp nhận đăng ngày 10 tháng 10 năm 2017

Tóm tắt: Nuclear factor-Y (NF-Y) là nhân tố phiên mã được tìm thấy ở hầu hết các loài thực vật, có vai trò quan trọng trong điều hòa và đáp ứng với tác nhân bất lợi. Trong nghiên cứu này, 17 gen mã hóa cho tiểu phần Nuclear factor-YB đã được xác định trên sắn (*Manihot esculenta* Crantz). Họ gen *NF-YB* phân bố trên các nhiễm sắc thể với tỷ lệ khác nhau. Phân tích cấu trúc cho thấy họ gen *MeNF-YB* rất đa dạng về chiều dài và sự sắp xếp exon/intron. Các gen cùng nằm trên một nhánh phân loại thường có cấu trúc exon/intron tương đồng nhau. Dựa trên dữ liệu microarray, hầu hết các gen *NF-YB* đều được tăng cường phiên mã tại ít nhất 1 mô chính. Đáng chú ý, *MeNF-YB12* có thể biểu hiện mạnh ở mô phân sinh đỉnh chồi, *MeNF-YB2* được tăng cường ở mô sẹo phôi hóa. Bên cạnh đó, một số gen như *MeNF-YB4/8/13/5* cũng biểu hiện mạnh ở mô sẹo phôi hóa. Kết quả này chứng tỏ rằng các gen *NF-YB* ở sắn có thể liên quan đến các quá trình sinh trưởng, phát triển của các mô phân sinh.

Từ khóa: Gen, Nuclear factor-YB, sắn, *Manihot esculenta*, tin sinh học.

1. Đặt vấn đề

Thực vật luôn chịu ảnh hưởng từ các yếu tố ngoại cảnh bất thuận. Rất nhiều cơ chế đáp ứng của thực vật đã được xác định, bản chất của các quá trình này là sự biểu hiện và điều hòa của những gen chức năng trong tế bào thực vật. Trong số đó, nhân tố phiên mã (transcription factor, TF) được chứng minh là một trong những nhóm protein quan trọng tham gia vào quá trình điều hòa sự hoạt động của gen, từ đó liên quan đến cơ chế đáp ứng của thực vật với tác nhân bất lợi.

Nuclear factor Y (NF-Y) là một trong những nhóm TF quan trọng và được tìm thấy ở tất cả các loài thực vật. Các nghiên cứu đã chứng minh rằng, NF-Y không chỉ liên quan đến quá trình sinh trưởng và phát triển ở thực vật mà còn tham gia vào cơ chế đáp ứng với các tác nhân bất lợi [1]. Về cấu trúc, NF-Y gồm 3 tiểu phần, NF-YA, NF-YB và NF-YC. Trong đó, NF-YB được đặc trưng bởi vùng gấp histone (histone fold motif, HFM) và lõi histone H2B (CBFD-NFYB-HMF, Pfam: PF00808) [2]. Đặc điểm này cho phép xác định một cách hệ thống về tiểu phần NF-YB ở các loài thực vật dựa vào một số công cụ tin sinh học trực tuyến.

Trong nghiên cứu này, nhóm tác giả tiến hành xác định các gen mã hóa NF-YB dựa trên hệ gen của sắn (*Manihot esculenta* Crantz) đã

*Tác giả liên hệ. ĐT.: 84-1675249612.

Email: lethithao52sh@gmail.com

<https://doi.org/10.25073/2588-1140/vnunst.4553>

được giải mã gần đây. Thông tin về chú giải họ gen, vị trí phân bố trên nhiễm sắc thể và đặc tính của gen đã được phân tích. Dữ liệu biểu hiện của các gen mã hóa NF-YB được tìm hiểu dựa trên dữ liệu microarray trên cổng thông tin Gene Expression Omnibus (GEO). Kết quả của nghiên cứu này sẽ cung cấp những dẫn liệu quan trọng về đặc tính của NF-YB, từ đó góp phần xác định họ TF NF-Y ở cây sắn.

2. Vật liệu và Phương pháp

2.1. Vật liệu nghiên cứu

Hệ gen của giống sắn mô hình AM560-2 [3] trên cơ sở dữ liệu Phytozome v11.0 [4].

2.2. Phương pháp nghiên cứu

- *Phương pháp xác định họ gen mã hóa NF-YB trên hệ gen sắn*: Thuật toán BlastP được sử dụng để xác định toàn bộ các protein có vùng bảo thủ *CBFD NFYB HMF* (Pfam, PF00808) [2] trên cơ sở dữ liệu của sắn [3]. Danh pháp và các thông tin về chú giải gen của họ *NF-YB* được thu thập thông qua cơ sở dữ liệu NCBI (Bioproject: PRJNA234389) [3].

- *Phương pháp phân tích cấu trúc gen*: Kích thước của từng gen thành viên của họ *NF-YB* được xác định bằng phần mềm Blast2GO [5]. Cấu trúc exon/intron được khai thác bằng công cụ GSDS 2.0 (Gene Structure Display Server) [6].

- *Phương pháp đánh giá mức độ biểu hiện của gen mã hóa NF-YB*: Dữ liệu biểu hiện của họ gen *NF-YB* ở các mô và cơ quan trong điều kiện thường được khai thác trên cơ sở GEO đã công bố (GEO accession: GSE82279) [7]. Trong nghiên cứu này, ngân hàng dữ liệu biểu hiện toàn hệ gen của sắn ở 4 loại mô, bao gồm mô sẹo phôi hóa (Friable Embryogenic Callus, FEC), tổ chức phát sinh phôi cấu tạo soma (Somatic Organized Embryogenic Structures, OES), mô phân sinh đỉnh rễ (Root Apical Meristem, RAM) và mô phân sinh đỉnh chồi (Shoot Apical Meristem, SAM) được truy cập để phân tích biểu hiện của các gen *NF-YB*

3. Kết quả nghiên cứu và thảo luận

3.1. Kết quả xác định và chú giải họ gen mã hóa tiểu phần NF-YB ở hệ gen sắn

Tìm kiếm trên cơ sở dữ liệu Phytozome [4], kết quả đã xác định được 17 trình tự có vùng bảo thủ PF00808 đặc trưng cho tiểu phần NF-YB (E-value < 1×10^{-6}). Thông tin định danh gen sau đó lần lượt được chú giải trên cổng thông tin NCBI chứa dữ liệu hệ gen của sắn đã được công bố gần đây (Bioproject: PRJNA234389) [3]. Như vậy, tổng số 17 gen mã hóa NF-YB đã được định danh và chú giải trên hệ gen sắn, thông tin về mã định danh, tên locus và gen tương đồng trên *Arabidopsis thaliana* được trình bày ở Bảng 1.

Bảng 1. Thông tin về chú giải của họ gen mã hóa tiểu phần NF-YB ở sắn

STT	Tên gen	Tên locus	Mã định danh	Tương đồng trên At
1	<i>MeNF-YB1</i>	Manes.01G215700	cassava4.1_022353m	AtNF-YB5
2	<i>MeNF-YB2</i>	Manes.01G022100	cassava4.1_018152m	AtNF-YB13
3	<i>MeNF-YB3</i>	Manes.03G193600	cassava4.1_018746m	AtNF-YB5
4	<i>MeNF-YB4</i>	Manes.03G141500	cassava4.1_025848m	AtNF-YB9
5	<i>MeNF-YB5</i>	Manes.05G117400	cassava4.1_018195m	AtNF-YB13
6	<i>MeNF-YB6</i>	Manes.05G065300	cassava4.1_029071m	AtNF-YB5
7	<i>MeNF-YB7</i>	Manes.08G002900	cassava4.1_017350m	AtNF-YB8
8	<i>MeNF-YB8</i>	Manes.08G068800	cassava4.1_032889m	AtNF-YB6
9	<i>MeNF-YB9</i>	Manes.08G119800	cassava4.1_026425m	AtNF-YB4

10	<i>MeNF-YB10</i>	Manes.09G168700	cassava4.1_032284m	AtNF-YB4
11	<i>MeNF-YB11</i>	Manes.09G080600	cassava4.1_017418m	AtNF-YB10
12	<i>MeNF-YB12</i>	Manes.10G041500	cassava4.1_017552m	AtNF-YB8
13	<i>MeNF-YB13</i>	Manes.11G162100	cassava4.1_017996m	AtNF-YB11
14	<i>MeNF-YB14</i>	Manes.12G066300	cassava4.1_016584m	AtNF-YB3
15	<i>MeNF-YB15</i>	Manes.12G157600	cassava4.1_024313m	AtNF-YB7
16	<i>MeNF-YB16</i>	Manes.13G077500	cassava4.1_026597m	AtNF-YB3
17	<i>MeNF-YB17</i>	Manes.15G013700	cassava4.1_026290m	AtNF-YB5

Tiếp theo, vị trí của 17 gen *NF-YB* ở sản cũng được xác định trên NCBI và được mô tả ở hình 1. Ở đây, tên gen mã hóa *NF-YB* được đặt theo vị trí phân bố của gen lần lượt trên các nhiễm sắc thể, từ *MeNF-YB1* đến *MeNF-YB17*. Kết quả cho thấy, các gen mã hóa *NF-YB* được phân bố trên hầu hết các nhiễm sắc thể với tỷ lệ khác nhau, nhiễm sắc thể số 2, 4, 6, 7, 14, 16, 17 và 18 không chứa gen *NF-YB*. Trong đó, có 3 gen *NF-YB* được cư trú trên nhiễm sắc thể số 8, trong khi các nhiễm sắc thể còn lại chỉ chứa 1 - 2 gen *NF-YB*.

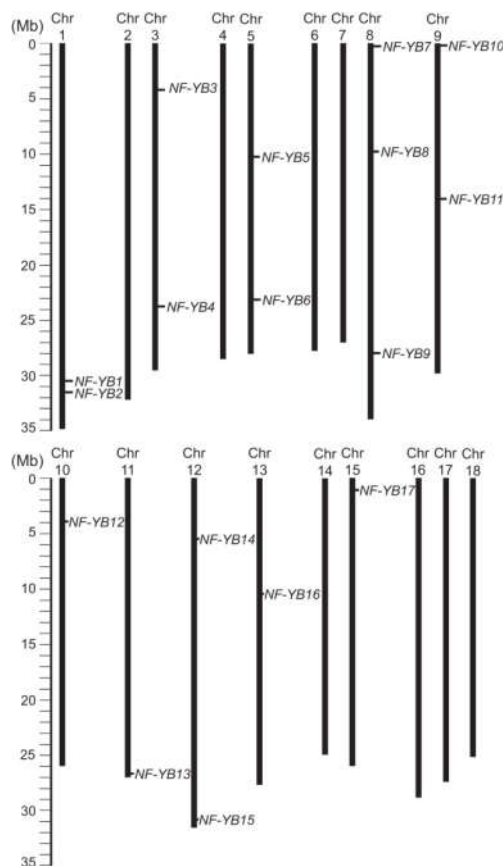
Kết quả xác định họ gen *NF-YB* ở sản được so sánh với một số nghiên cứu tương tự gần đây trên một số đối tượng cây trồng quan trọng. Trên đối tượng đậu tương (*Glycine max*), 32 gen mã hóa *NF-YB*, *GmNF-YB*, đã được xác định trong hệ gen [8]. Gần đây, kết quả nghiên cứu trên cây kê (*Setaria italica*) cũng đã ghi nhận được 15 gen *SiNF-YB* [9]. Các nghiên cứu này cho thấy rằng, họ gen *NF-YB* ở thực vật được xác định là 1 họ đa gen với số lượng thành viên rất đa dạng giữa các loài [2].

3.2. Kết quả phân tích cấu trúc gen mã hóa của họ *NF-YB* ở sản

Kết quả phân tích họ gen *NF-YB* ở sản cho thấy, kích thước vùng CDS của các gen tương đối ngắn, từ 408 nucleotide (*NF-YB5*, Manes.05G117400) đến 732 nucleotide (*NF-YB4*, Manes.03G141500). Trong khi đó, độ dài vùng genomic của các gen rất đa dạng, từ 426 nucleotide (*NF-YB10*, Manes.09G168700) đến khoảng hơn 6 kb (*NF-YB2*, Manes. 01G022100) (Hình 2).

Tiếp theo, số lượng exon và intron của gen *NF-YB* được xác định bằng công cụ GSDS [6]. Đầu tiên, cây phân loại được thiết lập dựa trên

căn trình tự tương đồng giữa các protein *NF-YB*, sau đó, trình tự chuỗi CDS và genomic của từng gen *NF-YB* tương ứng được sử dụng làm trình tự truy vấn trên công cụ GSDS [6]. Kết quả phân tích sự sắp xếp exon/intron được thể hiện ở hình 2. Có thể thấy rằng, họ *MeNF-YB* rất đa dạng về số lượng exon/intron.



Hình 1. Vị trí phân bố của gen mã hóa tiểu phần *NF-YB* trên hệ gen sản.

Các gen cùng nằm trên nhánh phân loại thường có cấu trúc gen tương tự nhau. Ví dụ, 2

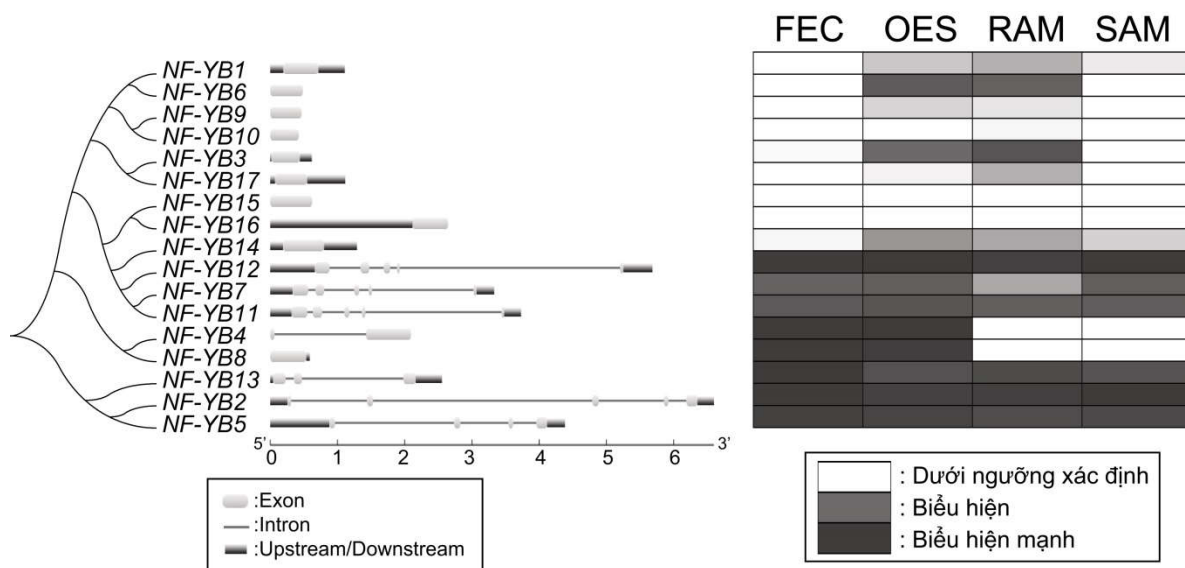
gen, *MeNF-YB7* và *MeNF-YB11* có cấu trúc 5 exon và 4 intron, trong khi rất nhiều gen, như *MeNF-YB1/6/9/10/13/15/16/17* chỉ có 1 exon. Kết quả này cho thấy rằng, các gen mã hóa NF-YB ở sắn khá đa dạng, hơn nữa, tỷ lệ A/T khá cao, chứng tỏ rằng các gen này có thể bị biến đổi trong quá trình tiến hóa để thực hiện nhiều chức năng trong tế bào.

3.3. Kết quả phân tích dữ liệu biểu hiện của các gen mã hóa NF-YB ở sắn

Dữ liệu biểu hiện của các gen thành viên của họ gen *NF-YB* được khai thác dựa trên nghiên cứu đã được công bố gần đây về ngân hàng biểu hiện toàn hệ gen trên một số mẫu mô và cơ quan quan trọng trên cây sắn (GEO accession: GSE82279) [7]. Trong nghiên cứu này, thông tin về mức độ biểu hiện

của các gen mã hóa NF-YB được tập trung khai thác trên 4 loại mô, bao gồm FEC, OES, RAM và SAM.

Kết quả cho thấy, hầu hết các gen đều có biểu hiện ở ít nhất một mẫu mô, chỉ có 2 gen *MeNF-YB15* và *MeNF-YB16* có mức biểu hiện dưới ngưỡng xác định. Trong số đó, *MeNF-YB12* và *MeNF-YB2* có xu hướng biểu hiện mạnh ở tất cả các mô, đặc biệt, *MeNF-YB12* biểu hiện mạnh ở SAM, trong khi *MeNF-YB2* biểu hiện mạnh ở FEC. Tương tự như vậy, *MeNF-YB4/8/13/5* cũng biểu hiện mạnh ở FEC. Những phân tích này đã chỉ ra rằng, các gen *NF-YB* ở sắn có thể liên quan đến các quá trình sinh trưởng, phát triển của các mô phân sinh. Nghiên cứu này sẽ được tiếp tục tiến hành nhằm tìm hiểu mức độ đáp ứng của họ gen *NF-YB* đối với tác nhân bất thuận.



Hình 2. Cấu trúc gen mã hóa và dữ liệu biểu hiện của họ NF-YB ở sắn.

4. Kết luận

Đã xác định được 17 gen mã hóa cho tiểu phần NF-YB ở sắn. Các gen này phân bố rải rác trên hệ gen của sắn.

Cấu trúc của họ gen *MeNF-YB* rất đa dạng về chiều dài và sự sắp xếp exon/intron. Các gen cùng nằm trên một nhánh phân loại thường có cấu trúc exon/intron tương đồng nhau.

Phân tích dữ liệu microarray cho thấy, hầu hết các gen *NF-YB* đều được tăng cường phiên mã tại ít nhất 1 mô chính. Trong đó, *MeNF-YB12* có thể biểu hiện mạnh ở SAM, trong khi *MeNF-YB2* được tăng cường ở FEC. Một số gen như *MeNF-YB4/8/13/5* cũng biểu hiện mạnh ở FEC. Điều này chứng tỏ rằng các gen *NF-YB* ở sắn có thể liên quan đến các quá trình sinh trưởng, phát triển của các mô phân sinh.

Tài liệu tham khảo

- [1] Mu J., Tan H., Hong S., Liang Y., Zuo J. *Arabidopsis* transcription factor genes NF-YA1, 5, 6, and 9 play redundant roles in male gametogenesis, embryogenesis, and seed development. *Mol Plant* (2013), 6: 188-201.
- [2] Laloum T., De Mita S., Gamas P., Baudin M., Niebel A. CCAAT-box binding transcription factors in plants: Y so many? *Trends Plant Sci* (2013), 18: 157-166.
- [3] Bredeson J. V., Lyons J. B., Prochnik S. E., Wu G. A., Ha C. M., Edsinger-Gonzales E., Grimwood J., Schmutz J., Rabbi I. Y., Egesi C., Nauluvula P., Lebot V., Ndunguru J., Mkamilo G., Bart R. S., Setter T. L., Gleadow R. M., Kulakow P., Ferguson M. E., Rounsley S., Rokhsar D. S. Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol* (2016), 34: 562-570.
- [4] Goodstein D. M., Shu S., Howson R., Neupane R., Hayes R. D., Fazo J., Mitros T., Dirks W., Hellsten U., Putnam N., Rokhsar D. S. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res* (2012), 40: D1178-1186.
- [5] Conesa A., Götz S., García-Gómez J. M., Terol J., Talón M., Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* (2005), 21: 3674-3676.
- [6] Hu B., Jin J., Guo A. Y., Zhang H., Luo J., Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* (2015), 31: 1296-1297.
- [7] Wilson M. C., Mutka A. M., Hummel A. W., Berry J., Chauhan R. D., Vijayaraghavan A., Taylor N. J., Voytas D. F., Chitwood D. H., Bart R. S. Gene expression atlas for the food security crop cassava. *New Phytol* (2017), 213: 1632-1641.
- [8] Quach T. N., Nguyen H. T., Valliyodan B., Joshi T., Xu D., Nguyen H. T. Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response. *Mol Genet Genomics* (2015), 290: 1095-1115.
- [9] Feng Z. J., He G. H., Zheng W. J., Lu P. P., Chen M., Gong Y., Ma Y. Z., Xu Z. S. Foxtail millet NF-Y families: genome-wide survey and evolution analyses identified two functional genes important in abiotic stresses. *Front Plant Sci* (2015), 6: 1142.

Identification of Genes Encoding Nuclear Factor-YB in Cassava (*Manihot esculenta* Crantz) using Bioinformatics

Chu Duc Ha¹, Le Thi Thao^{2,3}, Le Quynh Mai³, Pham Thi Ly Thu¹

¹Agricultural Genetics Institute, Pham Van Dong, North Tu Liem, Hanoi, Vietnam

²Rang Dong Light Source and Vacuum Flask Company, Ha Dinh, Thanh Xuan, Hanoi, Vietnam

³Faculty of Biology, VNU University of Science, 334 Nguyen Trai, Thanh Xuan, Hanoi, Vietnam

Abstract: Nuclear factor-Y (NF-Y) transcription factor can be found in most of plant species and is known to play important role in plant growth, development and stress response. In this study, 17 genes encoding NF-YB subunit were identified in cassava (*Manihot esculenta* Crantz). These genes were located in chromosomes with an uneven ratio. Analysis of gene structure indicated that *MeNF-YB* genes were varied in length and exon/intron organization. Genes in the same clade shared similar exon/intron structure. Based on available microarray database, most of *NF-YB* genes were up-regulated in at least 1 major tissue. Interestingly, *MeNF-YB12* was highly expressed in shoot apical meristem, while *MeNF-YB2* was specifically accumulated in friable embryogenesis callus. Additionally, some genes, such as *MeNF-YB4/8/13/5* were found to be strongly expressed in friable embryogenesis callus. Our results strongly indicated that *NF-YB* genes in cassava might be associated with various growth and developmental processes in meristem tissues.

Keywords: Gene, Nuclear factor-YB, cassava, *Manihot esculenta*, bioinformatics.