

## NGHIÊN CỨU CƠ CHẾ NHÂN RỘNG TRONG TIẾN HÓA CỦA CÁC GEN MÃ HÓA NHÂN TỐ PHIÊN MÃ NUCLEAR FACTOR-YB Ở CAM NGỌT (*Citrus sinensis*)

Chu Đức Hà<sup>1\*</sup>, Chu Thị Hiền Lương<sup>1,2</sup>, La Việt Hồng<sup>3</sup>,  
Thân Thị Hoa<sup>2</sup>, Phạm Thị Lý Thu<sup>1</sup>, Nguyễn Thị Thu Nga<sup>4</sup>

<sup>1</sup>Viện Di truyền Nông nghiệp - Viện Khoa học Nông nghiệp Việt Nam,

<sup>2</sup>Trường Đại học Nông Lâm Bắc Giang,

<sup>3</sup>Trường Đại học Sư phạm Hà Nội 2, <sup>4</sup>Trường Đại học Sư phạm - ĐH Thái Nguyên

### TÓM TẮT

Nhân tố phiên mã, trong đó có nhóm Nuclear factor-Y (NF-Y), gồm 3 tiểu đơn vị NF-YA, NF-YB và NF-YC, đóng vai trò quan trọng trong quá trình điều hòa biểu hiện gen liên quan đến sinh trưởng và phát triển ở thực vật. Trong nghiên cứu này, ba cặp gen lặp, *CsNF-YB11/-YB13*, *CsNF-YB3/-YB7* và *CsNF-YB9/-YB17*, đã được tìm thấy trong số 19 gen mã hóa NF-YB ở cam ngọt (*Citrus sinensis*). Các sự kiện lặp đều xuất phát từ hiện tượng chuyển lặp vùng cận đầu tận cùng trên các nhiễm sắc thể trong hệ gen cam ngọt. Trong quá trình tiến hóa, áp lực của chọn lọc tự nhiên đã kìm hãm đột biến điểm xảy ra trên các gen lặp. Trình tự bảo thủ của tiểu đơn vị NF-YB ở cam ngọt được cấu trúc từ vùng bám ADN, tương tác với NF-YA và NF-YC. Trong đó, một số axit amin đặc biệt đã được xác định trên trình tự này đã chỉ ra rằng các thành viên của họ NF-YB đóng vai trò cấu trúc NF-Y và có thể liên quan đến một số quá trình sinh lý ở cam ngọt.

**Từ khóa:** Cam ngọt, *Citrus sinensis*, gen lặp, Nuclear factor-YB, tiến hóa, tin sinh học

### MỞ ĐẦU

Dưới áp lực của chọn lọc tự nhiên, hệ gen của thực vật đã có những thay đổi trong việc phát sinh gen mới nhằm đáp ứng và chống chịu với sự thay đổi của điều kiện tự nhiên. Trong đó, nhóm nhân tố phiên mã (transcription factor, TF) đóng vai trò điều hòa sự biểu hiện của gen liên quan đến trao đổi chất và đáp ứng điều kiện bất lợi ở thực vật rất được chú ý [11]. Nghiên cứu về các nhóm TF ở thực vật luôn thu hút được nhiều sự quan tâm, bởi lẽ những kết quả thu được có thể góp phần làm rõ về cơ chế đáp ứng với điều kiện bất lợi ở thực vật.

NF-Y (Nuclear factor-Y) được biết đến như một trong số các nhóm TF phổ biến nhất ở thực vật [11]. Nhiều nghiên cứu đã cho thấy, NF-Y đóng vai trò quan trọng trong đời sống cũng như liên quan mật thiết đến tính chống chịu các điều kiện bất lợi ở thực vật [7], [9], [11]. Nhân tố phiên mã NF-Y, bao gồm 3 tiểu đơn vị NF-YA, NF-YB và NF-YC đã được ghi nhận trên cam ngọt (*Citrus sinensis*) [10],

một trong những loại cây ăn quả quan trọng hàng đầu ở Việt Nam hiện nay [1], [2]. Kết quả thu được đã cung cấp những dẫn liệu quan trọng về khả năng chống chịu ở đối tượng cây ăn quả này.

Với sự phát triển của các công cụ tin sinh học, tiểu đơn vị NF-YB ở *C. sinensis* đã được xác định với 19 thành viên [1]. Số lượng gen như vậy, so với 6 gen mã hóa tiểu đơn vị NF-YA tìm thấy được ở *C. sinensis* gần đây [2], đã đặt ra câu hỏi về sự nhân rộng của họ đa gen *CsNF-YB* diễn ra như thế nào trong quá trình tiến hóa? Trong nghiên cứu này, cơ chế tiến hóa của họ gen mã hóa tiểu đơn vị NF-YB ở cam ngọt đã được xem xét. Kết quả của nghiên cứu này sẽ cung cấp dữ liệu quan trọng để phân tích chức năng của NF-Y liên quan đến sinh trưởng, phát triển và chống chịu ở cam ngọt.

### VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

#### Dữ liệu nghiên cứu

Hệ gen và hệ protein của giống cam ngọt mô hình 'Valencia' [10] được cung cấp trên Phytozome (<http://phytozome.jgi.doe.gov/>) [5].

\* Tel: 0983 766070, Email: hachu\_amser@yahoo.com

Trình tự gen và protein của 19 thành viên tiểu đơn vị NF-YB ở cam ngọt được thu thập từ nghiên cứu trước đây [1].

### Phương pháp nghiên cứu

**Phương pháp dự đoán hiện tượng lặp gen:** Trình tự gen mã hóa NF-YB (định dạng .fasta) được sử dụng để căn trình tự tương đồng bằng công cụ MEGA [6]. Sự kiện lặp gen được xác định khi độ tương đồng (ở cấp độ nucleotide) giữa 2 gen lớn hơn 50% [9].

**Phương pháp xác định trị số thay thế đồng nghĩa, trái nghĩa và thời điểm lặp gen:** Trị số thay thế trái nghĩa Ka (nonsynonymous substitutions per non-synonymous site) và trị số thay thế đồng nghĩa Ks (synonymous substitutions per synonymous site) được xác định bằng cách phân tích trình tự nucleotide của 2 gen lặp trên phần mềm DNAsp [8]. Thời điểm xảy ra hiện tượng lặp gen được xác định tương đối theo công thức:  $T = Ks/2\lambda$  (T, đơn vị mya (million years ago), là thời điểm phát sinh hiện tượng lặp;  $\lambda$  là tỷ lệ xuất hiện Ks) [9].

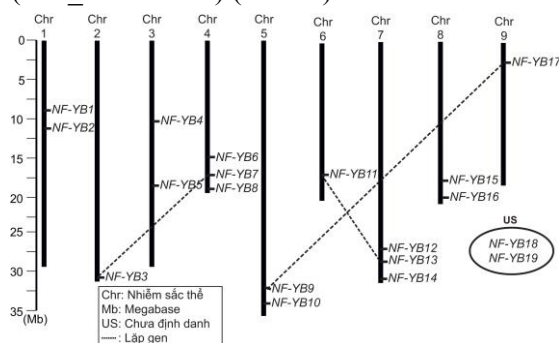
**Phương pháp phân tích vùng bảo thủ:** Vùng bảo thủ của NF-YB được xác định bằng cách căn trình tự tương đồng trên MEGA [6]. Thông số được cài đặt như sau, hình phạt khoảng trống đầu tiên (Gap open penalty) là 10, hình phạt khoảng trống phía sau (Gap extension penalty) là 0,2.

### KẾT QUẢ VÀ BÀN LUẬN

#### Kết quả dự đoán sự kiện lặp gen của họ gen mã hóa tiểu đơn vị NF-YB ở cam ngọt

Để xác định hiện tượng gen lặp, trình tự nucleotide của các gen *CsNF-YB* được sử dụng để căn trình tự tương đồng bằng MEGA [6]. Kết quả cho thấy, ba cặp gen lặp đã được xác định với mức độ tương đồng ở cấp độ nucleotide lớn hơn 50% (Bảng 1). Cụ thể, *CsNF-YB3* (XM\_015525536) được xác định lặp với *CsNF-YB7* (XM\_006475051), *CsNF-YB9* (XM\_006480045) được dự đoán lặp với *CsNF-YB17* (XM\_006489046), trong khi cặp gen lặp thứ 3 được tìm thấy lần lượt là *CsNF-*

*YB11* (XM\_006480533) và *CsNF-YB13* (XM\_006485877) (Hình 1).



**Hình 1.** Sự kiện lặp gen của họ gen mã hóa tiểu đơn vị NF-YB ở cam ngọt

Dựa trên vị trí phân bố trên hệ gen của *C. sinensis*, 3 cặp gen lặp đều xảy ra trên các nhiễm sắc thể khác nhau (segmental duplication). Điều này chứng tỏ sự kiện lặp ở họ gen mã hóa tiểu đơn vị NF-YB ở cam ngọt là kết quả của quá trình chuyển vùng cận đầu tận cùng của các nhiễm sắc thể (subtelomere) (Hình 1). Trước đó, hiện tượng này đã được ghi nhận như một giả thuyết để giải thích cho sự phát tán và nhân rộng của các gen [3]. Trên lúa miến (*Sorghum bicolor*), cơ chế nhân rộng của họ gen mã hóa TF NF-Y (*SbNF-YA*, *SbNF-YB*, *SbNF-YC*) cũng được chứng minh xuất phát từ hiện tượng lặp trên các nhiễm sắc thể khác nhau [9]. Những kết quả thu được cho thấy các gen mã hóa tiểu đơn vị NF-YB ở cam ngọt đều được phát sinh từ sự kiện lặp tại vùng cận đầu tận cùng trên các nhiễm sắc thể khác nhau. Để xem xét rõ hơn về cơ chế này, tác động của quá trình chọn lọc tự nhiên đến họ gen *CsNF-YB* tiếp tục được phân tích.

#### Kết quả phân tích áp lực của tiến hóa đến họ gen mã hóa NF-YB ở cam ngọt

Trong nghiên cứu này, trị số thay thế đồng nghĩa Ks và trái nghĩa Ka của 3 cặp gen lặp được xác định để xem xét áp lực của chọn lọc tự nhiên đến quá trình nhân rộng của họ gen *CsNF-YB* ở cam ngọt trong quá trình tiến hóa. Kết quả cho thấy tất cả các sự kiện lặp đều có giá trị Ka/Ks nhỏ hơn 1 (Bảng 1).

**Bảng 1.** Kết quả phân tích đặc điểm tiến hóa của họ gen *CsNF-YB* ở *C. sinensis*

Cặp gen lặp	Mức độ tương đồng (%)	Ks	Ka	Ka/Ks	Thời điểm lặp (mya)
<i>CsNF-YB3/CsNF-YB7</i>	54,8	1,14	0,52	0,46	87,95
<i>CsNF-YB9/CsNF-YB17</i>	50,8	0,53	0,21	0,40	40,97
<i>CsNF-YB11/CsNF-YB13</i>	72,0	1,96	0,24	0,12	150,50

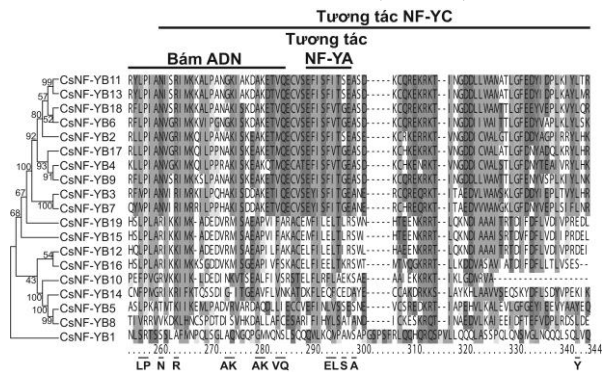
Cặp gen lặp *CsNF-YB3/-YB7*, *CsNF-YB7/-YB17* có trị số Ka/Ks khoảng 0,4 trong khi giá trị này của *CsNF-YB11/-YB13* thấp, đạt 0,12. Về mặt lý thuyết, cặp gen lặp có Ka/Ks nhỏ hơn 1 chứng tỏ chọn lọc tự nhiên đã bảo tồn sự toàn vẹn cấu trúc của gen [4]. Như vậy, tiến hóa đã kìm hãm sự phát sinh những đột biến điểm trên 3 cặp gen lặp của họ gen mã hóa tiểu đơn vị NF-YB ở cam ngọt. Trước đó, phân tích mức độ biểu hiện của họ gen *CsNF-YB* đã chỉ ra rằng, các gen lặp này biểu hiện rất mạnh ở một số bộ phận của cây cam ngọt [1]. Những bằng chứng này đã chỉ ra rằng các gen lặp đều được bảo toàn sự nguyên vẹn trong cấu trúc để bảo đảm chức năng trong hoạt động sống của loài *C. sinensis* trước sự thay đổi của ngoại cảnh.

Tuy nhiên, vấn đề đặt ra là tại sao *CsNF-YB13* có biểu hiện rất mạnh ở callus, hoa, lá và quả trong khi gen lặp *CsNF-YB11* chỉ có xu hướng biểu hiện mạnh ở 4 cơ quan trên?

Tương tự, *CsNF-YB3* biểu hiện đặc thù ở callus nhưng *CsNF-YB7* cũng chỉ có xu hướng biểu hiện tăng ở mô này [1]. Để trả lời câu hỏi này, thời điểm xảy ra sự kiện lặp của 3 cặp gen đã được xem xét. Kết quả đạt được cho thấy thời điểm tương đối xảy ra sự kiện lặp của 2 cặp gen *CsNF-YB3/-YB7* và *CsNF-YB11/-YB13* lần lượt cách đây khoảng 88 và 150 triệu năm (Bảng 1). Cặp gen lặp *CsNF-YB9/-YB17* được dự đoán xảy ra cách đây khoảng 41 triệu năm. Về mặt tiến hóa, cam ngọt được cho là tách khỏi bộ Bông cách đây khoảng 85 triệu năm [10]. Điều này phù hợp với giả thuyết rằng, sự kiện lặp *CsNF-YB11/-YB13* và *CsNF-YB3/-YB7* xảy ra trước khi *C. sinensis* tiến hóa, 1 trong 2 gen lặp sẽ mất chức năng (subfunctionalization), trong khi *CsNF-YB9/-YB17* xảy ra sau khi *C. sinensis* tách khỏi bộ Bông, 2 gen này vẫn giữ được chức năng (neofunctionalization).

**Kết quả phân tích cấu trúc vùng bảo thủ của tiểu đơn vị NF-YB ở cam ngọt**

Cấu trúc vùng bảo thủ của tiểu đơn vị NF-YB ở cam ngọt được phân tích bằng công cụ MEGA [6]. Kết quả đã chỉ ra rằng, hầu hết các thành viên trong cùng một nhánh đều chia sẻ trình tự polypeptit giống nhau, có chiều dài khoảng 89 axit amin, tương tự như trên các loài thực vật khác [7], [9], [11]. Hơn nữa, cấu trúc của NF-YB ở cam ngọt được cấu tạo từ 3 bộ phận, vùng bám ADN, vùng tương tác với tiểu đơn vị NF-YA và NF-YC (Hình 2).



**Hình 2.** Phân tích cấu trúc vùng bảo thủ của tiểu đơn vị NF-YB ở cam ngọt

Trình tự bảo thủ này của tiểu đơn vị NF-YB ở *C. sinensis* cũng đã được ghi nhận ở *S. bicolor* 4 cũng như trên các loài thực vật khác [7], [9], [11]. Phân tích trình tự bảo thủ đã tách biệt họ NF-YB ở cam ngọt thành 2 nhóm chính với sự có mặt của các gốc axit amin quan trọng. Một số chuỗi axit amin, leucine - proline (LP), asparagine (N), arginine (R), alanine - lysine (AK), valine - glutamine (VQ), axit glutamic (E), leucine (L), serine (S), alanine (A), tyrosine (Y), được tìm thấy trong cấu trúc của các thành viên của tiểu đơn vị NF-YB ở cam ngọt (Hình 2). Trước đó, các axit amin này đã được chứng minh rằng đóng vai trò quan trọng trong quá trình bám ADN và tương tác với 2 tiểu đơn vị NF-YA và NF-YC [7], [11]. Những dẫn liệu này cho thấy

NF-YB không chỉ thực hiện có chức năng cấu thành TF NF-Y mà còn có thể đóng vai trò trong các quá trình sinh trưởng và phát triển ở cam ngọt.

#### KẾT LUẬN

Đã xác định được 3 cặp gen lặp, *CsNF-YB11/-YB13*, *CsNF-YB3/-YB7* và *CsNF-YB9/-YB17* trong họ gen mã hóa NF-YB của cam ngọt. Hiện tượng chuyển lặp vùng cận đầu tận cùng trên các nhiễm sắc thể khác nhau là cơ chế chính của sự kiện lặp gen.

Các cặp gen lặp đều giữ nguyên được sự toàn vẹn trong cấu trúc gen dưới áp lực của chọn lọc tự nhiên. *CsNF-YB3/-YB7* và *CsNF-YB11/-YB13* xảy ra trước khi *C. sinensis* tiến hóa nên 1 trong 2 gen lặp mất chức năng, trong khi *CsNF-YB9/-YB17* vẫn giữ được chức năng do sự kiện này xảy ra sau khi *C. sinensis* tiến hóa.

Cấu trúc bảo thủ của NF-YB ở cam ngọt chứa cấu trúc vùng bám ADN, tương tác với tiểu đơn vị NF-YA và NF-YC. Sự có mặt của một số axit amin bảo thủ cho thấy NF-YB đóng vai trò cấu trúc TF NF-Y và có thể tham gia vào một số quá trình sinh lý ở cam ngọt.

#### TÀI LIỆU THAM KHẢO

- Chu Đức Hà, Nguyễn Thu Trang, Đoàn Thị Hải Dương, Vũ Thị Thu Hiền, Nguyễn Văn Giang, Phạm Thị Lý Thu, Lê Tiến Dũng (2017), "Phân tích *in silico* họ gen mã hóa yếu tố phiên mã Nuclear factor-YB trên cam ngọt (*Citrus sinensis*)", *Tạp chí Khoa học Công nghệ Nông nghiệp Việt Nam*, 3(76), tr. 22-26.
- Chu Duc Ha, Le Quoc Oai, Nguyen Van Giang, Pham Thi Ly Thu, Le Hung Linh (2016), "Insight into the Nuclear factor-YA transcription factor family in sweet orange (*Citrus sinensis*)", *J. Vietnam Agri. Sci. Tech.*, 1(2), pp. 26-31.
- Emanuel B. S., Shaikh T. H. (2001), "Segmental duplications: An 'expanding' role in genomic instability and disease", *Nat. Rev. Genet.*, 2(10), pp. 791-800.
- Galtier N., Gouy M., Gautier C. (1996), "SEAVIEW and PHYLO\_WIN: Two graphic tools for sequence alignment and molecular phylogeny", *Comput Appl. Biosci.*, 12(6), pp. 543-548.
- 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. (2012), "Phytozome: A comparative platform for green plant genomics", *Nucleic Acids Res.*, 40(Database issue), pp. D1178-D1186.
- Kumar S., Stecher G., Tamura K. (2016), "MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets", *Mol. Biol. Evol.*, 33(7), pp. 1870-1874.
- Laloum T., De Mita S., Gamas P., Baudin M., Niebel A. (2013), "CCAAT-box binding transcription factors in plants: Y so many?", *Trends Plant Sci.*, 18(3), pp. 157-166.
- Librado P., Rozas J. (2009), "DnaSP v5: A software for comprehensive analysis of DNA polymorphism data", *Bioinformatics*, 25(11), pp. 1451-1452.
- Malviya N., Jaiswal P., Yadav D. (2016), "Genome-wide characterization of Nuclear Factor Y (NF-Y) gene family of sorghum [*Sorghum bicolor* (L.) Moench]: A bioinformatics approach", *Physiol Mol. Biol. Plants*, 22(1), pp. 33-49.
- Xu Q., Chen L. L., Ruan X., Chen D., Zhu A., Chen C., Bertrand D., Jiao W. B., Hao B. H., Lyon M. P., Chen J., Gao S., Xing F., Lan H., Chang J. W., Ge X., Lei Y., Hu Q., Miao Y., Wang L., Xiao S., Biswas M. K., Zeng W., Guo F., Cao H., Yang X., Xu X. W., Cheng Y. J., Xu J., Liu J. H., Luo O. J., Tang Z., Guo W. W., Kuang H., Zhang H. Y., Roose M. L., Nagarajan N., Deng X. X., Ruan Y. (2013), "The draft genome of sweet orange (*Citrus sinensis*)", *Nature Genet.*, 45(1), pp. 59-66.
- Zanetti M. E., Ripodas C., Niebel A. (2017), "Plant NF-Y transcription factors: Key players in plant-microbe interactions, root development and adaptation to stress", *Biochim. Biophys. Acta.*, 1860(5), pp. 645-654.

## SUMMARY

**EVOLUTIONARY EXPANSION OF NUCLEAR FACTOR-YB GENE FAMILY IN SWEET ORANGE (*Citrus sinensis*)**

**Chu Duc Ha<sup>1\*</sup>, Chu Thi Hien Luong<sup>1,2</sup>, La Viet Hong<sup>3</sup>,  
Than Thi Hoa<sup>2</sup>, Pham Thi Ly Thu<sup>1</sup>, Nguyen Thi Thu Nga<sup>4</sup>**

<sup>1</sup>Agricultural Genetics Institute - Vietnam Academy of Agricultural Sciences

<sup>2</sup>Bac Giang Agriculture and Forestry University

<sup>3</sup>Hanoi Pedagogical University 2, <sup>4</sup>TNU - University of Education

Transcription factors, such as Nuclear factor-Y (NF-Y), formed by 3 subunits, NF-YA, NF-YB and NF-YC, were considered to naturally operate as regulators of biological processes in crop plants. In this study, 3 duplicated gene pairs, including CsNF-YB11/-YB13, CsNF-YB3/-YB7 and CsNF-YB9/-YB17, have been computationally predicted in 19 previously published members of CsNF-YB gene family in sweet orange (*Citrus sinensis*). The subtelomeric regions of chromosomes are more permissive to the expansion of all segmental CsNF-YB duplicated genes. During the evolution process, natural selection had positively prevented the mutagenesis in the duplicated pairs. The conserved domain of NF-YB subunit in sweet orange was characterized by 3 regions, including DNA binding, NF-YA interaction and NF-YC interaction domains. Among them, the existences of several specific amino acids in these sequences showed that whole members of NF-YB not only play significant role in the construction of NF-Y but also may be involved in various biological processes in sweet orange.

**Keywords:** *Sweet orange, Citrus sinensis, gene duplication, Nuclear factor-YB, evolution, bioinformatics.*

*Ngày nhận bài: 29/01/2018; Ngày phản biện: 22/02/2018; Ngày duyệt đăng: 27/4/2018*

\* Tel: 0983 766070, Email: hachu\_amser@yahoo.com