

Xác định và đánh giá mức độ biểu hiện của họ gen mã hóa tiểu phần Nuclear factor-YC ở cây sắn (*Manihot esculenta*)

Chu Đức Hà^{1*}, Nguyễn Đức Anh^{1,2}, Hoàng Thị Thao², Phạm Thị Lý Thu¹, Phạm Phương Thu³, Chu Thùy Dương², Nguyễn Thị Thu Phương², Phạm Xuân Hội¹

¹Viện Di truyền Nông nghiệp, Viện Khoa học Nông nghiệp Việt Nam

²Khoa Nông học, Trường Đại học Nông lâm Bắc Giang

³Khoa Sinh - Kỹ thuật nông nghiệp, Trường Đại học Sư phạm Hà Nội 2

Ngày nhận bài 2/4/2019; ngày chuyển phân biện 5/4/2019; ngày nhận phân biện 6/5/2019; ngày chấp nhận đăng 21/5/2019

Tóm tắt:

Nhân tố phiên mã Nuclear factor-Y (NF-Y) được cấu tạo bởi 3 tiểu phần NF-YA, NF-YB và NF-YC, là nhóm protein điều hòa tham gia vào nhiều quá trình quan trọng ở thực vật. Trong nghiên cứu này, tiểu phần NF-YC đã được phân tích ở cây sắn (*Manihot esculenta*) dựa trên các phân tích tin sinh học. Kết quả đã xác định được 14 gen mã hóa NF-YC ở genome cây sắn, đặt tên là *MeNF-YC*. Phân tích cấu trúc cho thấy, phần lớn các gen *MeNF-YC* chỉ có một exon, tương tự như ghi nhận ở các loài thực vật khác. Đánh giá dữ liệu transcriptome của họ *MeNF-YC* đã chứng minh phần lớn các gen thành viên có biểu hiện ở ít nhất một mô cơ quan, bộ phận trên cây sắn trong điều kiện thường. Đặc biệt, *MeNF-YC05* được xác định là gen có biểu hiện mạnh ở nhiều vị trí nhất (ở cả mô phân sinh đỉnh chồi, mô phân sinh chóp rễ, tổ chức phát sinh phôi cấu tạo soma và củ). Những kết quả này đã cung cấp các dữ liệu một cách toàn diện về họ gen mã hóa tiểu phần NF-YC ở sắn, từ đó cung cấp gen ứng viên cho nghiên cứu chức năng tiếp theo nhằm nâng cao tính chống chịu bất lợi ở cây sắn.

Từ khóa: biểu hiện gen, cấu trúc gen, nuclear factor-YC, sắn, tin sinh học.

Chỉ số phân loại: 4.6

Mở đầu

Nhân tố phiên mã (transcription factor - TF) là những nhóm protein điều hòa, có trình tự đặc hiệu giúp quá trình nhận biết và bám trên vùng promoter của gen, từ đó điều hòa sự biểu hiện của gen. Ở thực vật, các họ TF đóng vai trò quan trọng, tham gia vào nhiều quá trình sinh học thiết yếu, có thể kể đến như quá trình chín quả ở cây cà chua (*Solanum lycopersicum*) [1], điều hòa sự tương tác giữa rễ - hệ vi sinh vật đất [2] và cơ chế phát sinh phôi ở cây bông (*Gossypium hirsutum*) [3]. Trong đó, Nuclear factor-Y (NF-Y), cấu tạo từ 3 tiểu phần NF-YA, NF-YB và NF-YC, là một trong những nhóm TF đặc trưng, được ghi nhận có mặt ở tất cả các loài thực vật [4].

Với những thành tựu của các công nghệ “-omics”, các họ TF NF-Y đã được nghiên cứu trên nhiều cây trồng quan trọng, như lúa gạo (*Oryza sativa*) [5], ngô (*Zea mays*) [6] và đậu tương (*Glycine max*) [7]. Trong các ghi nhận gần đây,

2 tiểu phần của TF NF-Y, bao gồm NF-YA và NF-YB đã được phân tích trên đối tượng cây sắn (*Manihot esculenta*) [8-10]. Đây là 1 trong 5 loài cây trồng quan trọng ở Việt Nam, được sử dụng là thực phẩm, nguyên liệu chế biến thức ăn gia súc và sản xuất nhiên liệu sinh học. Tuy nhiên, chưa có ghi nhận nào về tiểu phần NF-YC trên đối tượng cây có củ có ý nghĩa này.

Trong nghiên cứu này, họ gen mã hóa tiểu phần NF-YC đã được xác định trên genome của giống sắn mô hình ‘AM560-2’. Dựa trên các thông tin được chú giải, cấu trúc của các gen mã hóa NF-YC đã được phân tích. Thêm vào đó, thông tin biểu hiện của các gen *NF-YC* trong điều kiện thường đã được khai thác dựa trên cơ sở dữ liệu transcriptome. Kết quả của nghiên cứu này đã cung cấp thêm những dẫn liệu quan trọng về họ TF NF-Y ở cây sắn, từ đó có thể xem xét và đề xuất những gen ứng viên cho phân tích chức năng gen nhằm nâng cao tính chống chịu bất lợi ở cây sắn nói riêng và thực vật nói chung.

*Tác giả liên hệ: Email: hachu_amsr@yahoo.com

Identification and *in silico* analysis of expression profiles of genes encoding Nuclear factor-YC subunit in cassava (*Manihot esculenta*) using the bioinformatics approach

Duc Ha Chu^{*}, Duc Anh Nguyen^{1,2}, Thi Thao Hoang²,
Thi Ly Thu Pham¹, Phuong Thu Pham³,
Thuy Duong Chu², Thi Thu Phuong Nguyen², Xuan Hoi Pham¹

¹Agricultural Genetics Institute, Vietnam Academy of Agricultural Sciences

²Faculty of Agronomy, Bac Giang Agriculture and Forestry University

³Faculty of Biology - Agricultural Technology, Hanoi Pedagogical University 2

Received 2 April 2019; accepted 21 May 2019

Abstract:

Nuclear factor-Y (NF-Y), composed of NF-YA, NF-YB and NF-YC subunits, is known as the regulatory protein involving in various important processes in plants. In this study, the NF-YC subunit was analysed in cassava (*Manihot esculenta*) based on the bioinformatic approach. As the results, a total of 14 genes encoding NF-YC, namely *MeNF-YC*, were identified in the cassava genome. Our structural analysis showed that the majority of *MeNF-YC* contained only one exon, similar as previously confirmed in other plant species. Based on the available transcriptome atlas, most of *MeNF-YC* genes were found to express in at least one tissue, organ in cassava plants in the normal condition. Notably, *MeNF-YC05* was defined as the highly-expressed gene in many tissues, including the shoot and root apical meristems, somatic embryogenic structures, and storage roots. Our results would provide the comprehensive information of the genes encoding NF-YC in cassava, thereby suggesting a list of candidate genes for studying the functional characterisation to improve the unfavorable condition tolerance of cassava.

Keywords: bioinformatics, cassava, gene expression, gene structure, Nuclear factor-YC.

Classification number: 4.6

Dữ liệu và phương pháp nghiên cứu

Dữ liệu

Dữ liệu genome và proteome của giống sắn mô hình ‘AM560-2’ giải mã gần đây [11] được khai thác trên NCBI. Dữ liệu transcriptome của giống sắn ‘TME 204’ được công bố gần đây (GEO accession: GSE82279) [12].

Phương pháp nghiên cứu

Phương pháp xác định NF-YC ở sắn: toàn bộ NF-YC ứng viên ở sắn được khai thác từ PlantTFDB [13]. Trình tự protein của các ứng viên được kiểm tra trên Pfam [14] để kiểm chứng sự có mặt của vùng bảo tồn NF-YC ở thực vật như trong ghi nhận gần đây [4].

Phương pháp chú giải thông tin gen NF-YC ở sắn: trình tự protein của NF-YC được BlastP vào proteome của sắn [11] trên NCBI để tìm kiếm mã gen, mã ARN và mã protein. Sau đó, trình tự gADN (genomic ADN), CDS (coding DNA sequence) và vị trí phân bố gen *NF-YC* được thu thập trên genome của sắn [11] tại NCBI.

Phương pháp phân tích cấu trúc gen NF-YC ở sắn: trình tự gADN và CDS của từng gen *NF-YC* ở sắn lần lượt được truy vấn trên PIECE [15] để phân tích cấu trúc exon/intron mục tiêu. Kết quả được mô hình hóa trên Illustrator. Sự sắp xếp của gen *NF-YC* được xác định dựa trên cây phân loại xây dựng theo phương pháp Neighbor-Joining trong MEGA [16].

Phương pháp đánh giá biểu hiện của gen NF-YC ở sắn: mã định danh gen *NF-YC* được sử dụng để khai thác trên transcriptome của sắn trong điều kiện thường [12]. Trong đó, các mẫu mô và bộ phận chính trên cây sắn được thu thập, bao gồm mô sẹo phôi hóa (Friable embryogenic callus - FEC), tổ chức phát sinh phôi cấu tạo soma (Somatic organized embryogenic structure - OES), mô phân sinh chóp rễ (Root apical meristem - RAM), mô phân sinh đỉnh chồi (Shoot apical meristem - SAM), củ (Storage root - SR), rễ sợi (Fibrous root - FR), thân (Shoot - S), chồi bên (Lateral bud - LB), lá (Leaf - L), gân lá (Mid vein - MV) và cuống lá (Petiole - P) [12].

Kết quả và thảo luận

Kết quả tìm kiếm và xác định tiểu phần NF-YC ở sắn

Để tìm kiếm toàn bộ các NF-YC ở sắn, cơ sở dữ liệu PlantTFDB [13] được rà soát để sàng lọc tất cả thông tin về tiểu phần này. Kiểm chứng vùng bảo thủ NF-YC trên Pfam [14], 14 protein ứng viên đã được xác định (bảng 1). Theo đó, các gen *NF-YC* được chú giải trên genome dựa

trên thuật toán BlastP trình tự amino acid tương ứng trên proteome của sắn [11] trên NCBI. Thông tin cơ bản về NF-YC ở sắn, bao gồm mã gen, mã mRNA, mã protein và mã locus được thể hiện ở bảng 1.

Bảng 1. Thông tin chú giải của họ gen *MeNF-YC* ở cây sắn.

TT	Tên gen	Mã gen ¹	Mã mRNA ¹	Mã protein ^{1,2}	Mã locus ¹
1	<i>MeNF-YC01</i>	Manes.01G063900	cassava4.1_027362m	XP_021604318	LOC110609202
2	<i>MeNF-YC02</i>	Manes.01G113300	cassava4.1_026359m	XP_021629508	LOC110627482
3	<i>MeNF-YC03</i>	Manes.01G176600	cassava4.1_013455m	XP_021607456	LOC110611429
4	<i>MeNF-YC04</i>	Manes.03G183200	cassava4.1_031452m	XP_021607269	LOC110611327
5	<i>MeNF-YC05</i>	Manes.04G022800	cassava4.1_015531m	XP_021609413	LOC110612883
6	<i>MeNF-YC06</i>	Manes.04G155300	cassava4.1_012635m	XP_021610750	LOC110613762
7	<i>MeNF-YC07</i>	Manes.06G120200	cassava4.1_015422m	XP_021614739	LOC110616616
8	<i>MeNF-YC08</i>	Manes.06G176100	cassava4.1_014053m	XP_021616875	LOC110618125
9	<i>MeNF-YC09</i>	Manes.09G165000	cassava4.1_014289m	XP_021623996	LOC110623383
10	<i>MeNF-YC10</i>	Manes.11G008800	cassava4.1_012312m	XP_021628782	LOC110626907
11	<i>MeNF-YC11</i>	Manes.12G012100	cassava4.1_026058m	XP_021631336	LOC110628830
12	<i>MeNF-YC12</i>	Manes.14G009500	cassava4.1_014024m	XP_021592144	LOC110599865
13	<i>MeNF-YC13</i>	Manes.14G050900	cassava4.1_034343m	XP_021591876	LOC110599671
14	<i>MeNF-YC14</i>	Manes.15G024300	cassava4.1_029608m	XP_021595951	LOC110602683

Ghi chú: ¹: thông tin được khai thác từ NCBI; ²: thông tin khai thác từ PlantTFDB.

Trước đây, họ tiểu phân NF-YC cũng đã được xác định trên một số đối tượng cây trồng khác. Cụ thể, 16 thành viên OsNF-YC đã được ghi nhận trên lúa gạo [5], trong khi ở ngô, Zhang và cs (2016) đã sàng lọc được 18 NF-YC [6]. Trước đó, họ NF-YC (gồm 15 GmNF-YC) đã được nghiên cứu trên đậu tương [7]. Như vậy, các gen mã hóa tiểu phân NF-YC ở thực vật nói chung là họ đa gen, với số lượng gen thành viên khác nhau giữa các đối tượng.

Tiếp theo, vị trí phân bố của các gen *NF-YC* được xác định trên genome của sắn [11] trên NCBI. Theo đó, họ gen *NF-YC* phân bố rải rác trên các nhiễm sắc thể (NST) của sắn với tỷ lệ khác nhau. Cụ thể, 3 gen *NF-YC* nằm trên NST1, trong khi NST4, 6 và 14 đều chứa hai gen *NF-YC*. Các NST3, 9, 11, 12, 14 và 15 lần lượt chỉ chứa duy nhất một gen *NF-YC*. Trong nghiên cứu này, tên của các gen mã hóa tiểu phân NF-YC được đặt theo vị trí phân bố trên genome với 'Me' ký hiệu cho *M. esculenta* (bảng 1, 2), tương tự như phương thức đặt tên cho MeNF-YA và MeNF-YB đã được ghi nhận gần đây ở sắn [8-10] cũng như các loài thực vật khác [5-7].

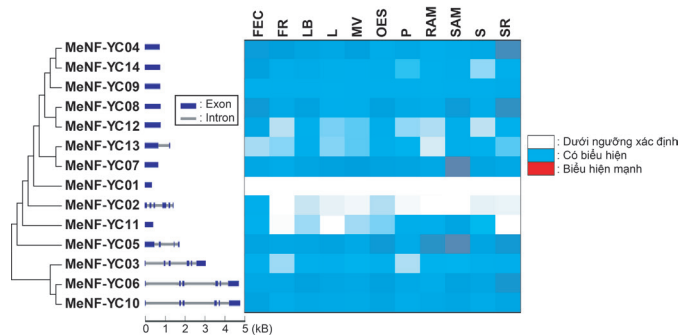
Bảng 2. Thông tin chú giải của họ gen *MeNF-YC* ở cây sắn

TT	Tên gen	Vị trí phân bố	Minh họa
1	<i>MeNF-YC01</i>	NST1: NC_035161.1 (18625791..18626144)	NST1
2	<i>MeNF-YC02</i>	NST1: NC_035161.1 (23354134..23356164)	NST1
3	<i>MeNF-YC03</i>	NST1: NC_035161.1 (27806813..27810509)	NST1
4	<i>MeNF-YC04</i>	NST3: NC_035163.1 (26997533..26999688, C)	NST3
5	<i>MeNF-YC05</i>	NST4: NC_035164.1 (2462001..2464098)	NST4
6	<i>MeNF-YC06</i>	NST4: NC_035164.1 (27771621..27776915)	NST4
7	<i>MeNF-YC07</i>	NST6: NC_035166.1 (22901983..22908231, C)	NST6
8	<i>MeNF-YC08</i>	NST6: NC_035166.1 (27506028..27508596)	NST6
9	<i>MeNF-YC09</i>	NST9: NC_035169.1 (27776414..27780827)	NST9
10	<i>MeNF-YC10</i>	NST11: NC_035171.1 (936525..941910, C)	NST11
11	<i>MeNF-YC11</i>	NST12: NC_035172.1 (1020980..1021585)	NST12
12	<i>MeNF-YC12</i>	NST14: NC_035174.1 (914713..917351, C)	NST14
13	<i>MeNF-YC13</i>	NST14: NC_035174.1 (4018243..4022194)	NST14
14	<i>MeNF-YC14</i>	NST15: NC_035175.1 (1853556..1855955)	NST15

Đáng chú ý, phần lớn họ gen *MeNF-YC* (11 trên 14) đều phân bố ở vùng cận đầu mút của các NST, ngoại trừ 3 gen *MeNF-YC* trên NST1 (bảng 2). Hiện tượng này cũng được quan sát thấy trên họ gen mã hóa tiểu phân NF-YA và NF-YB ở sắn [8-10] cũng như họ gen *NF-YC* trên các đối tượng khác [5-7].

Kết quả phân tích cấu trúc gen mã hóa tiểu phân NF-YC ở sắn

Họ *MeNF-YC* được phân tích trên PIECE [15] để xác định cấu trúc gen (exon/intron). Kết quả cho thấy, họ *MeNF-YC* có sự đa dạng về số lượng exon/intron trong cấu trúc của các gen thành viên (hình 1). Cụ thể, số lượng exon của họ gen *MeNF-YC* dao động từ một (8 gen), hai (1 gen), bốn (4 gen) đến sáu exon (4 gen) (hình 1). Đáng chú ý, các gen *MeNF-YC* nằm cùng một nhánh theo cây phân loại Neighbor-Joining thường có cấu trúc và kích thước gen tương tự nhau (hình 1).



Hình 1. Cấu trúc và mức độ biểu hiện của họ gen *MeNF-YC* trong điều kiện thường.

So sánh với số lượng exon trên *NF-YC* ở các loài thực vật khác cho thấy họ *NF-YC* có cấu trúc rất đa dạng. Ở cả chua, 13 trên tổng số 20 gen *SINF-YC* có một hoặc hai exon, trong khi số gen còn lại có cấu trúc dao động từ 3 đến 6 exon, duy nhất *Solyc02g091030* có 22 exon [1]. Trong khi đó, 12 thành viên của họ *ZmNF-YC* ở ngô (trên 18 gen) chỉ ghi nhận một exon, các gen còn lại chứa từ 2 đến 6 exon [6]. Các kết quả này cho thấy, họ *NF-YC* ở thực vật có cấu trúc gen đa dạng, tuy nhiên, một tỷ lệ lớn gen thành viên đều chỉ có một exon, tương tự như trong ghi nhận trước đây [4].

Kết quả đánh giá biểu hiện của các gen mã hóa tiểu phần *NF-YC* trong điều kiện thường

Trong nghiên cứu này, để tìm hiểu mức độ biểu hiện của họ gen *MeNF-YC* trong điều kiện thường, dữ liệu phiên mã ở 11 mẫu mô, cơ quan và bộ phận chính trên cây sắn đã được khai thác để xử lý [12]. Cụ thể, cây sắn ‘AM560-2’ đã được trồng trong điều kiện thường “tiêu chuẩn”, với chế độ ánh sáng 12h sáng - 12h tối, nhiệt độ tối ưu trong khoảng 28-32°C (ban ngày) và 25-27°C (ban đêm), độ ẩm 70% [12]. Mẫu thu thập tại các mô chính được sử dụng để tách chiết và giải trình tự RNA-Seq [12]. Gen được quy ước là ‘có biểu hiện’ khi giá trị Fragments per kilobase million (FPKM) >10, ‘biểu hiện mạnh’ khi FPKM >300, trong khi mức độ phiên mã của gen dưới ngưỡng phát hiện khi FPKM~1 [12]. Kết quả phân tích biểu hiện của các gen *MeNF-YC* ở 11 mẫu mô được minh họa ở hình 1.

Phần lớn các gen đều có biểu hiện ở ít nhất một mẫu mô, cơ quan chính trên cây trong điều kiện thường, ngoại trừ gen *MeNF-YC01* có mức độ phiên mã dưới ngưỡng phát hiện (hình 1). Có thể thấy rằng, một số gen *MeNF-YC* có xu hướng biểu hiện mạnh tại các vị trí trên cây (hình 1). Cụ thể, *MeNF-YC04*, *MeNF-YC06* và *MeNF-YC08* được ghi nhận có xu hướng biểu hiện đặc thù ở SR, trong khi *MeNF-YC07* có mức độ phiên mã mạnh ở SAM (hình 1). Đáng chú ý, gen *MeNF-YC05* có biểu hiện tương đối mạnh ở cả RAM, SAM, OES và SR (hình 1), cho thấy gen này có thể tham gia vào các quá trình diễn ra tại những mô này.

Kết luận

Khai thác trên cơ sở dữ liệu PlantTFDB đã xác định được gen mã hóa tiểu phần *NF-YC* ở cây sắn là họ đa gen, giống như ở các loài thực vật khác. Trong đó, 14 gen *MeNF-YC* đã được sàng lọc và phân tích.

Phân tích cho thấy, họ *MeNF-YC* có cấu trúc gen từ 1 đến 6 exon. Mặc dù có phần đa dạng về trật tự exon/intron nhưng một tỷ lệ lớn các gen *MeNF-YC* đều chỉ có một exon, tương tự như họ *NF-YC* ở các đối tượng cây trồng khác.

Khai thác dữ liệu transcriptome đã chỉ ra rằng, hầu hết các gen *MeNF-YC* có biểu hiện ở ít nhất một vị trí trên cây sắn trong điều kiện thường. Kết quả đã xác định được *MeNF-YC05* là gen có biểu hiện mạnh ở cả RAM, SAM,

OES và SR.

Các kết quả của nghiên cứu này sẽ cung cấp những dẫn liệu quan trọng cho các phân tích chức năng gen tiếp theo nhằm tìm kiếm các gen ứng viên liên quan đến quá trình đáp ứng bất lợi ở cây sắn.

TÀI LIỆU THAM KHẢO

- [1] S. Li, et al. (2016), “Genome-wide analysis of tomato NF-Y factors and their role in fruit ripening”, *BMC Genomics*, **17**, p.36.
- [2] M.E. Zanetti, C. Ripodas, A. Niebel (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.
- [3] Y. Chen, et al. (2018), “Genome-wide analysis of the NF-YB gene family in *Gossypium hirsutum* L. and characterization of the role of *GhDNF-YB22* in embryogenesis”, *Int. J. Mol. Sci.*, **19**, p.483.
- [4] T. Laloum, S. De Mita, P. Gamas, M. Baudin, A. Niebel (2013), “CCAAT-box binding transcription factors in plants: Y so many?”, *Trends Plant Sci.*, **18(3)**, pp.157-166.
- [5] W. Yang, Z. Lu, Y. Xiong, J. Yao (2017), “Genome-wide identification and co-expression network analysis of the *OsNF-Y* gene family in rice”, *Crop. J.*, **5(1)**, pp.21-31.
- [6] Z. Zhang, X. Li, C. Zhang, H. Zou, Z. Wu (2016), “Isolation, structural analysis, and expression characteristics of the maize nuclear factor Y gene families”, *Biochem. Biophys. Res. Commun.*, **478(2)**, pp.752-758.
- [7] T.N. Quach, et al. (2015), “Genome-wide expression analysis of soybean NF-Y genes reveals potential function in development and drought response”, *Mol. Genet. Genomics*, **290(3)**, pp.1095-1115.
- [8] Chu Đức Hà, Lê Thị Thảo, Lê Quỳnh Mai, Phạm Thị Lý Thu (2017), “Xác định các gen mã hóa Nuclear factor - YB trên sắn (*Manihot esculenta* Crantz) bằng công cụ tin sinh học”, *Tạp chí Khoa học, Đại học Quốc gia Hà Nội*, **33(1S)**, tr.133-137.
- [9] Chu Đức Hà, Lê Hoàng Thu Phương, Lê Thị Thảo, Hoàng Thị Thảo, Phạm Thị Lý Thu (2018), “Nghiên cứu cấu trúc của gen mã hóa Nuclear factor - YB ở sắn liên quan đến tính chống chịu điều kiện bất lợi”, *Tạp chí Khoa học Công nghệ Nông nghiệp Việt Nam*, **5(90)**, tr.5-9.
- [10] Duc Ha Chu, Thi Thuy Tam Do, Xuan Dac Le, Thi Ly Thu Pham (2017), “Genome-wide identification and annotation of the Nuclear factor YA gene family in cassava (*Manihot esculenta* Crantz)”, *Vietnam J. Sci., Technol. Eng.*, **59(3)**, pp.39-43.
- [11] J.V. Bredeson, et al. (2016), “Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity”, *Nat. Biotechnol.*, **34(5)**, pp.562-570.
- [12] M.C. Wilson, et al. (2017), “Gene expression atlas for the food security crop cassava”, *New Phytol.*, **213(4)**, pp.1632-1641.
- [13] J. Jin, F. Tian, D.-C. Yang, Y.-Q. Meng, L. Kong, J. Luo, G. Gao (2017), “PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants”, *Nucleic Acids Res.*, **45(D1)**, pp.D1040-D1045.
- [14] S. El-Gebali, et al. (2019), “The Pfam protein families database in 2019”, *Nucleic Acids Res.*, **47(D1)**, pp.D427-D432.
- [15] Y. Wang, et al. (2013), “PIECE: a database for plant gene structure comparison and evolution”, *Nucleic Acids Res.*, **41(D1)**, pp.D1159-D1166.
- [16] S. Kumar, G. Stecher, K. Tamura (2016), “MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets”, *Mol. Biol. Evol.*, **33(7)**, pp.1870-1874.