

NGHIÊN CỨU ĐẶC TÍNH LÝ HÓA VÀ XÂY DỰNG MÔ HÌNH CẤU TRÚC KHÔNG GIAN CỦA NHÓM NHÂN TỐ PHIÊN MÃ YẾU TỐ NHÂN Y Ở GIỐNG CHÓ DINGO (*Canis familiaris dingo*)

Chu Đức Hà¹, Nguyễn Thị Nho², Nguyễn Hữu Đức²,
Nguyễn Quốc Trung², Trần Thị Phương Liên³,
Lê Thị Ngọc Quỳnh⁴, Hà Thị Quyến¹,

TÓM TẮT

Yếu tố nhân Y (NF-Y) là một trong những nhóm protein điều hòa đóng vai trò nhân tố trung tâm trong các cơ chế điều hòa hoạt động của gen ở các loài trong sinh giới. Trong nghiên cứu này, thông tin về họ NF-Y ở giống chó Dingo (*Canis familiaris dingo*), bao gồm chú giải, đặc điểm lý hóa, cấu trúc gen, vùng domain đặc trưng, cấu trúc không gian, cấu trúc ARN, đã được tìm hiểu thông qua công cụ tin sinh học. Tiểu phần NF-YA gồm 347 axit amin và có trọng lượng phân tử 36,88 kDa, trong khi hai tiểu phần NF-YB và NF-YC có kích thước và trọng lượng phân tử lần lượt là 207 axit amin, 22,81 kDa và 402 axit amin, 44,69 kDa. Các tiểu phần không ổn định trong điều kiện ống nghiệm và có tính ưa nước, tương tự như ở các loài sinh vật khác. Phân tích cấu trúc cho thấy gen mã hóa NF-YA, NF-YB và NF-YC ở giống chó Dingo gồm 9, 6 và 10 exon. Kết quả của nghiên cứu này đã cung cấp những hiểu biết cơ bản về nhóm NF-Y ở giống chó Dingo.

Từ khóa: Chó Dingo, NF-Y, đặc tính lý hóa, cấu trúc, tin sinh học

I. ĐẶT VẤN ĐỀ

Nhân tố phiên mã (TF) là nhóm protein bám trên phân tử ADN, mã hóa bởi khoảng 7 - 10% số gen trong hệ gen của người và các loài động vật có vú, được chứng minh đóng vai trò thiết yếu trong điều hòa quá trình phiên mã, từ đó tham gia vào các quá trình sinh học diễn ra trong tế bào (Wilkinson *et al.*, 2017). Trong đó, yếu tố nhân - Y (NF-Y), một dạng phức tạp hợp của ba tiểu đơn vị (NF-YA, -YB và -YC), là nhóm TF được báo cáo tồn tại ở tất cả các loài sinh vật trong sinh giới (Li *et al.*, 2018). So với các nhóm TF khác, NF-Y xuất hiện sớm nhất trong quá trình tiến hóa, liên quan đến cơ chế đáp ứng và thích nghi của các loài sinh vật. Vì vậy, nghiên cứu về nhóm TF cơ bản này ở các loài động vật, cụ thể là phân tích đặc điểm cấu trúc và mô hình của ba tiểu phần, có thể cho phép có cái nhìn tổng quát hơn về sự tiến hóa trong sinh giới cũng như vai trò sinh học của NF-Y.

Dựa trên sự phát triển của công cụ giải trình tự gen thế hệ mới, thông tin di truyền của rất nhiều loài động vật đã được giải mã, từ đó cung cấp thông tin về các nhóm TF nói chung và NF-Y nói riêng. Đáng chú ý, các gen mã hóa TF NF-Y đã được nghiên cứu trên một số loài, điển hình như ruồi

giấm (*Drosophila melanogaster*) (Coustry *et al.*, 1998), chuột (*Rattus norvegicus*) (Yamanaka *et al.*, 2008) và người (*Homo sapiens*) (Currie *et al.*, 1998). Tuy nhiên, ghi nhận về TF NF-Y trên loài chó nói chung (*Canis familiaris*) (Field *et al.*, 2020) và giống chó Dingo (*C. familiaris dingo*) nói riêng (Sonu *et al.*, 2020) chưa được nghiên cứu trước đây.

Trong nghiên cứu này, kết quả xác định và phân tích đặc tính lý hóa của TF NF-Y giống chó Dingo đã được ghi nhận. Mức độ tương đồng trong cấu trúc của TF NF-Y giữa giống Dingo và loài chó đã được tìm hiểu, từ đó xây dựng mô hình cấu trúc không gian bằng thuật toán tin sinh học. Những dẫn liệu khoa học thu được từ nghiên cứu này đã định hướng cho phân tích chức năng gen mã hóa TF NF-Y ở giống chó Dingo.

II. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Hệ gen, hệ protein của loài *C. familiaris* (Mã số hệ tham chiếu: GCF_014441545.1) (Field *et al.*, 2020) và giống *C. familiaris dingo* (Mã số hệ tham chiếu: GCF_012295265.1) (Sonu *et al.*, 2020) được khai thác làm cơ sở dữ liệu cho nghiên cứu. Thông tin về cấu trúc vùng bảo thủ đặc trưng của NF-YA (Mã số:

¹ Khoa Công nghệ Nông nghiệp, Đại học Công nghệ, Đại học Quốc gia Hà Nội

² Khoa Công nghệ Sinh học, Học viện Nông nghiệp Việt Nam

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

⁴ Bộ môn Công nghệ Sinh học, Đại học Thủy lợi

PF02045), NF-YB và NF-YC (Mã số: PF00808) cho sinh giới (Li *et al.*, 2018) được sử dụng để trình tự đối chiếu cho các phép tìm kiếm trong nghiên cứu.

2.2. Phương pháp nghiên cứu

- Phương pháp tìm kiếm và xác định TF NF-Y: Vùng bảo thủ PF02045 và PF00808 đặc trưng cho ba tiểu phần (Li *et al.*, 2018) khai thác tại Pfam được sử dụng để đối chiếu trên nền tảng TF tích hợp (Zheng *et al.*, 2008) nhằm tìm các ứng viên tương đồng trong hệ protein của loài *C. familiaris* (Field *et al.*, 2020) và *C. familiaris dingo* (Sonu *et al.*, 2020). Giá trị cut-off được lựa chọn là E-value $\leq 1e-10$ theo ghi nhận trước đây (Chu *et al.*, 2018).

- Phương pháp chú giải TF NF-Y: Trình tự protein đầy đủ của TF NF-Y được đối chiếu trên hệ protein của và *C. familiaris dingo* (Sonu *et al.*, 2020) để thu thập mã định danh protein (proteinID), mã định danh gen (geneID) và mã định danh ARN (transcriptID). Trình tự đoạn mã hóa (CDS) và vùng gen (gDNA) của gen mã hóa ba tiểu đơn vị (định dạng .fasta) được khai thác cho các phân tích tiếp theo.

- Phương pháp xây dựng cấu trúc gen và vùng bảo thủ của TF NF-Y: Trật tự exon/intron được mô hình hóa bằng cách phân tích đoạn CDS và trình tự gDNA tương ứng trên giao diện hiển thị cấu trúc gen (GSDS) (Bo *et al.*, 2015), trong khi cấu trúc vùng bảo thủ của TF NF-Y được đánh giá bằng công cụ ClustalX.

- Phương pháp phân tích đặc tính lý hóa và dự đoán vị trí phân bố nội bào của TF NF-Y: Các đặc điểm cơ bản của TF NF-Y, bao gồm kích thước, trọng lượng phân tử (kDa), điểm đẳng điện, độ bất ổn định

và giá trị ưa nước trung bình, được khảo sát dựa trên trình tự protein đầy đủ của TF NF-Y thông qua công cụ ExPasy ProtParam. Các peptide tín hiệu đặc trưng cho ty thể (mTP), peptide tín hiệu cho hệ thống bao gói (SP) được phân tích bằng TargetP nhằm dự đoán sự phân bố nội bào của TF NF-Y.

- Phương pháp xây dựng mô hình cấu trúc không gian của TF NF-Y: Cấu trúc phân tử ARN được xây dựng dựa trên bản phiên mã của trình tự CDS bằng công cụ RNAstructure theo các bước trong nghiên cứu gần đây (Bellaousov *et al.*, 2013). Cấu trúc không gian bậc 3 của TF NF-Y được thiết kế dựa trên thuật toán mô phỏng trình tự protein đầy đủ với những NF-Y đã biết bằng công cụ Phyre2 (Kelley *et al.*, 2015).

2.3. Thời gian và địa điểm nghiên cứu

Nghiên cứu này được thực hiện từ tháng 11/2020 đến tháng 5/2021 tại Đại học Công nghệ (Đại học Quốc gia Hà Nội), Đại học Thủy lợi, Đại học Sư phạm Hà Nội 2, Viện Di truyền Nông nghiệp và Học viện Nông nghiệp Việt Nam.

III. KẾT QUẢ VÀ THẢO LUẬN

3.1. Xác định, chú giải và định danh TF NF-Y ở giống chó Dingo

Để tìm kiếm TF NF-Y ở hai đối tượng nghiên cứu, ba vùng đặc trưng cho ba tiểu đơn vị NF-YA, -YB và -YC tồn tại ở các loài trong sinh giới (Li *et al.*, 2018) được phân tích trên ITFP (Zheng *et al.*, 2008). Dựa trên độ tin cậy (E-value $\leq 1e-10$), ba tiểu phần cấu tạo TF NF-Y ở loài chó và giống Dingo đã được xác định (Bảng 1).

Bảng 1. Thông tin về TF NF-Y ở loài chó và giống chó Dingo

Loài	Tiểu đơn vị	ProteinID	GeneID	TranscriptID
<i>C. familiaris</i>	NF-YA	XP_005627357.1	474897	XM_005627300.4
	NF-YB	XP_005629363.1	475450	XM_005629306.3
	NF-YC	XP_022283452.1	475312	XM_022427744.1
<i>C. familiaris dingo</i>	NF-YA	XP_025274478.1	112641555	XM_025418693.2
	NF-YB	XP_025294767.1	112654441	XM_025438982.2
	NF-YC	XP_035554897.1	112642295	XM_035699004.1

Kết quả cho thấy ba tiểu phần cấu tạo TF NF-Y ở loài chó và giống chó Dingo được mã hóa bởi đơn gen, tương tự với các loài động vật khác, như *D. melanogaster* (Coustry *et al.*, 1998), *R. norvegicus*

(Yamanaka *et al.*, 2008) và *H. sapiens* (Currie *et al.*, 1998). Trong khi đó, họ TF NF-Y ở các loài thực vật nói chung, điển hình như cây *Cicer arietinum* lại được mã hóa bởi họ đa gen (Chu *et al.*, 2018). Điều

này có thể được giải thích do hiện tượng lặp hệ gen xảy ra ở một số loài sinh vật nhân chuẩn, bao gồm nấm men, lớp cá xương thật và thực vật dưới áp lực của quá trình tiến hóa, từ đó gây ra hiện tượng lặp trong các họ gen, trong khi ở hầu hết động vật bậc cao, các họ gen này thường tồn tại ở dạng đơn gen (Zheng *et al.*, 2018).

3.2. Phân tích đặc tính lý hóa và phân bố nội bào của TF NF-Y ở giống chó Dingo

Để xác định đặc điểm của TF NF-Y, trình tự protein đầy đủ được phân tích trên ExPasy

ProtParam. Kết quả cho thấy, tiểu đơn vị NF-YA có kích thước và trọng lượng phân tử lần lượt là 347 axit amin và 36,88 kDa, trong khi hai tiểu đơn vị NF-YB và NF-YC đạt 207 axit amin (tương ứng 22,81 kDa) và 402 axit amin (tương ứng 44,69 kDa). Giá trị điểm đẳng điện của TF NF-Y dao động từ khoảng axit (pI = 4,46) đến bazơ (pI = 9,99) (Bảng 2). Đáng chú ý, tất cả các giá trị độ bất ổn định của ba tiểu đơn vị của TF NF-Y đều lớn hơn 40, chứng tỏ các phân tử này đều không ổn định trong ống nghiệm, trong khi giá trị ưa nước trung bình đạt giá trị âm, chứng tỏ TF NF-Y có tính ưa nước.

Bảng 2. Khai thác đặc tính lý hóa của ba tiểu đơn vị của TF NF-Y ở giống chó Dingo

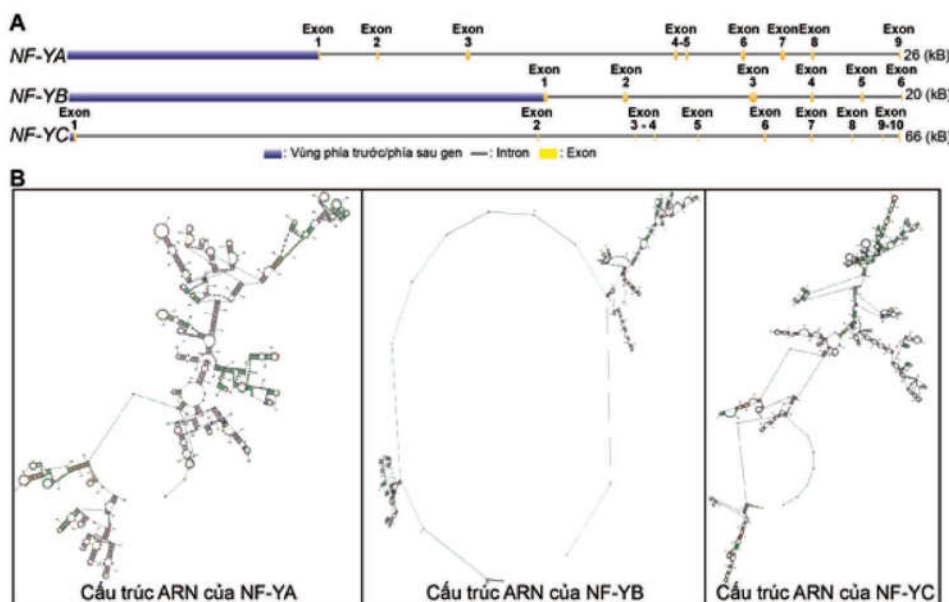
Tiểu đơn vị	Kích thước	mW	pI	II	GRAVY	mTP	SP	Tín hiệu khác
NF-YA	347	36,88	8,89	42,94	-0,37	0,076	0,046	0,932
NF-YB	207	22,81	4,46	42,16	-0,49	0,056	0,063	0,948
NF-YC	402	44,69	9,99	59,88	-0,55	0,659	0,034	0,425

Ghi chú: mW - Trọng lượng phân tử (kDa); pI - Điểm đẳng điện; II - Độ bất ổn định trong ống nghiệm; GRAVY - Giá trị ưa nước trung bình; mTP - Peptide tín hiệu đặc trưng cho ty thể; SP - Peptide tín hiệu cho hệ thống bao gói.

Các kết quả này đồng thuận với một số nghiên cứu trước đây trên động vật (Li *et al.*, 2018) và ngay cả trên đối tượng cây trồng (Chu *et al.*, 2018). Lấy ví dụ, ba tiểu phần NF-YA, NF-YB và NF-YC ở chuột đều bất ổn định trong ống nghiệm (giá trị bất ổn định lần lượt là 42,94; 47,80; 66,45) và có tính ưa nước (giá trị GRAVY đạt -0,37; -0,29; -0,48) (Yamanaka *et al.*, 2008). Trong khi đó, phân tích các

peptide tín hiệu đã chỉ ra rằng, tiểu phần NF-YC ở giống chó Dingo phân bố ở bào quan ty thể, trong khi không xác định được các peptide tín hiệu đặc trưng cho bào quan trên trình tự protein đầy đủ của hai tiểu phần NF-YA và NF-YB (Bảng 2).

3.3. Phân tích cấu trúc gen và mô hình hóa cấu trúc không gian của TF NF-Y ở giống chó Dingo



Hình 1. Cấu trúc (A) gen và (B) ARN của họ TF NF-Y ở giống chó Dingo

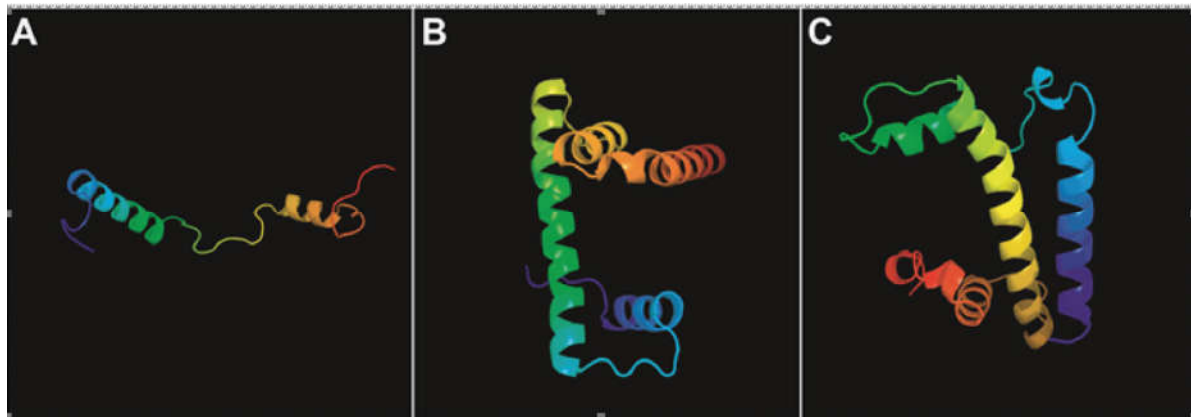
Trước hết, trình tự CDS và gDNA của từng gen được sử dụng để phân tích trật tự exon/intron. Kết quả cho thấy, gen mã hóa TF NF-Y có cấu trúc phân mảnh, với các đoạn exon xen kẽ vùng intron, tương tự như ghi nhận ở tất cả các loài sinh vật nhân chuẩn (Li *et al.*, 2018). Cụ thể, nghiên cứu đã xác định được gen *NF-YA* ở giống chó Dingo chứa 9 exon, trong khi gen *NF-YB* và *NF-YC* lần lượt bao gồm 6 và 10 exon (Hình 1A). Cấu trúc phân mảnh của họ gen mã hóa TF NF-Y ở giống chó Dingo nói riêng và động vật nói chung (Yamanaka *et al.*, 2008) thể hiện rõ rệt hơn so với ở giới thực vật. Ở thực vật bậc cao, ví dụ như loài *Cicer arietinum*, cấu trúc gen *NF-YA* được phân mảnh bởi 5 - 6 exon, trong khi gen *NF-YB* và *NF-YC* dao động từ 1 - 6 và 0 - 6 exon (Chu *et al.*, 2018). Hơn nữa, tỷ lệ CDS ở gen mã hóa TF NF-Y ở giống chó Dingo chỉ chiếm tỷ lệ rất thấp, từ 1,80 (*NF-YC*) đến 3,50% (*NF-YA*), tương tự như ở các loài động vật khác (Yamanaka *et al.*, 2008). Theo nghiên cứu trước đây, các loài động vật, do có khả năng di chuyển, đã phát triển hệ gen với kích thước bé hơn, gen có cấu trúc phân mảnh hơn (tỷ lệ CDS/gDNA thấp) để hạn chế tác động của đột biến điểm xảy ra (Kaplunovsky *et al.*, 2009).

Tiếp theo, cấu trúc của ARN và mô hình 3D của TF NF-Y ở giống chó Dingo lần lượt được trình bày ở hình 1B và 2. Theo đó, cấu trúc không gian của tiểu phần NF-YA ở có thể được thiết kế dựa trên mô hình 'c4aw1A' (Hình 2A), trong khi NF-YB và NF-YC lần lượt được xây dựng dựa trên hai mô hình, 'd1jfib' và 'c4g92C' (Hình 2B và C) (Narendja *et al.*,

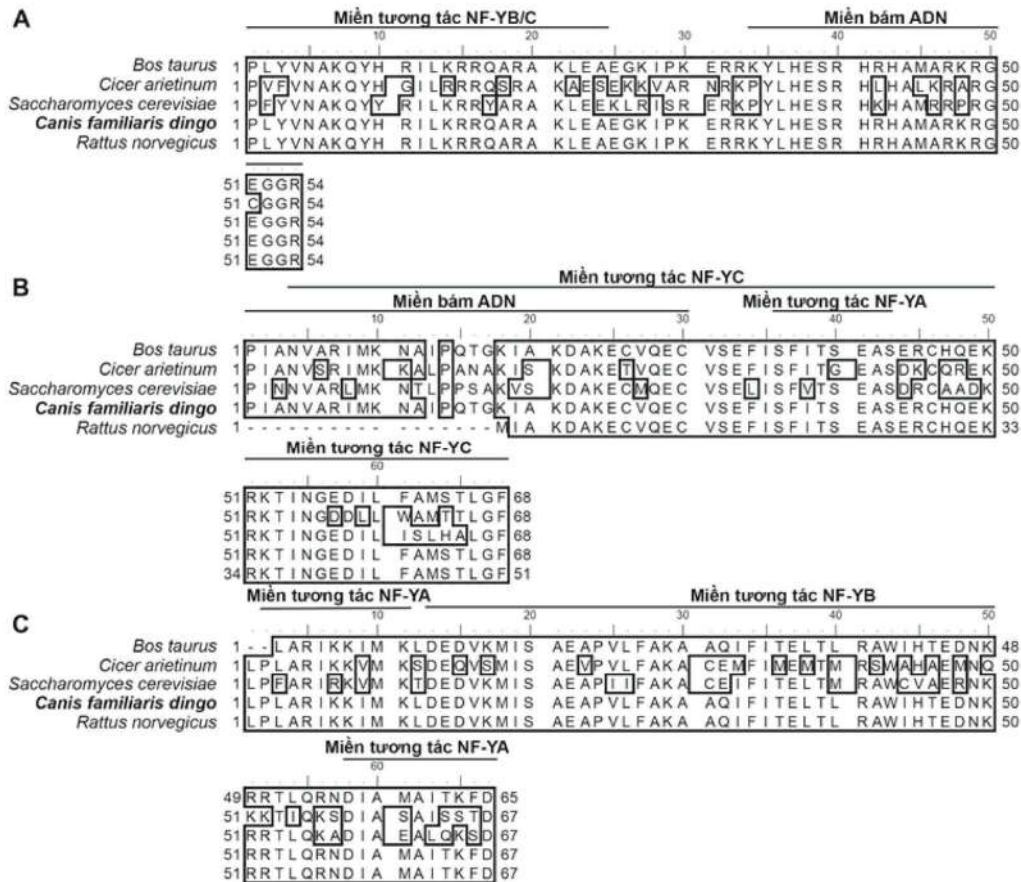
1999). Những kết quả này có thể đưa ra những giả thuyết cho việc xây dựng mô hình tương tác giữa ba tiểu phần để hình thành TF NF-Y ở giống chó Dingo.

Ngoài ra, cấu trúc vùng domain đặc trưng của TF NF-Y ở giống chó Dingo cũng đã được phân tích dựa trên mức độ bảo thủ so với các loài khác, bao gồm *R. norvegicus*, *B. taurus*, nấm men (*Saccharomyces cerevisiae*) và thực vật (đại diện là *C. arietinum*). Kết quả cho thấy vùng domain của ba tiểu phần NF-Y ở ba loài động vật được bảo thủ rất toàn vẹn, trong khi có một số sai khác được ghi nhận trong cấu trúc ở nấm men và đậu gà. Tuy nhiên, kết quả căn trình tự các vùng bảo thủ cũng đã chỉ ra một số amino acid đặc trưng cho chức năng của từng miền trong vùng domain, tương tự như ghi nhận trong nghiên cứu trước đây (Li *et al.*, 2018).

Vùng domain của tiểu phần NF-YA ở năm loài có cấu trúc gồm miền tương tác NF-YB/C, đặc trưng bởi N₅A₆, Q₈, R₁₁, L₁₃ và A₂₀, và miền bám ADN, đặc trưng bởi R₃₂, Y₃₅, H₃₇, R₄₀-H₄₁, H₄₃-A₄₄ và R₄₇. Vùng domain của tiểu phần NF-YB ở năm loài có cấu trúc gồm ba miền, bám ADN, tương tác với NF-YA và tương tác với NF-YC, chúng được đặc trưng bởi P₁, N₄, R₇, A₁₇-K₁₈, A₂₃-K₂₄, V₂₇-Q₂₈, S₃₆ và A₄₁. Vùng domain của tiểu phần NF-YC ở năm loài có cấu trúc gồm hai miền, tương tác với NF-YA và tương tác với NF-YB, được đặc trưng bởi A₃-R₄, K₇-K₈-I₉, L₂₆, A₂₈, I₃₅, L₃₈, I₅₉, A₆₂ và F₆₆.



Hình 2. Mô hình cấu trúc không gian của tiểu phần (A) NF-YA, (B) NF-YB và (C) NF-YC ở giống chó Dingo



Hình 3. Cấu trúc vùng bảo thủ của tiểu phần (A) NF-YA, (B) NF-YB và (C) NF-YC ở năm đối tượng nghiên cứu

IV. KẾT LUẬN VÀ ĐỀ NGHỊ

4.1. Kết luận

Ba tiểu phần của TF NF-Y ở giống chó Dingo được mã hóa bởi họ đơn gen. Trong đó, nghiên cứu đã xác định được ba tiểu phần NF-YA, NF-YB và NF-YC lần lượt có kích thước là 347 (36,88 kDa), 207 (22,81 kDa) và 402 axit amin (44,69 kDa), không ổn định và có tính ưa nước. Gen mã hóa TF NF-Y ở giống chó Dingo có cấu trúc phân mảnh, gồm 9 (NF-YA), 6 (NF-YB) và 10 exon (NF-YC). Ba tiểu phần của TF NF-Y ở giống chó Dingo chia sẻ cấu trúc domain đặc trưng tương tự như các loài khác trong sinh giới, gồm những miền chức năng với các axit amin bảo thủ.

4.2. Đề nghị

Nghiên cứu này sẽ tiếp tục nhằm tìm hiểu về cơ chế tương tác của các tiểu phần và cơ chế điều hòa gen của TF NF-Y để làm rõ hơn khả năng thích nghi môi trường của giống chó Dingo.

TÀI LIỆU THAM KHẢO

- Bellaousov, S., Reuter, J.S., Seetin, M.G., Mathews, D.H., 2013. RNAstructure: Web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res.*, 41: 471-474.
- Bo, H., Jinpu, J., An-Yuan, G., et al., 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*, 31 (8): 1296-1297.
- Country, F., Sinha, S., Maity, S.N., Crombrugge, B., 1998. The two activation domains of the CCAAT-binding factor CBF interact with the dTAFIII110 component of the Drosophila TFIID complex. *Biochem J.*, 331: 291-297.
- Currie, R., 1998. Biochemical characterization of the NF-Y transcription factor complex during B lymphocyte development. *J. Biol. Chem.*, 273 (29): 18220-18229.
- Field, M.A., Rosen, B.D., Dudchenko, O., et al., 2020. Canfam_GSD: De novo chromosome-length genome assembly of the German Shepherd Dog (*Canis lupus familiaris*) using a combination of long reads, optical mapping, and Hi-C. *Gigascience*, 9 (4): giaa027.

- Kaplunovsky, A., Ivashchenko, A., Bolshoy, A., 2009. Statistics of exon lengths in animals, plants, fungi, and protists. *Int. J. Biol. Life Sci.*, 5: 139-144.
- Kelley, L., Mezulis, S., Yates, C., *et al.*, 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, 10: 845-858.
- Li, G., Zhao, H., Wang, L., *et al.*, 2018. The animal nuclear factor Y: an enigmatic and important heterotrimeric transcription factor. *Am. J. Cancer Res.*, 8 (7): 1106-1125.
- Narendja, F.M., Davis, M.A., Hynes, M.J., (1999). AnCF, the CCAAT binding complex of *Aspergillus nidulans*, is essential for the formation of a DNase I-hypersensitive site in the 5' region of the amdS gene. *Mol. Cell. Biol.*, 19 (10): 6523-6531.
- Sonu, Y., Olga, D., Meera, E., *et al.*, 2020. Desert Dingo (*Canis lupus dingo*) genome provides insights into their role in the Australian ecosystem. *bioRxiv*, 11 (15): 384057.
- Wilkinson, A.C., Nakauchi, H., Göttgens, B., 2017. Mammalian transcription factor networks: Recent advances in interrogating biological complexity. *Cell Sys.*, 5 (4): 319-331.
- Yamanaka, T., Miyazaki, H., Oyama, F., *et al.*, 2008. Mutant huntingtin reduces HSP70 expression through the sequestration of NF-Y transcription factor. *EMBO J.*, 27 (6): 827-839.
- Zheng, G., Tu, K., Yang, Q., *et al.*, 2008. ITFP: an integrated platform of mammalian transcription factors. *Bioinformatics*, 24 (20): 2416-2417.
- Zheng, L., George, P., Sally, R., *et al.*, 2018. Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci.*, 115 (18): 4713-4718.

Analysis of physic-chemical properties and construction of the structural model of nuclear factor Y transcription factors in Dingo (*Canis familiaris dingo*)

Chu Duc Ha, Nguyen Thi Nho, Nguyen Huu Duc,
Nguyen Quoc Trung, Tran Thi Phuong Lien,
Le Thi Ngoc Quynh, Ha Thi Quyen

Abstract

Nuclear factor - Y (NF-Y) has been considered one of the major regulatory proteins that play a central role in regulating gene expression in the kingdom. Here, the information of NF-Y in Dingo (*Canis familiaris dingo*), including annotation, physic-chemical characteristics, gene structure, conserved domains, 3D models and RNA structures, has been investigated by the bioinformatic approach. We found that the NF-YA subunit comprises 347 amino acids in length and 36.88 kDa in molecular weight, while NF-YB and NF-YC subunits range from 207 (22.81 kDa) to 402 amino acids (44.69 kDa), respectively. Additionally, these proteins are unstable in the test tube and hydrophilic molecules as similar as in other organisms. Next, genes encoding NF-YA, NF-YB and NF-YC of Dingo contain 9, 6 and 10 exons, respectively. Taken together, our study could provide a basic understanding of the NF-Y in Dingo.

Keywords: Dingo, NF-Y, physic-chemical properties, structure, bioinformatics

Ngày nhận bài: 03/5/2021
Ngày phản biện: 18/5/2021

Người phản biện: TS. Nguyễn Hoàng Thịnh
Ngày duyệt đăng: 04/6/2021