CÔNG NGHỆ GEN

BIỂU HIỆN CỦA MỘT SỐ GEN LIÊN QUAN ĐẾN KHẢ NĂNG CHỊU HẠN Ở MỘT SỐ DÒNG CHÈ TẠI PHÚ THỌ

Nguyễn Hồng Chiên¹*, Lưu Ngọc Quyến¹, Nguyễn Thị Kim Linh¹, Mai Thị Phương Nga²

¹Viện Khoa học kỹ thuật Nông Lâm nghiệp miền núi phía Bắc

²Trường Đại học Khoa học và Công nghệ Hà Nội, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Dựa vào các kết quả nghiên cứu trước đây, 05 gen (GolS3, RHL41, bHLH102, DHN1 và DREB26) liên quan đến khả năng chịu hạn đã được lựa chọn để phân tích biểu hiện gen trong điều kiện hạn ở một số dòng chè trồng tại Phú Thọ. Cây chè (*Camellia sinensis*) tuổi 1 được xử lý hạn với dung dịch PEG 6000 20% (w/v) trong các khoảng thời gian 0h (đối chứng), 12h, 24h và 72h. Sau xử lý, mẫu lá trưởng thành (lá thứ 4-5) của ba dòng chè LVC4C1, PHN2C1 và VSHC7C2 được dùng để tách chiết RNA tổng số. Kết quả phân tích mức độ biểu hiện của các gen bằng kĩ thuật qRT-PCR cho thấy, tất cả 05 gen đều biểu hiện tăng ở điều kiện hạn, mức độ biểu hiện của các gen này khác nhau ở cùng một dòng chè và cũng khác nhau giữa các dòng chè. Đối với dòng chè LVC4C1, mức độ biểu hiện tương đối của 05 gen đều tăng ở tất cả các thời điểm xử lý, trong khi chỉ có 2 trong số 5 gen có mức độ biểu hiện tương đối tăng ở tất cả các thời điểm xử lý ở hai dòng chè PHN2C1 và VSHC7C2. Ngoài ra, trong ba dòng chè nghiên cứu, dòng LVC4C1 có 4/5 gen (RHL41, DHN1, bHLH102 và DREB26) có mức độ biểu hiện tương đối tăng mạnh nhất sau 72h xử lý. Những kết quả này có thể liên quan đến khả năng chịu hạn của dòng chè LVC4C1 tốt hơn so với hai dòng chè PHN2C1 và VSHC7C2.

Từ khóa: Cây chè, chống chịu hạn, mức độ biểu hiện gen, qRT-PCR, yếu tố phiên mã.

MỞ ĐẦU

Có nguồn gốc từ châu Á, cây chè (*Camellia sinensis* (L.) O. Kuntze) phân bố rộng rãi trên khắp thế giới, với hơn 5 triệu ha và được trồng ở hơn 60 quốc gia trên toàn thế giới. Cây chècó vị trí quan trọng trong nền kinh tế của nhiều nước, đem lại nguồn thu nhập và ngoại tệ quan trọng của các nước như Ấn Độ, Sri Lanka, Kenya, Malawi... Sản xuất chè chủ yếu tập trung ở châu Á (trong đó, sản xuất chè ở Trung Quốc và Nhật Bản chiếm khoảng 23%, Ấn Độ, Sri Lanka và Bangladesh, chiếm khoảng 44% tổng sản lượng chè trên thế giới). Năm 2023, các nước xuất khẩu chè chính gồm có Trung Quốc, Kenya, Sri Lanka, Ấn Độ. Việt Nam đứng thứ 10 trong số những nước xuất khẩu chè lớn nhất thế giới (Ridder, 2024).

Ở Việt Nam, chè là một trong những cây trồng quan trọng nhất, phục vụ tiêu dùng trong nước cũng như xuất khẩu. Năm 2023, diện tích trồng chè ở Việt Nam khoảng 123.000 ha với sản lượng chè khô đạt hơn 190 nghìn tấn. Cây chè được trồng chủ yếu ở các tỉnh miền núi phía Bắc và Tây Nguyên, đóng vai trò thiết yếu trong việc cung cấp sinh kế và kinh tế bền vững cho các vùng này. Ngành chè sử dụng khoảng 1,5 triệu lao động và kể từ năm 2010, đóng góp kinh tế hàng năm của xuất khẩuchè là hơn 200 triệu USD/năm (Van Ho *et al.*, 2019).

Hạn hán là một hạn chế lớn đối với sinh trưởng, năng suất và chất lượng của cây chè. Hạn hán có thể làm giảm sản lượng chè từ 14-33% và gây chết cây từ 6-19%. Hạn hán có thể dẫn đến giảm thế năng nước của các mô và tạo ra các loại oxy phản ứng tích tụ, gây hại nghiêm trọng cho các thành phần tế bào khác nhau. Các phản ứng của thực vật rất phức tạp, đặc biệt là ở cây thân gỗ lâu năm, có hàng trăm gen liên quan đến những phản ứng này. Các nghiên cứu nhằm làm sáng tỏ cơ chế phản ứng của cây chè với điều kiện khô hạn cũng đã được thực hiện. Nhiều yếu tố phiên mã và gen liên quan đến chất chuyển hóa được chứng minh là có liên quan đến phản ứng với stress hạn của thực vật. Ví dụ, các yếu tố phiên mã (Transcription Factors) CBF và DHN đều tham gia vào phản ứng lạnh, hạn hán và các stress phi sinh học khác (Hu et al., 2020). Một số nhiều yếu tố phiên mã khác (WRKY, bHLH, NAC, HSP, LEA) đã được chứng minh là được kích hoạt ở cây chè để ứng phó với lanh và han hán. Phân tích dữ liệu phiên mã ở cây chè cho thấy 12 họ TF (AP2/EREBP, bHLH, bZIP, HD-ZIP, HSF, MYB, NAC, WRKY, zinc-finger protein, SCL, ARR, và SPL) có thể đóng vai trò quan trọng trong việc cây chè phản ứng với hạn hán (Liu et al., 2016). Gần đây, Samarina và đồng tác giả (2020) sử dụng phương pháp phát hiện tương đồng toàn diện (kết hợp dữ liệu từ thư viện gen và phân tích transcriptome) họ đã xác định được 45 gen liên quan đến phản ứng với lanh và han hán ở cây chè, trong đó có 9 gene lần đầu được phát hiện và 36 gen đã được công bố. Sử dung gRT-PCR, nhóm nghiên cứu đã xác đinh được các gen GolS1 và GolS3 (galactinol syntases), RHL41 (zinc finger protein), DREB26 (dehydration response element-binding protein), DHN1, CAU1, Hydrolase22. BMY5 và bHLH102... được điều điều hòa biểu hiện tăng ở điều kiện hạn.

Do tác động của biến đổi khí hậu nên khí hậu ở Việt Nam những năm gần đây ngày càng bất thường, thể hiện qua các biểu hiện dị thường của các yếu tố nhiệt độ, lượng mưa, mực nước biển dâng và các hiện tượng thời tiết

cực đoan. Hạn hán xảy ra thường xuyên hơn đã tác động không nhỏ đến ngành sản xuất chè. Vì vậy, công tác nghiên cứu chọn tạo ra các giống chè có khả năng chịu hạn tốt là rất cần thiết. Tuy nhiên, cho đến nay, các nghiên cứu sâu về các gen liên quan đến khả năng chống chịu stress vô sinh nói chung và chống chịu hạn nói riêng trên cây chè vẫn còn rất ít. Nghiên cứu này phân tích biểu hiện của một số gen liên quan đến khả năng chịu hạn ở 3 dòng chè có khả năng chịu hạn khác nhau trồng tại Phú Thọ.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Cây chè tuổi 1 của ba dòng chè (*Camellia sinensis*): LVC4C1, PHN2C1 và VSHC7C2. Các cây được tạo ra bằng phương pháp giâm hom tại Viện Khoa học kĩ thuật nông lâm nghiệp miền núi phía Bắc. Dòng LVC4C1 có khả năng chịu hạn tốt, trong khi dòng hai dòng chè PHN2C1 và VSHC7C2 có khả năng chịu hạn ở mức trung bình khá.

Phương pháp

Xử lý hạn

Cây con được loại bỏ sạch giá thể ở phần rễ cây, sau đó chuyển vào dung dịch PEG 6000 20% (w/v), dung dịch chỉ gồm PEG 6000 và nước cất. Sau khoảng thời gian 0h (đối chứng), 12h, 24h và 72h tiến hành thu mẫu lá. Mỗi dòng chè thu khoảng 5-6 lá từ 03 cây. Mẫu lá được nhúng ngay vào N₂ lỏng, nghiền thành bột mịn, bảo quản trong tủ -80°C trước khi tách chiết RNA.

Tách chiết RNA tổng số và tổng hợp cDNA:

RNA tổng số được tách chiết từ lá chè theo phương pháp Muoki đã mô tả (Muoki *et al.*, 2012). Sau khi tách chiết RNA tổng số, các mẫu RNA được đánh giá cảm quan trên gel Agarose 1% bằng phương pháp điện di. Độ tinh sạch và nồng độ của RNA được kiểm tra và đánh giá bằng máy đo nồng độ RNA NanoDrop với tỷ lệ bước sóng A260/280 và A260/230. Các mẫu RNA sau đó được xử lý bằng bộ kit Turbo DNA-freeTM để loại bỏ DNA. Ở bước tiếp theo, phản ứng PCR được thực hiện với mồi β-tubulin (F- AGGTTCTGGGATGGGTACCT; R-CACATTGTTAGGGATCCACTCCAC)để xác nhận DNA đã bị cắt hoàn toàn bởi enzym DNase. Nếu mẫu RNA còn lẫn DNA, lúc này DNA sẽ là đoạn khuôn và sản phẩm PCR sau điện di sẽ xuất hiện vạch DNA có kích thước khoảng 2000 bp, nếu mẫu RNA sạch DNA thì không xuất hiện vạch DNA này. Sau đó, RNA tinh sạch được sử dụng để phiên mã ngược nhằm tạo ra DNA bổ sung (cDNA) bằng cách sử dụng cDNA Reverse Transcription Kit (Thermo Fisher Scientific).

qRT-PCR:

Mỗi ống phản ứng chứa 5 µL SYBR Green qPCR Mastermix 2X, 0,2 µL mồi mỗi loại, 1,0 µL cDNA và bổ sung nước đến thể tích 10 µL. Phản ứng real-time PCR được thực hiện theo chu trình nhiệt như sau: 95° C-10 phút, sau đó 40 chu kì gồm 2 bước: 95° C-15 giây, 60° C-30 giây và 72° C-30 giây, cuối cùng là bước kéo dài ở 72° C trong 10 phút. Mức độ biểu hiện tương đối của các gen được tính theo phương pháp $2^{-\Delta\Delta CT}$ (Livak, Schmittgen, 2001).

Mức ý nghĩa của so sánh mức độ biểu hiện tương đối được xác định giữa hai nhóm giá trị biểu hiện ở thời điểm ban đầu (0h) và các thời điểm xử lý hạn bằng hàm Tukey Test với độ tin cậy 95% với p < 0,05, phần mềm R 4.3.3.

TT	Tên mồi	Trình tự nucleotide (5'-3')	Gen đích	Kích thước				
1	GolS3	F-TGCTGGCATGTTTGTATTTGAG	Galactinol Synthase 3	142				
		R- IGGIGGGATAGGCIIGTAAAC						
		F- AACTGCTTGATGCTCCTCTC		105				
2	RHL41	R- AAGGTCTCCTCCCATTA	Responsive to high light 41	165				
		F- ACACCGATGAGGTGGAGGTA	Debuddie 4	4.40				
3	DHN1	R- AATCCTCGAACTTGGGCTCT	Denyarin 1	149				
		F- AGGCTCGACACAGACAATTC	bill literan existing factor bill 1400	404				
4	bHLH102	R- GACGATGACGACGATGATGAA	BHLH transcription factor BHLH 102	161				
-	DDEDOO	F- CCAGAGCCAAACAAAGCAATAC	Dehydration response element-binding	454				
5	DREB26	R- GGGTGGAATAAGAGCCTAACC	protein 26	151				
TT	Tên mồi	Trình tự nucleotide (5'-3')	Gen tham khảo	Kích thước				
4		F- GGCGGATCAAGTGTTGGAAGGGAG	TATA hav hinding protain gone	166				
1	IRA	R- ACGCTTGGGATTGTATTCGGCATTA	I A I A-box binding protein gene	100				

Bảng 1. Trình tự các mồi được sử dụng trong nghiên cứu

KẾT QUẢ VÀ THẢO LUẬN

Tách chiết và kiểm tra chất lượng RNA

Mẫu lá của 3 dòng chè ở các điều kiện: đối chứng (không xử lý hạn), xử lý hạn 12h, 24h và 72h được dùng để tách chiết RNA tổng số. Tách chiết RNA chè theo quy trình của Muoki (2012) cho kết quả tốt, RNA thu được nguyên vẹn. Tiếp theo, RNA tổng số được làm sạch để loại bỏ các protein, ion muối và các tạp chất khác, sau đó xử lý với enzym DNase để loại bỏ các sợi DNA mạch đôi.Kết quả sau đó được kiểm tra lại bằng điện di trên gel agarose (Hình 1).



Hình 1. Kết quả kiểm tra ARN tổng số trên gel agarose Ladder 1kb, mẫu ARN của các dòng chè được đánh số theo thứ tự từ 1 đến 12

Để khẳng định các mẫu RNA tổng số đã được loại bỏ hoàn toàn sợi DNA mạch đôi, phản ứng PCR được thực hiện với mồi β-tubulin. Gen β-tubulin trong hệ gen có kích thước khá lớn, với nhiều intron có kích thước khác nhau, nhưng sau quá trình phiên mã, các intron được loại bỏ, kích thước của gen này còn khoảng 1344bp (Wu *et al.*, 2016). Cặp mồi nhân gen β-tubulin đã được thiết kế để khuyếch đại gen này với kích thước khoảng 2000 bp khi vẫn còn intron (nếu sợi khuôn là DNA) và chỉ dài hơn 600bp khi đã loại bỏ intron (nếu sợi khuôn là DNA). Kết quả cho thấy, sau khi xử lý với enzym DNase, RNA của các mẫu đã sạch hoàn toàn, không còn lẫn DNA (Hình 2).

Độ tinh sạch của RNA cũng được kiểm tra, kết quả thu được tỷ lệ A260/280 và OD260/OD230 của 12 mẫu lần lượt nằm trong khoảng 1,8-1,9 và 1,9-2,1 cho thấy RNA tổng số thu được là tinh sạch. Nồng độ RNA tổng số ở các mẫu nằm trong khoảng >300-1000 ng/μL, đạt yêu cầu cho các thí nghiệm tiếp theo.



Hình 2. Kết quả kiểm tra RNA bằng PCR với mồi β-tubulin Sản phẩm PCR được điện di trên gel agarose 1%. Ladder 1kb; mẫu RNA của các dòng chè được đánh số theo thứ tự từ 1 đến 12; ĐC-: Đối chứng âm (không có đoạn khuôn); ĐC+: Đối chứng dương (đoạn khuôn là DNA của giống chè tím Nhật Bản)

Đánh giá mức độ biểu hiện của 5 gen đích

Chúng tôi lựa chọn 05 gen đích liên quan đến khả năng chịu hạn dựa vào các kết quả nghiên cứu trước đây. Sự biểu hiện tăng của các gen GoIS3, RHL41, bHLH102, DHN1 và DREB26 trong phản ứng với điều kiện hạn đã được báo cáo (Samarina *et al.*, 2020). Trong nghiên cứu này, gen CsTBP (TATA-box binding protein gene) được sử dụng làm gen tham chiếu. Gen CsTBP đã được chứng minh là một trong các gen tham chiếu có sự biểu hiện ổn định ở các mô, các giai đoạn phát triển của lá và dưới một số điều kiện kích thích ở các giống chè khác nhau (Wu *et al.*, 2016).

Hầu hết các gen trong nghiên cứu này đều có mức biểu hiện tăng ở các dòng chè tại các thời điểm thu mẫu sau xử lý hạn so với mẫu đối chứng (không xử lý hạn). Gen RHL41 biểu hiện mạnh nhất ở dòng chè LVC4C1, trong khi đó, gen GoIS3 biểu hiện mạnh nhất ở dòng VSHC7C2 sau 72h xử lý hạn (Hình 3).



Hình 3. Mức độ biểu hiện của 05 gen liên quan đến khả năng chịu hạn ở 03 dòng chè LVC4C1, PHN2C1 và VSHC7C2 ở các thời điểm xử lý hạn: 0h, 12h, 24h và 72h (lần lượt từ trái sang phải)

Các chữ cái khác nhau biểu hiện sự khác nhau có ý nghĩa thống kê giữa các thời điểm xử lý của cùng 1 gen với P<0,05

Để so sánh mức độ biểu hiện mạnh hay yếu của các gen ở từng dòng chè và giữa các dòng chè, mức độ biểu hiện tương đối của các gen đích ở thời gian xử lý hạn 0h (đối chứng) được đưa về giá trị là 1. Kết quả cho thấy, ở dòng chè LVC4C1, gen DHN1, GolS3 và RHL41 biểu hiện mạnh nhất sau 72h xử lý (tăng lần lượt là 5,9; 9,25 và 39,38 lần), trong khi hai gen bHLH102 và DREB26 biểu hiện mạnh nhất sau 12h xử lý hạn (tăng lần lượt là 5,9; 9,25 và 9,35 lần) (Hình 4). Tương tự, ở dòng chè VSHC7C2, 03 gen DHN1, GolS3 và RHL41 biểu hiện mạnh nhất sau 72h xử lý (tăng lần lượt là 3,93, 28,19 và 12,84 lần), gen DREB26 biểu hiện mạnh nhất sau 12h xử lý (tăng 3,19 lần), trái lại, gen bHLH102 biểu hiện giảm ở tất cả các thời gian xử lý 12h, 24h và 72h (giảm lần lượt là 0,53, 0,32 và 0,20 lần).Không giống hai dòng chè trên, các gen liên quan đến chịu hạn có kiểu biểu hiện khá khác ở dòng PHN2C1. Cụ thể, 04 gen RHL41, DHN1, bHLH102 và DREB26 biểu hiện mạnh nhất sau 12h xử lý (tăng 3,03 lần).

Galactinol synthase (GolS) đóng vai trò quan trọng và không thể thiếu trong các bước khởi đầu và quan trọng trong việc hình thành con đường sinh tổng hợp raffinose, có chức năngnhư chất bảo vệ thẩm thấu trong tế bào thực vật (Salvi *et al.*, 2020). Kết quả nghiên cứu của Samarina và đồng tác giả (2020) đã xác nhận cơ chế bảo vệ salicylate khỏisự tấn công của các gốc hydroxyl qua trung gian galactinol và raffinose là rất quan trọng đối với chống chịuhạn và lạnh ở cây chè.

bHLH (Basic helix-loop-helix) là một trong những họ nhân tố phiên mã quan trọng nhất và lớn nhất ở thực vật, đóng vai trò quan trọng trong việc điều hòa sinh trưởng và phát triển, phản ứng với stress.Ở cây *A. thaliana*, gen bHLH102 mã hóa cho một protein tín hiệu và kết quả nghiên cứu của Samarina và đồng tác giả (2020) đã xác nhận chức năng gen này là cần thiết ở cây chè.

DHN đã được chứng minh là gen downstream của gen CBFs và được điều hòa biểu hiện tăng trong điều kiện stress hạn và lạnh. Protein dehydrin (DHN) rất quan trọng đối với ổn định màng và ngăn ngừa sự kết tụ protein (Nguyen Hong Chien, 2016).

Samarina và đồng tác giả (2020) báo cáo rằng gen GoIS3, bHLH102 và DHN1 được cảm ứng bởi các kích thích căng thẳng và cho thấy mối tương quan tích cực đáng kể với khả năng chống chịu của cây đối với cả stress lạnh và hạn, trái lại, RHL41 và DREB26chỉ được điều hòa biểu hiện tăng ở điều kiện hạn. Kết quả này phù hợp với các kết quả trong nghiên cứu của chúng tôi khi tất cả 05 gen đều biểu hiện tăng ở điều kiện hạn. Ở dòng chè LVC4C1, các gen RHL41, DHN1, bHLH102 và DREB26 biểu hiện tăng ở tất cả các thời điểm xử lý hạn và tăng mạnh nhất sau 72h, khi so sánh với hai dòng chè PHN2C1 và VSHC7C2.



Hình 4. Biểu đồ so sánh mức độ biểu hiện tương đối của 05 gen liên quan đến chịu hạn sau 0h, 12h, 24h và 72h xử lý với PEG6000 ở 03 dòng chè LVC4C1, PHN2C1 và VSHC7C2

DREB26 là một đại diện của họ gen AP2/ERF, làm trung gian điều hòa sự phiên mã của các gen phản ứng với stress thẩm thấu (Parmar *et al.*, 2019). DREB26 thuộc nhóm phụ A5, mã hóa cho các protein ức chế (repressor proteins) để ngăn cản sự biểu hiện của các yếu tố phiên mã DREB khác. Điều này có nghĩa gen DREB26 có thể ngăn chặn các gen bảo vệ vàcác gen được cảm ứng bởi stress khi ở điều kiện bình thường. Trong nghiên cứu này, gen DREB26 biểu hiện giảm sau 24h và 72h xử lý hạn ở cả hai dòng chè PHN2C1 và VSHC7C2.

RHL41(liên quan đến zinc-finger protein Zat12), là mộtđại diện cho một nhóm nhỏ gen có phản ứng với nhiều stress môi trường khác nhau. Một số tác giả báo cáo rằng Zat12 hoạt động như một chất ức chế phiên mã CBF. Nghiên cứu của chúng tôi cho thấy mức độ biểu hiện của RHL41tăng lên mạnh mẽ trong thời gian xử lý hạn, cho thấy gen này có thể có chức năng cụ thể đối với phản ứng với stress hạn ở cây chè.

Mức độ biểu hiện tương đối của tất cả 05 gen đều tăng ở dòng chè LVC4C1 có thể liên quan đến khả năng chịu hạn tốt hơn của dòng chè này so với hai dòng chè PHN2C1 và VSHC7C2. Khi phân tích 11 chỉ tiêu giải phẫu cấu trúc lá để đánh giá khả năng chịu hạn của các dòng chè, dòng LVC4C1 cũng biểu hiện khả năng chịu hạn tốt hơn (giá trị hàm thành viên trung bình đạt 0,66) khi so sánh với hai dòng chè PHN2C1 (0,5) và VSHC7C2 (0,34).

KÉT LUẬN

Mức độ biểu hiện của 05 gen liên quan đến khả năng chịu hạn khác nhau ở cùng một dòng chè và cũng khác nhau giữa các dòng chè. Đối với dòng chè LVC4C1, mức độ biểu hiện tương đối của 05 gen đều tăng ở tất cả các thời điểm xử lý, trong khi chỉ có 2 trong số 5 gen có mức độ biểu hiện tương đối tăng ở tất cả các thời điểm xử lýở hai dòng chè PHN2C1 và VSHC7C2. Ngoài ra, trong ba dòng chè nghiên cứu, dòng LVC4C1 có 4/5 gen (RHL41, DHN1, bHLH102 và DREB26) có mức độ biểu hiện tương đối tăng mạnh nhất sau 72h xử lý.Những kết quả này có thể liên quan đến khả năng chịu hạn của dòng chè LVC4C1 tốt hơn so với hai dòng chè PHN2C1 và VSHC7C2.

Lời cám ơn: Công trình này được thực hiện với sự hỗ trợ kinh phí của đề tài "Nghiên cứu chọn tạo giống chè mới và giải pháp kỹ thuật canh tác phù hợp cho sản xuất hàng hóa tại một số vùng trồng chính".

TÀI LIỆU THAM KHẢO

Cheruiyot EK, Mumera LM, Ng'etich WK, Hassanali A and Wachira FN (2010). High fertilizer rates increase susceptibility of tea to water stress. *J Plant Nutr*, 33, 115–129.

Ridder M (2024). Leading tea exporters worldwide 2023. https://www.statista.com/statistics/264189/main-export-countries-for-tea-worldwide/

Hu Z, Ban Q, Hao J, Zhu X, Cheng Y, Mao J, Lin M, Xia E, Li Y (2020). Genome-wide characterization of the C-repeat binding factor (CBF) Gene family involved in the response to abiotic stresses in tea plant (Camellia sinensis). *Front Plant Sci*, 11:921.

Liu SC, Jin JQ, Ma JQ, Yao MZ, Ma CL, Li CF, et al. (2016). Transcriptomic analysis of tea plant responding to drought stress and recovery. PLoS ONE 11:e0147306.

Livak KJ, Schmittgen TD (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDCT method. *Methods*, 25(4): 402-408.

Muoki RC, Paul A, Kumari A, Singh K, Kumar S (2012). An improved protocol for the isolation of RNA from roots of tea (Camellia sinensis (L.) O. Kuntze). *Mol Biotechnol*, 52(1):82-8.

Nguyen Hong Chien (2016). Eucalyptus DREB regulation pathway: control of abiotic stress tolerance, plant development and wood formation. *PhD dissertation. LRSV Research Laboratory in Plant Sciences, Paul Sabatier-Toulouse III.*

Parmar R, Seth R, Singh P, Singh G, Kumar S, and Sharma RK (2019). Transcriptional profiling of contrasting genotypes revealed key candidates and nucleotide variations for drought dissection in *Camellia sinensis* (L.) O. Kuntze. *Sci Rep*, 9, 1-12.

Salvi P, Kamble NU, Majee M (2020). Ectopic over-expression of ABA-responsive Chickpea galactinol synthase (CaGolS) gene results in improved tolerance to dehydration stress by modulating ROS scavenging. *Environ Exp Bot*, 171, 103957.

Samarina LS, Bobrovskikh AV, Doroshkov AV, Malyukova LS, Matskiv AO, Rakhmangulov RS, Koninskaya NG, Malyarovskaya VI, Tong W, Xia E, Manakhova KA, Ryndin AV and Orlov YL (2020). Comparative expression analysis of stress-inducible candidate genes in response to cold and drought in tea plant [*Camellia sinensis* (L.) Kuntze]. *Front Genet*, 11:611283.

Van Ho B, Nanseki T & Chomei Y (2019). Profit efficiency of tea farmers: case study of safe and conventional farms in Northern Vietnam. *Environment, Development and Sustainability,* 21, 1695-1713.

Wu ZJ, Tian C, Jiang Q et al. (2016). Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (*Camellia sinensis*). Sci Rep, 6, 19748.

EXPRESSION OF SOME GENES RELATED TO DROUGHT TOLERANCE IN SOME TEA LINES IN PHU THO

Nguyen Hong Chien¹*, Luu Ngoc Quyen¹, Nguyen Thi Kim Linh¹, Mai Thi Phuong Nga²

¹Northern Mountainous Agriculture and Forestry Science Institute

²University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology

SUMMARY

Based on previous results, five genes (GolS3, RHL41, bHLH102, DHN1 and DREB26) related to drought tolerance were selected to analyze gene expression in three tea lines grown in Phu Tho under artificial drought conditions. 1-year tea plants of the tea lines LVC4C1, PHN2C1 and VSHC7C2 were treated with the PEG 6000 20% (w/v) solution for 0h (control), 12h, 24h and 72h. After drought treatments, the 4th-5th leaves were collected to extract the total RNA. The qRT-PCR results showed that all tested genes were significantly upregulated in response to drought in tea plants. The expression levels of these genes were different in the same tea line and among three tea lines. In the LVC4C1 line, the relative expression levels of 5 genes increased at all treatment times, while only 2 out of 5 genes had relative expression levels increased at all treatment times in PHN2C1 and VSHC7C2 lines. In addition, among the three tea lines, LVC4C1 line had 4/5 genes (RHL41, DHN1, bHLH102 and DREB26) with the strongest relative expression level after 72 hours of the drought treatment. These results may be related to the better drought tolerance of LVC4C1 line than PHN2C1 and VSHC7C2 lines.

Keywords: Tea plant, drought tolerance, gene expression level, qRT-PCR, transcription factor.

^{*} Author for correspondence: Tel. 0914788108; Email: donanvnvn@yahoo.com

LENGTH AND SEQUENCE HETEROPLASMY IN mtDNA D-LOOP REGION OF INDIVIDUALS FROM KINH POPULATION IN SOUTHERN VIETNAM

Nguyen Thi Lan Huong^{*}, Nguyen Ngoc Hoa Binh, Dang Thanh Xuan, Mai Quang Truong

Forensic Medicine Center of Ho Chi Minh City, HCMC, Vietnam

SUMMARY

In forensic science, the control region of human mitochondrial DNA (mtDNA), mainly the hypervariable regions I (HVI), II (HVII), and III (HVIII), which are located at positions 16024 to 576, are the most extensively studied. These regions provide a high index of polymorphism. However, high mutation rates in mtDNA can cause heteroplasmy, complicating the interpretation of mtDNA results. In spite of this, there are currently few studies that describe this issue. To gain further insights in mtDNA heteroplasmy, this work investigated the presence of heteroplasmy (length and sequence) in the hypervariable regions (HVI, HVII and HVIII) on D-loop segment of the mtDNA from 396 unrelated healthy Kinh individuals using standard Sanger sequencing method with separation by capillary electrophoresis. All the subjects displayed length and sequence heteroplasmies in the HVI, HVII and HVIII regions. From position 16180 to 16195, 15 patterns of polycytosine or C-stretch were observed (125 samples, 31.56%), 221 Kinh individuals (55.81%) were found to have similar pattern to the reference, and 8 patterns (50 samples, 12.63%) didn't have polycytosine but were different from the reference. From position 303 to 315, 8 patterns were observed, whereas 7CT6C and 8CT6C were the most frequently found. From position 514 to 523, there were 4 motifs of $(AC)_n$, including $(AC)_4$ to $(AC)_7$. From position 568 to 573, there were 6 samples (1.52%) with insertion of +2C, +3C, and +4C at np 573. In addition, there were 8 samples (2.02%) with sequence heteroplasmy, in which 7 positions were detected. The frequency of heteroplasmy was also calculated. The nomenclature of variants was established according to EMPOP guidelines. The study provided a new perspective with important consequences in medical, evolutionary and forensic fields.

Keywords: Control region, heteroplasmy, Kinh population, mitochondrial DNA.

INTRODUCTION

In forensic science, the nucleotide sequence analysis of the control region of human mtDNA, mainly in hypervariable regions I (HVI), II (HVII) and III (HVIII), has been categorized as a useful tool for personal identification and maternal testing. HVI is located at position 16,024 to 16,365, HVII at position 73 to 340 and HVIII at position 438 to 576 (Butler, 2011). These regions provide a high index of polymorphism. Due to the huge number of mtDNA copies existing in one human body and high mutation rate of mtDNA, all individuals are expected to exhibit some level of heteroplasmy, in which some copies of that individual's mtDNA have a different sequence than the others (Melton, 2004). In other words, heteroplasmy is the presence of more than one type of mtDNA in an individual (Melton, 2004). Two or more mtDNA populations may occur between cells in an individual, within a single cell, or within a single mitochondrion (Butler, 2011). Length heteroplasmy often occurs around the homopolymeric C-stretches while sequence heteroplasmy is typically detected by the presence of two nucleotides at a single site, showing up as overlapping peaks in a sequence electropherogram (Butler, 2011). Heteroplasmy has also been reported to remain stable over time in the same individuals and thus be inherited rather than age related (Lagerström-Fermér *et al.*, 2001). Although heteroplasmy can sometimes complicate the interpretation of mtDNA results, the presence of heteroplasmy at identical sites can improve the probability of a match (Ivanov *et al.*, 1996).

Understanding the levels of human mtDNA heteroplasmy is vital in several fields, especially in forensic science. However, currently there are few works that focus on this issue. To gain further insights in mtDNA heteroplasmy, this work investigated the presence of heteroplasmy (length and sequence) in the hypervariable regions (HVI, HVII and HVIII) on the D-loop segment of the mtDNA.

MATERIALS AND METHODS

Materials

The buccal cells were collected by sterile cotton swabs from 396 unrelated healthy Kinh individuals. This information was ascertained through inquiry before sampling. The data were collected from cases examined at the Forensic Medicine Center of Ho Chi Minh City from 2018 to 2022.

Methods

Extraction of genomic DNA: Genomic DNA was extracted from buccal swabs using the QIAamp DNA Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. The quantity of recovered DNA was determined using Invitrogen™ Qubit™ dsDNA HS Assay Kit (Thermo Fisher Scientific, USA).

PCR amplification, DNA purification, and sequencing:

PCR amplification: The extended D-loop region (about 1.3 kb in size, np 15973 to 632) was amplified with the primer sets of 15973F (5'-AACTCCACCATTAGCACCCAAAG-3') and 632R (5'-GTGAGCCCGTCTAAACATT-3') (Sigma, Germany). PCR was performed in 25 µL of a volume containing 6.75 µL of DNA solution (1-5 ng), 2.5 µL of each 5 µM primer, 0.75 µl DMSO 10%, and 12.5 µL Phusion[®] Hot Start Flex 2X Master Mix (Thermo Fisher Scientific, USA). Thermal cycling was carried out on the Veriti[™] 96-Well Thermal Cycler (Thermo Fisher Scientific, USA), began with 30 s at 98°C, followed by 30 cycles of 5 s at 98°C, 10 s at 55°C, and 30 s at 72°C, and the final extension at 72°C for 5 minutes.

Size selection and purification: All PCR products were analyzed by electrophoresis on a 1.5% agarose gel and visualized by GelRedTM Nucleic Acid Stain (Thermo Fisher Scientific, USA) staining under a UV transilluminator (Wealtec, USA). The PCR products were purified using QIAquick[®] Gel Extraction Kit (Qiagen, Germany).

Sequencing: The purified amplicons were sequenced by Sanger method using BigDye Terminator v3.1 Cycle Sequencing Kit (Thermo Fisher Scientific, USA) with 6 primers on the capillary electrophoresis benchtop instrument ABI 3130 Genetic Analyzer (Applied Biosystems, USA).

Statistical analysis: The data were preliminarily treated by Sequencing Analysis 5.4 software. The nucleotide sequences were analyzed and aligned with the revised Cambridge Reference Sequences (rCRS) using Sequencher 5.4.5 software. The frequencies of each heteroplasmy were facilitated by dividing the number of times the heteroplasmy observed in a population by the total number of examined samples at a particular genetic locus in the D-loop region.

RESULTS

Nucleotide variations in C-stretch at position 16180-16195 in HVI

All of the HVI variable sequences in 396 Kinh individuals were analyzed and compared with the rCRS using Sequencher 5.4.5 software and rechecked by the expanded rCRS (Figure 1). Some homopolymeric sequences were observed in some individuals in the control region between position 16180 and position 16195. There were 56 samples where nucleotide insertions occurred, causing the sequence of the 16180-16195 region to extend to 17 and 18 nucleotides instead of 16 nucleotides like the rCRS sequence.



Figure 1. Sequence analysis of the mitochondrial D-loop region

The electropherogram shows the differences between the sample and the rCRS (top row). Each sample was sequenced more than once to check for accuracy. Dots indicate where differences occurred.

Table 1. Sequence types of polyC1 (position 16180 to 16195) in the mtDNA control region for 396 Kinh individuals (Ref: AAAACCCCCTCCCCAT, bold; Base change: red, bold, underline; Symbol ":" means absence of one nucleotide)

Sequence	SNP	Number of Observations	Frequency
AAAACCCCCTCCCCAT	rCRS	221	0.5581
AAAACCCCC	16189C	12	0.0303
AAAACCCCC <u>C</u> CCCC <u>C</u> AT	16189C, 16193.1C	2	0.0051
AAAACCCCC <u>C</u> CCCC <u>CC</u> AT	16189C, 16193.1C, 16193.2C	5	0.0126
AAAACCCCC <u>C</u> CCC <u>T</u> CAT	16189C, 16192.1T	3	0.0076

AAAACCCCC	16189C, 16192T	4	0.0101
AAAACCCCCTCCC T AT	16193T	2	0.0051
AAAACCCCCTCC <u>T</u> CAT	16192T	28	0.0707
AAAACCCC <u>TC</u> CCC <u>T</u> CAT	16188T, 16189C, 16192.1T	1	0.0025
AAAACC <u>T</u> CC <u>C</u> CCCCAT	16186T, 16189C	1	0.0025
AAAAC <u>T</u> CCCTCCCCAT	16185T	10	0.0253
AAAA <u>T</u> CCCC <u>C</u> CCCCAT	16184T, 16189C	1	0.0025
AAAA <u>T</u> CCCCTCCCCAT	16184T	1	0.0025
AAA <u>C</u> CCCCC <u>C</u> CCCCAT	16183C, 16189C	3	0.0076
AAA <u>C</u> CCCCC <u>C</u> CCCC <u>C</u> AT	16183C, 16189C, 16193.1C	45	0.1136
AAA <u>C</u> CCCCC <u>C</u> CCC <u>T</u> CAT	16183C, 16189C, 16192.1T	1	0.0025
AAA <u>C</u> CCCCC <u>C</u> CC <u>T</u> CAT	16183C, 16189C, 16192T	1	0.0025
AAA <u>G</u> CCCCCTCCCCAT	16183G	2	0.0051
AA <u>CC</u> CCCCC <u>C</u> CCCCAT	16182C, 16183C, 16189C	39	0.0985
AA <u>G</u> ACCCCCTCCCCAT	16182G	5	0.0126
A <u>CCC</u> CCCCC <u>C</u> CCCCAT	16181C, 16182C, 16183C, 16189C	4	0.0101
<u>GACCCCCCCCCCCC</u> CCCCAT	16180G, 16182C, 16183C, 16189C	1	0.0025
AAA:CCCCCA	16183delA, 16189A	1	0.0025
AAA <u>CA</u> CCCC <u>C</u> CCCCAT	16183C, 16184A, 16189C	3	0.0076

In this study, it was revealed that the polycytosine stretch (C-stretch or polyC) observed in the HVI region had a total of 24 patterns of nucleotide sequence variations (Table 1). 15 patterns of C-stretch were observed (125 samples, 31.56%), whereas 221 Kinh individuals (55.81%) were found to be similar to the reference and 8 patterns (50 samples, 12.63%) didn't have polycytosine but were different from the reference. Interestingly, C-stretches with the nucleotide transitions from T to C at position 16189 were observed in 125 Kinh individuals (31.56%). In the C-stretches region, repetitions of cytosine with more than 9 nucleotides were observed in 112 Kinh individuals. Additionally, nucleotide transition (A \rightarrow G) at position 16180 was found in one Kinh individual and one individual with A deletion at position 16183 was identified.

Nucleotide variations in C-stretch at position 303-315 in HVII

In this study, 8 patterns were observed at position 303-315 in HVII region (Table 2). Notably, there was no sample which was identical to the rCRS. The insertions were observed within a stretch of seven Cs between 303 and 309 (+C, +2C, +3C) or within five Cs between 311 and 315 (+C) in the control region. At positions 303-309, 151 sequences with an insertion of a single C residue, 76 sequences with an insertion of two C residues and 5 sequences with an insertion of three C residues were identified. Finally, at 311-315, 391 control regions with 6 cytidine residues were found. It was notable that at position 303-309, there were 5 sequences with one C deletion and one sequence with 3C deletion. Additionally, nucleotide transition (T \rightarrow C) at position 310 which generated polycytosine were found in four Kinh individuals.

Table 2. Sequence variations from position 303 to position 315 in HVII of mitochondrial DNA	for 396 Kinh individuals
(Ref: CCCCCCCCCCC or 7CT5C, bold; Base change: red, bold, underline; Symbol ":" means al	bsence of one nucleotide)

Patterns	Sequence	SNP	Number of Observations	Frequency
7CT5C	сссссстссссс	rCRS	0	0
13C	22222 2 222222	310C	4	0.0101
4CT6C	СССС:::ТССССС <u>С</u>	307delC, 308delC, 309delC, 315.1C	1	0.0025
6CT5C	CCCCCC:TCCCCC	309delC	1	0.0025
6CT6C	ССССС:тССССС <u>С</u>	309delC, 315.1C	4	0.0101
7CT6C	сссссстссссс <u>с</u>	315.1C	154	0.3889
8CT6C	ссссссс <u>с</u> тссссс <u>с</u>	309.1C, 315.1C	151	0.3813
9CT6C	ссссссс <u>сс</u> тссссс <u>с</u>	309.1C, 309.2C, 315.1C	76	0.1919
10CT6C	ссссссс <u>ссс</u> тссссс <u>с</u>	309.1C, 309.2C, 309.3C, 315.1C	5	0.0126

Length variation with C-stretch at position 568-573 in HVIII

From position 568-573, there were 6 samples (1.52%) (Table 3) with insertion of +2C (3 samples), +3C (1 sample), and +4C (2 samples) at np 573. When the polyC has 10Cs or more, the downstream sequence was not recorded accurately because the signal was unclear, which was caused by an out-of-phase phenomenon during Sanger sequencing

Table 3. Nomenclature of SNPs from position 566 to position 576 in HVIII

(Ref: CACCCCCACA, bol	l; Base change: red, bold, underline)
-----------------------	---------------------------------------

Sequence	SNP	Number of Observations				
CACCCCCACA	rCRS	390				
CACCCCCC	573.1C, 573.2C	3				
CACCCCCC	573.1C, 573.2C, 573.3C	1				
CACCCCCC	573.1C, 573.2C, 573.3C, 573.4C	2				

AC repeats at nucleotide position 513-522 in HVIII

 $(AC)_n$ is a kind of length polymorphism in which the n represents the number of AC dinucleotide repeats (from position 513 to position 522 in the rCRS). Some previous studies recorded it as $(CA)_n$ motif. A total of four repeat motifs $[(AC)_4 \text{ to } (AC)_7]$ were detected: 195 sequences with $(AC)_4$, 196 with $(AC)_5$, 4 with $(AC)_6$ and 1 with $(AC)_7$ repeats (Table 4). $(AC)_4$ and $(AC)_5$ exhibited the most frequent occurrence whereas $(AC)_7$ was the rarest.

Table 4. Length variation from position 513 to position 522 in HVIII region of mitochondrial DNA for 396 Kinh individuals (*Ref: ACACACACAC or (AC)*₅, *bold*)

Sequence	Number of Observations	Frequency
ACACACAC or (AC) ₄	195	0.4924
ACACACACAC or (AC) ₅	196	0.4949
ACACACACACAC or (AC) ₆	4	0.0101
ACACACACACACAC or (AC)7	1	0.0025

Sequence heteroplasmy

There were 8 samples (2.02%) with sequence heteroplasmy: Y (T/C) and R(A/G) (Figure 2). Among the 8 individuals who possessed detectable sequence heteroplasmy, only one sample exhibited mixed bases at 2 positions (16111 C>Y, 198 C>Y), the remaining samples were heteroplasmic at one position. The sequence included 16111 C>Y, 16247 A>R, 16295 C>Y, 16299 A>R, 16311 T>Y, 198 C>Y, and 215 A>R. These SNP occurred at both HVI and HVII regions. The total number of sequence heteroplasmic positions recorded was 7 positions, of which 5 positions (71.43%) occur in the HVI region, the remaining 2 positions (28.53%) occur in the HVII region. No sequence heteroplasmy was detected in the remaining regions on the D-loop. Among the 7 heteroplasmic positions, position 215 was observed three times, the others occurred only once (Table 5).



Figure 2. Sequence heteroplasmy at 16311Y (T/C) and 215R (A/G)

Table 5.	Positions	along the	control reai	on at which	heteroplasm	v was observed
						,

Nucleotide position	Number of Observations	rCRS	Base change	
198	1	С	Y	
215	3	А	R	
16111	1	С	Y	
16247	1	А	R	
16295	1	С	Y	
16299	1	А	R	
16311	1	т	Y	

DISCUSSION

PolyC in HVI

The nucleotide transition (T>C) at position 16189, the nucleotide transversion (A>C) at position 16181 and/or 16182 and/or 16183 and the nucleotide C insertion at position 16193 resulted in the repetition of cytosine by more than 10 times. This region carries valuable genetic markers that could discriminate one ancestral population from another. The HVI C-stretch was observed in the Thai, Chinese Tu, Asian (Chinese, Taiwanese, Korean, and Japanese), European, and Amerindian Populations (Sangthong *et al.*, 2015). This homopolymeric stretch of cytosines create problems for DNA polymerases as they synthesize the complementary strand to the mtDNA template present in the PCR reaction (Butler, 2011). The impact of a 16189 T to C transition on the sequencing result downstream of the C-stretch region could be seen in Figure 3.



Figure 3. A sample with the C-stretch in HVI

Frequently, approximately 15% of Europeans get a T-to-C transition at position 16189, relative to the CRS (Bendall *et al.*, 1995). In the phylogenetic context, position 16189 was commonly observed in different mtDNA lineages (Finnila *et al.*, 2001). This position was among the 50 fastest evolving sites (Irwin *et al.*, 2009). 16189 T>C transition in the HVI region could be associated with disease (Bendall *et al.*, 1995). Chinnery *et al.* (2005) suggested that the 16184-16193 polyC tract has a very small role in the pathophysiology of type 2 diabetes (Chinnery *et al.*, 2005) while Liou *et al.* (2010) reported that the mtDNA 16189 variant could cause alteration of mtDNA copy number in human blood cells (Liou *et al.*, 2010).

PolyC in HVII

In the HVII region, at position 303-315, a polycytosine tract with a single thymidine inserted at position 310 creates length polymorphism among individuals, as well as variation within an individual, associated with aging and cancer. This sequence was involved in forming a persistent RNA-DNA hybrid that led to the initiation of mtDNA heavy-strand replication (Frigi *et al.*, 2009). The mtDNA length heteroplasmy in HVI and HVII regions played a role in determining mtDNA copy number (Zhao *et al.*, 2010). In the study, most samples showed insertions of C residues, located within a stretch of seven Cs between 303 and 309 or within five Cs between 311 and 315. There were 4 haplotypes containing 13C in the region 303-315. Substitution of T to C has also been reported. Zahidin *et al* (2018) reported 2 cases of T to C substitution, leading to polyC in the HVII region with 16C and 17C (Zahidin *et al.*, 2018). Similar to HVI, the phenomenon of C insertion in the region 303-315 also sometimes caused difficulty in sequencing (Figure 4).



Figure 4. C-stretch region of HVII region between 309 and 315 with substitution of T to C at np 310

PolyC in HVIII

C-stretch in HVIII is considered as a hotspot for length heteroplasmy in EMPOP. Thus, it is ignored for calculating haplotype frequency and forensic genetic indices (Brandstätter *et al.*, 2004). Accurate recording of the 6 Cs in this region was performed using CE, but this was not possible for longer than 10 Cs (Figure 5). Brandstätter *et al.* (2004) reported that polyC spanning nucleotides 568-573 were present in 5% of the individuals from Nairobi, Kenya. Irwin *et al.* (2009) recorded 3% of the samples showed length heteroplasmy with expansion of the C-stretch beyond the 6 Cs in most individuals.



Figure 5. Insertion of +4C at np 573

(AC)_n patterns in HVIII

Even though the HVI and HVII sequences were utilized in most forensic cases, it is remarkable that HVIII $(AC)_n$ repeat segment also plays a key role as an additional mtDNA marker in forensic investigations (Ivanov *et al.*, 1996). In some previous studies, $(AC)_5$ was the most common motif among Pakistani, Iraqi, Malaysian, Korean, Germans, Japanese, and Indian Muslims, while $(AC)_4$ was common among Thais, Venezuelans, Cameroonians and Urali Kuruman.

Sequence heteroplasmy

The detection of mixed positions that may represent sequence heteroplasmy from Sanger-based data in this study relied primarily on repeated visual inspection of properly aligned electropherogram traces by experienced examiners. Shared sequence heteroplasmy between maternal relatives can increase the strength of the mtDNA evidence in a case of historical significance (Ivanov *et al.*, 1996). Irwin *et al.* (2009) examined population-based mtDNA datasets developed via Sanger sequencing and identified sequence heteroplasmy in the control region in approximately 6% of individuals when buccal and blood cells were examined. The control region nucleotide positions at which sequence heteroplasmy has been most frequently observed are consistent with the positions with the highest substitution rates (Irwin *et al.*, 2009). However, there are a few notable exceptions. The variable rates are primarily accounted for by differences in the tissue types, and populations examined. The incidence of sequence heteroplasmy in tissues with high metabolic activity (e.g. 79% in muscle) has been generally higher than in blood or blood-derived specimens (4-8%) (Naue *et al.*, 2015).

Sequence heteroplasmy, which occurred in HVI and HVII, was similar to the study of Irwin *et al.*, 2009. Position 215 is not reported as an evolutionary fast site. This may be due to unusual mechanisms affecting variation at this site. Position 215 is situated close to the origin of heavy strand replication (Irwin *et al.*, 2009). Santos *et al.* (2008) reported that the 215G variant may induce a conformational change in the control region. Therefore, mutations at this position could affect mtDNA replication (Irwin *et al.*, 2009). In contrast to position 215, positions 198, 16111 and 16311 in this study are considered some of the fastest evolving sites according to the study of Irwin *et al.* (2009).

CONCLUSION

The study reported on heteroplasmy in the mtDNA control region in the Kinh population. From the 396 samples analyzed, examples of length heteroplasmy caused by polycytosine stretches were found in HVI, HVII and HVIII regions. Sequence heteroplasmy occurred at both HVI and HVII regions. The presence of these variants can be visually identified in sequence electropherograms generated by capillary electrophoresis. Whereas heteroplasmy may complicate interpretation of the results, it actually becomes a "signature" of the donor's mtDNA haplotype, which can strengthen the identification of a sample. The analysis of heteroplasmy in D-loop is potentially useful for medical, evolutionary, and forensic purposes.

REFERENCE

Bendall KE, Sykes BC (1995). Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. *Am J Hum Genet* 57: 248-256.

Brandstätter A, Peterson CT, Irwin JA, Mpoke S, Koech DK, Parson W, Parsons TJ (2004). Mitochondrial DNA control region sequences from Nairobi (Kenya): inferring phylogenetic parameters for the establishment of a forensic database. *Int J Legal Med* 118: 294-306.

Butler JM (2011). Advanced topics in forensic DNA typing: methodology. Academic Press: 405.

Chinnery PF, Elliott HR, Patel S, Lambert C, Keers SM, Durham SE, McCarthy Hitman GA, Hattersley AT, Walker M (2005). Role of the mitochondrial DNA 16184–16193 poly-C tract in type 2 diabetes. *Lancet* 366: 1650-1651, DOI:10.1016/S0140-6736(05)67492-2.

Finnila S, Lehtonene MS, Majamaa K (2001). Phylogenetic network for European mtDNA. Am J Hum Gen 68: 1475-1484.

Frigi S, Ennafaa H, Cherni L, Gaaied AE (2009). Investigation of polymorphisms in hypervariable region II (HVII) sequences of human mitochondrial DNA in two North-African populations. *Anthropologie* 107-118.

Hua Zhao, Jie Shen, Leonard Medico, Mary Platek, Christine B Ambrosone (2010). Length heteroplasmies in human mitochondrial DNA control regions and breast cancer risk. *Int J Mol Epidemiol Genet* 1(3): 184-192.

Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, Parsons TJ (2009). Investigation of heteroplasmy in the human mtDNA control region: A Synthesis of Observations from More Than 5000 Global Population Samples. *J Mol Evol* 68(5): 516-527.

Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996). Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. *Nat Genet* 12: 417-420.

Lagerström-Fermér M et al. (2001). Heteroplasmy of the human mtDNA control region remains constant during life. Am J Hum Genet 68: 1299-1301.

Liou CW, Lin TK, Chen JB, Tiao MM, Weng SW, Chen SD, Chuang YC, Chuang JH, Wang PW (2010). Association between a common mitochondrial DNA D-loop polycytosine variant and alteration of mitochondrial copy number in human peripheral blood cells. *J Med Genet* 47: 723-728.

Melton T (2004). Mitochondrial DNA heteroplasmy, Forensic Sci Rev 16: 1-20.

Naue J, Horer S, Sanger T, Strobl C, Hatzer-Grubwieser P, Parson W *et al.* (2015). Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA. *Mitochondrion* 20: 82-94.

Sangthong P, Janso A, Chinnabanchonchai N (2015). Sequence analysis of mitochondrial DNA hypervariable region I in Thai individuals. *Aust J Forensic Sci* 47(3): 345-354.

Santos C, Sierra B, Alvarez L, Ramos A, Fernandez E, Nogues R, Aluha MP (2008). Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. *J Mol Evol* 67: 191-200.

Zahidin MA, Omar WBW, Taib WRW, Japning JRR, Abdullah MT (2018). Sequence polymorphism and haplogroup data of the hypervariable regions on mtDNA in Semoq Beri population. *Data in Brief* 21: 2609-2615.

ĐA HÌNH TRÌNH TỰ VÙNG SIÊU BIẾN VÀ HIỆN TƯỢNG DỊ THỂ BỘ GEN TY THỂ CỦA CÁC CÁ THỂ DÂN TỘC KINH Ở MIỀN NAM VIỆT NAM

Nguyễn Thị Lan Hương^{*}, Nguyễn Ngọc Hòa Bình, Đặng Thanh Xuân, Mai Quang Trường

Trung tâm Pháp y Thành phố Hồ Chí Minh, Việt Nam

TÓM TẮT

Trong khoa học pháp y, vùng điều khiển của DNA ty thể người (mtDNA), gồm các vùng siêu biến I (HVI), II (HVII) và III (HVII), nằm ở vị trí 16024 đến 576, được nghiên cứu rộng rãi nhất. Những vùng này cung cấp chỉ số đa hình cao. Tuy nhiên, tỷ lệ đột biến cao ở mtDNA có thể gây ra hiện tượng dị thể, làm phức tạp việc giải thích kết quả mtDNA. Mặc dù vậy, hiện nay có rất ít nghiên cứu mô tả vấn đề này. Để hiểu rõ hơn về dị thể mtDNA, đề tài này đã nghiên cứu sự hiện diện của dị thể (chiều dài và trình tự) trong các vùng siêu biến (HVI, HVII và HVIII) trên đoạn D-loop của mtDNA từ 396 cá thể người Kinh khôe mạnh không có quan hệ, sử dụng phương pháp giải trình tự Sanger tiêu chuẩn dựa trên điện di mao quản. Tất cả các đối tượng đều hiển thị dị thể về chiều dài và trình tự ở các vùng HVI, HVII và HVIII. Từ vị trí 16180 đến 16195, có 15 kiểu polycytosine (125 mẫu, 31,56%), trong khi 221 cá thể Kinh (55,81%) được phát hiện giống với rCRS và 8 kiểu (50 mẫu, 12,63%) không có polycytosine nhưng khác với trình tự tham chiếu. Từ vị trí 514 đến 523 có 4 mô típ (AC)_n, gồm (AC)₄ đến (AC)₇. Từ vị trí 568 đến 573, 6 mẫu (1,52%) có chèn +2C, +3C, +4C tại vị trí 573. Ngoài ra, có 8 mẫu (2,02%) dị thể về trình tự, trong đó 7 vị trí được phát hiện. Danh pháp của các biến thể theo EMPOP đã được đề cập. Nghiên cứu này cung cấp một góc nhìn mới với những hệ quả quan trọng trong các lĩnh vực y tế, tiến hóa và pháp y.

Từ khóa: Dân tộc Kinh, DNA ty thể, hiện tượng dị thể, vùng điều khiển.

^{*} Author for correspondence: Tel: 0974487053; Email: ntlhuong226@gmail.com

GENETIC CHARACTERISTICS OF 22 a-STR LOCI IN THE VIETNAMESE POPULATION

Nguyen Thi Lan Huong, Vuong Gia Bao*, Mai Quang Truong

Forensic Medicine Center of Ho Chi Minh city, HCMC, Vietnam

SUMMARY

This study reported Short Tandem Repeat (STR) allele data from 4346 Vietnamese individuals across 22 autosomal STR (a-STR) loci: CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, Penta D and Penta E. Population samples consisted of 90% ethnic Kinh and 10% individuals of other ethnic minority or of mixed origin (by questionnaire data). Sample types included buccal swab, blood, hair, nail, tissue, tooth and bone. DNA extraction was performed using QIAamp DNA Mini Kit, QIAamp DNA Investigator Kit and Bone DNA Extraction kit. PCR reactions were carried out with the PowerPlex® Fusion System kit. The PCR products were subjected to capillary electrophoresis on ABI 3500 Genetic Analyzer. Electrophoretic results were analyzed using GeneMapper® ID-X software v1.4. Direct counting method and Excel software were applied to determine allele frequency and forensic genetic parameters. The allele distribution frequencies of 22 a-STR loci were established. The samples were in Hardy-Weinberg equilibrium according to the distribution of genotypes by 22 a-STR loci. The combined matching probability (CMP), combined power of discrimination (CPD), combined power of exclusion (CPE) and combined paternity index (CPI) achieved values of 6.2236×10^{-27} , 1, 0.99999999998952507 and 8.4811×10^8 , respectively. The a-STR loci have high discriminatory power and polymorphic information content, which demonstrates that a-STR has great potential for population biodiversity research, human identification, paternity testing, and forensic applications.

Keywords: Allele frequency, forensic genetic parameter, STR, Vietnamese population.

INTRODUCTION

The most widely used genetic markers in forensic DNA profiling (an examination of a DNA sample's sequence and/or length at discrete locations) are Short Tandem Repeats (STRs) or Simple Sequence Repeats (SSRs). STRs (also known as microsatellites) consist of units 2-6 bp long, repeated about 5 to 30 times. Therefore, when multiple loci are analyzed together, the number of possible combinations between STR loci as a group is extremely large. It is the highly variable number of repetitions of units in the STR marker region between individuals that has made STR markers the standard for individual identification. The small size of STR markers makes them much more suitable for forensics because shorter loci are more likely to be found intact (Butler, 2005). Each individual has a DNA profile on the autosomes (a-STR profile) and for each marker there is one paternal allele and one maternal allele. Therefore, the a-STR profile is individual-specific.

When comparing two DNA profiles, a match is always a statistical exercise. To determine the probability that a particular genotype might occur at random in a population as well as to ensure the reliability of DNA analysis based on STR markers, population data must be compiled to make an estimate of the frequency of each possible allele and genotype (Chakraborty, 1992). The number of STR loci and sample size must be large enough to obtain the precise and reliable results. Establishing a DNA database is vital for every forensic laboratory.

This study was carried out to initially create a DNA database, including allele frequencies and forensic genetic parameters for 22 a-STR loci in the Vietnamese population. The DNA database was used to serve the assessment work at the Forensic Medicine Center of Ho Chi Minh City in determining consanguinity relationships and solving complicated cases.

MATERIALS AND METHODS

Materials

Biological samples (buccal cell, blood, hair root, nail, tissue, bone, tooth) were collected for paternity testing and casework at the Forensic Medicine Center of Ho Chi Minh City (Vietnam) from 2016 to 2023. For this study, cases with excluded paternity were selected, including 4346 unrelated individuals (2451 men and 1895 women) of any age, living in Vietnam. Samples were randomly collected from across Vietnam, and all personal information was kept confidential.

Methods

Extraction of genomic DNA: Genomic DNA was extracted from buccal cells, blood and tissues with the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany), nails and hairs with QIAamp DNA Investigator Kit (Qiagen, Hilden, Germany), bones and teeth with Bone DNA Extraction Kit (Promega, USA) according to the manufacturer's instructions. The quantities of recovered DNA were determined using Invitrogen[™] Qubit[™] dsDNA HS Assay Kit (Thermo Fisher Scientific, USA).

Genotyping of autosomal STRs: The 22 a-STR loci, one Y-STR locus and the sex determination locus (Amelogenin gene) were simultaneously amplified using the PowerPlex® Fusion System kit (Promega, USA) on a Veriti[™] 96-Well thermal cycler (Applied Biosystems, USA). The PCR products were subjected to capillary electrophoresis on ABI 3500 Genetic Analyzer (Applied Biosystems, USA), following manufacturer's recommendations. Data were collected with Data Collection v1.0 software. Electrophoretic results were analyzed using GeneMapper® ID-X software v1.4 (Applied Biosystems, USA).

Statistical analysis: Allele frequency distribution for each locus and forensic genetic parameters (Matching Probability (MP), Power of Discrimination (PD), Polymorphic Information Content (PIC), Power of Exclusion (PE), Typical Paternity Index (TPI), observed heterozygosity (Ho), and expected heterozygosity (He)) were calculated according to some studies of Butler, 2005; Sensabaugh, 1982; Nei, 1987; Brenner, Morris, 1990.

RESULTS AND DISCUSSION

Allele frequencies and forensic genetic parameters of 22 a-STR loci

In the study, 320 alleles were detected with a mean of 14.55 alleles per locus, ranging from 8 alleles at TPOX to 27 alleles at Penta E. In 4346 individuals, the number of heterozygous individuals in the loci was 3 - 7 times greater than the number of homozygous individuals. The high number of heterozygous individuals showed that there was diversity about alleles and therefore there will be less chance for a random combination (Edwards *et al.*, 1992). The database of allele frequencies of 22 a-STR loci in 4346 individuals was shown in Table 1. The results showed that the allele with the highest frequency was allele 8 of the TPOX marker with a frequency of 0.5802 (Figure 1). Alleles with the lowest frequency were alleles that appeared only once with a frequency of 0.0001. These may be rare alleles and require further investigation with a larger sample size for accurate confirmation.



Figure 1. Illustration of DNA profile on autosomes *Left: Male DNA profile (XY), right: Female DNA profile (XX).*

Regional variation is one of the factors that contributes to improving the reliability of the matching probability of two DNA profiles. Allele frequencies vary across different regions and areas. For example, surveying the alleles of the TPOX locus, although allele 8 had the highest frequency, following by allele 11, values were achieved in turn of 0.4520 and 0.3571 in Japan (n=1501) (Fujii *et al.*, 2014), 0.5269 and 0.3353 in the Philippines (n=167) (Rodriguez *et al.*, 2015), 0.5746 and 0.2516 in Belarus (n=12225) (Tsybovskiia *et al.*, 2017), and 0.5802 and 0.2640 in the present study (n=4346). Or for allele 22 on the FGA locus, the frequency in Japan was 0.1965 (highest), in the Philippines it was 0.1617 (second highest after allele 23), in Belarus it was 0.2112 (highest) and in this study it was 0.1970 (highest).

Table 1. Allele frequencies and forensic genetic parameters of 22 a-STR loci in 4346 Vietnamese individuals

Allele	D3 S1358	D1 S1656	D2 S441	D10 S1248	D13 S317	Penta E	D16 \$539	D18 S51	D2 S1338	CSF1P O	Penta D	TH01	vWA	D21 S11	D7 S820	D5 S818	трох	D8 S1179	D12 S391	D19 S433	FGA	D22 S1045
4	-		-							-		-	-		-					0.0005		-
5						0.0413					0.0013	0.0003										
6					0.0001						0.0007	0.1291			0.0001		0.0001					
7					0.0012	0.0028	0.0001	0.0001		0.0068	0.0364	0.3221			0.0116	0.0321	0.0007					
7.2												0.0001										
7.4											0.0005											
8		0.0005	0.0001	0.0003	0.3479	0.0025	0.0024			0.0006	0.0645	0.0602			0.1629	0.0013	0.5802	0.0020				
8.1			0.0001			0.0002									0.0001							
9		0.0001	0.0023		0.1330	0.0114	0.2247	0.0005		0.0354	0.3445	0.3747			0.0637	0.0613	0.0995	0.0007		0.0072		
9.1			0.0145												0.0083							
9.2															0.0002							
9.3												0.0460										
10		0.0018	0.2178	0.0005	0.1245	0.0498	0.1291	0.0012		0.1949	0.1463	0.0656			0.1660	0.2438	0.0358	0.1550		0.0002		
10.1			0.0013												0.0008							
11	0.0002	0.0754	0.2951	0.0025	0.2138	0.2551	0.2816	0.0078		0.2793	0.1275	0.0018			0.3602	0.2693	0.2640	0.1338		0.0036		0.1667
11.1										0.0002					0.0010							
11.2											0.0032				0.0001					0.0005		
11.3			0.1271				0.0002															
12	0.0007	0.0436	0.1691	0.0609	0.1389	0.1205	0.2447	0.0618		0.3894	0.1343		0.0005		0.1910	0.2224	0.0186	0.1268		0.0467		0.0013
12.1																0.0006						
12.2																				0.0037		
12.3			0.0020																			
13	0.0024	0.0972	0.0163	0.3448	0.0321	0.0544	0.0957	0.1414	0.0001	0.0775	0.0958		0.0003		0.0304	0.1539	0.0010	0.1590		0.2531	0.0003	0.0030
13.1								0.0001														
13.2								0.0001												0.0425		
13.3			0.0001																			
14	0.0337	0.0873	0.1439	0.2382	0.0083	0.0893	0.0208	0.2006		0.0132	0.0332		0.2676		0.0033	0.0140		0.1612		0.2593		0.0361
14.1			0.0001																			
14.2																				0.1087		
14.3		0.0001																				
15	0.3102	0.2994	0.0097	0.2231	0.0003	0.0701	0.0006	0.2228		0.0023	0.0074		0.0293		0.0001	0.0013		0.1664	0.0116	0.0699		0.3104
15.2																				0.1550		
15.3		0.0029																				
15.4						0.0001																
16	0.3575	0.1948	0.0005	0.1034		0.0662		0.1585	0.0200	0.0002	0.0037		0.1568					0.0737	0.0048	0.0098	0.0021	0.2208
16.2																				0.0350		
16.3		0.0082																				
17	0.2269	0.0766		0.0240		0.0592		0.0629	0.1096		0.0002		0.2310					0.0171	0.0687	0.0006	0.0007	0.2321
17.2																				0.0033		
17.3		0.0723																	0.0002			
17.4						0.0005																
18	0.0617	0.0161		0.0023		0.0542		0.0385	0.0718		0.0001		0.2102					0.0039	0.1977		0.0115	0.0277
18.2																				0.0005	0.0001	
18.3		0.0185																	0.0005			
18.4						0.0015																
19	0 0064	0 0029				0.0411		0 0368	0 2215		0 0005		0 0866					0 0005	0 1944		0 0801	0.0013
19.1	0.0001	0.0025				0.0.111		0.0001	0.2215		0.0005		0.0000					0.0005	0.15 11		0.0001	0.0010
19.3		0 0023						0.0001											0.0001			
20	0.0003	0.0001				0.0301		0.0258	0.1079				0.0162						0.1870		0.0569	0.0007
20.2	0.0000	0.0001				0.0501		5.5250	5.2075				5.0102						0.1070		0.0008	5.0007
21						0.0182		0.0743	0.0400				0.0013						0.1261		0.1616	
21.2						0.0103		5.5275	5.5400				5.5015						0.1201		0.0148	
																					5.5140	

CÔNG NGHỆ GEN

51356 5141 51248 5137 553 5138 5138 511 5820 5818 5117 5391 5433 51 21.3 1.4 0.0001 0.0001 0.0001 0.0985 0.1970 22.2 0.0001 0.0001 0.001 0.0118 0.0118 23.4 0.0004 0.0040 0.1711 0.746 0.1416 23.2 0.0035 0.0024 0.1498 0.023 0.0133 24 0.0035 0.0024 0.1498 0.0021 0.0107 0.8800 25 0.0023 0.0009 0.6799 1 1 0.017 0.8000 25 0.0005 0.0058 0.0021 0.0117 0.8000 0.0078 26 0.0005 0.0058 0.0021 0.0013 0.0016 0.0024 0.0004 0.0024 0.0004 0.0024 0.0004 0.0024 0.0006 0.0024 0.0006 0.0024 0.0006 0.0024 0.0006 0.0024 0.0006 0.0024 0.0015 0.0004 0.0024 0.0004 <t< th=""><th>2</th></t<>	2
21.321.40.0010.0010.09850.0010.1970220.0010.0010.0180.018230.0040.0040.17110.07460.1416240.0050.00240.14980.0230.0230.01724.20.0030.0090.05790.01070.008025.30.0050.00580.00210.00130.00826.40.0050.00580.00210.0130.001270.0050.0060.00230.0060.002126.20.0050.0060.00230.0010.001270.0050.00580.00230.00230.001280.00020.00550.00550.0020.0023290.00050.00050.00550.00550.0055	045
21.4 0.001 0.009 0.003 0.001 0.0985 0.1970 22.2 0.001 0.001 0.018 23.2 0.0094 0.040 0.111 0.0766 0.1416 23.2 0.0035 0.0024 0.1393 0.0239 0.1427 24.4 0.035 0.0024 0.1498 0.0329 0.1427 24.2 0.0023 0.009 0.0579 0.0107 0.0080 25.2 0.0023 0.009 0.0579 0.0107 0.0076 25.3 0.0005 0.0058 0.0021 0.0013 0.0432 26.4 0.0005 0.0058 0.0021 0.0013 0.0432 26.4 0.0005 0.0058 0.0023 0.0151 0.0074 27.4 0.0002 0.0055 0.0085 0.0016 0.0074 28.4 0.0002 0.0555 0.0085 0.0016 0.0075	
22 0.0156 0.0090 0.0438 0.001 0.0985 0.1970 22.2 0.001 0.011 0.018 23 0.004 0.004 0.1711 0.0746 0.1146 23.2 0.0035 0.0024 0.1498 0.0239 0.1427 24.2 0.0035 0.009 0.0579 0.0107 0.0001 25.2 0.0023 0.009 0.0579 0.0107 0.0007 25.3 0.005 0.0058 0.0021 0.0013 0.0036 26.1 0.005 0.0058 0.0021 0.0013 0.0436 26.2 0.005 0.0058 0.0021 0.0013 0.0436 26.1 0.005 0.0058 0.0023 0.013 0.0436 26.2 0.005 0.0058 0.0023 0.0151 27.2 0.006 0.0023 0.0151 0.007 28.2 0.002 0.0535 0.0088 28.2 0.0055 0.015 0.0086	
22.2 0.0001 0.0118 23 0.0094 0.0040 0.1711 0.0746 0.1416 23.2 0.0035 0.0024 0.1498 0.0239 0.1427 24.2 0.0023 0.009 0.579 0.0107 0.800 25.2 0.0023 0.009 0.579 0.0107 0.800 25.3 0.005 0.0058 0.0021 0.0013 0.0435 26.1 0.005 0.0058 0.0021 0.0013 0.0435 26.2 0.0005 0.0058 0.0023 0.0013 0.0435 26.1 0.0005 0.0058 0.0023 0.0013 0.0435 26.2 0.0006 0.0023 0.0151 0.0024 0.0015 27.2 0.0002 0.0535 0.0008 0.0008 0.0008 28.2 0.0002 0.5557 0.0005 0.0015 0.0008	
23 0.0094 0.0040 0.1711 0.0746 0.1416 23.2 0.0035 0.0024 0.1498 0.0239 0.1427 24.2 0.0023 0.009 0.579 0.0107 0.0800 25.2 0.0025 0.0058 0.0021 0.0013 0.0076 25.3 0.0005 0.0058 0.0021 0.0013 0.0435 26.1 0.0005 0.0058 0.0021 0.0013 0.0435 26.2 0.0005 0.0006 0.0023 0.0013 0.0024 27 0.0006 0.0023 0.0151 0.0007 0.0007 28.2 0.0002 0.0535 0.0008 0.0008 0.0008 28.2 0.0002 0.0535 0.0008 0.0008 0.0008 0.0008 0.0015 0.0008 0.0008 0.0008 0.0015 0.0008 0.0015 0.0008 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015	
23.2 0.0035 0.0024 0.1498 0.0239 0.1427 24.2 0.0035 0.0009 0.0579 0.0107 0.0800 25.2 0.0025 0.0058 0.0021 0.0013 0.0068 25.3 0.0005 0.0058 0.0021 0.0013 0.0435 26.1 0.0005 0.0058 0.0021 0.0013 0.0435 26.2 0.0005 0.0006 0.0023 0.0013 0.0435 26.2 0.0005 0.0006 0.0023 0.0013 0.0024 27 0.0006 0.0023 0.0151 0.0007 0.0007 28 0.0002 0.0535 0.0008 0.0008 0.0008 0.0008 28.2 0.0002 0.0535 0.0008 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0008 0.0015 0.0015 0.0008 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015 0.0015	
24 0.0035 0.0024 0.1498 0.0239 0.1427 24.2 0.0023 0.0099 0.0579 0.0107 0.0800 25.2 0.0025 0.0058 0.0021 0.0013 0.0068 25.3 0.0005 0.0058 0.0021 0.0013 0.0435 26.1 0.0005 0.0001 0.0013 0.0024 26.2 0.0006 0.0023 0.0013 0.0024 27 0.0006 0.0023 0.0151 27.2 0.0006 0.0023 0.0015 28.2 0.0002 0.0535 0.0008 29.0 0.0587 0.0058 0.0058	
24.2 0.0023 0.0009 0.6579 0.0107 0.6800 25.2 0.0005 0.0058 0.0021 0.0013 0.0005 26.1 0.0005 0.0058 0.0021 0.0013 0.0023 26.2 0.0006 0.0001 0.0024 0.0024 27 0.0006 0.0023 0.0151 27.2 0.0006 0.0023 0.0151 28.2 0.0002 0.0535 0.0008 28.2 0.0005 0.0015 0.0005	
25 0.0023 0.0009 0.0579 0.0007 25.2 0.0078 0.0078 26.1 0.0005 0.0058 0.0021 0.0013 26.2 0.0006 0.0006 0.0023 0.0013 27.2 0.0006 0.0023 0.0151 28.2 0.0002 0.0555 0.0008 29.0 0.0015 0.0015	
25.2 0.0078 0.0078 25.3 0.0005 0.0021 0.0013 26.1 0.0001 0.0024 26.2 0.0006 0.0023 0.0151 27.2 0.0002 0.0053 0.0007 28.2 0.0002 0.0535 0.0008 29.0 0.0587 0.0058 0.0058	
25.3 0.0005 0.0058 0.0021 0.0013 0.0435 26.1 0.0001	
26 0.005 0.005 0.005 0.007 0.003 0.0435 26.1 0.001	
26.1 0.0001 26.2 0.0024 27 0.0006 0.0023 0.0151 27.2 0.0002 0.0535 0.0008 28.2 0.0015 0.0055 0.0005	
26.2 0.0024 27 0.0006 0.0023 0.0151 27.2 0.0002 0.0535 0.0008 28.2 0.0015 0.2587	
27 0.0005 0.0023 0.0151 27.2 0.0007 0.0535 0.0008 28.2 0.0015 0.2587	
28 0.0002 0.0535 0.0008 28.2 0.0015 0.2587	
28 0.002 0.003 0.003 0.003 28.2 0.0015 0.2587 0.2587	
29 0.2587	
23	
20.2	
29.2 0.0000	
30 0.2598	
30.2 0.0225	
30.3 0.0025	
31 0.0023	
31 2 0 0795	
32 0.0262	
32 0.0202	
32.2 0.1622	
33 0.0025	
33.2 0.0469	
34 0.0013	
34.2 0.0056	
35 0.0001	
35.2 0.0013	
Forensic genetic parameters	
MP 0.1284 0.0438 0.0693 0.0964 0.0796 0.0214 0.0800 0.0385 0.0335 0.1214 0.0576 0.1176 0.0712 0.0538 0.0822 0.0767 0.2309 0.0378 0.0400 0.0546 0.0277 0.0	1899
PD 0.8716 0.9562 0.9307 0.9036 0.9204 0.9786 0.9200 0.9615 0.9665 0.8786 0.9424 0.8824 0.9288 0.9462 0.9178 0.9233 0.7691 0.9622 0.9600 0.9454 0.9723 0.9	a101
PC 0.6662 0.8188 0.7705 0.7218 0.7488 0.8778 0.7502 0.8364 0.8485 0.6797 0.7889 0.6865 0.7673 0.7998 0.7433 0.7576 0.5258 0.8402 0.8105 0.8632 0.7	7342
PE 0.4499.06682.0609.05400.05580.07513.05514.07094.07117.04624.06263.04843.05893.06384.05555.05704.02767.05095.06095.06095.06095.0514.07	5493
TEI 17468 3.0563 2.5686 2.4404 2.2472 4.1078 2.2106 3.5048 3.5333 1.7059 2.1059 2.4361 2.755 2.2333 2.3104 4.2101 4.710 2.5077 3.3176 2.0513 2.558	1997
Ha 0.7106 0.8360 0.7004 0.7506 0.7707 0.8862 0.7840 0.8520 0.8621 0.7248 0.8008 0.7201 0.7076 0.8218 0.7745 0.7800 0.5821 0.8522 0.8560 0.7840 0.8520	7712
Ho 0.7138 0.8364 0.8053 0.7674 0.7775 0.8783 0.7738 0.8573 0.7246 0.7246 0.8145 0.7340 0.7048 0.8108 0.7761 0.7844 0.5874 0.8610 0.8403 0.8273 0.8703 0.7	7727
	. 21
10 19 16 10 10 27 10 22 14 11 17 9 11 20 16 10 8 12 15 18 25 10	—

The matching probability (MP) was highest at the TPOX locus (0.2309), lowest at the Penta E locus (0.0214). The power of Discrimination (PD), Polymorphic Information Content (PIC), Power of Exclusion (PE) and Typical Paternity Index (TPI) were highest at the Penta E locus with values of 0.9786, 0.8778, 0.7513, and 4.1078, respectively. The lowest values for PD, PIC, PE, and TPI were observed at the TPOX locus, with values of 0.7691, 0.5258, 0.2761, and 1.2119, respectively. The lowest and the highest observed heterozygosity (Ho) were obtained in the TPOX and Penta E loci (0.5874 and 0.8783, respectively). The expected heterozygosity (He) values ranged from 0.5821 (TPOX) to 0.8862 (Penta E). The heterozygosity of all loci was greater than 0.7, except for TPOX. These results indicated that these loci have high discriminatory power in the Vietnamese population.

Combined forensic genetic parameters

Combined matching probability (CMP) = 6.2236×10^{-27} . This means that the chance of any two people having exactly the same DNA profile is 1 in 1.6068×10^{-26} . That is the DNA profile of an individual was unique in the population. Thus, when these 22 STR loci are combined, it is reliable and accurate for individual identification, paternity testing, and research on population genetics.

A Combined Power of Discrimination (CPD) value of 1 indicates that the DNA profiling method can distinguish each individual in the population from all others. Understanding in another way, instead of looking at the probability of matching, the probability of "not matching" was interested in.

Combined Power of Exclusion (CPE) = 0.9999999999952507. This index indicated that a person's DNA profile differs from the DNA profile of a randomly selected person in the population by 99.9999998952507%.

Combined Paternity Index (CPI) = 8.4811×10^8 . This index is also often used for paternity testing. That means the number of similarities between a child and the biological father is 8.4811×10^8 times.

Almost all countries calculate their allele frequencies and standardize them to improve accuracy in conclusions. In Table 2, the combined matching probability (CMP) decreases when more markers are examined, from 7.29x10⁻¹⁶ (15 loci) to 1.948x10⁻³⁶ (31 loci). This means the possibility of having any two people with identical DNA profiles is increasingly unlikely as the number of markers examined increases. Similarly, the combined indices CPD, CPE and CPI all increase when increasing the number of markers and sample size. This means that the more markers are examined, the greater the chance of accurately identifying each individual and the lower the probability of errors in concluding consanguinity relationships. CPD and CPE values in this study indicate that the discriminatory power of the examined loci is considerably high. The study collected a large number of samples (n=4346) and analyzed with many loci simultaneously (22 loci), which improved the efficiency, reliability, and accuracy of the data.

One of the most important applications of the DNA database created from the study is to determine the accuracy (probability of paternity - POP) in lineal paternity testing. The more loci that are examined, the closer the POP value is to 100% if a match is obtained. For example, when comparing two DNA profiles of a father and a child, if 15 loci are examined, the POP value is 99.999992220147%. If 22 loci are examined, the POP value will be 99.999999994358%. However, if a new DNA profile does not have the observed allele in the database, the POP calculation cannot be performed. Therefore, the DNA database must be updated annually.

With the advancement of science and technology, a number of new technologies and genetic markers, such as next-generation sequencing and single nucleotide polymorphisms, have been widely used. However, due to the lack of databases for new genetic markers, STR profiling using older methods is still employed in forensic practice. In 1997, the Federal Bureau of Investigation (FBI) laboratory selected 13 STRs as its core loci (CODIS), which were expanded to 20 STRs in 2017. Commercial STR kits are based on these core loci. In recent years, these core loci have been supplemented by an increasing number of new non-CODIS loci to obtain additional genetic information and further improve discrimination. However, the newly approved STRs should be used with caution.

Parameter	СМР	CPD	CPE	CPI
This study	6.2236x10 ⁻²⁷ (n=4346, 22 loci)	1 (n=4346, 22 loci)	0.999999998952507 (n=4346, 22 loci)	8.4811x10 ⁸ (n=4346, 22 loci)
Shimada <i>et al.</i> , 2002		0.999999999999999999999998 (n=178, 16 loci)	0.99999994 (n=178, 16 loci)	
Trần Khánh Linh <i>et al.</i> , 2010		0.9999999999999491 (n=182, 12 loci)	0.9999923 (n=182, 12 loci)	
Liu <i>et al.</i> , 2013	7.29x10 ⁻¹⁶ (n=150, 15 loci)	0.9999999999999999914651 (n=150, 15 loci)	0.9999916368 (n=150, 15 loci)	
Anghel <i>et al.</i> , 2014		0.9999999999999999999999722 (n=336, 15 loci)	0.999999295 (n=336, 15 loci)	
Rodriguez <i>et al.</i> , 2015	8.21x10 ⁻²⁸ (n=176, 23 loci)		0.999996 (n=176, 23 loci)	
Park <i>et al.</i> , 2016	3.601×10 ⁻²⁶ (n=1000, 22 loci) 2.902×10 ⁻²⁸ (n=1000, 24 loci)		0.99999999971 (n=1000, 24 loci)	
Tsybovskiia <i>et</i> <i>al.</i> , 2017	7.089342×10 ⁻²² (n=12346, 18 loci)		0.999961 (n=12346, 18 loci)	

Table 2. Comparison of forensic genetic parameters with other studies (surveyed loci covered 13 CODIS)

Barrio <i>et al.</i> , 2019	2.022x10 ⁻²⁴ (n=496, 20 loci) 1.948x10 ⁻³⁶ (n=496, 31 loci)			1.473x10 ⁸ (n=496, 20 loci) 1.118x10 ¹² (n=496, 31 loci)
Vu <i>et al.</i> , 2021	1.8648x10 ⁻²⁷ (n=1000, 22 loci)	1 (n=1000, 22 loci)	0.99999999992 (n=1000, 22 loci)	

CONCLUSION

The study has established the allele frequency database and associated statistical parameters for 22 a-STR markers from a large sample (4346 unrelated individuals distributed among provinces scattered across Vietnam). The a-STR loci have high discriminatory power and polymorphic information content, which demonstrates that a-STR has great potential for population biodiversity research, human identification, paternity testing, and forensic applications.

REFERENCE

Anghel A, Enache A, Seclaman E, Gruin G, Ursoniu S, Alexa A, Antonescu M, Marian C (2014). Genetic polymorphism data on 15 autosomal STR markers in a Western Romanian population sample. *Leg Med (Tokyo)* 16(4): 238-240.

Barrio PA, Martín P, Alonso A, Müller P, Bodner M, Berger B, Parson W, Budowle B & DNASEQEX Consortium (2019). Massively parallel sequence data of 31 autosomal STR loci from 496 Spanish individuals revealed concordance with CE-STR technology and enhanced discrimination power. *Forensic Sci Int Genet* 42: 49-55.

Brenner CH, Morris JW (1990). Paternity Index Calculations in Single Locus Hypervariable DNA Probes: Validation and Other Studies Proceedings for The International Symposium on Human Identification 1989, *Promega Corporation*: 21-53.

Butler JM (2005). Forensic DNA Typing - Biology Technology DNA genetics of STR Markers, 2nd ed., Acadamic Press.

Chakraborty R (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing, *Hum Biol* 642: 141-159.

Edwards A, Hammond HA, Jin L, Caskey CT, Chakraborty R (1992). Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics* 12(2): 241-53.

Fujii K, Iwashima Y, Kitayama T, Nakahara H, Mizuno N, Sekiguchi K (2014). Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population. *Leg Med* 16(4): 234-237.

Liu J, Li SY, Yin JY, Zhang W, Gao B, Guo L, Qi R (2013). Genetic analysis of 15 STR loci from a Xibe population in Liaoning, Northeast of China. *Ann Hum Biol* 41(3): 244-248.

Nei M (1987). Molecular Evolutionary Genetics. New York: Columbia University Press.

Park HC, Kim K, Nam Y, Park J, Lee J, Lee H, Kwon H, Jin H, Kim W, Kim W, Lim S (2016). Population genetic study for 24 STR loci and Y indel (GlobalFiler[™] PCR Amplification kit and PowerPlex[®] Fusion system) in 1000 Korean individuals. *Leg Med* (*Tokyo*) 21: 53-57.

Rodriguez JJRB, Salvador JM, Calacal GC, Laude RP, Ungria MCA (2015). Allele frequencies of 23 autosomal short tandem repeat loci in the Philippine population. *Leg Med* 17(4): 295-297.

Sensabaugh G (1982). Biochemical markers of individuality. Saferstein R eds. Forensic Science Handbook. New York: Prentice-Hall, Inc: 338-415.

Shimada I, Brinkmann B, Nguyen QT, Hohoff C (2002). Allele frequency date for 16 STR loci in the Vietnamese population. Int J Legal Med 116(4): 246-248.

Tsybovskiia IS, Veremeichika VM, Kotovaa SA, Kritskayaa SV, Evmenenkob SA, Udinac IG (2017). Developing Forensic Reference Database by 18 Autosomal STR for DNA Identification in Republic of Belarus. *Russ J Genet* 53(2): 275-284.

Thi Thanh Hoan Vu, Thi Thu Mai Do, Thi Hoai Nguyen, Quoc Hai Luyen (2021). Allele frequencies of 23 short tandem repeat loci in the Vietnamese Kinh population. *Forensic Sci Int Rep.* https://doi.org/10.1016/j.fsir.2021.100210.

Trần Khánh Linh, Lê Thúy Quyên, Võ Đức Xuyên An và Phạm Hùng Vân (2010). Bước đầu khảo sát tần suất phân bố kiểu hình của các STR thường được sử dụng trong xét nghiệm dấu vân tay DNA ở người Việt Nam. *Tạp chí Y học Thành phố Hồ Chí Minh*, Tập 13, Phụ bản số 2: 116-120.

ĐẶC ĐIỂM DI TRUYỀN CỦA 22 CHỈ THỊ PHÂN TỬ a-STR TỪ CÁC CÁ THỂ NGƯỜI VIỆT NAM

Nguyễn Thị Lan Hương, Vương Gia Bảo^{*}, Mai Quang Trường

Trung tâm Pháp y Thành phố Hồ Chí Minh, Việt Nam

TÓM TẮT

Nghiên cứu này báo cáo dữ liệu alen STR từ 4346 người Việt Nam được khảo sát trên 22 locus STR nhiễm sắc thể thường (a-STR): CSF1PO, FGA, TH01, TPOX, vWA, D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, Penta D và Penta E. Dân số bao gồm 90% người Kinh và 10% là người dân tộc thiểu số khác hoặc có nguồn gốc hỗn hợp (theo dữ liệu bảng câu hỏi). Các loại mẫu bao gồm niêm mạc miệng, máu, tóc, móng tay, mô, răng, xương. Quá trình tách chiết DNA được thực hiện bằng cách sử dụng các bộ kit QIAamp DNA Mini Kit, QIAamp DNA Investigator Kit và Bone DNA Extraction Kit. Phản ứng PCR được thực hiện với bộ kit PowerPlex® Fusion System. Sản phẩm PCR được điện di mao quản trên máy phân tích di truyền ABI 3500. Kết quả điện di được phân tích bằng phần mềm GeneMapper® ID-X v1.4. Phương pháp đếm trực tiếp và phần mềm excel được áp dụng để xác định tần suất alen và các thông số di truyền pháp y. Bảng tần suất alen của 22 locus a-STR đã được xác lập. Các mẫu ở trạng thái cân bằng Hardy-Weinberg theo sự phân bố kiểu gen của 22 locus STR. Xác suất trùng khớp kết hợp (CMP), khả năng phân biệt kết hợp (CPD), khả năng loại trừ kết hợp (CPE) và chỉ số quan hệ huyết thống kết hợp (CPI) đạt các giá trị lần lượt là 6.2236x10⁻²⁷, 1, 0.999999998952507 và 8.4811x10⁸. Các locus a-STR có khả năng phân biệt và thông tin đa hình cao, chứng tỏ rằng a-STR có tiềm năng lớn cho nghiên cứu đa dạng sinh học quần thể, nhận dạng con người, xét nghiệm quan hệ huyết thống và các ứng dụng pháp y.

Từ khóa: Tần suất alen, thông số di truyền pháp y, STR, người Việt Nam.

Author for correspondence: Tel: 0765230982; Email: giabaovuong10@gmail.com

KHẢO SÁT PHƯƠNG PHÁP PCR-DGGE CHO PHÂN TÍCH ĐA DẠNG VI KHUẦN TRONG MÃU PHÂN CỦA TRỂ 6-24 THÁNG TUỔI

Nguyễn Thị Quý¹, Đào Trọng Khoa¹, Nguyễn Thị Thơm^{2,3}, Đỗ Thị Huyền^{1,2*}

¹Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Bệnh viện 19-8, Bộ Công an

TÓM TẮT

Đa dạng vi khuẩn và mất cân bằng vi khuẩn trong phân của trẻ 6-24 tháng tuổi được cho là có liên quan đến nhiều bệnh như tiêu chảy, tự kỷ, béo phì, dị ứng, chàm... Việc phân tích đa dạng cấu trúc vi khuẩn trong phân bằng phương pháp điện di trên gel gradient chất biến tính (DGGE) cho phép thực hiện trên một số lượng mẫu lớn, có khả năng so sánh cấu trúc vi sinh trong các mẫu và gợi ý được tình trạng mất cân bằng vi sinh cũng như xác định vi khuẩn gây bệnh có thể có. Để áp dụng được phương pháp DGGE trên đối tượng vi khuẩn trong mẫu phân trẻ khỏe mạnh và trẻ tiêu chảy kéo dài chưa rõ nguyên nhân, trong nghiên cứu này vùng V3, V6-V8 của gen 16S rDNA đã được khuếch đại bằng chương trình tough-up, tough-down và chương trình đặc hiệu từ hệ gen của 8 vi khuẩn đã biết, 5 đa hệ gen vi khuẩn của trẻ khỏe mạnh và 1 đa hệ gen của trẻ tiêu chảy kéo dài. Kết quả cho thấy, gel DGGE có nồng độ polyacrylamide 8% là thích hợp cho phân tách tốt, đa dạng vùng V3. Chương trình tough-up giúp khuếch đại được vùng V3 đa dạng hơn các chương trình PCR khác. Hàm lượng DNA gen V3 thích hợp nhất cho phân tích DGGE là 1 μg/giếng. Gel DGGE được nhuộm với SYBR-Gold cho độ nét và độ nhạy tốt, đảm bảo cho việc quét ảnh phân tích đánh giá đa dạng vi sinh vật trong các mẫu nghiên cứu.

Từ khóa: DGGE, chương trình PCR tough-up, đa dạng vi khuẩn trong phân, tiêu chảy kéo dài, vùng gen V3.

MỞ ĐẦU

Ở Việt Nam, tiêu chảy là một trong 10 bệnh phổ biến và đứng thứ 4 về tỷ lệ tử vong ở trẻ dưới 5 tuổi và tập trung chủ yếu ở trẻ 6-24 tháng tuổi. Tùy theo điều kiện y tế và dịch tễ, khoảng 3-20% trẻ tiêu chảy cấp diễn tiến thành tiêu chảy kéo dài. Mặc dù Bệnh viện Nhi trung ương đã áp dụng các phương pháp hiện đại như real-time PCR chẩn đoán 24 tác nhân gây tiêu chảy thông thường nhưng cũng có tới 47,39% trẻ không tìm được ra nguyên nhân (Vũ Thị Thu Hà et al., 2016). Gần đây, bằng phương pháp giải trình tự thế hệ mới tìm hiểu đa hệ gen vi khuẩn trong phân của trẻ tiêu chảy ở một số nước, các nhà nghiên cứu nhận thấy sự thay đổi cấu trúc hệ vi sinh vật của trẻ (rối loan hệ vi sinh) là nguyên nhân dẫn tới trẻ tiêu chảy kéo dài, khó hồi phục. Phương pháp giải trình tự gen thế hệ mới là phương pháp đánh giá được toàn diện nhất sự mất cân bằng hệ vi sinh của trẻ nhưng phương pháp này chỉ đánh giá được trên từng cá thể, với chi phí tốn kém và mất nhiều thời gian, khó so sánh tình trang mất cân bằng trên số lương lớn trẻ trong cùng thời điểm. Mặc dù được xem là phương pháp có khả năng phân tích thấp hơn phương pháp giải trình tự nhưng các phương pháp dấu vân tay phân tử (molecular fingerprinting techniques) như điện di trên gel gradient biến tính (denaturing gradient gel electrophoresis -DGGE), đa hình đô dài đoan giới han (restriction fragment length polymorphism - RFLP) hay đa hình cấu hình sợi đơn (single strand conformation polymorphism - SSCP) vẫn được sử dụng thường xuyên có giá trị để có thể so sánh sự biến đông vi sinh trong hê sinh thái liên guan đến các yếu tố tác đông khác nhau (Westaway et al., 2021). Phân tích DGGE vùng V3, V6-V8 (gọi tắt là vùng V68) của gen 16S rDNA thường được sử dụng trong đánh giá đa dạng vi khuẩn trong phân của trẻ để tìm ra sự mất cân bằng vi sinh trong bệnh như dị ứng (Štšepetova et al., 2007, Bisgaard et al., 2011) hay benh eczema (Forno et al., 2008), tiêu chảy (Mai et al., 2006).

Do phân có hệ vi sinh đa dạng nhất trong các hệ sinh thái, đặc biệt ở cơ thể người (Costea *et al.*, 2017) nên việc đánh giá đa dạng vi sinh, so sánh, xác định sự biến động vi sinh giữa nhóm chứng và nhóm bệnh bằng phương pháp DGGE cần được khảo sát. Các nghiên cứu trước đã chỉ ra nồng độ chất biến tính (gồm urea và formamide deionized, trong đó 100% chất biến tính gồm 7 M urea, 40% formamide deionized) thích hợp cho phân tách DGGE vùng gen V3 là 35-70% và vùng V68 là 40-60%. Một số tác giả đã cho rằng gel được nhuộm bằng bạc có độ nhạy tốt (Qiu *et al.,* 2012) nhưng một số tác giả lại cho rằng việc nhuộm gel DGGE bằng ethidium bromide và SYBR-Gold cho độ nhạy và độ nét tốt hơn (Huang, Fu, 2005). Vì vậy, trong nghiên cứu này chúng tôi tiến hành

khảo sát một số yếu tố như nồng độ gel polyacrylamide, vùng gen V3, V68 của gen 16S rDNA, hàm lượng DNA và phương pháp thích hợp cho phân tích gen bằng DGGE để có thể đánh giá đa dạng vi khuẩn trong mẫu phân trẻ 6-24 tháng tuổi tốt nhất. Kết quả của công trình này sẽ định hướng phương pháp cho phân tích đa dạng vi khuẩn trong các mẫu phân của trẻ khỏe mạnh và trẻ tiêu chảy kéo dài, xem xét sự mất cân bằng vi khuẩn trong phân của trẻ tiêu chảy kéo dài.

NGUYÊN VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu

DNA hệ gen của 8 chủng vi khuẩn bao gồm *Escherichia coli* ATCC27117, *Bacillus cereus, B. subtilis* 168, *Staphylococcus aureus, Salmonella enterica* Typhymurium (viết tắt là *S.* Typhimurium), *S. enterica* Enteritidis ATCC 13076 (viết tắt là S. Enteritidis), *Listeria fermentum, Enterococcus faecium* đã được tách chiết trong nghiên cứu trước (Nguyễn Hồng Dương *et al.,* 2023) được sử dụng làm khuôn khuếch đại gen 16S rDNA trước khi khuếch đại vùng gen V3, V68 trong thí nghiệm lựa chọn nồng độ gel polyacrylamide thích hợp cho phân tách DGGE hai vùng gen.

DNA đa hệ gen của vi khuẩn trong mẫu phân của trẻ khỏe mạnh H1 (bé gái 8 tháng tuổi), trẻ tiêu chảy kéo dài chưa rõ nguyên nhân B1 (bé gái 11 tháng tuổi bị tiêu chảy trên 14 ngày) được tách chiết từ nghiên cứu trước (Nguyen *et al.*, 2023). Hai mẫu DNA đa hệ gen H1 và B1 được sử dụng để khảo sát phương pháp PCR khuếch đại vùng V3 cho việc đánh giá đa dạng và lựa chọn phương pháp nhuộm gel DGGE. DNA đa hệ gen vi khuẩn trong phân của 4 trẻ khỏe mạnh 6-24 tháng tuổi (ký hiệu H2-H5) dùng để khuếch đại vùng V3 cho thí nghiệm đánh giá hàm lượng gen V3 thích hợp cho phân tích DGGE. Việc thu mẫu phân cho nghiên cứu đã được Hội đồng Đạo đức của Bệnh viện Nhi trung ương chấp thuận cho phép theo Giấy chứng nhận số 1875/BVNTWW-HĐĐĐ ngày 10/7/2023.

Phương pháp PCR khuếch đại vùng gen V3, V68 và gen 16S rDNA từ hệ gen/đa hệ gen vi khuẩn

Gen 16S rDNA của vi khuẩn được khuếch đại bằng cặp mồi 27-F và 1527-R từ DNA hệ gen/đa hệ gen của vi khuẩn. Vùng gen V3, V68 được khuếch đại từ sản phẩm PCR gen 16S rDNA, sử dụng các trình tự mồi V3 (Ariefdjohan *et al.*, 2010, Fei *et al.*, 2016) V68 (Mai *et al.*, 2006) được liệt kê trong Bảng 1 trong đó trình tự clamp là 5' CGC CCG CGC GCG GGC GGG GGG GGG GGG GCA CGG GGG 3'. PCR được tiến hành trong tổng thể tích 25 µl gồm 2,5 µl đệm PCR 10x, 2,0 µl dNTP (2 mM), 1,0 µl mồi nỗi loại (10 µM), 5,0 µl DNA khuôn (2 ng/µl), 0,25 µl DreamTaq DNA polymerase (5 U/µl, Fermentas). Chương trình PCR sử dụng cho khảo sát khuếch đại đa dạng vùng gen V3, V68 gồm ba chương trình. Chương trình đặc hiệu (ký hiệu là N) có nhiệt độ bắt cặp mồi với sợi khuôn là 55°C, chương trình tough-up (ký hiệu là U) có nhiệt độ gắn mồi tăng dần trong mỗi 5 chu kỳ từ 50°C, 52°C, 55°C và cuối cùng là 20 chu kỳ có nhiệt độ gắn mồi ở 58°C. Chương trình tough-down tương tự như chương trình tough-up nhưng nhiệt độ bắt cặp mồi với sợi khuôn giảm dần từ 58°C xuống 52°C, cuối cùng là 20 chu kỳ với nhiệt độ mồi bắt cặp sợi khuôn là 50°C. Tổng số chu kỳ trong các chương trình PCR là 35 chu kỳ (Nguyễn Hồng Dương *et al.*, 2023). Sản phẩm PCR được kiểm tra bằng điện di trên gel agarose 0,8% (Sambrook *et al.*, 2001). Hàm lượng DNA và chất lượng DNA được đo bằng máy Nanodrop 2000 (Implen).

Gen	Kích thước	Tên mồi	Trình tự mồi
Gen 16S rDNA	1500 bp	27-F	5'-GAG TTT GAT CCT GGC TCAG-3'
		1527-R	5'-AGA AAG GAG GTG ATC CAGCC-3'
Vùng gen V3	180 bp	BA338F-GC	5'-GC-clamp GAC TCC TAC GGG AGG CAG CAG-3'
		UN518R	5'-ATT ACC GCG GCT GCT GG-3'
Vùng gen V6-V8	433 bp	U968F-GC	5'-GC-clamp GAA CGC GAA GAA CCT TAC-3'
		L1401R	5'-GCG TGT GTA CAA GAC CC-3'

Bảng 1 Danh cách mồi dùng	r để khuốch đại con	169 PDNA C	ác vùng gọn V3	1/6_1/9
Dariy I. Darin Sacii inoi uun	a de kildecil dai gell	IUS IDNA, C	ac vully yell vo.	v0-v0

Phân tích vùng gen V3, V68 bằng DGGE

Vùng gen V3, V68 sau khi khuếch đại được phân tích bằng điện di DGGE như được mô tả bởi Hovda và đồng tác giả (Hovda *et al.,* 2007) trên hệ thống máy VS20-DGGETC system (Cleaver Scientific Ltd, England). Trong đó vùng V3 được phân tích trên gel có nồng độ chất biến tính từ 35% đến 70%, vùng V68 được phân tích trên gel có nồng độ chất biến tính từ 35% đến 70%, vùng V68 được phân tích trên gel có nồng độ chất biến tính từ 35% đến 70%, vùng V68 được phân tích trên gel có nồng độ chất biến tính từ 40 đến 60%. Nồng độ polyacrylamide trên bản gel được khảo sát ở hai nồng độ là 6% và 8%. Điện di được tiến hành trong đệm TAE 1x (Eppendorf AG, Hamburg, Germany) ở 60°C, 20V trong 10 phút sau đó tăng lên 70V trong 18 giờ. Sau khi điện di, gel được thử nghiệm nhuộm với ba phương pháp là ethidium bromide (0,5 mg/l) hoặc SYBR-Gold (S11494, Thermo Scientific, USA) pha loãng 10000 lần trong đệm TAE 1x,

pH 7,6 và nhuộm bạc theo phương pháp số 2 của Qiu và đồng tác giả năm 2012 (Qiu et al., 2012).

KẾT QUẢ VÀ THẢO LUẬN

Khảo sát nồng độ gel polyacrylamide cho phân tích vùng V3, V68

Để đảm bảo độ ổn định cho việc khuếch đại vùng V3, V68 của 8 chủng vi khuẩn khảo sát, và của đa hệ gen vi khuẩn trong mẫu H1, B1, gen 16S rDNA và đa gen 16S rDNA của các vi khuẩn đã được khuếch đại bằng PCR sử dụng mồi và chương trình PCR đặc hiệu từ hệ gen của các chủng khảo sát và đa hệ gen của mẫu H1, B1. Kết quả (Hình 1) cho thấy gen 16S rDNA có cùng kích thước khoảng 1,5 kb đã được khuếch đại thành công từ 8 chủng vi khuẩn và hai DNA đa hệ gen B1, H1. Gen được khuếch đại đặc hiệu, không có băng phụ (Hình 1A) trong đó mẫu đối chứng âm không có băng nào được khuếch đại. Sử dụng sản phẩm PCR gen 16S rDNA này làm khuôn, các vùng V3, V68 của vi khuẩn cũng đã được khuếch đại thành công với hàm lượng DNA thu được lớn (khoảng 100 ng/µl). Vùng gen V3 được khuếch đại có kích thước thấp hơn DNA chuẩn 250 bp, tương ứng với kích thước mong đợi là 433 bp (Hình 1B).

Các vùng gen V3, V68 đã được phân tích trên gel DGGE với nồng độ polyacrylamide 6%, tuy nhiên, chỉ có vùng V68 xuất hiện các băng DNA nhưng rất mờ, không rõ ràng (Hình 1C) còn vùng V3 hoàn toàn không nhìn thấy băng điện di (kết quả không được trình bày). Khi tăng nồng độ gel polyacrylamide lên 8%, các băng DNA vùng V3, V68 của các vi khuẩn đã thể hiện rõ ở tất cả các mẫu (Hình 1D, E). Vùng V3 có khả năng phân tách khá tốt, đặc biệt ở khả năng phân biệt các vi khuẩn trong cùng chi. Vùng V3 của các vi khuẩn ở các chi khác nhau dừng ở các vị trí cách nhau khá xa trên gel điện di và quan sát rõ hơn ở đường chạy Mix (Hình 1D). Trình tự có quãng đường di chuyển dài hơn thể hiện trình tự giàu GC hơn so với các trình tự khác do các trình tự này có nhiệt độ nóng chảy cao hơn nên dừng ở vùng gel có nồng độ chất biến tính cao hơn. Theo đó, vùng V3 của *E. faecium* di chuyển quãng đường ngắn nhất, tiếp đó là vùng V3 của *L. fermentum, B. cereus, B. subtilis, E. coli, S.* Typhimurium, S. Enteritidis và cuối cùng là S. aureus. Đáng chú ý, vùng V3 của *B. cereus, B. subtilis* có trình tự gần giống nhau nên nằm gần nhau. Vùng V3 của S. Typhimurium và S. Enteritidis cùng ở vị trí như nhau trên điện di đồ vì vùng V3 của hai vi khuẩn này tương đồng nhau về trình tự vì cùng trong một loài *S. enterica* (Hình 1D). Điều đó cho thấy, phân tích vùng V3 trên DGGE của các chủng vi khuẩn này có khả năng đánh giá độ đa dạng và quan hệ di truyền của các vi khuẩn.



Hình 1. Lựa chọn nồng độ gel polyacrylamide cho phân tách cùng V3, V68

A: Điện di đồ phân tích gen/đa gen 16S rDNA khuếch đại từ DNA hệ gen của 8 chủng vi khuẩn và đa hệ gen vi khuẩn trong mẫu phân của trẻ khỏe mạnh và trẻ tiêu chảy kéo dài. B: Điện di đồ phân tích vùng gen V3, V68 được khuếch đại từ khuôn là sản phẩm PCR gen 16S rDNA; C: Điện di đồ DGGE trên gel polyacrylamide 6% phân tách vùng V68; D, E: Điện di đồ DGGE trên gel polyacrylamide 8% phân tách vùng V3, V68 tương ứng. (-) Đối chứng âm không chứa DNA khuôn; M: DNA chuẩn 1 kb (Fermentas); 1-8: Sản phẩm PCR có nguồn gốc tương ứng từ E. coli ATCC 27117, B. cereus, B. subtilis 168, S. aureus, S. Typhymurium, S. Enteritidis ATCC 13076, L. fermentum, E. faecium. H1, B1: Tương ứng với DNA đa hệ gen từ mẫu phân của trẻ khỏe mạnh; trẻ tiêu chảy kéo dài; Mix: Hỗn hợp vùng V3 của 8 chủng chỉ thị từ 1 đến 8. Ngược lại, mặc dù vùng V3 được phân tách tốt trên gel DGGE ở nồng độ gel polyacrylamide 8% nhưng vùng V68 của 8 chủng vi khuẩn không được phân tách tốt và không đặc trưng cho chi, loài. Điển hình, vùng V68 của *S*. Typhimurium nằm ở vị trí cao nhất trên gel sau đó đến vùng V68 của *E. coli* ATCC27117 và cuối cùng là *B. subtilis* 168. Tuy nhiên, vùng V68 của các vi khuẩn *Salmonella, L. fermentum* lại nằm cùng trên một vùng, và rất gần, khó phân biệt với vùng V68 của *S. aureus, E. faecium, B. cereus*. Vì vậy, việc sử dụng vùng gen V68 cho đánh giá đa dạng vi khuẩn và phân biệt vi khuẩn trong các mẫu môi trường, đặc biệt là trong phân nơi có cộng đồng vi khuẩn rất phức tạp là rất khó khăn. Vùng gen V68 của các chủng cũng không được cải thiện. Do vậy, vùng V3 sẽ được sử dụng để khảo sát phương pháp nhuộm gel DGGE và hàm lượng DNA đưa lên gel. Hỗn hợp V3 của 8 chủng được trộn với hàm lượng tương đương để làm đối chứng dương trong các thí nghiệm tiếp theo.

Khảo sát phương pháp nhuộm gel DGGE

Trong các nghiên cứu trước, chương trình tough-up đã được chứng minh có khả năng khuếch đại đa gen 16S rDNA từ đa hê gen vi khuẩn trong mẫu phân đa dạng hơn chương trình tough-down và chương trình đặc hiệu. Chương trình tough-down cũng cho độ đa dạng của gen 16S rDNA tốt hơn chương trình đặc hiệu (Nguyễn Hồng Dương et al., 2023). Trong nghiên cứu này, đa gen 16S rDNA của vi khuẩn từ mẫu phân của trẻ khỏe mạnh và trẻ tiêu chảy kéo dài được khuếch đại bằng chương trình tough-up. Đa gen 16S rDNA này được dùng làm khuôn cho khuếch đại vùng V3 bằng ba chương trình tough-up, tough-down và chương trình đặc hiệu. Kết quả (Hình 2A) cho thấy vùng gen V3 đã được khuếch đại ở cả ba chương trình nhưng vùng gen V3 có kích thước đa dạng hơn, hàm lượng cao hơn ở mẫu PCR sử dụng chương trình tough-up. Trên bản điện di, vùng gen V3 khuếch đại bằng chương trình tough-up có thêm các băng phụ đậm hơn, nét hơn so với sản phẩm khuếch đại bằng chương trình đặc hiệu (chỉ có một băng đặc hiệu duy nhất, với lượng thấp hơn khi điện di với cùng thể tích mẫu). Kết quả này cũng tương đồng với các nghiên cứu trước, trong đó ở các chu kỳ đầu của chương trình PCR, nhiệt độ gắn mồi thấp đã giúp khuếch đại được các trình tự có số bản copy thấp có trong mẫu và số lượng mồi bám được vào sợi khuôn tăng, giúp khuếch đại DNA với hàm lượng lớn (Pechgit et al., 2011). Kết quả thử nghiệm nhuộm gel DGGE bằng các phương pháp khác nhau cho thấy, trong ba gel được nhuộm bằng SYBR-Gold, ethidium bromide và nhuộm bạc thì bản gel được nhuộm bằng bạc cho độ phát hiện băng kém nhất, các băng không rõ ràng, không nét và đây cũng là phương pháp nhuôm phức tap nhất (Hình 2D). Phương pháp nhuôm SYBR-Gold và nhuộm ethidium là phương pháp nhuộm đơn giản nhưng cho độ nét của băng tốt hơn, phổ băng phát hiện cũng nhiều hơn (Hình 2B, C). Trong đó, phương pháp nhuộm bằng SYBR-gold cho băng rõ nét nhất, đa dạng nhất. Tương đồng với kết quả này, nghiên cứu trước đây so sánh việc sử dụng phương pháp nhuôm gel bằng SYBR-Gold và ethidium cũng cho thấy SYBR-Gold có độ nhạy tốt hơn (Huang, Fu, 2005, Tuma et al., 1999). Phương pháp nhuôm gel bằng ethidium bromide rẻ nhưng đôc. Vì vây, phương pháp nhuôm gel bằng SYBR-Gold được lựa chọn cho các thí nghiệm sau.



Hình 2. Điện di đồ phân tích sản phẩm khuếch đại vùng gen V3 từ mẫu phân trẻ khỏe mạnh và trẻ tiêu chảy kéo dài (A) và DGGE phân tích vùng gen được nhuộm bằng ethidium bromide (B), SYBR-Gold (C) và bạc (D)

M: DNA chuẩn 1 kb (Fermentas); (-) Đối chứng âm không chứa DNA khuôn; H1: DNA khuôn có nguồn gốc từ trẻ khỏe mạnh; B1: DNA có nguồn gốc từ mẫu phân của trẻ tiêu chảy; N. U, D: Tương ứng là DNA được khuếch đại bằng chương trình đặc hiệu, tough-up và tough-down; (+): Đối chứng dương.

Kết quả phân tích DGGE cũng cho thấy, ở mẫu phân trẻ tiêu chảy kéo dài, vùng V3 được khuếch đại bằng chương trình tough-up cho độ đa dạng băng cao hơn so với chương trình tough-down. Điều đó chứng tỏ để phân tích đa dạng vi khuẩn trong một mẫu, việc sử dụng chương trình tough-up sẽ cho kết quả tốt hơn là hai chương trình còn lại. So sánh giữa mẫu chứng và mẫu bệnh, trẻ bị tiêu chảy có xu hướng giảm mức độ đa dạng vi khuẩn. Kết quả trên Hình 2 cũng chỉ ra rằng mẫu từ trẻ khỏe mạnh có số băng nhiều hơn mẫu bệnh nên mẫu khỏe mạnh có xu hướng có hệ vi sinh vật đa dạng hơn.

Khảo sát hàm lượng DNA vùng V3 thích hợp cho phân tích DGGE

Trong các mẫu phân, vi sinh vật trong mẫu rất đa dạng. Vì vậy, việc lựa chọn hàm lượng DNA thích hợp cho phân tích DGGE là cần thiết. Trong các thí nghiệm trên, hàm lượng V3 được phân tích trên mỗi giếng là 1 µg.

CÔNG NGHỆ GEN

Trong thí nghiệm này, để đánh giá chính xác hơn hàm lượng DNA thích hợp cho phân tích DGGE, vùng V3 gen 16S rDNA của đa hệ gen vi khuẩn từ 5 trẻ khỏe mạnh H1-H5 đã được khuếch đại và phân tích DGGE. Kết quả (Hình 3) cho thấy vùng V3 của đa gen vi khuẩn đã được phân tách khá tốt trên gel DGGE, các băng DNA đều thể hiện rõ nét trong cả 5 mẫu. Mặc dù là mẫu khỏe mạnh nhưng đa dạng vùng V3 ở các mẫu khác nhau là khác nhau về cả số lượng băng và vị trí của các băng DNA trên điện di đồ (Hình 3A). Điều này chứng tỏ vi khuẩn ở các mẫu chứng cũng rất đa dạng. Dựa trên phổ băng DNA có thể thấy đa dạng vi khuẩn trong mẫu H1 khác hẳn với 4 mẫu còn lại. Mẫu H3, H4 có sự tương đồng nhau về phổ băng hơn cả. Như vậy, rất có thể đa dạng vi khuẩn mẫu H3, H4 giống nhau hơn so với mẫu H1. Nhìn trên hai bản điện di thì có thể thấy hàm lượng DNA thích hợp đưa lên giếng DGGE là 1 µg vì ở hàm lượng này các băng thể hiện rõ nét, không bị nhòe phủ sang các băng bên canh tao thuận lợi cho việc quét băng đánh giá đa dạng trong các mẫu. Khi tăng lên hàm lượng 2 µg/giếng điện di, các băng có sự giao thoa, nhòe với các băng lân cận làm cho độ nét của băng kém hơn. Theo tìm hiểu của tác giả, không có tài liêu khoa học nghiên cứu khảo sát hàm lượng DNA thích hợp cho phân tích DGGE nhưng trong môt diễn đàn, các nhà nghiên cứu đã thảo luân, chia sẻ có thế sử dung hàm lương DNA vùng V3 từ 250 ng-750 ng/giếng điện di là đủ. Tuy nhiên, để nghiên cứu được đa dạng vi sinh trong mẫu sinh thái, việc tăng hàm lượng DNA cho phân tích là cần thiết. Vì vậy, để phân tích đa dạng vi khuẩn dựa trên vùng V3, lượng DNA thích hợp để đưa lên giếng cho phân tách là 1 µg/giếng.



Hình 3. Điện di đồ DGGE phân tích vùng gen V3 của vi khuẩn trong mẫu phân từ 5 trẻ khỏe mạnh với hàm lượng DNA khác nhau 1 μg/giếng (A) và 2 μg/giếng (B)

(+): Đối chứng dương; H1-H5: DNA có nguồn gốc từ mẫu phân của trẻ khỏe mạnh H1-H5.

KÉT LUẬN

Vùng V3 gen 16S rDNA của vi khuẩn trong mẫu phân trẻ khỏe mạnh và trẻ tiêu chảy kéo dài đã được phân tích thành công bằng phương pháp PCR-DGGE trên gel polyacrylamide 8%, nồng độ chất biến tính 35-70% với hàm lượng DNA được đưa vào phân tích trên mỗi giếng điện di là 1 μg. Phổ băng DNA của V3 trên điện di đồ ở các mẫu khác nhau là khác nhau thể hiện sự đa dạng của vi sinh vật có trong mẫu.

Lời cảm on: Công trình được hỗ trợ kinh phí từ để tài nghiên cứu mã số ĐTĐLCN.63/22 (Bộ Khoa học và Công nghệ) và trang thiết bị của Phòng Thí nghiệm trọng điểm Công nghệ gen, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (VAST).

TÀI LIỆU THAM KHẢO

Ariefdjohan MW, Savaiano DA, Nakatsu CH (2010). Comparison of DNA extraction kits for PCR-DGGE analysis of human intestinal microbial communities from fecal specimens. *Nutr J*, 9: 23.

Bisgaard H, Li N, Bonnelykke K, Chawes BLK, Skov T, Paludan-Müller G, Stokholm J, Smith B, Krogfelt KA (2011). Reduced diversity of the intestinal microbiota during infancy is associated with increased risk of allergic disease at school age. *J Allergy Clin Immunol*, 128: 646-652.e5.

Costea PI, Zeller G, Sunagawa S, Pelletier E, Alberti A, Levenez F, Tramontano M, Driessen M, Hercog R, Jung FE, Kultima JR, Hayward MR, Coelho LP, Allen-Vercoe E, Bertrand L, Blaut M, Brown JRM, Carton T, Cools-Portier S, Daigneault M, Derrien M, Druesne A, de Vos W, Finlay BB, Flint HJ, Guarner F, Hattori M, Heilig H, Luna RA, van Hylckama Vlieg J, Junick J, Klymiuk I, Langella P, Le Chatelier E, Mai V, Manichanh C, Martin JC, Mery C, Morita H, O'Toole PW, Orvain C, Patil KR, Penders J, Persson S, Pons N, Popova M, Salonen A, Saulnier D, Scott KP, Singh B, Slezak K, Veiga P, Versalovic J, Zhao L, Zoetendal EG, Ehrlich SD, Dore J, Bork P (2017). Towards standards for human fecal sample processing in metagenomic studies. *Nat Biotechnol.* 35 (11):1069-1076.

Fei P, Li L, Cai X, Zhang X, Bai HJ, Jiang YJ, Feng Z, Guo L (2016). Differences in the biodiversity of the fecal microbiota of infants with Rotaviral diarrhea and healthy infants. *Jundishapur J Microbiol*, 9: e32356.

Forno E, Onderdonk AB, McCracken J, Litonjua AA, Laskey D, Delaney ML, Dubois AM, Gold DR, Ryan LM, Weiss ST, Celedon JC, (2008). Diversity of the gut microbiota and eczema in early life. *Clin Mol Allergy*. 6: 11.

Hovda MB, Sivertsvik M, Tore Lunestad B, Lorentzen G, Rosnes JT (2007) Characterisation of the dominant bacterial population in modified atmosphere packaged farmed halibut (*Hippoglossus hippoglossus*) based on 16S rDNA-DGGE. Food

Microbiol, 24: 362-371.

Huang Q, Fu WL (2005) Comparative analysis of the DNA staining efficiencies of different fluorescent dyes in preparative agarose gel electrophoresis. *Clin Chem Lab Med*, 43: 841–842.

Mai V, Braden CR, Heckendorf J, Pironis B, Hirshon JM (2006). Monitoring of stool microbiota in subjects with diarrhea indicates distortions in composition. *J Clin Microbiol* 44: 4550-4552.

Nguyễn Hồng Dương, Nguyễn Thị Quý, Ngọc Thu Thảo, Nghiêm Đỗ Như Thảo, Phùng Thị Bích Thủy, Đỗ Thị Huyền (2023). Nghiên cứu phương pháp PCR-RFLP gen 16S rRNA để đánh giá đa dạng vi khuẩn trong phân trẻ tiêu chảy không rõ nguyên nhân. *Tạp chí Y học Việt Nam*, 526: 182-189.

Nguyen TQ, Phung TBT, Ngoc TT, Do TH (2023). A comparison of five methods for effective extraction of bacterial metagenomic DNA from stools of children suffer from diarrhea with unknown cause. *Biol Forum - Int J*, 15: 1-9.

Qiu S, Chen J, Lin S, Lin X (2012). A comparison of silver staining protocols for detecting DNA in polyester-backed polyacrylamide gel. *Braz J Microbiol*, 43: 649–652.

Sambrook J, Fritsch EF, Maniatis T (2001). Molecular cloning: a laboratory manual. 11. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Štšepetova J, Sepp E, Julge K, Vaughan E, Mikelsaar M, de Vos WM (2007). Molecularly assessed shifts of *Bifidobacterium* ssp. and less diverse microbial communities are characteristic of 5-year-old allergic children. *FEMS Immunol Med Microbiol*, 51: 260-269.

Tuma RS, Beaudet MP, Jin X, Jones LJ, Cheung CY, Yue S, Singer VL (1999). Characterization of SYBR Gold nucleic acid gel stain: a dye optimized for use with 300-nm ultraviolet transilluminators. *Anal Biochem*, 268: 278-288.

Vũ Thị Thu Hà, Nguyễn Tuấn Khiêm, Tăng Chí Thượng, Trần Thị Mộng Hiệp (2016). Khảo sát đặc điểm bệnh tiêu chảy kéo dài ở trẻ em tại bệnh viện Nhi đồng 2. Nghiên Cứu Y Học - Y Học TP Hồ Chí Minh, 20: 96-102.

Westaway JAF, Huerlimann R, Miller CM, Kandasamy Y, Norton R, Rudd D (2021). Methods for exploring the faecal microbiome of premature infants: a review. *Matern Health Neonatol Perinatol*, 7: 11.

INVESTIGATION OF PCR-DGGE METHOD FOR ANALYZING BACTERIAL DIVERSITY IN FECAL SAMPLES FROM CHILDREN OF 6-24 MONTHS OLD

Nguyen Thi Quy¹, Dao Trong Khoa¹, Nguyen Thi Thom^{2,3}, Do Thi Huyen^{1,2*}

¹Institute of Biotechnology, Vietnam Academy of Science and Technology

²Graduate University of Science and Technology, Vietnam Academy of Science and Technology

³19-8 Hospital, Ministry of public security

SUMMARY

Bacterial diversity and bacterial dysbiosis in stools of children at 6-24 months old are related to many diseases such as diarrhea, autism, obesity, allergies, eczema, etc. Analysis of bacterial diversity by denaturing gradient gel electrophoresis (DGGE) method allows spontaneously performing on a large number of samples, capable of comparing the structure of microbial diversity in the samples and suggesting the status of bacterial imbalance or possible pathogens presence in the feces. To successfully apply the DGGE method for analysis of the bacterial structure and diversity in the feces of healthy children and persistent diarrheal children with undetected-pathogens, in this study V3, V6-V8 regions of the 16S rDNA gene were amplified from the genomic DNA of 8 bacteria, 5 bacterial metagenomic DNA extracted from the feces of healthy children and a metagenomic DNA from feces of a child with persistent diarrhea, using tough-up, tough-down and specific PCR programs. The results showed that DGGE gel contained 8% polyacrylamide concentration was suitable for separating diverse V3 regions of bacteria. The V3 region that was amplified by tough-up program exhibited more diverse than one amplified by other PCR programs. The V3 DNA content of 1 µg per well was enough for analyzing this region by DGGE. DGGE gel stained with SYBR-Gold presented good sharpness and sensitivity for analysis of bacterial diversity in research samples.

Keywords: DGGE, PCR tough-up program, bacterial diversity, persistent diarrhea, V3 gene region.

Author for correspondence: Tel: +84-4-7917021; Email: dohuyen@ibt.ac.vn

KHẢO SÁT *in silico* MỘT SỐ ĐẶC ĐIỀM CỦA BETA-GLUCOSIDASE TỪ DỮ LIỆU TRÌNH TỰ DNA METAGENOME 8.4 GB CỦA VI KHUẦN TRONG DẠ CỎ DÊ

Đỗ Thị Huyền^{1,2*}, Nguyễn Minh Chiến^{2,3}, Trương Nam Hải¹

¹Viện Công nghệ sinh học - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Công ty TNHH Dược Hanvet

TÓM TẮT

Trong nghiên cứu trước, metagenomic DNA của vi khuẩn trong dạ cỏ dê đã được giải trình tự với kích thước dữ liệu là 8,4 Gb và 164,644 gen đã được dự đoán. Sử dụng các bộ dữ liệu CAZy, KEGG, COG, 384 trình tự mã hóa cho beta-glucosidase đã được khai thác trong đó có 131 trình tự được ước đoán là trình tự hoàn thiện. Phần lớn (93,1%) trình tự có kích thước lớn trên 700 amino acid, có chỉ số pI trong khoảng 5,0-6,5 mang vùng trung tâm xúc tác thuộc họ GH3 và có vùng phụ trợ FN3. Các enzyme được ước đoán có chỉ số alkaline cao, trong đó 67,9% enzyme có khả năng hoạt động ở vùng pH kiềm (chỉ số alkaline ≥ 0.6). Bốn enzyme có chỉ số pI cao nhưng được ước đoán hoạt động tốt ở vùng pH axit và 3 enzyme có chỉ số pI cao được ước đoán hoạt động ở vùng pH axit và 3 enzyme tiết, có trình tự tín hiệu tiết ở phía đầu N. Hơn nửa số enzyme (69 enzyme) được ước đoán bền ở nhiệt độ trên 65°C. Hai enzyme có kích thước thấp nhất được ước đoán có khả năng hoạt động ở pH kiềm. Các biến số độ dài gen, giá trị pI, giá trị alkaline, khả năng bền nhiệt, cấu trúc enzyme là các biến độc lập, chưa xác định được mối liên quan với nhau. Nghiên cứu này đã cung cấp những thông tin chung về một số đặc điểm cơ bản của beta-glucosidase được ước đoán dựa trên phân tích tin sinh học, có giá trị cho định hướng khai thác gen và biểu hiện gen.

Từ khóa: Beta-glucosidase, chỉ số alkaline, dữ liệu 8,4 Gb, in silico, vi khuẩn trong dạ cỏ dê.

MỞ ĐẦU

Beta-glucosidase (BGL, EC 3.2.1.21) là enzyme quan trọng được ứng dụng nhiều trong nhiều ngành công nghiệp, nông nghiệp, thực phẩm và y dược. BGL có khả năng xúc tác chuyển hóa gốc glucosyl từ đầu không khử của cả ba loại liên kết aryl-, alkyl và beta-glucoside trong disaccharide và cả các chuỗi oligosaccharide ngắn thành glucose và các chuỗi polysacchradide ngắn hơn. Enzyme được dùng trong chuyển hóa sinh khối lignocellulose cho sản xuất cồn sinh học. Ngoài ra, enzyme có vai trò quan trọng trong nhiều quá trình sinh học ở các cơ thể sống khác nhau và được ứng dụng rộng rãi trong công nghiệp thực phẩm, công nghiệp hương liệu, tái chế giấy, khử độc bột sắn và tăng tốc độ đường hóa, chuyển hóa chất xơ trong thức ăn gia súc, gia cầm, tăng hiệu quả sử dụng thức ăn, hấp thụ chất dinh dưỡng từ thức ăn cho động vật. Ở trong một số điều kiện nhất định, beta-glucosidase còn được sử dụng trong tổng hợp oligosaccharide, aryl-, alkyl-beta-glucoside. Đây là các hợp chất có tính ổn định cao, thân thiện với môi trường, bề mặt không bị ion hóa và tự phân hủy sinh học, được ứng dụng nhiều trong chế tạo thuốc, vaccine, sản phẩm làm đẹp, công cụ trong chấn đoán,... (Seeberger, Werz, 2007; Tran *et al.*, 2023; Ma *et al.*, 2024). Vì vậy enzyme này rất được các nhóm nghiên cứu trên thế giới quan tâm nghiên cứu, tìm kiếm để ứng dụng vào công nghiệp.

Hiện nay, việc ứng dụng BGL vào thực tế còn nhiều hạn chế do enzyme thường có hoạt độ thấp, chỉ hoạt động tốt nhất ở nhiệt độ 40–70°C và pH 4,5–5, nhưng các ứng dụng công nghiệp thường cần enzyme hoạt động ở phố pH rộng và nhiệt độ nằm ngoài giới hạn này. Ví dụ, các bước tiền xử lý sinh khối trong quá trình sản xuất nhiên liệu sinh học thường hoạt động ở nhiệt độ trên 80°C (Rocha *et al.*, 2023). Để tăng hương vị cho nước ép trái cây, các enzyme có hoạt tính tối ưu ở pH acid 2,8–3,8 sẽ thích hợp để giải phóng các chất bay hơi có chứa kết glycoside (Godse *et al.*, 2021). BGL thường rất nhạy cảm với glucose nên dễ bị ức chế nếu sản phẩm chuyển hóa tăng cao (Ahmed *et al.*, 2017). Vì vậy, việc nghiên cứu mở rộng tìm kiếm các nguồn gen mã hóa cho các BGL có đặc tính quý, được ước đoán từ dữ liệu DNA metagenome từ các hệ sinh thái tiềm năng là rất cần thiết. Hiện nay, công nghệ Metagenomics đã được ứng dụng để giải trình tự đa hệ gen của vi sinh vật trong hệ sinh thái đã cho ra những bộ dữ liệu gen rất lớn từ nguồn vi sinh vật chưa nuôi cấy được. Đây là nguồn gen rất mới, chưa được khai thác, mang những thông tin mới về cấu trúc của các loại protein/enzyme khác nhau. Cùng với sự phát triển của công nghệ thông tin, việc khảo sát cấu trúc và tính chất của enzyme từ dữ liệu lớn này có thể thực hiện được để đưa ra những thông tin mới về cấu trúc cũng như đặc điểm của nhóm enzyme trong hệ sinh thái, cho phép lựa chọn enzyme có đặc tính quý cho nghiên cứu ứng dụng.

NGUYÊN VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu nghiên cứu

Dữ liệu DNA metagnome có kích thước 8,4 Gb của vi khuẩn trong dạ cỏ dê thu được từ giải trình tự bằng hệ thống HiSeq Illumina 2500 (Illumina, San Diego, Hoa Kỳ). Từ 8,4 Gb dữ liệu, chúng tôi đã lắp ráp và ước đoán được 164,644 gene. Sau đó sử dụng công cụ BLASTall, các gen này đã được so sánh với cơ sở dữ liệu CAZy (carbohydrate-active enzymes database- CAZy, http://www.cazy.org), Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa *et al.*, 2008), Cluster of Orthologous Groups (COG) (Powell *et al.*, 2011) để chú giải chức năng của từng gen (Do *et al.*, 2018a, 2018b).

Nghiên cứu khai thác các trình tự mã hóa beta-glucosidase từ bộ dữ liệu DNA metagenome

Từ các kết quả chú giải chức năng dựa trên bộ dữ liệu CAZy, KEGG, COG, các trình tự mã hóa cho BGL sẽ được trích xuất ra một file riêng. Các thông tin về mã gen, trình tự gen, trình tự amino acid suy diễn từ gen được trích xuất bằng công thức Vlookup trong Microsorf Excel. Độ tương đồng của trình tự amino acid với các trình tự trong dữ liệu KEGG/CAZy/NR NCBI sẽ được trích xuất theo cùng trình tự.

Nghiên cứu đánh giá đa dạng phổ pl và chỉ số axit/kiềm dự đoán của beta-glucosidase

Khối lượng phân tử và giá trị pl của mỗi trình tự enzyme BGL được ước đoán bằng công cụ ước đoán trong EXPASY (https://web.expasy.org/compute_pi/). Khoảng pH hoạt động của enzyme được ước đoán dựa trên phần mềm AcalPred (http://lin.uestc.edu.cn/server/AcalPred). Đây là công cụ có thể đạt được độ chính xác tổng thể là 94,01% (Fan *et al.*, 2013). Khi một trình tự amino acid nhập vào, phần mềm sẽ tự động tính toán và cho ra kết quả cuối cùng về khả năng chịu kiềm/axit của protein sau vài phút. Kết quả cuối là một giá trị trong khoảng giới hạn từ 0 đến 1 theo mức tăng của chỉ số truy vấn. Nếu việc khảo sát mức độ hoạt động trong vùng alkaline thì trình tự nào có chỉ số càng gần 1 là trình tự có khả năng hoạt động tốt ở vùng alkaline. Do vậy, đối với 1 trình tự, tổng giá trị alkaline va acid sẽ bằng 1.

Dự đoán khả năng chịu nhiệt của enzyme

Dựa trên thành phần và trình tự của các amino acid, liên kết hidro, liên kết Vanderwaals, tương tác kỵ nước và đặc điểm của các enzyme từ các sinh vật sống trong điều kiện môi trường có nhiệt độ cao (Ebrahimi *et al.*, 2011), Phần mềm của TBI (www.tbi.org.tw/tools) xây dựng trên số liệu của 150.000 protein chịu nhiệt độ khác nhau trong ngân hàng NCBI để dự đoán khả năng chịu nhiệt, dựa trên nguyên tắc tương đồng. Khả năng chịu nhiệt được dự đoán ở ba mức là trên 65°C, 55-65°C và dưới 55°C.

Dự đoán cấu trúc của enzyme

Một số công cụ dự đoán cấu trúc được ứng dụng nhiều trong những năm gần đây là Swiss-Prot (có tích hợp AlphaFold), Phyre2 hay Comparative modelling. Mỗi công cụ có ưu điểm riêng, phù hợp với từng nhóm trình tự có độ tương đồng với cấu trúc có sẵn khác nhau. Công cụ Swiss-Prot cho phép ước đoán cấu trúc phân tử của các trình tự mới, có độ tương đồng thấp, khó ước đoán được bằng Phyre2. Phyre2 là công cụ dự đoán cấu trúc bậc ba của các trình tự dựa trên nguyên lý nhận biết vùng cuộn xoắn theo các tương tác phân tử đối với các trình tự có độ tương đồng với cấu trúc đã biết trên 30%. Phương pháp Comparative modelling có thể giúp dự đoán cấu trúc phân tử đối với các trình tự có độ tương đồng cao trên 50% so với trình tự đã biết. Trong nghiên cứu này, các enzyme được dự đoán cấu trúc dựa trên trình tự amino acid bằng công cụ Phyre2 (http://www.sbg.bio.ic .ac.uk/~phyre2, truy cập tháng 6/2024) với chế độ mô hình hóa mở rộng (intensive) do chúng có độ tương đồng củ trức đã biết khoảng trên 30%. Các vùng bảo thủ của enzyme được ước đoán dựa trên công cụ ước đoán vùng bảo thủ của gói công cụ trực tuyến BLAST.

KẾT QUẢ VÀ THẢO LUẬN

Khai thác các trình tự mã hóa beta-glucosidase

Từ 164,644 gene, chúng tôi đã thu được 387 trình tự mã hóa betaglucosidase. Trong số các trình tự này chỉ có 131 trình tự hoàn thiện (chiếm 29,2%) 88 trình tự thiếu đầu 3', 60 trình tự thiếu đầu 5' và 103 trình tự thiếu cả hai đầu. Trong số các trình tự hoàn thiện có 43 trình tự được chú giải chức năng dựa trên dữ liệu COG, 113 trình tự chú giải chức năng bởi KEGG (đạt 86,25%), 131 trình tự được chú giải bởi CAZy (đạt 100%). Như vậy, dữ liệu CAZY cho đạt tỉ lệ chú giải gen thành công cao nhất, đạt 100%.



Hình 1. Biểu đồ phân bố độ dài (A) và pl (B) của 131 trình tự amino acid của beta-glucosidase từ bộ dữ liệu DNA metagenome trong dạ cỏ dê

Kết quả khảo sát độ dài các trình tự cho thấy, trình tự gen có độ dài ngắn nhất là 257 amino acid và dài nhất là 894 amino acid. Trong đó có 122 trình tự (chiếm 93,1%) có độ dài trên 700 amino acid. Phần lớn các trình tự có chỉ số pl nghiêng về phía axit nhẹ, nằm trong khoảng 5,0-6,5 chỉ có 1 trình tự có chỉ số pl thấp dưới 5 và 6 trình tự có pl cao trên 7 (Hình 1B). Trình tự có kích thước thấp nhất được ước đoán có chỉ số pl thấp nhất. Không có mối liên hệ nào giữa độ dài và giá trị pl của beta-glucosidase.

Thông thường các trình tự beta-glucosidase GH1 thường có kích thước khoảng 400-500 amino acid. Điển hình như BGL GH1 của *Trichoderma* spp. thường có kích thước trong khoảng 455 đến 466 amino acid (Mohamad Sobri *et al.*, 2020). Các beta-glucosidase GH3 thường có kích thước lớn hơn do phân tử thường có 2 module cho hoạt tính xúc tác là GH3N và GH3C cùng với các vùng phụ trợ khác giúp enzyme bám lên cơ chất (Nguyen *et al.*, 2021). Trong nghiên cứu này, tất cả các trình tự đều được CAZy ước đoán là beta-glucosidase GH3 và phần lớn có thêm vùng FN3 ở phía sau vùng GH3C. Đây là cấu trúc khá điển hình của enzyme beta-glucosidase trong dạ cỏ dê. Các nghiên cứu cũng cho thấy điểm đắng điện của các beta-glucosidase trong bộ dữ liệu nghiêng về vùng axit nhẹ, trong khi đó các enzyme beta-glucosidase GH1 của *Trichoderma* spp. có điểm đẳng điện dao động trong khoảng 5,10-5.53 (Mohamad Sobri *et al.*, 2020).

Khảo sát khả năng hoạt động ở vùng pH kiềm của enzyme

Theo ước đoán bằng phần mềm AcalPred, các enzyme beta-glucosidase trong dạ có dê có khả năng hoạt động trong vùng pH rất rộng, từ vùng pH rất axit (với chỉ số alkaline thấp nhất (0,05)) đến vùng hoạt động rất kiềm (chỉ số alkaline cao nhất là 1). Nếu phân loại chỉ số alkaline thành 3 vùng là vùng axit (chỉ số alkaline ≤0,4) vùng trung tính (lớn hơn 0,4 và nhỏ hơn 0,6) và vùng kiềm (từ 0,6 đến 1,0) thì 30 enzyme hoạt động ở vùng axit, 12 enzyme hoạt động ở vùng trung tính và 89 enzyme (chiếm 67,9%) hoạt động ở vùng pH kiềm (Hình 2). Quan sát trên biểu đồ (Hình 2B) có thể thấy, sự phân bố enzyme hoạt động ở các vùng pH khác nhau là ngẫu nhiên, không có mối liên quan tới điểm đẳng điện của enzyme. Tuy nhiên, có 4 enzyme có giá trị pI cao lại hoạt động ở vùng pH cao và cũng không nhìn thấy enzyme có pI thấp hoạt động ở vùng pH thấp (Bảng 1). Kết quả khảo sát mối liên quan giữa độ dài trình tự và giá trị alkaline của enzym cũng cho thấy hai đại lượng này không có mối liên hệ với nhau (kết quả không được trình bày).



Hình 2. Biểu đồ phân bố số lượng các trình tự beta-glucosidase theo chỉ số alkaline (A) và sự tương quan giữa chỉ số alkaline với pl của 131 trình tự

Bảng 1. Một số đặc điểm của 4 trình tự có pl cao hoạt động trong vùng pH thấp và 3 trình tự pl cao hoạt động trong
vùng pH cao trong bộ dữ liệu DNA metagenome của vi khuẩn trong dạ cỏ dê

Mã trình tự	Độ dài (amino acid)	Khối lượng (kDa)	pl	Vùng xúc tác	Vùng đầu C	Độ tương đồng (%)	Mã CaZy/ KEGG	Điểm số acid	Điểm số alkaline	Tm
[denovogenes]_3474	819	90,6	8,62	GH3		40,93	CBK66838	0,29	0,71	>65
[denovogenes]_17198	493	54,1	8,35	GH3	FN3	46,36	K05349	0,78	0,22	>65
[denovogenes]_5059	741	81,7	8,32	GH3	FN3	44,48	K05349	0,21	0,79	55-65
[denovogenes]_11003	580	64,1	8,25	GH3	FN3	86,14	K05349	0,65	0,35	55-65
[denovogenes]_2683	869	97,0	8,11	GH3	FN3	76,46	K01188	0,95	0,05	55-65
[denovogenes]_2632	874	96,9	8,1	GH3	FN3	76,62	K01188	0,77	0,23	>65
[denovogenes]_4363	775	84,4	7,48	GH3	FN3	60,57	K05349	0,06	0,94	55-65

Dựa trên phần mềm SignalP 5.0 của Trường Đại học Kỹ thuật Đan Mạch (https://services.healthtech.dtu.dk /services/SignalP-5.0/), trong số 7 trình tự ở trên, các trình tự có kích thước lớn hơn 700 amino acid đều có chứa tín hiệu tiết, còn hai trình tự kích thước nhỏ không chứa tín hiệu tiết. Tất cả các trình tự đều có hai vùng xúc tác của GH3N và GH3C trong đó 5 trình tự có vùng FN3 được cấu tạo bởi phiến beta. Các trình tự đều có cấu trúc đặc thù của beta-glucosidase gồm cấu trúc TIM (β/α)8 barrel, (α/β)6 barrel (Hình 3). Các cấu trúc beta glucosidase đều tương đồng 31-51% so với cấu trúc của beta-glucosidase trong dữ liệu PDB. Trình tự FN3 đã được nghiên cứu có khả năng làm tăng giá trị pl cho enzyme (Nguyen *et al.*, 2021). Enzyme mã hóa từ gen [denovogenes]_3474 có cấu trúc tương đồng 33% với glycosyl hydrolase family 3 (Kim *et al.*, 2011). Enzyme mã hóa từ gen [denovogenes]_5059 có cấu trúc tương đồng 37% với beta-glucosidase chịu nhiệt của *Clostridium thermocellum* (Almeida, Muniz, 2022). Quan sát cấu trúc của các enzyme, chúng tôi không nhận thấy các điểm khác biệt giữa các enzyme có pl cao nhưng có khả năng hoạt động ở vùng pH axit và enzyme có pl cao nhưng dự đoán hoạt động trong vùng pH bazơ.



Hình 3. Cấu trúc dự đoán của hai beta-glucosidase điển hình mã hóa từ hai trình tự gen

Khảo sát khả năng bền nhiệt của các enzyme beta-glucosidase

Dựa trên phần mềm TBI (www.tbi.org.tw/tools), khả năng bền nhiệt của enzyme thông qua chỉ số Tm đã được ước đoán. Kết quả, 100% các trình tự được ước đoán có khả năng bền nhiệt cao từ 55°C trở lên và không có trình tư nào bền nhiệt thấp dưới 55°C. Trong đó 53% số lượng enzyme (69 enzyme) được ước đoán bền ở nhiệt độ trên 65°C, đồng nghĩa là 53% enzyme sẽ bị mất hoạt tính khi nhiệt độ tăng lên trên 65°C và 47% số enzyme còn lại bị mất hoạt tính ở nhiệt độ 55-65°C. Kết quả khảo sát mối liên hệ giữa độ dài trình tự với khả năng bền nhiệt của enzyme cho thấy, hai đại lượng này không có mối quan hệ vởi nhau. Trong đó đáng chú ý là hai enzyme có trình tự ngắn nhất lại được dự đoán hoạt động ở trong vùng pH kiềm. Trình tự [denovogenes]_68027 tạo thành từ 257 amino acid là trình tự ngắn nhất trong bộ dữ liệu, tiếp theo đó là trình tự [denovogenes]_39321 có 352 amino acid. Kết quả khảo sát cấu trúc của enzyme bằng Phyre2 cho thấy cả hai trình tự có cấu trúc tương đồng cao với beta-glucosidase trong đó enzyme [denovogenes] 68027 được ước đoán có cấu trúc tương đồng 52% so với cấu trúc của beta-glucosidase bglpcc1 của Prevotella copri (Zhou et al., 2023). Enzyme mã hóa từ gen còn lại có cấu trúc tương đồng 45% với cấu trúc của beta-glucosidase của Kluyveromyces marxianus (Yoshida, Hidaka, 2010). Cả 2 enzyme đều có cấu trúc FN3 ở vùng đầu C và một phần của cấu trúc (α/β)6 barrel ở đầu N. Do vậy rất có thể enzyme bị thiếu đi vùng GH3N. Tuy nhiên, để chắc chẳn đây có phải là enzyme hoàn thiên, có thể hiên chức năng sinh học hay không thì cần phải được nghiên cứu tiếp bằng sinh học phân tử tái tổ hợp gen.



A

Hình 4. Biểu đồ phân bố các enzyme beta-glucosidase dựa theo độ dài, giá trị alkaline và khả năng bền nhiệt của các enzyme (A), cấu trúc ước đoán của hai enzyme mã hóa từ hai gen điển hình (B, C)

KÉT LUẬN

Trong số 131 trình tự hoàn thiện mã hóa cho beta-glucosidase được khai thác từ dữ liệu giải trình tự đa hệ gen của vi khuẩn trong dạ cỏ dê, 100% trình tự có vùng xúc tác GH3, 93,1% trình tự có kích thước lớn trên 700 amino acid, 67,9% enzyme có khả năng hoạt động ở vùng pH kiềm (chỉ số alkaline ≥0,6), 52,7% enzyme ước đoán bền ở nhiệt độ trên 65°C. Kết quả này có ý nghĩa cho lựa chọn trình tự mã hóa enzyme có đặc tính mong muốn cho biểu hiện protein dạng tái tố hợp.

Lời cảm ơn: Công trình được hỗ trợ kinh phí từ đề tài nghiên cứu mã số NCVCC08.05/24-25 do PGS. TS Đỗ Thị Huyền làm chủ nhiệm.

TÀI LIỆU THAM KHẢO

Ahmed A, Nasim F, ul-Hassan, Batool K, Bibi A (2017). Microbial β -glucosidase: Sources, production and applications. *J Appl Environ Microbiol* 5: 31–46.

Almeida L, Muniz J (2022). RCSB PDB - 7MS2: Three-dimensional structure of a GH3 beta-glucosidase from *Clostridium* thermocellum in complex with glycerol.

Do TH, Dao TK, Nguyen KHV, Le NG, Nguyen TMP, Le TL, Phung TN, van Straalen NM, Roelofs D, Truong NH (2018a). Metagenomic analysis of bacterial community structure and diversity of lignocellulolytic bacteria in Vietnamese native goat rumen. *Asian-Australas J Anim Sci* 31: 738–747.

Do TH, Le NG, Dao TK, Nguyen TMP, Le TL, Luu HL, Nguyen KHV, Nguyen VL, Le LA, Phung TN, van Straalen NM, Roelofs D, Truong NH (2018b). Metagenomic insights into lignocellulose-degrading genes through Illumina-based de novo sequencing of the microbiome in Vietnamese native goats' rumen. *J Gen Appl Microbiol* 64: 108–116.

Ahmed A, Nasim F ul-Hassan, Batool K, Bibi A (2017). Microbial β -glucosidase: Sources, production and applications. *J Appl Environ Microbiol*, 5: 31–46.

Almeida L, Muniz J (2022). RCSB PDB - 7MS2: Three-dimensional structure of a GH3 beta-glucosidase from *Clostridium thermocellum* in complex with glycerol.

Ebrahimi M, Lakizadeh A, Agha-Golzadeh P, Ebrahimie E, Ebrahimi M (2011). Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PLoS ONE*,6.

Fan GL, Li QZ, Zuo YC (2013). Predicting acidic and alkaline enzymes by incorporating the average chemical shift and gene ontology informations into the general form of Chou's PseAAC. *Process Biochem*, 48: 1048–1053.

Godse R, Bawane H, Tripathi J, Kulkarni R (2021). Unconventional β -glucosidases: A promising biocatalyst for industrial biotechnology. *Appl Biochem Biotechnol*, 193: 2993–3016.

Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res*, 36: D480-484.

Kim Y, Chhor G, Bearden J, Joachimiak A (2011). X-ray diffraction data for the crystal structure of glycoside hydrolase from *Synechococcus* (3SQL). *IRRMC*.

Ma LJ, Liu X, Guo L, Luo Y, Zhang B, Cui X, Yang K, Cai J, Liu F, Ma N, Yang FQ, He X, Shi SP, Wan JB (2024). Discovery of plant chemical defence mediated by a two-component system involving β-glucosidase in *Panax* species. *Nat Commun*, 15: 602.

Mohamad Sobri MF, Abd-Aziz S, Abu Bakar FD, Ramli N (2020). *In-silico* characterization of glycosyl hydrolase family 1 β-glucosidase from *Trichoderma asperellum* UPM1. *Int J Mol Sci*, 21: 4035.

Nguyen KHV, Dao TK, Nguyen HD, Nguyen K. H., Nguyen TQ, Nguyen TT, Nguyen TMP, Truong NH, Do TH (2021). Some characters of bacterial cellulases in goats' rumen elucidated by metagenomic DNA analysis and the role of fibronectin 3 module for endoglucanase function. *Anim Biosci*, 34: 867–879.

Powell S, Szklarczyk D, Trachana K, Roth A., Kuhn M, Muller J, Arnold R, Rattei T, Letunic, Doerks T, Jensen LJ, von Mering C, Bork P (2011). eggNOG v3.0: orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res,* 40: D284–D289.

Rocha REO, Mariano DCB, Almeida TS, CorrêaCosta LS, Fischer PHC, Santos LH, Caffarena ER, da Silveira CH, Lamp LM, Fernandez-Quintero ML, Liedl KR, de Melo-Minardi RC, de Lima LHF (2023). Thermostabilizing mechanisms of canonical single amino acid substitutions at a GH1 β-glucosidase probed by multiple MD and computational approaches. *Proteins*, 91: 218–236.

Seeberger PH, Werz DB (2007). Synthesis and medical applications of oligosaccharides. Nature, 446: 1046–1051.

Tran TNA, Son JS, Awais M, Ko JH, Yang DC, Jung SK (2023). β-Glucosidase and its application in bioconversion of ginsenosides in *Panax ginseng. Bioeng Basel Switz*, 10: 484.

Yoshida E, Hidaka M (2010). RCSB PDB - 3AC0: Crystal structure of Beta-glucosidase from *Kluyveromyces marxianus* in complex with glucose.

Zhou J, Zhu H, Zhang R, Huang Z (2023). RCSB PDB - 8IDZ: Beta-glucosidase BgIPcC1.

In silico PREDICTION OF SOME PROPERTIES OF BETA-GLUCOSIDASE MINED FROM METAGENOMIC DNA DATA 8.4 GB OF BACTERIA IN GOATS RUMEN

Do Thi Huyen^{1,2*}, Nguyen Minh Chien^{2,3}, Truong Nam Hai¹

¹Institute of Biotechnology–Vietnam Academy of Science and Technology

²Graduate University of Science and Technology - Vietnam Academy of Science and Technology

³Hanvet Pharmaceutical Company Limited

SUMMARY

In a previous study, metagenomic DNA of bacteria in goats' rumen was sequenced to yield of 8.4 Gb, from which 164 644 genes were predicted. Using CAZy, KEGG, COG databases, 384 genes coding for betaglucosidase were mined, of which 131 sequences were estimated to be complete sequences. The majority proportion (93.1%) of the enzymes coded by the complete genes had long size of over 700 amino acids. The significant portion of the enzymes possessed pI value in the range of 5.0-6.5, contained a catalytic domain belonging to GH3 family and an auxiliary region FN3. The enzymes were estimated to have a high alkaline score, of which 67.9% of enzymes were predicted to be capable of exhibiting activity in the alkaline pH region (alkaline index ≥ 0.6). Among them, four enzymes have high pI scores but were estimated to work well in the acidic pH region and three enzymes had high pI scores but were estimated to work well in the alkaline pH region. Of these seven enzymes, five enzymes were identified as secretory enzymes because of containing a signal peptide sequence at the N-terminus. More than half of the enzymes (69 enzymes) were estimated to be stable at temperatures up to 65° C. The two enzymes with the lowest sizes were predicted to be able to function at alkaline pH. The amino acid length, pI value, alkaline value, Tm and enzyme structure were independent variables, no relationship between each others were observed. This study has provided general information about some basic characteristics of beta-glucosidase estimated based on bioinformatics analysis, which is valuable for gene mining and expression.

Keywords: DGGE, PCR tough-up program, bacterial diversity, persistent diarrhea, V3 gene region.

^{*} Author for correspondence: Tel: 0437917021; Email: dohuyen@ibt.ac.vn

NHÂN DÒNG VÀ PHÂN TÍCH TRÌNH TỰ GEN glpF MÃ HÓA PROTEIN XUYÊN MÀNG HÕ TRỢ VẬN CHUYỄN GLYCEROL Ở Escherichia coli

Phạm Thị Kiều Duyên¹, Nguyễn Thị An Hòa¹, Nguyễn Hà Mi¹, Quyền Mỹ Linh¹, Nguyễn Quang Huy^{1,2}, Kiều Cẩm Nhung³, Chử Lương Luân^{2,*}

¹Khoa Sinh học, Trường Đại học Khoa học Tự Nhiên, Đại học Quốc gia Hà Nội

²Phòng Thí nghiệm Trọng điểm Công nghệ Enzym và Protein, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

³Bộ Khoa học và Công nghệ

TÓM TẮT

Biodiesel là nguồn năng lượng tái tạo, hứa hẹn có thể thay thế cho nhiên liệu hóa thạch truyền thống bởi rư điểm thân thiện với môi trường. Sự phát triển của công nghiệp biodiesel trên phạm vi toàn cầu đã dẫn đến dư thừa một trữ lượng lớn glycerol thô - sản phẩm phụ từ quá trình sản xuất. Điều này gây mất cân bằng giá thành glycerol thương mại và được xem là thách thức đối với tính bền vững của công nghiệp biodiesel. Từ đó, mục tiêu chuyển đổi glycerol thô thành sản phẩm có giá trị cao đã trở thành vấn đề cấp thiết ở nhiều quốc gia. *Escherichia coli (E. coli)* đã được chứng minh có thể sử dụng glycerol như một nguồn carbon thay thế thông qua sự kết hợp giữa glycerol kinase (GlpK) và glycerol 3- phosphate dehydrogenase (GlpD). Tuy nhiên, sự biểu hiện của gen aquaglyceroporin (*glpF*) là mâu chốt quan trọng, làm tăng tốc độ sinh trưởng và sự hình thành sinh khối. GlpF có cấu trúc bảo tồn, tính chọn lọc cao và được biết đến là protein vận chuyển duy nhất xúc tác cho quá trình khuếch tán glycerol qua màng trong của *E. coli*. Trong nghiên cứu này, gen *glpF* từ *E. coli* chủng DH5 alpha đã được nhân dòng vào vector pJET 1.2/ blunt và giải trình tự theo phương pháp Sanger. Trình tự gen đích được giải mã gồm 846 nucleotide và tương đồng 100% với trình tự gen tham chiếu. Kết quả so sánh trình tự amino acid suy diễn và phân tích cấu trúc protein bậc 3 cho kết quả tương đồng 100% với độ tin cậy là tuyệt đối. Kết quả nghiên cứu này cung cấp cơ sở khoa học, tạo nguyên liệu di truyền để phát triển chủng vi sinh vật tái tổ hợp giúp tăng cường biểu hiện gen *glpF* mã hóa cho protein xuyên màng hỗ trợ vận chuyển glycerol.

Từ khóa: Escherichia coli, glycerol, *glpF*, trình tự gen, nhân dòng.

MỞ ĐẦU

Biodiesel là nhiên liêu có thể tái chế, được sản xuất từ dầu thực vật hoặc mỡ động vật với mục tiêu giảm thiểu ô nhiễm môi trường và thay thế cho nhiên liệu hóa thach. Tuy nhiên, quá trình chế tạo dầu biodiesel dẫn đến dự thừa một lượng lớn glycerol - sản phẩm phụ của quá trình sản xuất. Theo tính toán, mỗi 9 kg dầu biodiesel được sản xuất từ quá trình este hóa sẽ tạo ra 1 kg glycerol thô (Zhou et al., 2018). Mặc dù, từng được coi là loại "sản phẩm phu" có giá tri, glycerol thô đang nhanh chóng trở thành một "sản phẩm thải" do chi phí xử lý cao và ảnh hưởng tới giá thành của glycerol tinh khiết. Theo nghiên cứu, sản lượng glycerol thô trên thị trường tăng nhanh và có liên quan trực tiếp tới sự phát triển của biodiesel. Mặt khác, glycerol thô chưa qua tinh chế không thể thải ra môi trường vì đốt trực tiếp sẽ tạo ra khí độc (Gupta & Kumar, 2012). Việc phát triển một quy trình bền vững với mục đích chuyển đổi glycerol thành sản phẩm có giá tri đang là vấn đề cấp thiết của các lĩnh vực kinh tế, năng lượng và công nghệ (Chiang et al., 2020). Hiện nay, phương pháp chuyển đổi glycerol thường được tiếp cân bao gồm chuyển hóa hóa học và đặc biệt là lên men sử dụng vi sinh vật cải biến di truyền. Một số vi khuẩn có thể chuyển hóa glycerol thuộc chi E. coli, Klebsiella, Enterobacter, Glucanobacter, Clostridium, Candida, Aspergillus. Glycerol thô thường được chuyển đổi thành các sản phẩm có giá trị như: ethanol, propionic acid, biogas, citric acid, ethanol, acid docosahexaenoic và một số acid béo omega-3 khác (Chilakamarry et al., 2021). Với lợi ích về chi phí và bảo vệ môi trường, công nghệ DNA tái tổ hợp được xem là lĩnh vực triển vọng trong nghiên cứu chuyển hóa glycerol thô.

E. coli mang plasmid thường được sử dụng trong lĩnh vực nghiên cứu tái tổ hợp với ưu điểm về thao tác di truyền và chi phí mở rộng sản xuất. Vi khuẩn này có khả năng sử dụng một số hợp chất làm nguồn carbon và glucose được ưu tiên sử dụng do ái lực cao với hệ thống phosphotransferase (PTS). Mặt khác, glucose cung cấp nhiều năng lượng cho tế bào thông qua việc khởi động quá trình đường phân dưới dạng glucose 6-phosphate và được tiêu thụ qua chu trình Krebs (Martínez-Gómez *et al.*, 2012). Tuy nhiên, *E. coli* có thể sinh trưởng trong môi trường giàu glycerol và sử dụng chúng như một nguồn carbon thay thế. Mặt khác, người ta còn chứng minh rằng glycerol có ảnh hưởng tích cực tới thời gian sống của tế bào và năng suất protein tái tổ hợp.

Glycerol có thể đi qua màng tế bào thông qua cơ chế khuếch tán thụ động không đặc hiệu, tích lũy ở tế bào chất và có thể di chuyển theo gradient nồng độ ra khỏi tế bào. Ngoài ra, glycerol cũng được xem là ví dụ điển hình cho

quá trình khuếch tán qua kênh vận chuyển trung gian ở E. coli. Tương tự như các aquaporin, GIpF là một kênh xuyên màng có vai trò thúc đẩy quá trình thẩm thấu glycerol. Sau khi qua màng tế bào, glycerol trong tế bào chất trải qua quá trình phosphoryl hóa cần ATP và được xúc tác bởi glycerol kinase (GlpK). Sản phẩm của quá trình được giữ lại trong tế bào dưới dạng glycerol-3phosphate (G3P). Quá trình trình khử hydro diễn ra và dihydroxyacetone phosphate (DHAP) được tạo thành. G3P đóng vai trò là chất cảm ứng của operon glpFK và giúp E. coli có thể sử dụng glycerol hiệu quả. Mặt khác, G3P không phải chất nền cho GlpF nên được giữ lại bên trong tế bào mà không phụ thuộc gradient nồng độ (Sweet et al., 1990). GIpF thuộc họ aquaporin (AQP), chứa chuỗi -Asn-Pro-Ala- bảo tồn, đã được xác định rõ về cấp độ di truyền và các đặc tính chức năng (Stroud et al., 2003). GlpF gồm 1 xoắn dài xuyên màng với một vị trí uốn cong ở giữa và 2 chuỗi xoắn alpha xuyên màng một nửa, xiên chéo qua màng hình thành một phần của lỗ chọn lọc glycerol. GlpF có tính chọn lọc cao, không cho phép các hợp chất tích điện đi qua giúp duy trì gradient điện hóa trên màng. Trong một số nghiên cứu, glpF được xác đinh mã hóa cho protein xuyên màng GlpF và nằm chung operon với *glpK* mã hóa cho glycerol kinase (GlpK) (Sweet et al., 1990). Hai protein GlpF và GlpK có quan hệ tương tác, sự xuất hiện của GlpF sẽ kích hoat glycerol kinase (Voegele et al., 1993). Thêm vào đó, tốc độ vận chuyển glycerol qua GlpF nhanh hơn từ 100-1000 lần so với các phương thức vận chuyển khác (Heller et al., 1980). Tại Việt Nam, đã có nghiên cứu tập trung vào tối ưu giá thành sản xuất biodiesel thay vì tận dụng nguồn sản phẩm có giá trị, sử dụng đối tượng: vi tảo biển, mỡ động vật. Tuy nhiên, những đối tượng nghiên cứu này còn giới hạn về tốc độ sinh trưởng, mùa vụ và khả năng mở rộng sản xuất. Chưa có nghiên cứu nào sử dụng công nghiệp DNA tái tổ hợp trong nghiên cứu chuyển hóa trữ lượng glycerol thô dư thừa từ công nghiệp biodiesel.

Tóm lại, biểu hiện *glpF* được coi là mấu chốt quan trọng, làm tăng tốc độ sinh trưởng và sự hình thành sinh khối. Trong nghiên cứu này, chúng tôi tiến hành nhân dòng và phân tích trình tự *glpF*, bước đầu tạo nguyên liệu di truyền cho phát triển chủng tái tổ hợp mang biểu hiện gen quan tâm. Đây là một trong những nghiên cứu đầu tiên tại Việt Nam tiếp cận tới định hướng tận thu glycerol thô từ công nghiệp biodiesel sử dụng công nghệ tái tổ hợp.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Chủng vi khuẩn và plasmid

E. coli chủng DH5 alpha (Invitrogen, Hoa Kỳ) được sử dụng cho mục đích tách chiết DNA tổng số và thiết kế các bước của thí nghiệm nhân dòng. Vector tạo dòng dùng sử dụng trong nghiên cứu là pJET 1.2/blunt (Invitrogen, Hoa Kỳ).

Tách chiết genomic DNA

E. coli DH5 alpha được nuôi lắc tăng sinh trong 5 mL LB lỏng (160 vòng/phút, 37°C qua đêm). DNA tổng số của *E. coli* DH5 alpha được tách chiết bằng bộ sinh phẩm thương mại Genomic DNA Prep Kit (BioFact, Hàn Quốc). Sản phẩm tinh sạch được điện di trên gel agarose 1% có bổ sung thuốc nhuộm RedSafeTM Nucleic Acid Staining Solution 20000x (iNtRON Biotechnology, Hàn Quốc). Kết quả điện di được quan sát dưới tia tử ngoại.

Thiết kế cặp mồi đặc hiệu khuếch đại gen glpF

Cặp mồi đặc hiệu khuếch đại gen *glpF* được chúng tôi nghiên cứu thiết kế dựa trên trình tự đã được công bố trên ngân hàng gen (GenBank: U00096.3). Bộ mồi được tổng hợp bằng phương pháp hóa học bởi Công ty TNHH MTV Sinh Hóa Phù Sa (PHUSA Biochem) và có trình tự: *glpF*-fw (5' – ATGAGTCAAACATCAACCTTG – 3'); *glpF-Afl* II- rv (5' –CTTAAGTTACAGCGAAGCTTTTTG – 3'). Phản ứng PCR khuếch đại gen *glpF* sử dụng bộ sinh phẩm Platinum™ SuperFi II PCR Master Mix (2X) (Invitrogen, Hoa Kỳ) chứa *Pfu* DNA polymerase biểu hiện hoạt tính exonuclease 3'-5' có khả năng đọc sửa và khuếch đại chính xác gen đích. Quá trình khuếch đại (PCR) được thực hiện bằng thiết bị luân nhiệt SimpliAmp™ Thermal Cycler (Applied Biosystem, Hoa Kỳ) theo chu trình: biến tính 95°C trong 5 phút; 35 chu kỳ với mỗi chu kỳ là 95°C trong 30 giây, 55°C trong 45 giây và 72°C trong 1 phút; cuối cùng giữ 72°C trong 10 phút. Sản phẩm PCR có dạng đầu bằng, kích thước theo tính toán là 846 bp. Kết quả khuếch đại được kiểm tra bằng điện di trên gel agarose TAE 1% ở điện áp 100V trong 25 phút. Tinh sạch sản phẩm PCR sử dụng Monarch DNA Gel Extraction Kit (NEB, Hoa Kỳ). Kiểm tra kích thước sản phẩm tinh sạch bằng kỹ thuật điện di trên gel agarose. Trong nghiên cứu này, công cụ ImageJ (Hoa Kỳ) được sử dụng với mục đích phân tích hình ảnh điện di.

Phản ứng gắn glpF vào vector nhân dòng

Sau khi tinh chế, đoạn gen *glpF* được chèn trực tiếp vào vector pJET 1.2/blunt (Invitrogen, Hoa Kỳ) theo tỉ lệ: 1 μL vector pJET 1.2/blunt, 5 μL 2X reaction buffer, 1 μL T4 ADN ligase, 1,5 μL sản phẩm tinh sạch gen *glpF*, bổ sung thêm nuclease free water để đạt thể tích cuối cùng là 10 μL. Phản ứng nối được ủ ở nhiệt độ phòng (25°C) trong thời gian 2 giờ trước khi được sử dụng cho mục đích biến nạp.

Biến nạp và PCR khuẩn lạc sàng lọc thể biến nạp

Toàn bộ sản phẩm gắn được trộn với 100 μL tế bào khả biến *E. coli* DH5 alpha. Hỗn hợp được giữ trên đá lạnh 15 phút, sau đó sốc nhiệt ở 42°C trong 1 phút và đặt lên đá lạnh trong 2 phút. Bổ sung 300 μL LB lỏng vào hỗn

hợp biến nạp, nuôi lắc tăng sinh ở 37°C, 160 rpm trong ít nhất 1 giờ. Cấy trải 200 μL hỗn hợp biến nạp trên môi trường LB đặc có bổ sung 50 μg/mL ampicillin và nuôi qua đêm trong tủ ấm 37°C. Chọn 10 khuẩn lạc riêng rẽ trên đĩa thạch LB, hòa tan vào 40 μL nước cất khử trùng và dùng làm khuôn cho phản ứng PCR sử dụng cặp mồi của bộ kit nhân dòng (pJET fw/rv).

Tách chiết, tinh sạch và định lượng plasmid tái tổ hợp

Khuẩn lạc chứa DNA đoạn chèn đúng kích thước được nuôi riêng rẽ ở 37°C, qua đêm trong 5 mL môi trường LB lỏng chứa kháng sinh ampicillin. Sau đó, 5 mL dịch nuôi cấy được ly tâm 12000 vòng/ phút trong thời gian 2 phút để thu cặn tế bào. DNA plasmid được tinh sạch sử dụng kit Exprep Plasmid SV (Gene All, Hàn Quốc). Sản phẩm tinh sạch được định lượng bằng đo độ hấp thụ ánh sáng tử ngoại bước sóng A₂₆₀.

Kiểm tra plasmid tái tổ hợp bằng enzyme cắt giới hạn

Vector tái tổ hợp pJET-*glpF* được ủ với 2 enzyme giới hạn là *Bgl*II và *Afl*I (NEB, Hoa Kỳ). Phản ứng cắt được thực hiện theo tỉ lệ: 1µL buffer 10X tương ứng, 1 µL DNA plasmid đã được tinh sạch, 1 µL enzyme giới hạn và bổ sung thêm nuclease free water để đạt thể tích cuối cùng là 10 µL. Ủ mẫu ở 37°C trong thời gian 45 phút trước khi điện di kiểm tra sản phẩm cắt.

Giải trình tự plasmid tái tổ hợp và phân tích bằng các công cụ tin sinh

Plasmid tái tổ hợp sau khi tinh sạch được dùng cho mục đích giải trình tự Sanger sử dụng cặp mồi đi kèm bộ kit nhân dòng pJET fw/rv. Giải trình tự được thực hiện tại công ty 1st BASE (Malaysia). Phân tích kết quả giải trình tự bằng phần mềm BioEdit. Công cụ SnapGene được sử dụng để đưa ra trình tự amino acid suy diễn. Sử dụng CLUSTAL OMEGA cho mục đích so sánh trình tự gen và trình tự amino acid với trình tự tham chiếu. Phân tích cấu trúc không gian ba chiều của protein được thực hiện qua phần mềm Protein Homology/analogy Recognition Engine V 2.0. Prankweb cũng được dùng để xác định trung tâm hoạt động dựa vào cấu trúc 3D của protein GlpF quan tâm. Xác định cấu trúc xoắn xuyên màng dự đoán sử dụng công cụ TMHMM 2.0. Bên cạnh đó, điểm đẳng điện (pI) và khối lượng phân tử (Mw) được tính toán bằng ExPASy pI Calculator.

KẾT QUẢ VÀ THẢO LUẬN

Kết quả PCR khuếch đại gen glpF và định lượng sản phẩm tinh sạch

Chúng tôi nghiên cứu tiến hành nhân bản đoạn gen mã hóa GIpF sử dụng cặp mồi được thiết kế đặc hiệu với khuôn là DNA tổng số của *E. coli* DH5 alpha (Hình 1A, giếng 1). Kết quả điện di cho thấy, band đích sáng rõ, kích thước 846 bp phù hợp với tính toán ban đầu (Hình 1B, giếng 1), không có band phụ, đối chứng âm không xuất hiện band. Điều này chứng tỏ cặp mồi do nhóm nghiên cứu thiết kế có độ đặc hiệu cao. Sau khi tinh chế, kết quả đo nanodrop chứng tỏ sản phẩm tinh sạch gen *glpF* có độ sạch tốt và nồng độ tinh sạch đạt yêu cầu (Bảng 1, STT 1). Mặt khác, trên gel agarose TAE 1% xuất hiện band đích sáng rõ, kích thước theo tính toán khoảng 846 bp (Image J, Hoa Kỳ) phù hợp với giả thiết, không có band phụ (Hình 1B, giếng 2). Từ đó kết luận, thiết kế thành công cặp mồi khuếch đại gen *glpF* từ genome *E. coli* DH5 alpha và đoạn gen đích sau khi tinh chế hoàn toàn phù hợp làm nguyên liệu cho các thí nghiệm tiếp theo.



Hình 1. Kết quả điện di

Giếng M: Thang chuẩn DNA (A) Sản phẩm tách DNA tổng số, giếng 1: DNA tổng số của E. coli chủng DH5 alpha; (B) Sản phẩm khuếch đại và tinh sạch gen glpF, giếng (-): đối chứng âm, giếng 1: sản phẩm PCR glpF từ hệ gen của E. coli chủng DH5-α sử dụng cặp mồi glpF- fw/Afl II- rv, giếng 2: sản phẩm tinh sạch gen glpF; (C) Sản phẩm cắt pJET-glpF bằng enzyme cắt giới hạn, giếng 1,4: Không cắt, giếng 2,5: cắt với enzyme BgIII, giếng 3,6: cắt với enzyme AfIII.

Kết quả nhân dòng gen mã hóa GIpF bằng vector pJET 1.2/ blunt

Sản phẩm PCR có dạng đầu bằng nên gen *glpF* nên sau tinh chế có thể được nối trực tiếp vào vector nhân dòng pJET 1.2/ blunt theo tỉ lệ đã được mô tả trong phần phương pháp. Sau đó, vector tái tổ hợp được biến nạp vào tế bào khả biến *E. coli* chủng DH5 alpha. Tại vị trí đa điểm trên vector nhân dòng có chứa gen gây độc *Eco47IR* bị gián đoạn bởi đoạn chèn, trình khởi động Plac được sửa đổi cho phép biểu hiện gen *Eco47IR* ở mức độ vừa
phải, đủ để chọn lọc plasmid. Vì vậy, chỉ những vi khuẩn mang DNA tái tổ hợp mới có thể mọc thành khuẩn lạc. Do đó, hoàn toàn có thể bỏ qua bước sàng lọc xanh/ trắng tốn kém. Kết quả biến nạp cho thấy, khuẩn lạc trên đĩa LB - ampicillin xuất hiện tròn, rõ, rời rạc và không có khuẩn lạc vệ tinh, phù hợp để thực hiện sàng lọc bằng PCR khuẩn lạc (Hình 2A). Tuy nhiên, *glpF* có sẵn trong hệ gen của vi khuẩn *E. coli* DH5 alpha nên không thể sử dụng cặp mồi do chúng tôi thiết kế để kiểm tra khuẩn lạc. Mặt khác, cặp mồi pJET fw/rv bám vào hai đầu vector nhân dòng và bị gián đoạn bởi vùng MCS nên thích hợp cho mục đích kiểm tra đoạn chèn. Kết quả sàng lọc khuẩn lạc bằng phương pháp PCR cho thấy cả 10/10 khuẩn lạc được chọn đều xuất hiện band đích (972bp) (Hình 2B, giếng 1-10). Sự chênh lệch kích thước đến từ việc chúng tôi đã nghiên cứu sử dụng cặp mồi của vector nhân dòng, phù hợp với giả thiết ban đầu.

Kết quả kiểm tra plasmid tái tổ hợp bằng enzyme cắt giới hạn

Trước hết, 2/10 khuẩn lạc đã được chọn lọc từ phản ứng PCR được tinh sạch plasmid theo quy trình đã mô tả trong phần phương pháp. Kết quả kiểm tra hấp thụ ánh sáng tử ngoại bước sóng A₂₆₀ lần lượt là 210 ng/ μL và 160 ng/ μL, hoàn toàn phù hợp cho các thí nghiệm tiếp theo. Hai enzyme giới hạn được sử dụng để kiểm tra plasmid tái tổ hợp là *Bg*/II và *Af*/II với mục đích kiểm tra kích thước plasmid, vị trí và kích thước đoạn chèn. Kết quả điện di đồ sản phẩm cắt với enzyme giới hạn đã được thể hiện (Hình 1C), kích thước band đích tiếp tục được tính toán sử dụng công cụ Image J. Trong đó, enzyme cắt mở vòng *Af*/II chứa 1 điểm cắt trên vector pJET 1.2/ blunt tại vị trí 750/751 bp nên khi cắt sẽ tạo ra 1 band đích kích thước khoảng 3820 bp (Hình 1C, giếng 3,6). Mặt khác, vector nhân dòng mang 2 vị trí nhận biết enzyme *Bg*/II tại vị trí (337/338 bp và 383/384 bp) kế bên MCS nên được sử dụng để kiểm tra kích thước đoạn chèn. Do đó, khi cắt với *Bg*/II sẽ cho 2 band đích có kích thước lần lượt khoảng 2974 bp (vector) và 846 bp (đoạn chèn) (Hình 1C, giếng 2,5). Bên cạnh đó, kết quả điện di vector tái tổ hợp không cắt với enzyme xuất hiện band kích thước lớn, có thể do giải thích do plasmid tồn tại ở trạng thái siêu xoắn, một dạng lý tưởng cho mục đích giải trình tự (Hình 1C, giếng 1,4). Bước đầu kết luận gen *glpF* đã được chèn chính xác vào vị trí MCS và được biến nạp thành công vào tế bào khả biến.





Hình 2. Kết quả PCR khuẩn lạc sàng lọc thể biến nạp M: Thang chuẩn DNA, 1-10: khuẩn lạc 1-10.

Hình 3. Trung tâm hoạt động của GlpF (Vùng màu đỏ thể hiện trung tâm hoạt động).

Kết quả giải trình tự gen glpF và phân tích sử dụng các công cụ tin sinh

Chúng tôi tiếp tục nghiên cứu giải trình tự nucleotide để khẳng định vector tái tổ hợp pJET-*glpF* đã mang chính xác đoạn gen quan tâm. Kết quả so sánh trình tự bằng công cụ tin sinh CLUSTAL OMEGA cho thấy gen *glpF* có kích thước 846bp và tương đồng 100% với trình tự gen *glpF* đã công bố trên ngân hàng gen (GenBank: AAA23886.1). Các nghiên cứu trước đó thường sử dụng chủng gốc *E. coli* K12 MG1655, *Corynebacterium glutamicum, Pseudomonas aeruginosa* hoặc các chủng vi khuẩn khác phục vụ cho mục đích tách nguồn gen. Mặc dù *E. coli* DH5 alpha là một dạng dẫn xuất của chủng gốc *E. coli* K12 MG1655 và gen *glpF* có tính bảo tồn cao. Theo dự đoán, gen *glpF* có độ tương đồng khoảng 99% về trình tự nucleotide giữa 2 chủng vi khuẩn này, sự sai khác có thể do các biến thể nhỏ hoặc các đột biến điểm trên gen. Kết quả nghiên cứu này của chúng tôi cho thấy gen *glpF* từ *E. coli* DH5 alpha được kiểm tra trình tự trong nghiên cứu tương đồng tuyệt đối với gen này ở chủng *E. coli* K12 MG1655, chứng tỏ tính bảo tồn tiến hóa cao của gen.

sp GlpF E.coli.K12	MSQTSTLKGQCIAEFLGTGLLIFFGVGCVAALKVAGASFGQWEISVIWGLGVAMAIYLTA	60
sp GlpF E.coli	MSQTSTLKGQCIAEFLGTGLLIFFGVGCVAALKVAGASFGQWEISVIWGLGVAMAIYLTA	60

sp GlpF E.coli.K12	GVSGAHLNPAVTIALWLFACFDKRKVIPFIVSQVAGAFCAAALVYGLYYNLFFDFEQTHH	120
sp GlpF E.coli.	GVSGAHLNPAVTIALWLFACFDKRKVIPFIVSQVAGAFCAAALVYGLYYNLFFDFEQTHH	120

sp GlpF E.coli.K12	IVRGSVESVDLAGTFSTYPNPHINFVQAFAVEMVITAILMGLILALTDDGNGVPRGPLAP	180
sp GlpF E.coli.	IVRGSVESVDLAGTFSTYPNPHINFVQAFAVEMVITAILMGLILALTDDGNGVPRGPLAP	180

sp GlpF E.coli.K12	LLIGLLIAVIGASMGPLTGFAMNPARDFGPKVFAWLAGWGNVAFTGGRDIPYFLVPLFGP	240
sp GlpF E.coli.	LLIGLLIAVIGASMGPLTGFAMNPARDFGPKVFAWLAGWGNVAFTGGRDIPYFLVPLFGP	240

sp GlpF E.coli.K12	IVGAIVGAFAYRKLIGRHLPCDICVVEEKETTTPSEQKASL 281	
sp GlpF E.coli.	IVGAIVGAFAYRKLIGRHLPCDICVVEEKETTTPSEQKASL 281	



Hình 4. Kết quả so sánh trình tự amino acid và xây dựng cấu trúc không gian 3 chiều dự đoán GlpF từ *E. coli* DH5 alpha (A) Kết quả so sánh trình tự amino acid (Dấu sao thể hiện trình tự tương đồng) (B) Mô trình cấu trúc không gian 3 chiều GlpF trên E. coli chủng DH5 alpha (Cấu trúc xoắn α được thể hiện bằng hình lát mỏng dạng cuộn).

Từ trình tự nucleotide, chúng tôi sử dụng phần mềm SnapGene để đưa ra trình tự amino acid suy diễn và tiếp tục dùng CLUSTAL OMEGA cho mục đích so sánh trình tự. Kết quả cho thấy, chiều dài protein dự đoán là 281 amino acid, tương đồng 100% với trình tự tham chiếu trên RCSB PDB (Mã: 1FX8) (Hình 4A). Điểm đẳng điện (pl) theo tính toán bằng công cụ ExPASy pl Calculator là 6,18 và khối lượng phân tử là 29,78 kDa, phù hợp với nghiên cứu đã được công bố (Weissenborn et al., 1992). Bên cạnh đó, GlpF hoạt động tốt ở môi trường từ trung tính đến kiềm thấp, pH được xác định trong khoảng 6,5 đến 7,5 và hoạt động ở nhiệt độ phát triển điển hình của vi khuẩn (37°C).

GlpF là protein màng có cấu trúc quyết định chức năng, vì vậy mô phỏng cấu trúc 3D và dự đoán chuỗi xuyên màng là cần thiết. Chúng tôi sử dụng phần mềm Protein Homology/analogy Recognition Engine V 2.0 để dự đoán cấu trúc không gian 3 chiều đặc trưng của protein GlpF (Hình 4B). Kết quả cho thấy, protein quan tâm có cấu trúc tương đồng với protein hỗ trợ vận chuyển glycerol ở *E. coli* (GlpF) với độ tin cậy 100%. Cụ thể, cấu trúc không gian 3 chiều dự đoán của protein GlpF đặc trưng với cấu trúc 6 chuỗi xoắn alpha (lần lượt từ TM1 đến TM6), đầu N và đầu C nằm trong tế bào chất. Các chuỗi xoắn này tạo thành kênh hẹp chọn lọc ở trung tâm protein. Kênh này được lót bằng các gốc axit amin phân cực và tích điện, tạo điều kiện tương tác với glycerol. Kết quả phân tích bằng TMHMM 2.0 cho thấy: 6 chuỗi xoắn alpha (TM1 đến TM6) xuyên màng dự đoán (Hình 5), phù hợp với mô tả về cấu trúc đặc trưng của protein quan tâm. Thêm vào đó, GlpF là một protein thuộc họ aquaporin nên thường không có peptide tín hiệu. GlpF được tổng hợp với các miền xuyên màng đã có sẵn, vì vậy không cần peptide tín hiệu để tích hợp vào màng. Thực tế, chưa có nghiên cứu nào trước đây báo cáo về peptide tín hiệu của GlpF vì chức năng định hướng vào màng được coi là một quá trình sinh tổng hợp thông thường. Thay vào đó, cấu trúc miền xuyên màng của protein này là đặc điểm quyết định chức năng và thường được quan tâm nghiên cứu.



Hình 5. Kết quả dự đoán cấu trúc xuyên màng của GIpF (màu tím thể hiện trạng thái xuyên màng dự đoán).

Khi sử dụng Prankweb, 40 amino acid được xác định liên quan tới trung tâm hoạt động của protein GlpF dựa vào cấu trúc không gian ba chiều. Không giống như các enzyme có vị trí "active site" cụ thể, trung tâm hoạt động của GlpF nằm ở kênh trung tâm, giúp vận chuyển và tương tác với glycerol (Hình 3). Các gốc amino acid chính được xác định có vai trò quan trọng trong việc quyết định tính đặc hiệu và tạo điều kiện thuận lợi cho quá trình vận chuyển glycerol bao gồm: Arginine (ARG 206) - tương tác với các nhóm hydroxyl của glycerol; Asparagine (ASN 68, ASN 203) - hình thành liên kết với glycerol, loại trừ phân tử lớn hơn; Serine (SER 63, SER 136) - duy trì cấu hình của kênh. Có thể nói, chức năng của GlpF nằm ở việc tăng quá trình vận chuyển glycerol qua màng tế bào thay vì xúc tác phản ứng sinh hóa với một vị trí trung tâm hoạt động như với các loại enzyme truyền thống.

KÉT LUẬN

Nghiên cứu đã khuếch đại và nhân dòng thành công gen *glpF* trên đối tượng *E. coli* DH5 alpha. Gen *glpF* có chiều dài 846 bp mã hóa cho protein xuyên màng (GlpF) gồm 281 amino acid. Kết quả phân tích trình tự nucleotide và trình tự amino acid có độ tương đồng là 100% so với trình tự tham chiếu đã được công bố trước đó trên ngân hàng gen. Cấu trúc không gian ba chiều của protein dự đoán và các chuỗi xoắn xuyên màng đã được mô phỏng hoàn toàn phù hợp với cấu trúc đặc trưng của GlpF. Các gốc amino acid chịu trách nhiệm chính cũng đã được đề cập. Đây là những kết quả bước đầu trong việc tạo vật liệu di truyền cho các nghiên cứu biểu hiện và đánh giá vai trò của protein GlpF trong chuyển hóa glycerol ở *E. coli* cải biến di truyền.

TÀI LIỆU THAM KHẢO

Chiang CJ, Ho YJ, Hu MC, Chao YP (2020). Rewiring of glycerol metabolism in *Escherichia coli* for effective production of recombinant proteins. *Biotechnol Biofuels*, *13*, 205.

Chilakamarry CR, Sakinah A MM, Zularisam AW, Pandey A (2021). Glycerol waste to value added products and its potential applications. *Syst Microbiol Biomanufacturing*, *1*(4): 378-396.

Gupta M, Kumar N (2012). Scope and opportunities of using glycerol as an energy source. *Renew Sustain Energy Rev*, *16*(7): 4551-4556.

Heller KB, Lin EC, Wilson TH (1980). Substrate specificity and transport properties of the glycerol facilitator of *Escherichia coli*. J Bacteriol, 144(1): 274-278.

Martínez-Gómez K, Flores N, Castañeda HM, Martínez-Batallar G, Hernández-Chávez G, Ramírez OT, Gosset G, Encarnación S, Bolivar F (2012). New insights into *Escherichia coli* metabolism: carbon scavenging, acetate metabolism and carbon recycling responses during growth on glycerol. *Microb Cell Fact*, *11*(1): 46.

Stroud RM, Nollert P, Miercke L (2003). The glycerol facilitator GlpF its aquaporin family of channels, and their selectivity. Adv Protein Chem, 63: 291-316.

Sweet G, Gandor C, Voegele R, Wittekindt N, Beuerle J, Truniger V, Lin EC, Boos W (1990). Glycerol facilitator of *Escherichia coli*: Cloning of *glpF* and identification of the *glpF* product. *J Bacteriol*, *172*(1): 424-430.

Voegele RT, Sweet GD, Boos W (1993). Glycerol kinase of *Escherichia coli* is activated by interaction with the glycerol facilitator. *J Bacteriol*, 175(4): 1087-1094.

Weissenborn DL, Wittekindt N, Larson TJ (1992). Structure and regulation of the *glpFK* operon encoding glycerol diffusion facilitator and glycerol kinase of *Escherichia coli* K-12. *J Bio Chem* 267(9): 6122-6131.

MOLECULAR CLONING AND SEQUENCE ANALYSIS OF glpF GENE ENCODING THE GLYCEROL FACILITATOR IN *Escherichia coli*

Pham Thi Kieu Duyen¹, Nguyen Thi An Hoa¹, Nguyen Ha Mi¹, Quyen My Linh¹, Nguyen Quang Huy^{1,2}, Kieu Cam Nhung³, Luan Luong Chu^{2*}

¹Faculty of Biology, VNU University of Science, Vietnam National University, Hanoi

²National Key Laboratory of Enzyme and Protein Technology, VNU University of Science, Vietnam National University, Hanoi

³*Ministry of Science and Technology*

SUMMARY

Biodiesel, a renewable energy source, promises to replace conventional fossil fuels for its environmental benefits. The global expansion of the biodiesel industry has led to an excessive redundancy of crude glycerol - a major byproduct of the production process. This issue has imbalanced commercial glycerol's price and posed a significant challenge to the sustainability of the biodiesel industry. Therefore, the conversion of crude glycerol into high-value products has become an urgent issue in many countries. Escherichia coli (E. coli) has been demonstrated to utilize glycerol as a carbon source by the combined action of glycerol kinase (GlpK) and glycerol 3-phosphate dehydrogenase (GlpD). However, the expression of aquaglyceroporin gene (glpF) plays a crucial role in enhancing growth and biomass formation of bacteria. GlpF exhibits a conservative structure, high selectivity, and is known as the sole protein transport catalyzing glycerol diffusion across the E. coli inner membrane. In this study, the glpF gene from E. coli DH5 α was successfully cloned in the pJET 1.2/blunt vector and sequenced by the Sanger method. The decoded sequence of target gene consisted of 846 nucleotides and was 100% homologous to the reference gene sequence. Furthermore, the comparison of the deduced amino acid sequence and the analysis of the protein's three-dimensional structure also revealed 100% similarity with absolute confidence. These findings of the study would provide scientific foundation, generate genetic resources for developing recombinant microorganisms which enhance the expression of the glpF gene encoding a transmembrane protein that facilitates glycerol transport.

Keywords: Escherichia coli, glycerol, glpF, molecular cloning, sequence analysis.

Author for correspondence: Tel: 0985464929; Email: luancl@vnu.edu.vn

ĐÁNH GIÁ HIỆU QUẢ SỬ DỤNG CỦA BỘ KIT TẠO DÒNG DỰA TRÊN TRÌNH TỰ TƯƠNG ĐỒNG eClone

Nguyễn Thị Mỹ Trinh^{*}, Nguyễn Văn Hậu

Phòng Thí nghiệm Công nghệ sinh học phân tử, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

TÓM TẮT

Hiện nay, tạo dòng DNA trở thành một phương pháp cơ bản trong các nghiên cứu thuộc lĩnh vực sinh học và công nghệ sinh học, và do đó, nhiều phương pháp tạo dòng đã được phát triển nhằm tăng hiệu suất tạo dòng. Cũng trong xu thế này, phòng thí nghiệm CNSH Phân tử, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh đã phát triển thành công bộ kit tạo dòng eClone dựa trên sự tái tổ hợp tương đồng giữa các đoạn trình tự tương đồng trên vector và DNA. Trong nghiên này, chúng tôi tiến hành đánh giá hiệu quả của bộ kit trong việc nối ba đoạn gene có độ dài 300 bp, 1000 bp, 1500 bp vào vector pET-28a(+) nhờ sự có mặt của các trình tự tương đồng dài 10-20 nu ở hai đầu của gene và vector. Đồng thời, khi so sánh hiệu quả với bộ kit thương mại GeneBuilder (GenScript, Hoa Kỳ), chúng tôi nhận thấy mặc dù số lượng khuẩn lạc thu nhận được từ bộ kit tạo dòng eClone có thấp hơn so với bộ kit thương mại nhưng tỷ lệ khuẩn lạc dương tính không khác biệt đáng kể. Bên cạnh đó, cho đến nay đã có trên 200 gene đã được tạo dòng thành công bằng phương pháp eClone tại nhiều đơn vị kinh doanh và nghiên cứu khác nhau. Các kết quả này cho thấy tiềm năng ứng dụng của bộ kit eClone trong các nghiên cứu liên quan đến tạo dòng phân tử.

Từ khóa: eClone, trình tự tương đồng, tạo dòng DNA, pET-28a(+), hiệu suất tạo dòng.

MỞ ĐẦU

Tạo dòng phân tử hoặc tạo dòng DNA là quá trình chèn một đoạn DNA vào một vector, sau đó biến nạp vào tế bào chủ để khuếch đại đoạn DNA mục tiêu hoặc biểu hiện protein tái tổ hợp. Từ khi thành công vào năm 1972, tạo dòng DNA ngày càng được sử dụng rộng rãi, nhanh chóng trở thành một kĩ thuật cơ bản, quan trọng trong lĩnh vực sinh học và công nghệ sinh học (Cohen, 2013). Do đó, nhiều phương pháp tạo dòng đã được các phòng thí nghiệm phát triển để nâng cao hiệu quả và giảm chi phí.

Phương pháp tạo dòng truyền thống sử dụng enzyme cắt giới hạn (Restriction enzyme - RE) và DNA ligase, bao gồm các bước chính: 1, sử dụng các RE để cắt đoạn DNA mục tiêu và vector; 2, khuếch đại DNA mục tiêu và chèn vào vector đã cắt mở vòng bằng cách sử dụng enzyme DNA ligase; 3, Biến nạp hỗn hợp nối vào tế bào chủ; và 4, sàng lọc dòng mục tiêu (Lorsch, 2013). Mặc dù phương pháp này đơn giản, linh hoạt, chi phí thấp, nhưng có nhiều hạn chế như một số trường hợp không thể tìm được RE phù hợp do đoạn gene mục tiêu chứa nhiều vị trí nhận biết của RE; điều kiện phản ứng của các RE khác nhau nên làm giảm hiệu suất tạo dòng nếu chọn chiến lược cắt vector bằng 2 enzyme và thu nhận nhiều khuẩn lạc không chứa gene do vector tự nối.

Để cải thiện hiệu quả tạo dòng DNA, các nhà khoa học sử dụng các bộ kit tạo dòng đã được phát triển và hiện có trên thị trường như Cloning Independent Cloning (New England Biolabs, Hoa Kỳ), In-fusion kit (Takara, Nhật Bản), Cold-fusion kit (System Biosciences, Hoa Kỳ), hoặc Genbuilder™ DNA Assembly (Genscript Biotech, Hoa Kỳ) thay vì sử dụng ligase. Các kit tạo dòng này này hoạt động dựa trên sự hiện diện của vùng tương đồng ở mỗi đầu của đoạn DNA mục tiêu và vector đã cắt mở vòng dưới tác động của các enzyme đặc hiệu (Valla *et al.*, 2014). Ưu điểm của phương pháp này so với các phương pháp truyền thống là đạt hiệu quả tạo dòng cao, không cần cắt trình tự mục tiêu với RE và tiết kiệm thời gian. Tuy nhiên, do sử dụng các enzyme tinh sạch, các bộ kit này hiện được bán với giá cao nên vẫn chưa được sử dụng rộng rãi trong các phòng thí nghiệm tại Việt Nam.

Với mục đích thúc đẩy nghiên cứu Sinh học và Công nghệ sinh học trong nước, Phòng Thí nghiệm Công nghệ sinh học phân tử, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh đã nghiên cứu và phát triển bộ kit tạo dòng eClone dựa trên quá trình tái tổ hợp tương đồng, giúp việc tạo dòng gene trở nên đơn giản, hiệu quả và tiết kiệm chi phí. Trên cơ sở đó, nghiên cứu này được thực hiện với mục đích đánh giá hiệu quả tạo dòng của bộ kit eClone.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Bộ kit tạo dòng eClone được sản xuất và cung cấp bởi Phòng thí nghiệm Công nghệ Sinh học phân tử, Trường Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh. Bộ kit gồm 2 thành phần: enzyme

eClone và buffer 10X. Plasmid pET-28a(+) (Merck Biosciences, Hoa Kỳ) được thu nhận từ tế bào *E. coli* DH5α/pET-28a(+). Tế bào *E. coli* DH5α (Thermofisher Scientific, Hoa Kỳ) được sử dụng cho mục đích tạo dòng. Thang DNA 1 kb được cung cấp bởi hãng Bioline (Canada).

Chiến lược tạo dòng

Bộ kit eClone cho phép nối một đoạn gene bất kỳ vào plasmid theo cơ chế tái tổ hợp tương đồng theo quy trình được trình bày ở Hình 1A. Trong đó, plasmid sẽ được cắt mở vòng bằng 1 hoặc 2 enzyme cắt giới hạn. Trong nghiên cứu này, plasmid pET-28a(+) được sử dụng và được cắt mở vòng bằng enzyme *Bam*HI. Song song đó, gene mục tiêu được khuếch đại bằng phương pháp PCR với cặp mồi đặc hiệu, trong đó mỗi mồi sẽ chứa 3 trình tự đặc trưng theo thứ tự từ đầu 5' đến đầu 3' như sau: 1) 10-20 nucleotide (nu) giống với trình tự ở đầu tương ứng trên plasmid đã cắt mở vòng; 2) trình tự của RE được dùng để cắt mở vòng plasmid; và 3) 18-20 nu đặc hiệu với gene để khuếch đại gene mục tiêu (Hình 1B). Trong nghiên cứu này, 3 gene có độ dài khác nhau A (~300 bp), B (~1000 bp) và C (~1500 bp) được sử dụng để đánh giá hiệu quả tạo dòng của bộ kit eClone.

Sau đó, gene mục tiêu được nối vào plasmid thông qua quá trình tái tổ hợp giữa các trình tự tương đồng ở 2 đầu của gene và plasmid dưới sự xúc tác của enzyme eClone. Sản phẩm nối được biến nạp vào tế bào vi khuẩn *E. coli* khả nạp. Cuối cùng, hỗn hợp biến nạp được trãi trên đĩa chứa môi trường LB có bổ sung nhân tố chọn lọc (Kanamycin). Các khuẩn lạc phát triển trên môi trường được kiểm tra bằng phương pháp PCR khuẩn lạc và xác nhận lại bằng phương pháp PCR plasmid.



(A)



Chuẩn bị gene mục tiêu và plasmid

Ba gene A, B, C được khuếch đại từ mạch khuôn được đặt tổng hợp bằng phương pháp PCR với các cặp mồi tương ứng được thiết kế theo nguyên tắc đã trình bày ở trên dưới sự xúc tác của Taq DNA polymerase (Bioline,

Canada). Để làm rõ hơn về cách thiết kế mồi, trình tự cặp mồi dùng để khuếch đại gene A mang 15 nu tương đồng với plasmid được lấy làm ví dụ và thể hiện ở Hình 1C, gồm mồi F: 5'-ACGGAGCTCGAATTCgg atcc**AAATATTATGGTAATGGTG-3'** và mồi R: 5'-ACAGCAAATGGTCGC ggatcc**TTTGCTAACATGAACGC**-3' (ký tự chữ hoa: trình tự giống với plasmid, ký tự chữ thường: vị trí nhận diện của BạmHI, và ký tự chữ hoa đậm: trình tự đặc hiệu với gene). Sản phẩm PCR sau đó được tinh sạch bằng phương pháp phenol/choloroform. Nồng độ DNA sau đó được xác định dựa trên độ hấp thụ ở bước sóng 260 nm bằng máy Nanodrop (Thermo Fisher, Hoa Kỳ).

Để cắt mờ vòng plasmid, phản ứng gồm Buffer Tango 1X, 0,3 U/µL *Bam*HI, và 4000 ng/µL plasmid pET-28a(+) được chuẩn bị và ủ ở 37°C trong 16 giờ. Sau khi hoàn thành, phản ứng được ủ ở 65°C trong 10 phút để bất hoạt enzyme *Bam*HI. Cuối cùng, sản phẩm cắt được tinh sạch bằng bộ kit EZ-10 Spin Column DNA PCR purification (Bio Basic, Canada).

Thực hiện phản ứng nối gene vào plasmid bằng bộ kit eClone

Thiết lập 10 µL phản ứng nối gồm 37,5 ng/µL plasmid pET-28a(+) đã cắt mở vòng, gene mục tiêu, 1X eClone buffer, 1 µL enzyme eClone. Trong đó, lượng gene mục tiêu được bổ sung để đạt tỷ lệ gene:plasmid là 3:1. Phản ứng nối được tiến hành ở 37°C trong 15 phút.

Chuẩn bị tế bào khả nạp

Tế bào *E. coli* DH5α khả nạp được chuẩn bị theo quy trình đã được mô tả trước đó (Tu, 2008). Tính khả nạp của tế bào được đánh giá bằng cách biến nạp 1 μg plasmid pET-28a(+) vào 100 μL dịch tế bào khả nạp. Sau đó, trải tế bào trên đĩa Luria Bertani có bổ sung 30 μg/mL Kanamycin (LB-Kan) và đếm tổng số khuẩn lạc xuất hiện trên đĩa. Hiệu quả biến nạp của tế bào được ghi nhận bằng số khuẩn lạc xuất hiện trên 1 μg plasmid (cfu/μg).

Biến nạp và đánh giá hiệu quả tạo dòng của bộ kit eClone

Bổ sung toàn bộ 10 μL phản ứng nối vào 1 eppendorf chứa 100 μL tế bào khả nạp theo quy trình được mô tả bởi Tu (2008). Hỗn hợp biến nạp được trãi trên đĩa môi trường LB-Kan, sau đó đĩa được ủ ở 37°C trong 16-20 giờ. Số lượng khuẩn lạc xuất hiện trên đĩa được đếm và ghi nhận lại.

Sau đó, 10 khuẩn lạc được chọn ngẫu nhiên để kiểm tra sự hiện hiện của plasmid tái tổ hợp pET-28a(+)-A/B/C bằng phương pháp PCR khuẩn lạc với các cặp mồi đặc hiệu đã được sử dụng để thu nhận gene. Các khuẩn lạc cho kết quả dương tính sẽ được nuôi cấy trong môi trường lỏng LB-Kan và plasmid được thu nhận bằng phương pháp SDS-Kiềm. Sự hiện diện của gene mục tiêu trong plasmid được kiểm tra lại bằng phương pháp PCR plasmid với cặp mồi gồm mồi F trên gene (mồi dùng để thu gene) và mồi R trên T7 terminator của plasmid.

Tỷ lệ khuẩn lạc dương tính được tính theo công thức: Tỷ lệ khuẩn lạc dương tính (%)= (b/a) x (d/c), trong đó: a là số lượng khuẩn lạc được chọn để tiến hành PCR khuẩn lạc, b là số lượng khuẩn lạc cho kết quả dương tính với PCR khuẩn lạc, c là số lượng plasmid được kiểm tra bằng phương pháp PCR plasmid, d là số lượng plasmid cho kết quả dương tính với PCR plasmid

So sánh hiệu quả tạo dòng của bộ kit eClone và GenBuilder

Gene A được khuếch đại bằng phản ứng PCR với cặp mồi đặc hiệu chứa 15 nucleotide tương đồng với plasmid. Quá trình tạo dòng được thực hiện như quy trình được đề cập ở trên, trong đó phản ứng nối gene vào plasmid bằng eClone được thực hiện ở 37°C trong 15 phút trong khi phản ứng nối bằng GenBuilder được thực hiện ở 50°C trong 15 phút. Sự khác nhau giữa số khuẩn lạc thu nhận được và tỷ lệ dương tính của bộ kit eClone và GenBuilder được phân tích thống kê bằng phần mềm Graphpad prism 7.0 sử dụng phép kiểm định unpaired T-test.

KẾT QUẢ VÀ THẢO LUẬN

Thu nhận gene và plasmid

Ba gene A (300bp), B (1000bp) và C (1500bp) được khuếch đại với mồi đặc hiệu chứa vùng trình tự tương đồng lần lượt là 10 nu, 15 nu và 20 nu. Kết quả điện di trên gel agarose 1,2% (Hình 2A) cho thấy một vạch ở mỗi giếng có kích thước phù hợp với kích thước lý thuyết của gene và không xuất hiện vạch phụ, cho thấy chúng tôi đã khuếch đại thành công các gene mục tiêu với các mồi được thiết kế và phản ứng PCR đặc hiệu cho các gene.

Song song đó, chúng tôi cũng tiến hành thu nhận và cắt mở vòng plasmid pET-28a(+) bằng enzyme cắt giới hạn *Bam*HI. Kết quả điện di trên gel agarose cho thấy plasmid sau khi cắt mở vòng cho một vạch duy nhất và sắc nét nằm trong khoảng giữa hai vạch 5000 bp và 6000 bp của thang DNA, tương ứng với kích thước lý thuyết là 5369 bp (Hình 2B). Như vậy, bước đầu chúng tôi khẳng định được đã thành công trong việc cắt mở vòng plasmid. Tuy nhiên, do hạn chế về giới hạn phát hiện DNA của phương pháp điện di trên gel agarose và nhuộm DNA bằng ethidium bromide, chúng tôi tiến hành bước kiểm tra sâu hơn để đánh giá hiệu quả của phản ứng cắt bằng cách biến nạp sản phẩm cắt vào tế bào *E. coli* DH5α. Kết quả biến nạp cho thấy số khuẩn lạc xuất trên đĩa LB-Kan rất ít, chỉ 0-1 khuẩn lạc ở mỗi lần lặp lại. Như vậy, chúng tôi khẳng định được phản ứng cắt mở vòng plasmid thành công với hiệu suất cao.



Hình 2. Kết quả thu nhận gene A, B, C (A) và cắt mở vòng plasmid pET-28a(+) bằng enzyme BamHI (B)

Kiểm tra hiệu quả biến nạp của tế bào khả nạp

Do hiệu quả tạo dòng phụ thuộc nhiều vào tính chất của tế bào khả nạp nên chúng tôi tiến hành đánh giá sức sống cũng như tính khả nạp của tế bào. Sức sống của tế bào được đánh giá bằng cách nuôi cấy trên đĩa môi trường LB không chứa kháng sinh và LB-Kan. Tính khả nạp của tế bào được đánh giá thông qua số lượng thể biến nạp xuất hiện trên môi trường LB - Kan sau khi được biến nạp plasmid pET-28a(+). Kết quả cho thấy tế bào khả nạp *E. coli* DH5α có hiệu quả biến nạp **1,6 x 10⁶ cfu/µg.** Hiệu quả biến nạp của tế bào *E. coli* DH5α trong sản phẩm thương mại của một số công ty như Thermo Fisher, Abmgood từ khoảng 10⁶ đến 10⁸ cfu/µg. Như vậy, chúng tôi đã chuẩn bị tế bào khả nạp có hiệu quả biến nạp cao và thích hợp để tạo dòng gene mục tiêu.

Đánh giá hiệu quả tạo dòng của bộ kit eClone đối với các gene có kích thước khác nhau và chiều dài vùng trình tự tương đồng khác nhau

Nghiên cứu Shen và Huang (1986) cho thấy hiệu suất của quá trình tái tổ hợp tương đồng phụ thuộc vào độ dài của trình tự tương đồng. Các nghiên cứu sau này, điển hình là nghiên cứu năm 1995 của Fujitani và đồng tác giả (1995) chỉ ra mối quan hệ tuyến tính giữa chiều dài vùng tương đồng và hiệu suất quá trình tái tổ hợp. Do đó, chúng tôi tiến hành đánh giá hiệu quả tạo dòng của bộ kit eClone đối với các gene có chiều dài khác nhau (300bp, 1000bp và 1500bp) và độ dài đoạn tương đồng khác nhau (10 nu, 15 nu và 20 nu). Hiệu quả tạo dòng được đánh giá dựa trên 2 tiêu chí: số khuẩn lạc xuất hiện trên đĩa biến nạp và tỉ lệ khuẩn lạc dương tính được tính.

Kết quả tạo dòng (Hình 3) cho thấy đối với gene A có kích thước 300bp, việc sử dụng mồi có kích thước vùng tương đồng 15 nucleotide cho số lượng khuẩn lạc xuất hiện trên đĩa LB-Kan cũng như tỷ lệ khuẩn lạc dương tính cao nhất. Nhìn chung, đối với cả 3 độ tương đồng, tỷ lệ khuẩn lạc dương tính mang gene A đều trên 85%. Đối với gene B có kích thước 1000 bp, số lượng khuẩn lạc xuất hiện trên đĩa LB-Kan cũng như tỷ lệ khuẩn lạc dương tính cao nhất ứng với mồi có kích thước vùng tương đồng 10 nu và tỷ lệ khuẩn lạc dương tính đều trên 60% đối với việc sử dụng mồi có chiều dài vùng tương đồng ngắn (10 và 15 nu). Đối với gene C có kích thước 1500 bp, việc sử dụng mồi có kích thước vùng tương đồng 10 nucleotide cho số lượng khuẩn lạc tương đối cao và tỷ lệ khuẩn lạc dương tính cao nhất, nhưng chỉ thu được 25 khuẩn lạc với tỉ lệ dương tính 63,33 ± 6,22%.

Như vậy, trong phạm vi khảo sát ở nghiên cứu này, chúng tôi kết luận được: đối với gene có kích thước 300 bp, độ dài đoạn tương đồng tối ưu là 15 nu, trong khi đối với gene có kích thước cao hơn (1000 nu và 1500 nu), kích thước vùng tương đồng tối ưu là 10 nu. Về kích thước gene, gene có kích thước nhỏ cho hiệu suất (số khuẩn lạc và tỷ lệ dương tính) cao hơn gene có kích thước lớn. Nhưng nhìn chung, với việc sử dụng bộ kit eClone, chúng tôi đã thành công trong việc tạo dòng gene có kích thước lên đến 1500 bp với độ dài vùng trình tự tương đồng là 10-20 nu. Bên cạnh đó, 3 plasmid mang gene A, B, C được lựa chọn ngẫu nhiên để kiểm tra độ chính xác của trình tự được chèn bằng phương pháp giải trình tự Sanger, kết quả cho thấy cả 3 đoạn gene A, B, C được chèn vào các plasmid này có độ chính xác 100% so với trình tự lý thuyết.

Nghiên cứu của Jacobus và đồng tác giả (2015) cho thấy tỉ lệ khuẩn lạc dương tính tăng khi tăng chiều dài vùng tương đồng và thích hợp nhất với đoạn 20 bp. Nghiên cứu của Jacobus sử dụng gene có kích thước 500bp và vector pUC19 (2314 bp), nồng độ plasmid từ 100 ng - 250 ng với tỉ lệ 2:1. Đối chiếu với những khảo sát trong đề tài này, cho thấy sự khác nhau trên hai kết quả thu được là khó tránh khỏi khi sử dụng những điều kiện nghiên cứu khác nhau. Một mặt, nghiên cứu của tác giả trên sử dụng tế bào khả nạp có hiệu suất khả nạp và số lượng khuẩn lạc sàng lọc cao hơn. Mặt khác, kích thước plasmid khi biến nạp vào tế bào chủ *E. coli* bằng phương pháp sốc nhiệt có thể ảnh hưởng đến hiệu suất biến nạp (Hanahan, 1983). Đề tài của chúng tôi khảo sát trên các kích thước gene khác nhau cho góc nhìn rộng hơn khi so sánh các kích thước của vùng tương đồng. Vì vậy, với kết quả tạo dòng trong đề tài, chúng tôi xác nhận được bộ kit tạo dòng eClone để tạo dòng với kích thước vùng tương đồng là 10 nu (hoặc 15 nu trong trường hợp đoạn gene có kích thước ngắn). Kết quả của chúng tôi cũng có khác biệt so với kết quả nghiên cứu của Fujitani (1995), tác giả kết luận rằng hiệu suất tái tổ hợp tuyến tính với chiều dài đoạn tương đồng. Mặc dù nguyên nhân khác biệt vẫn chưa rõ, chúng tôi giả thiết rằng khi DNA mạch đơn được tạo thành ở bước đầu của quá trình tái tổ hợp, việc kéo dài đoạn tương đồng trong nghiên cứu của sú trúc thứ cấp ở đoạn DNA mạch đơn này dẫn tới hiện tượng giảm hiệu suất tạo dòng.



CÔNG NGHỆ GEN

Hình 3. Ảnh hưởng của chiều dài gene và chiều dài vùng tương đồng đến hiệu quả tạo dòng

So sánh hiệu quả tạo dòng giữa bộ kit đang phát triển và bộ kit GenBuilder

Kết quả so sánh hiệu quả tạo dòng bằng bộ kit eClone và bộ kit tái tổ hợp tương đồng thương mại GenBuilder (Genscript) cho thấy số khuẩn lạc xuất hiện trên môi trường chọn lọc khi sử dụng bộ kit thương mại GenBuilder nhiều hơn so với bộ kit eClone, tuy nhiên tỷ lệ dương tính của cả 2 phương pháp đều không có sự khác biệt có ý nghĩa thống kê (Bảng 1).

Bảng 1. So sánh h	iệu quả tạo dòng	với bộ kit thương mại
-------------------	------------------	-----------------------

	Số khuẩn lạc	Tỷ lệ dương tính (%)
Bộ kit thương mại GenBuilder	241 ± 15,9	83 ± 5,8
Bộ kit hiện tại	83 ± 9,8****	87 ± 11,5 ^{ns}

Ký hiệu: ns: p>0,5, không khác biệt có ý nghĩa thống kê; ****: p<0.001.

Thống kê số lượng gene đã tạo dòng thành công bằng bộ kit eClone

Bộ kit tạo dòng eClone được chúng tôi sử dụng thường xuyên trong việc tạo dòng DNA, đồng thời gửi đi một số đơn vị nghiên cứu khác để đánh giá hiệu quả tạo dòng của bộ kit này. Cho đến nay, theo kết quả thống kê chúng tôi thu nhận được, đã có 222 dòng được tạo thành công nhờ bộ kit eClone, với kích thước gene từ 162 bp cho đến 5000 bp và kích thước plasmid từ 2686 bp đến khoảng 7500 bp (Bảng 2). Trong đó, tất cả các dòng đã được tạo ra tại Trường ĐH Khoa học Tự nhiên, ĐHQG-HCM đã được giải trình tự, kết quả cho thấy tất cả các đoạn DNA được chèn vào plasmid có trình tự chính xác 100% so với trình tự lý thuyết. Tỷ lệ dương tính có nhiều biến động lớn, từ 8-100%, chủ yếu do nguyên liệu dùng để tạo dòng như plasmid cắt mở vòng, gene mục tiêu và tế bào *E. coli* khả nạp chưa được chuẩn hoá như trong các thí nghiệm trước. Bên cạnh đó, kết quả thử nghiệm tại công ty sinh hoá Phù Sa Biochem cho thấy đã thành công trong việc chia một đoạn gene kích thước 3022 bp thành 3 mảnh nhỏ khác nhau và tiến hành nối đồng thời vào plasmid pUC19 để tạo ra plasmid tái tổ hợp chứa toàn bộ gene này (Bảng 3, STT 18). Kết quả này chứng minh được rằng bộ kit eClone có khả năng nối nhiều đoạn gene vào một plasmid dựa trên trình tự tái tổ hợp tương đồng giữa chúng, một hiệu quả khó có thể có khi sử dụng phương pháp tạo dòng dựa trên DNA ligase truyền thống.

Bang 2. Bang thong ke ket qua tạo dong với nhiều gene và plasmid khác n

STT	Kích thước gene (bp)	Plasmid	Kích thước plasmid (bp)	Số khuẩn lạc	Tỷ lệ dương tính	STT	Kích thước gene (bp)	Plasmid	Kích thước plasmid (bp)	Số khuẩn lạc	Tỷ lệ dương tính
Trường Đại học Khoa học Tự nhiên, ĐHQG TP. Hồ Chí Minh					Công	ty Phù Sa	Biochem				
1	162	pET-22b(+)	5493	19	67%	Kết qu	ả QC				
2	1008	pET-22b(+)	5493	62	40%	16	600	pUC19	2686	74-144	90-100%
3	1008	pYSag	6378	18	60%	17	435	pJET1.2	2974	54-85	100%

HỘI NGHỊ KHOA HỌC TOÀN QUỐC VỀ CÔNG NGHỆ SINH HỌC 2024

4	1032	pET-22b	5493	1	100%
5	528	pQE60	3431	>300	40%
6	420	pPICZαA	3593	117	60%
7	109	pET-28a(+)	5369	4	50%
8	2274	pET-22b(+)	5493	6	17%
9	2274	pYSag	6378	3	100%
10	441	pPICZαA	6378	6	83%
11	900	pYSag	6378	24	100%
12	1371	pYSag	6378	40	100%
13	723	pPICZαA	6378	109	17%
14	267	p427-TEF	7464	17	100%
15	75	p427-TEF	7464	44	86%

Khảo sát nối nhiều đoạn (nối 3 đoạn gene vào 1 vector)								
18	3022	pUC19	2686	300	8%			
Tạo dòng cho mục đích thương mại								
19- 220 Tạo dòng thành công 202 gene có kích thước từ 200- 5000 bp								
Trường Đại học Quốc tế, ĐHQG TP. Hồ Chí Minh								
221	221 300 3000 20-250 60-10							
Trung tâm Công nghệ sinh học Tp.Hồ Chí Minh								
222	714	5335			100%			

KÉT LUẬN

Các kết quả đánh giá về hiệu quả tạo dòng bằng bộ kit eClone của chúng tôi cho thấy chúng tôi đã tạo dòng thành công 3 gene có kích thước khác nhau 300 bp, 1000 bp và 1500 bp vào plasmid pET-28a(+) với hiệu quả khác nhau tuỳ theo kích thước gene và kích thước vùng tương đồng. Khi so sánh hiệu quả tạo dòng giữa bộ kit eClone với bộ kit thương mại GenBuilder (Genscript), chúng tôi nhận thấy tuy bộ kit eClone cho số khuẩn lạc xuất hiện trên đĩa sàng lọc thấp hơn nhưng tỷ lệ dương tính của 2 phương pháp này là tương tự nhau. Các kết quả đánh giá độc lập tại các đơn vị nghiên cứu khác cũng cho thấy họ đã tạo dòng thành công rất nhiều đoạn gene khác nhau khi sử dụng bộ kit này. Những kết quả này cũng cấp những bằng chứng cho thấy tiềm năng thương mại lớn của bộ kit eClone.

Lời cảm ơn: Nghiên cứu được tài trợ bởi Đại học Quốc gia Thành phố Hồ Chí Minh trong khuôn khổ đề tài mã số DS2020-18-01.

TÀI LIỆU THAM KHẢO

Cohen SN (2013) DNA cloning: a personal view after 40 years. PNAS 110(39): 15521-15529.

Fujitani Y, Yamamoto K, Kobayashi I (1995) Dependence of frequency of homologous recombination on the homology length. *Genetics* 140(2): 797-809. https://www.ncbi.nlm.nih.gov/pubmed/7498755

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166(4): 557-580. https://www.ncbi.nlm.nih.gov/pubmed/6345791

Jacobus AP, Gross J (2015) Optimal cloning of PCR fragments by homologous recombination in *Escherichia coli. PLoS One* 10(3): e0119221. https://doi.org/10.1371/journal.pone.0119221

Lorsch J (2013) Laboratory methods in enzymology: RNA (Vol. 530). Academic Press.

Shen P, Huang HV (1986) Homologous recombination in Escherichia coli: dependence on substrate length and homology. *Genetics* 112(3): 441-457. https://www.ncbi.nlm.nih.gov/pubmed/3007275

Tu AHT (2008) Transformation of *Escherichia coli* made competent by calcium chloride protocol. *Am Soc Microbiol 8*: 1542-1546.

Valla S, Lale R (2014) DNA cloning and assembly methods (Vol. 119). Springer.

EVALUATING THE EFFICIENCY OF HOMOLOGY-BASED CLONING KIT eClone

Nguyen Thi My Trinh^{*}, Nguyen Van Hau

Laboratory of Molecular Biotechnology, University of Science, Vietnam National University - Ho Chi Minh city

SUMMARY

Nowadays, DNA cloning has been widely used as the basic and critical technique in biology and biotechnology. Due to its importance, a great number of methods has been being developed by multiple laboratories to improve the efficiency and reduce the cost of cloning. For the same purpose, the Laboratory of Molecular Biotechnology, VNUHCM-University of Science has developed an easy, effective and economical cloning kit, named eClone. This kit provides a simple cloning method that allows you to insert one or more DNA fragment(s) into a linearized vector due to the presence of homologous sequences on them. In this study, we investigated the cloning efficiency of this kit. Using eClone kit, we succeeded in cloning 3 genes of 300 bp, 1000 bp, and 1500 bp long containing 10-20 nu homologous sequences into vector pET-28a(+). Besides, when comparing the cloning efficiency of eClone kit and the commercial kit GenBuilder (Genscript, USA), we found that although the eClone kit yielded a lower number of colonies in the selective medium than Genbuilder kit, the possitive rates of these two kits were similar. Importantly, more than 200 genes were successfully cloned into different vectors using the eClone kit, as reported by some other research groups. Taken together, these results suggested the commercial potential of this cloning kit.

Keywords: eClone kit, homologous sequence, DNA cloning, pET-28a(+), cloning efficiency.

Author for correspondence: Tel: 0937750861; Email: ntmtrinh@hcmus.edu.vn

TẠO DÒNG BỔ SUNG (COMPLEMENTATION LINES) TỪ CÂY ĐỘT BIẾN MẤT CHỨC NĂNG GEN DO CHÈN ĐOẠN T-DNA Ở Arabidopsis thaliana

Lê Nguyễn Tiểu Ngọc

Viện Công nghệ sinh học và Môi trường, Trường Đại học Tây Nguyên

TÓM TẮT

Phương pháp tạo dòng bổ sung (complementation lines) trên nền cây đột biến với mục đích khôi phục lại gen đã bị knock-out do đoạn T-DNA chèn vào gen mục tiêu được xem là công cụ hữu ích trong nghiên cứu vai trò chức năng gen đối với quá trình sinh trưởng, phát triển cũng như đáp ứng stress ở thực vật. Trong nghiên cứu này, *Arabidopsis thaliana* mang đột biến *CMAL* được dùng làm đối tượng nghiên cứu. Gen *CMAL* (Chloroplast <u>mra</u>W-Like) đóng vai trò thiết yếu trong quá trình sinh tổng hợp lục lạp, quang hợp và sinh trưởng, phát triển của cây. Sự chèn đoạn T-DNA vào vùng intron của gen *CMAL* tạo ra cây đột biến *cmal* biểu hiện kiểu hình còi cọc, sức sinh trưởng kém, ra hoa chậm, số lượng hạt tạo thành ít. Gen *CMAL* ngoại lai được nhân dòng trong vector đích pBI121, sau đó plasmid *CMAL*/pBI121 được biến nạp vào vi khuẩn *E. coli* và *Agrobacterium tumefaciens.* Dòng vi khuẩn *A. tumefaciens* mang plasmid tái tổ hợp *CMAL*/pBI121 được dùng để chuyển vào cây đột biến *cmal* bằng phương pháp xâm nhập chân không. Các dòng bổ sung (*CMAL*-comlines) tạo ra được sàng lọc bằng cách gieo hạt trên môi trường dinh dưỡng Murashige và Skoog (MS, 1962) bổ sung các chất kháng sinh, đồng thời kiểm tra sự hiện diện của gen *CMAL* trong cây bằng kỹ thuật RT-PCR.

Từ khóa: Arabidopsis thaliana, CMAL, complementation lines, đột biến, nhân dòng, T-DNA.

MỞ ĐẦU

Ở thực vật, việc tạo ra các thể đột biến bằng phương pháp chèn đoạn T-DNA có ý nghĩa quan trọng trong nghiên cứu vai trò chức năng của gen đối với quá trình sinh trưởng, phát triển cũng như đáp ứng các stress phi sinh học (hạn, lạnh, nóng, úng, UV) hoặc/và sinh học (sâu bệnh, nấm, virus) từ môi trường bên ngoài (O'Malley, Ecker, 2010). Trong phản ứng chèn đoạn T-DNA vào hệ gen thực vật, khi nghiên cứu các đột biến của các gen mục tiêu, các đoạn T-DNA được chèn ngẫu nhiên vào các vị trí khác nhau, chẳng hạn như vùng điều hòa (promoter) ở vùng thượng nguồn (5'UTR upstream), vùng mang mã di truyền (exon), vùng không mang mã di truyền (intron), trước và sau điểm kết thúc (stop codon), vùng chứa đuôi poly A ở vùng hạ nguồn (3'UTR downstream) (Azpiroz-Leehan, Feldmann, 1997; Radhamony *et al.*, 2005). Sự tương tác giữa các vùng trong gen mục tiêu có thể gây nên những thay đổi về kiểu hình ở các thể đột biến như thay đổi về hình thái, sức sống, sức sinh trưởng, phát triển, khả năng đáp ứng stress, điều này thể hiện rõ vai trò chức năng của gen trong suốt chu kỳ sống của thực vật (Koncz *et al.*, 1992).

Khi phân tích kiểu hình của các thể đột biến gây ra bằng phương pháp chèn đoạn T-DNA, trong trường hợp chỉ có một vị trí chèn đoạn biểu hiện kiểu hình, để chứng minh kiểu hình đó thực sự gây ra bởi sự knock-out gen mục tiêu, một vài phương pháp được dùng như phương pháp phân tích dị bội nhiễm sắc thể hay phương pháp tạo dòng bổ sung (complementation lines (com-lines)) từ cây đột biến để phục hồi chức năng gen bị knock-out (Radhamony *et al.*, 2005). Trong đó, phương pháp tạo các dòng bổ sung thường được sử dụng phổ biến do tính chính xác, độ tin cậy và đạt hiệu quả cao. Bên cạnh đó, việc sử dụng các dòng bổ sung để so sánh, phân tích hình thái, tính trạng ở mức độ cơ thể, tế bào hoặc/và dưới tế bào được xem là công cụ hữu ích trong nghiên cứu di truyền.

Trong nghiên cứu này, cây *Arabidopsis thaliana* đột biến do chèn đoạn T-DNA làm mất chức năng gen *CMAL* được chọn làm đối tượng nghiên cứu. *CMAL* là một gen trong nhân mã hóa cho protein CMAL, protein này định vị trong lục lạp và hoạt động như một rRNA methyltransferase, giúp hình thành gốc methyl tại vị trí C₁₃₅₂ dạng N4-methylcytidine (m⁴C) trên phân tử 16S rRNA trong lục lạp (Zou *et al.*, 2020). Phân tích kiểu hình các thể đột biến này cho thấy, chỉ có vị trí đoạn T-DNA chèn vào vùng intron thứ 2 gây ra sự thay đổi kiểu hình so với cây đối chứng. Thể đột biến đồng hợp tử *cmal* biểu hiện kiểu hình kém phát triển, cây lùn, còi cọc, lá nhỏ và có màu vàng nhạt, rễ ngắn và số lượng hạt tạo thành ít (Ngoc *et al.*, 2020). Để chứng minh kiểu hình này thực sự do sự knock-out gen *CMAL* gây ra, việc tạo ra các dòng bổ sung *CMAL*-comlines là điều cần thiết. Nghiên cứu này cung cấp cái nhìn tổng quan về quy trình tạo dòng bổ sung từ các thể đột biến mất chức năng gen mục tiêu gây ra bởi sự chèn đoạn T-DNA ở cây *Arabidopsis*.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu

Cây Arabidopsis thaliana: Hạt giống các cây hoang dại đối chứng Columbia-0 (Col-0) và cây đột biến chèn đoạn T-DNA (SAIL_564_E05 (*cmal*)) được cung cấp bởi Trung tâm Tài nguyên Sinh vật Arabidopsis (Columbus, OH, USA).

Gen CMAL: Thông tin về gen như trình tự nucleotide, trình tự acid amin được thu thập từ Ngân hàng dữ liệu cây Arabidopsis thaliana (TAIR - Home Page (arabidopsis.org)). Cây Arabidopsis được trồng, thu mẫu và tách chiết RNA bằng bộ KIT tách chiết thực vật Plant RNeasy (GeneAll, Seoul, Hàn Quốc), sau đó RNA tổng số được sử dụng để tổng hợp cDNA dùng cho tất cả các thí nghiệm về sau.

Vector pGEM-T: Được sử dụng làm vector nhân dòng ở vi khuẩn *E. coli*. Cấu trúc của vector pGEM-T (Promega, USA) thuận tiện cho việc gắn và nhân dòng với các sản phẩm PCR, vector pGEM-T mang các gen kháng ampicillin và gen LacZ giúp dễ dàng sàng lọc dựa vào màu sắc khuẩn lạc mọc trên môi trường chọn lọc.

Vector pET-22b(+): Được sử dụng làm vector nhân dòng và biểu hiện gen ở vi khuẩn *E. coli*. Vector pET-22b(+) mang các gen kháng ampicillin và gắn đuôi 6 Histone vào đoạn gen lạ, điều này giúp dễ dàng tinh sạch protein tái tổ hợp.

Vector pBI121: Được sử dụng làm vector nhân dòng và biểu hiện gen ở thực vật. Vector pBI121 (Clontech, www.clontech.com) mang gen kháng kanamycin giúp sàng lọc trên môi trường nuôi cấy thực vật.

Phương pháp nghiên cứu

Tạo dòng vi khuẩn A. tumefaciens mang plasmid tái tổ hợp CMAL/pBI121

* Nhân dòng vi khuẩn E. coli mang plasmid tái tổ hợp CMAL/pGEM-T và CMAL/pET-22b(+)

Đoạn cDNA mã hóa gen *CMAL* (1305 bp) được nhân dòng trong hệ thống vector pGEM-T, và plasmid tái tổ hợp được biến nạp vào tế bào *E. coli* DH5α bằng phương pháp sốc nhiệt, sau đó vi khuẩn được sàng lọc trên môi trường chứa kháng sinh, tám khuẩn lạc được lựa chọn và thực hiện phản ứng PCR colony để kiểm tra sự hiện diện của gen chuyến. Sau khi xác định được dòng vi khuẩn mong muốn, khuẩn lạc được nuôi cấy, plasmid được tách và giải trình tự để xác định đúng dòng vi khuẩn mang *CMAL*/pGEM-T (Macrogen, Hàn Quốc).

Sau đó, vi khuẩn chứa vector *CMAL*/pGEM-T được nhân dòng, gen *CMAL* được khuếch đại và đưa vào vector pET-22b(+). Quá trình biến nạp và nuôi cấy vi khuẩn được thực hiện như phương pháp chuyển gen vào vector pGEM-T. Sau khi nuôi cấy qua đêm, các khuẩn lạc được chọn ngẫu nhiên và được xác nhận bằng phản ứng PCR colony, các khuẩn lạc có băng DNA trùng khớp với đoạn DNA mục tiêu được chọn lọc, khuẩn lạc được nuôi cấy lỏng và lắc qua đêm, plasmid được tách và giải trình tự để xác định đúng dòng vi khuẩn mang *CMAL*/pET-22b(+) (Macrogen, Hàn Quốc).

* Tạo dòng vi khuẩn *E. coli* mang plasmid tái tổ hợp CMAL/pBI121

Gen *CMAL* được khuếch đại bằng phản ứng PCR từ plasmid tái tổ hợp *CMAL*/pET-22b(+) và nhân dòng trong hệ thống vector pBI121 bằng cách sử dụng các vị trí của enzyme hạn chế *Xbal/Smal*, sau đó phản ứng tạo dòng *CMAL*/pBI121 được biến nạp vào tế bào *E. coli* DH5α bằng sốc nhiệt, nuôi cấy qua đêm trên môi trường LB chứa kanamycin (50 µg/mL). Tám khuẩn lạc được chọn ngẫu nhiên và được xác nhận bằng phản ứng PCR colony, khuẩn lạc có băng DNA trùng khớp với đoạn DNA mục tiêu được chọn lọc, nuôi cấy lỏng và lắc qua đêm, sau đó plasmid được tách chiết và được gửi giải trình tự để xác định đúng dòng vi khuẩn mang plasmid tái tổ hợp (Macrogen, Hàn Quốc).

* Biến nạp plasmid tái tổ hợp CMAL/pBI121 vào vi khuẩn Agrobacterium tumefaciens

Sau khi xác định được gen *CMAL* được chuyển thành công vào vector pBI121, plasmid *CMAL*/pBI121 được chuyển vào vi khuẩn *A. tumefaciens* GV3101 bằng phương pháp đông/rã đông (Jyothishwaran *et al.*, 2006), và vi khuẩn được nuôi cấy qua đêm trên môi trường YEP sàng lọc có bổ sung kanamycin (50 µg/mL) và rifampicin (50 µg/mL). Các khuẩn lạc được chọn và thực hiện PCR để kiểm tra sự hiện diện của gen chuyển. Tiếp đó, khuẩn lạc được chọn lọc và nuôi trường YEP (+ kanamycin + rifampicin) để tăng sinh nhằm thu sinh khối, chuẩn bị cho thí nghiệm tiếp theo.

* Chuẩn bị vi khuẩn cho chuyển gen

Vi khuẩn *A. tumefaciens* GV3101 mang vector *CMAL*/pBI121 được tăng sinh (từ một khuẩn lạc) qua đêm trong 10,0 mL môi trường YEP có bổ sung kanamycin (50 μg/mL) và rifampicin (50 μg/mL) ở 28°C, tốc độ lắc 250 vòng/phút trong điều kiện tối. Sau đó, 1 mL huyền phù được chuyển sang 500 mL YEP có bổ sung kháng sinh và nuôi cấy trong điều kiện tương tự cho đến khi giá trị đo OD₆₀₀ đạt từ 0,8-1,2. Tiếp đó, sinh khối vi khuẩn được ly tâm ở tốc độ 3.000 vòng/phút tại 4°C trong 15 phút. Sinh khối vi khuẩn được huyền phù với 500 mL môi trường MS lỏng (pH 5,7, điều chỉnh với KOH), bổ sung với 15 g sucrose, 0,25 g MES, 100 μL Silwet và 100 μM Acetosyringone. Huyền phù được trộn đều bằng máy khuấy từ.

Chuyển vector CMAL/pBI121 vào thực vật qua trung gian vi khuẩn A. tumefaciens

* Chuẩn bị chậu cây để chuyển gen

Hạt cây đột biến *cmal* được gieo trên chậu đất (hỗn hợp vermiculite, rêu than bùn và đá trân châu theo tỷ lệ 3:1:1), các chậu được đặt trong phòng tối trong thời gian 3 ngày ở 4°C, sau đó các chậu được chuyển sang phòng nuôi cây ở điều kiện ánh sáng dài ngày (chu kỳ 16h sáng/8h tối) ở nhiệt độ 20 – 22°C. Sau 4-6 tuần, các cụm hoa đã xuất hiện và cây đã sẵn sàng cho việc chuyển gen.

* Chuyển vector CMAL/pBI121 vào thực vật qua trung gian vi khuẩn A. tumefaciens

Vi khuẩn *A. tumefaciens* mang plasmid tái tổ hợp *CMAL*/pBI121 được dùng để chuyển vào cây đột biến *cmal* bằng phương pháp xâm nhập chân không (Bechtold và Pelletier, 1998). Các chậu cây sau khi chuyển được đặt trong điều kiện phòng nuôi cây, và tiếp tục chăm sóc cho đến khi cây khô và thu hạt. Hạt được thu hoạch theo từng cây riêng lẻ. Hạt thu được từ cây được chuyển gen được xem như là hạt thế hệ F1.

Sàng lọc các dòng bổ sung (com- lines)

Hạt F1 được khử trùng bề mặt và gieo trên đĩa petri chứa 20 mL môi trường MS có bổ sung kanamycin (50 μg/mL) và carbenicillin (50 μg/mL). Các hạt nảy mầm và mọc lên cây con có lá màu xanh là những cây chuyển gen. Sau 3 tuần nuôi cấy, chuyển các cây con sang chậu đất và nuôi trong điều kiện phòng nuôi cây cho đến khi thu hạt thế hệ F2. Các cây com-line tiếp tục được sàng lọc cho đến khi thu được hạt thế hệ F3, F4 là các cây đồng hợp tử. Mỗi cây tương ứng với một dòng. Sự hiện diện của gen *CMAL* trong cây đột biến *cmal* và trong các dòng com-lines được xác nhận bằng phản ứng RT-PCR với các đoạn mồi đặc hiệu được liệt kê trong Bảng 2.

Thực hiện phản ứng PCR

* Thực hiện phản ứng PCR colony

Phản ứng PCR colony được thực hiện để kiểm tra sự có mặt của gen *CMAL* trong các vector sử dụng ở cả vi khuẩn *E. coli* và vi khuẩn *A. tumefaciens*. Trong nghiên cứu này, các đoạn mồi được dùng được liệt kê trong Bảng 1. Phản ứng được thực hiện như sau: dùng đầu típ (10 μ L) tách lấy khuẩn lạc và nhúng vào hỗn hợp phản ứng PCR bao gồm 10 μ L Taq master mix (2X), 1 μ L của 10 μ M mỗi primer và 8 μ L H₂O. Chu trình nhiệt bao gồm 1 chu kỳ biến tính (95°C/5 phút); 35 chu kỳ tổng hợp DNA (95°C/30 giây; 56°C/30 giây; 72°C/1 phút) và bước tổng hợp cuối cùng 72°C/10 phút.

	- · · ·	••••••		
Tên mồi	Trình tự mồi (5' – 3')	Plasmid tái tổ hợp	Kích thước sản phẩm (bp)	
F_CMAL_T	gagctcCATGGCGGGAGTGATTAG		1214	
R_CMAL_T	aagcttCAACTTCTGAATCACTCTTAAC	CIMAL/pGEIN-1	1314	
F_CMAL_pET	gageteCATGGCGGGAGTGATTAG		1914	
R_CMAL_pET	aagcttCAACTTCTGAATCACTCTTAAC	C///AL/pet-220(+)	1314	
F_CMAL_pBI	tctagaATGGCGGGAGTGATTAGG		1244	
R_ CMAL_pBI	cccgggTCAGTGGTGGTGGTGGTGGTG	CIVIAL/PBITZT	~1044	

Bảng 1. Trình tự các đoạn mồi trong phản ứng PCR

*Chú thích: Chữ viết thường, in đậm là trình tự nhận biết của các enzyme hạn chế.

* Thực hiện phản ứng RT-PCR

Để kiểm tra sự biểu hiện của gen *CMAL*, 100 mg mẫu tươi hoặc đông lạnh được dùng để tách chiết RNA tổng số bằng bộ KIT tách chiết thực vật Plant RNeasy. 500 ng RNA tổng số được sử dụng để tổng hợp cDNA bằng bộ KIT Transcription (Applied Biosystems, Mỹ). Chu trình nhiệt cho phản ứng RT-PCR bao gồm 1 chu kỳ biến tính (95°C/5 phút); 30 chu kỳ tổng hợp DNA (94°C/30 giây; 55°C/30 giây; 72°C/30 giây) và bước tổng hợp cuối cùng 72°C/10 phút. Trong nghiên cứu này, các đoạn mồi đặc hiệu được sử dụng được liệt kê trong Bảng 2.

Tên mồi	Trình tự mồi (5' – 3')	Kích thước sản phẩm (bp)
RT_F_CMAL	ATCATTCAAAGCCACTCTG	220
RT_R_CMAL	GGCCACGAATTCACTATATCTTC	330
RT_F_TUBULIN	CTCAAGAGGTTCTCAGCAGTA	492
RT_R_TUBULIN TCACCTTCTTCATCCGCAGTT		403

Bảng 2. Trình tự các đoạn mồi trong phản ứng RT-PCR

KẾT QUẢ VÀ THẢO LUẬN

Thu nhận gen CMAL từ cDNA cây Arabidopsis đối chứng

Cây đối chứng Col-0 được trồng, thu mẫu và tách chiết RNA tổng số, sau đó phản ứng phiên mã ngược được thực hiện để tổng hợp cDNA. Gen *CMAL* với chiều dài đầy đủ 1305 bp được khuếch đại bằng phản ứng PCR với các đoạn mồi được dùng để nhân dòng trong vector pGEM-T. Kết quả sản phẩm PCR trên **Hình 1** cho thấy, giếng 3,4 xuất hiện các băng DNA có chiều dài tương ứng 1317 bp – đúng với kích thước gen *CMAL* và hai vùng trình tự nhận biết. Phần gel chứa DNA mong muốn được cắt, sau đó DNA được tinh sạch và thu nhận bằng bộ KIT tách chiết (GeneAll, Hàn Quốc).



Hình 1. Kết quả điện di sản phẩm PCR

M: thang chuẩn 1kb Plus (NEB); 1,2: Mẫu không chứa gen CMAL; 3-4: Mẫu có chứa gen CMAL

Tạo dòng vi khuẩn A. tumefaciens mang plasmid tái tổ hợp CMAL/pBI121

* Tạo dòng vi khuẩn E. coli mang plasmid tái tổ hợp CMAL/pBI121

Gen *CMAL* được nhân dòng trong vector pGEM-T mang gen kháng ampicillin và gen LacZ. Sau quá trình biến nạp và nuôi cấy qua đêm trên môi trường LB có bổ sung X-gal. Tám khuẩn lạc màu trắng được chọn lọc và thực hiện phản ứng PCR colony (**Hình 2**). Kết quả sản phẩm điện di cho thấy, từ giếng 1 đến giếng 6 có sự xuất hiện một băng DNA được khuếch đại với kích thước khoảng 1317 bp, tương đương với kích thước mong muốn (gen *CMAL* và hai vùng trình tự nhận biết), điều này khẳng định gen *CMAL* đã được chuyển thành công vào vector pGEM-T. Plasmid tái tổ hợp *CMAL*/pGEM-T được gửi giải trình tự, sau đó, kết quả trình tự được so sánh sự tương đồng với trình tự gốc của gen *CMAL* được thu thập từ Ngân hàng dữ liệu cây *Arabidopsis* (arabidopsis.org) bằng phần mềm CLUSTALW (https://www.genome.jp) để xác định chính xác gen được chuyển vào là gen *CMAL*.





M: Thang chuẩn 1kb Plus (NEB), 1-6: Khuẩn lạc chứa gen CMAL; 7-8: Khuẩn lạc không chứa gen CMAL.

Sau khi xác định được dòng vi khuẩn mang vector *CMAL*/pGEM-T, thông thường có thể tiến hành bước kế tiếp là khuếch đại và nhân dòng gen *CMAL* trong vector đích pBI121 biểu hiện ở thực vật. Tuy nhiên, trong nghiên cứu này, để thuận tiện cho việc tinh sạch protein tái tổ hợp nhằm thực hiện các phản ứng phân tích hoạt tính của protein CMAL, tác giả đã nhân dòng gen *CMAL* trong vector biểu hiện pET-22b(+) giúp dễ dàng tinh sạch protein tái tổ hợp dựa vào đuôi His-tag. Gen *CMAL* được đưa vào vector pET-22b(+) tại vị trí các enzyme giới hạn Sacl/HindIII, plasmid tái tổ hợp *CMAL*/pET-22b(+) được biến nạp vào tế bào *E. coli* BL21, vi khuẩn được nuôi

cấy, plasmid được tách chiết và giải trình tự để xác định đúng dòng vi khuẩn mang plasmid tái tổ hợp *CMAL*/pET-22b(+).

Để gen *CMAL* ngoại lai biểu hiện được trong tế bào thực vật, vector pBI121 được sử dụng để tạo dòng, plasmid tái tổ hợp *CMAL*/pBI121 được biến nạp vào tế bào *E. coli*, và sự có mặt của gen *CMAL* được kiểm định bằng phản ứng PCR colony (**Hình 3A**). Kết quả sản phẩm điện di cho thấy, tất cả 8 giếng đều xuất hiện một băng DNA có kích thước khoảng 1344 bp, kích thước này phù hợp với kích thước tính toán mong muốn, do đoạn gen được chuyển (*CMAL-6His*) là đoạn gen tái tổ hợp bao gồm gen *CMAL* và phần đuôi His-tag gắn vào từ vector pET-22b(+). Plasmid *CMAL*/pBI121 được giải trình tự và so sánh sự tương đồng với trình tự gốc của gen *CMAL* bằng phần mềm CLUSTALW để xác định chính xác gen *CMAL* được chuyển thành công vào vector pBI121.



Hình 3. Kết quả điện di sản phẩm PCR từ khuẩn lạc chứa plasmid tái tổ hợp CMAL/pBI121

(A) Vi khuẩn E. coli, M: Thang chuẩn 1 kb (NEB), 1-8: Khuẩn lạc chứa gen CMAL; (B) Vi khuẩn A. tumefaciens, M: Thang chuẩn 1 kb Plus (NEB), 1-4: Khuẩn lạc chứa gen CMAL.

* Biến nạp plasmid tái tổ hợp CMAL/pBI121 vào vi khuẩn A. tumefaciens

Sau khi chọn được dòng vi khuẩn mục tiêu, plasmid *CMAL*/pBI121 được chuyển vào vi khuẩn *A. tumefaciens* theo phương pháp đông/rã đông, và nuôi cấy chọn lọc trên môi trường chứa kháng sinh thích hợp. Chọn ngẫu nhiên bốn khuẩn lạc và thực hiện phản ứng PCR colony để kiểm tra tính chính xác của gen chuyển (Hình 3B). Kết quả sản phẩm điện di cho thấy, cả 4 giếng gel đều xuất hiện một băng DNA với kích thước khoảng 1344 bp, điều này chứng minh plasmid *CMAL*/pBI121 đã được chuyển thành công vào vi khuẩn *A. tumefaciens*. Tiếp đó, một trong bốn khuẩn lạc được chọn lọc và nuôi cấy tăng sinh trong môi trường YEP có bổ sung các kháng sinh thích hợp để chuẩn bị cho thí nghiệm chuyển vi khuẩn *A. tumefaciens* mang vector *CMAL*/pBI121 vào cây đột biến *cmal.*

Tạo dòng bổ sung CMAL-comlines bằng phương pháp chuyển gen qua vi khuẩn A. tumefaciens

Chậu cây đột biến *cmal* được sử dụng để chuyển plasmid *CMAL*/pBI121 thông qua vi khuẩn *A. tumefaciens*. Sau khi chuyển, các chậu cây được đặt trong phòng nuôi cây và tiếp tục chăm sóc đến khi thu hạt. Hạt F1 được gieo trên môi trường chọn lọc, và các cây con có lá thật màu xanh được chuyển sang đất trồng để thu hạt F2. Đến thế hệ F2, tiếp tục gieo hạt và sàng lọc các đĩa chứa cây con sinh trưởng trên môi trường chọn lọc theo tỷ lệ 3:1 (3 come-line: 1 đột biến). Chọn những cây com-line từ đĩa có tỷ lệ 3:1 chuyển sang chậu đất và tiếp tục trồng cho đến khi thu hạt thế hệ F3, F4.

Để chứng minh sự thay đổi kiểu hình ở cây *cmal* do sự bất hoạt của gen *CMAL*, phải tạo ra được ít nhất 02 dòng đồng hợp tử *CMAL*-comline ở thế hệ F3 hoặc F4 có kiểu hình giống nhau và giống với kiểu hình của cây đối chứng. Trong thí nghiệm này, những phân tích khái quát về hình thái và quá trình sinh trưởng, phát triển của cây cho thấy các cây comline có kiểu hình tương tự như cây đối chứng. Kết quả này khẳng định kiểu hình ở cây đột biến *cmal* là do sự bất hoạt của gen *CMAL* gây ra. Bên cạnh đó, sự có mặt của gen *CMAL* ở các dòng bổ sung cũng đã được kiểm định bằng phản ứng RT-PCR (Hình 4), đây là cơ sở khoa học chứng minh rằng gen *CMAL* ngoại lai đã được chuyển thành công vào cây đột biến *cmal*. Như vậy, hoạt động của gen *CMAL* ngoại lai ở các dòng bổ sung đã bù lại gen *CMAL* bị mất đi trong cây đột biến, cho nên các dòng bổ sung *CMAL*-comline có kiểu hình tương tự như cây đối chứng.

CÔNG NGHỆ GEN



Hình 4. (A), Sự biểu hiện của gen CMAL trong cây được đánh giá bằng RT-PCR; (B), Các cây được chuyển trồng trong đất sau 20 ngày. Col-0: Cây đối chứng, cmal: Cây đột biến, com1-4: Các cây CMAL-comline thế hệ F4; Tubulin: Gen tham chiếu.

KÉT LUẬN

Gen *CMAL* ngoại lai được nhân dòng trong vector pBI121 và được chuyển thành công vào *A. thaliana* đột biến *cmal* thông qua vi khuẩn *A. tumefaciens*. Các dòng bổ sung (*CMAL*-comline) thu được có kiểu hình giống cây hoang dại đối chứng (Col-0) và sự hiện diện của gen *CMAL* trong cây *A. thaliana* chuyển gen được xác nhận bằng RT-PCR. Điều này chứng minh rằng gen *CMAL* ngoại lai đã bổ sung cho hoạt tính của CMAL trong cây đột biến *cmal*, và gen này đóng vai trò quan trọng trong quá trình sinh trưởng và phát triển của *A. thaliana*.

Lời cảm ơn: Xin chân thành cảm ơn GS.TS. Hunseung Kang đã tài trợ về tài chính và đưa ra những lời khuyên quý báu để hoàn thành nghiên cứu này.

TÀI LIỆU THAM KHẢO

Azpiroz-Leehan R, Feldmann KA (1997). T-DNA insertion mutagenesis in Arabidopsis: going back and forth. *Trends Genet*, 13: 152–156.

Bechtold N, Pelletier G (1998). In planta Agrobacterium-mediated transformation of adult *Arabidopsis thaliana* plants by vacuum infiltration. *Methods Mol Biol*, 82: 259-266.

Jyothishwaran G, Kotresha D, Selvaraj T, Srideshikan SM, Rajvanshi PK, Jayabaskaran C (2006). A modified freeze-thaw method for efficient transformation of *Agrobacterium tumefaciens*. *Current Science*, 93(6): 770-772.

Koncz C, Németh K, Rédei GP, Schell J (1992). T-DNA insertional mutagenesis in Arabidopsis. Plant Mol Biol, 20: 963-976.

Ngoc LNT, Park SJ, Huong TT, Lee KH, Kang HS (2021). N4-methylcytidine rRNA methylation in chloroplasts is crucial for chloroplast function, development, and abscisic acid response in *Arabidopsis. J Integr Plant Biol*, 63: 570-582.

O'Malley RC, Ecker JR (2010). Linking genotype to phenotype using the Arabidopsis unimutant collection. *The Plant Journal*, 61: 928-940.

Radhamony RN, Prasad AM, Srinivasan R (2005). T-DNA insertional mutagenesis in *Arabidopsis*: a tool for functional genomics. *Electron J Biotechnol*, 8(1): 82-106.

Zou M, Mu Y, Chai X, Ouyang M, Yu LJ, Zhang L, Meurer J, Chi W (2020). The critical function of the plastid rRNA methyltransferase, CMAL, in ribosome biogenesis and plant development. *Nucleic Acids Res*, 48: 3195-3210.

GENERATION OF COMPLEMENTATION LINES IN THE LOSS-OF-FUCTION MUTANT DUE TO T-DNA INSERTION IN *Arabidopsis thaliana*

Le Nguyen Tieu Ngoc^{*}

Institute of Biotechnology and Environment, Tay Nguyen University

SUMMARY

The generation of the complementation lines in the loss-of-function mutants to complement a knock-out gene due to a T-DNA fragment inserted into the target gene is considered a useful tool in researching of the functional role of gene during plant growth and development as well as stress response. The *cmal* mutant was used in this study. In *Arabidopsis thaliana*, *CMAL* gene (Chloroplast <u>mra</u>W-Like) plays a crucial role in chloroplast biogenesis, photosynthesis, and plant growth and development. The insertion of T-DNA into the second intron of *CMAL* gene to generate the *cmal* mutant displayed growth-defect phenotypes with stunted stems, retarded growth, delayed flowering time, the formed seed number was small. Foreign *CMAL* was cloned into the target pBI121 vector in *E. coli*, plasmid *CMAL*/pBI121 was subsequently transformed into *E. coli* and *Agrobacterium tumefaciens*. After that, *A. tumefaciens* carrying *CMAL*/pBI121 recombinant plasmids were used to transform the *cmal* mutant using vacuum infiltration method. The resulting complementation lines (CMAL-comlines) were screened by sowing the seeds on Murashige and Skoog (MS, 1962) medium supplemented with antibiotics, and the presence of *CMAL* gene was confirmed via RT-PCR.

Keywords: Arabidopsis thaliana, CMAL, complementation lines, mutant, cloning, T-DNA.

Author for correspondence: Tel: 0865769027; Email: Intngoc@ttn.edu.vn

NGHIÊN CỨU ĐIỀU KIỆN BIỀU HIỆN ULVAN LYASE TÁI TỔ HỢP TỪ VI KHUẨN BIỀN *Formosa agariphila* VỚI HOẠT TÍNH PHÂN CẮT ULVAN

Trần Nguyễn Hà Vy^{1*}, Cao Thị Thúy Hằng¹, Huỳnh Hoàng Như Khánh¹, Nguyễn Thị Thuận¹, Phạm Đức Thịnh¹, Trần Hoàng Hải²

¹Viện Nghiên cứu và Ứng dụng Công nghệ Nha Trang, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Ulvan lyase là enzyme xúc tác thủy phân ulvan thông qua cắt các liên kết glycoside trong mạch chính của phân tử ulvan, polysaccharide sulfate hóa được chiết xuất từ rong lục. Hiện nay, ulvan có khối lượng phân tử thấp thu hút sự quan tâm chú ý đặc biệt của các nhà khoa học do chúng thể hiện nhiều hoạt tính sinh học quý như chống oxy hóa, virus, ung thư, chống tăng lipid máu; kháng viêm, tăng hoạt tính điều hòa miễn dịch... Vì thế, hướng tìm kiếm các công cụ hiện đại để bẻ ngắn mạch ulvan là một trong những hướng nghiên cứu có giá trị khoa học và mang tính thực tiễn cao. Trong nghiên cứu này, gene mã hóa ulvan lyase *FA28* có nguồn gốc từ vi khuẩn biển *Formosa agariphila* được tuyển chọn trên dữ liệu online, được thiết kế tái tổ hợp vào *E. coli* BL21 (DE3), được nghiên cứu điều kiện biểu hiện gồm môi trường biểu hiện, nồng độ IPTG và nhiệt độ. Kết quả cho thấy *E. coli* BL21(DE3)-*FA28*-pET-28a(+) biểu hiện tốt trong môi trường LB, 15 h, 20 °C, nồng độ IPTG cảm ứng 2 mM. Ulvan lyase tái tổ hợp thu được có kích thước 48 kDa và có hoạt tính trên cơ chất ulvan chiết từ rong lục (*Ulva lactuca*). Đối chứng âm được sử dụng trong nghiên cứu này là chủng vi khuẩn *E. coli* BL21(DE3) mang plasmid *FA28*-pET-28a(+) không được cảm ứng biểu hiện, đã không biểu hiện protein mục tiêu và không có hoạt tính thủy phân cơ chất ulvan trong tất cả các thí nghiệm. Kết quả này cung cấp thêm một công cụ tiềm năng để điều chế ulvan khối lượng phân tử thấp định hướng ứng dụng trong lĩnh vực y được.

Từ khóa: ulvan lyase, Escherichia coli BL21 (DE3)-FA28-pET-28a(+); biểu hiện gene; tái tổ hợp.

MỞ ĐẦU

Ulvan là sulfate polysaccharide có trong rong lục chi *Ulva* và *Enteromorpha*. Cũng giống như các sulfate polysaccharide từ rong biển khác, ulvan có cấu trúc rất phức tạp, nó được cấu tạo bởi các thành phần chủ yếu là các đường rhamnose, xylose, các acid glucuronic, acid iduronic và nhóm sulfate. Trong những năm gần đây một số công bố khoa học đã cho thấy cấu trúc và thành phần của ulvan đóng vai trò quan trọng trong chức năng sinh học của chúng. Các nghiên cứu này chỉ ra ulvan có trọng lượng phân tử thấp thể hiện hoạt tính sinh học cao hơn, ví dụ như có hoạt tính chống oxy hóa mạnh hơn (Qi *et al.*, 2005), hoạt động điều hòa miễn dịch tốt hơn (Peasure *et al.*, 2016; Tabarse *et al.*, 2018), và tác dụng chống tăng lipid máu cao hơn (Pengzhan *et al.*, 2003).

Để điều chế oligosaccharide dạng ulvan, các phương pháp xử lý truyền thống như hóa học và vật lý đã được sử dụng (Zhu *et al.*, 2021). So với các phương pháp bẻ ngắn mạch polysaccharide truyền thống thì phương pháp cắt mạch bằng enzyme có các điều kiện phản ứng nhẹ hơn (axit thấp đến trung tính) và giúp giảm thiểu tác hại không tốt cho môi trường. Cùng với đó, enzyme có tính đặc hiệu phản ứng cao nên các oligosaccharide tạo thành vẫn giữ được cấu trúc toàn vẹn ban đầu (Zhu *et al.*, 2021; Wijesinghe *et al.*, 2012). Do đó hướng nghiên cứu sử dụng enzyme như là một công cụ tiềm năng để chuyển hóa các polysaccharide thành các oligosaccharide với các hoạt tính sinh học ưu việt hơn là một trong những hướng nghiên cứu có giá trị khoa học và mang tính thực tiễn cao.

Ulvan lyase, một loại enzyme có khả năng phân hủy ulvan. Ulvan lyase thuộc họ polysaccharide lyase, có khả năng cắt các liên kết glycosid trong polysaccharide theo cơ chế loại bỏ liên kết β. Ulvan lyase phá vỡ liên kết giữa rhamnose 3-sulfat và acid uronic trong ulvan, tạo ra oligosaccharide với axit uronic không bão hòa ở đầu không khử.

Theo hiểu biết của chúng tôi, tất cả các ulvan lyase đã xác định đặc tính xúc tác đều được phân lập và tinh chế từ vi khuẩn biển và là enzyme tái tổ hợp (Qin *et al.*, 2018; Gao *et al.*, 2019; Rodrigues *et al.*, 2024). Đó là do, ulvan lyase là một loại enzyme khó thu nhận và tinh sạch từ các nguồn tự nhiên do dễ bị mất hoạt tính trong quá trình tinh sạch. Trong phương pháp sản xuất enzyme tái tổ hợp, hệ biểu hiện *E. coli* là phổ biến nhất để tổng hợp protein tái tổ hợp do nhiều lợi ích, bao gồm dễ nuôi cấy, tốc độ sinh trưởng nhanh, đặc điểm di truyền được nghiên cứu kỹ lưỡng, sự tồn tại của nhiều vector tách dòng và chủng đột biến thương mại cho hệ thống này, cũng như việc thu nhận protein mục tiêu dễ dàng. Để thuận tiện cho quá trình tinh sạch, ulvan lyase cần được tổng hợp một cách hiệu quả và ổn định, cùng với cải tiến quá trình sinh khối tế bào. Biểu hiện protein ngoại lai trong chủng chủ bị ảnh hưởng bởi nhiều yếu tố, trong đó, điều kiện lên men có tác động lớn, dẫn đến hiệu suất

tổng hợp protein không hiệu quả, sai lệch cấu trúc protein và mất chức năng sinh học. Việc này có thể được khắc phục bằng cách kiểm soát chặt chẽ các thông số trong quá trình lên men.

Vì vậy, trong nghiên cứu này, chúng tôi tập trung khảo sát các yếu tố (môi trường lên men, nồng độ chất cảm ứng, nhiệt độ biểu hiện, pH môi trường và thời gian lên men) ảnh hưởng đến khả năng biểu hiện của gene ulvan lyase *FA28*, được tìm kiếm trên ngân hàng dữ liệu online từ *Formosa agariphila* KMM 3901^T, một vi khuẩn biển thuộc nhóm *Favobacterium*, được phân lập từ một loài rong lục ở biển Nhật Bản (Glasson *et al.*, 2017).

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Chủng *E. Coli* DH 5alpha được sử dụng làm chủng nhân bản vector tái tổ hợp. Chủng *E. Coli* BL21 được sử dụng làm chủng biểu hiện ulvan lyase tái tổ hợp. Khuôn gene được tổng hợp nhân tạo từ trình tự tối ưu hoá cho việc biểu hiện trong hệ chủng *E. coli* BL21 (DE3). Plasmid pET-28a(+)được sử dụng để dòng hoá mang gene *FA28* mã hóa ulvan lyase có trình tự protein được lấy từ ngân hàng dữ liệu mã số (WP_038530530), và sự biểu hiện gene được kiểm soát bởi promotor T7 thông qua chất cảm ứng IPGT. Gene này đã được thiết kế mã hóa protein không có tín hiệu peptide ở đầu N-teminal nhưng bổ sung thêm 10x histidine để thuận lợi cho quá trình tinh sạch enzyme tái tổ hợp. Tất cả các chủng vi sinh vật và plasmid được cung cấp bởi Khoa Công nghệ kỹ thuật Sinh học và Y sinh học, Trường Đại học Khoa học Kỹ thuật Đan Mạch, Đan Mạch.

Phương pháp

Tạo dòng chủng E. coli BL21 (DE3) mang vector tái tổ hợp FA28-pET-28a(+)

Trình tự gen thiết kế được tổng hợp bởi hãng Genscript (Piscataway, NJ, USA) và khuếch đại bằng phản ứng PCR với cặp mồi đặc hiệu mang trình tự nhận biết của hai enzyme cắt giới hạn Xhoi và Ndel. Sản phẩm PCR được kiểm tra trên gel agarose 1%, nhuộm màu bằng ethidium bromide và phân tích hình ảnh điện di bằng hệ thống Dual Intensity UV Transillumninator. Sau đó, sản phẩm PCR được tách và gửi đi giải trình tự để kiểm tra lại độ chính xác của gene mục tiêu (Macrogen Europe B.V., Hà Lan). Sản phẩm plasmid tái tổ hợp được biến nạp vào tế bào E. coli DH5alpha bằng phương pháp sốc nhiệt. Sau đó được trải trên môi trường Luria Bertani Broth (LB) có chứa 100 μg/mL kháng sinh kanamycin, nuôi ủ 37 °C trong 24 giờ. Các khuẩn lạc chèn thành công plasmid có chứa gen mã hóa được nuôi tăng sinh, thu sinh khối tế bào và tách chiết plasmid tổ hợp bằng bộ kits Mix & Go Kit (Zymo Research).

Vector tái tổ hợp pET28a(+) cùng được cắt bởi cặp enzyme giới hạn Xhoi và Ndel để tạo đầu dính tương hợp với gen ulvan lyase. Sau đó, sản phẩm PCR được xử lý bởi cặp enzyme cắt giới hạn Xhoi và Ndel để tạo đầu dính và thực hiện phản ứng nối nhờ vào enzyme T4 DNA ligase, nối với pET-28a(+) đã được cắt mở vòng bởi hai enzyme tương ứng.

Plasmid *FA28*-pET-28a(+) được biến nạp vào chủng *E. coli* BL21 (DE3) bằng Mix & Go Kit (Zymo Research). Sau khi biến nạp, chủng *E. coli* BL21 (DE3) mang vector tái tổ hợp *FA28-pET-28a*(+), được gọi là chủng *E. coli* BL21 (DE3) / *FA28-pET-28a*(+) được nuôi cấy trên môi trường thạch LB có kháng sinh kanamycin nồng độ cuối 100 μg/mL để tách dòng và chọn lọc dòng vi khuẩn mang gene mục tiêu. Các chủng vi khuẩn mọc được trên môi trường LB có chứa 100 μg/mL kháng sinh kanamycin được lựa chọn để kiểm tra kết quả tạo dòng thông qua kỹ thuật PCR, sản phẩm PCR được gửi đi giải trình tự để kiểm tra lại độ chính xác của gene *FA28* ulvan Iyase (Macrogen Europe B.V., Hà Lan) (Tran *et al.*, 2022).

Biểu hiện FA28 ulvan lyase tái tổ hợp

Sau khi kiểm tra đúng gene mục tiêu đã được tạo dòng thành công, quá trình biểu hiện được thực hiện như sau: Khuẩn lạc *E. coli* BL21 (DE3)/ *FA28*-pET-28a(+) được chọn lọc và nuôi tăng sinh trong 5 mL môi trường LB lỏng có bổ sung 50 µg/mL kanamycin, 34 µg/mL chloramphenicol và 0,05% (w/v) arabinose ở nhiệt độ 37 °C, lắc 180 rpm qua đêm. Sau đó, dịch nuôi cấy vi khuẩn được cấy chuyển sang môi trường LB lỏng 500 mL với tỉ lệ 1:20 (v/v) và tiếp tục nuôi cấy lắc ở 37°C, 180 rpm cho đến khi OD₆₀₀ đạt giá trị 0,8-1,0 thì bổ sung chất cảm ứng IPTG (Isopropyl β-D-1-thiogalactopyranoside) ở các nồng độ khác nhau ở 20°C, tốc độ lắc 180 rpm. Sau 15 h cảm ứng, sinh khối vi khuẩn được thu nhận bằng ly tâm với tốc độ 7.000 vòng/phút trong 15 phút ở 4°C. Đánh giá khả năng biểu hiện của ulvan Iyase bằng điện di hỗn hợp sinh khối tế bào (whole cells) theo SDS-PAGE và nhận diện protein tái tổ hợp bằng phản ứng Western blot.

Đối chứng âm được sử dụng trong nghiên cứu này là chủng vi khuẩn *E. coli* BL21 (DE3) mang plasmid *FA28*-pET-28a(+) không được cảm ứng biểu hiện.

Ảnh hưởng điều kiện biểu hiện gene FA28 ulvan lyase tái tổ hợp

Hai môi trường khác nhau gồm LB (yeast extract 0,5%, peptone 1,0%, NaCl 1,0%, pH 7,5) và môi trường tự cảm ứng AT (0,6% w/w Na₂HPO₄; 0,3% w/w KH₂PO₄; 2,0% w/w tryptone; 0,5% w/w yeast extract; 0,5% w/w NaCl; 0,06% v/v glycerol; 0,05% w/v glucose và 0,04% w/v α -lactose, pH 7,5) được khảo sát cho sự biểu hiện gene *FA28* mã hóa ulvan lyase. Đối với môi trường LB được thực hiện theo các thông số của điều kiện chuẩn nêu trên;

đối với môi trường AT, quá trình cấy chuyển từ ống 5 mL sang môi trường biểu hiện AT được thực hiện ở 20 °C, 180 rpm trong 20 h. Song song với quá trình khảo sát điều kiện môi trường biểu hiện, tiến hành kiểm tra sự biểu hiện ulvan lyase theo thời gian với các mốc thời gian là 5 h, 10 h,15 h và 20 h. Môi trường và thời gian biểu hiện phù hợp được lựa chọn cho các thí nghiệm tiếp theo. Sự biểu hiện của ulvan lyase bằng điện di hỗn hợp sinh khối tế bào (whole cells) theo SDS-PAGE và Western blot.

Đối với môi trường biểu hiện LB, nồng độ chất cảm ứng được khảo sát là IPTG 0-2 mM, nhiệt độ biểu hiện được đánh giá tại các điểm 14°C; 16°C; 18°C; 20°C; 25°C; và 30°C. Sự biểu hiện của ulvan lyase được đánh giá bằng điện di enzyme thô thu nhận được theo SDS-PAGE và Western blot và hoạt tính enzyme theo phương pháp C-PAGE trên cơ chất fucoidan chiết xuất từ rong *S. mcclurei.*

Chiết xuất enzyme thô

Kết thúc quá trình biểu hiện, tiến hành ly tâm dịch lên men (7.000 rpm, 15 phút, 4°C) để thu nhận sinh khối tế bào. Huyền phù sinh khối tế bào trong dịch đệm 20 mM Tris-HCl, 250 mM NaCl, 20 mM imidazole, pH 7,4, sau đó bổ sung thêm lyzozyme 0,2 mg/mL để hỗ trợ quá trình phá vỡ tế bào. Các tế bào được phá vỡ bằng sóng siêu âm (UP400S Hielscher, Teltow, Germany) chu kỳ 0,5 và cường độ 100% trong 2 phút, sau đó để ở nhiệt độ phòng 15 phút. Dịch enzyme thô được thu nhận bằng cách ly tâm thu dịch nổi (19.000 rpm, 4°C, 20 phút), dịch nổi được lọc qua màng lọc 0,22 µm để loại xác tế bào. Enzyme thô thu nhận được tiến hành đo hàm lượng protein theo Bradford với chất chuẩn là albumin huyết thanh bò (Bradford, 1976), sau đó thực hiện SDS-PAGE và Western blot.

Thử nghiệm hoạt tính enzyme bằng C-PAGE (Carbohydrate-Polyacrylamide Gel Electrophoresis)

Hỗn hợp phản ứng chứa dung dịch enzyme có nồng độ 0,5 g/L trong môi trường đệm Tris-HCl 20 mM, pH 7,4, 250 mM NaCl và 10 mM CaCl₂ (đệm A), và 1% w/v fucoidan trong đệm A. Các mẫu được ủ ở 35°C trong vòng 24 h. Hỗn hợp phản ứng (5 μL) được pha trộn với 5 μL đệm tải (20% glycerol và 0,02% phenol đỏ). Sau đó, sử dụng 5 μL mẫu để thực hiện điện di trên gel polyacrylamide 20% (w/v) dày 1 mm với đệm chạy là Tris-borate 100 mM, pH 8,3 trong 1 h. Quá trình nhuộm gel được thực hiện qua hai bước: đầu tiên là với dung dịch chứa 0,05% alcian blue 8GX (Panreac, Barcelona, Tây Ban Nha) trong 2% axit acetic trong 45 phút, sau đó là với dung dịch 0,01% O-toluidine blue (Sigma-Aldrich, Steinheim, Đức) trong 50% ethanol và 1% axit acetic. Mẫu chuẩn (đối chứng dương) là sản phẩm phản ứng của FFA2 trên fucoidan từ rong nâu *Fucus evanescens* do Khoa Công nghệ kỹ thuật Sinh học và Y sinh học, Trường Đại học Khoa học Kỹ thuật Đan Mạch, Đan Mạch cung cấp.

KẾT QUẢ VÀ THẢO LUẬN

Tạo dòng và biểu hiện FA28 ulvan lyase tái tổ hợp

Plasmid *FA28*-pET-28a(+) được biến nạp vào chủng *E. coli* BL21 (DE3), khuẩn lạc mọc trên môi trường thạch LB chọn lọc có chứa 100 µg/mL kanamycin được lựa chọn để tiến hành PCR khuếch đại đoạn gene *FA28* nhằm kiểm tra lại sự thành công của quá trình tạo dòng. Kết quả điện di DNA cho thấy sản phẩm PCR của gene *FA28* ulvan lyase có kích thước khoảng 1324 bp, tương đồng với kích thước của DNA được thiết kế (Hình 1). Đồng thời sản phẩm PCR cũng được gửi đi giải trình tự và sau đó so sánh trình tự với trình tự gene ban đầu để kiểm tra độ chính xác của gene đã được biểu hiện, sự tương đồng cao giữa gene trước và sau khi biểu hiện. Như vậy, rõ ràng FA28 ulvan lyase đã được tạo dòng thành công, khuẩn lạc đều đã mang vector tái tổ hợp.

Kết quả biểu hiện protein FA28 ulvan lyase ở điều kiện tiêu chuẩn được thể hiện ở Hình 2. Từ kết quả điện di trên SDS-PAGE (Hình 2A) cho thấy, protein FA28 ulvan lyase biểu hiện có khối lượng phân tử khoảng 48 kDa. Đây cũng chính là khối lượng phân tử tương đương với khối lượng phân tử protein mục tiêu FA28. Để xác định chính xác sự có mặt của protein tái tổ hợp mong muốn, chúng tôi tiến hành phản ứng lai miễn dịch Western blot (Hình 2B), kết quả cho thấy, xuất hiện vạch màu đỏ tương ứng với protein kích thước 48 kDa ngay tại vị trí biểu hiện của protein mục tiêu FA28 ulvan lyase, do chủng vi khuẩn *E. coli* BL21 (DE3)-*FA28*-pET-28a(+) sản xuất. Phản ứng màu sẽ xuất hiện khi có sự liên kết giữa protein tái tổ hợp mang đuôi His-tag và kháng thể monoclonal anti-polyHistidine peroxidase-conjugated antibody. Ngoài ra, khi tiến hành thí nghiệm tương tự với chủng vi khuẩn *E. coli* BL21 (DE3) mang plasmid *FA28*-pET-28a(+) không được cảm ứng biểu hiện, chúng tôi nhận thấy không có xuất hiện vạch protein mục tiêu (vị trí 48 kDa) ở trên gel điện di SDS-PAGE và phản ứng lai miễn dịch Western blot cũng không xuất hiện bất cứ vệt có màu đỏ nào (số liệu không nêu lên ở bài báo này).



37 25 20 15 10

Hình 1. Điện di sản phẩm PCR gene FA28 ulvan lyase Thang chuẩn kích thước DNA (GeneRulerTM 1 kb DNA Ladder, Thermo Sientific, Mỹ).



kDa 1

A

в

(A) - Điện di SDS-PAGE; (B) - Western blot. Trong đó: 1. Thang chuẩn; 2. Mẫu sinh khối sau biểu hiện.

Ảnh hưởng của điều kiện biểu hiện gene FA28 ulvan lyase tái tổ hợp

Kết quả khảo sát 02 công thức môi trường AT và LB thể hiện như Hình 3 cho thấy, ở các môi trường khác nhau thì sự biểu hiện FA28 ulvan lyase tái tổ hợp là khác nhau. Môi trường AT, là môi trường tự cảm ứng, bao gồm nguồn nitơ cao, nguồn carbon glucose/glycerol nhưng lượng ulvan lyase thu được trong 300 µL môi trường lên men thấp hơn so với môi trường LB (Hình 3). Do đó, môi trường LB được lựa chọn cho sự biểu hiện FA28 ulvan lyase tái tổ hợp do lượng protein được tổng hợp là nhiều nhất và ổn định nhất. LB là một môi trường cơ bản cho sự biểu hiện của protein tái tổ hợp, dễ chuẩn bị, thành phần môi trường cung cấp đầy đủ các peptide, axit amin thiết yếu và muối khoáng natri cho sự sinh trưởng của vi khuẩn *E. coli.* Kết quả khảo sát thời gian biểu hiện FA28 ulvan lyase cho thấy, sau 5 h lên men, enzyme đã biểu hiện và lượng protein tăng dần theo thời gian, sau 10-15 h cảm ứng protein FA28 ulvan lyase tái tổ hợp dược tổng hợp nhiều và biểu hiện mạnh và không thay đổi nhiều khi tăng thời gian biểu hiện lên 20 h. Do đó trong khoảng thời gian 15 h sau khi cảm ứng được lựa chọn là thời gian tối ưu cho sự biểu hiện FA28 ulvan lyase tái tổ hợp.

Đối chứng âm được sử dụng trong nghiên cứu này là chủng vi khuẩn *E. coli* BL21 (DE3) mang plasmid *FA28*pET-28a(+) không được cảm ứng biểu hiện. Kết quả thể hiện trên Hình 3, ở vị trí đối chứng âm – được kí hiệu là C cho thấy, hoàn toàn không có xuất hiện vạch protein mục tiêu (vị trí 48 kDa) ở trên gel điện di SDS-PAGE, trong khi các protein khác của chủng vi khuẩn *E. coli* đều được thể hiện. Điều này chứng tỏ, vạch protein kích thước 48 kDa chính là protein mục tiêu.

Protein tái tổ hợp được điều khiển tổng hợp bởi promoter T7 trên vector pET-28a(+). Promoter này được cảm ứng bởi sự có mặt của IPTG trong môi trường nuôi cấy. Nồng độ chất này ảnh hưởng đến mức độ biểu hiện protein tái tổ hợp mục tiêu. Ở nồng độ quá cao nó gây độc cho tế bào, trong khi ở nồng độ quá thấp protein thường được tổng hợp kém. Do đó, chúng tôi đã tiến hành bổ sung IPTG với các nồng độ khác nhau là 0,5 mM; 1 mM; 1,5 mM: 2 mM; 2,5 mM và 3 mM, kết quả thể hiện ở Hình 4A. Kết quả cho thấy với nồng độ 2 mM, IPTG FA28 ulvan lyase được tổng hợp mạnh nhất. Khi tăng IPTG lên đến 3 mM quá trình biểu hiện hầu như không đối. Do đó, chúng tôi chọn nồng độ cảm ứng IPTG 2 mM cho những nghiên cứu tiếp theo.







St - Thang protein chuẩn, AT - môi trường tự cảm ứng, LB - môi trường lên men E. coli được cảm ứng bởi IPTG; 5 h,10 h,15 h, 20 h - thời gian lên men, C – Không cảm ứng biểu hiện (ITPG 0mM).

Hình 4. Ảnh hưởng của điều kiện lên men đến sự biểu hiện và hoạt tính của FA28 ulvan lyase tái tổ hợp (A) Nồng độ IPTG, (B) Nhiệt độ.

Nhiệt độ là yếu tố quan trọng ảnh hưởng đến hiệu suất và chất lượng protein tái tổ hợp. Một trong những nguyên nhân là do nhiệt độ làm ảnh hưởng đến tính ổn định của plasmid trong tế bào. Nhiệt độ càng cao thì khả năng

CÔNG NGHỆ GEN

dào thải plasmid của chủng chủ càng lớn. Đồng thời nhiệt độ còn làm cho mRNA dễ bị phá hủy. Để thu protein FA28 ulvan lyase tái tổ hợp có lượng lớn mà không lẫn quá nhiều protein của *E. coli*, nhằm làm tăng hiệu quả hoặc giảm lược cho quá trình tinh sạch này, chúng tôi tiến hành khảo sát khả năng tổng hợp protein FA28 ulvan lyase ở các điều kiện nhiệt độ khác nhau tại 14°C, 16°C, 18°C, 20°C, 25°C và 30°C. Kết quả của thí nghiệm này được thể hiện như trên Hình 4B. Từ kết quả trên cho thấy rằng, nhiệt độ nuôi cấy càng tăng thì mức độ biểu hiện của protein tái tổ hợp và các protein khác của *E. coli* càng tăng. Tuy nhiên, hoạt tính của ulvan lyase bị ảnh hưởng rất lớn bởi nhiệt độ, khi sử dụng enzyme trong khảo nghiệm này để thử hoạt tính, chúng tôi nhận thấy enzyme thể hiện hoạt tính tốt nhất khi biểu hiện ở 20°C (số liệu không nêu lên ở bài báo này). Do đó, với mục đích thu nhận protein FA28 ulvan lyase tái tổ hợp thể hiện hoạt tính tốt chúng tôi chọn nhiệt độ cho quá trình biểu hiện là 20°C.

Kết quả hoạt tính thủy phân ulvan của FA28 ulvan lyase tái tổ hợp



Hình 5. Hoạt tính của của FA28 ulvan lyase

Ulac – ulvan được điều chế từ rong lục Ulva lactuca; St - đối chứng dương của điện di C-PAGE; R - Sản phẩm của phản ứng thủy phân ulvan bởi FA28 ulvan lyase; C – đối chứng âm: phản ứng giữa dịch chiết thô của vi khuẩn E. coli BL21 (DE3) không được biểu hiện và cơ chất ulvan.

Ulvan lyase FA28 chiết xuất thô được sử dụng để thử nghiệm khả năng bẻ mạch ulvan chiết xuất từ loài rong *U. lactuca* thu nhận tại vùng biển Nha Trang-Khánh Hoà, kết quả thể hiện ở Hình 5. Kết quả cho thấy sự khác biệt của mẫu ulvan chưa thủy phân (Ulac) và mẫu thủy phân (R). Bên cạnh đó, mẫu đối chứng âm (C) hoàn toàn không xảy ra phản ứng, do đó chứng tỏ rằng chỉ có ulvan lyase tái tổ hợp thể hiện hoạt tính chứ không phải thành phần protein nào khác trong *E. Coli* tái tổ hợp. Hình ảnh C-PAGE ở Hình 5 cho thấy, ulvan lyase FA28 đã bẻ mạch ulvan thành những oligomer sulfate hóa có kích thước khác nhau trên C-PAGE. Điều này cho thấy tiềm năng ứng dụng của ulvan lyase FA28 để điều chế ulvan chiết xuất từ rong lục Việt Nam. Quá trình phân cắt để tạo oligosaccharide dạng ulvan này còn được đánh giá thông qua phương pháp quang phổ. Hoạt tính ulvan lyase được xác định bằng sự gia tăng các sản phẩm không bão hòa (liên kết đôi trong hỗn hợp phản ứng), khi đo độ hấp thụ ở bước sóng 235 nm (Manns *et al.*, 2013). Kết quả ở Hình 5 cho thấy, ulvan lyase FA28 chiết thô từ dịch nội bào đã thể hiện rõ hoạt tính thuỷ phân ulvan từ rong lục *U. lactuca* trong thí nghiệm lặp lại 3 lần.

KÉT LUẬN

Trong nghiên cứu này, chúng tôi đã thành công trong nghiên cứu tạo dòng và biểu hiện ulvan lyase FA28 bằng môi trường LB và thời gian để enzyme biểu hiện tốt là 15 h sau khi cảm ứng bằng IPTG ở nồng độ 2 mM ở nhiệt độ 20 °C. Enzyme có kích thước là 48 kDa, đồng thời enzyme này đã bẻ mạch ulvan thành các oligomer sulfate hóa có kích thước khác nhau. Điều này chứng tỏ tiềm năng ứng dụng của ulvan lyase FA28 trong việc điều chế ulvan từ rong lục Việt Nam. Cần tiến hành thêm các nghiên cứu xác định đặc điểm xúc tác của enzyme và tối ru hóa quá trình tạo oligosaccharide để có thể thu nhận được sản phẩm hiệu quả cao nhất. Nghiên cứu này cung cấp những thông tin quan trọng về biểu hiện và hoạt tính của ulvan lyase FA28, đồng thời hướng tới cho các nghiên cứu tiếp theo về ứng dụng của enzyme này trong lĩnh vực công nghệ sinh học và y học.

Lời cảm ơn: Kết quả bài báo được hỗ trợ thực hiện bằng nguồn kinh phí từ Đề tài thuộc các hướng KHCN ưu tiên cấp Viện Hàn lâm Khoa học và Công nghệ Việt Nam, mã số VAST06.01/23-24.

TÀI LIỆU THAM KHẢO

Gao J, Du C, Chi Y, Zuo S, Ye H, Wang P (2019). Cloning, Expression, and Characterization of a New PL25 Family Ulvan Lyase from Marine Bacterium *Alteromonas* sp. A321. *Mar Drugs*, 17(10): 568.

Glasson CRK, Sims IM, Carnachan SM, et al. A cascad- ing biorefinery process targeting sulfated polysaccharides (ulvan) from *Ulva ohnoi. Algal Res*, 2017; 27: 383-391.

Mann AJ, Hahnke RL, Huang S, Werner J, Xing P, Barbeyron T (2013). The genome of the alga associated marine flavobacterium *Formosa agariphila* KMM 3901 reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol*, 79: 6813- 6822.

Peasura N, Laohakunjit N, Kerdchoechuen O (2016). Assessment of biochemical and immunomodulatory activity of sulphated polysaccharides from *Ulva intestinalis*. Int J Biol Macromol, 91: 269–277.

Pengzhan Y, Ning L (2003). Antihyperlipidemic effects of different molecular weight sulfated polysaccharides from *Ulva pertusa* (Chlorophyta). *Pharmacol Res*, 48: 543–549.

Qi H, Zhao T (2005). Antioxidant activity of different molecular weight sulfated polysaccharides from *Ulva pertusa* Kjellm (Chlorophyta). *J Appl Phycol*, 17, 527–534.

Qin HM, Xu P, Guo Q, Cheng X, Gao D, Sun D, Zhu Z, Lu F (2018). Biochemical characterization of a novel ulvan lyase from *Pseudoalteromonas* sp. strain PLSV. *RSC Adv*, 8(5): 2610-2615.

Rodrigues VJ, Jouanneau D, Fernandez-Fuentes N et al (2024). Biochemical characterisation of a PL24 ulvan lyase from seaweed-associated *Vibrio* sp. FNV38. *J Appl Phycol*, 36: 697-711.

Tabarsa M,You S, Dabaghian EH, Surayot U (2018). Water-soluble polysaccharides from *Ulva intestinalis*: Molecular properties, structural elucidation and immunomodulatory activities. *J Food Drug Anal.* 26: 599-608.

Tran VHN, Nguyen TT, Meier S, Holck J, Cao HTT, Van TTT, Meyer AS, Mikkelsen MD (2022) The Endo-α(1,3)-Fucoidanase Mef2 Releases Uniquely Branched Oligosaccharides from *Saccharina latissima* Fucoidans. *Marine Drugs*, 20(5): 305.

Wijesinghe W, Jeon YJ (2012). Enzyme-assistant extraction (EAE) of bioactive components: A useful approach for recovery of industrially important metabolites from seaweeds: A. review. *Fitoterapia*, 83: 6–12.

Zhu B (2021). Marine oligosaccharides originated from seaweeds: Source, preparation, structure, physiological activity and applications. *Crit Rev Food Sci Nutr*, 61: 60-74.

STUDY ON EXPRESSION CONDITIONS OF RECOMBINANT ULVAN LYASE FROM MARINE BACTERIA *Formosa agariphila* WITH ULVAN DEGRADING ACTIVITY

Tran Nguyen Ha Vy^{1*}, Cao Thi Thuy Hang¹, Huynh Hoang Nhu Khanh¹, Nguyen Thi Thuan¹, Pham Duc Thinh¹, Tran Hoang Hai²

¹Nha Trang Institute of Technology Research and Application, VAST

²Graduate University of Science and Technology, VAST

SUMMARY

Ulvan lyase is an enzyme that catalyzes the hydrolysis of ulvan by cleaving glycosidic linkages within the main chain of the ulvan molecule, a sulfated polysaccharide extracted from green seaweed. Nowadays, low molecular weight ulvan attracts special attention from modern scientists because it has many valuable biological activities such as: anti oxidation, anti virus, anti cancer, anti inflammatory, anti increased blood lipids, anti inflammation, increased immune regulatory activity... Therefore, the study on modern tools to degrade ulvan into ulvan oligosaccharides is one of the research directions with high scientific value and practicality. In this study, the gene encoding the ulvan lyase *FA28*, sourced from the marine bacterium *Formosa agariphila*, was selected based on online data and recombined into *E. coli* BL21 (DE3). The expression conditions, including the expression medium, IPTG concentration, and temperature were investigated. The results showed that *E. coli* BL21 (DE3)/ *FA28* pET 28a(+) expressed well in LB medium, for 15 hours at 20 °C, 2 mM IPTG. The reassembled ulvan lyase obtained had a size of 48 kDa and exhibited activity on ulvan substrates extracted from green seaweed *Ulva lactuca*. The negative control used in this study was the *E. coli* strain BL21 (DE3) carrying the uninduced expression plasmid *FA28* pET-28a(+), which did not express the target protein in all experiments. Besides, the negative control sample had absolutely no enzyme-substrate reaction, thus proving that only recombinant ulvan lyase showed activity and not any other protein components in *E. coli* recombinant. These findings provide an additional potential tool for the production of low molecular weight ulvan, with potential applications in the field of medicine.

Keywords: Ulvan lyase, E. coli BL21 (DE3)-FA28-pET-28a(+), gene expression, recombinant.

Author for corresspondence: Tel: 0901973350; Email: havy158@gmail.com

XÁC ĐỊNH VÀ PHÂN TÍCH CÁC GENE HSP70 Ở CÂY CỦ CẢI ĐƯỜNG (Beta vulgaris L.) BẰNG PHƯƠNG PHÁP TIN SINH HỌC

Lê Thị Mận, Trần Thị Mai Lan, Chu Thị Bích Ngọc,

Nguyễn Thị Thanh Hương, Nguyễn Phương Quý, Cao Phi Bằng*

Trường Đại học Hùng Vương

TÓM TẮT

Ở thực vật, họ gene Hsp70 đóng vai trò quan trọng trong quá trình sinh trưởng, phát triển và phản ứng với các điều kiện bất lợi phi sinh học. Nghiên cứu này nhằm xác định và phân tích đặc điểm họ Hsp70 ở cây củ cải đường. Kết quả nghiên cứu cho thấy có 32 thành viên họ gene Hsp70 đã được xác định trong hệ gene của cây củ cải đường. Ngoại trừ 3 gene (*BvHsp70-11, BvHsp70-20* và *BvHsp70-24*), tất cả các BvHsp70 còn lại đều là các gene phân mảnh với số lượng exon nằm trong khoảng từ 2 đến 14. Các BvHsp70 mang các đặc điểm hóa lí đặc trưng đã được phát hiện ở các Hsp70 ở nhiều loài thực vật khác như phần lớn các BvHsp70 có tính acid, ưa nước. Các BvHsp70 được phát hiện có thể phân bố trong ti thể (BvHsp70-28 và BvHsp70-32), lục lạp (BvHsp70-19 và BvHsp70-27), lưới nội chất (BvHsp70-23 và BvHsp70-25) và trong dịch bào (các BvHsp70 còn lại). Dựa trên kết quả phân tích cây phát sinh chủng loại, các BvHsp70 được xếp vào 5 nhóm, từ nhóm I đến nhóm V, trong đó nhóm V (phân bố trong dịch tế bào) có số lượng thành viên nhiều nhất. Những phát hiện này sẽ là cơ sở cho các nghiên cứu sâu hơn về tách dòng gene, phân tích biểu hiện, chức năng của các gene Hsp70 ở cây củ cải đường.

Từ khóa: Cây Củ cải đường, cấu trúc gene, cây phát sinh chủng loại, đặc điểm hóa-lí, Hsp70.

MỞ ĐẦU

Củ cải đường (*Beta vulgaris* L.) là một nguồn thực phẩm giàu chất dinh dưỡng bao gồm vitamin, khoáng chất, chất xơ, protein và nhiều chất có hoạt tính sinh học, chủ yếu bao gồm betalain và các thành phần khác có hoạt tính chống oxy hóa như coumarin, carotenoid, sesquiterpenoids, triterpenes và flavonoid (Singh *et al.*, 2014; Mirmiran *et al.*, 2020). Củ cải đường và các sản phẩm của nó mang lại nhiều lợi ích cho sức khỏe và có thể giúp ngăn ngừa nhiều loại bệnh như phòng ngừa tiểu đường, bảo vệ gan, hạ huyết áp và làm lành vết thương (Domínguez *et al.*, 2017, 2018; Bangar *et al.*, 2022; Mirmiran *et al.*, 2020)). Vì vậy, các nghiên cứu để nâng cao năng suất, chất lượng và khả năng chống chịu với các điều kiện môi trường bất lợi của cây Củ cải đường rất được chú trọng.

Các protein sốc nhiệt 70 kD (HSP70) đại diện cho một nhóm phân tử được phân bố rộng rãi trong tất cả các giới của sự sống, đóng vai trò sinh học quan trọng trong sự sinh trưởng, phát triển và khả năng chống stress của thực vật (Usman et la., 2017, Chen et al., 2019, Berka et al., 2022, Xu et al., 2024). Các Hsp70 có chức năng duy trì cấu trúc gấp nếp, ngăn cản sự kết tủa protein khi tế bào thực vật bị stress, do đó, các Hsp70 có vai trò rất quan trọng đối với tính chống chịu của thực vật (Usman et al., 2017). Đến nay họ gene Hsp70 đã được xác định và phân tích ở nhiều loài thực vật khác nhau, chẳng hạn như Arabidopsis thaliana (18 gene) (Lin et al., 2001), khoai tây (Solanum tuberosum) (20 gene) (Liu et al., 2018), diễm mạch (Chenopodium quinoi) (16 gene) (Liu et al., 2018), bắp cải (Brassica oleracea var. capitata) (52 gene) (Su et al., 2019), đậu xanh (Vigna radiata) (32 gene) (Jasrotia et al., 2019), bí đỏ (Cucurbita moschata) (21 gene) (Davoudi et al., 2022). Tuy nhiên họ gene này chưa được nghiên cứu ở Củ cải đường.

Nghiên cứu này hướng tới mục tiêu xác định và phân tích đặc điểm các gene Hsp70 bằng phương pháp tin sinh học ở cây Củ cải đường. Kết quả nghiên cứu này sẽ là cơ sở cho các nghiên cứu chuyên sâu về tách dòng gene, phân tích chức năng của các gene trong họ HSP70 và chọn giống cây Củ cải đường trong đáp ứng các điều kiện bất lợi của môi trường.

PHƯƠNG PHÁP NGHIÊN CỨU

Co sở dữ liệu: Hệ gene của cây Củ cải đường (*Beta vulgaris* L.) đã được giải trình tự (Dohm *et al.*, 2014, McGrath *et al.*, 2020) được đặt trên trang phytozome (https://phytozome-next.jgi.doe.gov/info/Bvulgarisssp_vulgaris_EL10_2_2).

Phương pháp nghiên cứu:

Phương pháp xác định gene Hsp70 trong hệ gene cây củ cải đường: Chương trình TBLASTN (Gertz et al., 2006) được sử dụng để tìm kiếm các gene tương đồng trên toàn hệ gene của cây Củ cải đường, các protein Hsp70 của

cây *Arabidopsis thaliana* (mã số Genbank: CAB85987, CAB85986, AAF14038, BAB02269, AAF18501, CAB45063, BAA97012, AAD15393, CAB37531, CAB89371, AAF88019, BAB08435, AAB70400, AAG51503, AAG52240, AAG52244, AAD30257) (Lin *et al.*, 2001; Sung *et al.*, 2001) được sử dụng làm khuôn dò.

Phân tích in silico các đặc điểm lí, hóa lí thuyết của các protein: Các đặc điểm vật lí, hóa học của các protein nghiên cứu được phân tích nhờ phần mềm của ExPASy (Expert Protein Analysis System; Gasteiger et al., 2003).

Dự đoán vị trí phân bố dưới tế bào: Vị trí phân bố dưới tế bào của các Hsp70 của cây củ cải đường được xác định *in silico* bằng cách sử dụng công cụ Yloc (Briesemeister *et al.*, 2010).

Phân tích in silico cấu trúc của các gene mã hóa Hsp70: Cấu trúc intron-exon của các gene được xây dựng bằng công cụ Gene structure display server v2 (GSDS) (Hu et al., 2015). gDNA và CDS được đưa vào máy chủ hiển thị cấu trúc gene (GSDS) để xây dựng cấu trúc exon/intron.

Tạo cây phả hệ: Các trình tự protein Hsp70 của các loài *Arabidopsis*, diêm mạch và củ cải đường được sắp hàng bằng cách dùng phần mềm MAFFT (Katoh *et al.*, 2013). Cây phát sinh chủng loại được xây dựng bằng phần mềm MEGA 11 nhờ sử dụng phương pháp Maximum Likelihood và tuân theo các tham biến: mẫu Jones-Taylor-Thornton (JTT), dữ liệu được xử lí là tất cả các vị trí và phương pháp Bootstrap với 1000 lần lặp lại (Tamura *et al.* 2021).

KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Kết quả xác định của họ gene Hsp70 và đặc điểm hóa lí của các BvHsp70

Tổng số 32 gene của họ Hsp70 đã được xác định từ hệ gene của cây Củ cải đường (Bảng 1). Như vậy, số lượng thành viên của họ HSP70 ở cây củ cải đường bằng với họ gene này ở cây đậu xanh (*Vigna radiata*) (32 gene) (Jasrotia *et al.*, 2019), nhưng lớn hơn ở cây *A. thaliana* (18 gene) (Lin *et al.*, 2001), bí đỏ (*Cucurbita moschata*) (21 gene) (Davoudi *et al.*, 2022), cũng như cây sắn (*Manihot esculanta*) (22 gene) (Đồng Thị Xiêm *et al.*, 2022). Số lượng gene Hsp70 ở cây củ cải đường cũng lớn hơn ở cây diêm mạch (16 gene theo nghiên cứu trước đây (Liu *et al.*, 2018), 24 gene theo nghiên cứu bổ sung của chúng tôi). Tuy nhiên, số lượng gene Hsp70 của cây củ cải đường thơn ở cây cải bắp (*Brassica oleracea* var. *capitata*) (52 gene) (Su *et al.*, 2019).

Bảng 1. Đặc điểm cấu trúc, hóa lí của các gene trong họ Hsp70 trong hệ gene cây củ cải đường

Gene	Locus	GS (bp)	PL (aa)	mW (kDa)	pl	GRAVY	SLC
BvHsp70-01	Bevul.1G090100	2921	599	66,14	6,19	-0,274	С
BvHsp70-02	Bevul.1G090200	3410	642	71,20	5,62	-0,308	С
BvHsp70-03	Bevul.1G090400	3165	624	69,30	6,10	-0,305	С
BvHsp70-04	Bevul.1G090500	2888	650	71,98	5,28	-0,278	С
BvHsp70-05	Bevul.1G090600	3197	631	69,83	5,11	-0,344	С
BvHsp70-06	Bevul.1G090700	2260	627	70,24	5,57	-0,415	С
BvHsp70-07	Bevul.1G090800	5770	442	50,58	6,60	-0,527	С
BvHsp70-08	Bevul.1G090900	2972	631	69,70	5,24	-0,334	С
BvHsp70-09	Bevul.1G091000	2717	627	69,96	6,51	-0,281	С
BvHsp70-10	Bevul.1G091100	2255	624	69,65	5,63	-0,375	С
BvHsp70-11	Bevul.1G091200	1713	570	64,52	7,92	-0,418	С
BvHsp70-12	Bevul.1G091300	5028	638	70,96	5,46	-0,332	С
BvHsp70-13	Bevul.1G091400	2639	602	66,78	7,51	-0,250	С
BvHsp70-14	Bevul.1G091500	3557	623	69,93	6,25	-0,290	С
BvHsp70-15	Bevul.1G091700	5270	641	71,18	5,60	-0,332	С
BvHsp70-16	Bevul.1G091800	2222	637	71,30	6,41	-0,345	С
BvHsp70-17	Bevul.1G092000	2405	579	64,72	5,45	-0,266	С
BvHsp70-18	Bevul.1G092400	4191	582	65,55	5,81	-0,398	С
BvHsp70-19	Bevul.1G202400	3832	700	75,41	5,14	-0,320	Ср
BvHsp70-20	Bevul.1G209700	1968	655	71,78	5,37	-0,423	С
BvHsp70-21	Bevul.2G099600	2859	651	71,30	5,13	-0,395	С
BvHsp70-22	Bevul.2G099800	2758	647	71,02	5,10	-0,427	С
BvHsp70-23	Bevul.3G034500	3517	665	73,34	4,98	-0,449	ER
BvHsp70-24	Bevul.3G138300	1716	571	62,07	5,28	0,056	С
BvHsp70-25	Bevul.4G205100	4079	664	73,77	5,15	-0,470	ER

BvHsp70-26	Bevul.5G261000	8357	894	99,20	5,17	-0,421	С	
BvHsp70-27	Bevul.6G153600	3908	713	75,75	5,16	-0,302	Ср	
BvHsp70-28	Bevul.6G225800	3672	679	72,79	5,43	-0,282	М	
BvHsp70-29	Bevul.7G208800	3027	648	71,28	5,17	-0,474	С	
BvHsp70-30	Bevul.8G056100	14003	737	82,24	5,31	-0,392	С	
BvHsp70-31	Bevul.9G203400	4941	850	93,50	5,20	-0,413	С	
BvHsp70-32	Bevul.9G216900	3359	678	72,59	5,53	-0,309	М	

Chú thích: GS = Kích thước gene, PL = Chiều dài phân tử protein, MW = Khối lượng phân tử protein, pl = Điểm đẳng điện, GRAVY= Độ ưa nước trung bình, SCL = Vị trí phân bố dưới tế bào, bp: Base pair, aa = Amino acid, C = Dịch bào, ER = Lưới nội chất, M = Ti thể, Cp = Lục lạp.

Kích thước các gene trong họ này dao động từ 1713 bp (*BvHsp70-11*) đến 14003 bp (*BvHsp70-30*). Các protein suy diễn có chiều dài trong khoảng từ 442 cho đến 894 aa, khối lượng phân tử dao động từ 50,58 đến 99,20 kDa. Chỉ số pl của các protein suy diễn từ các gene trong họ này nằm trong khoảng từ 4,98 tới 7,92, trong đó phần lớn protein có tính acid ngoại trừ BvHsp70-11 (pl=7,92) và BvHsp70-13 (pl=7,51). Chỉ số Gravy của 31 trên tổng số 32 protein thuộc họ này nhỏ hơn 0 (ngoại trừ BvHsp70-24), điều đó cho thấy rằng các phân tử này là protein ưa nước (Bảng 1). Các đặc điểm này khá tương đồng với các đặc điểm của Hsp70 của cây *A. thaliana* (Lin *et al.*, 2001), sắn (Đồng Thị Xiêm *et al.*, 2022), ca cao (La, 2022).

Dự đoán vị trí phân bố dưới tế bào của các protein

Kết quả dự đoán vị trí phân bố dưới tế bào đã chỉ ra rằng BvHsp70-19 và BvHsp70-27 phân bố trong lục lạp, BvHsp70-28 và BvHsp70-32 phân bố trong ti thể, BvHsp70-23 và BvHsp70-25 phân bố trên màng lưới nội chất. Các phân tử Hsp70 còn lại được dự đoán phân bố trong dịch bào (Bảng 1). Kết quả nghiên cứu của chúng tôi tương đồng với các kết quả nghiên cứu trước đó ở cây ca cao (La, 2022), diêm mạch (Liu *et al.*, 2018).

Đặc điểm cấu trúc gene của họ BvHsp70 gene

Ngoại trừ 3 gene *BvHsp70-11, BvHsp70-20* và *BvHsp70-24* mã hóa liên tục, các *BvHsp70* còn lại đều là các gene phân mảnh, số exon của các gene BvHsp70 phân mảnh này nằm trong khoảng từ 2 đến 14 (Hình 1). Có 14 trên tổng số 32 gene *BvHsp70* có 2 exon, 6 gene có 3 exon (*BvHsp70-01, BvHsp70-04, BvHsp70-13, BvHsp70-14, BvHsp70-15, BvHsp70-17)*, 1 gene có 5 exon (*BvHsp70-28)*, 2 gen có 6 exon (*BvHsp70-25, BvHsp70-32*), 4 gene có 8 exon (*BvHsp70-19, BvHsp70-23, BvHsp70-27, BvHsp70-30*), 1 gene thuộc họ này có 9 exon (*BvHsp70-31*), 14 exon (*BvHsp70-26*) (Hình 1).



Hình 1. Cấu trúc gene của họ BvHsp70 trong hệ gene cây củ cải đường

Hiện tượng cấu trúc exon đa dạng của các gene BvHsp70 này cũng được ghi nhận trong các báo cáo trước đây của họ gene Hsp70 ở các loài thực vật khác như cây đậu tương (*Glycine max* L.) (Zhang, 2015), khoai tây

(Solanum tuberosum) (Liu et al., 2018), bắp cải (Brassica oleracea) (Su et al, 2019), ca cao (Theobroma cacao) (La, 2022), củ cải (Raphanus sativus L.) (Pan et al., 2024).

Phân tích phát sinh chủng loại họ Hsp70 ở cây Củ cải đường

Để tìm hiểu mối quan hệ tiến hóa của các gene Hsp70, phân tích phát sinh chủng loại đã được tiến hành trên cơ sở các trình tự protein Hsp70 của cây Củ cải đường và cây *A. thaliana* (Hình 2).



Hình 2. Sơ đồ cây phát sinh chủng loại được xây dựng từ các Hsp70 của cây củ cải đường (Bv), diêm mạch (Cq) và *A. thaliana (At)*

Cây phát sinh chủng loại cho phép phân loại các Hsp70 thành 5 nhóm. Nhóm I bao gồm hai thành viên (BvHsp70-19 và BvHsp70-27), tương đồng với các Hsp70 phân bố ở lạp thể của cây *A.thaliana*. Nhóm II gồm hai thành viên (BvHsp70-28 và BvHsp70-32), tương đồng với các Hsp70 phân bố ở ti thể của cây *A.thaliana*. Nhóm III cũng gồm hai thành viên (BvHsp70-23 và BvHsp70-25), tương đồng với các Hsp70 gắn với màng lưới nội chất của cây *A.thaliana*. Nhóm IV gồm bốn thành viên (BvHsp70-26, BvHsp70-26, BvHsp70-30, và BvHsp70-31), tương đồng với các Hsp110 của cây *A.thaliana*. Nhóm cuối cùng, nhóm V, gồm 22 thành viên, tương đồng với các Hsp70 phân bố trong dịch tế bào của cây *A. thaliana* (Hình 2). Sự sắp xép các Hsp70 của cây củ cải đường vào 5 nhóm tương đồng với ở cây sắn và *A. thaliana* với 5 nhóm (Đồng Thị Xiêm *et al.*, 2022). Tuy nhiên, sự sắp xếp thành 5 nhóm này khác với ở cây củ cải (*R. sativus*) (Pan *et al.*, 2024), ở loài này, các Hsp70 được sắp xếp thành 6 nhóm. Ở cây diêm mạch, các Hsp70 cũng được sắp xếp thành 6 nhóm (I – VI) (Liu *et al.*, 2018), tuy nhiên trong nghiên cứu ở cây diêm mạch, chỉ có 16 Hsp70 được phát hiện nên số lượng Hsp70 ở nhóm phân bố trong dịch hệu đính. Rất đáng chú ý rằng, ở các nhóm I, II, III và IV, số lượng Hsp70 của cây củ cải đường nhiều hơn 5,5 lần so với ở cây diêm mạch. Kết quả nghiên cứu này gọi ý rằng có một quá trình nhân gene Hsp70 nhóm V diễn ra mạnh mẽ ở cây củ cải đường.

KẾT LUẬN

Trong hệ gene của cây Củ cải đường có 32 thành viên của họ Hsp70 đã được xác định. Chiều dài các gene trong họ này dao động từ 1713 đến 14003 pb. Các protein suy diễn có chiều dài trong khoảng từ 442 đến 894 aa, khối lượng phân tử dao động từ 50,58 đến 99,20 kDa. Phần lớn các protein trong họ này có tính acid ngoại trừ 02 đại diện. Có 31 trên tổng số 32 protein thuộc họ này là protein ưa nước. Các gene này được xếp vào 5 nhóm trong đó nhóm V có số lượng thành viên nhiều nhất. Ngoại trừ 3 gene (*BvHsp70-11, BvHsp70-20* và *BvHsp70-24*) thì tất cả các *BvHsp70* còn lại đều là các gene phân mảnh, với số exon của các gene này nằm trong khoảng từ 2 đến 14. Phân tích *in silico* họ gene *Hsp70* ở cây củ cải đường có ý nghĩa lớn, là tiền đề cho việc tách dòng gene và phân tích chức năng của các gene trong họ Hsp70 ở loài cây này trong đáp ứng với các điều kiện stress.

Lời cảm ơn: Công trình này được hoàn thành với sự hỗ trợ kinh phí từ chương trình nghiên cứu khoa học cơ bản của Trường Đại học Hùng Vương (Đề tài mã số: 16/2022/HĐKH.HV16.2022).

TÀI LIỆU THAM KHẢO

Bangar SP, Sharma N, Sanwal N, Lorenzo JM, Sahub JK (2022). Bioactive potential of beetroot (*Beta vulgaris*), *Food Res Int* 158, August 2022, 111556 https://doi.org/10.1016/j.foodres.2022.111556

Berka M, Kopecká R, Berková V, Brzobohatý B, Černý M (2022). Regulation of heat shock proteins 70 and their role in plant immunity. *J Exp Bot*, 73(7): 1894-1909. doi:10.1093/jxb/erab549.

Chen X, Shi L, Chen Y, Zhu L, Zhang D, Xiao S, Aharoni A, Shi J, Xu J (2019). Arabidopsis HSP70-16 is required for flower opening under normal or mild heat stress temperatures. *Plant, Cell* doi:https://doi.org/10.1111/pce.13480.

Briesemeister S, Rahnenfuhrer J, Kohlbacher O (2010). YLoc--an interpretable web server for predicting subcellular localization. *Nucleic Acid Res*, 38(Web Server issue), W497-502. doi:10.1093/nar/gkq477

Davoudi M, Chen J, Lou Q (2022). Geneome-Wide Identification and Expression Analysis of Heat Shock Protein 70 (HSP70) Gene Family in Pumpkin (Cucurbita moschata) Rootstock under Drought Stress Suggested the Potential Role of these Chaperones in Stress Tolerance. *Int J Mol Sci*, 23(3): 1918. https://doi.org/10.3390/ijms23031918

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, Rupp O, Sörensen TR, Stracke R, Reinhardt R, Goesmann A, Schulz B, Stadler PF, Schmidt T, Gabaldón T, Lehrach H, Weisshaar B, Himmelbauer H (2014). The geneome of the recently domesticated crop plant sugar beet (Beta vulgaris). *Nature*, 505: 546–549.

Đồng Thị Xiêm, Nguyễn Thị Oanh, Trần Thị Thanh Huyền, Trần Thị Mai Lan, Lê Thị Mận, Nông Thị Thu Huyền, Cao Phi Bằng (2022). Xác định và phân tích các gene mã hóa Hsp70 ở cây sắn (Manihot esculanta) bằng phương pháp tin sinh học, *Báo cáo khoa học về nghiên cứu và giảng dạy sinh học ở Việt Nam - Hội nghị khoa học quốc gia lần thứ 5,* 874-883 DOI: 10.15625/vap.2022.0098

Domínguez R, Cuenca E, Maté-Muñoz J, García-Fernández P, Serra-Paya N, Estevan MC, Herreros PV, Garnacho-Castaño MV (2017). Effects of beetroot juice supplementation on cardiorespiratory endurance in athletes. *System Rev Nut*, 9(1): 43.

Domínguez R, Maté-Muñoz JL, Cuenca E, García-Fernández P, Mata-Ordoñez F, Lozano-Estevan MC, et al (2018). Effects of beetroot juice supplementation on intermittent high-intensity exercise efforts. *J Int Soc Sports Nutr*, 15(1): 2.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, 31: 3784-3788.

Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol*, 4: 41. doi:10.1186/1741-7007-4-41.

Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2015). GSDS 2.0: An upgraded gene visualization server. *Bioinfor*, 31(8): 1296-1297. doi:10.1093/bioinformatics/btu817.

Jasrotia RS, Jaiswal S, Yadav PK, Raza M, Iquebal MA, Rai A, Kumar D (2019). Geneome-Wide Analysis of HSP70 Family Protein in Vigna radiata and Coexpression Analysis Under Abiotic and Biotic Stress. *J Comput Biol*, 27(5): 738-754. doi:10.1089/cmb.2019.0166.

Katoh K, Standley (2013). MAFT multiple sequence alignmen software version 7: improvements in performance and usability. *Mol biol evol*, 30: 772-800

La VH (2022). Geneome-Wide Identification and Analysis of Heat Shock Protein 70 Family in Theobroma cacao, *Pak J Biol Sci*, 25 (7): 608-618, 2022

Lin BL, Wang JS, Liu HC, Chen RW, Meyer Y, Barakat A, Delseny M (2001). Geneomic analysis of the Hsp70 superfamily in *Arabidopsis thaliana. Cell Stress Chaperones*, 6(3): 201-208. doi:10.1379/1466-1268(2001)0062.0.co;2.

Liu J, Pang X, Cheng Y, Yin Y, Zhang Q, Su W, Hu B, Guo Q, Ha S, Zhang J, Wan H (2018). The Hsp70 gene family in Solanum tuberosum: Geneome-wide identification, phylogeney, and expression patterns. *Sci Rep*, 8, 16628(2018). https://doi.org/10.1038/s41598-018-34878-7.

Liu J. Wang R, Liu W, Zhang H, Guo Y, Wen R (2018). Genome-Wide Characterization of Heat-Shock Protein 70s from Chenopodium quinoa and Expression Analyses of Cqhsp70s in Response to Drought Stress. *Genes* (Basel), 9(2). doi:10.3390/genes9020035

McGrath JM, Funk A, Galewski P, Ou S, Townsend B, Davenport K, Daligault H, Johnson S, Lee J, Hastie A, Darracq A, Willems G, Barnes S, Liachko I, Sullivan S, Koren S, Phillippy A, Wang J, Lu T, Pulman J, Childs K, Yocum A, Fermin D, Mutasa-Göttgenes E, Stevanato P, Taguchi K, Shu S, Naegele R, Dorn K. (2020). A contiguous de novo geneome assembly of sugar beet EL10 (*Beta vulgaris* L.). *DNA Research.* doi: https://doi.org/10.1101/2020.09.15.298315

Mirmiran P, Houshialsadat Z, Gaeini Z, Bahadoran Z, Azizi F (2020). Functional properties of beetroot (*Beta vulgaris*) in management of cardio-metabolic diseases. *Nutrition & Metabolism*, 17,3. https://doi.org/10.1186/s12986-019-0421-0

Pan X, Zheng Y, Lei K, Tao W, Zhou N (2024). Systematic analysis of Heat Shock Protein 70 (HSP70) gene family in radish and potential roles in stress tolerance. *BMC Plant Biol* 24, 2. https://doi.org/10.1186/s12870-023-04653-6

Singh B, Hathan BS (2014). Chemical composition, functional properties and processing of beetroot-a review. *Int J Sci Eng Res*, 5(1): 679-684.

Su H, Xing M, Liu X, Fang Z, Yang L, Zhuang M, Zhang Y, Wang Y, Honghao Lv (2019). Geneome-wide analysis of HSP70 family genes in cabbage (*Brassica oleracea* var. capitata) reveals their involvement in floral development. *BMC Geneomics*, 20, 369. https://doi.org/10.1186/s12864-019-5757-3

Sung D Y, Vierling E, Guy CL (2001). Comprehensive expression profile analysis of the Arabidopsis Hsp70 gene family. *Plant Physiol*, 126(2): 789-800. https://doi.org/10.1104/pp.126.2.789

Tamura K, Stecher G, Kumar S (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol*, 38(7): 3022-3027. doi:10.1093/molbev/msab120.

Xu T, Zhou H, Feng J, Guo M, Huang H, Yang P, Zhou J (2024). Involvement of HSP70 in BAG9-mediated thermotolerance in Solanum lycopersicum, *Plant Physiol Biochem*, Volume 207, https://doi.org/10.1016/j.plaphy.2024.108353

Usman MG, Rafii MY, Martini MY, Yusuff OA, Ismail MR, Miah G (2017). Molecular analysis of Hsp70 mechanisms in plants and their function in response to stress. *Biotechnol Genet Eng Rev*, 33(1): 26-39. doi:10.1080/02648725.2017.1340546.

Zhang L, Zhao HK, Dong QL, Zhang YY, Wang YM, Li HY, Xing GJ, Li QY, Dong YS (2015). Geneome-wide analysis and expression profiling under heat and drought treatments of HSP70 gene family in soybean (*Glycine max* L.). *Front Plant Sci*, Vol. 6. 10.3389/fpls.2015.00773.

IDENTIFYING AND ANALYZING HSP70 GENES IN SUGAR BEET (*Beta vulgaris* L.) BY USING BIOINFORMATICS METHODS

Le Thi Man, Tran Thi Mai Lan, Chu Thi Bich Ngoc, Nguyen Thi Thanh Huong, Nguyen Phuong Quy, Cao Phi Bang*

Hung Vuong University

SUMMARY

The Hsp70 family plays an important role in plant growth, development, and response to adverse abiotic conditions. This study aims to identify and analyze the characteristics of the Hsp70 family in sugar beet. The results show that there are 32 genes of the Hsp70 gene family identified in the genome of sugar beet. Most *BvHsp70* genes are fragmented, with the number of exons ranging from 2 to 14, except for three genes, *BvHsp70-11, BvHsp70-20* and *BvHsp70-24, respectively*. The BvHsp70s possess characteristic physio-chemical properties that have been observed in Hsp70s in many other plant species, such as acidity and hydrophilicity. BvHsp70s were found to be distributed in mitochondria (BvHsp70-28 and BvHsp70-32), chloroplasts (BvHsp70-19 and BvHsp70-27), endoplasmic reticulum (BvHsp70-23 and BvHsp70-25), and in the cytosol (the remaining BvHsp70s). Based on phylogenetic analysis, the BvHsp70s are classified into five groups, from Group I to Group V, with Group V (cytoplasm localized) having the highest number of members. These findings provide a basis for further studies on gene cloning, expression analysis, and the functions of Hsp70 genes in sugar beet.

Keywords: Sugar beet, gene structure, phylogenetic tree, physio-chemical properties, Hsp70.

^{*} Author for correspondence: Tel: 0904922412; Email: phibang.cao@hvu.edu.vn

PHÂN LẬP VÀ PHÂN TÍCH GEN MÃ HÓA β-1,3-GLUCANASE TỪ Bacillus sp. 41

Nguyễn Thị Thanh Huyền^{1,2}, Lê Thị Nhật Anh¹, Trần Quốc Dung^{3*}, Nguyễn Đức Huy^{1*}

¹Viện Công nghệ sinh học, Đại học Huế

²Trường THPT chuyên Lê Quý Đôn, Đông Hà, Quảng Trị

³Trường Đại học Sư phạm, Đại học Huế

TÓM TẮT

β-1,3-glucanase là một enzyme tiềm năng trong hệ thống phòng thủ chống lại sự tấn công của các loại nấm gây hại ở thực vật. Trong nghiên cứu này, gen mã hóa cho một β-1,3-glucanase (*Bgl*) từ chủng *Bacillus* sp. 41 đối kháng *Aspergilus niger* đã được khuếch đại, tạo dòng, giải trình tự nucleotide và phân tích cấu trúc thành công. Kết quả thu được cho thấy gen có kích thước 655 bp, mã hoá cho một chuỗi polypeptide có kích thước 214 aa. Trình tự *Bgl* chứa đoạn mã hóa cho chuỗi tín hiệu peptide dài 29 aa, chứng tỏ *Bgl* mã hóa cho protein ngoại bào. Cấu trúc không gian dự đoán của protein này được mô tả tương đồng cao với β-1,3-1,4-glucanase ở *Bacillus subtilis* với 55% trình tự amino acid là phiến β. Dựa trên các đặc điểm *Bgl* cũng như cấu trúc protein mã hóa cho thấy *Bgl* đóng vai trò quan trọng trong hệ gen của *Bacillus*. Do đó, phân lập thành công gen *Bgl* là một dữ liệu di truyền quan trọng hỗ trợ cho các nghiên cứu tiếp theo trong việc hiểu rõ cơ chế đối kháng của vi khuẩn *Bacillus* sp. đối với các loại nấm gây hại. Bên cạnh đó, kết quả tạo dòng thành công *Bgl* cung cấp nguyên liệu di truyền để tạo ra enzyme tái tổ hợp thông qua biểu hiện trong vật chủ thích hợp.

Từ khóa: β-1,3-glucanase, Bacillus, Bgl, tạo dòng.

ĐẶT VẤN ĐỀ

Trong quá trình sinh trưởng và phát triển, các loài thực vật đã xây dựng một loạt các cơ chế phòng thủ nhằm chống lại sự tấn công của các tác nhân sinh học có nguy cơ gây hại cho bản thân. Để cân bằng khả năng phòng vệ với các hoạt động sinh lý diễn ra bên trong cơ thể, thực vật đã sử dụng các biện pháp sinh hoá bao gồm: sản xuất phytoalexin tăng cường sự phát triển của thành tế bào (lignin hoá và suberin hoá), kích thích sự biểu hiện của các protein chống tác nhân gây hại (protein PR) (Bhuiyan *et al.*, 2009; Tiku, 2020). Protein PR được báo cáo lần đầu tiên vào năm 1970 trong nghiên cứu về lá thuốc lá bị nhiễm *Tobacco mosaic* virus (Van Loon *et al.*, 1970). Hiện nay, 17 họ protein PR đã được công nhận dựa trên sự tương đồng về trình tự amino acid, đặc tính sinh học và cơ chế hoạt động của enzyme (Sinha *et al.*, 2014). Các PR thế hiện đa dạng chức năng đối với thực vật như: hoạt tính kháng nấm, diệt côn trùng, kháng virus, kháng khuẩn, vận chuyển chất lỏng, kích thích nảy mầm ở hạt, phát triển ống phấn và sự thụ tinh. β -1,3-glucanase là một protein PR2 điển hình được biết đến với chức năng tăng sức đề kháng của thực vật trước sự tấn công của nấm bệnh. Tuy nhiên, không phải loài thực vật nào cũng có khả năng sản sinh ra enzyme này để chống lại nấm bệnh, do đó, việc tìm kiếm sinh vật có khả năng sản sinh enzyme này hỗ trợ việc phòng chống nấm bệnh ở thực vật là cần thiết (Worrall *et al.*, 1992).

Ở các giới sinh vật khác như nấm, β-1,3-glucanase tham gia vào quá trình mở rộng của tế bào và giải phóng bào tử (El-Katatny *et al.*, 2001); ở vi khuẩn enzyme này phá hủy thành tế bào nấm để sử dụng làm nguồn cung cấp dinh dưỡng (Watanabe *et al.*, 1992). Nhiều nghiên cứu đánh giá về khả năng kiểm soát sinh học nấm bởi β-1,3-glucanase từ vi sinh vật mang lại hiệu quả cao đã được công bố, trong đó đề cập nhiều phải kể đến là các loài thuộc chi *Bacillus* như *B. circulans, B. amyloliquefaciens, B. subtilis,...* (Magnin-Robert *et al.*, 2007; Otsuka *et al.*, 2022; Duy, 2023; Masilamani *et al.*, 2013; Phến at al.,2016).

Trong nghiên cứu này, chúng tôi báo cáo về quá trình nhân bản và giải trình tự của gen mã hoá β-1,3-glucanase phân lập từ chủng *Bacillus* sp. 41. Dựa trên kết quả giải trình tự, cấu trúc bậc 2 của protein chức năng đã được dự đoán và đối chiếu với dữ liệu của ngân hàng PDB. Kết quả của báo cáo là tiền đề cho các nghiên cứu chuyên sâu trong công tác tuyển chọn, sản xuất các protein tái tổ hợp dạng PR2 phục vụ phòng trừ nấm bệnh trong sản xuất nông nghiệp và công nghiệp.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu

Chủng vi khuẩn *Bacillus* sp. 41 được phân lập từ đất nông nghiệp Tỉnh Quảng Trị được bảo quản và lưu trữ tại Viện Công nghệ sinh học, Đại học Huế (Huyền *et al.*, 2024). Chủng vi khuẩn được nuôi cấy và lưu trữ trong môi

trường LB chứa 1% pepton, 0,5% dịch chiết nấm men, 1% NaCl. Chủng vi khuẩn *Escherichia coli* TOP10 được nuôi cấy trên môi trường LB rắn chứa 2% agar.

Phương pháp nghiên cứu

Phương pháp tách chiết DNA tổng số

DNA tổng số của chủng vi khuẩn *Bacillus* sp. 41 được tách chiết dựa trên phương pháp CTAB theo mô tả của Sambrook và Russell (2003) với một số điều chỉnh: Vi khuẩn sau 18 giờ nuôi cấy trong môi trường LB với tốc độ lắc 180 vòng/phút được ly tâm 10.000 vòng/phút trong 5 phút. Sinh khối được rửa lại bằng nước cất vô trùng để loại bỏ hoàn toàn môi trường nuôi cấy và tái huyền phù trong 500 µL đệm CTAB (100 mM Tris-HCI, pH 8; 1,4 M NaCI; 20 mM EDTA; 2% CTAB; 0,2% β-mercaptoethanol). Tế bào được ủ ở 65°C trong 15 phút. Dịch chiết tế bào chứa DNA tổng số được thu hồi bằng ly tâm trong 5 phút với tốc độ 10.000 vòng/phút. DNA tổng số được tách chiết và tinh sạch trong 700 µL hỗn hợp gồm: phenol, chloroform, isopropanol (tỷ lệ 25:24:1), trộn đều bằng vortex và phân tách bằng ly tâm 5 phút với tốc độ 10.000 vòng/phút. Khoảng 500 µL pha trên được chuyển sang ống 1,5 mL mới và DNA tổng số được kết tủa với 2 lần thể tích ethanol tinh khiết, rửa kết tủa bằng ethanol 70% và hòa tan trong 50 µL đệm TE. DNA tổng số sau đó được kiểm tra chất lượng bằng điện di trên gel agarose 1% với thuốc nhuộm 6× GelRed Loading Buffer (ABT-VN).

Phân lập gen β-1,3-glucanase từ Bacillus sp. 41

Để phân lập gen mã hoá β-1,3-glucanase từ DNA tổng số của chủng Bacillus sp. 41 đã được tách chiết, trình tự mồi đặc hiệu được thiết kết dựa trên các phân tích trước đó về gen β-1,3-glucanase và dữ liệu hệ gen chủng vi khuẩn Bacillus halotolerans Y6 (GenBank: MH643779). Trình tự mồi xuôi là BgIF 5'-ATGTCTACTGCATCAGCT CA-3' và trình tự mồi ngược BgIR 5'-TTATTTTTTGTATAGCGC-3'. Phản ứng khuếch đại chuỗi polymerase (PCR) được tiến hành với thành phần phản ứng gồm 50 ng DNA tổng số, 10 pmol mồi xuôi, 10 pmol mồi ngược, 1x MyTaq™ Mix (Bioline, Đức) trong thể tích phản ứng 20 μL. Chu trình nhiệt PCR được thực hiện ở 95°C trong 10 phút; 30 chu kỳ với 95°C trong 60 giây, 55°C trong 30 giây và 72°C trong 90 giây; kéo dài ở 72°C trong 10 phút. Sản phẩm PCR được điện di trên gel agarose 1% trong đệm TAE (Tris acetic acid EDTA) với hiệu điện thế 70V trong 30 phút, sử dụng thuốc nhuộm 6x GelRed Loading Buffer (ABT-VN).

Tạo dòng gen β-1,3-glucanase

Sản phẩm PCR có kích thước phù hợp được phân tách và tinh sạch từ gel agarose bằng Kit tách chiết – TopPURE® PCR/GEL DNA Purification Kit (HI-412) (ABT, Việt Nam). Sản phẩm PCR sau tinh sạch được điện di kiểm tra trên gel agarose 1% trước khi tạo dòng và giải trình tự. Khoảng 40 ng sản phẩm PCR sau tinh sạch được ủ với 50 ng pGEM®T-Easy vector (Promega, Mỹ) trong phản ứng có 1× đệm gắn, 3 đơn vị T4 DNA ligase ở 4°C qua đêm. Hỗn hợp gắn được biến nạp vào *E. coli* TOP10 bằng phương pháp sốc nhiệt. Tế bào *E. coli* TOP10 được nuôi cấy trên môi trường LB rắn có bổ sung 50 µg/mL ampicillin (Amp), 50 µg/mL 5-Bromo-4-Chloro-3-Indolyl –D-Galactopyranoside (X-gal) và 100 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) trong 16 giờ ở 37°C. Thể tái tổ hợp được chọn lọc dựa vào sự xuất hiện khuẩn lạc có viền màu xanh hoặc màu trắng trên đĩa, trong đó khuẩn lạc màu trắng có chứa vector tái tổ hợp.

Khuẩn lạc *E. coli* TOP10 màu trắng xuất hiện được kiểm tra bằng phản ứng PCR khuẩn lạc với cặp mồi đặc hiệu gen β -1,3-glucanase như mô tả ở trên. Các khuẩn lạc cho kết quả sản phẩm PCR đặc trưng được nuôi cấy trong 5 mL môi trường LB có bổ sung ampicillin qua đêm ở 37°C. Tế bào *E. coli* TOP10 được thu nhận bằng ly tâm lạnh ở 4°C với tốc độ 10.000 vòng/phút trong 5 phút. DNA plasmid được tách chiết sử dụng Kit tách chiết – TopPURE® Plasmid DNA Extraction Kit (HI-142) (ABT, Việt Nam) theo hướng dẫn của nhà sản xuất. Sự hiện diện gen β -1,3glucanase trong thể tái tổ hợp với vector pGEM®T-Easy được kiểm tra bằng enzyme cắt giới hạn *KpnI* (Thermoscientific, Mỹ) và phản ứng PCR sử dụng cặp mồi M13 (M13F: 5'-GTAAAACGACGGCCAGT-3'; M13R: 5'-CGGATAACAATTTCACACAGG-3') được thiết kế trên vector pGEM®T-Easy. Sản phẩm PCR bằng cặp mồi M13 được gửi đi giải trình tự theo phương pháp Sanger tại Công ty TNHH DNA Sequencing, Việt Nam.

Phân tích trình tự gen β-1,3-glucanase

Trình tự nucleotide gen β -1,3-glucanase được so sánh và đối với chiếu với trình tự gen tương ứng từ chủng *B. halotolerans* Y6 sử dụng phần mềm Clustal Omega (https://www.ebi.ac.uk/jdispatcher/msa/clustalo). Trình tự amino acid suy diễn của gen β -1,3-glucanase được thu nhận bằng công cụ protein translate tool (https://web.expasy.org/translate/). Sự hiện diện của trình tự nucleotide mã hóa tín hiệu peptide của gen β -1,3-glucanase được xác định sử dụng công cụ SignalP-5.0 (https://services.healthtech.dtu.dk/services/SignalP-5.0/). Cuối cùng, cấu trúc không gian 3D của protein β -1,3-glucanase được dự đoán bằng phần mềm Protein Homology/analogy Recognition Engine V 2.0 (Kelley *et al.*, 2015), điểm đẳng điện sử dụng phần mềm IPC2 (https://ipc2.mimuw.edu.pl/).

KẾT QUẢ VÀ THẢO LUẬN

Kết quả tách chiết DNA tổng số

Bacillus sp. 41 sau 18 giờ nuôi cấy trong môi trường LB lỏng, tốc độ lắc 180 vòng/ phút được thu sinh khối và tách chiết DNA tổng số bằng phương pháp CTAB theo mô tả của Sambrook và Russell (2003). DNA tổng số sau

khi tách chiết được điện di trên agarose gel 1%. Kết quả điện di ở hình 1 cho thấy DNA tổng số thu được là một băng duy nhất, có kích thước lớn hơn 10 kb với nồng độ 500 ng/µL, sạch, ít đứt gãy, rõ nét. Chất lượng DNA tổng số đảm bảo để làm nguyên liệu cho những thí nghiệm tiếp theo.



Hình 1. Kết quả điện di DNA tổng số

M: Hypper Ladder 1kb (Bioline); 1: DNA tổng số Bacillus sp. 41

Khuếch đại gen β-1,3-glucanase

Gen mã hoá β -1,3-glucanase (Bgl) được khuếch đại từ DNA tổng số của chủng Bacillus sp. 41 với cặp mồi đặc hiệu Bgl. Kết quả phản ứng PCR cho thấy sản phẩm khuếch đại với kích thước khoảng 600 bp. Do đó, sản phẩm PCR được tinh sạch từ gel agarose 1% tạo nguồn nguyên liệu cho bước tạo dòng và giải trình tự gen.



Hình 2. Kết quả PCR với cặp mồi đặc hiệu

M: Hypper Ladder 1kb (Bioline); 1: Sản phẩm PCR với cặp mồi đặc hiệu từ DNA tổng số Bacillus sp. 41

Tạo dòng gen Bgl

Sản phẩm PCR mang gen *Bgl* sau tinh sạch được gắn vào vector pGEM®-T Easy theo phương pháp TA-cloning dưới sự xúc tác của enzyme T4 Ligase và biến nạp vào *E. coli* TOP10. Các khuẩn lạc đơn màu trắng sau 16 giờ cấy trải trên đĩa LB có bổ sung Amp, IPTG và X-gal được sàng lọc bằng phương pháp PCR khuẩn lạc đơn với cặp mồi đặc hiệu nhằm kiểm tra sự có mặt của gen mã hoá có trong vector pGEM®-T Easy tái tổ hợp. Trình tự mồi M13 được thiết kế nằm ở hai đầu vùng tạo dòng của vector pGEM®-T Easy cho phép lựa chọn chính xác các khuẩn lạc có mang gen *Bgl.* Kết quả điện di gel agarose 1% được trình bày ở Hình 3.



Hình 3. Kết quả PCR với cặp mồi M13



M: Hypper Ladder 1kb (Bioline); 1: Sản phẩm PCR với cặp mồi M13 với khuẩn lạc đơn mang gen Bgl M: Hypper Ladder 1kb (Bioline); 1: vector pGEM®-T Easy tái tổ hợp; 2: vector pGEM®-T Easy tái tổ hợp được cắt bằng enzyme cắt hạn chế Kpnl

Phân tích trên gel cho thấy có một band DNA có kích thước khoảng 800 bp (bao gồm một đoạn khoảng 217 bp nằm trên vector pGEM®-T Easy và gen mã hoá 655 bp). Các khuẩn lạc dương tính được chuyển sang môi trường LB lỏng có bổ sung ampicillin (50 µg/mL) trong 16 giờ ở 37°C, sau đó tiến hành tách plasmid và gửi đi giải trình tự.

Kết quả trình bày ở Hình 4 cho thấy, có hai band ở vị trí 1700 bp và 2000 bp, sự sai khác này có thể do vector tái tổ hợp ở trạng thái siêu xoắn (do cấu trúc đóng vòng của plasmid tái tổ hợp). Khi được cắt bằng enzyme cắt hạn chế *Kpn*I (sự hiện diện của enzyme nằm trên trình tự nucleotide của gen *BgI*), thu được 1 band duy nhất ở vị trí 4000 bp (vector pGEM®-T Easy: 3015 bp, *BgI*: 655 bp). Chứng tỏ, gen *BgI* đã được tái tổ hợp thành công vào hệ thống vector tạo dòng pGEM®-T Easy.

Giải trình tự gen

Kết quả giải trình tự cho thấy gen *Bgl* phân lập được có độ tương đồng 99% so với gen mã hoá β -1,3-glucanase của *B. velezensis* BY6 đã được công bố trên GenBank (CP051011). Trình tự gen *Bgl* mã hoá protein cho chuỗi polypeptide có chiều dài 214 amino acid, khi so sánh với trình tự β -1,3-glucanase từ chủng *B. halotolerans* Y6 (MH643779) nhận thấy có 23 amino acid khác nhau ở các vị trí 17, 19, 24, 41, 63, 67, 72, 100, 104, 115, 121, 127, 129, 133, 134, 139, 161, 170, 172, 175, 178, 179, 183 (Hình 5). Sự sai khác này không làm ảnh hưởng đến chức năng của enzyme. Protein *Bgl* có trọng lượng phân tử là 26,87 kDa và pl là 5,28. Phân tích trình tự tín hiệu peptide bằng công cụ SignalP-5.0, cho kết quả *Bgl* chứa trình tự mã hóa cho chuỗi tín hiệu peptide gồm 29 amino acid.

MH643779.1 Bacillus	QTGGSFFEPFNSYNSGLWEKANGKSNGDMFNCTWRANNVSLTSSGEMRLALTSPSYNKFD QTGGSFFEPFNSYNSGFWQKANGYSNGDMFNCTWRANNVSVTSSGEMRLALTSPSYNKFD ************************************	60 60
MH643779.1 Bacillus	CGDNRSGQTYGEGLYEVRMKPAKNTGIVSSFFTYTGPTDRTPWHEIDIEFLGKDSTKVQF CGENRSAQTYGYGLYEVRMKPAKNTGIVSSFFTYTGPTDGTPWDEIDIEFLGKDTTKVQF	120 120
MH643779.1	**:***.*** ****************************	180
Bacillus	NYYTNGAGNHEKVADLGFDAANAYHTYAFDWQPNSIKWYVDGQLKHTATSQIPTTPGKIM :***** *:***:.****:********************	180
MH643779.1 Bacillus	MNFWNGIGVDDWLGSYNGVNPLYAHYDWVRYTKK 214 MNLWNGIGVDDWLGSYNGVNPLYAHYDWVRYTKK 214	

Hình 5. So sánh trình tự amino acid suy diễn của gen Bgl từ chủng Bacillus sp. 41

(.) Dấu chấm được sử dụng để thể hiện sự thay thế bảo thủ; (*) Dấu hoa thị biểu thị các vị trí được bảo toàn hoàn toàn trên tất cả các chuỗi; (:) Dấu hai chấm đại diện cho sự thay thế bán bảo tồn; () Khoảng trắng đại diện cho không bảo tồn

Qua phân tích cấu trúc protein tương đồng và nhận diện các vùng chức năng cho thấy Bgl có cấu trúc tương tự β -1,3-1,4 glucanase *B. subtilis* 168 với mã số PDB: 305S với độ tin cậy 100% (Hình 6). Kết quả cho thấy cấu trúc bậc 2 dự đoán của Bgl chiếm 55% vị trí gập cuốn phiến β (Furtado *et al.*, 2021).



Hình 6. Cấu trúc dự đoán của Bgl

KÉT LUẬN

Trong nghiên cứu này, gen *Bgl* đã được phân lập thành công từ DNA tổng số của chủng vi khuẩn *Bacillus* sp. 41 bằng cặp mồi *Bgl*. Gen mã hoá có chiều dài 655 bp, trình tự amino acid suy diễn có chiều dài 214 aa. Protein Bgl được dự đoán có cấu trúc và chức năng tương tự enzyme β -1,3-1,4 glucanase *B. subtilis* 168.

Lời cảm ơn: Nghiên cứu được tài trợ bởi Chương trình học bổng đào tạo thạc sĩ, tiến sĩ trong nước của Quỹ đổi mới sáng tạo Vingroup (VINIF), mã số VINIF.2022.TS056 và Đề tài cấp Bộ Giáo dục và Đào tạo, mã số CT2022.09.DHH.01.

TÀI LIỆU THAM KHẢO

Bhuiyan N, Selvaraj G, Wei Y, & King J (2009). Role of lignification in plant defense. Plant Signal Behav, 4(2): 158-159.

Duy L (2023). Evaluation of antifungal activity against *Corynespora cassiicola* by bacteria isolated from soil in the root zone of cucumber plants. *Biodiversitas*, 24(12): 6584-6591.

El-Katatny M, Gudelj M, Robra K, Elnaghy M, Gübitz G (2001). Characterization of a chitinase and an endo-β-1, 3-glucanase from *Trichoderma harzianum* Rifai T24 involved in control of the phytopathogen *Sclerotium rolfsii. Appl Microbiol Biotechnol*, 56: 137-143.

Furtado G, Ribeiro L, Santos C, Tonoli C, De Souza A, Oliveira R, Ward R (2011). Biochemical and structural characterization of a β -1, 3–1, 4-glucanase from *Bacillus subtilis* 168. *Process Biochem*, 46(5): 1202-1206.

Kelley L, Mezulis S, Yates C, Wass M, Sternberg M (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*, 10(6): 845-858.

Magnin-Robert M, Trotel-Aziz P, Quantinet D, Biagianti S, Aziz A (2007). Biological control of Botrytis cinerea by selected grapevine-associated bacteria and stimulation of chitinase and β -1, 3 glucanase activities under field conditions. *Eur J Plant Pathol*, 118: 43-57.

Masilamani R, Sharma OP, Muthuvel SK, Natarajan S (2013). Cloning, expression of b-1,3-1,4 glucanase from *Bacillus subtilis* SU40 and the effect of calcium ion on the stability of recombinant enzyme: *in vitro* and *in silico* analysis. *Bioinformation*, 6;9(19):958-62.

Nguyễn Thị Thanh Huyền, Nguyễn Đức Huy, Nguyễn Bảo Hưng, Lê Thị Nhật Anh, Nguyễn Thị Minh Nga, Trần Quốc Dung (2024). Khả năng đối kháng bệnh thán thư do *Colletotrichum fructicola* CL5 trên ớt của β-1,3-glucanase sinh tổng hợp từ *Bacillus* spp.. *Tạp chí Khoa học Đại học Huế: Nông nghiệp và Phát triển nông thôn*, 133(3A): 49-60.

Otsuka Y, Sato K, Yano S, Kanno H, Suyotha W, Konno H, Taira T (2022). GH-16 type β -1, 3-glucanase from *Lysobacter* sp. MK9-1 enhances antifungal activity of GH-19 type chitinase, and its glucan-binding domain binds to fungal cell-wall. *J Appl Glycosci*, 69(3): 49-56.

Sambrook S, Russell D (2003). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, NY.

Sinha M, Singh R, Kushwaha G, Iqbal N, Singh A, Kaushik S, Singh T (2014). Current overview of allergens of plant pathogenesis related protein families. *Sci World J*, 2014: 543195.

Tiku A (2020). Antimicrobial compounds (phytoanticipins and phytoalexins) and their role in plant defense. Co-evolution of secondary metabolites, Reference Series in Phytochemistry. Springer, Cham.

Trần Vũ Phến, Trần Ánh Lụa, Đinh Ngọc Trúc(2016). Khả năng kích kháng lưu dẫn của vi khuẩn *Bacillus* spp. đối với bệnh cháy lá lúa do nấm Pyricularia oryzae trong điều kiện nhà lưới. *Tạp chí Khoa học Trường Đại học Cần Thơ*, Số chuyên đề: Nông nghiệp, 3: 249-257.

Van Loon L, Van Kammen A (1970). Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana tabacum* var. 'Samsun'and 'Samsun NN': II. Changes in protein constitution after infection with tobacco mosaic virus. *Virology*, 40(2): 199-211.

Vogeli-Lange R, Frundt C, Hart C M, Nagy F, Meins F (1994). Developmental, hormonal, and pathogenesis-related regulation of the tobacco class I β -1, 3-glucanase B promoter. *Plant Mol Biol*, 25: 299-311.

Watanabe T, Kasahara N, Aida K, Tanaka H (1992). Three N-terminal domains of beta-1,3-glucanase A1 are involved in binding to insoluble beta-1,3-glucan. *J Bacteriol*, 174(1): 186–190.

Worrall D, Hird D, Hodge R, Paul W, Draper J, Scott R (1992). Premature dissolution of the microsporocyte callose wall causes male sterility in transgenic tobacco. *The Plant Cell*, 4(7): 759-771.

ISOLATION AND ANALYSIS OF AN β -1,3-GLUCANASE FROM Bacillus sp. 41

Nguyen Thi Thanh Huyen^{1,2}, Le Thi Nhat Anh¹, Tran Quoc Dung^{3*}, Nguyen Duc Huy^{1*}

¹Institute of Biotechnology, Hue University

²Le Quy Don High School, Dong Ha, Quang Tri

³University of Education, Hue University

SUMMARY

 β -1,3-glucanase is a potential enzyme involved in the defense system against harmful fungi in plants. In the present study, a gene encoding for β -1,3-glucanase (Bgl) isolated from Bacillus sp. 41 strain which inhibited harmful Aspergilus niger was succesfully amplified, cloned, performed nucleotide sequecing and analyzed the structure. The results indicated that the gene consisted of 655 bp, encoding for a polypeptide chain of 214 aa. The functional prediction of Bgl indicated the gene encodes for a signal peptide of 22 aa length demonstrated that Bgl encodes for an extracellular protein. The predicted structure of translated protein was highly similar to the β -1,3-1,4 glucanase in Bacillus subtilis, containing 55% amino acid chain of β -sheet. The properties of Bgl and secondary structure of translated protein propose Bgl is importance factor in Bacillus genome. Thus, the successful isolation of the Bgl gene provides important genetic data that supports for further study to elucidate the antagonistic mechanism between Bacillus sp. to harmful fungi. In addition, these results provide genetic sources for the expression of recombinant enzyme in appropriate hosts.

Keywords: β-1,3-glucanase, Bacillus, Bgl, cloning.

Author for correspondence: Tran Quoc Dung; Email: tranquocdung@hueuni.edu.vn or Nguyen Duc Huy; Email: ndhuy@hueuni.edu.vn

NGHIÊN CỨU XÁC ĐỊNH, PHÂN TÍCH ĐẶC TÍNH VÀ ĐÁNH GIÁ BIỀU HIỆN CỦA NHÓM GENE MÃ HÓA JASMONIC ACID CARBOXYL METHYLTRANSFERASE Ở CÂY SẮN (*Manihot esculenta*)

Phùng Trường Trinh¹, Lê Đức Chiến¹, Đào Thị Chúc¹, Nguyễn Quốc Trung², Tống Văn Hải², Lê Thị Ngọc Quỳnh³, Đồng Huy Giới², La Việt Hồng⁴, Chu Đức Hà^{1*}, Lê Huy Hàm¹

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

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

³Bộ môn Công nghệ Sinh học, Trường Đại học Thủy lợi

⁴Viện Nghiên cứu Khoa học và Ứng dụng, Trường Đại học Sư phạm Hà Nội 2

TÓM TẮT

Jasmonic acid carboxyl methyltransferase (JMT) là một trong những nhóm enzyme quan trọng tham gia vào các quá trình sinh học xảy ra ở cây trồng, đặc biệt liên quan đến đáp ứng stress sinh học. Nghiên cứu này được thực hiện nhằm làm rõ thông tin của nhóm JMT trên đối tượng cây sắn (*Manihot esculenta*), một trong những cây trồng quan trọng ở Việt Nam hiện nay. Theo đó, tổng số 23 thành viên của nhóm JMT đã được xác định trên dữ liệu của cây sắn. Bằng việc sử dụng các công cụ tin sinh học, hầu hết các thành viên của nhóm JMT đã được định danh thông tin về mã định danh. Kết quả phân tích trên công cụ Expasy cho thấy, nhóm JMT ở cây sắn có tính chất lý hóa đa dạng, tương tự như nhóm JMT ở các loài thực vật khác. Đáng chú ý, nghiên cứu này đã khai thác dữ liệu biểu hiện của các gene mã hóa nhóm JMT có mức độ biểu hiện thay đổi đa dạng ở mẫu lá lây nhiễm bệnh sọc nâu virus. Trong đó, tổng số 8 gene mã hóa các thành viên của nhóm JMT, bao gồm 4 gene tăng cường biểu hiện (fold-change $\geq 2,00$) và 4 gene kìm hãm biểu hiện (fold-change $\leq -2,00$) đã được xác định. Tóm lại, kết quả của nghiên cứu này đã cung cấp những dẫn luận quan trọng trong việc đề xuất các ứng viên phục vụ phân tích chức năng gene ở nhóm JMT trên cây sắn.

Từ khóa: Sắn, jasmonic acid carboxyl methyltransferase, sọc nâu virus, tin sinh học.

MỞ ĐẦU

Cây sắn (*Manihot esculenta*), còn gọi là khoai mì, là một loại cây trồng có nguồn gốc từ Nam Mỹ, đặc biệt là khu vực Brazil (Guira *et al.*, 2017). Từ lâu, cây sắn đã được du nhập và trồng rộng rãi ở nhiều vùng nhiệt đới và cận nhiệt đới trên thế giới, trong đó có Việt Nam (Malik *et al.*, 2020). Cây sắn được biết đến với khả năng chịu hạn tốt và dễ trồng trên nhiều loại đất khác nhau, kể cả đất nghèo dinh dưỡng (Okogbenin *et al.*, 2013). Về mặt dinh dưỡng, củ sắn là nguồn cung cấp tinh bột chính, chứa nhiều carbohydrate, cung cấp năng lượng quan trọng cho con người (Li *et al.*, 2017). Cây sắn đóng vai trò quan trọng trong đời sống kinh tế và xã hội của nhiều quốc gia. Ngoài việc làm thực phẩm cho con người, sắn còn được sử dụng làm thức ăn chăn nuôi và nguyên liệu cho ngành công nghiệp sản xuất bột ngọt, cồn sinh học và các sản phẩm hóa học khác. Nhờ vào những đặc điểm này, cây sắn đã trở thành một cây trồng đa dụng và quan trọng trong sản xuất nông nghiệp bền vững.

Trong quá trình sinh trưởng và phát triển của thực vật, jasmonic acid carboxyl methyltransferase (JMT) được ghi nhận là một enzyme quan trọng trong quá trình sinh tổng hợp các chất điều hòa sinh trưởng ở thực vật, đặc biệt là jasmonic acid và các dẫn xuất của jasmonic acid (Seo *et al.*, 2001). Theo đó, JMT chịu trách nhiệm methyl hóa jasmonic acid, tạo thành methyl jasmonate, một hợp chất có vai trò thiết yếu trong nhiều quá trình sinh lý và đáp ứng sinh học của thực vật (Seo *et al.*, 2001; Qi *et al.*, 2016). Cấu trúc của JMT bao gồm một chuỗi polypeptide có trung tâm hoạt động (Methyltransf_7, mã định danh Pfam: PF03492) gắn kết với S-adenosyl-L-methionine (Ogawa *et al.*, 2001), nguồn cung cấp nhóm methyl cho phản ứng methyl hóa (Seo *et al.*, 2001; Cheong, Choi, 2003). Enzyme này thực hiện việc chuyển nhóm methyl từ S-adenosyl-L-methionine đến jasmonic acid, tạo thành methyl jasmonate (Seo *et al.*, 2001; Cheong, Choi, 2003). Do vậy, thông qua sự có mặt của JMT, jasmonic acid và các dẫn xuất của jasmonic acid tham gia vào cơ chế phản ứng phòng vệ đối với các tác nhân gây bệnh như côn trùng, nấm và vi khuẩn, giúp tăng cường khả năng kháng bệnh của thực vật (Qi *et al.*, 2016). Gần đây, nhóm JMT đã được báo cáo trên một số đối tượng cây trồng, như lúa gạo (*Oryza sativa*) (Qi *et al.*, 2016), cà chua (Solanum lycopersicum) (Wei *et al.*, 2021), *Hedychium coronarium* (Yue *et al.*, 2024) và chè (*Camellia sinensis*) (Guo, Qiao, 2020). Tuy nhiên, chưa có thông tin được ghi nhận về nhóm JMT trên cây sắn, mặc dù dữ liệu của đối tượng cây trồng này đã được báo cáo gần đây (Bredeson *et al.*, 2016).
Nghiên cứu này được thực hiện nhằm xác định, phân tích đặc tính và đánh giá biểu hiện của nhóm JMT trên dữ liệu của cây sắn. Theo đó, các gene mã hóa cho nhóm JMT được sàng lọc và chú giải trong genome tham chiếu của cây sắn. Cuối cùng, mức độ biểu hiện của các gene mã hóa thành viên của nhóm JMT được phân tích dựa trên dữ liệu microarray.

DỮ LIỆU VÀ PHƯƠNG PHÁP

Dữ liệu nghiên cứu

Genome và proteome của giống sắn AM560-2 (Mã số NCBI RefSeq: GCF_001659605.2) (Prochnik *et al.*, 2012; Bredeson *et al.*, 2016) được khai thác trên NCBI và Phytozome (Goodstein *et al.*, 2012).

Dữ liệu microarray phân tích trên mẫu lá sắn lây nhiễm bệnh sọc nâu virus (Mã số GEO NCBI: GSE56467) (Maruthi *et al.*, 2014) được thu thập trên GEO NCBI.

Phương pháp nghiên cứu

Sàng lọc và định danh: Cấu trúc bảo tồn Methyltransf_7 (mã định danh Pfam: PF03492), đặc trưng cho JMT ở thực vật (Ogawa *et al.*, 2001) trên dữ liệu Pfam (Mistry *et al.*, 2021) được khai thác để truy vấn trên proteome của cây sắn (Bredeson *et al.*, 2016) nhằm xác định các protein có cấu trúc tương đồng theo mô tả trong nghiên cứu trước đây (Chu *et al.*, 2024). Theo đó, tất cả các trình tự protein (≥ 100 amino acid) được thu thập và sàng lọc với giá trị E-value ≤ 1e-10. Các ứng viên sau đó được đối chiếu (BlastP) trên genome của cây sắn (Bredeson *et al.*, 2016) nhằm chú giải và định danh thông tin của nhóm JMT. Trình tự protein, gene mã hóa được thu thập cho các phân tích *in silico* tiếp theo.

Phân tích tính chất lý hóa của protein: Trình tự protein đầy đủ của các JMT được truy vấn trên Expasy Protparam (Gasteiger *et al.*, 2003) dựa theo mô tả trong nghiên cứu trước đây (Chu *et al.*, 2024). Các đặc tính cơ bản của protein, như kích thước (gốc amino acid), trọng lượng phân tử (kDa), điểm đẳng điện, độ bất ổn định, chỉ số béo và độ ưa nước được khai thác. Trong đó, giá trị điểm đẳng điện > 7 (tính base), < 7 (tính acid), trong khi độ bất ổn định), < 40 (ổn định) và độ ưa nước > 0 (kị nước), < 0 (ưa nước) (Gasteiger *et al.*, 2003).

Khai thác dữ liệu biểu hiện gen: Mức độ biểu hiện của các gene mã hóa nhóm GATA được phân tích thông qua cơ sở dữ liệu GEO NCBI (Barrett *et al.*, 2013) dựa theo mô tả trong nghiên cứu trước đây (Chu *et al.*, 2024). Theo đó, dữ liệu RNA-Seq từ mẫu lá sắn bị nhiễm bệnh sọc nâu virus (Mã số GEO NCBI: GSE56467) (Maruthi *et al.*, 2014) được sử dụng. Mức độ biểu hiện của các gene mã hóa JMT được tính toán bằng giá trị fold-change, dựa trên tỷ lệ giữa giá trị RPKM (Reads Per Kilobase per Million mapped reads) của gene tương ứng ở mẫu lá nhiễm virus và mẫu lá đối chứng. Gene *PP2A* được sử dụng làm gene tham chiếu trong phân tích dữ liệu RNA-Seq như mô tả trong nghiên cứu trước đây (Moreno *et al.*, 2011).

Phân tích số liệu: Mức độ biểu hiện của gene được quy ước dựa theo giá trị fold-change như mô tả trong nghiên cứu trước đây (Chu *et al.*, 2024). Theo đó, gene được quy ước là có đáp ứng với stress với |fold-change| ≥ 2 (*p-value* < 0.05), với gene tăng cường biểu hiện (fold-change $\ge 2,00$) và gene kìm hãm biểu hiện (fold-change $\le -2,00$).

KẾT QUẢ VÀ THẢO LUẬN

Xác định và chú giải thông tin của nhóm JMT ở cây sắn

Để xác định nhóm JMT, cấu trúc bảo tồn Methyltransf_7 (Ogawa *et al.*, 2001) trên Pfam (Mistry *et al.*, 2021) được truy vấn trên dữ liệu của cây sắn (Bredeson *et al.*, 2016). Kết quả sàng lọc (E-value ≤ 1e-10) cho thấy, tổng số 23 thành viên của nhóm JMT đã được ghi nhận trên cây sắn. Theo đó, thông tin định danh của nhóm JMT, bao gồm mã gene, mã protein và mã locus được xác định trên dữ liệu di truyền của cây sắn (Bảng 1). Trong đó, 3 gene, bao gồm *Manes.01G141701, Manes.10G070800* và *Manes.15G135800* chưa được chú giải về mã protein và mã locus.

Gần đây, nhóm JMT đã được báo cáo trên một số đối tượng cây trồng chính. Cụ thể, tổng số 24 thành viên của nhóm JMT đã được xác định trên loài thực vật mô hình 2 lá mầm Arabidopsis. Trong khi đó, nhóm JMT ở cây lúa gạo đã được báo cáo gồm 41 thành viên (Qi *et al.*, 2016). Ở loài *Populus tremula*, nhóm JMT chứa 28 thành viên, trong khi nhóm JMT ở Salvia miltiorrhiza gồm 30 thành viên. Ở *loài H. coronarium*, tổng số 12 thành viên của nhóm JMT đã được ghi nhận (Yue *et al.*, 2024). Như vậy, số lượng thành viên của nhóm JMT ở cây sắn (23), ít hơn ở Arabidopsis (24), *P. tremula* (28), *S. miltiorrhiza* (30) và lúa gạo (41), và nhiều hơn so với *H. coronarium* (12). Các nghiên cứu này cho thấy số lượng thành viên của nhóm JMT ở các loài thực vật tương đối đa dạng.

STT	Tên gene	Mã gene	Mã protein	Mã locus				
1	Manes.01G114200	XP_021605389	XM_021749697	LOC110609869				
2	Manes.01G138400	XP_021622782	XM_021767090	LOC110622544				

Bảng 1. Định danh nhóm JMT ở cây sắn

CÔNG NGHỆ GEN

3	Manes.01G138500	XP_021621696	XM_021766004	LOC110621713
4	Manes.01G138600	XP_021624312	XM_021768620	LOC110623617
5	Manes.01G141701	OAY60821	-	-
6	Manes.02G096900	XP_021601892	XM_021746200	LOC110607126
7	Manes.02G107100	XP_021604376	XM_021748684	LOC110609241
8	Manes.03G136100	XP_021606188	XM_021750496	LOC110610534
9	Manes.04G074231	XP_043812210	XM_043956275	LOC110614162
10	Manes.05G156400	XP_021612254	XM_021756562	LOC110614876
11	Manes.06G008900	XP_021617130	XM_021761438	LOC110618288
12	Manes.06G009000	XP_021616954	XM_021761262	LOC110618177
13	Manes.10G070800	OAY39144	-	-
14	Manes.10G070900	XP_021624991	XM_021769299	LOC110624182
15	Manes.13G061650	XP_043805338	XM_043949403	LOC110629170
16	Manes.15G066100	XP_021594262	XM_021738570	OC110601438
17	Manes.15G111600	XP_021595968	XM_021740276	LOC110602697
18	Manes.15G135700	XP_021594332	XM_021738640	LOC110601490
19	Manes.15G135800	KAG8637562	-	-
20	Manes.15G135900	XP_021594238	XM_021738546	LOC110601425
21	Manes.17G084900	XP_021598742	XM_021743050	LOC110604759
22	Manes.17G085100	XP_043808378	XM_043952443	LOC110604925
23	Manes.18G145282	XP_043808821	XM_043952886	LOC110610538

Phân tích đặc tính lý hóa của nhóm JMT ở cây sắn

Để khai thác tính chất của nhóm JMT ở cây sắn, trình tự protein được phân tích trên Expasy Protparam (Gasteiger *et al.*, 2003) để tính toán kích thước, trọng lượng phân tử, điểm đẳng điện, độ bất ổn định, chỉ số béo và độ ưa nước. Kết quả được mô tả ở Bảng 2.

Kết quả phân tích cho thấy kích thước của nhóm JMT ở cây sắn dao động từ 332 (Manes.10G070800) - 393 (Manes.15G111600) gốc amino acid, tương ứng với trọng lượng phân tử đạt từ 37,56 - 44,95 kDa. Phân tích giá trị điểm đẳng điện cho thấy nhóm JMT ở cây sắn đều có tính acid (< 7), dao động từ 5,06 (Manes.15G135800) - 6,43 (Manes.17G085100). Trong khi đó, 2 (trên tổng số 23) thành viên của nhóm JMT ở cây sắn, bao gồm Manes.01G114200 và Manes.15G135900, thể hiện tính ổn định (độ bất ổn định < 40), trong khi 21 (trên tổng số 23) thành viên còn lại của nhóm JMT ở cây sắn không có tính ổn định (độ bất ổn định > 40). Chỉ số béo của các thành viên trong nhóm JMT ở cây sắn dao động từ 75,47 (Manes.17G085100) - 99,21 (Manes.01G114200). Bên cạnh đó, phân tích độ ưa nước của các thành viên trong nhóm JMT ở cây sắn dao động từ 75,47 (Manes.01G5100) - 99,21 (Manes.01G114200). Bên cạnh đó, phân tích độ ưa nước của các thành viên trong nhóm JMT ở cây sắn cho thấy các phân tử này đều có tính ưa nước, với giá trị độ ưa nước < 0, đạt từ -0,35 (Manes.06G009000) - -0,06 (Manes.01G141701).

Bảng 2	. Đặc tín	h lý hóa	i của nhóm	JMT	ở cây sắn
--------	-----------	----------	------------	-----	-----------

STT	Tên protein	Kích thước	Trọng lượng	Điểm đẳng điện	Độ bất ổn định	Chỉ số béo	Độ ưa nước
1	Manes.01G114200	392	44,24	5,86	36,44	99,21	-0,11
2	Manes.01G138400	359	40,44	5,55	53,17	86,63	-0,16
3	Manes.01G138500	367	41,59	5,48	48,16	85,69	-0,15
4	Manes.01G138600	374	42,35	5,50	45,58	81,55	-0,17
5	Manes.01G141701	355	39,96	5,80	52,27	91,75	-0,06
6	Manes.02G096900	366	41,33	5,57	44,35	82,81	-0,19
7	Manes.02G107100	372	41,34	5,27	42,16	77,07	-0,24
8	Manes.03G136100	382	42,30	6,07	51,48	82,75	-0,23

HÔI NGHI KHOA	HOC TOÀN QI	UỐC VỀ CÔNG	NGHÊ SIN	HHOC 2024
	1190 10/11/44	000 12 00110		

9	Manes.04G074231	361	41,04	5,77	50,84	86,76	-0,22
10	Manes.05G156400	365	40,76	5,21	54,21	80,08	-0,22
11	Manes.06G008900	356	39,69	5,21	45,38	79,27	-0,17
12	Manes.06G009000	359	40,41	5,80	43,66	80,36	-0,35
13	Manes.10G070800	332	37,56	5,14	41,23	79,28	-0,29
14	Manes.10G070900	372	41,83	5,65	42,52	86,24	-0,20
15	Manes.13G061650	364	41,19	5,57	44,93	92,12	-0,19
16	Manes.15G066100	385	42,31	5,69	48,40	87,14	-0,13
17	Manes.15G111600	393	44,02	6,10	45,03	79,75	-0,33
18	Manes.15G135700	364	40,48	5,97	50,16	82,06	-0,20
19	Manes.15G135800	392	44,95	5,06	48,93	91,53	-0,08
20	Manes.15G135900	359	40,74	5,63	38,51	82,81	-0,19
21	Manes.17G084900	353	39,76	5,13	41,15	86,97	-0,07
22	Manes.17G085100	380	42,74	6,43	50,20	75,47	-0,23
23	Manes.18G145282	366	41,21	5,17	50,35	82,92	-0,21

Ghi chú: Kích thước (gốc amino acid); Trọng lượng (kDa).

Trong các nghiên cứu trước đây, đặc tính của nhóm JMT ở một số đối tượng thực vật cũng đã được ghi nhận. Ví dụ, nhóm JMT ở loài *H. coronarium* có kích thước từ 366 - 387 gốc amino acid, tương ứng với trọng lượng phân tử đạt từ 38197,02 - 43220,25 Da (Yue *et al.*, 2024). Các thành viên của nhóm JMT ở *H. coronarium* cũng thể hiện tính acid, gửi giá trị điểm đẳng điện đạt từ 4,96 - 6,11 (Yue *et al.*, 2024). Tương tự, kích thước và trọng lượng phân tử của nhóm JMT ở loài *Neolamarckia cadamba* được ghi nhận từ 134 - 726 gốc amino acid và 15,20 - 82,60 kDa. Giá trị điểm đẳng điện của nhóm JMT ở *N. cadamba* đạt từ 4,71 (acid) - 8,58 (base). Ở cà chua, nhóm JMT có kích thước từ 127 - 410 gốc amino acid, trong khi trọng lượng phân tử được báo cáo có giá trị từ 14,37 - 45,38 kDa (Wei *et al.*, 2021). Các protein này có điểm đẳng điện từ 4,80 (tính acid) - 7,69 (tính base) (Wei *et al.*, 2021). Các kết quả này cho thấy, nhóm JMT ở các loài thực vật có tính chất lý hóa, về kích thước, trọng lượng phân tử, điểm đẳng điện, chỉ số béo, độ bất ổn định và độ ưa nước, tượng đối đa dạng.

Đánh giá biểu hiện của nhóm JMT ở mẫu lá lây nhiễm bệnh sọc nâu virus ở cây sắn

Sọc nâu virus là một trong những bệnh nguy hiểm trên cây sắn (Tomlinson *et al.*, 2018). Các nghiên cứu trước đây đã ghi nhận về vai trò của nhóm JMT trong đáp ứng stress sinh học ở một số đối tượng cây trồng (Qi *et al.*, 2016). Theo đó, nghiên cứu này đã tiến hành tái phân tích dữ liệu RNA-Seq mô tả mức độ biểu hiện của các gene trong điều kiện lây nhiễm nhân tạo bệnh sọc nâu virus trên lá của cây sắn (Maruthi *et al.*, 2014). Kết quả phân tích giá trị fold-change của các gene mã hóa nhóm JMT được mô tả ở Hình 1.



Hình 1. Biểu hiện của các gene mã hóa nhóm JMT ở các mẫu lá lây nhiễm nhân tạo bệnh sọc nâu virus trên cây sắn

Nhìn chung, các gene mã hóa nhóm JMT có sự thay đổi mức đô phiên mã đa dang khi tiến hành lây nhiễm bênh sọc nâu virus trên mẫu lá của cây sắn. Cụ thể, 15 (trên tổng số 23) gene, bao gồm Manes.01G114200, Manes.01G138400. Manes.01G138500. Manes.01G141701. Manes.02G096900. Manes.03G136100. Manes.04G074231. Manes.05G156400. Manes.10G070800, Manes.10G070900, Manes.13G061650. Manes.15G066100, Manes.15G135800, Manes.15G135900 và Manes.18G145282, không thể hiện sự đáp ứng có khác biệt (|fold-change| < 2) ở mẫu lá khi lây nhiễm bệnh sọc nâu virus. Đáng chú ý, nghiên cứu đã chỉ ra 4 (trên tổng số 23) gene, bao gồm Manes.01G138600, Manes.02G107100, Manes.06G009000 và Manes. 15G111600 kìm hãm biểu hiện (fold-change < -2,00) ở mẫu lá lây nhiễm bệnh, với giá trị fold-change đạt lần lượt là -5,59; -5,94; -2,24 và -6,71-fold. Trong khi đó, 4 (trên tổng số 23) gene, bao gồm Manes.06G008900, Manes.15G135700, Manes.17G084900 và Manes.17G085100, thể hiện mức độ phiên mã được tăng cường ở mẫu lá lây nhiễm bênh soc nâu virus, với giá tri fold-change đat lần lượt là 2,38; 5,36; 3,13 và 7,60-fold.

KÉT LUẬN

Nghiên cứu này đã xác định và chú giải thông tin của nhóm jasmonic acid carboxyl methyltransferase (JMT) trong cây sắn, phát hiện 23 thành viên của nhóm này, bao gồm thông tin về mã gene, mã protein và mã locus, với ba gene chưa được chú giải hoàn toàn. Phân tích đặc tính lý hóa của nhóm JMT cho thấy các protein có kích thước từ 332 đến 393 gốc amino acid, trọng lượng phân tử từ 37,56 đến 44,95 kDa, điểm đẳng điện từ 5,06 đến 6,43, và tính ưa nước. Phân tích dữ liệu RNA-Seq cho thấy mức độ biểu hiện của các gene mã hóa nhóm JMT thay đổi đa dạng khi cây sắn bị nhiễm bệnh sọc nâu virus, với 4 gene bị kìm hãm và 4 gene khác tăng cường phiên mã. Điều này đề xuất cơ chế điều hòa âm tính và điều hòa dương tính của nhóm JMT trong đáp ứng sinh học của cây sắn đối với bệnh sọc nâu virus, cung cấp hiểu biết sâu hơn về cơ chế phản ứng của cây trước stress sinh học.

TÀI LIỆU THAM KHẢO

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2013) NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res* 41: (Database issue) D991-995.

Bredeson JV, Lyons JB, Prochnik SE, Wu GA, Ha CM, Edsinger-Gonzales E, Grimwood J, Schmutz J, Rabbi IY, Egesi C, Nauluvula P, Lebot V, Ndunguru J, Mkamilo G, Bart RS, Setter TL, Gleadow RM, Kulakow P, Ferguson ME, Rounsley S, Rokhsar DS (2016) Sequencing wild and cultivated cassava and related species reveals extensive interspecific hybridization and genetic diversity. *Nat Biotechnol* 34: (5) 562-570.

Cheong JJ, Choi YD (2003) Methyl jasmonate as a vital substance in plants. Trends Genet 19: (7) 409-413.

Chu HD, Yen TTH, Pham CT, Quynh LTN, Huyen TTT, Trung NQ, Gioi DH, Tien TV, Cao PB (2024) Identification of Two Enzymes for Trehalose Synthesis and Their Potential Function in Growth and Development in Peanut (Arachis hypogaea). *Journal of Tropical Life Science* 14: (1) 83-94.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31: (13) 3784-3788.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res* 40: (Database issue) D1178-D1186.

Guira F, Some K, Kabore D, Sawadogo-Lingani H, Traore Y, Savadogo A (2017) Origins, production, and utilization of cassava in Burkina Faso, a contribution of a neglected crop to household food security. *Food Sci Nutr* 5: (3) 415-423.

Guo Y, Qiao D (2020) Genome-wide identification and expression analysis of SABATH methyltransferases in tea plant (Camellia sinensis): insights into their roles in plant defense responses. 15: (10) 1804684.

Li S, Cui Y, Zhou Y, Luo Z, Liu J, Zhao M (2017) The industrial applications of cassava: current status, opportunities and prospects. J Sci Food Agric 97: (8) 2282-2290.

Malik AI, Kongsil P, Nguyen VA, Ou W, Sholihin, Srean P, Sheela MN, Becerra Lopez-Lavalle LA, Utsumi Y, Lu C, Kittipadakul P, Nguyen HH, Ceballos H, Nguyen TH, Selvaraj Gomez M, Aiemnaka P, Labarta R, Chen S, Amawan S, Sok S, Youabee L, Seki M, Tokunaga H, Wang W, Li K, Nguyen HA, Nguyen VD, Ham LH, Ishitani M (2020) Cassava breeding and agronomy in Asia: 50 years of history and future directions. *Breed Sci* 70: (2) 145-166.

Maruthi MN, Bouvaine S, Tufan HA, Mohammed IU, Hillocks RJ (2014) Transcriptional response of virus-infected cassava and identification of putative sources of resistance for cassava brown streak disease. *PloS One* 9: (5) e96642.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res* 49: (D1) D412-D419.

Moreno I, Gruissem W, Vanderschuren H (2011) Reference genes for reliable potyvirus quantitation in cassava and analysis of Cassava brown streak virus load in host varieties. *J Virol Methods* 177: (1) 49-54.

Ogawa M, Herai Y, Koizumi N, Kusano T, Sano H (2001) 7-Methylxanthine Methyltransferase of Coffee Plants: GENE ISOLATION AND ENZYMATIC PROPERTIES *. *Journal of Biological Chemistry* 276: (11) 8213-8218.

Okogbenin E, Setter TL, Ferguson M, Mutegi R, Ceballos H, Olasanmi B, Fregene M (2013) Phenotypic approaches to drought in cassava: review. *Front Physiol* 4: 93.

Prochnik S, Marri PR, Desany B, Rabinowicz PD, Kodira C, Mohiuddin M, Rodriguez F, Fauquet C, Tohme J, Harkins T, Rokhsar DS, Rounsley S (2012) The Cassava Genome: Current Progress, Future Directions. *Trop Plant Biol* 5: (1) 88-94.

Qi J, Li J, Han X, Li R, Wu J, Yu H, Hu L, Xiao Y, Lu J, Lou Y (2016) Jasmonic acid carboxyl methyltransferase regulates development and herbivory-induced defense response in rice. *J Integr Plant Biol* 58: (6) 564-576.

Seo HS, Song JT, Cheong JJ, Lee YH, Lee YW, Hwang I, Lee JS, Choi YD (2001) Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci U S A* 98: (8) 4788-4793.

Tomlinson KR, Bailey AM, Alicai T, Seal S, Foster GD (2018) Cassava brown streak disease: historical timeline, current knowledge and future prospects. *Mol Plant Pathol* 19: (5) 1282-1294.

Wei X, Tao K, Zhang J, Lu S, Chen S, Liao J (2021) Identification of SABATH Family Members in Solanum lycopersicum and Their Expression Patterns Under Abiotic/Biotic Stresses. *Plant Molecular Biology Reporter* 39: (2) 403-418.

Yue Y, Zhang X, Wang L, He J, Yang S, Li X, Yu Y, Yu R, Fan Y (2024) Identification and Characterization of Jasmonic Acid Methyltransferase Involved in the Formation of Floral Methyl Jasmonate in Hedychium coronarium. Vol. 13 ed.^eds.),

IDENTIFICATION, CHARACTERIZATION AND EXPRESSION PROFILES OF GENES ENCODING JASMONIC ACID CARBOXYL METHYLTRANSFERASE IN CASSAVA (*Manihot esculenta*)

Phung Truong Trinh¹, Le Duc Chien¹, Dao Thi Chuc¹, Nguyen Quoc Trung², Tong Van Hai², Le Thi Ngoc Quynh³, Dong Huy Gioi², La Viet Hong⁴, Chu Duc Ha^{1*}, Le Huy Ham¹

¹Faculty of Agricultural Technology, University of Engineering and Technology, Vietnam National University, Hanoi

²Faculty of Biotechnology, Vietnam National University of Agriculture

³Department of Biotechnology, Thuyloi University

⁴Institute of Scientific Research and Application, Hanoi Pedagogical University 2

SUMMARY

Jasmonic acid carboxyl methyltransferase (JMT) is one of the important enzyme groups involved in biological processes in plants, particularly related to the response to biotic stress. This study was conducted to clarify the information of the JMT group in cassava (*Manihot esculenta*), one of the important crops in Vietnam today. Accordingly, a total of 23 members of the JMT group were identified in cassava data. Using bioinformatics tools, most members of the JMT group have been annotated with annotated identifiers. Our results from the Expasy tool showed that the JMT family in cassava has diverse physicochemical properties, similar to the JMT group in other plant species. Notably, this study exploited the expression data of genes encoding the JMT family under artificial infection conditions of cassava brown streak virus on cassava leaf samples. The results showed that the genes encoding the JMT group exhibited diverse expression levels in the cassava leaf samples infected with the brown streak virus. Among them, a total of 8 genes encoding the JMT group members, including 4 genes with induced expression (fold-change ≥ 2.00) and 4 genes with reduced expression (fold-change ≤ -2.00), were surveyed. Taken together, the results of this study have provided important insights for proposing candidates for functional gene analysis in the JMT group in cassava.

Keywords: Cassava, jasmonic acid carboxyl methyltransferase, cassava brown strike virus, bioinformatics.

Author for correspondence: Tel: 0983766070; Email: cd.ha@vnu.edu.vn

PHÁT HIỆN VÀ ĐỊNH LƯỢNG Tomato necrotic ringspot virus GÂY HẠI TRÊN CÀ CHUA (Solanum lycopersicum) DỰA TRÊN TRÌNH TỰ GEN NUCLEOCAPSID PROTEIN

Huỳnh Nguyễn Minh Nghĩa, Nguyễn Thị Kim Thoa, Nguyễn Xuân Dũng*

Trung tâm Công nghệ sinh học Thành phố Hồ Chí Minh

TÓM TẮT

Tomato necrotic ringspot virus (TNRV) là virus gây ra bệnh đốm vòng trên cà chua mới được phát hiện gần đây. Các nghiên cứu liên quan đến việc phát hiện virus này hiện vẫn còn hạn chế về số lượng và chủ yếu tập trung ở mức độ định tính. Bài báo này trình bày kết quả phát hiện và định lượng *TNRV* bằng kỹ thuật real-time RT-PCR. Trong đó, trình tự gen nucleocapsid protein của virus được sử dụng cho thiết kế mồi và thiết lập phản ứng real-time RT-PCR. Kết quả cho thấy phản ứng real-time RT-PCR có khả năng khuếch đại và định lượng virus thông qua phân đoạn gen có kích thước 101 bp và tương đồng 98% với trình tự gen đã được công bố của *TNRV*. Phản ứng có giới hạn định lượng đạt mức 1000 bản sao/µL dựa trên đường chuẩn (y = -3,22x + 38,87) có hệ số tương quan $R^2 = 0,99$. Kết quả áp dụng phản ứng đã thiết lập cho việc phát hiện và định lượng virus TNRV trong 10 mẫu cà chua thu thập từ đồng ruộng đã xác định được 02 mẫu có sự hiện diện của của virus với giá trị định lượng cao nhất đạt mức 2,48 × 10⁸ số bản sao/µL. Kết quả này cho thấy phản ứng real-time RT-PCR được thiết lập trong nghiên cứu này có thể được áp dụng vào thực tế để hỗ trợ cho công tác quản lý virus gây bệnh trên cà chua.

Từ khóa: Bệnh virus, cà chua, nucleocapsid protein, Tomato necrotic ringspot virus, real-time PCR.

MỞ ĐẦU

Cà chua (Solanum lycopersicum L.) là một loại cây trồng có giá trị kinh tế đang được trồng phổ biến tại Việt Nam. Diện tích trồng cà chua hằng năm đạt 15.000-17.000 ha với năng suất ước tính khoảng 15 - 17 tấn/ha và hơn 30 tấn/ha đối với một số vùng chuyên canh (Tuan, Mao, 2015). Tuy nhiên, cà chua lại khá nhạy cảm với các loại bệnh do nấm, vi khuẩn và virus gây ra (Jolly, 2016). *Tomato necrotic ringspot virus* (TNRV), là virus thuộc nhóm *Tospovirus*, gây ra bệnh đốm vòng trên cây cà chua, được phát hiện lần đầu tiên tại Thái Lan (Seepiban et al., 2011; Chiemsombat *et al.*, 2010). TNRV được phát hiện tại Việt Nam vào năm 2021 (Hoàng Việt Hậu *et al.*, 2021) và hiện đang tiếp tục gây hại trên nhiều khu vực trồng cà chua. Phương pháp phát hiện TNRV đang được sử dụng hiện nay chủ yếu là định tính thông qua ELISA, RT-PCR hoặc ELISA-RT-PCR (Channarong Seepiban *et al.*, 2011; Hoàng Việt Hậu *et al.*, 2021; Saengsoon Charoenvilaisiri *et al.*, 2014), do đó không thể xác định được lượng virus có trong mẫu. Trái lại, real-time RT-PCR là phương pháp có thể đồng thời phát hiện và định lượng virus với độ nhạy cao (Jeong *et al.*, 2014) và đã được sử dụng trong nghiên cứu kiểm soát virus gây bệnh trên cây trồng (Nageswara-Rao *et al.*, 2013). Việc áp dụng real-time RT-PCR để phát hiện và định lượng *TNRV* sẽ giúp nâng cao hiệu quả kiểm soát virus này. Nghiên cứu này được thực hiện với mục đích thiết lập quy trình phát hiện và định lượng *TNRV* dựa trên trình tự gen nucleocapsid protein bằng phương pháp real-time RT-PCR.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Thu thập nguồn mẫu nhiễm virus

Mẫu lá được thu thập dựa trên các triệu chứng đặc trưng của bệnh đốm vòng cà chua (lá xoăn, nhăn nheo, có đốm xanh đậm đen xen với xanh vàng, mặt dưới lá có các đốm đen hoại tử) (Seepiban *et al.*, 2011; Chiemsombat *et al.*, 2010) tại các vườn trồng cà chua ở Lâm Đồng. Sau khi thu thập, mẫu được rửa sạch với nước, bảo quản ở nhiệt độ -80°C trong tủ âm sâu (Sanyo, Nhật Bản).

Chủng vi khuẩn NEB 10-beta Competent *E. coli* được sử dụng cho thí nghiệm nhân dòng gen (kèm theo bộ Kit NEB PCR Cloning).

Thiết kế mồi phát hiện virus

Mồi phát hiện virus được thiết kế dựa trên trình tự gen của virus được công bố trên ngân hàng gen GenBank (FJ946835, HM113532, FN995637, FN806775, FJ947153, KY563368) sau khi tiến hành so sánh để chọn lọc vùng bảo tồn. Sau đó, khả năng bắt cặp của mồi với các trình tự gen của các virus khác được kiểm tra bằng phần mềm Blast NCBI.

Ly trích RNA virus và khuếch đại gen

RNA được ly trích từ mẫu lá cà chua (0,1 g) nhiễm virus bằng bộ Kit GeneJET Plant RNA purification mini (Thermo Fisher Scientific, Hoa Kỳ) theo hướng dẫn của nhà sản xuất. RNA thu được sau ly trích sẽ được pha loãng đến nồng độ 100 ng/μL để sử dụng cho khuếch đại trình tự gen nucleocapsis protein. Phản ứng RT-PCR khuếch đại gen (25 μ) gồm có 1X Dreamtaq Green PCR Master Mix, 2 U/μL RevertAid Reverse Transcriptase, 0,4 U/μL RiboLock RNase Inhibitor, 0,4 μM mỗi loại mồi xuôi và mồi ngược, và 4 ng/μL RNA mẫu. Chương trình RT-PCR bao gồm 1 chu kỳ ở 50°C/15 phút; 1 chu kỳ ở 95°C/2 phút; 40 chu kỳ ở 95°C/30 giây, Ta/30 giây, 72°C/60 giây và 1 chu kỳ ở 72°C/5 phút, với T_a là nhiệt độ bắt cặp mồi được thiết lập thấp hơn nhiệt độ nóng chảy 3°C. Sản phẩm PCR được điện di trên gel agarose (1,5%) trong 30 phút/100V, sau đó nhuộm với Ethidium Bromide (0,4 μg/mL) trong 30 phút và quan sát trên máy chụp gel (Geldoc, Biorad, Hoa Kỳ).

Dòng hóa và kiểm tra trình tự gen

Phân đoạn gen mục tiêu được gắn vào vector pMiniTTm 2.0 (sử dụng Kit NEB[®] PCR Cloning), sau đó biến nạp vào vi khuẩn NEB 10-beta Competent *E. coli* bằng phương pháp sốc nhiệt. Vi khuẩn sau biến nạp được sàng lọc trên môi trường LB rắn bổ sung 100 mg/L ampicillin, ở 37[°]C trong 24 giờ trước khi kiểm tra bằng phản ứng PCR khuẩn lạc (sử dụng mồi Cloning Analysis Primer, F: 5'-ACCTGCCAACCAAAGCGAGAAC-3', R: 5'-TCAGGGTTATTGTCTCATGAGCG-3'). Các dòng vi khuẩn thu nhận sau biến nạp được tách plasmid (sử dụng Kit DNA-SpinTM Plasmid DNA Purification, iNtRON, Hàn Quốc) và gửi mẫu giải trình tự.

Chuẩn bị mẫu DNA plasmid

Plasmid mang đoạn gen mục tiêu được sử dụng làm mẫu chuẩn xác định hàm lượng DNA và tính số lượng bản sao gen theo công thức (Staroscik, 2018):

Số bản sao gen =
$$\frac{Khối \, lượng \, DNA \times 6,022 \times 10^{23}}{Chiều \, dài \, đoạn \, gen \times 1 \times 10^9 \times 650}$$

Trong đó: Đơn vị khối lượng DNA: ng, chiều dài đoạn gen: bp.

DNA plasmid sau khi xác định số bản sao sẽ được pha loãng thành các mẫu có nồng độ từ 10⁶ đến 10⁰ bản sao/µL theo hệ số bậc 10.

Thiết lập phản ứng và dựng đường chuẩn cho real-time PCR

Phản ứng real-time RT-PCR khuếch đại trình tự gen mục tiêu trong mẫu RNA được thiết lập với thể tích 25 μL, bao gồm 1X Maxiam SYBR Green/Rox qPCR Master Mix, 2 U/μL RevertAid Reverse Transcriptase, 4 U/μL RiboLock RNase Inhibitor, 0,4 μM mỗi loại mồi xuôi và mồi ngược và 0,4 ng/μL RNA mẫu. Phản ứng được thực hiện với 40 chu kỳ khuếch đại (tổng hợp cDNA: [50°C/15 phút]; khuếch đại cDNA: [95°C/2 phút]; 40 x [95°C/30 giây, 54°C/30 giây, 72°C/60 giây]; [95°C/10 giây, 65°C 1 phút, 97°C 1 giây]; [37°C 30 giây]) và phân tích kết quả trên máy real-time PCR (LightCycler[®] 96, Roche).

Đường chuẩn được xây dựng dựa trên giá trị biến thiên chu kỳ ngưỡng (∆Ct) và log (bản sao/µL) tương ứng của mẫu thu được từ phản ứng real-time PCR (có thành phần và chương trình nhiệt tương tự phần khuếch đại cDNA của phản ứng real-time RT-PCR) khuếch đại mẫu các mẫu DNA plasmid có nồng độ từ 10¹ đến 10⁶ bản sao/µL.

Tạo mẫu chuẩn RNA

Mẫu RNA ly trích từ lá cà chua nhiễm bệnh (được xác định trước đó bằng RT-PCR) được định lượng bằng phản ứng real-time RT-PCR để xác định nồng độ virus có trong mẫu, sau đó pha loãng thành các mẫu có nồng độ từ 10⁵ đến 10⁰ bản sao/µL theo hệ số bậc 10.

Xác định ngưỡng phát hiện (LOD), giới hạn định lượng (LOQ) và xác định khả năng định lượng thực tế của phản ứng real-time RT-PCR

LOD được xác định bằng cách lặp lại 3 lần phản ứng real-time RT-PCR trên các mẫu RNA có nồng độ từ 10⁰ đến 10⁵ bản sao/µL. Xác định nồng độ thấp nhất mà phản ứng có thể phát hiện được và lặp lại 10 lần phản ứng real-time RT-PCR cho trường hợp này. LOD sẽ được xác lập nếu tất cả 10 lần lặp lại đều cho kết quả dương tính. Nếu số lần cho kết quả dương tính nhỏ hơn 10, thực hiện lại việc xác định LOD với mẫu có nồng độ cao hơn liền kề.

LOQ được xác định trên mẫu có nồng độ biết trước (bao gồm một mẫu ở mức LOD và hai mẫu cao hơn LOD) bằng cách lặp lại 10 lần phản ứng real-time RT-PCR định lượng. Dựa vào kết quả định lượng, tính độ lệch chuẩn (SD) và độ lệch chuẩn tương đối (RSD) theo công thức (European Network of GMO laboratories, 2011):

$$SD = \sqrt{\frac{\sum Xi^2 - \frac{(\sum Xi)^2}{n}}{n-1}}$$
 và RSD $= \frac{SD}{\overline{X}} \times 100$

Với $\sum Xi^2$: tổng bình phương các giá trị đo được tại một giá trị số bản sao cho trước; $\sum Xi$: tổng các giá trị đo được tại một nồng độ; n: số lần lặp lại tại một nồng độ; X: giá trị trung bình các lần đo được tại một nồng độ: $\overline{X} = \frac{\sum Xi}{n}$.

Nồng độ mẫu thấp nhất mà tại đó phản ứng cho kết quả dương tính và có giá trị RSD ≤ 25% được xác định là LOQ.

Khả năng định lượng thực tế của phản ứng real-time RT-PCR được đánh giá thông qua việc định lượng virus TRNV trên mẫu thu thập.

KẾT QUẢ VÀ THẢO LUẬN

Thiết kế mồi

Kết quả so sánh các trình tự gen nucleocapsis protein đã được công bố của virus *TNRV* cho thấy vùng trình tự từ vị trí nucleotide 258 đến 341 (theo trình tự FN995637) có sự tương đồng giữa các trình tự lớn hơn so với các vùng còn lại (Hình 1). Do đó, vùng trình tự này đã được chọn để thiết kế mồi.

FJ946835.1	CACTTTCACTATCTGATATAAATGATCCTGTAATGTAAT	240
HM113532.1	CTACAGTCCTATACTTCCT	19
FN995637.1		27
FN806775.1		27
FJ947153.1		27
KY563368.1		27
E1946835 1		300
HM113532.1	CTCCTAGTTGGCTTGCTAGTACTCACACTATTGCTCCCAAAATCATCAATGACTGGCTTG	79
FN995637.1		51
FN806775.1	AACGAGAAAATCAAGGAATTGCTA-	51
FJ947153.1	AACGAGAAAATCAAGGAATTGCTA-	51
KY563368.1	AACGAGAAAATCAAGGAATTGCTA- * * * ***** ** ***	51
F3946835.1	GATCCAAATGCTTTCTCAAGATGAGCAATCTGTTCTGAGTATTTCTTCAAGGATATTGAG	360
HM113532.1	GATCCAAATGCTTTCTCAAGATGAGCAATCTGTTCTGAGTACTTCTTCAAGGATATTGAG	139
FN995637.1	CATTGAAATAGAGGAGAAGCAGACATTGAAATAGAGCTTGAT	87
FN806775.1	CATTGAAATAGAGGAGAAGCAGACATTGAAATAGAGCTTGAT	87
FJ947153.1	CATTGAAATAGAGCTTGAT	87
KY563368.1	CATTGAAATAGAGCTCGAT	87
FJ946835.1	CTAGCTGTTCCAGGTGAGCTAGTCTTCAAAATAGCTGTTGCAGATTCGAACAATTTCTTC	420
HM113532.1	CTAGCTGTTCCAGGTGAGCTAGTCTTCAAAATAGCTGTTGCAGATTCGAACAATTTCTTC	199
FN995637.1	GAAGCTACACCAGGGTTCAGCTTTTCTAAATTCTATGATGAAAAACAAAC	139
FN806775.1	GAAGCTACACCAGGATTCAGCTTTTCTAAATTCTATGATGAAAACGAACAGG	139
FJ947153.1	GAAGCTACACCAGGATTCAGCTTTTCTAAATTCTATGATGAAAAACAAAC	139
KY563368.1	GAAGCTACACCAGGATTCAGCTTTTCTAAATTCTATGATGAAAACAAAC	139
	**** ***** * * *** * ** ** *	
FJ946835.1	AAATCACCGGTAAATTCAACTTGAGCAGCATCACACACT	459
HM113532.1	AAATCACCGGTAAATTCAACTTGAGCAGCATCACACACT	238
FN995637.1	ACATCTTCGGAAAATTTACATTCAACAATGGAATAACTATACTAAAGAGTAGAAAGCAGA	199
FN806775.1	ACATCTTCGGAAAATTTACATTCAACAATGGAATAACTATACTAAAGAGTAGAAAGCAGA	199
FJ947153.1	ACATCTTCGGAAAATTTACGTTCAACAATGGAATAACTATACTGAAGAGTAGAAAGCAGA	199
KY563368.1	ACATCTTCGGAAAATTTACGTTCAACAATGGAATAACTATACTAAAGAGTAGAAAGCAGA	199
FJ946835.1	CGAGCAACCTTACATAGCTGTTCATATGTAGAAAAGTTTTTTATTCCGAGA	510
HM113532.1	CGAGCAACCTTACATAGCTGTTCATATGTAGAAAAGTTTTTTATTCCGAGA	289
FN995637.1	TATTTGCAGCCTGCAAGAACAATCAATTCGTCTTTTGTGGAAAGCAAATAGCACAAAA	257
FN806775.1	TATTTGCAGCCTGCAAGAACAATCAATTCGTCTTTTGTGGAAAGCAAATAGCACAAAA	257
FJ947153.1	TATTTGCAGCCTGCAAGAGCAATCAATTCGTCTTTTGTGGAAAGCAGATAGCACAAAA	257
KY563368.1	TATTTGCAGCCTGCAAGAGCAATCAATTTGTCTTCTGTGGAAAGCAGATAGCACAAAA	257
FJ946835.1	TTCTGTTTTTTGACAT-TCTGGAAATATGCAAGAGAGAAAGCTATAGGTGCCATTGCAGG	569
HM113532.1	TTCTGTTTTTTGACAT-TCTGGAAATATGCAAGAGAGAAAGCTATAGGTGCCATTGCAGG	348
EN995637.1	TICIGACACTGCTGATGAAAAGACATGGACATTCAAGAGAACTGAAGCTGTCCTTAGAGT	317
ENGOCTTE 1	TTCTGACACTGCTGATGAAAAGACATGGACATTCAAGAGAACTGAAGCTGTCCTTAGAGT	217
F3047153.1		317
FJ94/155.1	TTCTGACACTGCTGATGAAAAGACATGGACATTCAAGAGAACTGAAGCCGTCCTTAGAGT	517
KY563368.1	TTCTGACACTGCTGATGAAAAGACATGGACATTCAAGAGAACTGAAGCTGTCCTTAGAGT ***** * ** ** ** ** ** ** ** ** ** ** *	317
FJ946835.1	GAGACTGGACAATAAAGGCAAAGGACCGCCGACACAGCATCAGCCTGGTTGCACAGGA	629
HM113532.1	GAGACTGGACAATAAAGGCAAAGGACCGCCGACACACAGCATCAGCCTGGTTGCACAGGA	408
FN995637.1	TCTCATGGCAAAGATGGTTGAAGAATGCACAACCCAGAAACT	359
FN806775.1	TCTCATGGCAAAGATGGTTGAAGAATGCACAACCCAGAAACT	359
F1947153.1	TCTCATGGCAAAGATGGTTGAAGAATGCACAACCCCAGAAACT	359
KY563368 1	TETEATGGCAAAGATGGTTGAAGAATGCACAACCCAGAAACT	359
	*** ** * *** * * *** * *	555
E104692E 1	ATCANANTIEGENGETACATTCANCCCETANGCANTANCCATAGGANGTGACATAAGTTT	690
1 3 3 4 0 6 3 3 . 1		469
HH113532.1	ATCAMATTGGGAGGTACATTCAACCCGTAAGCAATAACCATAGGAAGTGACATAAGTTT	468
FN995637.1	CLAGCAAGATATGTATTCAAAGCTTATGTCACTTCCTATGGTCATTGCTTACGGGT	415
FN806775.1	CCAGCAAGATATGTATTCAAAGCTTATGTCACTTCCTATGGTCATTGCTTACGGGT	415
FJ947153.1	CCAGCAAGATATGTATTCAAAACTTATGTCACTTCCTATGGTTATTGCTTATGGGT	415
KY563368.1	CCAGCAAGATATGTATTCAAAACTTATGTCACTTCCTATGGTTATTGCTTACGGGT	415

Hình 1. Kết quả so sánh các trình tự gen nucleocapsis protein của TNRV đã được công bố trên ngân hàng gen NCBI

Mồi được thiết kế với chiều dài 21 đến 23 nucleotide, tỷ lệ %GC từ 42,86 đến 43,48%, nhiệt độ nóng chảy từ 58,05 đến 58,94°C (Bảng 1) và được dự đoán có khả năng phát hiện được các chủng virus khác nhau do được thiết kế trên vùng tương đồng giữa các trình tự đã công bố.

Tên	Trình tự (5' - 3')	Độ dài	%GC	Tm (°C)	Sản phẩm (bp)
	TCTGACACTGCTGATGAAAAGAC	23	43,48	58,94	101
INRV	AGTTTCTGGGTTGTGCATTCT	21	42,86	58,05	101

Kết quả phân tích khả năng bắt cặp với các trình tự gen virus công bố trên NCBI (Hình 2) cho thấy mồi chỉ có khả năng bắt cặp với các trình tự gen *TNRV* mà không hề bắt cặp với trình tự gen của các virus khác (Hình 2); điều này cho thấy mồi có tính đặc hiệu cao.

Khuếch đại phân đoạn gen mục tiêu của virus

Phản ứng RT-PCR đã khuếch đại được phân đoạn DNA xuất hiện ở ở vị trí khoảng 100-200 bp, phù hợp với sản phẩm dự kiến (101 bp) của cặp mồi TNRV (Hình 3). Kết quả này bước đầu cho thấy đã khuếch đại được phân đoạn gen phù hợp với phân đoạn gen mục tiêu. Tuy nhiên, không thể chắc chắn trình tự đã được khuếch đại có đúng là trình tự gen của virus hay không vì mồi vẫn có khả năng bắt cặp và khuếch đại một trình tự RNA khác có trong mẫu RNA. Vì vậy, phân đoạn DNA (kích thước khoảng trên 100 bp) này đã được dòng hóa và giải trình tự để kiểm tra.

Description ▼	Scientific Name	Max Score ▼	Total Score ▼	Query Cover	E value ▼	Per. Ident	Acc. Len	Accession
Tomato necrotic ringspot virus isolate CT1 non-structural movement protein and glycoprotein genes. complete cds	<u>Tomato necrotic</u> ringspot virus	36.2	36.2	100%	4.7	100.00%	4726	<u>KY007015.1</u>
Tomato necrotic ringspot virus isolate C94M non-structural movement protein and glycoprotein genes, complete cds	<u>Tomato necrotic</u> ringspot virus	36.2	36.2	100%	4.7	100.00%	4736	<u>KY007014.1</u>
Tomato necrotic ringspot virus isolate chilli-CR segment M. complete sequence	<u>Tomato necrotic</u> ringspot virus	36.2	36.2	100%	4.7	100.00%	4724	<u>MW256414.1</u>
Tomato necrotic ringspot virus isolate TT1 segment M, complete seguence	Tomato necrotic ringspot virus	36.2	36.2	100%	4.7	100.00%	4716	FJ947152.1
Tomato necrotic ring virus segment M. complete sequence	Tomato necrotic ring virus	34.2	34.2	94%	19	100.00%	4725	ME469049.1

Hình 2. Kết quả phân tích khả năng bắt cặp của mồi với các trình tự gen virus bằng phần mềm BLAST NCBI



Hình 3. Kết quả điện di sản phẩm RT-PCR khuếch đại gen virus TNRV

Ghi chú: 1. chứng âm, 2. chứng dương (khuếch đại RNA của CyMV, 257 bp), 3. kiểm chứng PCR (297 bp); L. thang DNA, 4-8. RNA ly trích từ lá cà chua

Tạo dòng và kiểm tra trình tự của đoạn gen khuếch đại

Do vector pMiniT 2.0 có mang gen kháng ampicilin nên sau khi được chèn phân đoạn gen mục tiêu và biến nạp vào vi khuẩn, các dòng vi khuẩn biến nạp sẽ phát triển và tạo khuẩn lạc trên môi trường chứa ampicilin. Kết quả biến nạp với sự xuất hiện của các khuẩn lạc trên môi trường LB bổ sung 100 mg/L ampicilin cho thấy quá trình biến nạp đã thành công. Tuy nhiên, để đảm bảo vector đã được chèn đúng phân đoạn gen mục tiêu, các khuẩn lạc tiếp tục được kiểm tra với cặp mồi Cloning Analysis. Kết quả kiểm tra đã thu được phân đoạn DNA có kích thước phù hợp với dự kiến (410 bp), bao gồm kích thước đoạn chèn (101 bp) và kích thước đoạn gen mục tiêu (Hình 4).



Hình 4. Kết quả điện di sản phẩm PCR kiểm tra khuẩn lạc với cặp mồi pMiniT[™] 2.0

Ghi chú: L. Thang DNA, 1. Chứng âm, 2-4. Mẫu khuẩn lạc

Kết quả giải trình tự và so sánh trên ngân hàng gen NCBI cho thấy đoạn gen thu được tương đồng 98% với đoạn gen tương ứng của *TNRV* đã được công bố (Hình 5); chứng tỏ đoạn gen đã khuếch đại được đúng là gen nucleocapsis protein của *TNRV*.

Description	Scientific Name ▼	Max Score ▼	Total Score ▼	Query Cover	E value ▼	Per. Ident ▼	Acc. Len ▼	Accession
Tomato necrotic ringspot virus isolate P13 nucleocapsid protein (N) gene, complete cds	Tomato necrotic ringspot virus	176	176	100%	6e-40	98.02%	846	<u>KY563368.1</u>
Tomato necrotic ring virus segment S. complete sequence	Tomato necrotic ring virus	176	176	100%	6e-40	98.02%	3017	MF469050.1
Tomato necrotic ringspot virus isolate T1 NSs protein and nucleocapsid protein genes, complete cds	Tomato necrotic ringspot virus	176	176	100%	6e-40	98.02%	3016	KM887842.1
Tomato necrotic ringspot virus isolate 94 NSs protein and nucleocapsid protein genes, complete cds	Tomato necrotic ringspot virus	176	176	100%	6e-40	98.02%	3017	<u>KM887841.1</u>
Tomato necrotic ringspot virus isolate chilli-CR segment S, complete sequence	Tomato necrotic ringspot virus	176	176	100%	6e-40	98.02%	3013	MW256415.1

Hình 5. Kết quả so sánh trình tự đoạn gen khuếch đại với các trình tự gen *TNRV* công bố trên ngân hàng Gen NCBI

Thiết lập phản ứng real-time PCR

Phản ứng real-time PCR phát hiện *TNRV* cho sản phẩm khuếch đại được thể hiện bằng đường biểu diễn tín hiệu huỳnh quang. Trong đó, đường biểu diễn tín hiệu của các mẫu đều vượt trên tín hiệu nền (Hình 6), với chu kỳ ngưỡng lần lượt là 14,87; 18,69; 23,30. Sản phẩm khuếch đại có nhiệt độ nóng chảy (Tm) ở khoảng 77,5^oC (Hình 7). Điều này cho thấy việc thử nghiệm phản ứng real-time PCR khuếch đại phân đoạn gen mục tiêu đã thực hiện thành công.



Hình 6. Biểu đồ khuếch đại phân đoạn gen mục tiêu dựa trên tính hiệu huỳnh quang

Ghi chú: Trục tung (Y) - Cường độ tín hiệu huỳnh quang, trục hoành (X) - số chu kỳ (Ct). 1. Chứng âm, 2, 3, 4. Mẫu RNA



Hình 7. Biểu đồ phân tích nhiệt độ nóng chảy phân đoạn gen mục tiêu

Ghi chú: 1. Chứng âm; 2, 3, 4. Mẫu RNA

Thiết lập đường chuẩn cho phản ứng real-time PCR

Đường chuẩn đã được thiết lập dựa trên giá trị chu kì ngưỡng (Ct) thu được khi thực hiện phản ứng real-time PCR với các mẫu chuẩn có số lượng bản sao DNA thay đổi từ $10^6 - 10^1$ bản sao/µL (Bảng 2). Phương trình đường chuẩn có dạng y = -3,22x + 38,87 với hệ số tương quan R² = 0,99 (Hình 8).

Mẫu (bản sao/µL)	Log số bản sao	Ct
10 ⁶	6	19,29
10 ⁵	5	23,44
10 ⁴	4	25,65
10 ³	3	28,96
10 ²	2	32,30
10 ¹	1	35,88
40	TNRV ^y	= -3,22x + 38,8 R ² = 0,99

Bảng 2. Giá trị Ct của phản ứng khuếch đại gen



Hình 8. Biểu đồ đường chuẩn của phản ứng khuếch đại phân đoạn gen mục tiêu

Xác định giới hạn phát hiện của phản ứng real-time RT-PCR

Với 6 mẫu RNA (có nồng độ từ 10⁵-10⁰ bản sao RNA/µL) được sử dụng, phản ứng real-time RT-PCR chỉ cho tín hiệu khuếch đại ở các mẫu có nồng độ từ 10³ đến 10⁵ bản sao/µL, với chu kỳ ngưỡng có giá trị thay đổi từ 26,23 đến 20,91 (Bảng 3). Điều này cho thấy mẫu có nồng độ thấp nhất có thể phát hiện là 10³ bản sao/µL.

Kết quả thực hiện phản ứng lặp lại 10 lần với mẫu có nồng độ 10³ bản sao/µL đều thu được tín hiệu khuếch đại ở tất cả các lần lặp lại với giá trị chu kỳ ngưỡng trung bình ở mức 26,23 (Hình 9).

Như vậy, ngưỡng phát hiện của phản ứng được xác định ở mức 10³ bản sao/µL. Với ngưỡng phát hiện này, mặc dù có thể phát hiện được sự hiện diện của virus trong cây ở giai đoạn sớm nhưng vẫn còn thấp so với các kết quả phát hiện các virus khác trên cà chua như *Tomato black ring virus*, *Tomato yellow leaf curl virus* (Harper *et al.*, 2011; Huỳnh Nguyễn Minh Nghĩa *et al.*, 2020). Một thí nghiệm tối ưu hóa để nâng cao độ nhạy của phản ứng là cần thiết trong giai đoạn tiếp theo của nghiên cứu này.





Hình 9. Biểu đồ tín hiệu huỳnh quang khuếch đại gen với mẫu có nồng độ 103 bản sao/µ

Ghi chú: N. Chứng âm; 1. Mẫu 103 bản sao/µL

Xác định giới hạn định lượng của phản ứng real-time RT-PCR

Giới hạn định lượng của phản ứng real-time RT-PCR được phân tích với mẫu ở các nồng độ 10³, 10⁴, 10⁵ bản sao RNA/µL. Kết quả cho thấy, giá trị RSD của phản ứng đạt mức thấp hơn 25% ở mẫu có nồng độ 10³ bản sao/µL (Bảng 4). Như vậy, giới hạn định lượng của phản ứng có thể được xác định ở mức 10³ bản sao/µL.

Số bản sao lý thuyết	Số bản sao thực nghiệm	RSD (%)
10 ³	5,63 × 10 ³	18,87
10 ⁴	1,29 × 10 ⁴	10,05
10 ⁵	4,73 × 10 ⁵	12,25

Bảng 4. Kết quả xác định giới hạn định lượng phản ứng real-time RT-PCR khuếch đại gen

Định lượng virus trong mẫu cà chua

Virus có trong mẫu lá cà chua được định lượng dựa trên trên đường chuẩn đã thiết lập và chu kì ngưỡng trung bình của mẫu thử nghiệm (Bảng 5). Trong 10 mẫu được phân tích, chỉ có 02 mẫu LĐ-01 và LĐ-07 có sự hiện diện của virus TNRV với giá trị lần lượt là 8,52 × 10⁵ và 2,48 × 10⁸ số bản sao/µL.

Kí hiệu mẫu	Ct trung bình	Giá trị định lượng (Số bản sao/µL)					
LÐ-01	17,36	8,52 × 10⁵					
LÐ-02	-	-					
LÐ-03	-	-					
LÐ-04	-	-					
LÐ-05	-	-					
LÐ-06	-	-					
LÐ-07	9,61	2,48 × 10 ⁸					
LÐ-08	-	-					
LÐ-09	-	-					
LÐ-10	-	-					

Bảng 5. Giá trị Ct trung bình và định lượng tương ứng của mẫu lá cà chua thu thập

KÉT LUẬN

Đã thiết lập được quy trình phát hiện virus *TNRV* gây bệnh trên cà chua bằng real-time RT-PCR với ngưỡng phát hiện và định lượng đều ở mức 10^3 bản sao/µL. Quy trình phát hiện được sử dụng để định lượng virus trong các mẫu cà chua với lượng virus dao động từ 8,52 × 10^5 đến 2,48 × 10^8 số bản sao/µL.

TÀI LIỆU THAM KHẢO

Tuan NM, Mao NT (2015). Effect of plant density on growth and yield of Tomato (Solanum lycopersicum L.) at Thai Nguyen, Vietnam. Int J Plant Soil Sci, 7(6): 357 - 361.

Jolly B (2014). Tomato yellow leaf curl virus: A serious threat to tomato plants world wide. J Plant Pathol Microbiol, 7(4): 1-5.

Chiemsombat P, Sharman M, Srivilai K, Campbell P, Persley D, Attathom S (2010). A new tospovirus species infecting Solanum esculentum and Capsicum annuum in Thailand. Aus Plant Dis Note, 5(1): 75 - 78.

Seepiban C, Gajanandana O, Attathom T, Attathom S (2011). *Tomato necrotic ringspot virus*, a new tospovirus isolated in Thailand. *Arch Virol*, 156: 263 - 274.

Hoang Viet Hau, Ho Minh Hien, Luu The Trung, Nguyen Khoa Tuong, Le Ngoc Trieu (2021). Khảo sát thành phần và tình trạng virus gây bệnh hại cà chua tại vùng trồng chính của tỉnh Lâm Đồng, *Tạp chí nông nghiệp và phát triển nông thôn,* 1: 59-66.

Charoenvilaisiri S, Seepiban C, Bhunchoth A, Warin N, Luxananil P, Gajanandana O (2014). Development of a multiplex RT-PCR ELISA to identify four distinct species of tospovirus, *J Virol Meth*, 202: 54-63.

Jeong JJ, Ju HJ, Noh J (2014). A review of detection methods for the plant viruses. Res Plant Dis, 20(3): 173-181.

Nageswara-Rao M, Kwit C, Agarwal S, Patton MT, Skeen JA, Yuan JS, Stewart CN (2013). Sensitivity of a real-time PCR method for the detection of transgenes in a mixture of transgenic and non-transgenic seeds of papaya (*Carica papaya* L.). *BMC Biotechnol*, 13: 1-11.

Staroscik A (2018). Calculator for determining the number of copies of a template. URI Genomics & Sequencing Center, accessed 26 January 2018. Available from: http://cels.uri.edu/gsc/cndna.html.

European Network of GMO laboratories (2011). Verification of analytical methods for GMO testing when implementing interlaboratory validated methods. *JRC Scientific and Technical reports*, EUR 24790 EN.

Harper SJ, Delmiglio C, Ward LI, Clover GR (2011). Detection of *Tomato black ring virus* by real-time one-step RT-PCR. *J Virol Meth*, 171(1), 190-194.

Huynh Nguyen Minh Nghia, Nguyen Vinh, Duong Hoa Xo, Nguyen Xuan Dung (2020). Phát hiện *Tomato yellow leaf curl Kanchanaburi virus* gây hại trên cây cà chua bằng real-time PCR. Báo cáo toàn văn Hội nghị công nghệ sinh học toàn quốc: 168-174.

DETECTION AND QUANTIFICATION OF *Tomato necrotic ringspot virus* AFFECTING TOMATO (*Solanum lycopersicum*) BASED ON NUCLEOCAPSID PROTEIN GENE

Huynh Nguyen Minh Nghia, Nguyen Thi Kim Thoa, Nguyen Xuan Dung^{*}

Biotechnology Center of Ho Chí Minh City

SUMMARY

Tomato necrotic ringspot virus (TNRV) is a recently discovered virus that causes ring spot disease in tomatoes. Research for detecting this virus is still limited in quantity and mainly focuses on the qualitative level. This article shows the results of detection and quantification of TNRV using real-time RT-PCR technology. The viral nucleocapsid protein gene sequence served as the basis for designing primers and conducting real-time RT-PCR. The results showed that the real-time RT-PCR is capable of amplifying and quantifying the virus based on a gene segment (101 bp) with 98% sequence homology to the published gene of TNRV. The reaction has a limit of quantification of 1000 copies/µL based on the calibration curve (y = -3.22x + 38.87) with a correlation coefficient of $R^2 = 0.99$. The results of applying the established reaction for the detection and quantification TNV in ten collected samples of tomato from the field identified two samples with presence of the virus with highest quantification value reaching 2.48×10^8 copies/µL. These findings suggest that the real-time RT-PCR reaction established in this study can be utilized in practice to aid in managing tomato viruses.

Keywords: Viral disease, nucleocapsid protein gene, Tomato necrotic ringspot virus, real-time PCR.

^{*} Author for correspondence: Tel: 0903800438, Email: nxdung.snn@tphcm.gov.vn

ĐÁNH GIÁ TÍNH SINH MIỄN DỊCH CỦA PROTEIN TÁI TỔ HỢP MÃ HÓA VÙNG RBD CỦA VIRUS VIÊM PHẾ QUẢN TRUYỀN NHIỄM Ở GÀ

Nguyễn Thị Trà¹, Lê Thị Trà My¹, Chu Thanh Tâm¹, Trịnh Thái Vy¹, Ngô Hồng Dương¹, Nguyễn Thị Thu Hiền¹, Lê Thị Kim Xuyến¹, Đoàn Thị Thanh Hương¹, Hoàng Thị Thu Hằng^{1,2,*}, Phạm Bích Ngọc^{1,*}

¹Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Học viện Khoa học và Công nghệ Việt Nam, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Bệnh viêm phế quản truyền nhiễm (Infectiuos Bronchitis, IB) là một bệnh cấp tính nguy hiểm trên gà, do infectious bronchitis virus (IBV) gây ra. Hiên nay, bênh viêm phế quản truyền nhiễm ở gà đang là mối lo ở các nước chăn nuôi gà tập trung phát triển manh ở châu Á trong đó có Việt Nam. Tiêm phòng vaccine được xem là biên pháp hiêu quả nhất để kiểm soát sư bùng phát của dịch bênh này. Tuy nhiên, các vaccine phòng IBV hiên nay ở nước ta đa phần là vaccine nhập ngoại từ Trung Quốc, Hàn Quốc, Ân Đô... hoặc một số công ty thuốc thú y trong nước (MTV, AVAC...) nhập chủng IBV từ nước ngoài về để sản xuất vaccine nhược độc đông khô đơn giá hoặc nhị giá (kết hợp vacxin phòng bệnh New Castle), dẫn tới hiệu quả bảo hộ thấp hoặc không còn khả năng bảo hộ trước những biến chủng IBV mới. Trong nghiên cứu trước, protein RBD-VAC dung hợp motif GCN4pII đã được biểu hiện và sản xuất thành công từ lá thuốc lá *Nicotiana benthamiana*. Trong nghiên cứu này, khả năng kích thích đáp ứng miễn dịch của protein RBD-VAC-pII được đánh giá trên mô hình chuột, gà thí nghiệm và công cường độc bằng chủng IBHYM. Huyết thanh của chuột và gà được thu nhận để đánh giá đáp ứng sinh kháng thể IgG và IgY đặc hiệu RBD-VAC-pII thông qua phản ứng ELISA và Western blot. Kết quả cho thấy protein RBD-VAC-pII tái tổ hợp đã kích thích sinh kháng thể đặc hiệu trên động vật thí nghiệm, đồng thời đã ngăn chặn và giảm thiểu các triệu chứng cho gà trước thử thách công cường độc ở mức độ gần như ngang bằng với vaccine thương mại. Những kết quả này chứng tỏ tiểm năng của việc sử dụng kháng nguyên RBD-VAC-pII trong phát triển vaccine tiểu đơn vị phòng IBV.

Từ khoá: Hiệu giá kháng thể, IBV, tính sinh miễn dịch, vaccine thực vật, vùng gắn kết thụ thể.

ĐẶT VẤN ĐỀ

Bệnh viêm phế quản truyền nhiễm (infectious bronchitis, IB) là một bệnh cấp tính nguy hiểm trên gà, xảy ra quanh năm và trên mọi lứa tuổi ở gà, với tỷ lệ mắc bệnh 50-100%, gây chết 0-25%. Bệnh đặc biệt nguy hiểm đối với gà con dưới 1 tháng tuổi, và gây thiệt hại nghiêm trọng trên gà nuôi lấy trứng giống và trứng thương phẩm. Các dấu hiệu đầu tiên dễ thấy nhất bao gồm: thở khó kèm tiếng ran, chảy dịch mũi, gà con có ủ rũ, giảm ăn, bỏ ăn, đứng túm vào nhau. Gà lớn có biểu hiện giảm ăn nhẹ, thở hồn hển, ho, xuất huyết khí quản và chảy nước mũi. Nhiều trường hợp phát hiện bệnh tích trên khí quản, phổi, thận và ống dẫn trứng ở gà đẻ, sản lượng trứng giảm có thể tới 50-70%. Tác nhân gây bệnh viêm phế quản truyền nhiễm (IBV) thuộc họ *Coronaviridae*, giống Gammacoronavirus, cấu tạo bên ngoài giống hình "vương miện". Virus có đường kính xấp vỉ 120 nm với lớp vỏ bọc ngoài, bề mặt có gai hình dạng chùy dài khoảng 20 nm, bên trong chứa bộ gene là chuỗi RNA đơn dương (Cavanagh, 2007; Cook *et al.*, 2012; Bande *et al.*, 2017).

Cho tới nay, tất cả các loại vaccine thương mại được sử dụng phổ biến là vaccine sống nhược độc hoặc vaccine bất hoạt. Vaccine IBV nhược độc có một số hạn chế cần được khắc phục như khả năng lại độc và tổn thương mô (Jackwood *et al.*, 2017). Khả năng sản sinh kháng thể của vaccine bất hoạt tương đối thấp so với vaccine nhược độc và không tạo được đáp ứng qua trung gian tế bào T. Do đó, thông thường, vaccine bất hoạt phòng IBV được tiêm lần thứ hai sau khi đã kích thích miễn dịch bằng vaccine nhược độc, dẫn tới gia tăng giá thành của vaccine. Mặc dù hứa hẹn nhiều tiềm năng và khắc phục được một số hạn chế của vaccine nhược độc, những vaccine sử dụng hệ vector virus, vaccine tiểu đơn vị, hay vaccine DNA vẫn có một số hạn chế như các đáp ứng miễn dịch không mong muốn đối với hệ vector, sự giảm khả năng thu nhận kháng nguyên của các tế bào trình diện kháng nguyên, sự biến đổi kháng nguyên sau dịch mã, gấp cuộn và glycosyl hóa không hoàn chỉnh trong tế bào chủ ảnh hưởng đến tính sinh miễn dịch của kháng nguyên đích (Bande *et al.*, 2015). Vaccine tiểu đơn vị được sản xuất từ thực vật sử dụng phương pháp biểu hiện tạm thời có nhiều ưu điểm, trong đó ưu điểm lớn nhất là có thể tạo ra một lượng lớn các protein tái tố hợp trong thời gian ngắn (vài ngày đến vài tuần) sau khi biến nạp, chính vì vậy đáp ứng kịp thời với dịch bệnh xảy ra.

Vùng gene đầu 3' của virus mã hóa cho 4 protein cấu trúc: S (gai), M (màng), N (nucleoprotein) và protein vỏ (E) trong đó S1 là yếu tố quyết định khả năng gây bệnh và tính đặc hiệu của virus đối với tế bào vật chủ. Miễn dịch bảo hộ, phản ứng ức chế ngưng kết hồng cầu (HI) và kháng thể trung hòa virus đều được tạo ra bởi protein S1 (Li *et al.*, 2008). Tiểu phần S1 chứa vùng gắn kết thụ thể (Receptor Binding Domain-RBD, là phần biến đổi nhiều nhất của protein S và là cơ sở để xác định được kiểu huyết thanh (serotype) phân biệt giữa các chủng IBV khác nhau (Abro *et al.*, 2012; Promkuntod *et al.*, 2014). Điều này cho thấy tiềm năng của việc phát triển vaccine tiểu đơn vị phòng chống IBV dựa vào tiểu phần S, đặc biệt là vùng RBD. Trong nghiên cứu trước, chúng tôi đã biểu hiện và tinh sạch thành công protein RBD-VAC_pII từ lá thuốc lá *N. benthamiana* (Nguyen *et al.*, 2023). Trong nghiên cứu này, khả năng kích thích sinh miễn dịch của kháng nguyên RBD-VAC-pII có nguồn gốc từ thực vật được đánh giá trên chuột nhắt trắng và gà Leghorn. Các kết quả nghiên cứu cho thấy kháng nguyên tái tổ hợp này đã kích thích sinh kháng thể đặc hiệu. Đặc biệt, nó còn giúp gà giảm thiểu và chống lại các triệu chứng do virus gây ra. Những kết quả này cho thấy tiềm năng của việc sử dụng kháng nguyên RBD-VAC-pII có nguồn gốc từ thực vật trong việc phát triển vaccine tiểu đơn vị phòng IBV.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu

Nguồn nguyên liệu thực vật: Lá thuốc lá đã biểu hiện thành công protein tái tổ hợp RBD-VAC-pII do phòng Công nghệ ADN ứng dụng, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam cung cấp (Nguyen *et al.*, 2023).

Nguồn nguyên liệu động vật: Chuột cái BALB/C 6 tuần tuổi được cung cấp bởi Phòng Thử nghiệm sinh học, Viện Công nghệ sinh học, Viện Hàn lâm khoa học và Công nghệ Việt Nam.

Trứng gà Leghorn trắng sạch bệnh được nhập khẩu từ Công ty VALO Biomedia (Đức), ấp nở và nuôi tại phòng Miễn dịch học, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Chủng IBHYM gây bệnh tại Hưng Yên, Việt Nam phân lập năm 2021 cung cấp bởi phòng Miễn dịch học, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam (Roan *et al.*, 2023).

Phương pháp nghiên cứu

Phương pháp tinh sạch protein tái tổ hợp bằng sắc kí ái lực (Immobilized affinity chromatography, IMAC)

Các protein đích tái tổ hợp sau khi được kiểm tra đã biểu hiện thành công trên cây thuốc lá thì tiến hành tinh sạch bằng sắc kí ái lực theo công bố của Nguyen và đồng tác giả (2023). Lá được nghiền thành bột mịn với nitơ lỏng, sau đó bổ sung đệm binding buffer ((50 mM Na₂HPO₄, 300 mM NaCl, 100 mM Na₂SO₃, 1% Triton X100, pH 8.0) với tỉ lệ 1:3 để tách chiết. Dịch chiết được làm sạch bằng các ly tâm 2 lần trong 60 phút, ở 4 °C với tốc độ 13.000 rpm. Dịch chiết sau đó được trộn với Ni-NTA agarose (Qiagen, Singapore) qua đêm ở 4°C. Đưa hỗn hợp lên cột (Disposable Gravity Flow Columns, Marvelgent Biosciences, USA) và rửa cột với đệm rửa (50 mM Na₂HPO₄, 300 mM NaCl, 100 mM Na₂SO₃, 10 mM Imidazole, pH 8.0). Protein sau đó được hoà tan khỏi cột với đệm hòa tan (50 mM Na₂HPO₄, 300 mM NaCl, 500 mM Imidazole, pH 8.0) và được đổi sang đệm PBS không chứa NaCl qua đêm ở 4°C. Sau đó, protein tái tổ hợp được cô bằng cột cô protein cutoff 10K (Thermo Fisher Scientific, USA), được xác định bằng Western Blot và bảo quản với glycerol 30% (Merk, Đức) ở nhiệt độ -30°C. Các protein sau khi tinh sạch bằng IMAC tiếp tục được tinh sạch bằng phương pháp sắc ký lọc gel (Size exclusion chromatography, SEC) bằng cột SuperoseTM 6 increase 10/300GL column (GE Healthcare, USA) theo công bố của Phan và đồng tác giả (2017).

Phương pháp Western Blot

Nồng độ của protein tái tổ hợp tinh sạch bằng IMAC và SEC được xác định bằng phương pháp lai miễn dịch Western Blot. Protein tái tổ hợp cùng đối chứng dương thương mại SARS-CoV2 S1 (Sino Biological, China) được điện di SDS-PAGE trên gel polyacrylamide-SDS 12%, protein được chuyển qua màng bằng máy chuyển màng của Bio rad ở chế độ 25 V qua đêm. Màng chứa kháng nguyên được phủ bằng 5% sữa tách béo pha trong dung dịch PBS trong 2 giờ. Dùng kháng thể anti-His tag (Invitrogen, USA) với độ pha loãng 1.000 lần trong dung dịch sữa tách béo 5% phủ màng trong 2 giờ. Tiếp tục sử dụng kháng thể thứ cấp anti-mouse IgG có gắn Horseradish Peroxidase (HRP) (Invitrogen, USA) với độ pha loãng 400 lần trong sữa tách béo 5% phủ màng trong 2 giờ. Rửa màng bằng dung dịch PBS 3 lần, mỗi lần 10 phút. Hiện màu trong dung dịch hiện màu có chứa cơ chất Diaminobenzidine (DAB) (BioWorld, USA). Nồng độ protein tái tổ hợp sau khi được tinh sạch IMAC và SEC xác định được nồng độ bằng máy chụp AmershamTM Imager 680 và phần mềm ImageQuant TL 8.0 (Cytiva, USA).

Phương pháp gây đáp ứng miễn dịch trên chuột

Các protein sau khi tinh sạch IMAC, PBS và vaccine thương mại (Jovac IB H120- Jordan Bio-industries center) được sử dụng để đánh giá tính sinh miễn dịch trên chuột. Kháng nguyên, PBS được trộn đều với tá dược Emulsigen-D (MVP adjuvants, USA) bằng ống trộn theo tỉ lệ 4:1. Chuột cái trắng BALB/C 6 tuần tuổi được tiêm

dưới da háng ba lần, mỗi lần tiêm cách nhau 14 ngày, mỗi con được tiêm 5 ug protein, mỗi nhóm 6 con. Huyết thanh từ máu chuột được thu vào ngày thứ 7 sau lần tiêm thứ 2 và thứ 3 (Ho et al., 2022)

Phương pháp đánh giá đáp ứng kháng thể lgG ở chuột thí nghiệm bằng phản ứng ELISA

Đáp ứng kháng thể IgG được sản sinh từ các con chuột sau khi gây đáp ứng miễn dịch được đánh giá bằng phản ứng ELISA theo quy trình như sau: Đĩa microtiter (Immunoplate Maxibinding, SPL, Hàn Quốc) được phủ với protein tái tổ hợp tinh sạch SEC trong đệm PBS với nồng độ 30 ng/well và ủ qua đêm ở 4°C. Sau đó rửa đĩa lại 10 lần với dung dịch rửa PBST. Đĩa được phủ trong dung dịch PBS chứa 5% sữa không béo trong 2 giờ, lắc đều trên máy lắc. Đĩa tiếp tục được rửa 5 lần với PBST. Tiếp đến, đĩa được ủ với huyết thanh của chuột trong 2 giờ. Rửa đĩa 10 lần với PBST sau khi hết thời gian ủ với huyết thanh. Đĩa được ủ với kháng thể anti-mouse IgG có gắn Horseradish Peroxidase (HRP) (Invitrogen, USA) với độ pha loãng 2000 lần trong 2 giờ. Màng được rửa lại bằng dung dịch PBST 10 lần. Tín hiệu được phát hiện bằng dung dịch có chứa cơ chất 3,3',5,5'-Tetramethylbenzidine (TMB) (Thermo, USA) trong 20 phút. Sản phẩm phản ứng cho màu xanh, sau đó cố định màu bằng HCI 1M, màu vàng của phản ứng được đo ở bước sóng 450nm bằng máy Multiskan™ SkyHigh (Thermo Scientific, USA).

Phương pháp gây đáp ứng miễn dịch và công cường độc trên gà

Nhóm gà Leghorn trắng sạch bệnh được sử dụng để gây đáp ứng miễn dịch bằng cách tiêm dưới da cổ. Nhóm gà tiêm protein tái tổ hợp RBD-VAC-pII (7 con) mỗi con gà được tiêm với 200 µl chứa 5 µg protein tinh sạch IMAC và tá dược Emulsigen-D ở lần 1 vào ngày thứ 7 và lần 2 vào ngày thứ 21 sau khi gà nở. Gà đối chứng âm (6 con) được tiêm với 200 µl chứa PBS và tá dược Emulsigen-D. Nhóm đối chứng dương (7 con) được tiêm 200 µl vaccine thương mại H120 Jovac theo liều hướng dẫn của nhà sản xuất. Huyết thanh của gà được thu nhận sau 3 tuần tiêm mũi 2 để đánh giá sự có mặt của kháng thể IgY đặc hiệu với protein tái tổ hợp IBV. Sau khi tiêm mũi 2 để đánh giá sự có mặt của kháng thể IgY đặc hiệu với protein tái tổ hợp IBV. Sau khi tiêm mũi 2 được 24 ngày, tiến hành công cường độc trên gà. Chủng IBHYM được dùng để công cường độc với liều 0.2 ml 1000 EID₅₀ theo đường nhỏ mắt mũi. Huyết thanh được thu sau 7 và 14 ngày công cường độc để sử dụng tiến hành các thí nghiệm ELISA đánh giá đáp ứng miễn dịch IgY. Các bước thực hiện giống như phương pháp đánh giá đáp ứng kháng thể IgG ở chuột như đã nêu ở trên nhưng thay thế huyết thanh chuột bằng huyết thanh gà, kháng thể thứ cấp là anti-chicken IgY có gắn HRP (Invitrogen, USA) với độ pha loãng 1:5000.

Phương pháp xử lý thống kê

Phân tích thống kê cho ELISA được tiến hành sử dụng phần mềm Excel và SPSS. Sự sai khác giữa giá trị trung bình được biểu diễn là X±SD. Giá trị p<0,05 được xác định là sai khác có ý nghĩa thống kê.

KẾT QUẢ VÀ THẢO LUẬN

Tinh sạch protein VAC-RBD tái tổ hợp bằng phương pháp sắc kí ái lực (IMAC)

Để phục vụ cho mục đích đánh giá khả năng gây đáp ứng miễn dịch của protein RBD-VAC-pII tái tổ hợp trên mô hình động vật thí nghiệm, lá thuốc lá sau biến nạp được tinh sạch bằng phương pháp sắc kí ái lực (IMAC), sau đó được tinh sạch bằng phương pháp sắc kí lọc gel (SEC) để sử dụng cho các phản ứng ELISA hay Western blot.



Hình 1. Kết quả Western blot kiểm tra protein tái tổ hợp sau khi tinh sạch bằng IMAC

Thang chuẩn protein được đặt của hãng Thermo Fisher Scientific (Cat.No. 26619). 40 μL dịch protein tái tổ hợp sau khi tinh sạch và cô bằng cột cô protein kích thước 10 kDa cut-off (Thermo Fisher Scientific, USA) được tra vào các giếng chỉ ra trên hình. Protein SARS-CoV2 S1 được sử dụng làm đối chứng dương với các nồng độ khác nhau được chú thích trên mỗi giếng (300, 600, 900, 1200 và 1500 ng/μL, Sino Biological, China).

Tinh sạch protein tái tổ hợp bằng phương pháp IMAC theo công bố của Nguyen và đồng tác giả (2023) trước đó với sự thay đổi về khối lượng lá thuốc lá ban đầu là 220 g thay vì 110-140 g. Sau khi tiến hành thu được protein RBD-VAC-pII thì định lượng nồng độ bằng phần mềm ImageQuant TL 1D v8.2.0 với đối chứng dương là protein thương mại SARS-CoV2 S1 (Sino Biological, China). Kết quả ở Hình 1 cho thấy rằng, protein RBD-VAC-pII đã được tinh sạch thành công từ thực vật sử dụng phương pháp sắc ký ái lực (IMAC) đúng với kích thước tính toán lý thuyết của protein này (khoảng 40 kDa ở dạng monomer) và có một số băng vạch có kích thước lớn hơn do trạng thái trimer, glycosyl hóa với nồng độ là 932 ng/µl.

Đánh giá tính sinh kháng thể đặc hiệu của kháng nguyên RBD-VAC-pll trên mô hình chuột thí nghiệm bằng phản ứng ELISA

Để phục vụ cho việc đánh giá tính sinh miễn dịch của protein tái tổ hợp RBD-VAC-pII trên chuột, kháng nguyên RBD-VAC-pII tinh sạch bằng IMAC, PBS và vaccine thương mại được trộn với chất bổ trợ, sau đó được sử dụng theo sơ đồ gây đáp ứng miễn dịch trên chuột được thể hiện ở Hình 2A. Đáp ứng sản sinh kháng thể IgG đặc hiệu trong huyết thanh của các nhóm chuột sau tiêm được đánh giá bằng phản ứng ELISA. Kết quả ELISA ở Hình 2B cho thấy không phát hiện được kháng thể đặc hiệu trong huyết thanh của nhóm tiêm PBS. Trong khi đó, nhóm chuột tiêm protein tái tổ hợp có hiệu giá kháng thể lớn hơn đáng kể so với nhóm tiêm vaccine đối chứng về mặt thống kê. Sau cả 2 và 3 mũi tiêm, hiệu giá kháng thể IgG của nhóm RBD-VAC-pII đều xấp xỉ 17 log₂, trong khi đó vaccine thương mại có hiệu giá sau 2 và 3 lần tiêm lần lượt là 10.44 và 12.04 (log₂), chứng tỏ được rằng kháng nguyên tái tổ hợp RBD-VAC-pII đã kích thích sinh kháng thể IgG đặc hiệu trên mô hình chuột thí nghiệm.



Hình 2. Đánh giá đáp ứng sinh kháng thể IgG đặc hiệu trong huyết thanh chuột thí nghiệm bằng phản ứng ELISA

(A). Sơ đồ gây đáp ứng miên dịch trên chuột. (B). Hiệu giá kháng thể đặc hiệu lgG trong huyết thanh chuột sau khi tiếm. Huyết thanh thu từ chuột của 3 nhóm được đánh giá hàm lượng kháng thể đặc hiệu bằng phản ứng ELISA gián tiếp. Hiệu giá IgG của mỗi mẫu huyết thanh là hệ số pha loãng cao nhất cho giá trị OD_{450nm} cao hơn giá trị cut-off. Số liệu được trình bày dưới dạng trung bình ±SD cho mỗi nhóm chuột, ***: P<0.001.</p>

Đánh giá tính sinh miễn dịch và khả năng bảo hộ của kháng nguyên RBD-VAC-pll tái tổ hợp trên mô hình gà thí nghiệm

Theo kết quả đánh giá trên mô hình chuột thí nghiệm đã cho thấy kháng nguyên RBD-VAC-pll tái tổ hợp có khả năng kích thích sinh hàm lượng kháng thể cao ở chuột. Đây là loại protein tái tổ hợp có nguồn gốc từ chủng vaccine thương mại, là tiểu phần RBD đầu tiên của IBV được biểu hiện thành công trong cây thuốc lá, chứng minh tính khả thi của mô hình thiết kế (Nguyen *et al.*, 2023). Do đó, chúng tôi quyết định tiếp tục sử dụng kháng nguyên này để tiến hành các thí nghiệm tiếp theo trên gà. Để đánh giá được khả năng gây đáp ứng miễn dịch cũng như khả năng bảo hộ gà của protein tái tổ hợp, các nhóm gà thí nghiệm được bố trí như Hình 3. Sau khi tiêm protein tái tổ hợp hoặc vaccine thương mại, gà phát triển khỏe mạnh, không có sự khác biệt giữa các nhóm thí nghiệm. Huyết thanh được thu ở thời điểm 21 ngày sau khi tiêm vaccine, 7 ngày và 14 ngày sau công cường độc.



Hình 3. Sơ đồ gây đáp ứng miễn dịch trên mô hình gà thí nghiệm

Quan sát thể trạng và các triệu chứng của gà trong khi tiến hành thí nghiệm chúng tôi có một số kết quả triệu chứng lâm sàng của gà ở ngày thứ 7 sau công cường độc như Bảng 1 dưới đây:

		Triệu chứng			_
STT	Nhóm	Chảy nước mũi, hắt hơi, thở khò khè	Mệt mỏi, ủ rũ, run lạnh	Chậm lớn	Điểm thể trạng
1	Nhóm tiêm PBS	++++	+++	+	1
2	Nhóm tiêm RBD_VAC_pII	+	-	-	2
3	Nhóm tiêm vaccine thương mại	+	-	-	2

Bang 1. Triệu chững làm sáng của gã thì nghiệm ở ngày thứ v sau công cướng đồ				
Build II thou onally fall out of a full the light fill the four out of a one of a	Bang 1. Trieu chứng	lam sang của g	a thi nghiêm ở ngày thứ	y sau cong cương đóc

-: Không có dấu hiệu; +: Có dấu hiệu (số lượng dấu cộng thể hiện mức độ tăng dần của triệu chứng).

Gà thuộc nhóm tiêm PBS biểu hiện các triệu chứng điển hình của bệnh viêm phế quản truyền nhiễm (chảy nước mũi, hay hắt hơi, thở khò khè, gà ủ rũ, đứng co cụm, chậm lớn). Những triệu chứng này rất rõ ràng trong vòng 1-10 ngày đầu, sau đó có xu hướng giảm nhưng vẫn kéo dài tới ngày 14 của thí nghiệm. Ngược lại, gà thuộc nhóm tiêm protein tái tổ hợp RBD-VAC-pII và nhóm tiêm vaccine thương mại chỉ có những biểu hiện nhẹ; những biểu hiện này biến mất dần trong khoảng thời gian 7-10 ngày sau công cường độc. Như vậy, có thể khẳng định được rằng, tương tự như tiêm vaccine thương mại, việc tiêm protein tái tổ hợp này có thể giúp gà giảm các triệu chứng, hoặc tránh được các tác động không mong muốn khi bị nhiễm IBV. Song song với việc quan sát thể trạng và biểu hiện bệnh của các nhóm gà thí nghiệm, chúng tôi đánh giá khả năng sản sinh kháng thể đặc hiệu IgY đặc hiệu bằng phản ứng ELISA và Western blot. Kết quả ở Hình 4 cho thấy hiệu giá kháng thể IgY trong huyết thanh gà nhóm PBS rất thấp, gần như không được phát hiện và thấp hơn nhiều so với hai nhóm còn lại. Trong khi đó, nhóm gà tiêm protein tái tổ hợp có hiệu giá kháng nguyên tái tổ hợp ổn định ở cả ba thời điểm trước công, sau công 7, 14 ngày với giá trị xấp xỉ 15. Trong khi đó, nhóm tiêm vaccine có hiệu giá là 11.8 trước khi công, tăng nhẹ lên 12.3 sau khi công cường độc được 7 ngày và giảm xuống còn 11.3 log₂ sau khi công cường độc 14 ngày.



Hình 4. Hiệu giá kháng thể đặc hiệu IgY ở gà trước và sau khi công cường độc

Huyết thanh thu từ gà của 3 nhóm được đánh giá hàm lượng khánh thể đặc hiệu bằng phản ứng ELISA gián tiếp. Hiệu giá IgY của mỗi mẫu huyết thanh là hệ số pha loãng cao nhất cho giá trị OD_{450nm} cao hơn giá trị cut-off. Số liệu được trình bày dưới dạng trung bình ±SD cho mỗi nhóm gà, ***: P<0.001.





Song song với phản ứng ELISA, chúng tôi cũng tiến hành thí nghiệm Western blot để đánh giá khả năng sinh tháng thể đặc hiệu IgY ở các nhóm gà thí nghiệm. Đúng như dự đoán và phù hợp với kết quả của phản ứng ELISA như đã trình bày ở trên, Hình 5 thể hiện nhóm được tiêm protein tái tổ hợp sản sinh kháng thể rất mạnh mẽ tại cả ba thời điểm thu mẫu huyết thanh (nhất là sau công cường độc 7 ngày). Ở nhóm được tiêm PBS trước khi công hầu như không quan sát thấy sự xuất hiện của tín hiệu, sau công 7 và 14 ngày thì có xuất hiện tín hiệu nhưng không đáng kể so với nhóm tiêm protein tái tổ hợp. Kết quả này cho thấy rằng, protein tái tổ hợp RBD-VAC-pII có khả năng kích thích sinh kháng thể đặc hiệu rất mạnh. Điều này có thể được lí giải bởi vùng biểu hiện có chứa epitope quan trọng có tính kháng nguyên mạnh của IBV là RBD (receptor binding domain). Nó cũng đã được chứng minh ở nghiên cứu trước đây của Promkuntod và đồng tác giả (2014). Ngoài ra, việc biểu hiện trên hệ thống thực vật có một số ưu điểm như khả năng cải biến protein sau dịch mã, độ an toàn cao và có thể dễ dàng nhân rộng, nhanh mô hình để đáp ứng được dịch bệnh (Gidding *et al.*, 2001). Do đó, đây là một kháng nguyên tiềm năng cho sản xuất vaccine tái tổ hợp có nguồn gốc từ thực vật để phòng chống IBV.

KẾT LUẬN

Đã tinh sạch thành công protein tái tổ hợp RBD-VAC-pII bằng phương pháp IMAC và SEC với hàm lượng là 932 ng/µl từ 220 g lá sau biến nạp. Đặc biệt, đã đánh giá được khả năng sản sinh kháng thể của protein tái tổ hợp tinh sạch bằng IMAC trên mô hình chuột và gà thí nghiệm, với hiệu giá kháng thể IgG, IgY lần lượt xấp xỉ 17 và 15 (log₂). Cùng với đó chứng minh được khả năng chống chịu tốt của gà tiêm RBD-VAC-pII trước thử thách công cường độc của chủng virus IBHYM là tương đương với vaccine thương mại (Jovac IB H120). Những kết quả mở rộng khả năng sản xuất protein đích dựa trên hệ thống biểu hiện tạm thời ở thực vật, đồng thời chứng tỏ kháng nguyên RBD-VAC-pII có nguồn gốc từ lá thuốv lá có thể là một ứng viên tiềm năng để phát triển vaccine tiểu đơn vị phòng IBV tại Việt Nam.

Lời cảm ơn: Nghiên cứu được hỗ trợ kinh phí từ đề tài "Nghiên cứu đánh giá đặc tính sinh học của protein Spike tái tổ hợp từ thực vật của virus gây bệnh viêm phế quản truyền nhiễm ở gà (Avian Infectious virus-IBV)", cấp Viện Hàn lâm Khoa học và Công nghệ Việt Nam, mã số đề tài: VAST02.03/21-22. Nghiên cứu có sử dụng thiết bị Phòng Thí nghiệm Trọng điểm Công nghệ gen, Phòng Công nghệ Tế bào Thực vật và sự hỗ trợ của phòng Miễn dịch học, Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

TÀI LIỆU THAM KHẢO

Abro SH, Renström LHM, Ullman K, Belák S, Baule C (2012). Characterisation and analysis of the full- length genome of a strain of the European QX-like genotype of infectious bronchitis virus. *Arch Virol*, 157: 1211-1215.

Bande F, Arshad SS, Hair Bejo M, Moeini H, Omar AR (2015). Progress and challenges toward the development of vaccines against avian infectious bronchitis. *J Immunol Res*, 2015:424860.

Cavanagh D (2007). Coronavirus avian infectious bronchitis virus. Vet Res, 38: 281-297.

Cook JKA, Jackwood M, Jones RC (2012). The long view: 40 years of infectious bronchitis research. Avian Pathology, 4(3): 239-250.

Do TR, Nguyen TK, Luu MD, Nguyen TTH, Le TH, Le TH, Le TKX, Doan TH (2023). Whole genome sequencing analysis of Avian Infectious bronchitis virus isoled in Hung Yen province in 2021. *Vietnam J Biotechnol.* 21(4): 645-654.

Giddings G, Allison G, Brokks D, Carter A (2001). Transgenic plants as factories for biopharmaceuticals. *Nat Biotechnol*, 18: 1151-1155.

Ho TT, Trinh VT, Tran HX, Le PTT, Nguyen TT *et al* (2022). The immunogenicity of plant-based COE-GCN4pII protein in pigs against the highly virulent porcine epidemic diarrhea virus strain from genotype 2. *Front Vet Sci*, 9: 940395.

Jackwood MW (1999). Current and future recombinant viral vaccines for poultry. Advances in Veterinary Medicine, 41: 517–522.

Li L, Kang H, Liu P, Makkinje N, Williamson ST, Leibowitz JL, Giedroc DP (2008). Structural lability in stem-loop 1 drives a 5' UTR-3' UTR interaction in coronavirus replication. *J of Mol Bio.* 377: 790-803.

Nguyen TT, Ngo HD, Trinh TV, Le TTM, Nguyen TTH, Pham TV, Doan TTH, Pham BN, Chu HH, Hoang TTH (2023). Study on the transient expression of infectious bronchitis virus spike protein in *Nicotiana Benthamiana* leaves. *Vietnam J Biotechnol.* 21(3): 1-11.

Phan HT, Ho TT, Chu HH, Vu TH, Gresch U, Conrad U (2017). Neutralizing immune responses induced by oligomeric H5N1hemagglutinins from plants. *Vet Res.* 48:53.

Promkuntod N, van Eijndhoven REW, de Vrieze G, Gröne A, Verheije MH (2014). Mapping of the receptor binding domain and amino acids critical for attachment in the spike protein of avian coronavirus infectious bronchitis virus. *Virol*, 448:26-32.

INVESTIGATING THE IMMUNOGENICITY OF THE RECOMBINANT AVIAN IBV RECEPTOR-BINDING DOMAIN

Nguyen Thi Tra¹, Le Thi Tra My¹, Chu Thanh Tam¹, Trinh Thai Vy¹, Ngo Hong Duong¹, Nguyen Thi Thu Hien¹, Le Thi Kim Xuyen¹, Doan Thi Thanh Huong¹, Hoang Thi Thu Hang^{1,2*} Pham Bich Ngoc^{1*}

¹Institute of Biotechnology, Vietnam Academy of Science and Technology

²Graduate University of Science and Technology

ABSTRACT

Infectious Bronchitis (IB) is an acute disease affecting chickens caused by the avian infectious bronchitis virus (IBV). Currently, infectious bronchitis is a concern in highly developed chicken-raising countries in Asia, including Vietnam. Vaccination is considered the most effective measure to control IB outbreaks. However, the current IBV vaccines available in Vietnam are mostly imported vaccines from China, Korea, and India, in the form of monovalent or New Castle-combined bivalent vaccines. These imported IBV vaccine strains led to low or no protection against new IBV strains. In our previous research, RBD-VAC fused with GCN4pII motif was successfully expressed in *Nicotiana benthamiana*. In this study, the immunogenicity of RBD-VAC-pII was evaluated in mice and experimental chickens challenged with the IBHYM strain. Sera from mice and chickens were collected for the evaluation of RBD-VAC-pII-specific IgG and IgY antibody responses by ELISA and Western blot. The results showed that the recombinant RBD-VAC-pII protein stimulated the production of specific antibodies in experimental animals, prevented and minimized symptoms in chickens against virus challenge. These results demonstrate the potential of using the RBD-VAC-pII plant-derived recombinant protein in the development of subunit vaccines against IBV.

Keywords: Antibody titer, IBV, immunogenicity, plant-derived vaccines, receptor-binding doman.

Author for correspondence: Tel: 0975394838; Email: thuhanghoang2010@gmail.com. Tel: 0912247887; Email: pbngoc@ibt.ac.vn

TẠO DÒNG VÀ BIỀU HIỆN PROTEIN TÁI TỔ HỢP hMMP-1 (HUMAN MATRIX METALLOPROTEINASE 1) TRONG *E. COLI* BL21

Nguyễn Trọng Bình^{1*}, Trần Phú Hiển², Chương Thị Ngọc Hiếu¹, Nguyễn Thiện Phương¹

¹Trung tâm Công nghệ sinh học Thành phố Hồ Chí Minh

²Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

TÓM TẮT

MMP-1 là protein họ Matrix metalloproteinase, thuộc nhóm collagenase. Protein có khả năng phân hủy collagen loại I, II và III, các phân tử ECM (fibronectin, gelatin, aggrecan, laminin, perlecan và vitronectin) và một số protein hòa tan khác. MMP-1 đóng vai trò quan trọng trong quá trình phát triển, hình thành mô, sửa chữa các vết thương và cải thiện xơ gan... Hoạt động bất thường của MMP-1 có mối liên quan chặt chẽ đến sự xuất hiện và phát triển khối u, bệnh viêm khóp, bệnh Alzheimer... MMP-1 có nhiều vai trò quan trọng nhưng hiện nay MMP-1 có giá thành trên thị trường khá cao. Trong nghiên cứu này với mục tiêu sản xuất ra lượng lớn protein MMP-1 phục vụ cho nghiên cứu và ứng dụng. Đầu tiên, gen mã hóa MMP-1 được khuếch đại bằng phản ứng PCR rồi nối với vector pET28a. Sau khi kiểm tra sự hiện diện của gen bằng phương pháp PCR khuẩn lạc, cắt kiểm tra với hai enzyme cắt giới hạn *EcoRI* và *Hind*III ,vector pET28a–MMP-1 được hóa biến nạp để biểu hiện trong *E. coli* BL21 (DE3). Protein được thu nhận sau khi cảm ứng bằng 0.4 mM IPTG ở các điều kiện 37°C trong 6 giờ; 28°C trong 16 giờ; 16°C trong 18 giờ. Protein sau thu nhận, tinh sạch bằng phương pháp sắc ký ái lực với hạt Ni Sepharose và kiểm tra bằng SDS-PAGE, Western blot. Kết quả nghiên cứu đã tạo dòng, biểu hiện thành công protein MMP-1 tái tổ hợp tồn tại dạng thể vùi trong tế bào chất ở cả ba điều kiện khảo sát, tinh sạch và thu được 3,297 mg protein MMP-1 với kích thước khoảng 47 kDa trong 1 lít dịch lên men, độ tinh sạch 92,32% và hiệu suất tinh sạch 38,72%.

Từ khóa: Collagen, E. coli BL21 (DE3), Matrix metalloproteinase, MMP-1, pET28a.

ĐẶT VẤN ĐỀ

Human Matrix Metalloproteinase 1 (MMP-1) còn được gọi là collagenase 1, là một enzyme đóng vai trò quan trọng trong nhiều quá trình sinh học thiết yếu. MMP-1 giúp phá vỡ các thành phần của ma trận ngoại bào (ECM), chẳng hạn như collagen và elastin, cho phép các tế bào di chuyển, tái cấu trúc mô và sửa chữa tổn thương. MMP-1 có liên quan đến quá trình phát triển phôi thai, hình thành cơ quan và phân hóa tế bào. Bên cạnh đó, MMP-1 cũng đóng vai trò quan trọng trong phản ứng viêm và quá trình miễn dịch, giúp các tế bào miễn dịch di chuyển đến vị trí bị nhiễm trùng và tiêu diệt vi sinh vật. Trong nhiều loại ung thư, sự biểu hiện quá mức của MMP-1 thúc đẩy sự phát triển, di căn và xâm lấn của khối u. Đối với sự phát triển một số bệnh như bệnh viêm khớp, MMP-1 có liên quan đến sự thoái hóa sụn khớp trong viêm khớp. Với bệnh tim mạch, MMP-1 có liên quan đến sự hình thành mảng bám trong động mạch, có thể dẫn đến đau tim và đột quy.

MMP-1 là một glycoprotein bao gồm một chuỗi polypeptide đơn có trọng lượng phân tử khoảng 52 kDa. Gen mã hóa protein MMP-1 được xác định nằm trên nhiễm sắc thể 11 tại vị trí q22,2 với độ dài là 1971 bp với 10 exon và 9 intron. Cấu trúc của MMP-1 bao gồm một vùng xúc tác chứa các ion kim loại kẽm cần thiết cho hoạt động enzyme, một vùng liên kết collagen giúp MMP-1 gắn vào collagen, và một vùng đuôi tham gia vào sự tương tác với các protein khác. Ngoài ra, MMP-1 có thể được kích hoạt bởi các protein khác, ví dụ như plasmin và proteinase-3. Do đóng vai trò quan trọng, MMP-1 là chủ đề của nhiều nghiên cứu gần đây nhằm hiểu rõ hơn về chức năng của nó và phát triển các phương pháp điều trị mới cho các bệnh liên quan đến MMP-1. Đối với bệnh ung thư, các nhà nghiên cứu đang nghiên cứu các phương pháp nhắm mục tiêu MMP-1 để điều trị ung thư, chẳng hạn như sử dụng các chất ức chế MMP-1 hoặc các liệu pháp miễn dịch nhắm mục tiêu MMP1. MMP-1 có liên quan đến sự hình thành mảng bám trong động mạch, vì vậy kiểm soát MMP-1 hỗ trợ điều trị các bệnh tim mạch. Một số nghiên cứu gần đây cũng cho thấy ứng dụng của MMP-1 để làm chậm hoặc ngăn chặn sự tiến triển của bệnh viêm khớp, thúc đẩy phục hồi vết thương do kích hoạt tiểu cầu có khả năng phân huỷ collagen dạng sợi (Austin *et al.*, 2013) và ngăn ngừa hoặc làm chậm sự tiến triển của bệnh Alzheimer (Zhang *et al.*, 2023).

Nhìn chung, MMP-1 là một enzyme quan trọng tham gia vào nhiều quá trình sinh học thiết yếu. Nhờ khả năng phân huỷ collagen mà MMP-1 được đồng sử dụng với ghép tế bào gốc trung mô từ tủy xương mở ra cơ hội điều trị cho các bệnh nhân xơ gan (Du *et al.*, 2018); hỗ trợ điều trị sẹo lồi, sẹo phì đại; điều trị các vết loét ở bệnh nhân tiểu đường do hỗ trợ quá trình tân sinh mạch máu, di chuyển của tế bào trong giai đoạn tăng sinh, làm co và tái cấu trúc mô ở giai đoạn sửa chữa (Ayuk *et al.*, 2016). Việc sử dụng MMP-1 trong thực tế lâm sàng vẫn đang

được nghiên cứu và phát triển. Một hướng nghiên cứu cũng ứng dụng MMP-1 là dùng để phát triển các mô hình mới để nghiên cứu bệnh lý, chẳng hạn như ung thư và bệnh tim mạch. Trong nghiên cứu này, chúng tôi tiến hành tạo dòng, biểu hiện và tinh sạch protein MMP-1 tái tổ hợp nhằm mục tiêu sản xuất lượng lớn protein phục vụ cho việc ứng dụng nghiên cứu.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Chủng chủ và plasmid

Chủng *Escherichia coli* DH5α được sử dụng để nhân bản vector. Chủng *Escherichia coli* BL21 (DE3) được sử dụng để biểu hiện vector tái tổ hợp. Plasmid pET28a+ được sử dụng làm vector biểu hiện protein tái tổ hợp nhờ vào T7 promoter,trình tự mã hóa lacl và các operator lac. Vì thế việc biểu hiện protein mục tiêu có thể được thúc đẩy bằng isopropyl β-D-1-thiogalactopyranoside (IPTG), một chất tương tự lactose. Các chủng vi sinh và plasmid được cung cấp từ công ty Novagen, Đức. Thành phần nuôi cấy vi khuẩn *Escherichia coli* tuân theo hướng dẫn từ nhà sản xuất Biobasic, Canada.

Phương pháp

Tạo dòng E. coli DH5α mang gen *mã hóa protein MMP-*1

Thiết kế trình tự gen 1101 bp của MMP-1 từ vị trí 306 đến 1407 dựa trên cơ sở dữ liệu Genbank: NM_002421.4. Để hỗ trợ cho quá trình cắt, trình tự gen có chứa vị trí cắt của hai enzyme cắt giới hạn *EcoR*I và *Hind*III (12 bp). Đoạn gen mục tiêu được khuếch đại bằng phương pháp phusion PCR tổng thể tích phản ứng là 50 µL. Sau quá trình biến tính ban đầu ở 95°C trong 2 phút; 10 chu kì khuếch đại đầu tiên như sau: biến tính ở 95°C trong 30 giây và kéo dài ở 72°C trong 60 giây; 20 chu kỳ khuếch đại tiếp theo được thực hiện như sau: biến tính ở 95°C trong 30 giây, bắt cặp ở 60°C trong 30 giây và kéo dài ở 72°C trong 60 giây; bước kéo dài ở 72°C trong 60 giây; bước kéo dài sau cùng ở 72°C trong 10 phút và kết thúc phản ứng giữ ở 10°C.

Sản phẩm PCR khuếch đại gen mục tiêu sau khi được tinh sạch bằng GeneJET Gel Extraction Kit (Thermo Fisher ScientificTM). Sản phẩm PCR và plasmid pET28a lần lượt được xử lý với hai enzyme cắt giới hạn *EcoR*I và *Hind*III. Sản phẩm cắt được điện di trên gel agarose 0,8% và được tinh sạch bằng GeneJET Gel Extraction Kit (Thermo Fisher ScientificTM). Tiến hành nối gen MMP-1 với plasmid pET28a dưới tác động của enzyme T4 DNA ligase. Sản phẩm nối được biến nạp vào tế bào *E. coli* DH5α bằng phương pháp hóa biến nạp và sàng lọc trên môi trường LB agar có kháng sinh Kanamycin nồng độ 50 µg/mL. Các khuẩn lạc dự tuyến sẽ được kiểm tra lại bằng phản ứng PCR khuẩn lạc với mồi xuôi đặc hiệu cho pET28a (T7 promoter) và mồi ngược đặc hiệu cho gen MMP-1. Để chắc chắn biểu hiện đúng protein MMP-1 đã nhân dòng chúng tôi tiến hành tách, tinh sạch plasmid pET28a-MMP-1 và lập phản ứng cắt kiểm tra cũng như giải trình tự gen bằng phương pháp Sanger.

Biểu hiện, tinh sạch và phát hiện protein tái tổ hợp MMP-1

Vector biểu hiện pET28a giúp tạo ra protein tái tổ hợp có gắn đuôi 6xHis ở đầu N, hỗ trợ quá trình tinh sạch bằng Ni Sepharose. Sau khi biến nạp thành công pET28a-MMP-1 vào chủng *E.coli* BL21 (DE3), tiến hành nuôi cấy sơ cấp *E. coli* BL21 (DE3)/pET28a-MMP-1 trong môi trường LB/Kanamycin ở 37°C, 250 vòng/phút, nuôi cấy thứ cấp đến khi OD 600 nm đạt từ 0,6-0,8 thì bổ sung IPTG 0,4 mM. Quá trình biểu hiện protein khảo sát ở các điều kiện 37°C trong 6 giờ, 28°C trong 16 giờ, 16°C trong 18 giờ với tốc độ lắc 250 vòng/phút. Sau đó, tiến hành ly tâm dịch nuôi cấy để thu sinh khối vi khuẩn. Sinh khối được huyền phù với đệm (20 mM Tris-HCI, 300 mM NaCI, 10 mM imidazole, 8M Urea, pH 8.0) phá mẫu bằng sóng siêu âm và thu riêng các tổng tế bào, phân đoạn dịch tế bào, cặn sau ly tâm. Sau đó, tiến hành kiểm tra bằng phương pháp điện di SDS-PAGE 12,5% và nhuộm với Coomassie Brilliant Blue G-250. Các mẫu đối chứng: *E. Coli* BL21 (DE3)/ pET28a và *E. coli* BL21 (DE3)/

Sau khi xác nhận đã biểu hiện thành công protein MMP-1, dịch protein sẽ được loại bỏ sinh khối bằng phương pháp ly tâm tốc độ cao và lọc với kích thước lỗ lọc 0,45 µm. Dịch protein sau lọc được tinh sạch theo bộ kit ProBondTM Purification system. Rửa cột với đệm rửa có nồng độ imidazole 20 mM. Sau đó, các protein trên cột được tái gấp cuộn bằng các dung dịch đệm có nồng độ ure giảm dần. Cuối cùng, protein trên cột sẽ được thôi giải bằng dung dịch thôi giải có nồng độ Imidazole cao (500 mM). Protein sau khi tinh sạch được kiểm tra bằng phương pháp điện di SDS-PAGE 12.5% và nhuộm với Coomassie Brilliant Blue G-250.

Ngoài ra, sự hiện diện của protein tái tổ hợp còn được xác nhận bằng phương pháp Western Blot với kháng thể đặc hiệu. Kháng thể sơ cấp sử dụng (Anti-6X His tag® antibody [HIS.H8]) đặc hiệu với protein MMP-1 tái tổ hợp và kháng thể thứ cấp (Goat Anti-Mouse IgG H&L (HRP)) được đánh dấu bằng horseradish peroxidase. Màu được phát hiện nhờ phản ứng xúc tác của peroxidase trên cơ chất luminol và H₂O₂ ở bước sóng 425 nm. Kích thước protein MMP-1 tái tổ hợp được dự đoán khoảng 47 kDa.

KẾT QUẢ VÀ THẢO LUẬN

Tạo dòng *E. coli* DH5α mang gen mã hóa protein MMP-1

Từ Hình 1A cho thấy gen mục tiêu MMP-1 sau khi được khuếch đại và điện di trên gel agarose 0,8% thu nhận được duy nhất một đoạn gen có kích thước khoảng 1113 bp tương ứng với kích thước khi thiết kế của gen MMP-1. Điều này chứng tổ rằng đã khuếch đại được đoạn gen mục tiêu mã hóa cho protein MMP-1 tái tổ hợp.

Gen MMP-1 và plasmid pET28a+ được nối với nhau sau khi tạo đầu dính bằng cặp enzyme cắt giới hạn là *Eco*RI và *Hind*III. Sản phẩm nối được biến nạp vào vi khuẩn *E. coli* DH5α. Sàng lọc bước đầu plasmid pET28a+-MMP-1 dựa trên môi trường nuôi cấy LB có chứa kháng sinh Kanamycin (50 µg/ ml), những khuẩn lạc dự tuyển được kiểm tra bằng phương pháp PCR với mồi T7 và MMP-1 cho kết quả điện di kích thước khoảng 1313 bp bằng kích thước của gen MMP-1 (1113 bp) cộng với chiều dài của plasmid không có gen chèn khi được khuếch đại bằng đoạn mồi T7 promoter (Hình 1B). Các khuẩn lạc dương tính được tiếp tục nuôi cấy và tách plasmid để cắt kiểm tra. Plasmid pET28a-MMP-1 khi cắt với enzyme cắt giới hạn *Eco*RI và *Hind*III cho ra kích thước 1113 bp phù hợp với gen mục tiêu (Hình 1C). Ngoài ra, Kết quả giải trình tự gen cho thấy trình tự gen giống với trình tự khi thiết kế gen MMP-1 (Dữ liệu không trình bày).



Hình 1. Kết quả nhân dòng gen mã hóa protein MMP-1. Sản phẩm khuếch đại gen MMP-1 bằng cặp mồi đặc hiệu cho gen (A). Kết quả PCR khuẩn lạc bằng mồi xuôi T7 promoter đặc hiệu cho vector pET28a và mồi ngược đặc hiệu cho gen (B). Kết quả cắt kiểm tra plasmid bằng enzyme cắt giới hạn EcoRI và HindIII (C). M. Thang kích thước DNA.

Biểu hiện, tinh sạch và phát hiện protein tái tổ hợp MMP-1

Sau khi nhân dòng thành công, plasmid pET28a-MMP-1 được biến nạp vào chủng vi khuẩn *Escherichia coli* BL21 (DE3), và biểu hiện trong điều kiện có chất cảm ứng IPTG và đối chứng không có chất cảm ứng. Biểu hiện ở các nhiệt độ 37°C, 28°C và 16°C. Tế bào *E. coli* được phá vỡ mẫu bằng sóng siêu âm và kiểm tra protein sau phá mẫu bằng điện di SDS-PAGE 12,5% và nhuộm với Comassive Brilliant G-250.

Kết quả phân tích ở nhiệt độ 37°C (Hình 2A) cho thấy có một vạch protein biểu hiện vượt mức ở giếng có bổ sung IPTG pET28a-MMP-1 (+) với kích thước khoảng 47 kDa đúng với kích thước MMP-1 đã dự đoán và không thấy sự biểu hiện của vạch này ở đối chứng âm là plasmid pET28a (+) không chứa gen mục tiêu, có cảm ứng IPTG. Điều kiện biểu hiện được kiểm soát bởi chứng dương là plasmid pET28+-EGFP chứa gen EGFP, sự biểu hiện vượt mức của gen EGFP (kích thước ~ 35kDa) quan sát được ở giếng có bổ sung chất cảm ứng IPTG. Bên cạnh đó kết quả điện di cho thấy có sự khác biệt khi không có chất cảm ứng và có chất cảm ứng (Hình 2A, Hình 3A, Hình 4A) ở cả 3 nhiệt độ khảo sát. Khi không có cảm ứng IPTG thì protein tổng của *E.coli* được biểu hiện nhiều hơn nhưng khi có cảm ứng IPTG thì protein mục tiêu lại được biểu hiện nhiều hơn. Protein MMP-1 sau khi biểu hiện chủ yếu nằm ở dạng thể vùi trong pha cặn sau ly tâm với kích thước 47 kDa, điều này thể hiện ra trong kết quả điện SDS-PAGE (Hình 2B).



Hình 2. Kết quả điện di SDS-PAGE của tổng protein biểu hiện khi không có cảm ứng IPTG (-) và không có cảm ứng IPTG (+) ở 37ºC (A) và biểu hiện của protein mục tiêu ở các pha tổng tế bào, cặn tế bào và dịch tế bào sau ly tâm (B). M là thang phân tử lượng protein

Kết quả tương tự cũng xảy ra ở nhiệt độ biểu hiện protein ở 28°C, protein MMP-1 vẫn được tế bào biểu hiện vượt mức ở 28°C với kích thước khoảng 47 kDa và protein nằm ở chủ yếu ở dạng thể vùi (Hình 3B). Tuy nhiên, ở nhiệt độ này đã thấy có xuất hiện protein biểu hiện ở dạng tan rõ hơn so với 37°C. Sự biểu hiện vượt mức của đối chứng dương pET28a+EGFP cũng được ghi nhận ở nhiệt độ 28°C.



Hình 3. Kết quả điện di SDS-PAGE của tổng protein biểu hiện khi có cảm ứng IPTG và không có cảm ứng IPTG ở 28°C (A) và biểu hiện của protein mục tiêu ở các pha tổng tế bào, cặn tế bào và dịch tế bào sau ly tâm (B). M là thang phân từ lượng protein

Kết quả phân tích SDS-PAGE (Hình 4B) cho thấy ở 16°C protein mục tiêu vẫn được biểu hiện vượt mức với kích thước khoảng 47 kDa nhưng lượng protein thấp hơn ở 37°C và 28°C. Tuy nhiên, protein MMP-1 có biểu hiện ở dạng dịch nổi ở 16°C có tỷ lệ lớn nhất ở cả 3 nhiệt độ. Bên cạnh đó, sự biểu hiện vượt mức của đối chứng dương pET28a+EGFP cũng được ghi nhận ở nhiệt độ 16°C.



Hình 4. Kết quả điện di SDS-PAGE của tổng protein biểu hiện khi có cảm ứng IPTG và không có cảm ứng IPTG ở 16ºC (A) và biểu hiện của protein mục tiêu ở các pha tổng tế bào, cặn tế bào và dịch tế bào sau ly tâm (B). M là thang phân tử lượng protein

Protein sau khi biểu hiện và chọn ra nhiệt độ thích hợp thì protein MMP-1 sẽ được mang đi biểu hiện ở thể tích lớn và được tinh sạch bằng hạt Ni Sepharose, với chất cạnh tranh là imidazole. Sau khi qua trình tinh sạch, protein sẽ được kiểm tra bằng phương pháp SDS-PAGE với gel có nồng độ 12,5% acrylamide và đánh giá đặc hiệu bằng phương pháp Western blot với kháng thể kháng 6xHis tag.



Hình 5. Kết quả điện di SDS-PAGE sau tinh sạch (A) và Western Botting sau tinh sạch (B). M là thang phân tử lượng protein

Kết quả Western Blot cũng chỉ ra được đối với đối chứng dương pET28a-EGFP có kháng thể chống 6xHis cũng cho kết quả một vạch đúng kích thước khoảng 35 kDa, chứng tỏ quá trình thực hiện Western blot thành công. Kết quả điện di và kết quả Western Blot sau tinh sạch chỉ ra rằng MMP-1 tái tổ hợp đã được biểu hiện thành công trên chủng *E*. coli BL21(DE3)/pET28a-MMP-1 dưới dạng dung hợp với đuôi 6xHis tag và phần lớn protein MMP-1 thu được có một vạch đúng kích thước 47 kDa (Hình 5). Nồng độ protein được định lượng tương đối bằng phương pháp Bradford và mức độ tinh sạch của dung dịch protein thu được được phân tích hình ảnh bằng phần mềm ImageJ. Kết quả lượng protein thu được khoảng 3,297 mg, độ tinh sạch 92,32% trong 1 lít dịch lên men, hiệu suất tinh sạch là 38,72%.

So với nghiên cứu của nhóm tác giả Kumar L trước đây, biểu hiện được proMMP-1 với kích thước 54 kDa dạng tan và vùng xúc tác 28 kDa dạng thể vùi với lượng protein sau tin sạch là 3,5 mg, nồng độ 0,5 mg/ mL từ 1 g vi khuẩn (Kumar *et al.*, 2018). Thì trong nghiên cứu này chúng tôi biểu hiện được protein MMP-1 với kích thước 47 kDa tồn tại ở dạng thể vùi và dạng tan, lượng protein sau khi tinh sạch từ thể vùi là 3,297 mg, độ tinh sạch 92,32% trong 1 lít lên men và hiệu suất tinh sạch 38,52%. Ngoài ra, chúng tôi còn xác định được phần trăm protein mục tiêu so với tổng protein biểu hiện ở từng pha tổng, cặn và dịch sau ly tâm, đây là điểm sáng và chưa ghi nhận ở các công bố nghiên cứu biểu hiện MMP-1 tái tổ hợp trước nay.

Việc protein MMP-1 có thể được biểu hiện ở cả 37°C, 28°C, 16°C cho thấy rằng khoảng nhiệt độ mà nó có thể được biểu hiện trong *E. coli* BL21 (DE3) rất rộng. Ngoài ra, vì đa phần protein MMP-1 biểu hiện ở dạng thể vùi sau khi được biểu hiện giúp bảo vệ protein khỏi sự phân hủy của các protease trong tế bào chất của *E. coli*, tránh thất thoát protein do bị phân huỷ. Tuy nhiên, hoạt tính sinh học của protein có thể bị ảnh hưởng sau khi tái gấp cuộn do quá trình gấp cuộn sai, ảnh hưởng đến hàm lượng protein MMP-1 có hoạt tính thực tế thu được. Do đó, hướng nghiên cứu tiếp theo mà nhóm nghiên cứu nhắm đến sau bước đầu biểu hiện thành công MMP-1 tái tổ hợp là tiến hành thử nghiệm hoạt tính của protein MMP-1 thu được và tiến hành tối ưu hóa quy trình tinh sạch để thu được hàm lượng cao protein MMP-1 có hoạt tính tối ưu.

KÉT LUẬN

Chúng tôi đã nhân dòng thành công đoạn gen mã hóa cho protein MMP-1 với trình tự đúng hoàn toàn với trình tự gen đã thiết kế; tạo thành công dòng tế bào *E. Coli* BL21(DE3) mang vector pET28a-MMP-1 có khả năng biểu hiện protein MMP-1 tái tổ hợp. Nghiên cứu cho thấy protein MMP-1 biểu hiện vượt mức dưới dạng thể vùi trong *E. coli* BL21 (DE3) ở cả 3 nhiệt độ nghiên cứu và có biểu hiện dạng tan ở 28°C và 16°C khi có cảm ứng bằng IPTG, kích thước protein thu được khoảng 47 kDa. Thêm vào đó chúng tôi cũng đã tinh sạch được protein mục tiêu sau khi biểu hiện. Nghiên cứu đã bước đầu tạo được nguồn cung cấp MMP-1 tái tổ hợp cho các nghiên cứu tiếp theo hướng đến việc phát triển sản phẩm cho ứng dụng nghiên cứu tiếp theo.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Trung tâm Công nghệ sinh học Thành phố Hồ Chí Minh.

TÀI LIỆU THAM KHẢO

Austin KM, Covic, L, Kuliopulos A (2013). Matrix metalloproteases and PAR1 activation. Blood, 121(3): 431-9.

Ayuk SM, Abrahamse H, Houreld NN (2016). The role of matrix metalloproteinases in diabetic wound healing in relation to photobiomodulation. *J Diabetes Res*,2: 89765-6.

Dai L, Mugaanyi J, Cai X Dong M, Lu C, Lu C (2022). Comprehensive bioinformatic analysis of MMP1 in hepatocellular carcinoma and establishment of relevant prognostic model. *Sci Rep*, 12(1): 136-139.

Du C, Jiang M, Wei X, Qin J, Xu H, Wang Y, Zhang Y, Zhou D, Xue H, Zheng S, Zeng W (2018). Transplantation of human matrix metalloproteinase-1 gene-modified bone marrow-derived mesenchymal stem cell attenuates CCL4-induced liver fibrosis in rats. *Int J Mol Med*, 41(6): 3175-3184.

Ita M, Singh S, Troche H, Welch R, Winkelstein B (2022). Intra-articular MMP-1 in the spinal facet joint induces sustained pain and neuronal dysregulation in the DRG and spinal cord, and alters ligament kinematics under tensile loading. *Frontiers in Bioengineering and Biotechnology*, 10: 926675.

Kumar L, Colomb W, Czerski J, Cox CR, Sarkar SK (2018). Efficient protease based purification of recombinant matrix metalloprotease-1 in *E. coli. Protein Expr Purif*, 148: 59-67.

Li J, Brick P, O'Hare MC, Skarzynski T, Lloyd LF, Curry VA, Clark IM, Bigg HF, Hazleman BL, Cawston TE, Blow, DM (1995). Structure of full-length porcine synovial collagenase reveals a C-terminal domain containing a calcium-linked, four-bladed;-propeller. *Structure*, 3(6): 541-549.

Spurlino JC, Smallwood AM, Carlton DD, Banks TM, Vavra KJ, Johnson JS, Cook ER, Falvo J, Wahl RC, Pulvino TA (1994). 1.56 A structure of mature truncated human fibroblast collagenase. *Proteins*, 19(2): 98-109.

Zhang H, Sun C, He B, Zhang X, Hao H, Hou Y, Li A, Wang Y, Wang Y (2023). Macrophage migration inhibitory factor promotes expression of matrix metalloproteinases 1 and 3 in spinal cord astrocytes following gecko tail amputation. *J Integrat Neurosci*, 22: 29.

CLONING, EXPRESSION AND PURIFICATION OF GENE ENCODING PROTEIN hMMP-1 (HUMAN MATRIX METALLOPROTEINASE 1)

Nguyen Trong Binh^{1*}, Tran Phu Hien², Chuong Thi Ngoc Hieu¹, Nguyen Thien Phuong¹

¹Biotechnology Center of Ho Chi Minh City

²Ho Chi Minh City University of Science

SUMARY

MMP-1 is a member of the matrix metalloproteinase family, specifically within the collagenase group. This protein has the ability to break down collagen types I, II, and III, as well as several other important proteins in the extracellular matrix. MMP-1 is involved in tissue formation, wound healing, and liver cirrhosis. However, abnormal MMP-1 activity is associated with conditions such as tumors, arthritis, and Alzheimer's disease. Despite its important roles, MMP-1 is currently quite expensive. In this study, we aimed to produce large quantities of MMP-1 protein for research and applications. To achieve this, we cloned the gene encoding the MMP-1 protein into the pET28a vector and used the *E. coli* BL21 strain (DE3) for protein expression. The gene encoding MMP-1 was amplified by PCR and then inserted into the pET28a vector. After confirming the presence of the gene by colony PCR and restriction enzyme digestion (*EcoR*I and *Hind*III), the pET28a-MMP-1 vector was expressed in *E. coli* BL21 (DE3) after transformation. We induced protein expression using IPTG at different temperatures (37°C for 6 hours, 28°C for 16 hours, and 16°C for 18 hours). The expressed proteins were harvested, purified by affinity chromatography, analyzed using SDS-PAGE and Western blot and quantified by the Bradford measurement method. The results showed successful cloning and expression of recombinant MMP-1 protein in *E. coli*. We obtained 3,297 mg of MMP-1 protein per liter of fermentation liquid, with a purity of 92.32% and purification efficiency of 38.72%.

Keywords: Collagen, E. coli BL21 (DE3), Matrix metalloproteinase, MMP-1, pET28a.

Author for correspondence: Tel: 0904516 382, Email: ntbinh.snn@tphcm.gov.vn or ntbinhbiomedicine@gmail.com

GIẢI MÃ VÀ PHÂN TÍCH GEN S1 CỦA CHỦNG VIRUS PORCINE EPIDEMIC DIARRHEA GÂY BỆNH TIÊU CHẢY CẤP TRÊN LỢN TẠI TỈNH HƯNG YÊN NĂM 2023

Lưu Minh Đức^{1,2}, Đỗ Thị Roan^{1,2}, Nguyễn Thị Khuê^{1,2}, Don Nguyen³, Đoàn Thị Thanh Hương^{1,2*}

¹Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Squalicum High School Bellingham, Washington, USA

TÓM TẮT

Virus PED gây bệnh tiêu chảy cấp chủ yếu ở lợn con chưa cai sữa. Tỷ lệ nhiễm bệnh trong đàn khá cao, tỷ lệ chết cũng tương đối cao gây nên nhiều thiệt hại cho ngành chăn nuôi. Kích thước hệ gen virus khoảng 28 kb mã hóa cho các gen quan trọng cho quá trình nhân bản và hoàn thiện cấu tạo virus. Gen kháng nguyên spike chứa tiểu phần S1 bao gồm phần protein dẫn và hai vùng thụ thể liên kết. Gen S1 của chủng PEDVHY1 và PEDVHY2 thu nhận tại Hưng Yên năm 2023 đều có kích thước 2205 nucleotide mã hóa cho 735 amino acid thuộc genotype G2b. Protein S1 chứa nhiều đột biến, đặc biệt là đột biến chèn đoạn và xóa đoạn so với chủng virus nhược độc vaccine. Vùng gen S1 của hai chủng nghiên cứu ít có sai khác hơn so với các chủng thuộc genotype khác. Phân tích phả hệ nguồn gốc cho thấy hai chủng PEDV nghiên cứu có quan hệ gần gũi với nhau trong cây phả hệ và cùng nhóm với các chủng PEDV thu thập tại Việt Nam thu nhận trong giai đoạn 2013 đến 2018, thuộc genotype G2b, khác nhóm với chủng virus vaccine DR13 và CV 777 nhược độc.

Từ khóa: Gen S1, genotype, Hưng Yên, PEDV.

MỞ ĐẦU

Bệnh tiêu chảy trên lợn (PED-Porcine Epidemic Diarrhea) do virus tiêu chảy cấp ở lợn (PEDV-Porcine Epidemic Diarrhea Virus) gây ra. Đây là một bệnh truyền nhiễm với tốc độ lây lan nhanh, lợn mắc bệnh có triệu chứng nôn mửa, tiêu chảy cấp tính, mất nước nghiêm trọng dẫn đến tử vong. Tỷ lệ tử vong của lợn dưới một tuần tuối đặc biệt cao, tỷ lệ tử vong của lợn trưởng thành thấp hơn (Pospischil *et al.*, 2002).

Bệnh PED lần đầu tiên xuất hiện tại Anh vào năm 1971, sau đó nhanh chóng lan ra các khu vực châu Âu bao gồm Thụy Sĩ, Đức, Pháp, Hà Lan và Bulgaria (Pensaert, De Bouck, 1978). Tại châu Á, PED lần đầu tiên được ghi nhận ở Nhật Bản vào năm 1982, sau đó là hàng loạt các quốc gia lớn như Trung Quốc năm 1986, Ấn Độ năm 2003 và Thái Lan năm 2007 (Temeeyasen *et al.*, 2013). Năm 2013, một đợt bùng phát dịch lớn đã xảy ra tại Mỹ, khiến 8 triệu con lợn mới sinh tử vong (Stevenson *et al.*, 2013). PEDV được báo cáo tại Việt Nam lần đầu tiên vào năm 2009 và các nghiên cứu cho thấy chủng PEDV gây nên đợt bùng phát dịch bệnh có mối quan hệ họ hàng với các chủng PEDV có nguồn gốc Trung Quốc (Nguyễn Tất Toàn *et al.*, 2012; Duy *et al.*, 2013). Cho đến nay, hầu hết các khu vực chăn nuôi lợn tại cả nước đều ghi nhận sự có mặt của PEDV. Mặc dù đã được sử dụng vaccine nhưng nhiều trang trại vẫn xảy ra bệnh gây thiệt hại rất nặng nề. Từ đó cho thấy các chủng virus thực địa đã biến đổi về đặc tính di truyền, dẫn đến không còn tương đồng về tính kháng nguyên-miễn dịch với các chủng virus vaccine đang sử dụng.

PEDV là một loại virus RNA sợi đơn dương, thuộc phân chi *Pedecovirus*, chi *Alphacoronavirus* trong họ *Coronaviridae,* thuộc bộ *Nidovirales* (Schoch *et al.*, 2020). Virus có cấu trúc hình cầu hoặc đa hình với đường kính 95–190 nm, có vỏ bọc ngoài cùng bao gồm các phần nhô ra hình gậy, hình tam giác có chiều dài 18–23 nm. Bộ gen PEDV dài khoảng 28 kb, bao gồm 7 khung đọc mở (ORF), được sắp xếp theo thứ tự 5'UTR-ORF1a-ORF1b-S-ORF3-E-M-N-3'UTR và đuôi poly (A) (Lee 2019). ORF1a, ORF1b chiếm khoảng 2/3 bộ gen, ở vị trí gần đầu 5' và mã hóa 16 protein phi cấu trúc (nsps). ORF3 nằm giữa gen S và E có vai trò mã hóa các protein phụ, đồng thời là yếu tố quyết định độc lực ở PEDV. Gen E giúp hình thành và giải phóng vỏ virus (Park *et al.*, 2008). Gen S mã hóa glycoprotein, được chia thành miền S1 và S2, đóng vai trò quan trọng trong liên kết với các thụ thể tế bào. Đặc biệt vùng gen S1 kích thích vật chủ sản xuất kháng thể trung hòa và là vùng có nhiều đột biến nhất trong hệ gen nên được sử dụng chủ yếu cho các nghiên cứu về dịch tễ học phân tử, phả hệ nguồn gốc và phát triển vaccine thế hệ mới (Sun *et al.*, 2007). Mặc dù PEDV được báo cáo chỉ có một kiểu huyết thanh, nhưng về mặt di truyền được chia thành hai nhóm: nhóm 1 (Genotype1: là nhóm cổ điển) và nhóm 2 (Genotype 2: là nhóm độc lực cao) (Li *et al.*, 2012). Việc bổ sung thêm các nghiên cứu về đặc điểm phân tử, di truyền của virus là vô cùng cần thiết để phục vụ cho công tác phòng bệnh. Để góp phần tìm hiểu rõ hơn về các chủng PEDV đang lưu hành, chúng tôi tiến hành giải mã và phân tích gen kháng nguyên S1 của một số chủng virus thực địa thu

thập tại Hưng Yên năm 2023. Đây là một trong các tỉnh phía Bắc có ngành chăn nuôi lợn phát triển cũng như bị ảnh hưởng nặng nề bởi dịch bệnh. Kết quả thu nhận được phân tích và so sánh với các chủng PEDV tham chiếu trên thế giới.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Mẫu bệnh phẩm

Mẫu bệnh phẩm là phân và niêm mạc ruột của lợn bệnh khoảng 2 đến 3 tuần tuổi có triệu chứng nôn, đi ngoài phân lỏng bị ốm hoặc đã chết. Mẫu bệnh phẩm được mổ khám tại các trang trại và giữ lạnh trước khi chuyển đến phòng Miễn dịch học, Viện Công nghệ Sinh học. Mẫu bệnh phẩm được pha loãng bằng đệm PBS để có dạng huyền phù. Li tâm hỗn hợp ở tốc độ 2000 vòng/phút trong 30 phút. Dịch nổi phía trên được dùng để tách chiết RNA tổng số sử dụng bộ kit QIAamp Viral RNA Mini Kit (Qiagen, Đức) theo hướng dẫn của nhà sản xuất. Các mẫu được kiểm tra dương tính PEDV bằng cặp mồi chẩn đoán sẽ được lựa chọn để giải mã và phân tích gen S1.

Tổng hợp cDNA, PCR và dòng hóa

cDNA được tổng hợp từ RNA virus (khoảng 1 μg) bằng bộ kit RevertAid First Strand cDNA Synthesis (Thermo, Mỹ) theo hướng dẫn của nhà sản xuất. Phản ứng PCR được thực hiện nhằm chẩn đoán PEDV, sử dụng cặp mồi: PEDVF (5'-TTCTGAGTCACGAACAGCCA-3'), PEDVR (5'-CATATGCAGCCTGCTCTGAA-3') nằm trên gen S (Park *et al.*, 2008) thu sản phẩm PCR có kích thước 651 bp. Cặp mồi thứ 2 nhằm khuếch đại phân đoạn DNA chứa toàn bộ gen S1 gồm: PEDVS1F (5'-GCTAGTGCGTAATAATGACGCCA-3') và PEDVS1R (5'-ACAGAGCCTGTGTGGTGTA-3') thu sản phẩm PCR có kích thước khoảng 2,4 kb. Phản ứng PCR được thực hiện trên tổng thể tích 50 μL bao gồm: 25 μL DreamTaq PCR Master Mix (2X), 2 μL mỗi loại mồi (10 pmol/μL) và 3 μL cDNA khuôn và nước tinh khiết cho đủ 50 μL. Chu trình nhiệt của phản ứng PCR như sau: 95°C – 5 phút, tiếp theo là 35 chu kỳ [95°C – 30 giây, 55°C – 30 giây, 72°C – 3 phút] và 72°C – 10 phút. Sản phẩm PCR dương tính có chất lượng tốt được tinh sạch bằng bộ kit GeneJET PCR Purification Kit (Thermo, Mỹ) sau đó gắn vào vector tách dòng bằng bộ kit TA CloningTM Kit (Thermo, Mỹ). Sản phẩm ligation được chuyển nạp vào tế bào khả biến DH5α.

Phân tích trình tự gen và phả hệ nguồn gốc

DNA plasmid tái tổ hợp được giải trình tự bằng kit BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Thermo fisher, Mỹ). Để thu được toàn bộ chuỗi DNA từ DNA plasmid trên, chúng tôi thiết kế thêm một mồi bên trong gen S1 (PEDVS-F1: 5'-AATTGCATTGGTATGCTGC-3') để giải trình tự. Chuỗi nucleotide được xử lý bằng chương trình Seqed1.3, so sánh bằng chương trình AssemblyLIGN1.9 và MacVecter8.2 (Accelrys Inc). Các trình tự tương ứng với vùng gen S1 đăng ký tại Ngân hàng gen được sử dụng để so sánh đối chiếu với chuỗi gen nghiên cứu, sử dụng chương trình GENEDOC2.7 (http://www.nrbsc.org/gfx/genedoc/). Phân tích phả hệ nguồn gốc bằng chương trình MEGAX sử dụng phương pháp kết nối liền kề (Neighbor-joinning) với giá trị bootstrap 1000 lần lặp lại (Kumar *et al.*, 2018).

KẾT QUẢ VÀ THẢO LUẬN

Kết quả sàng lọc mẫu bệnh phẩm có chứa PEDV

Bằng cặp mồi chẩn đoán, chúng tôi phát hiện 5 mẫu dương tính với PEDV, là các mẫu cho sản phẩm PCR kích thước khoảng 0,6 kb đúng như dự tính (Hình 1).



Hình 1. Kết quả điện di phát hiện PEDV bằng cặp mồi chẩn đoán PEDVF-PEDVR

Giếng M: Thang DNA chuẩn (DNA của thực khuẩn thể λ được cắt bằng enzyme HindIII. Giếng 1, 2, 3, 4, 5, 6, 7: Kết quả PCR bằng cặp mồi chẩn đoán của các mấu PEDV thu nhận tại các trang trại đại diện cho các mẫu nghiên cứu. Trong năm mẫu dương tính, chúng tôi lựa chọn hai mẫu có chất lượng tốt nhất và đại diện cho hai huyện khác nhau của Hưng Yên (là mẫu số 2 và mẫu số 6) để giải mã gen S1. Mẫu được ký hiệu là PEDVHY1 và PEDVHY2. Hai mẫu này sẽ được dùng để nhân gen S1 bằng cặp mồi đặc hiệu PEDVS1F - PEDVS1R.

Kết quả thu nhận và giải trình tự gen S1

Sản phẩm PCR gen S1 của hai chủng PEDVHY1 và PEDVHY2:

Kết quả điện di sản phẩm PCR gen S1 và sản phẩm cắt DNA plasmid tái tổ hợp bằng enzyme giới hạn *Eco*RI được trình bày trong Hình 2. Sản phẩm PCR thu được từ mẫu 2 và mẫu 6 (tương ứng với mẫu bệnh phẩm ký hiệu PEDVHY1 và PEDVHY2) có chất lượng tốt, dòng hóa thành công vào vector pCR2.1. DNA plasmid tái tổ hợp được cắt kiểm tra bằng enzyme giới hạn *Eco*RI cho hai băng, một băng có kích thước 3.9 kb là vector pCR2.1, một băng có kích thước khoảng 2.4 kb bằng kích thước sản phẩm PCR gen S1 đem dòng hóa. Các DNA plasmid này sẽ được chọn để giải trình tự.





Giếng M: thang DNA chuẩn (DNA của thực khuẩn thể λ được cắt bằng enzyme HindIII).

Hình 2A: Giếng 1: kết quả PCR khuếch đại gen S1 của mẫu PEDVHY1; Giếng 2: kết quả PCR khuếch đại gen S1 của mẫu PEDVHY2

Hình 2B: Giếng 1: Kết quả điện di DNA plasmid của chủng PEDVHY1 cắt bằng enzyme giới hạn EcoRI; Giếng 2: Kết quả điện di DNA plasmid của chủng PEDVHY2 cắt bằng enzyme giới hạn EcoRI.

DNA plasmid được giải trình tự bằng phương pháp Sanger. Kết quả đã thu nhận được toàn bộ gen S1 của hai chủng PEDVHY1 và PEDVHY2, gồm 2205 nucleotide mã hóa cho 735 amino acid.

Trình tự amino acid (suy diễn) của hai chủng nghiên cứu được so sánh với các chủng: XJ-YLGL-2022021, IBT-VN, HBEZ3, X-J-WLMQ, KNU-1308, DR13 wild type (G1a), CV777 wild type (G1a), DR13 nhược độc (G1b), CV777 nhược độc (G1b) đang được sử dụng làm vaccine hiện nay (Bảng 1).

Kết quả cho thấy hai chủng nghiên cứu chỉ sai khác nhau ở bốn vị trí amino acid (L120R, E484P, K638E và P720S) trên protein S1 trong khi chứa nhiều sai khác lớn (xóa đoạn, chèn đoạn) so với các chủng virus nhược độc vaccine.

Protein S1 chứa tổng số 71 vị trí sai khác về amino acid giữa các chủng đại diện cho các genotype. Trong nghiên cứu này, chúng tôi tìm hiểu các sai khác amino acid giữa chủng CV777 thực địa (AF353511) (G1a), CV777 nhược độc vaccine (KT323979) (G1b), DR13 thực địa (JQ023162) (G1b) thu nhận tại Hàn Quốc năm 2009, DR13 nhược độc vaccine ((JQ023162) (G1b), HBZ3 thực địa (KY775054) thu nhận tại Trung Quốc năm 2016 (G2c), XJ-WLMQ-202203 (OR026668) thu nhận tại Trung Quốc năm 2022 (G2a), KNU-1308 (KJ451043) thu nhận tại Hàn Quốc năm 2013 (G2b), IBT-VN (MT198679) thu nhận tại Việt Nam năm 2018 và hai chủng nghiên cứu là PEDVHY1, PEDVHY 2 thu nhận tại tỉnh Hưng Yên năm 2023.

Tiểu phần protein S1 có chứa chuỗi peptide tín hiệu (SP), hai vị trí liên kết thụ thể thụ thể đầu tận N (S1-NTD) và thụ thể đầu tận C (S1-CTD) bao trùm cả vùng COE (CO-26K equivalent).

Trong vùng peptide tín hiệu, các chủng virus thuộc genotype G1 chứa ba vị trí sai khác về amino acid so với các chủng virus thuộc genotype G2 (aa 2: R/T↔K/S, aa 5: I↔N/T, aa 15: P/L↔S).

Trong vùng liên kết thụ thể thứ nhất (S1-NTD) phát hiện 34 sai khác lớn về amino acid giữa các genotype. Đồng thời xuất hiện đột biến chèn amino acid của chủng virus thực địa so với chủng vaccine tại vị trí 56 (I/T), 139 (D/N)

và 157 (Y/H). Tại vị trí amino acid 59 – 61, chủng thực địa được chèn thêm 3 amino acid (HGV/QGV) so với chủng vaccine. Trong khi đó, chủng thuộc G1a và G1b lại có thêm 2 amino acid (DI/NI) ở vị trí 163 – 164 so với chủng thuộc G2a, G2b, G2c. Tại các vị trí khác là các đột biến điểm đặc trưng giữa các genotype (Bảng 1). Các đột biến chèn/xóa nằm chủ yếu ở vùng siêu biến đổi của chuỗi peptide S1.

Trong vùng liên kết thụ thể thứ 2 (S1-CTD) chứa 7 sai khác, trong đó chủng CV777 (G1a) thực địa chứa nhiều sai khác nhất so với các chủng còn lại. Tại vị trí 554, 599, 638, chủng virus vaccine hoàn toàn khác với hai chủng virus nghiên cứu. Đặc biệt, ở vị trí 638 chỉ riêng chủng PEDVHY1 là K trong khi các chủng khác thuộc cùng genotype 2 là E và các chủng thuộc genotype 1 là Q/E.

Epitope COE-CO26K là một trong những epitope trung hòa có trên protein S1 của PEDV. Amino acid tại vùng epitope này có sự đương đồng cao giữa các chủng PEDV thuộc nhóm G2 nhưng có nhiều vị trí sai khác so với nhóm G1. Các chủng PEDV thuộc nhóm G2 là những chủng có độc lực cao, đang lưu hành hiện nay tại nhiều quốc gia trên thế giới. Kết quả nghiên cứu cho thấy hai chủng PEDV nghiên cứu là những chủng có độc lực cao, có nguy cơ lây lan cao tại các trang trại chăn nuôi lợn.

Những sai khác kể trên giữa các chủng vaccine và chủng thực địa có thể là nguyên nhân dẫn đến việc virus tăng hay giảm khả năng liên kết với tế bào chủ, ảnh hưởng đến khả năng bảo hộ của vaccine.

So sánh tỷ lệ đồng nhất nucleotide dựa trên trình tự gen S1 cho thấy, tỷ lệ đồng nhất nucleotide giữa chủng PEDVHY1 và chủng PEDVHY2 cao, 99.66%. Tuy nhiên, khi so sánh với một số chủng PEDV phân lập tại Việt Nam giai đoạn 2013 đến 2018 thì tỷ lệ tương đồng đạt thấp hơn, dao động từ 96.55% đến 97.32%. Điều này cho thấy đã có sự biến đổi trong gen kháng nguyên S1 của PEDV thực địa đang lưu hành tại Việt Nam so với các chủng trước đây.

Bảng 1. Vị trí sai khác amino acid thuộc protein S	1 giữa các chủng PEDV đại diện các genotype
và hai chủng	nghiên cứu

Vị trí	1	-SF	P-19											S 1	-NTC)									
(CV777(G1a)	2	5	10	15	27-29	30	56	57	59-61	62	64	68-72	74	82	84	86	87	89	118	120	132	133	138	139	157
CV777(G1a)	R	Ι	L	Ρ	QST	Т	-	М		S	S	GTGIE	А	L	Y	D	S	Q	Ν	Ι	К	Т	V	-	Y
DR3(G1a)	R	I	F	L	QST	I	-	М		s	S	GTGIE	D	L	Y	D	S	Q	Ν	Т	к	т	V	-	Y
CV777 (at) (G1b)	т	I	F	L	QST	I	-	М		S	S	GTGIE	D	L	Y	D	S	Q	S	Т	к	т	V	-	-
DR3(at) (G1b)	т	I	F	L	QST	Т	-	М		S	S	GTGIE	D	L	Y	D	S	Q	S	Т	к	т	V	-	-
HBEZ3 (G2c)	к	Ν	F	S	QST	Т	-	М		S	S	GTGIE	А	L	Y	D	А	Q	Ν	Т	к	т	V	-	Y
XJ-WLMQ (G2a)	к	Т	F	S	SAN	т	Т	Е	HGV	Ν	т	AGQHP	А	L	н	R	G	н	Ν	т	Ν	т	А	D	н
KNU-1308 (2b)	S	Т	F	S	SAN	т	Т	Е	QGV	Ν	т	AGQPH	А	V	н	R	G	н	Ν	т	к	т	А	Ν	н
IBT-VN	S	Т	F	S	SAN	т	Т	Е	QGV	Ν	т	AGQPH	А	L	н	R	G	н	Ν	т	к	т	А	D	н
PEDVHY1	S	Т	F	S	SAN	т	т	Е	QGV	Ν	т	AGQHP	А	L	н	R	G	н	Ν	т	к	А	А	Ν	н
PEDVHY2	s	т	F	S	SAN	т	т	Е	QGV	Ν	т	AGQHP	А	L	н	R	G	н	Ν	т	к	А	А	Ν	н

Vị trí	S1-NTD									S1-CTD (COE-CO26K)										
(CV777(G1a)	158	159	160-162	163-164	187	201-203	211	228	230-231	237	248-249	270	522	541	554	599	610	617	638	640
CV777(G1a)	М	R	DGK	DI	L	RRS	Т	Y	EP	Т	DS	L	А	F	Т	G	А	L	Е	I
DR3 (G1a)	L	Q	DGK	NI	L	NRS	т	Y	EP	S	DS	L	А	F	т	G	Е	F	Е	V
CV777 (at) (G1b)	L	Q	DGK	NI	Т	NRS	т	Y	EP	S	DS	L	А	F	т	G	Е	F	Q	V
DR3(at) (G1b)	L	Q	DGK	NI	Т	NRS	т	Y	EP	S	DS	L	А	F	т	G	Е	F	Q	V
HBEZ3 (G2c)	М	Q	DGK	NI	L	NRS	т	Y	EP	S	DS	L	S	F	S	S	Е	F	Е	V
XJ-WLMQ (G2a)	М	S	EHS		F	SGG	Е	S	QL	L	EL	L	S	L	S	S	Е	F	Е	V
KNU-1308 (2b)	М	S	EHS		F	SGG	Е	S	QP	I	EP	L	S	F	S	S	Е	F	Е	V
IBT-VN	М	S	EHS		F	SGG	Е	S	QP	I	EP	L	А	F	S	S	Е	F	Е	V
PEDVHY1	М	s	EHS		F	SGG	Е	S	QP	Т	EP	V	А	F	S	S	Е	F	к	V
PEDVHY2	М	S	EHS		F	SGG	Е	S	QP	Т	EP	V	А	F	s	S	Е	F	Е	V

Chú thích: vị trí amino acid trong bảng 2.1 là vị trí của chủng CV777 thực địa (AF353511).

So sánh với các chủng thuộc genotype G1 (cường độc và vaccine), hai chủng PEDV phân lập tại Hưng Yên năm 2023 có tỷ lệ tương đồng thấp (từ 90.79% đến 91.12%). Kết quả cho thấy có sự khác biệt đáng kể về thành phần nucleotide giữa các chủng PEDV thuộc các genotype khác nhau, từ đó thấy được tầm quan trọng của công tác giám sát dịch tế học phân tử và sử dụng vaccine phòng bệnh PED tại Việt Nam.

KẾT QUẢ PHÂN TÍCH PHẢ HỆ NGUỒN GỐC

Cây phả hệ được xây dựng dựa trên trình tự gen S1 của 54 chủng PEDV, trong đó có hai chủng nghiên cứu PEDVHY1 và PEDVHY2, 52 chủng còn lại đại diện cho các genotype đã và đang lưu hành trên toàn cầu. Các chủng tham chiếu được tải xuống từ Ngân hàng gen NCBI theo mã số GenBank của từng chủng.

Cây phả hệ gồm năm nhóm chính tương ứng với năm genotype: G1a, G1b, G2a, G2b, G2c (Hình 3).

Nhóm thứ nhất gồm 29 chủng PEDV thuộc genotype G2b, bao gồm các chủng virus phân lập tại nhiều quốc gia bao gồm: Mỹ, Hàn Quốc, Mexico, Trung Quốc, Việt Nam (từ năm 2013 đến năm 2018), và hai chủng PEDVHY1 và PEDVHY2 thu nhận mới năm 2023 tại tỉnh Hưng Yên trong nghiên cứu này. Điều này chứng tỏ sự hiện diện của PEDV G2b khắp toàn cầu trong thời gian dài và tồn tại cho tới tận ngày nay.

Nhóm thứ hai gồm các chủng thuộc genotype G2a. Nhóm này gồm các chủng của Trung Quốc phân lập trong khoảng 4 năm trở lại đây gây ra nhiều ổ dịch lớn tại Trung Quốc và lan sang các nước lân cận. Tuy nhiên, tại Việt Nam vẫn chưa ghi nhận thấy sự lưu hành phổ biến của các chủng PEDV thuộc genotype G2a này.

Nhóm thứ ba gồm các chủng PEDV thuộc genotype G1a. Nhóm này chứa chủng PEDV cổ điển CV777 phân lập từ năm 1978 tại Bỉ. Ngoài ra còn có các chủng phân lập tại Hàn Quốc năm 1998 và phân lập tại Đan Mạch phân lập năm 2017. Chủng Avac-PEDV-98 phân lập năm 2021 của Việt Nam thuộc nhóm này.

Nhóm thứ tư gồm các chủng PEDV thuộc genotype G1b. Nhóm này chứa các chủng virus vaccine nhược độc được biến đổi từ các chủng cường độc thuộc genotype G1a. Trong đó chủng CV777 dại (genotype G1a) sau quá trình nhược độc hóa thành chủng CV777 nhược độc thuộc G1b, chủng DR13 có nguồn gốc G1a, sau khi cấy chuyển 100 đời đã thành chủng nhược độc DR13 thuộc G1b. Đây là hai trong số nhiều chủng virus vaccine đang được sử dụng rộng rãi hiện nay.

Nhóm thứ năm gồm các chủng PEDV thuộc genotype G2c. Nhóm này gồm các chủng thu nhận tại Mỹ, Pháp, Trung Quốc từ khoảng những năm 2010 trở lại đây.

Hai chủng PEDV nghiên cứu (PEDVHY1 và PEDVHY2) phân lập năm 2023 tại tỉnh Hưng Yên thuộc genotype G2b, không cùng nhóm với hai chủng virus vaccine CV777 và DR13. Điều này là nguyên nhân dẫn đến việc đàn lợn vẫn mắc bệnh mặc dù đã sử dụng vaccine để phòng bệnh.



Hình 3. Phả hệ nguồn gốc PEDV dựa trên trình tự gen S1

Chú thích: Các nhóm khác nhau được đánh dấu bằng màu nền khác nhau. Chủng nghiên cứu PEDVHY1 và PEDVHY2 được đánh dấu bằng hình mũi tên.

KÉT LUẬN

Nghiên cứu đã giải mã và phân tích trình tự gen S1 của hai chủng PEDV thu nhận tại tỉnh Hưng Yên năm 2023. Kết quả phân tích phả hệ nguồn gốc cho thấy cả hai chủng PEDV nghiên cứu đều thuộc genotype G2b, là chủng virus đang lưu hành phổ biến hiện nay ở Việt Nam và trên thế giới. Hai chủng PEDV thực địa có 45 vị trí sai khác về amino acid trong protein kháng nguyên S1 so với các chủng PEDV thuộc các genotype khác và với hai chủng virus vaccine đang sử dụng hiện nay.

Lời cảm ơn: Nghiên cứu được tài trợ từ Nhiệm vụ Quỹ gen cấp quốc gia "Khai thác và phát triển nguồn gen virus (CSFV và PEDV) để chế tạo kit chẩn đoán và phục vụ sản xuất vắc-xin phòng bệnh dịch tả lợn cổ điển và bệnh tiêu chảy cấp ở lợn" mã số NVQG-2023/DT.04 do TS. Đoàn Thị Thanh Hương chủ nhiệm.

TÀI LIỆU THAM KHẢO

Duy DT, Toan NT, Puranaveja S, Thanawongnuwech R (2013). Genetic characterization of porcine epidemic diarrhea virus (PEDV) isolates from southern Vietnam during 2009-2010 outbreaks. *Thai J Vet Med*, 41: 55-56.

Lee C (2019). Porcine Viruses: From Pathogenesis to Strategies for Control. Zakaryan H, editor. Norfolk: Caister Academic Press; Porcine epidemic diarrhoea virus: 107-134.

Li W, van Kuppeveld FJM, He Q, Rottier PJM, Bosch BJ (2016). Cellular entry of the porcine epidemic diarrhea virus. Virus Res, 226: 117-127.

Li W, Li H, Liu Y, Pan Y, Deng F, Song Y, Tang X, He Q (2012). New variants of porcine epidemic diarrhea virus, China, 2011. *Emerg Infect Dis*, 18: 1350-1353.

Park SJ, Moon HJ, Luo Y, Kim HK, Kim EM, Yang JS, Song DS, Kang BK, Lee CS, Park BK (2008). Cloning and further sequence analysis of the ORF3 gene of wild- and attenuated-type porcine epidemic diarrhea viruses. *Virus Genes*, 36: 95-104.

Pensaert M, De Bouck P (1978) A new coronavirus-like particle associated with diarrhea in swine. Arch Virol, 58: 243-247.

Pospischil A, Stuedli A, Kiupel M (2002). Update on porcine epidemic diarrhea. J Swine Health Prod, 10: 81-85.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution*, *35*(6):1547-1549.

Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, Sharma S, Soussov V, Sullivan JP, Sun L, Turner S, Karsch-Mizrachi I (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxford)*: baaa062.

Stevenson GW, Hoang H, Schwartz KJ, Burrough ER, Sun D, Madson D, Cooper VL, Pillatzki A, Gauger P, Schmitt BJ, Koster LG, Killian ML, Yoon KJ (2013). Emergence of Porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J Vet Diagn Investig of Publ Am Assoc Vet Lab Diagn Inc*, 25(5): 649-654.

Sun DB, Feng L, Shi HY, Chen JF, Liu SW, Chen HY (2007). Spike protein region (aa 636-789) of porcine epidemic diarrhea virus is essential for induction of neutralizing antibodies. *Acta Virol.* 51(3): 149-156.

Temeeyasen G, Srijangwad A, Tripipat T, Tipsombatboon P, Piriyapongsa J, Phoolcharoen W, Chuanasa T, Tantituvanont A, Nilubol D (2013). Genetic diversity of ORF3 and spike genes of porcine epidemic diarrhea virus in Thailand. *Infect Genet Evol*, 21: 205-213.

SEQUENCING AND ANALYSIS S1 GENE OF PORCINE EPIDEMIC DIARRHEA VIRUS ISOLATED IN HUNG YEN IN 2023

Luu Minh Duc^{1,2}, Do Thi Roan^{1,2}, Nguyen Thi Khue^{1,2}, Don Nguyen³, Doan Thi Thanh Huong^{1,2*}

¹Institute of Biotechnology, Vietnam Academy of Science and Technology

²Graduate University of Science and Technology, Vietnam Academy of Science and Technology

³Squalicum High School Bellingham, Washington, USA

SUMMARY

PED virus causes acute diarrhea mainly in unweaned piglets. The infection rate in herds and the mortality rate are high, causing a lot of damage to the livestock industry. The genome of virus is about 28 kb in size, encodes important genes for replication and complete virus creation. The antigen protein spike contains an S1 subunit contains a signal protein and two receptor-binding domains. The S1 genes of both strains PEDVHY1 and PEDVHY2 collected in Hung Yen in 2023 are 2205 nucleotides in size, encoding 735 amino acids and belong to genotype G2b. The S1 protein contains many mutations, especially insertion and deletion mutations compared to vaccine strains. The S1 gene region of the two studied trains has little difference, while they have many mutations compared to other genotypes. Phylogenetic analysis shows that the two studied PEDV strains are closely related in the family tree and are in the same group with Vietnamese strains collected from 2013 to 2018 belonging to genotype G2b, different genotype from attenuated vaccine virus strains: DR13 and CV777.

Keywords: S1 gene, genotype, Hung Yen, PEDV.

^{*} Author for correspondence: Tel: 0988904605; Email: doantthuong74@gmail.com

ĐÁNH GIÁ ĐA DẠNG DI TRUYỀN CỦA MỘT SỐ GIỐNG LẶC LÀY (*Trichosanthes cucumerina* L.)

Trương Thị Hồng Hải^{1*}, Hồ Thị Hoàng Nhi¹, Sonexay Rasphone², Hồ Ngọc Hân¹

¹Viện Công nghệ sinh học, Đại học Huế, Việt Nam

²Trường Đại học Savannakhet, Lào

TÓM TẮT

Lặc lày (*Trichosanthes cucumerina* L.) là một loại rau có giá trị dinh dưỡng cao và cũng là một loại thảo dược trị bệnh quan trọng. Ở Việt Nam, loài cây này được trồng và sử dụng ngày càng nhiều. Để nâng cao năng suất và chất lượng của lặc lày, chọn lọc các giống địa phương và giống lai F1 mang các ưu thế lai phù hợp với điều kiện Việt Nam nói chung và Thừa Thiên Huế nói riêng cân được chú trọng phát triển để phục vụ công tác chọn tạo giống trong thời gian tới. Tuy nhiên, các kiến thức về đa dạng di truyền của lặc lày vẫn còn rất hạn chế ở Việt Nam. Ở nghiên cứu này, bốn giống lặc lày, trong đó có 1 giống địa phương được thu thập ở Savannakhet, Lào; 2 giống F1 của Thái Lan và 1 giống F1 của Việt Nam, đã được tách chiết, tinh sạch DNA, sau đó các mồi RAPD được sàng lọc và thực hiện chỉ thị RAPD để đánh giá sự khác biệt di truyền giữa bốn giống. Các chỉ số PIC, EMR, MI, Rp, na, ne, h và I đã cho thấy 18 mồi UBC-RAPD được chọn lọc là thích hợp cho nghiên cứu sự đa dạng di truyền. Đã có 199 băng khuếch đại được tạo ra, trong đó 91 băng là băng đa hình. Mồi UBC#475 cho thấy mức độ đa hình cao nhất với các chỉ số PIC, EMR, MI, Rp lần lượt là 0,357; 10,286; 3,673 và 8,000. Ngoài ra, kết quả phân tích cây phân cụm dựa vào phương pháp UPGMA đã chỉ ra rằng giống lặc lày lai F1 (HUIB_Tc2) có sự khác biệt di truyền so với 3 giống còn lại là giống địa phương (HUIB_Tc1) và hai giống thuần (HUIB_Tc2 và HUIB_Tc3).

Từ khóa: Chỉ thị DNA, đa dạng di truyền, lặc lày, RAPD, Trichosanthes cucumerina.

MỞ ĐẦU

Lặc lày (*Trichosanthes cucumerina* L.) thuộc họ Cucurbitaceae, là một loài cây leo thường được dùng làm rau và thuốc thảo dược (Yang *et al.*, 2023). Chúng phân bố nhiều ở các vùng ôn đới châu Á như Trung Quốc và vùng nhiệt đới như Bangladesh, Ấn Độ, Nepal, Pakistan, Sri Lanka, Myanmar, Việt Nam, Indonesia; Malaysia, Philippines và Australia (Sandhya *et al.*, 2010).

Đây là loài cây giàu chất dinh dưỡng với 18 amino acid đã được tìm thấy trong quả, bao gồm: glycine, alanine, serine, proline, valine, threonine, isoleucine, leucine, aspartate, lysine, methionine, histidine, arginine, glutamate, phenylalanine, tyrosine, tryptophan và cystine (Osuagwu *et al.*, 2022). Đồng thời, chúng cũng chứa hàm lượng rất cao các chất béo, chất xơ, carb, khoáng chất, vitamin A và vitamin E. Tác dụng dược lý và trị liệu của lặc lày là nhờ vào flavonoid, carotenoids, phenolic acid, chất xơ hòa tan và không hòa tan và các khoáng chất quan trọng có rất nhiều trong cây. Trong y học cổ truyền của Sri Lanka và Ấn Độ, lặc lày được sử dụng trong điều trị khó tiêu; sốt; mụn nhọt; vết loét và các bệnh phát ban trên da như bệnh chàm, viêm da, vẩy nến, viêm, loét và tiểu đường. Bên cạnh đó, trong toàn bộ rễ, lá, quả và hạt của lặc lày có hoạt tính kháng khuẩn, chống viêm, trị giun sán, bảo vệ dạ dày và chống oxy hóa (Bobade *et al.*, 2022).

Lặc lày được trồng và sử dụng ngày càng rộng rãi ở nhiều nước Đông Nam Á. Ở Việt Nam, lặc lày còn có các tên dân dã khác như mướp nhật, lặc lè, mướp rừng hay bầu rắn. Trên thị trường các nước Việt Nam, Lào và Thái Lan, hiện có nhiều giống lặc lày với màu sắc, hình thái, kích thước khác nhau, và tồn tại ở nhiều dạng như giống địa phương, giống thuần và giống lai F1. Điều này tạo nên sự đa dạng biến dị trong quá trình phát triển giống. Hiện nay, xu hướng tạo ra các giống lặc lày F1 mang ưu thế lai về năng suất và chất lượng ngày càng được phát triển (Islam *et al.*, 2022). Tuy nhiên để chọn lọc và tạo ra các giống ưu việt này, việc nắm vững kiến thức về đặc điểm di truyền là điều kiện tiên quyết (Ilakiya *et al.*, 2022). Để nâng cao năng suất và chất lượng của lặc lày, chọn lọc các giống địa phương và giống lai F1 mang các ưu thế lai phù hợp với điều kiện Việt Nam nói chung và Thừa Thiên Huế nói riêng cần được chú trọng phát triển để phục vụ công tác chọn tạo giống trong thời gian tới.

Trong số các loại chỉ thị DNA, RAPD (Random Amplified Polymorphic DNA) là chỉ thị trội thường được sử dụng trong nghiên cứu đa dạng di truyền với nhiều ưu điểm như: nhanh chóng, tiết kiệm, đơn giản, hiệu quả; chỉ cần một lượng nhỏ DNA khuôn mẫu và không cần biết trước trình tự DNA cũng như trình tự mồi (Ghorpade *et al.*, 2022). Bên cạnh đó, trên đối tượng chi *Trichosanthes* đã có nhiều công bố cho thấy việc ứng dụng chỉ thị RAPD trong phân tích đa dạng di truyền (Goswami *et al.*, 2009; Adhikari *et al.*, 2014). Chính vì vậy, trong nghiên cứu này chúng tôi đã sử dụng chỉ thị RAPD để đánh giá sự đa dạng di truyền của bốn giống lặc lày (giống địa phương, giống thuần và giống lai F1) được thu thập từ ba nước Việt Nam, Lào và Thái Lan.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Bốn giống lặc lày sử dụng trong nghiên cứu này đã được thu thập từ ba nước Việt Nam, Lào và Thái Lan (Bảng 1). Trong đó, có một giống địa phương (HUIB_Tc1), hai giống thuần (HUIB_Tc2 và HUIB_Tc3) và một giống lai F1 (HUIB_Tc4). Bên cạnh đó, 100 mồi UBC-RAPD (University of British Columbia, Công ty Bioneer, Korea) và các hóa chất cần thiết đã được cung cấp bởi Trung tâm Nghiên cứu và Ứng dụng kỹ thuật Y sinh tiên tiến, Viện Công nghệ sinh học, Đại học Huế.

STT	Mã Code	Tên giống	Nguồn gốc của giống
1	HUIB_Tc1	Lặc lày địa phương	Savannakhet, Lào
2	HUIB_Tc2	Lặc lày AAA	Công ty TNHH CHUA YONG SENG SEED, Bangkok, Thái Lan
3	HUIB_Tc3	NAKEE	Công ty TNHH SEEDLIND, Nakompathom, Thái Lan
4	HUIB_Tc4	Lặc lày lai F1 RADO 248	Công ty TNHH Hạt giống Rạng Đông, Hồ Chí Minh, Việt Nam

Phương pháp

Tách chiết và tinh sạch genomic DNA

Genomic DNA từ lá lặc lày được tách chiết theo phương pháp CTAB (Cetyl trimethylammonium bromide) (Doyle và Doyle, 1987). Sau đó, genomic DNA thu được sẽ được nhuộm với SYBR Green I (Invitrogen, USA), điện di trên 1% agarose gel và đo OD bằng máy quang phổ Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, USA). Mẫu DNA có giá trị A260/A280 nằm trong khoảng 1,8-2,0 được coi là tinh khiết và đạt yêu cầu cho thí nghiệm. Trong khi đó, DNA chưa đạt sẽ được đưa qua cột silica và kiểm tra lại.

Chỉ thị RAPD

Để đánh giá sự đa dạng di truyền giữa các giống lặc lày ở các quốc gia khác nhau (Việt Nam, Lào và Thái Lan), chúng tôi đã chọn một mẫu genomic DNA lặc lày Việt Nam (HUIB_Tc4) và một mẫu có nguồn gốc từ nước khác (HUIB_Tc1) để khảo sát mồi. Đầu tiên, hai mẫu genomic DNA này được pha loãng về nồng độ khoảng 15-18 ng/µL để sàng lọc 100 mồi UBC-RAPD. Trong đó, 18 mồi đã được lựa chọn dựa vào sản phẩm khuếch đại DNA có số băng khuếch đại lớn, các băng rõ nét và có băng đa hình (Bảng 2). Các mồi này sau đó sẽ được sử dụng để PCR cho cả 4 giống lặc lày.

	•		
STT	Tên mồi UBC-RAPD	Trình tự 5'-3'	Nhiệt độ nóng chảy (T _m)
1	UBC#424	ACGGAGGTTC	32°C
2	UBC#428	GGCTGCGGTA	34°C
3	UBC#434	TCGCTAGTCC	32°C
4	UBC#442	CTACTCGGTT	30°C
5	UBC#452	CTAATCACGG	30°C
6	UBC#453	AGTACAAGGG	30°C
7	UBC#458	CTCACATGCC	32°C
8	UBC#460	ACTGACCGGC	34°C
9	UBC#463	AGGCGGAAGC	34°C
10	UBC#465	GGTCAGGGCT	32°C
11	UBC#466	TTCTTAGCGG	30°C
12	UBC#467	AGCACGGGCA	34°C
13	UBC#469	CTCCAGCAAA	30°C
14	UBC#475	CCAGCGTATT	30°C
15	UBC#476	TTGAGGCCCT	32°C
16	UBC#495	CTTTCCTTCC	30°C
17	UBC#497	GCATAGTGCG	32°C
18	UBC#498	GACAGTCCTG	32°C

Bảng 2. Trình tự của các mồi UBC-RAPD đã được chọn lọc

Mỗi phản ứng khuếch đại DNA (PCR) (10 μL) bao gồm các thành phần: 16,75 mM MgCl₂ (Bioline-Meridian, UK); 6,7 mM deoxyribonucleotide triphosphate mix (Bioline-Meridian, UK); 2X MyTaq Mix (Meridian Bioscience, USA); 10 pmol của mồi UBC-RAPD; 5-10 ng/μL của DNA tổng số và nước cất. Chu kỳ phản ứng được thực hiện lần lượt theo các bước: 95°C trong 5 phút (1 chu kỳ); 95°C trong 1 phút, 35°C trong 2 phút, 72°C trong 2 phút (40 chu kỳ) và 72°C trong 10 phút (1 chu kỳ). Sản phẩm PCR sau đó được nhuộm với SYBR Green I rồi điện di trên 2% agarose gel và quan sát dưới ánh sáng UV của hệ thống đọc gel (Vilber, Pháp). GeneRuler DNA Ladder Mix (Thermo Scientific) được sử dụng để ước lượng kích thước của các đoạn DNA (marker chuẩn).

Phân tích dữ liệu

Ma trận nhị phân được xây dựng bằng cách: ký hiệu số "1" đối với các băng DNA có xuất hiện (băng rõ ràng, không biến dạng), ký hiệu số "0" nếu không xuất hiện băng hoặc băng quá mờ. Kích thước mỗi băng DNA sẽ được ước tính dựa trên marker chuẩn. Dữ liệu ma trận này sau đó được dùng để tính toán các chỉ số đánh giá mồi như: tổng số băng (TB), số băng đa hình (PB), số băng đơn hình (MB), tỉ lệ băng đa hình (PPB (%)), chỉ số đa dạng di truyền - PIC (Polymorphism Information Content), chỉ số sai khác của mỗi cặp mồi - Rp (Resolving power), chỉ số đa dạng trung bình của các locus đa hình - MI (Marker Index) và chỉ số EMR (Effective multiplex ratio). Ngoài ra, ma trận này cũng được phân tích bằng phần mềm POPGENE 1.32 để tìm ra chỉ số mức độ đa dạng kỳ vọng (Nei's gene diversity - h), chỉ số đa dạng Shannon (Shannon's information index - I), số allele quan sát được (n_a) và số allele hữu hiệu (Effective number of allele - n_e) (Kumar *et al.*, 2014; Rasphone *et al.*, 2022). Tiếp đó, ma trận nhị phân được sử dụng để tính khoảng cách di truyền và phát triển cây phân cụm UPGMA trong phần mềm PHYLIP 3.698.

KẾT QUẢ VÀ THẢO LUẬN

Cả 4 giống lặc lày đều có băng genomic DNA đậm, rõ nét, không bị smear, nồng độ sau khi tinh sạch trên 20 ng/µL với giá trị A260/A280 trên 1,8 (Hình 1). Do đó, các DNA này thích hợp để sử dụng cho PCR.

18 mồi UBC-RAPD được chọn lọc trước đó đã được sử dụng để khuếch đại 4 mẫu DNA lặc lày. Kết quả cho thấy cả 18 mồi đều có khả năng khuếch đại DNA tốt, với 199 băng DNA được tạo thành (trung bình 11,056 băng/mồi). Có 6 mồi RAPD cho tỉ lệ băng đa hình trên 50% là UBC#434, UBC#442, UBC#452, UBC#463, UBC#475 và UBC#498. Trong đó, UBC#475 là mồi có số băng khuếch đại đa hình (PB) nhiều nhất (12 băng đa hình) và có tỉ lệ băng đa hình trên 50% là UBC#476 có tỉ lệ băng đa hình thấp nhất (14,286%). Có 91 băng đa hình được tạo ra ở mỗi mồi (trung bình 5,056 băng đa hình/mồi) và tỉ lệ các băng đa hình trung bình là khá thấp (chỉ 45,23%) (Hình 2, Bảng 3). Tỉ lệ đa hình này cao hơn so với nghiên cứu trước đó của Adhikari và đồng tác giả (2014) trên đối tượng *Trichosanthes dioica* R. (31,03%), nhưng thấp hơn so với kết quả nghiên cứu đa hình các giống *Trichosanthes anguina* L. do Alam và đồng tác giả (2011) tiến hành (55,13%). Sự chênh lệch này là do sự khác biệt về nguồn gốc đối tượng nghiên cứu.





Hình 1. Genomic DNA của 4 giống lặc lày sau khi tách chiết và tinh sạch



Trong đó: UBC#475 là mồi có nhiều băng đa hình nhất và UBC#476 là mồi có ít băng đa hình nhất. Ngôi sao dùng để chỉ vị trí các băng đa hình. M: GeneRuler DNA Ladder Mix (Thermo Scientific).
HỘI NGHỊ KHOA HỌC TOÀN QUỐC VỀ CÔNG NGHỆ SINH HỌC 2024

	Bảng 3. Các chỉ số đánh giá mồi UBC-RAPD và chỉ số đa dạng di truyền của quần thể 4 giống lặc lày												
STT	Tên mồi UBC-RAPD	тв	РВ	PPM (%)	MB	PIC	EMR	МІ	Rp	n _a	n _e	h	I
1	UBC#424	14	5	35,714	9	0,152	1,786	0,271	3,500				
2	UBC#428	14	7	50,000	7	0,205	3,500	0,719	4,500				
3	UBC#434	11	7	63,636	4	0,250	4,455	1.114	4,000				
4	UBC#442	8	6	75,000	2	0,375	4,500	1,688	6,000				
5	UBC#452	9	5	55,556	4	0,236	2,778	0,656	3,500				
6	UBC#453	9	3	33,333	6	0,167	1,000	0,167	3,000				
7	UBC#458	11	3	27,273	8	0,136	0,818	0,112	3,000				
8	UBC#460	10	5	50,000	5	0,225	2,500	0,563	4,000				
9	UBC#463	13	7	53,846	6	0,221	3,769	0,834	4,500				
10	UBC#465	11	2	18,182	9	0,080	0,364	0,029	1,500				
11	UBC#466	9	3	33,333	6	0,139	1,000	0,139	2,000				
12	UBC#467	17	6	35,294	11	0,147	2,118	0,311	4,000				
13	UBC#469	10	4	40,000	6	0,163	1.600	0,260	2,500				
14	UBC#475	14	12	85,714	2	0,357	10,286	3,673	8,000				
15	UBC#476	7	1	14,286	6	0,071	0,143	0,010	1,000				
16	UBC#495	13	5	38,462	8	0,163	1,923	0,314	3,500				
17	UBC#497	8	4	50,000	4	0,203	2,000	0,406	2,500				
18	UBC#498	11	6	54,545	5	0,216	3,273	0,707	3,500				
Т	rung bình	11,056	5,056	45,23	6,00	0,195	2,656	0,665	3,583	1,457	1,300	0,173	0,257
	SD									0,499	0,375	0,202	0,292

TB: Tổng số băng, PB: Số băng đa hình, MB: Số băng đơn hình, PPB: Tỉ lệ băng đa hình (%), PIC: Chỉ số đa dạng di truyền, Rp: Chỉ số sai khác của mỗi cặp mồi, MI: Chỉ số đa dạng trung bình của các locus đa hình, EMR: Chỉ số multiplex hữu hiệu, n_a: Số allele quan sát được, n_e: Số allele hữu hiệu, h: Chỉ số mức độ đa dạng kỳ vọng, I: Chỉ số đa dạng Shannon

Khi chỉ số MI càng cao thì việc ứng dụng một kĩ thuật nào đó (ở nghiên cứu này là chỉ thị RAPD) để đánh giá lượng lớn các băng được khuếch đại là hiệu quả hơn so với việc chỉ dựa vào các băng đa hình (Powell *et al.*, 1996). Trung bình chỉ số MI ở mỗi mồi là 0,665. Trong đó, chỉ số MI cao nhất là 3,673 ở mồi UBC#475, tiếp theo là mồi UBC#434 và UBC#442 với giá trị MI lần lượt là 1,114 và 1,688. Trong khi đó, UBC#476 là mồi có chỉ số MI thấp nhất (MI = 0,01) (Bảng 3).

Một chỉ thị phân tử hiệu quả là khi nó phân nhóm tốt các kiểu gen trong quần thể. Chỉ số Rp là đơn vị thể hiện sự tương quan giữa các kiểu gen với chỉ thị phân tử DNA, khi Rp càng cao thì chỉ thị phân tử phân nhóm kiểu gen càng hữu hiệu và ngược lại (Prevost và Wilkinson, 1999). Bên cạnh đó, chỉ số EMR cũng là thông số chỉ ra mức độ hiệu quả của mồi (Srisamoot và Padsri, 2018). Dựa vào đó, mồi UBC#475 là chỉ thị phân tử tốt nhất để phân chia kiểu gen của các giống lặc lày do có giá trị Rp và EMR cao nhất (Rp = 8,0 và EMR = 10,286) (Bảng 3).

Theo nghiên cứu trước đó của Botstein và đồng tác giả (1980), một mồi có chỉ số PIC > 0,5 thì được coi là có mức đa hình cao. Ngược lại, chỉ số PIC nằm trong khoảng 0,25 - 0,5 cho mức đa hình trung bình và với chỉ số PIC < 0,25 cho mức đa hình thấp. UBC#434, UBC#442 và UBC#475 là ba mồi duy nhất cho thấy hệ số PIC nằm trong khoảng 0,25 - 0,5 và thể hiện mức độ đa hình trung bình. Trong khi đó, kết quả phân tích tất cả các mồi cho thấy mức độ đa hình thấp, với chỉ số PIC trung bình là 0,195 và hầu hết các mồi đều có chỉ số PIC nhỏ hơn 0,25 (Bảng 3). Tuy nhiên, giá trị PIC trung bình này là khá tương đồng với công bố của Adhikari và đồng tác giả (2014) (PIC = 0,192). Đồng thời, các giá trị n_a, n_e, h và I lần lượt là 1,457; 1,300; 0,173 và 0,257 trong nghiên cứu của chúng tôi là khá thấp. Điều này cho thấy sự ít đa dạng di truyền trong quần thể bốn giống lặc lày. Mặc dù vậy, kết quả này là cao hơn so với kết quả của Adhikari và đồng tác giả (n_a = 1,3103; n_e = 1,1804; h = 0,1110 và I = 0,1670). Vì vậy, 18 mồi UBC-RAPD được sử dụng là hiệu quả, thể hiện được mức độ đa hình và thích hợp để đánh giá sự đa dạng di truyền của quần thể 4 giống lặc lày.

CÔNG NGHỆ GEN

	HUIB_Tc1	HUIB_Tc2	HUIB_Tc3	HUIB_Tc4					
HUIB_Tc1	0,0000	0,0195	0,0157	0,0217					
HUIB_Tc2	0,0195	0,0000	0,0169	0,0190					
HUIB_Tc3	0,0157	0,0169	0,0000	0,0191					
HUIB_Tc4	0,0217	0,0190	0,0191	0,0000					





Hình 3. Cây UPGMA thể hiện mối quan hệ di truyền giữa 4 giống lặc lày

Kết quả ở Bảng 4 cho thấy rằng sự khác biệt di truyền giữa 4 giống lặc lày là không nhiều. Khoảng cách di truyền giữa 4 giống dao động từ 0,0157 đến 0,0217. Giống HUIB_Tc1 và HUIB_Tc4 có sự khác biệt lớn nhất, với khoảng cách di truyền là 0,0217. Trong khi đó, giống HUIB_Tc1 và HUIB_Tc3 ít khác biệt nhất (khảng cách di truyền là 0,0157). Bên cạnh đó, cây phân cụm UPGMA chia các giống lặc lày thành hai nhóm, nhóm 1 bao gồm ba giống là HUIB_Tc1, HUIB_Tc2 và HUIB_Tc3; nhóm 2 chỉ gồm giống HUIB_Tc4 (Hình 3). Vì các giống thuộc cùng một nhóm trên cây UPGMA sẽ cho thấy sự tương đồng di truyền cao hơn nên có thể nhận định rằng giống HUIB_Tc4 có sự khác biệt di truyền nhiều hơn so với 3 giống còn lại. Nguyên nhân để giải thích là vì HUIB_Tc4 là giống lai F1, do đó nó mang nhiều biến dị hơn so với các giống địa phương (HUIB_Tc1) và giống thuần (HUIB_Tc2 và HUIB_Tc3). Trong công bố trước đó của Islam và đồng tác giả (2022) đã chỉ ra rằng ngày càng nhiều giống F1 mang các ưu thể lai được tạo ra trên thị trường, với mục đích tăng năng suất và chất lượng của lặc lày.

Trong nghiên cứu này, mặc dù các giống được thu thập từ ba nước khác nhau là Việt Nam, Lào và Thái Lan nhưng sự khác biệt di truyền là không lớn. Điều này là do ngoài yếu tố vị trí địa lý thì các yếu tố khác như khả năng phát tán và các ảnh hưởng từ môi trường sống cũng có thể tạo nên sự khác biệt di truyền trong quần thể.

KÉT LUẬN

Sự ít khác biệt di truyền giữa 4 giống lặc lày đã được đánh giá thông qua 18 mồi UBC-RAPD trong nghiên cứu này. Giá trị của các chỉ số PIC, EMR, MI, Rp, na, ne, h và I đã cho thấy các mồi được sử dụng là thích hợp. Trong số 18 mồi UBC-RAPD, UBC#475 là mồi cho kết quả đa hình cao nhất, với các chỉ số PIC, EMR, MI, Rp lớn nhất lần lượt là 0,357; 10,286; 3,673 và 8,000. Đặc biệt, giống lặc lày lai F1 (HUIB_Tc4) đã cho thấy có sự khác biệt di truyền so với 3 giống còn lại bao gồm một giống địa phương (HUIB_Tc1) và hai giống thuần (HUIB_Tc2 và HUIB_Tc3).

Lời cảm ơn: Nghiên cứu này được thực hiện với sự hỗ trợ của nhóm nghiên cứu tiêu biểu Đại học Huế, mã số: NCTB.DHH.2024.03.

TÀI LIỆU THAM KHẢO

Adhikari S, Biswas A, Bandyopadhyay TK, Ghosh PD (2013). A preliminary report on the genetic variation in pointed gourd (*Trichosanthes dioica* Roxb.) as assessed by random amplified polymorphic DNA. *Acta Biol Hung*, 65(2): 156-164.

Alam SS, Jahan N, Habib MA, Islam MN (2011). Cytogenetical and molecular characterization of five commercial varieties in *Trichosanthes anguina* L. *Cytologia*, 77(2):155-162.

Bobade AA, Thatte CV, Tijare RB (2022). *Trichosanthes cucumerina*: A perspective on various medicinal uses or activities. *GSC Biol Pharm Sci*, 20(03): 141-147.

Botstein D, White RL, Skolnick M, Davis R (1980). Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, 32(3): 314-331.

Doyle J, Doyle J (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull, 19(1): 11-15.

Ghorpade BB, Salokhe SS, Ghuge VG, Adsul AT (2022). Molecular Characterization of Different Soybean [*Glycine max* (L). Merril] genotypes by RAPD Markers. *J Biotechnol Bioinform Res*, 4(4): 1-6.

Goswami S, Tripathi V, Kumar N, Prakash A (2009). Molecular characterisation of dioecious *Trichosanthes dioica* roxb. using RAPD markers. *Indian J Genet*, 69(1): 76-78.

Ilakiya T, Premalakshmi V, Arumugam T, Sivakumar T (2022). Variability analysis in tomato (*Solanum lycopersicum* L.) crosses under drought stress. *J Appl Nat Sci*, 14(SI): 49-52.

Islam MR, Rahman MM, Pramanik S, Ferdousi J (2022). Heterosis studies in snake gourd (*Trichosanthes cucumerina* var. *anguina* L.). *Bangladesh J Agril Res*, 47(1): 1-12.

Kumar A, Mishra P, Singh SC, Sundaresan V (2014). Efficiency of ISSR and RAPD markers in genetic divergence analysis and conservation management of *Justicia adhatoda* L, a medicinal plant. *Plant Syst Evol*, 300: 1409-1420.

Osuagwu AN, Aguoru CU, Omoigui LO, Olasan JO (2022). Amino acid Profile of *Trichosantes cucumerina* (L.) from four Geopolitical Zones in Nigeria. *J Exp Molec Biol*, 23(1), 37-45.

Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey SV, Rafalski A (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed,* 2: 225-238.

Prevost A, Wilkinson MJ (1999). A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. *Theor Appl Genet*, 98: 107-112.

Sandhya S, Vinod KR, Sekhar JC, Aradhana R, Nath VS (2010). An updated review on *Tricosanthes cucumerina* L. Int J Pharm Sci Rev Res, 1(2): 56-60.

Srisamoot N, Padsri I (2018). Assessing genetic diversity of some Anthurium andraeanum Hort. cut-flower cultivars using ISSR markers. Genom Genet, 11(1&2): 1-8.

Rasphone S, Ho NTH, Dang LT, Nguyen BLQ, Truong HTH (2022). Genetic diversity analysis of black pepper (*Piper* spp.) with RAPD markers. *HUJOS: Natural Science*, 131(1D): 49-59.

Yang JY, Chien YY, Chiu YC, Mejia HM, Tan CM (2023). Chapter 7 - Diversity, distribution, and status of phytoplasma diseases in Taiwan. Tiwari AK, Caglayan K, Al-Sadi AM, Azadvar M, Abeysinghe S, eds. Phytoplasma Diseases in Asian Countries, Diversity, Distribution, and Current Status. *Academic Press*, 1: 149-168.

EVALUATION OF GENETIC DIVERSITY OF SOME SNAKE GOURD VARIETIES (*Trichosanthes cucumerina* L.)

Truong Thi Hong Hai^{1*}, Ho Thi Hoang Nhi¹, Sonexay Rasphone², Ho Ngoc Han¹

¹Institute of Biotechnology, Hue University, Vietnam

²Savannakhet University, Laos

SUMMARY

Snake gourd (*Trichosanthes cucumerina* L.) is a vegetable with high nutritional value and an important medicinal herb. In Vietnam, this plant is grown and used more and more. To improve the productivity and quality of snake gourd, the selection of local varieties and F1 hybrid varieties with hybrid advantages suitable for the conditions of Vietnam in general and Thua Thien Hue in particular should be focused on developing to serve the needs of farmers. However, knowledge about the genetic diversity of snake gourd is still very limited in Vietnam. In this study, DNA of four snake gourd varieties, including one local variety collected in Savannakhet, Laos; two F1 varieties from Thailand and one F1 variety from Vietnam, were extracted and purified. Then, RAPD primers were screened and RAPD markers were performed to evaluate the genetic differences between varieties. The PIC, EMR, MI, Rp, na, ne, h, and I indexes showed that 18 selected UBC-RAPD primers were suitable for studying genetic diversity. There were 199 amplified bands created, of which 91 were polymorphic. UBC#475 showed the highest level of polymorphism with PIC, EMR, MI, and Rp indexes of 0.357, 10.286, 3.673, and 8.000, respectively. In addition, the results of clustering tree analysis based on the UPGMA method have shown that the F1 hybrid variety (HUIB_Tc4) has genetic differences compared to the remaining three varieties, the local variety (HUIB_Tc1) and two pure varieties (HUIB_Tc2 and HUIB_Tc3).

Keywords: DNA marker, genetic diversity, RAPD, snake gourd, Trichosanthes cucumerina.

^{*} Author for correspondence: Tel: 0961423419; Email: tthhai@hueuni.edu.vn

THIẾT KẾ CẦU TRÚC VECTOR CRISPR/Cas9 CHỈNH SỬA GEN TÍN HIỆU CỦA HOÓC MÔN BRASSINOSTEROID Ở CÂY CÀ CHUA

Lã Văn Hiền^{1,2*}, Trương Thanh Tùng³, Lã Thị Thảo⁴, Nguyễn Văn Tịnh⁵

¹Viện Khoa học Sự sống, Đại học Thái Nguyên

²Trung tâm Nghiên cứu cây trồng thích ứng với biến đổi khí hậu, Trường Đại học Nông Lâm, Đại học Thái Nguyên

³Viện Công nghệ sinh học và Công nghệ thực phẩm, Trường Đại học Nông Lâm, Đại học Thái Nguyên

⁴Khoa Công nghệ Sinh học, Học viện Nông Nghiệp Việt Nam

⁵Khoa Khoa học cơ bản, Trường Đại học Y-Dược Buôn Ma Thuột

TÓM TẮT

Yếu tố tín hiệu của hoóc môn brassinosteroid, BRI1 EMS SUPPRESSOR 1/BRASSINAZOLE RESISTANT 1 (BES1/BZR1) Homolog 3 (BEH3) tham gia điều hòa sinh trưởng, phát triển, cũng như tăng cường khả năng chống chịu của cây trồng với điều kiện bất lọi. Tuy nhiên, thông tin về chức năng của BEH3 trong phản ứng của thực vật với stress vẫn còn hạn chế. Cà chua là loại cây trồng mang lại hiệu quả kinh tế cao thứ tư trên thế giới và cũng là một trong những cây trồng được sử dụng trong nghiên cứu di truyền. Do thiếu ổn định của đột biến gen *SlBEH3* ở cà chua nên người ta biết rất ít về chức năng của *SlBEH3* trong phản ứng với các stress sinh học và phi sinh học. Ở nghiên cứu này, chúng tôi đã phân lập gen *SlBEH3* từ cà chua 'MicroTom' với kích thước 1501 bp. Phân tích trình tự nucleotide cho thấy độ tương tự 61,4% giữa *AtBEH3* và *SlBEH3* (*Solyc03g005990.3.1*). Hơn nữa, chúng tôi đã xác định được vị trí sgRNA nằm trên exon I của gen *SlBEH3* và tạo ra cấu trúc vector pAMG4723 mang sgRNA chỉnh sửa gen *SlBEH3* bằng phương pháp ghép nối Golden Gate.

Từ khóa: Brassinosteroid, Cas9, chỉnh sửa gen, MicroTom, SlBEH3.

MỞ ĐẦU

Brassinosteroid (BR) là các hoóc môn steroid thiết yếu điều chỉnh sự sinh trưởng, phát triển và phản ứng của thực vật với áp lực môi trường. Tín hiệu hoóc môn BR và các yếu tố phiên mã của BR tham gia điều chỉnh phản ứng của BR, sinh tổng hợp BR đã được mô tả (Wang *et al.*, 2002; Yu *et al.*, 2011). Cho đến nay, mối liên kết chéo giữa tín hiệu BR và stress phi sinh học đã được tìm hiểu. Trong đó, các protein liên quan đến tín hiệu BR, bao gồm BES1/BZR1 (BRI1 EMS SUPPRESSOR 1/BRASSINAZOLE RESISTANT 1), BIN2 (BRASSINOSTEROID INSENSITIVE 2) và BEH3 (BES1/BZR1 Homolog 3), có thể điều chỉnh phản ứng đầu ra tín hiệu ABA để thích ứng với tình trạng căng thẳng khi mất nước. Gần đây, Nguyen và đồng tác giả (2021) đã báo cáo chức năng về gen tương đồng BES1/BZR1, được chỉ định là BEH3 (đối với BES1/BZR1 Homolog 3) với những mô tả về chức năng sinh lý và phân tử của gen *AtBEH3* như một yếu tố điều chỉnh tiêu cực của stress thẩm thấu, và liên kết với BR truyền tín hiệu ở *Arabidopsis thaliana*. Hơn nữa, một số đột biến *AtBEH3* biểu hiện kiểu hình chống chịu hạn (Noguchi *et al.*, 1999; Feng *et al.*, 2015). Tuy nhiên, các cơ chế mà BEH3 tham gia vào quá trình truyền tín hiệu BR để điều chỉnh sự sinh trưởng, phát triển của thực vật và phản ứng với môi trường vẫn chưa được biết rõ.

Nhiều nghiên cứu đã chứng minh hệ thống CRISPR/Cas9 có tiềm năng ứng dụng để cải thiện các tính trạng mong muốn của cây. Công cụ này có thể tạo ra các đột biến đích di truyền qua các thế hệ. Ưu điểm nổi bật của công nghệ này đó là tạo ra các đột biến gen đích một cách chính xác mà không cần sự có mặt của gen ngoại (Ricroch *et al.*, 2017). Báo cáo gần đây cho thấy, đột biến gen BEH3 bằng CRISPR/Cas9 đã tạo ra kiểu hình chống chịu hạn tốt ở cây *A. thaliana* và cây đậu tương (Nguyen *et al.*, 2021; Park *et al.*, 2022). Ở nghiên cứu này, chúng tôi sẽ thiết kế cấu trúc vector chỉnh sửa gen *SIBEH3* bằng CRISPR/Cas9 ở cây cà chua, với mục tiêu tìm kiếm và thiết kế được sgRNA chỉnh sửa gen *SIBEH3* hiệu quả. Đây sẽ là vật liệu quan trọng để tiếp tục chỉnh sửa gen ở cây cà chua, nhằm tạo ra các dòng cà chua chỉnh sửa gen chống chịu với điều kiện biến đổi khí hậu.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Giống cà chua MicroTom, do phòng thí nghiệm nghiên cứu chức năng gen thực vật, Đại học quốc gia Chonnam-Hàn Quốc cung cấp. Hệ thống cấu trúc vector pICSL01009, pICH47751, pICH7732, pICH41766, và pAGM4732 được mua từ hãng Addgene (https://www.addgene.org/).

Tách chiết RNA và phân lập gen SIBEH3 từ cây cà chua MicroTom

RNA tổng số được phân lập từ 50 mg lá tươi bằng kít RNA plant easy (Qiagen, Đức). cDNA được tổng hợp bằng hệ thống phiên mã ngược GoScript (Promega, Hoa Kỳ). Mồi khuếch đại đoạn gen *SIBEH3* sử dụng trong PCR bao gồm mồi xuôi (F:5'-TCAACTTCAAGCCGATCCAG-3'; và mồi ngược (R: 5'-CTATCTAGTGCTAGAGTTGC-3'). Phản ứng PCR được bắt đầu ở 94°C trong 5 phút, tiếp theo là 35 chu kỳ 94°C trong 30 giây, 60°C trong 30 giây và 72°C trong 30 giây, với thời gian kéo dài cuối cùng là 72°C trong 5 phút. Sản phẩm PCR-SIBEH3 được kiểm tra trên gel agarose 1%, chạy điện di ở 80 vol,100 mA trong 30 phút. Sản phẩm PCR được hiển thị trên máy UV-Vis với bước sóng 365 nm.

Thiết kế RNA định hướng (sgRNA)

Trình tự gRNA đặc hiệu cho gen *SIBEH3* được thiết kế bằng phần mềm CRISPR MultiTargeter (http://www.multicrispr.net/), dựa trên mô tả bởi Liang và đồng tác giả (2016). Hàm lượng GC của các gRNA được phân tích bằng phần mềm Genetyx 4.0.

Thiết kế cấu trúc vector mang sgRNA chỉnh sửa gen SIBEH3

Cấu trúc vector pAMG4723 mang sgRNA chỉnh sửa gen *SIBEH3* được thực hiện dựa trên phương pháp lắp ráp Golden Gate (Weber *et al.*, 2011; Engler *et al.*, 2014). Cụ thể, các *sgRNA* được khuếch đại bằng phản ứng oligo phosphprylation 10 µL gồm sgRNA oligo forward 1 µL, sgRNA oligo reverse 1 µL, đệm 10X T4 polynucleotide kinase 1 µL, T4 polynucleotide kinase (NEB) 0,5 µL, nước 6,5 µL. Phản ứng ủ ở 37°C (30 phút), 95°C (5 phút), và giảm nhiệt độ xuống 25°C. Các sgRNA1 và sgRNA2 tiếp tục được ghép nối vào vector pICSL01009::AtU6p để tạo cấu trúc pICSL01009::*AtU6p::sgRNA-SIBEH3* theo phản ứng cắt - nối của Weber và đồng tác giả (2011) bao gồm: đệm 10X NEB T4 ligase 1 µL, enzym T4 ligase 0,5 µL, sgRNA plasmid 1 µL, Oligos 1 µL, nước 6,5 µL. Cấu trúc vector này chuyển vào vi khuẩn E.coli và kiểm tra kết quả biến nạp bằng phản ứng PCR và cắt enzyme giới hạn *Bsal.*

Tiếp theo, sản phẩm PCR tinh sạch của sgRNA1 và sgRNA2 được ghép nối vào vectơ tiếp nhận pICH47751 trong phản ứng Golden Gate (Weber *et al.*, 2011) với thành phần gồm có 10X T4 Ligase Buffer 0,5 µL, 10X BSA 0,5 µL, PCR-sgRNA1 (100 ng) hoặc PCR-sgRNA2 (100 ng) 1 µL, pICH47751 (100 ng) 1 µL, pICSL01009_u6pro (100 ng) 1 µL, Bsal 0,5 µL, T4 DNA ligase 0,5 µL, và nước 5 µL. Quá trình ghép nối được thực hiện ở phản ứng ủ ở 37°C (5 phút), 16°C (10 phút), thực hiện trong 10 chu kỳ và tăng nhiệt độ lên 50°C (5 phút) và 80°C (5 phút), lưu giữ ở 10°C. Cấu trúc vector tố hợp trên được biến nạp vào *E.coli*, nuôi trên môi trường LB để chọn khuẩn lạc trắng và kiểm tra bằng PCR, cắt enzyme giới hạn *Bsal*.

Các vector tái tổ hợp mang promoter U6, gen chọn lọc NPTII, protein Cas9, và cấu trúc vector tái tổ hợp pICH47751-sgRNA1/sgRNA2 được cắt - nối vào vector pAGM4723 với phản ứng ghép nối Golden Gate (Weber *et al.*, 2011; Engler *et al.*, 2014) bao gồm 10X T4 Ligase Buffer 1 μL, pICH47732-NPTII (100 ng) 1 μL, pICH47742-35S:Cas9 (100 ng) 1 μL, pICH47751-sgRNA1 (100 ng) 1 μL, pICH47751-sgRNA2 (100 ng) 1 μL, pICH47766-linker (100 ng) 1 μL, pAGM4723 (100 ng) 1 μL, Bpil 0,5 μL, T4 DNA Ligase 0,5 μL, và nước 2 μL. Phản ứng Golden Gate được ủ ở máy ủ nhiệt theo quy trình như trên. Biến nạp cấu trúc vector tái tôt hợp trên vào tế bào *E.coli* và nuôi trên môi trường LB đặc có bổ sung kanamycin 50 μg/ml, và nuôi qua đêm ở 37°C. Chọn lọc khuẩn lạc trắng và kiểm tra bằng phản ứng PCR với cặp mồi đặc hiệu cho *sgRNA-SIBEH3*, Cas9, và cắt enzyme giới hạn *Bpil*.

Sản phẩm vector tái tổ hợp *pAGM4723::Cas9::sgRNA-SIBEH3* (100 ng) được biến nạp vào vi khuẩn *Agrobacterium tumefaciens* LBA4404 khả biến (200 µL), theo phương pháp sốc nhiệt. Chọn lọc vi khuẩn trên môi trường LB chứa 50 mg/L kanamycin. Thu nhận khuẩn lạc màu trắng và kiểm tra khuẩn lạc bằng PCR với cặp mồi đặc hiệu *sgRNA-SIBEH3* và giải trình tự.

KẾT QUẢ VÀ THẢO LUẬN

Kết quả xác định trình tự gen SIBEH3 ở cây cà chua MicroTom

Để tìm kiếm trình tự gen SIBEH3, code *Solyc03g005990.3.1* được sử dụng cho tìm kiếm trên cơ sở dữ liệu về cây cà chua EnsemblPlants (https://plants.ensembl.org/). Kết quả tìm kiếm cho thấy *Solyc03g005990.3.1* nằm trên nhiễm sắc thể (NST) số 3, có chứa cả vùng không dịch mã (5'UTR và 3'UTR), các vùng exon và intron. Trong đó, *Solyc03g005990.3.1* có chứa 3 vùng exon và 5 vùng intron. Vùng mã hóa cho gen *Solyc03g005990.3.1* có chiều dài 1501 bp (Hình 1A).

RNA tổng số được tách chiết bằng kít RNA plant easy của Qiagen từ lá cây cà chua và được phân tích trên gel agarose 2%. Kết quả điện di ở hình 1 cho thấy RNA tổng số thu được đảm bảo độ tinh sạch, RNA vẫn nguyên vẹn (Hình 1B). RNA được sử dụng cho phản ứng tổng hợp cDNA. Kết quả kiểm tra nồng độ cDNA bằng máy Nanodrop cho thấy nồng độ thu được đạt dao động từ 1000 - 1600 ng/µl.

Gen *SIBEH3* được nhân bản bằng PCR với khuôn là DNA của mẫu lá cà chua non (giai đoạn 6 lá), sử dụng cặp mồi được thiết kế dựa vào trình tự nucleotide của gen *SIBEH3* được công bố trên Genbank. Kết quả điện di trên gel agarose 1% cho thấy sản phẩm PCR chỉ xuất hiện một băng DNA duy nhất có kích thước xấp xỉ 1,5 kb tương ứng với kích thước lý thuyết của đoạn gen *SIBEH3* (1501 bp) (Hình 1C). Như vậy, gen *SIBEH3* mong muốn đã được phân lập thành công từ RNA tổng số của mẫu lá non.



Hình 1. Kết quả xác định vị trí và phân lập gen SIBEH3

(A) Vị trí gen SIBEH3 (Solyc03g005990.3.1). (B) RNA tổng số tách chiết từ mẫu lá. (C) Sản phẩm PCR-SIBEH3 ở giống cà chua MicroTom điện di trên gel agarose 1%. M: Thang chuẩn 1 kb. 1 và 3 mẫu RNA được PCR với cặp mồi đặc hiệu gen SIBEH3.

Kết quả nhận diện sự tương đồng và xác định vùng axit amin mã hóa protein *SIBEH3* chứa một khung đọc mở duy nhất mã hóa protein 666 axit amin. So sánh trình tự axit amin phân tích từ 9 gen cho thấy *Solyc03g005990.3.1* có mức độ tương đồng cao 61,4% so với gen *AtBEH3* (At1g75080) và cây phân loại (Hình 2A), có trình tự amino acid tương đồng đáng kể với các thành viên đã biết của họ protein BES1-N terminal active domain. Các giá trị tương đồng thể của 40-53% nhận dạng và độ tương tự 40,7%–61,4% đã được quan sát giữa *AtBEH3* và SIBEH3 (*Solyc03g005990.3.1*) tương ứng. Sự sắp xếp của *AtBEH3* và BES1-N-terminal active domain suy ra các chuỗi axit amin và mức độ tương đồng của gen *SIBEH3* biểu thị bằng cây phân loại (Hình 2B).





(A) Hiển thị các trình tự nucleotide tương đồng ở Arabidopsis BEH3 (At1g7508), và một số gen Solyc03g005990.3.1, Solyc02g071990.3.1, Solyc07g062260.3.1, Solyc02g063010.3.1, Solyc04g079980.3.1, Solyc12g089040.3.1, Solyc10g07390.2.1, và Solyc08g005780.4.1 từ cà chua MicroTom. Màu đen và màu xám tương ứng biểu thị các axit amin giống hệt và tương tự nhau. Khoảng trống được sử dụng để tối ưu hóa việc căn chỉnh. (B) Cây phát sinh chủng loại mô tả mối quan hệ tương đồng giữa các gen BEH3 ở Arabidopsis thaliana và cà chua MicroTom. Các số tại các điểm nhánh biểu thị giá trị bootstrap sau 1000 lần lặp.

Thiết kế trình tự sgRNA định hướng chỉnh sửa gen SIBEH3

Hoạt động chỉnh sửa gen bằng CRISPR/Cas9 dựa trên khả năng cắt chính xác DNA sợi đôi tại vị trí mong muốn của phức hệ protein-RNA CRISPR/Cas9, bao gồm hai thành phần chính: (1) protein Cas9 có hoạt tính endonuclease và (2) phân tử sgRNA (single guide RNA) có vai trò dẫn đường cho phức hệ đến đúng vị trí DNA cần cắt. Tính chính xác của hệ thống CRISPR/Cas9 được quyết định bởi đoạn trình tự dài ~20 nucleotide (crRNA) trên phân tử sgRNA (Bortesi và Fischer, 2015). Để phục vụ nghiên cứu chỉnh sửa gen *SIBEH3* bằng CRISPR/Cas9, trình tự crRNA được thiết kế nhằm gây đột biến tại các vị trí exon I trên gen *SIBEH3*. Bằng phần mềm CRISPR MultiTargeter, 227 trình tự (20 nu) trên gen *SIBEH3* đã được xác định nằm phía trước vị trí nhận biết bảo thủ PAM (protopacer adjacent motif-NGG) (Dữ liệu không được trình bày). Để đảm bảo khả năng duy trì hoạt tính và tính đặc hiệu của hệ thống CRISPR/CAS9 trong nhân tế bào, gRNA phải có hàm lượng GC trong trình tự crRNA từ 30-80%, có khả năng hình thành cấu trúc bậc II ổn định (duy trì cầu trúc vòng loop RAR, SL2 và SL3), tổng số cặp bazơ (TBP-total base pairs) < 13, số cặp bazơ liên tục (CBP-consecutive base pairs) < 8, số cặp bazơ bên trong cấu trúc crRNA (IBP-internal base pairs) < 7 (Liang *et al.*, 2016). Kết quả phân tích bằng phần mềm Genetyx 4.0 và Mfold 2.3 cho thấy gRNA đảm bảo việc hình thành phức hệ CRISPR/CAS9 ổn định có hoạt tính trong nhân tế bào thực vật (Hình 3).



Hình 3. Thiết kế sgRNAs và cấu trúc vector mang sgRNAs chỉnh sửa gen SIBEH3

(A) Xác định vị trí và trình tự của sgRNAs trên vùng nucleotide mã hóa (CDS) cho gen SIBEH3. (B) Cấu trúc vector pAMG4723 mang sgRNA1 và sgRNA2. Vị trí mục tiêu cho cấu trúc Cas9 (màu xanh dương), PAM (màu tím), và vị trí mồi khuếch đại cho cả vùng gen SIBEH3 (màu xanh đậm).

Cấu trúc vector mang sgRNA chỉnh sửa gen SIBEH3



Hình 4. Sơ đồ mô tả quá trình thiết kế cấu trúc vector pAMG4723 mang sgRNAs chỉnh sửa gen SIBEH3 theo phương pháp cắt – nối Golden Gate (Weber *et al.*, 2011; Engler *et al.*, 2014)

(A) Khuếch đại sgRNA và ghép nối lần 1. (B) Ghép nối lần 2 để tạo cấu trúc vector pAMG4723 mang sgRNA1 và sgRNA2 hoàn chỉnh. (C) Cấu trúc vector pAMG4723 mang sgRNAs chỉnh sửa gen SIBEH3.

Trình tự sgRNA-SIBEH3 được khuếch đại thông qua việc gắn nối vào vector pICH86966::AtU6p để tạo cấu trúc tổ hợp pICH86966::AtU6p::sgRNA-SIBEH3 nhờ enzym cắt giới hạn Bsal. Kết quả kiểm tra khuẩn lạc bằng chuỗi

trùng hợp (PCR) cho thấy hai khuẩn lạc có một băng duy nhất kích thước 0,75 kb (Hình 5A). Sản phẩm PCR của sgRNA-SIBEH3 trên được tinh sạch và thiết lập phản ứng cắt nối sử dụng enzym giới hạn Bsal cho cấu trúc pICSL01009:AtU6p (Spec^R) và cấu trúc vector mục tiêu ở ghép nối lần 1 pICH47751 (Carb^R) để tạo thành cấu trúc tái tổ hợp mang trình tự sgRNA-SIBEH3. Cấu trúc này sẽ được biến nạp vào vi khuẩn *E.coli* DH5α. Kết quả PCR chọn lọc ngẫu nhiên một khuẩn lạc cho thấy có một băng vạch DNA duy nhất xuất hiện, kích thước khoảng 0,75 kb (Hình 5B). Để xác nhận, chúng tôi tiếp tục cắt plasmid bằng enzym giới hạn *Basl*, kết quả điện di sản phẩm cắt trên gel 1% có hai băng vạch DNA xuất hiện, trong đó có một băng DNA kích thước 0,75 kb của gen sgRNA-SIBEH3 và băng DNA kích thước 5 kb của vector pICH47751 (Hình 5C).

Ở bước ghép nối lần 2 cấu trúc *pICH47751::sgRNA-SIBEH3* sẽ được tiến hành dung hợp với các tổ hợp cấu trúc khác mang gen chọn lọc kháng sinh kanamycin (pICH47732::NPTII), Cas9 (pICH47742::Cas9), và phần liên kết (pICH41766-linker) cùng với cấu trúc vector pAGM4723 biểu hiện ở cây cà chua, để tạo cấu trúc hoàn chỉnh mang sgRNA chỉnh sửa gen *SIBEH3* (Hình 4C). Quá trình cắt nối được tiến hành dưới sự xúc tác của enzyme cắt giới hạn *Bpil (BbsI)*. Kết quả kiểm tra PCR khuẩn lạc tái tổ hợp cho thấy sự hiện diện của gen Cas9, kích thước 1,1 kb (Hình 5D) và *sgRNA-SIBEH3* (0,75 kb, Hình 5E). Các khuẩn lạc này tiếp tục được kiểm tra bằng phản ứng cắt enzym giới hạn *Bpil*, kết quả xác nhận rằng gen Cas9, *sgRNA-SIBEH3* đã được chèn vào vector pAGM4723 (12 kb) (Hình 5F). Kết quả giải trình tự vector pAGM4723 tái tổ hợp cho thấy sgRNA1/sgRNA2 đã được chèn ghép nối thành công vào vector pAGM4723 (Hình 6). Quy trình lắp ráp Golden Gate tuân theo các quy tắc thiết kế đơn giản và rõ ràng, đồng thời cho phép cắt nối nhiều mảnh có kích thước khác nhau vào một khung đọc mở (Weber *et al.*, 2011; Engler *et al.*, 2014). Sản phẩm cuối cùng có thể được tích hợp trực tiếp vào hệ thống nhân bản Multisite GatewayTM thông qua vector nhị thể pAGM4723 được sử dụng rộng rãi, mang lại sự linh hoạt hơn khi sử dụng gen chuyển ở cây cà chua (Brooks *et al.*, 2014).



Hình 5. Kết quả kiểm tra sự hiện diện của sg RNA ở cấu trúc vector pAMG4723

(A) và (B) Khuẩn lạc được kiểm tra PCR một phần đoạn gen SIBEH3 chứa vị trí sgRNA được nối vào vector pICH47751 (ghép nối lần 1). (C) Cắt kiểm tra plasmid bằng enzyme cắt giới hạn Bsal. (D) Kết quả kiểm tra PCR khuẩn lạc tái tổ hợp của protein Cas9, kích thước 1,1 kb. Kiểm tra khuẩn lạc bằng phản ứng PCR và cắt enzym giới hạn Bpil đối với đoạn gen mang sgRNA ghép nối vào vector pAMG4723 (ghép nối lần 2).



Vị trí mồi xuôi khuếch đại vector pAGM4723_F (màu nền xanh đậm), vị trí mồi xuôi khuếch đại vector pAGM4723_R (màu hồng), sgRNA1 (chữ đậm màu xanh đậm), sgRNA2 (chữ đậm màu đỏ), linker (màu nền xanh lá cây), <u>ACTA</u> vị trí nhận diện protein cas9.

KÉT LUẬN

Ở nghiên cứu này, gen *SIBEH3* liên quan đến tín hiệu của hoóc môn BR đã được phân lập và giải trình tự đầy đủ. Vùng trình tự đã phân lập chứa exon I đã được chọn để thiết kế một cấu trúc gRNA phục vụ nghiên cứu chỉnh sửa gen *SIBEH3* bằng công nghệ CRISPR/Cas9. Hai yếu tố gồm có sgRNA1/sgRNA2 và protein Cas9 đã được ghép nối thành công vào hệ thống vector pAMG4723. Cấu trúc vector này sẽ tiếp tục được chuyển vào cây cà chua để đánh giá hiệu quả chỉnh sửa gen của các sgRNA và làm rõ thêm vai trò của gen *SIBEH3* đối với quá trình điều hòa sinh trưởng, phát triển và đáp ứng stress của cây cà chua ở các nghiên cứu tiếp theo.

Lòi cảm ơn: Nghiên cứu được thực hiện dưới sự hỗ trợ của đề tài cấp Quốc gia theo nhiệm vụ Nghị định thư, mã số NĐT/KR/23/10. Nhóm nghiên cứu cảm ơn sự hỗ trợ và cung cấp vật liệu của giáo sư Cheol Soo Kim, Khoa Nông nghiệp và Khoa học sự sống, Đại học quốc gia Chonnam, Hàn Quốc.

TÀI LIỆU THAM KHẢO

Belhaj K, Chaparro-Garcia A, Kamoun S, Nekrasov V (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods*, 9(1): 39.

Bortesi L, Fischer R (2015). The CRISPR/Cas9 system for plant genome editing and beyond. Biotechnol Adv, 33(1): 41-52.

Brooks C, Nekrasov V, Lippman ZB, Eck JV (2014). Efficient gene editing in tomato in the first generation using the Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-Associated9 System. *Plant Physiol*, 166: 1292-1297.

Engler C, Youles M, Gruetzner R, Ehnert TM, Werner S, Jones JDG, Patron NJ, Marillonnet S (2014). A Golden Gate modular cloning toolbox for plants. ACS Synth Biol, 3: 839-843.

Feng Y, Yin Y, Fei S (2015). Down-regulation of BdBRI1, a putative brassinosteroid receptor gene produces a dwarf phenotype with enhanced drought tolerance in *Brachypodium distachyon. Plant Sci*, 234: 163-173.

Liang G, Zhang H, Lou D, Yu D (2016). Selection of highly efficient sgRNAs for CRISPR/Cas9-based plant genome editing. Sci Rep, 6: 21451.

Nguyen TV, Park CR, Lee KH, Lee S, Kim CS (2021). BES1/BZR1 Homolog 3 cooperates with E3 ligase AtRZF1 to regulate osmotic stress and brassinosteroid responses in *Arabidopsis J Exp Bot*, 72: 636-653.

Noguchi T, Fujioka S, Choe S, Takatsuto S, Yoshida S, Yuan H, Feldmann KA, Tax FE (1999). Brassinosteroid-insensitive dwarf mutants of Arabidopsis accumulate brassinosteroids. *Plant Physiol*, 121: 743-752.

Park CR, Nguyen VT, Min JH, Sang HK, Lim GH, Kim CS. 2022. Isolation and functional characterization of soybean BES1/BZR1 Homolog 3-Like 1 (GmBEH3L1) associated with dehydration sensitivity and brassinosteroid signaling in *Arabidopsis thaliana*. *Plants*, 11(19): 2565.

Ricroch A, Clairand P, Harwood W (2017). Use of CRISPR systems in plant genome editing: toward new opportunities in agriculture. *Emerg Top Life Sci*, 1:169-182.

Wang ZY, Nakano T, Gendron J, He J, Chen M, Vafeados D, Yang Y, Fujioka S, Yoshida S, Asami T, Chory J (2002). Nuclearlocalized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. *Dev Cell*, 2: 505-513.

Yu X, Li L, Zola J, Aluru M, Ye H, Foudree A, Guo H, Anderson S, Aluru S, Liu P, Rodermel S, Yin Y (2011). A brassinosteroid transcriptional network revealed by genome wide identification of BESI target genes in *Arabidopsis thaliana*. *Plant J*, 65: 634-646.

CONSTRUCTION OF THE CRISPR/Cas9 VECTOR EDITING BRASSINOSTEROID HORMONE SIGNALING GENE IN TOMATO

La Van Hien^{1,2*}, Truong Thanh Tung³, La Thi Thao⁴, Nguyen Van Tinh⁵

¹Institute of Life Science, Thai nguyen University

²Center of Research Crop for adaptation to Climate change (CRCC), TNU- University of Agriculture and Forestry (TUAF)

³Institute of Biotechnology and Food Technology, TNU- University of Agriculture and Forestry (TUAF)

⁴Vietnam National University of Agriculture

⁵*Faculty of Basic Science, BMU- University of Pharmacy*

SUMMARY

The signaling factor of the brassinosteroid hormone, BRI1 EMS SUPPRESSOR 1/BRASSINAZOLE RESISTANT 1 (BES1/BZR1) Homolog 3 (BEH3) regulates plant growth and development, as well as enhancing plant resistance to stress conditions. However, information on the function of BEH3 in plant responses to environments stress is still limited. Tomatose are the fourth most economically productive crop in worldwide and are also one of the best characterized model plants used in genetic research. Due to the lack of stability of the *SlBEH3* gene mutation in tomato, little is known about the function of *SlBEH3* in tomato response to biotic and abiotic stresses. In this study, we isolated the *SlBEH3* gene from 'MicroTom' tomato with a size of 1501 bp. Nucleotide sequence analysis revealed 61.4% similarity between *AtBEH3* and *SlBEH3* (*Solyc03g005990.3.1*). Furthermore, we identified the sgRNA located on exon I of the *SlBEH3* gene and created the vector construct pAMG4723 carrying the sgRNA mutated *SlBEH3* gene using the Golden Gate assembly method.

Keywords: Brassinosteroid, Cas9, gene editing, MicroTom, SlBEH3.

Author for correspondence: Tel: 0395787043; Email: lavanhien@tuaf.edu.vn

PHÂN TÍCH ĐA HÌNH ĐƠN NUCLEOTIDE CỦA THỤ THỂ KHỨU GIÁC cOR52H9 VÀ cOR9S13 TRÊN CHÓ NGHIỆP VỤ

Lê Thị Huệ¹, Phạm Thị Khánh Linh^{1,2}, Nguyễn Ngọc Hưng³, Lê Văn Trọng³, Đoàn Thị Thanh Hương^{1,2*}

¹Viện Công nghệ sinh học, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

²Trường Đại học Khoa học và Công nghệ Hà Nội, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

³Bộ Tư lệnh Cảnh sát cơ động, Bộ Công an

TÓM TẮT

Môt số loài chó được quân đôi và cảnh sát sử dụng vì chúng có khứu giác phi thường, nhay hơn nhiều so với con người. Nhờ đặc điểm này nên chúng có thể phát hiện ma túy, chất nổ và thậm chí là định vị người. Từ hàng trăm năm đến nay chó đã trở thành một phần thiết yếu trong nhiều hoạt động quân sự và cảnh sát, góp phần đáng kể vào an ninh và an toàn công cộng. Các đa hình nucleotide đơn (SNP) trong gen thụ thể khứu giác là những yếu tố chính góp phần tạo nên sự khác biệt về khả năng khứu giác giữa các loài chó. Những biến thể di truyền này có thể ảnh hưởng đáng kể đến khả năng phát hiện và phân biệt giữa các mùi hương khác nhau của chó. Chính vì vậy áp dụng các phương pháp sinh học phân tử có thể cải thiện đáng kể việc lựa chọn và đào tạo các đội chó chuyên nghiệp có hiệu quả cao. Trong nghiên cứu này, chúng tôi đã khuếch đại và phân tích các SNPs trong gene cOR52H9 và cOR9S13 của 20 chó giống Malinois của Bộ Tư lệnh Cảnh sát cơ động, Bộ Công an. Kết quả giải trình tự cho thấy, các trình tự gene cOR52H9 và cOR9S13 có tỷ lệ tương đồng cao trên 99% so với các trình tự gene đã được công bố trên Ngân hàng gene. Gene cOR52H9 có 8 điểm Single nucleotide polymorphisms (SNP) ở các vị trí: G58A, A180G, G250A, T315A, G598A, T642G, G678A và G1042T. Gene cOR9S13 có 8 điểm Single nucleotide polymorphisms (SNP) ở các vị trí: T125A, C375T, T390C, G398A, C416T, C546G, C566A và G592A. Trong số các SNPs được xác định, chỉ có một con chó xuất hiện 1 SNP bất lợi ở vị trí 592 của gene cOR9S13 (G592A), là gen quy định khả năng khứu giác; trong khi có 2 cá thể lại xuất hiện 2 SNPs bất lợi trên gene cOR52H9 nhưng ở mức độ vừa. Việc nghiên cứu sử dụng các SNPs ở các gen mã hóa cho thụ thể khứu giác của chó sẽ giúp chon loc được những chú chó thuần chủng có khả năng phát hiện mùi tốt nhất cho công tác bảo tồn, phát triển nguồn gene và huấn luyện.

Từ khóa: Chó nghiệp vụ, cOR52H9, đa hình đơn nucleotit, gene thụ thể khứu giác.

MỞ ĐẦU

Chó đánh hơi đã được sử dụng rộng rãi trong quân đội, công an và dân sự trên toàn thế giới. Việc tuyển chọn chó để đưa vào huấn luyện trước đây chỉ dựa vào các đặc điểm ngoại hình. Tuy nhiên thực tế cho thấy việc đánh giá chỉ dựa trên ngoại hình là chưa đủ để lựa chọn được những cá thể chó tốt nhất cho mục tiêu phát hiện ma túy, chất nổ... (Maejiima *et al.*, 2007). Để khắc phục nhược điểm trên, thời gian gần đây đã có một số nghiên cứu bước đầu về khả năng sử dụng kỹ thuật sinh học phân tử phục vụ cho ngành nuôi chó nghiệp vụ của công an (Lesniak *et al.*, 2008).

Các thụ thể khứu giác đã được Buck và Axel phát hiện năm 1990 (Buck, Axel, 1991). Người ta đưa ra giả thuyết rằng khả năng khứu giác của chó được xác định bởi tính đa hình trong gene thụ thể khứu giác (OR). Quá trình phân biệt mùi bắt đầu ở biểu mô thần kinh khứu giác nằm trong khoang mũi. Chất tạo mùi kích hoạt các thụ thể khứu giác trên bề mặt tế bào của tế bào thần kinh khứu giác, và truyền tín hiệu tiếp theo đến não. Các gene OR thuộc siêu họ GPCR chứa khoảng 1300 gene ở chó (Olender *et al.*, 2004) và được cho là họ gene có quy mô lớn nhất trong bộ gene của động vật có vú (Malnic *et al.*, 2004). Cấu trúc của thụ thể khứu giác bao gồm các phần sau: E-TM1-IC1-TM2-EC1-TM3-IC2-TM4-EC2-TM5-EC3-TM6-IC3-TM7-I. Trong đó, E là vùng đầu –NH2 nằm bên ngoài màng tế bào; TM là phần xuyên màng; IC là phần nội bào; EC là phần ngoại bào; và I là đầu COOH nằm trong tế bào (Tacher *et al.*, 2005). Kích thước lớn của họ OR cho thấy rằng sự phân biệt ban đầu giữa các chất tạo mùi khác nhau phụ thuộc vào ái lực liên kết có chọn lọc của OR (Zhao and Firestein, 1999).

Trong các OR có 2 gene có khả năng ảnh hưởng đến khả năng khứu giác của chó phục vụ trong các lĩnh vực liên quan đến phát hiện chất nổ và ma túy, đó là: cOR52H9 (Canis lupus familiaris olfactory receptor family 52 subfamily H member 9) và cOR9S13 (Canis lupus familiaris olfactory receptor family 9 subfamily S member 13). Các nghiên cứu trước đây đã tìm thấy mối liên hệ đáng kể giữa các đa hình nucleotide đơn trên gene cOR52H9, cOR9S13, cOR10H1 với khả năng phát hiện mùi (Yang *et al.*, 2022). Cụ thể các SNPs T414G, G450A và G814T trên gene cOR52H9 và các SNPs C331T, G336A, C465T, G592A, G759A và T762C trên gene cOR9S13 đã được chứng minh có liên quan đến khả năng khứu giác của chó (Đỗ Văn Thu *et al.*, 2017; Lesniak *et al.*, 2008; Yang *et al.*, 2022). Đặc biệt đột biến tại vị trí 592 của gene OR9S13 có sự biến đổi nucleotide từ G > A, dẫn đến thay đổi amino acid từ Alanine > Threonine ở vùng cấu trúc EC2 sẽ làm giảm khả năng khứu giác của chó.

Trong nghiên cứu này, chúng tôi đã thu nhận và giải trình tự gene cOR52H9 và cOR9S13 của 20 chó giống Malinois. Từ đó, xác định đa hình kiểu gene/allen của hai vùng gene trên để đánh giá khả năng khứu giác của các chó nghiên cứu.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Mẫu bệnh phẩm: là mẫu máu của 20 chó giống Malinois của Bộ Tư lệnh Cảnh sát cơ động, Bộ Công an. Máu chó được lấy trong ống chống đông và chuyển về phòng thí nghiệm càng sớm càng tốt ngay sau khi lấy. Sau khi về phòng thí nghiệm, ống đựng máu được bảo quản ở điều kiện nhiệt độ 4°C để sử dụng cho tách chiết DNA hệ gene.

Tách chiết DNA tổng số

Tách chiết DNA hệ gene từ máu bằng bộ kit DNeasy Blood and Tissue (Qiagen, Đức) theo hướng dẫn của nhà sản xuất. Các bước tách chiết cụ thể như sau: Bước 1: Bổ sung 180 µL Buffer ATL vào mẫu máu, vortex và ủ 56°C trong 30 phút; bước 2: Thêm 20 µL Proteinase K, vortex và ủ ở 56°C cho đến khi mẫu được ly giải hoàn toàn; bước 3: Bổ sung 200 µL Buffer AL vào mẫu, vortex và ủ mẫu ở 70°C trong 10 phút; bước 4: Thêm 200µl ethanol (96-100%), đảo nhẹ mẫu; bước 5: chuyển cột và ly tâm 12000 vòng/phút trong 2 phút, bỏ dịch phía dưới ống hứng; bước 6: Thêm 500µl Buffer AW1 và ly tâm 12000 vòng/phút trong 2 phút, bỏ dịch dưới ống hứng; bước 7: Thêm 500 µL Buffer AW2 và ly tâm 12000 vòng/phút trong 3 phút, bỏ dịch dưới ống hứng; bước 8: Ly tâm làm khô cột 2 lần với tốc độ 12000 vòng/phút, mỗi lần 2 phút; bước 9: Bổ sung 30 µL Buffer AE vào giữa cột, ủ 2 phút ở nhiệt độ phòng, sau đó ly tâm ở 12000 vòng/phút trong 1 phút để thu DNA tổng số.

Mẫu DNA tổng số chứa hệ gene của chó đã tách chiết được trên điện di gel agarose 1% và đo quang phổ hấp thụ ở bước sóng 260/280nm trên máy NanoDrop lite (hãng Thermo scientific) để đánh giá chất lượng.

Phản ứng PCR khuếch đại gene cOR52H9 và cOR9S13

Vùng gene cOR52H9 được thu nhận bằng hai cặp mồi cOR52H9-F1 - cOR52H9-R1 và cOR52H9-F2 - cOR52H9-R2. Vùng gene cOR9S13 được thu nhận bằng cặp mồi cOR9S13-F - cOR9S13-R (Bảng 1).

Gene	Tên mồi	Trình tự mồi	Kích thước dự tính								
	cOR52H9-F1	3'-TCTTTCTTCATTTGAGCCCAGG-5'	1.4 kb								
	cOR52H9-R1	3'-TTTCCTCCCTTCCCATACTTTG-5'	1,4 KD								
cOR52H9	cOR52H9-F2	3'-ATCGCTAATGTGTCCTCAGGG-5'	1 kb								
	cOR52H9-R2	3'-ACTTCTCCTTCAGTGACTCTCC-5'	T KD								
oOP0612	cOR9S13-F	3'-ACTTGACCCACTTAATCTGCCA-5'	1146								
000013	cOR9S13-R	3'-AGTGGCAATCTCCGTATCCTC-5'	1,1 KD								

Bảng 1. Các cặp mồi sử dụng trong phân lập gene cOR52H9 và cOR9S13

Phản ứng PCR được thực hiện trên tổng thể tích 50 µL bao gồm: 25 µL PCR Master Mix (2X), 2 µL mỗi loại mồi (10pmol/µL) và 3 µL cDNA khuôn và nước tinh khiết cho đủ 50 µL. Chu trình nhiệt của phản ứng PCR như sau: 94°C-5 phút, tiếp theo là 35 chu kỳ [94°C-1 phút, 52°C-1 phút, 72°C-2 phút] và 72°C-10 phút. Sản phẩm PCR được kiểm tra bằng điện di trên agarose 1%.

Giải trình tự và xác định đa hình kiểu gene

Sản phẩm PCR dương tính có chất lượng tốt được tinh sạch và giải trình tự trực tiếp tại công ty TNHH dịch vụ và thương mại Nam Khoa bằng phương pháp Sanger. Chuỗi nucleotitide được xử lý bằng chương trình Seqed1.3, so sánh bằng chương trình AssemblyLIGN1.9 và MacVecter8.2 (Accelrys Inc.) trên máy tính Mactintosh. Các trình tự nucleotide và amino acid thu nhận được so sánh đối chiếu với chuỗi gene tương ứng đăng ký tại Ngân hàng gene bằng chương trình GENEDOC2.7 (http://www.nrbsc.org/gfx/genedoc/) để xác định các vị trí đột biến trong gene.

KẾT QUẢ VÀ THẢO LUẬN

Các chó nghiên cứu được cung cấp đầy đủ thông tin cho nghiên cứu. DNA tổng số của các mẫu máu sau tách chiết được kiểm tra bằng điện di trên agarose 1% và đo trên máy quang phổ hấp thụ đều đảm bảo đạt tiêu chuẩn cho các phân tích tiếp theo.

Kết quả thu nhận và phân tích gene cOR52H9

Từ DNA tổng số tách từ mẫu máu đã thu nhận được sản phẩm PCR vùng gene cOR52H9 của 20 chó nghiên cứu, với kích thước lần lượt là 1,4 kb (với cặp mồi cOR52H9-F1 - cOR52H9-R1) và sản phẩm PCR kích thước khoảng 1 kb (với cặp mồi cOR52H9-F2 - cOR52H9-R2) (Hình 1 và Hình 2).



Hình 1. Hình ảnh điện di sản phẩm PCR khuếch đại gene cOR52H9 của chó với cặp mồi cOR52H9-F1 - cOR52H9-R1 Ghi chú: Giếng M là thang DNA chuẩn Lambda/HindIII, Giếng 1-20 là sản phẩm PCR của 20 mẫu chó nghiên cứu



Hình 2. Hình ảnh điện di sản phẩm PCR khuếch đại gene cOR52H9 của chó với cặp mồi cOR52H9-F2 - cOR52H9-R2 Ghi chú: Giếng M là thang DNA chuẩn Lambda/HindIII, Giếng 1-20 là sản phẩm PCR của 20 mẫu chó nghiên cứu

Gene cOR52H9 có kích thước 1170 nucleotide bao gồm mã mở đầu (ATG) và mã kết thúc (TGA). Kết quả giải trình tự gene của 20 cá thể chó giống Malinois cho thấy có tỷ lệ đồng nhất cao (từ 99.3% đến 100%) khi so sánh với nhau và so với các trình tự gene tương ứng đăng kí trên Ngân hàng gene, và phát hiện được 8 điểm SNPs bao gồm các vị trí: G58A, A180G, G250A, T315A, G598A, T642G, G678A và G1042T (Hình 3). So sánh với kết quả nghiên cứu của Đỗ Văn Thu và đồng tác giả (2017), chúng tôi đã phát hiện được thêm 2 SNPs khác bao gồm G58A và A181G. Tuy nhiên các sai khác về nucleotide không làm sai khác về amino acid của OR (Bảng 1), từ đó suy đoán không bị ảnh hưởng đến khả năng khứu giác. Bên cạnh đó, theo Đỗ Văn Thu và cộng sự, kiểu gene TT ở vị trí T315A có điểm phát hiện ma túy cao (Đỗ Văn Thu và cộng sự, 2017). Kết quả phân tích trình tự gene cho thấy 18/20 cá thể chó nghiên cứu có kiểu gene đồng hợp tử TT ở vị trí này, chứng tỏ có khả năng ngửi mùi tốt. Bên cạnh đó, ba SNPs được chứng minh có ảnh hưởng đến khả năng khứu giác của chó ở mức độ vừa gồm T642G, G678A và G1042T; trong đó các kiểu gen tương ứng là GG, AA và TT có điểm phát hiện ma túy thấp hơn so với các kiểu gen khác (Đỗ Văn Thu *et al.*, 2017). G680A mã hóa cho amino acid Valine tại đoạn xuyên màng 4 (TM4), SNP tại vi trí G1042T làm thay đổi amino acid Valine thành Leucine tại đoạn xuyên màng thứ 7 (TM7), có vai trò đối với khả năng phân biệt mùi của của thụ thể khứu giác (Liu et al., 2003). Kết quả nghiên cứu trên 20 chó nghiệp vụ cho thẩy có 14 cá thể chó nghiên cứu có chứa kiểu gen GA ở SNP G678A và 18 cá thể chó nghiên cứu có chứa kiểu gen GT ở SNP G1042T – là các kiểu gen có điểm phát hiện ma túy cao; 06 cá thể chó nghiên cứu có chứa kiểu gen AA ở SNP G678A và và 02 cá thể chó nghiên cứu có chứa kiểu gen TT ở SNP G1042T - là các kiểu gen có điểm phát hiện ma túy thấp. Tổng hợp số liệu phân tích toàn bộ gen cOR52H9 cho thấy chỉ có hai cá thể chó số 14 và số 20 mang hai đồng thời hai SNPs bất lợi ở mức độ vừa về khả năng khứu giác trên gene cOR52H9. Tuy nhiên cần tiếp tục có những nghiên cứu mở rộng về mối liên quan này (Đỗ Văn Thu et al., 2017).

	6	0	186	_	•	320	600	640	-	680	1040
XM-005633593:	lee ter	AAI	ATCGC	AATGTG	GGATCGGGA	I CCITIIGIE	A GI GCC	CIGACIC	AGGIGAICA	ETTS IGA SGAGC	icgaaa igi ictij
Canis-lupu :	A			G			. A	G.		. .A	· · · · · · · · · · · · · · · ·
Canis-lupu :	A.			G.		A					
Canis-lupu :	A.			G.			. A			. .A.	
Mischka-br ;	A.			G.		A		G.			T
SID07034-2 ;	A.			G.			. A				
Tasha-bree :											
UNSW-BAS-2 :	A			G		A				A.	т
Zoey-breed :											
Canis-fami :						A		G		A.	
1.EGON :											
2.EVAN										Δ	
3.BUDHA :											
4.CLEO :											
5.APOLLO :											
6.HITTER :											
7.FRODO :											· · · · · · · · · · · · · · ·
8.CIGO ;								· · · · · · · · · · · · · · · · · · ·			
9.KIWI ;								· · · · · · · · · · · · · · · · · · ·			
10.CASPER :	- 11						. A	· · · · · · · · · · · · · · · · · · ·			
11.BERETTA :								· · · · · · · · · · · · · · · · · · ·			
12.Honey :				~ ~ ~ ~ ~ ~ ~		• • • • • • • • • • • •		· · · · · · · · · · · · · · · · · · ·			
13.Wanda :				G		• • • • • • • • • • • •		· · · · · · · · · · · · · · · · · · ·		A	
14.Cica :				G				· · · · · · · · · · · · · · · · · · ·		A.	
15.Amma :	· · · · ^			····	· · · · · · · · · · · · · · · · · · ·	*******					тт
16.ANGELIN :					· · · · · · · · · · · · · · · · · · ·	• • • • • • • • • • • •					
17.ASKA :					· · · · · · · · · · · · · · · · · · ·	· <mark>·</mark> · · · · · · · · · · ·					
18.IDA :					· · · · · · · · · · · · · · · · · · ·						
19.CIRCE :					· · · · · · · · · · · · · · · · · · ·						
20.AMON				G						1	
	A.			в.		A					
	_								• • • • • • • • •		. T.

Hình 3. Các vị trí đa hình (SNPs) trên gene cOR52H9 của 20 chó nghiên cứu

Bảng 2. Vị trí các SNPs và amino acid tương ứng trong cOR52H9

SNPs	G58A	A180G	G250A	T315A	G598A	T642G	G678A	G1042T
Amino acid	Val20	Leu60	Asp84	lle105	Val200	Pro214	Val226	Val348

Kết quả thu nhận và phân tích gene cOR9S13

Sản phẩm PCR khuếch đại gene cOR9S13 từ DNA hệ gene có kích thước khoảng 1,1 kb tương đương với kích thước dự kiến (Hình 4).



Hình 4. Kết quả điện di sản phẩm PCR thu nhận vùng gene cOR9S13 của 20 mẫu chó giống Malinois nghiên cứu

Kết quả so sánh trình tự vùng gene cOR9S13 của 20 cá thể chó giống Malinois nghiên cứu đã phát hiện được 8 điểm đa hình đơn nucleotide (SNPs) bao gồm các vị trí: T125A, C375T, G398A, C416T, C546G, C566A, G592A và C843T (Hình 5). Đột biến thay đổi nucleotide trên gene cOR9S13 đã dẫn đến thay đổi amino acid ở 6 vị trí: V42E, S133N, T139I, F182L, L189M và A198T.

HỘI NGHỊ KHOA HỌC TOÀN QUỐC VỀ CÔNG NGHỆ SINH HỌC 2024



Hình 5. Các vị trí đa hình (SNPs) trên gene cOR9S13 của 20 chó nghiên cứu

Theo các công bố trước đây, 6 SNPs trên gene cOR9S13 liên quan đến khả năng khứu giác của chó bao gồm các vị trí 331, 336, 465, 592, 759, và 762, tương ứng với các vị trí amino acid 111, 112, 155, 198, 253 và 254 (Lesinak *et al.*, 2008). Ở các vị trí này nếu xuất hiện kiểu gene đồng hợp tử đột biến (AA) dẫn đến thay đổi amino acid thì sẽ làm mất đi khả năng khứu giác của chó. Kết quả giải trình tự gene cho thấy cả 20 chó nghiên cứu đều không có đột biến xuất hiện kiểu gene đồng hợp tử AA ở cả 6 vị trí này. Trong 8 SNPs được phát hiện ở 20 cá thể chó nghiên cứu, chỉ có 1 SNP G592A liên quan đến khả năng khứu giác của chó.

Đột biến trong gene cOR9S13 ở vị trí 592 từ G thành A (dẫn đến thay đổi từ Alanine thành Threonine ở vị trí 198 tương ứng với vùng chức năng EC2) đã được chứng minh có ảnh hưởng có ý nghĩa thống kê (p<0,001) về hiệu suất đánh hơi của những chó săn Đức (German shepherd) trong thời kỳ huấn luyện (Lesinak *et al.*, 2008; Sacharczuk *et al.*, 2019). Trong 20 cá thể chó kiểm tra có 1/20 con mang đột biến bất lợi ở vị trí này. Vị trí thay đổi này đã được xác định có tác động tiêu cực đến khả năng khứu giác của chó (Lesinak *et al.*, 2008; Sacharczuk *et al.*, 2019). Kết quả của chúng tôi cũng phù hợp với các kết quả nghiên cứu của thế giới khi không phát hiện thấy có đột biến thay thế amino acid ở 5 vị trí còn lại (111Leu, 112Leu, 155Cys, 253Val và 254Thr).

Tổng hợp kết quả phân tích đa hình trên hai gene cOR52H9 và cOR9S13

Trong tổng số 14 SNPs liên quan đến khả năng khứu giác của chó (8 SNPs của gene cOR52H9 và 6 SNPs của gene cOR9S13), có 2/20 cá thể chó (số 14 và số 20) mang đột biến bất lợi ở mức độ vừa tại vị trí 680 (AA) và 1040 (TT) của gene cOR52H9; và 1/20 cá thể chó (số 11) có đột biến ở vị trí 592 của gene cOR9S13 (G>A) dẫn đến sự thay đổi về amino acid Ala>Thr, là đột biến ảnh hưởng xấu đến khứu giác của chó. Các chó còn lại đều không có đột biến bất lọi, cho thấy có nguồn gene tốt. Đây là điều thuận lợi cho công tác huấn luyện chó của đơn vị.

Chó Malinois của Bỉ được biết đến là loài chó có thân hình đẹp, cơ bắp, có sức bền và nhanh nhẹn. Bên cạnh đó giống chó này còn có bản tính rất thông minh, linh hoạt, khả năng cảnh giác và định hướng cao do khứu giác rất nhạy; đồng thời có khả năng huấn luyện cao. Chính vì vậy chó giống Malinois được đánh giá là có khả năng làm việc xuất sắc, đặc biệt trong các vai trò như cảnh sát, tìm kiếm cứu nạn, các hoạt động quân sự, phát hiện ma túy, chất nổ... Lựa chọn sớm được các chó tiềm năng có thể giảm chi phí và công sức, thời gian của các chương trình huấn luyện chó (Yang *et al.*, 2016).

Các gen OR của chó có tính đa hình (Robin *et al.*, 2009). Sự lựa chọn tích cực trên các gen OR dường như đã thúc đẩy sự tiến hóa của các gen OR và sự thay đổi khả năng khứu giác (Benbernou *et al.*, 2011). Do yêu cầu cấp thiết trong lĩnh vực an ninh và quân sự, trên thế giới đã có nhiều nghiên cứu và thảo luận về ảnh hưởng của các gene OR lên khả năng khứu giác của chó (Tacher *et al.* 2005, Lesniak *et al.* 2008, Yang *et al.* 2016). Các nghiên cứu đã chỉ ra rằng các SNPs của các gen OR có thể ảnh hưởng đến khả năng khứu giác của các giống chó khác nhau (Tacher *et al.* 2005, Lesniak *et al.* 2008, Yang *et al.* 2016). Từ đó đã đề xuất vai trò của các alen đặc hiệu trong phát hiện mùi và mối liên kết giữa đa hình nucleotide đơn (SNPs) và hiệu quả nhận biết mùi.

Khả năng phát hiện mùi của chó nghiệp vụ phụ thuộc vào nhiều yếu tố: môi trường làm việc, công tác huấn luyện và yếu tố di truyền của mỗi cá thể, trong đó khả năng phát hiện mùi của chó phụ thuộc vào nhiều gene (Sacharczuk *et al.*, 2019; Robin *et al.*, 2009). Việc phối hợp tốt giữa các yếu tố trên sẽ là cơ sở để tạo được những cá thể chó nghiệp vụ có chất lượng tốt, hiệu suất làm việc cao. Bên cạnh đó việc lựa chọn được các cá thể thuần chủng có kiểu gen tốt còn có giá trị đặc biệt trong công tác bảo tồn nguồn gen và cung cấp những cá thể chó tiềm năng cho công tác huấn luyện.

KÉT LUẬN

Nghiên cứu đã giải mã thành công hai gene cOR52H9 và cOR9S13 của 20 cá thể chó giống Malinois. Trình tự gene thu được tương đồng cao (99.3% đến 100%) với các trình tự gene tương ứng trên Ngân hàng gene. Gene cOR52H9 và cOR9S13 trên 20 cá thể chó nghiên cứu có 8 điểm đa hình đơn nucleotide (SNP). Trong đó 2/20 cá thể chó có mang 02 SNPs bất lợi ở mức độ vừa đối với khả năng phát hiện mùi trên gen cOR52H9 (ở vị trí G680A và G1042T: là AA và TT); và 1/20 cá thể chó có một SNP bất lợi đến khả năng khứu giác của chó trên gen cOR9S13 (ở vị trí G592A: G>A dẫn đến thay đổi amino acid Ala>Thr). Cần tiếp tục mở rộng nghiên cứu trên nhiều gene khác để hoàn thiện kỹ thuật chọn lọc giống chó nghiệp vụ đạt chất lượng như mong muốn.

Lời cảm ơn: Nghiên cứu được tài trợ từ Đề tài "Nghiên cứu thuần chủng bảo tồn, phát triển nguồn gene chó giống Malinois phục vụ công tác nghiệp vụ của lực lượng Công an nhân dân", mã số: BH.2018.K02.01.

TÀI LIỆU THAM KHẢO

Benbernou N, Robin S, Tacher S, Rimbault M, Rakotomanga M, Galibert F, Notes A (2011). cAMP and IP3 signaling pathways in HEK293 cells transfected with canine olfactory receptor genes. *J Hered* 102 (Suppl 1): S47–S61.

Buck L, and Axel R (1991). A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1), 175-187.

Do Van Thu, Doan Viet Binh, Nguyen Trong Chi, Le Xuan Phong, Nguyen Ngoc Hung, Le Thi Hue, Tran Xuan Khoi, Vo Thi Ninh (2017). Phân tích đa hình và đánh giá tương quan di truyền gene thụ thể khứu giác cOR52H9 liên quan đến khả năng phát hiện mùi của chó nghiệp vụ. *Tạp chí Sinh học*, 39(1).

Lesniak A, Walczak M, Jezierski T, Sacharczuk M, Gawkowski M, Jaszczak K (2008) Canine Olfactory Receptor Gene Polymorphism and Its Relation to Odor Detection Performance by Sniffer Dogs. *J Hered*, 99(5): 518-527.

Liu AH, Zhang X, Stolovitzky GA, Califano A, Firestein SJ (2003) Motifbased construction of a functional map for mammalian olfactory receptors. *Genomics*, 81: 443-456.

Maejiima M, Inoue-Murayama M, Tonosaki K, Matsuura N, Kato S, Saito Y, Weiss A, Murayama Y, Ito S (2007) Traits and genotypes may predict the successful training of drug detection dogs. Applied Animal Behavior Science, 107: 287–298.

Malnic B, Godfrey PA, Buck LB (2004). The human olfactory receptor gene family. PNAS, 101(8), 2584-2589.

Olender T, Fuchs T, Linhart C, Shamir R, Adams M, Kalush F, Keh M, Lancet D (2004) The canine olfactory subgenome. *Genomics.* 83: 361-372.

Robin S, Tacher S, Rimbault M, Vaysse A, Dréano S, André C, Hitte C, Galibert F (2009) Genetic diversity of canine olfactory receptors. *BMC Genom*, 10:21.

Tacher S, Quignon P, Rimbault M, Dréano S, André C, Galibert F (2005) Olfactory Receptor Sequence Polymorphism Within and Between Breeds of Dogs. *J Hered*, 96(7):812-816.

Yang M, Zhang HX, Geng GJ, Wang FJ, Liu CW, Liu JL (2022). Artificial Selection Drives SNPs of Olfactory Receptor Genes into Different Working Traits in Labrador Retrievers. *Genet Res*, 8319396.

Yang M, Geng GJ, Zhang W, Cui L, Zhang HX, Zheng JL (2016). SNP genotypes of olfactory receptor genes associated with olfactory ability in German Shepherd dogs. *Animal Genet*, 47 (2): 240-244.

Sacharczuk M, Walczak M, Adamkiewicz E, Walasek A, Ensminger J, Presch M, and Jezierski T (2019) Polymorphism of olfactory and neurotransmitters receptor genes in drug and explosives detection dogs can be associated with differences in detection performance. *Appl Animal Behav Sci, 215,* 52–60.

Zhao H, Firestein S (1999). Vertebrate odorant receptors. Cell Mol Life Sci CMLS, 56, 647-659.

SINGLE NUCLEOTIDE POLYMORPHISM ANALYSIS OF OLFACTORY RECEPTORS cOR52H9 AND cOR9S13 IN POLICE DOGS

Le Thi Hue¹, Pham Thi Khanh Linh^{1,2}, Nguyen Ngoc Hung³, Le Van Trong³, Doan Thi Thanh Huong^{1,2*}

¹Institute of Biotechnology, Vietnam Academy of Science and Technology

²University of Science and Technology of Hanoi, Vietnam Academy of Science and Technology

³Mobile Police High Command, Ministry of Public Security

SUMMARY

Some dog species are utilized by the military and police because they possess extraordinary olfactory senses, much more sensitive than humans. Thanks to this characteristic, they can detect drugs, explosives, and even track individuals, making dogs an essential part of many military and police operations, significantly contributing to public security and safety. Single nucleotide polymorphisms (SNPs) in olfactory receptor genes are key factors contributing to differences in olfactory abilities among dog species. These genetic variations can significantly impact a dog's ability to detect and differentiate between different scents. Therefore, applying molecular biology methods can significantly improve the selection and training of highly effective professional dog teams. In this study, we amplified and analyzed SNPs in the cOR52H9 and cOR9S13 genes of 20 trained dogs of the Malinois breed from the Mobile Police High Command, Ministry of Public Security. The sequence results show that the gene sequences cOR52H9 and cOR9S13 have a high similarity rate of over 99% compared to sequences deposited in the gene bank. Gene cOR52H9 exhibits 8 Single Nucleotide Polymorphisms (SNPs) at positions: G58A, A180G, G250A, T315A, G598A, T642G, G678A, and G1042T. Gene cOR9S13 has 8 SNPs at positions: T125A, C375T, T390C, G398A, C416T, C546G, C566A, and G592A. Among the identified SNPs, only one dog showed a disadvantageous SNP at position 592 of the cOR9S13 gene (G592A), which regulates its sniffing ability; meanwhile, 2 individuals exhibited 2 disadvantageous SNPs, albeit to a moderate extent. Research using SNPs in genes encoding olfactory receptors of dogs will help selectively breed dogs with the best scent detection capabilities for training purposes.

Keywords: Working dogs, cOR52H9, single nucleotide polymorphism, olfactory receptor gene.

^{*} Author for correspondence: Tel: 0988904605; Email: doantthuong74@gmail.com

PHÂN TÍCH CÁC ĐẶC ĐIỂM CỦA HỌ PROTEIN VACUOLAR IRON TRANSPORTER (VIT) VÀ DỮ LIỆU BIỂU HIỆN HỌ GEN *OsVIT* Ở LÚA BẰNG CÔNG CỤ TIN SINH HỌC

Lê Quỳnh Mai^{*}, Phạm Minh Duy, Đỗ Minh Ân

Khoa Sinh học, Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

TÓM TẮT

Sắt là một trong các nguyên tố thiết yếu với sinh trưởng và phát triển của thực vật. Để hấp thu, vận chuyển và kiểm soát sự tích lũy sắt trong cơ thể, thực vật sử dụng nhiều protein khác nhau trong đó có protein vận chuyển ở không bào (Vacuolar Iron Transporter - VIT). Nhờ các protein ho VIT, sắt có thể được đưa vào tích trữ ở không bào, giúp giảm tránh các tác động bất lợi của hiện tượng dư thừa sắt. Nhóm protein này còn có các đồng hình là Vacuolar Iron Transporter-like (VTL) cũng tham gia quá trình vận chuyển sắt vào không bào. Các VIT và VTL được gọi chung là họ VIT. Ở lúa gao, mới chỉ có hai protein VIT là OsVIT1 và OsVIT2 được phát hiện và phân tích. Trong nghiên cứu này, 9 protein họ VIT gồm VIT1, VIT2 và 7 VTL được xác định ở lúa trên cơ sở so sánh với EgVIT1, đồng thời được phân tích cùng các protein họ VIT ở cây một lá mầm khác là ngô và ở cây mô hình Arabidopsis thaliana. Các dẫn liêu cho thấy các protein họ VIT ở lúa dài từ 69 đến 252 aa, khối lượng phân tử từ 7,263 đến 26,551 kDa, điểm đẳng điện trong khoảng 4,69 đến 8,57. Hầu hết các protein họ VIT ở lúa đều có 5 vùng xuyên màng và đều được xác định biểu hiện ở màng không bào. Trình tự các protein được so sánh, xác đinh các motif tương đồng từ đó xây dựng cây phát sinh cho thấy mối quan hê gần gũi giữa các protein VIT ở cây một lá mầm so với cây hai lá mầm. Dữ liệu biểu hiện của 4 trong số các gen họ VIT ở lúa cho thấy các gen OsVIT2, OsVTL1 và OsVTL2 có biểu hiện mạnh ở hạt trong giai đoạn sớm sau đó giảm dần khi hạt phát triển. Đặc biệt, OsVIT1 biểu hiện mạnh ở rễ và lá mà không biểu hiện ở bất kỳ giai đoạn nào của hạt. OsVIT1 có tiềm năng là gen đích cho các nghiên cứu cải tiến hàm lương sắt trong hat.

Từ khóa: Họ gen OsVIT, không bào, mức độ biểu hiện gen, vacuolar iron transporter, vận chuyển sắt.

ĐẶT VẤN ĐỀ

Ở thực vật, nguyên tố sắt thiết yếu cho sinh trưởng và phát triển, liên quan đến các quá trình như quang hợp, tổng hợp diệp lục, chuỗi truyền điện tử và nhiều phản ứng oxi hóa khử (Aung and Masuda, 2020). Tuy nhiên, tích tụ quá nhiều sắt trong tế bào chất dẫn đến sự hình thành các gốc oxi hóa phản ứng (reactive oxygen species-ROS) qua phản ứng Fenton, làm phá hủy DNA và các bào quan (Winterbourn, 1995; Gayomba *et al.*, 2015). Sắt dư thừa được thực vật chuyển đến các cơ quan dự trữ nhằm tránh gây tổn thương các mô. Không bào cũng được thực vật sử dụng để dự trữ sắt. Họ protein VIT (Vacuolar Iron Transporter) gồm các thành viên được xác định tham gia vận chuyển ion sắt vào không bào ở thực vật. Họ protein này gồm hai nhóm chính: nhóm VIT và nhóm VIT-like (VTL). Các protein nhóm VTL khác với VIT ở điểm không có không có vùng chức năng là domain bám kim loại (metal binding domain-MBD) (Gollhofer *et al.*, 2014; Walton *et al.*, 2020).Trình tự các protein VIT trong nhóm cây côn lại. Các protein thuộc họ VIT ở các loài thực vật tham gia vào nhiều hoạt động khác nhau (Connorton *et al.*, 2017; Cao, 2019).

AtVIT1 là thành viên đầu tiên của họ protein VIT được phát hiện từ cây mô hình *Arabidopsis thaliana* (Kim *et al.*, 2006). OsVIT1 và OsVIT2 ở lúa gạo (*Oryza sativa*) được xác định biểu hiện mạnh ở lá và tham gia vận chuyển sắt từ cơ quan hấp thu (rễ) đến cơ quan tích trữ (lá, hạt) (Zhang *et al.*, 2012). Tuy có nhiều nghiên cứu về chức năng nhưng cấu trúc bậc ba cũng như cơ chế vận chuyển sắt qua protein VIT lại ít được làm rõ. Nghiên cứu đầu tiên về cấu trúc của VIT đã xác định cấu trúc tinh thể của EgVIT1, protein VIT từ *Eucalyptus grandis* (Kato *et al.*, 2019). Cấu trúc tinh thể của EgVIT1 đã khẳng định dự đoán cấu hình bậc hai của các protein VIT, gồm 5 vùng xoắn xuyên màng và domain bám ion kim loại (MBD) nằm giữa hai vùng xuyên màng thứ 2 và thứ 3. Hoạt động của các protein VIT khác dựa trên sự tương đồng trình tự với EgVIT1 cũng đã được dự đoán trong nghiên cứu này của Kato và cộng sự (Kato *et al.*, 2019). Một nghiên cứu khác cũng đã sử dụng mô hình tương đồng với EgVIT1 để xây dựng cấu trúc ba chiều của AtVIT1 và OsVIT1 (Krishna *et al.*, 2023).

Ở lúa gạo, cho đến nay chỉ có OsVIT1 và OsVIT2 là hai gen VIT được xác định và phân tích, chưa có gen nào thuộc nhóm VTL ở lúa được nghiên cứu. Nghiên cứu của Zhang và đồng tác giả (2012) cho thấy OsVIT1 và

OsVIT2 đều biểu hiện mạnh ở lá, các protein OsVIT1 và OsVIT2 đều nằm trên màng không bào. Tuy nhiên mức độ biểu hiện của hai gen này thay đổi khác nhau với điều kiện môi trường thừa hoặc thiếu sắt. Biểu hiện gen OsVIT2 tăng mạnh trong môi trường thừa sắt và giảm mạnh khi cây bị thiếu sắt, còn mức độ biểu hiện OsVIT1 không có sự biến động lớn trong vòng 7 ngày từ khi xử lý sắt dù là tăng hay giảm hàm lượng. Nghiên cứu cũng đã thực hiện chuyển gen nhằm loại bỏ các gen OsVIT, kết quả cho thấy các cây chuyển gen có hàm lượng sắt giảm đi ở lá nhưng lại tăng lên ở phôi, đồng thời khả năng sinh trưởng không bị ảnh hưởng. Zhang và đồng tác giả (2012) dự đoán rằng OsVIT1 và OsVIT2 giúp điều hoà phân bố sắt giữa các mô trong cây.

Trong bài viết này các đặc điểm của họ protein VIT ở lúa gạo được xác định và trên cơ sở so sánh các motif của các protein này cùng với các protein cùng họ ở ngô (*Zea mays*) và *A. thaliana*, cây phát sinh phân nhóm các protein họ VIT đã được dựng cho thấy mối quan hệ gần gũi giữa chúng. Mức độ biểu hiện của một số gen trong họ *VIT* ở lúa cũng được khai thác để làm sáng tỏ phần nào vai trò của các VIT tương ứng với sự tích lũy sắt ở các mô, cơ quan. Các dẫn liệu này có thể được sử dụng để lựa chọn gen đích phù hợp cho các nghiên cứu cải tiến hàm lượng dinh dưỡng trong hạt gạo.

VẬT LIỆU VÀ PHƯƠNG PHÁP

Vật liệu nghiên cứu

Dữ liệu về protein và gen thuộc họ Vacuolar Iron Transporter (VIT) của lúa, ngô và *A. thaliana* trên cơ sở dữ liệu Phytozome (Goodstein *et al.*, 2012). Dữ liệu về biểu hiện gen được khai thác từ cơ sở dữ liệu GEO trên NCBI có mã số GSE6893.

Phương pháp nghiên cứu

Xác định và dự đoán một số thông số hóa lý các protein thuộc họ VIT ở lúa gạo

Trình tự gióng hàng rút gọn (seed) của domain VIT1 (mã: PF01988) từ cơ sở dữ liệu Pfam (Mistry *et al.*, 2021) được sử dụng để tìm kiếm các trình tự protein thuộc họ VIT trên proteome: Oryza sativa v.70-rice trong cơ sở dữ liệu Phytozome bằng công cụ BLAST (Goodstein *et al.*, 2012) với ngưỡng giá trị E-1. Các trình tự thu được từ tìm kiếm trên được kiểm tra bằng công cụ MotifFinder để khẳng định có motif thuộc domain VIT1. Tên các protein, số vùng xuyên màng và dự đoán vị trí trong tế bào được tìm kiếm trên cơ sở dữ liệu UniProt (2022) qua các mã gen tương ứng. Các thông số hóa lý như: khối lượng phân tử (MW), điểm đẳng điện (pl), độ kị nước trung bình (GRAVY), chỉ số bất ổn định (II) được tính toán bằng công cụ ProtParam (Wilkins *et al.*, 1999).

Xây dựng cây phát sinh chủng loại và xác định các motif bảo thủ của họ protein VIT ở lúa, ngô và A. thaliana

Trình tự các thành viên họ protein VIT được tìm kiếm ở *O. sativa, A. thaliana* và *Z. mays* cùng với trình tự của EgVIT1 (mã UniProtKB: P0DO17) được gióng hàng bằng công cụ MEGA11 (Tamura *et al.*, 2021) theo phương pháp MUSCLE. Cây phát sinh chủng loại của họ protein VIT được xây dựng từ thông tin sắp gióng cột nhiều trình tự bằng công cụ IQ-TREE (Nguyen *et al.*, 2015) theo phương pháp Maximum Likelihood, 1000 bootstrap UFBoot (Hoang *et al.*, 2018) và được biểu diễn bằng công cụ iTOL (Letunic and Bork, 2024). Các motif bảo thủ ở các protein VIT được xác định bằng công cụ MEME với các thông số: độ dài motif tối thiểu: 6; độ dài motif tối đa: 50; ngưỡng E-value tối đa: E-1. Vị trí các motif trên trình tự các protein VIT được biểu diễn bằng công cụ TBtool (Chen *et al.*, 2023).

Phân tích dữ liệu biểu hiện họ gen VIT ở lúa gạo

Biểu hiện họ gen *VIT* ở lúa gạo qua các giai đoạn phát triển khác nhau được phân tích thông qua dữ liệu microarray (mã GEO: GSE6893). Dữ liệu biểu hiện các gen có trong bộ dữ liệu này được lấy log2, chuẩn hóa và biểu diễn dưới dạng bản đồ nhiệt (heatmap) bằng công cụ TBtool (Chen *et al.*, 2023).

KẾT QUẢ VÀ THẢO LUẬN

Các protein VIT ở lúa vừa tương đồng vừa đa dạng so với ở ngô và Arabidopsis

Thông tin gióng hàng rút gọn của domain VIT1 (mã Pfam: PF01988) được dùng làm truy vấn trên cơ sở dữ liệu Phytozome bằng công cụ BLAST đối với proteome của lúa *O. sativa.*

Kết quả đã xác định được 9 protein ở lúa với các đặc điểm dự đoán ở Bảng 1. Họ protein VIT ở lúa gồm 2 thành viên thuộc nhóm VIT là OsVIT1 và OsVIT2 và 7 thành viên thuộc nhóm VTL. Các protein nhóm VIT đã được xác định có độ dài và kích thước phân tử ít chênh lệch (OsVIT1 dài 252 aa, khối lượng phân tử khoảng 26,5 kDa và OsVIT2 dài 246 aa, khối lượng phân tử gần 26,2 kDa). Còn ở nhóm VTL, các thông số độ dài (69-249 amino acid) và kích thước phân tử (7,2-25,1 kDa) có phân bố rộng hơn. Hầu hết các protein thuộc họ VIT ở lúa trên đều có 5 vùng xuyên màng (trừ OsVTL6 và OsVTL7 không có thông tin). Các phân tử được xác định có điểm đẳng điện dao động từ 4,69 đến 8,57. Phần lớn các protein có giá trị GRAVY dương (8/9 phân tử) thể hiện là các phân tử kị nước. Có 5/9 phân tử có chỉ số bất ổn định thấp dưới 40. Các protein có thông tin về vị trí nội bào trên cơ sở dữ liệu UniProtKB đều ở màng không bào.

	Bang 1. mong un cac protein thuộc hộ VIT ở Oryza Sativa											
Tên protein	UniProtKB ID	Độ dài (a. a.)	MW (Da)	Số vùng xuyên màng	рІ	GRAVY	II	Vị trí nội bào				
OsVIT1	Q6MWE5	252	26551	5	5,73	0,221	35,97	ton				
OsVIT2	Q6ERE5	246	26161	5	4,9	0,268	39,04	ton				
OsVTL1	Q6H658	232	22876	5	5,02	0,499	44,76	ton				
OsVTL2	B7F138	189	19107	5	6,55	0,685	32,16	ton				
OsVTL3	Q84ZM7	249	25100	5	7,70	0,336	61,42	ton				
OsVTL4	Q53PN2	216	21236	5	7,97	0,560	37,43	ton				
OsVTL5	Q7XTL7	208	20981	5	8,57	0,623	28,91	ton				
OsVTL6	Q10M16	87	9330	-	5,45	-0,387	68,51	-				
OsVTL7	Q2QWM2	69	7263	-	4,69	0,130	41,17	-				

Chú thích: MW: Khối lượng phân tử, pl: Điểm đẳng điện, GRAVY: Độ kị nước trung bình, II: Chỉ số bất ổn định, ton: Màng không bào; "-": Không có thông tin trên cơ sở dữ liệu.

Cây phát sinh chủng loại và các motif bảo thủ ở họ protein VIT

Cây phát sinh chủng loại của họ protein VIT được xây dựng từ 23 trình tự protein được tìm kiếm ở 3 loài gồm 9 protein ở lúa (*O. sativa*), 8 protein ở ngô (*Z. mays*) và 6 protein ở *A. thaliana* cùng với trình tự của EgVIT1 từ *E. grandis* (Hình 1). Cây không gốc thu được gồm 2 nhánh tách biệt với độ tin cậy tuyệt đối (chỉ số bootstrap 100): nhánh I gồm các protein thuộc nhóm VIT và nhánh II gồm các protein nhóm VTL. Trong nhánh I, 2 protein VIT từ 2 loài cây hai lá mầm là AtVIT1 và EgVIT1 tạo thành một nhánh con với độ tin cậy cao (chỉ số bootstrap 91). Những protein VIT còn lại trong nhánh I thuộc về 2 loài một lá mầm (*O. sativa* và *Z. mays*) cũng tạo thành một nhánh con với độ tin cậy trung bình (chỉ số bootstrap 65). Điều này cho thấy nhóm protein VIT của cùng nhóm cây một lá mầm hoặc hai lá mầm gần gũi với nhau hơn so với nhóm còn lại. Xu hướng khác biệt giữa cây một lá mầm và hai lá mầm ở họ protein VIT cũng có thể thấy ở nhánh II - nhóm VTL. 5 AtVTL tạo thành phân nhánh Ila còn các VTL từ *O. sativa* và *Z. mays* tạo thành phân nhánh IIb, tuy nhiên độ tin cậy của 2 phân nhánh này đều không cao. Xu hướng khác biệt này cũng đã được nhận thấy trong các nghiên cứu trước đây (Connorton *et al.*, 2017; Cao, 2019).



Hình 1. Cây phát sinh chủng loại họ protein VIT ở lúa, ngô và A. thaliana

Các chỉ số bootstrap được hiển thị cạnh mỗi nhánh của cây phát sinh chủng loại, những chỉ số bootstrap dưới 50 không được hiển thị.



Hình 2. Phân bố các motif bảo thủ của họ protein VIT ở lúa, ngô và A. thaliana

Tìm kiếm bằng công cụ MEME đã xác định được 10 motif bảo thủ có ý nghĩa thống kê (E-value < 0.1) ở 24 protein kể trên. Sự phân bố của các motif này trên các trình tự protein được thể hiện trong Hình 2. Kết quả này phù hợp với cây phát sinh chủng loại khi các protein thuộc cùng nhánh trên cây phát sinh chủng loại thì sự phân bố motif của chúng tương đồng với nhau hơn. Cụ thể, phân bố các motif 2-1-10-5-7-3-6-4 gần như bảo thủ ở nhóm VIT, riêng ZmVIT1 không có các motif số 2, số 10 và số 5. Cụm các motif 2-1-3-5-4 bảo thủ ở nhóm các AtVTL, trong khi ở các VTL còn lại, chỉ có motif số 5 xuất hiện ở tất cả các thành viên. Điều này cũng cho thấy trình tự các VTL có sự đa dạng cao hơn so với nhóm VIT.

Kết quả phân tích dữ liệu biểu hiện họ gen VIT ở lúa gạo

Dữ liệu biểu hiện của họ gen VIT ở lúa gạo trong các giai đoạn phát triển khác nhau được trích xuất và phân tích từ bộ dữ liệu microarray (mã GEO: GSE6893), các mức độ biểu hiện được đánh giá là có ý nghĩa so với đối chứng trong các nghiên cứu gốc. Có 4 thành viên họ gen VIT ở lúa gạo có dữ liệu biểu hiện gen trong bộ dữ liệu, gồm OsVIT1, OsVIT2, OsVTL1 và OsVTL2, biểu hiện của các gen này được biểu diễn dưới dạng bản đồ nhiệt (Hình 3). Nhìn chung, các gen này có sự khác biệt với nhau về phân bố biểu hiện gen. OsVIT1 và OsVTL2 biểu hiện mạnh ở rễ, còn biểu hiện của OsVIT2 và OsVTL1 ở rễ chỉ ở mức trung bình và thấp. OsVIT1 và OsVIT2 đều biểu hiện mạnh ở lá, tương đồng với kết quả nghiên cứu của Zhang và đồng tác giả (2012). Các gen VIT hầu như đều biểu hiện thấp ở hoa giai đoạn sớm (Young inflorescence Inflorescence P3), sau đó tăng lên khi hoa phát triển. OsVTL1, OsVIT2 và OsVIT2 biểu hiện mạnh ở hạt giai đoạn sớm (Seed S1-Seed S3), giảm dần khi hạt phát triển. Ngược lại, OsVIT1 biểu hiện thấp ở tất cả các giai đoạn của hạt.



Hình 3. Bản đồ nhiệt thể hiện mức độ biểu hiện các gen họ VIT ở lúa gạo

Mức độ biểu hiện được thể hiện theo thang màu bên phải. Chú thích: SAM (shoot apical meristem): mô phân sinh đỉnh ngọn, leaf: lá, root: rễ, inflorescence: hoa, seed: hạt, seedling: cây con, young: non, mature: trưởng thành, P2-P6: các giai đoạn hoa theo thứ tự thời gian, S1-S5: các giai đoạn hạt theo thứ tự thời gian.

KÉT LUẬN

Họ protein VIT ở lúa gạo đã được phân tích *in silico* một cách toàn diện. Các thông số hóa lý, vị trí trong tế bào, cây phát sinh chủng loại và các motif bảo thủ ở 9 thành viên họ protein VIT ở lúa gạo cùng với họ protein VIT ở ngô và *A. thaliana* đã được phân tích, cho thấy nhóm protein VIT có sự bảo thủ cao và các protein VIT của cây một lá mầm có xu hướng gần gũi với nhau hơn so với ở cây hai lá mầm. Dữ liệu biểu hiện của một số gen *VIT* ở lúa đã được thu thập và phân tích cho thấy vai trò tiềm năng trong việc làm gen đích cho các nghiên cứu cải tiến hàm lượng sắt trong hạt. Cụ thể, dữ liệu biểu hiện cho thấy các gen *OsVIT2, OsVTL1* và *OsVTL2* có biểu hiện mạnh ở hạt trong giai đoạn sớm sau đó giảm dần khi hạt phát triển. Biểu hiện của gen *OsVIT1* mạnh ở rễ từ giai đoạn sóm, tăng dần theo sự phát triển của lá nhưng đặc biệt giai đoạn nào của hạt cũng có mức biểu hiện rất thấp. *OsVIT1* có tiềm năng là gen đích cho các nghiên cứu cải tiến hàm lượng sắt trong hạt.

TÀI LIỆU THAM KHẢO

Aung M S and H Masuda (2020) How Does Rice Defend Against Excess Iron?: Physiological and Molecular Mechanisms. *Front Plant Sci.* 11: 1102.

Cao J (2019) Molecular Evolution of the Vacuolar Iron Transporter (*VIT*) Family Genes in 14 Plant Species. *Genes* (Basel) 10 (2): 144.

Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, Liu Y, Feng J, Chen H, He Y and Xia R (2023) TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol Plant* 16 (11): 1733-1742.

Connorton JM, Jones ER, Rodríguez-Ramiro I, Fairweather-Tait S, Uauy C and Balk J (2017) Wheat Vacuolar Iron Transporter TaVIT2 Transports Fe and Mn and Is Effective for Biofortification. *Plant Physiol.* 174 (4): 2434-2444.

Gayomba SR, Zhai Z, Jung HI and Vatamaniuk OK (2015) Local and systemic signaling of iron status and its interactions with homeostasis of other essential elements. *Front Plant Sci.* 6: 716.

Gollhofer J, Timofeev R, Lan P, Schmidt W and Buckhout TJ (2014) Vacuolar-Iron-Transporter1-Like proteins mediate iron homeostasis in Arabidopsis. *PLoS One* 9 (10): e110468.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N and Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40 (Database issue): D1178-86.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ and Vinh LS (2018) UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol.* 35 (2): 518-522.

Kato T, Kumazaki K, Wada M, Taniguchi R, Nakane T, Yamashita K, Hirata K, Ishitani R, Ito K, Nishizawa T and Nureki O (2019) Crystal structure of plant vacuolar iron transporter VIT1. *Nat Plants* 5 (3): 308-315.

Kim SA, Punshon T, Lanzirotti A, Li L, Alonso JM, Ecker JR, Kaplan J and Guerinot ML (2006) Localization of iron in Arabidopsis seed requires the vacuolar membrane transporter VIT1. *Science* 314 (5803): 1295-8.

Krishna TPA, Ceasar SA and Maharajan T (2023) Biofortification of Crops to Fight Anemia: Role of Vacuolar Iron Transporters. *J Agric Food Chem.* 71 (8): 3583-3598.

Letunic I and Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 52 (W1): W78-W82.

Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD and Bateman A (2021) Pfam: The protein families database in 2021. *Nucleic Acids Res.* 49 (D1): D412-D419.

Nguyen LT, Schmidt HA, von Haeseler A and Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32 (1): 268-74.

Walton JH, Kontra-Kováts G, Green RT, Domonkos Á, Horváth B, Brear EM, Franceschetti M, Kaló P and Balk J (2020) The *Medicago truncatula* Vacuolar iron Transporter-Like proteins VTL4 and VTL8 deliver iron to symbiotic bacteria at different stages of the infection process. *New Phytol.* 228 (2): 651-666.

Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD and Hochstrasser DF (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* 112: 531-52.

Winterbourn CC (1995) Toxicity of iron and hydrogen peroxide: the Fenton reaction. Toxicol Lett. 82-83: 969-74.

Tamura K, Stecher G and Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 38 (7): 3022-3027.

The UniProt Consortium (2022) UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research 51 (D1): D523-D531.

Zhang Y, Xu YH, Yi HY and Gong JM (2012) Vacuolar membrane transporters OsVIT1 and OsVIT2 modulate iron translocation between flag leaves and seeds in rice. *Plant J.* 72 (3): 400-10.

ANALYSIS OF CHARACTERISTICS OF VACUOLAR IRON TRANSPORTER (VIT) PROTEIN FAMILY AND EXPRESSION OF OsVIT GENE FAMILY IN RICE USING BIOINFORMATICS TOOLS

Le Quynh Mai*, Pham Minh Duy, Do Minh An

Faculty of Biology, University of Science, Vietnam National University in Hanoi

SUMMARY

Iron is one of the essential elements for plant growth and development. To absorb, transport and control iron accumulation, plants use many different proteins, including vacuolar iron transporters (VIT). By VIT proteins, iron can be stored in vacuoles, helping to reduce the negative effects of iron excess. This group of proteins also has isoforms, Vacuolar Iron Transporter-like (VTL) proteins, which also participates in the process of transporting iron into the vacuole. VITs and VTLs are collectively known as the VIT family. In rice, only two VIT proteins, OsVIT1 and OsVIT2, have been discovered and analyzed. In this study, 9 VIT family proteins in rice including VIT1, VIT2 and 7 VTLs were identified based on comparison with EgVIT1, and continuously compared with VIT family proteins in an other monocotyledonous plant, maize and model plant of Arabidopsis thaliana. Data show that the VIT family proteins in rice were from 69 to 252 aa long, molecular weight from 7.263 to 26.551 kDa, and isoelectric point from 4.69 to 8.57. Most of the VIT family proteins in rice have five transmembrane domains and were determined expression in the tonoplast. The protein sequence comparison and homologous motif identification showed a close relationship between VIT proteins in monocot plants compared to dicot plants. A phylogenetic tree was built. The expression levels of four of the VIT family genes in rice showed that the OsVIT2, OsVTL1 and OsVTL2 genes over-expressed in seeds during the early stages and then gradually decreased the expression during the developing of seeds. In particular, OsVIT1 was strongly expressed in roots and leaves but did not express in seeds at any stage. OsVIT1 has potential as a target gene for studies to improve iron content in rice grains.

Keywords: OsVIT gene family, vacuole, gene expression level, vacuolar iron transporter (VIT), iron transport.

^{*} Author for correspondence: Tel: 0947485588; Email: lequynhmai80@vnu.edu.vn / lequynhmai80@gmail.com

NGHIÊN CỨU MỐI QUAN HỆ DI TRUYỀN CỦA MỘT SỐ GIỐNG HOA HỒNG (*Rosa* spp.) CỔ Ở MIỀN BẮC VIỆT NAM DỰA TRÊN TRÌNH TỰ VÙNG *trn*H-*psb*A

Ngô Văn Đức, Hoàng Hải Đăng, Triệu Phương Mai, Phạm Quốc Toàn, Nguyễn Thị Bích Hường, Lê Thị Kiều Trinh, Nguyễn Minh Phương

Trường Đại học Công Thương Thành phố Hồ Chí Minh

TÓM TẮT

Nghiên cứu này được thực hiện nhằm giải và phân tích trình tự vùng *trnH-psbA*, từ đó đánh giá mối quan hệ di truyền của các giống/loài hoa hồng được trồng lâu đời ở miền Bắc Việt Nam. Mồi khuếch đại DNA được thiết kế dựa trên trình tự DNA chi *Rosa* tham khảo từ NCBI Genbank. DNA tổng số được ly trích từ mẫu lá bằng phương pháp CTAB, từ đó vùng *trnH-psbA* được khuếch đại bằng phương pháp PCR và giải trình tự bằng phương pháp giải trình tự Sanger. Mười sáu mẫu hoa hồng đại diện cho các giống hoa hồng cổ đã được thu thập, ly trích DNA tổng số, khuếch đại gen và giải trình tự thành công. Các trình tự vùng *trnH-psbA* của các mẫu hoa hồng thu được trong nghiên cứu cùng với 10 trình tự tham khảo từ Genbank được phân tích mức độ tương đồng và xây dựng cây phát sinh loài. Kết quả cho thấy các trình tự vùng gene *trnH-psbA* của các mẫu hoa hồng được phân tích có mức độ tương đồng cao nhất (từ 99,6% đến 100%) với các loài *Rosa chinensis, Rosa henryi* và *Rosa maximowicziana.* Khoảng cách di truyền giữa các giống hoa hồng cổ miền Bắc Việt Nam ở mức thấp, nằm trong khoảng 0,000-0,145.

Từ khóa: Mã vạch DNA, quan hệ di truyền, hoa hồng, Rosa, trnH-psbA.

MỞ ĐẦU

Hoa hồng là nhóm cây cảnh thuộc chi *Rosa* (họ Rosaceae), một trong những chi cây cảnh lớn nhất và quan trọng nhất trong lịch sử kinh tế và văn hóa của loài người. Chi này bao gồm 150-200 loài với khoảng 18.000 giống hoa hồng thương mại, có nguồn gốc phân bố tự nhiên ở các vùng ôn đới của Bắc bán cầu, hiện được trồng rộng rãi ở khắp các châu lục (Fougère-Danezan *et al.*, 2015). Với nhu cầu ngày càng tăng về các giống hoa hồng mới trên thị trường hoa thế giới, một số lượng lớn các giống hoa hồng đã và đang liên tục được chọn, tạo giống nhằm cải tiến chất lượng của giống như đặc điểm ra hoa, khả năng thích nghi với các môi trường sống mới (Bendahmane *et al.*, 2013). Tại Việt Nam, hoa hồng bản địa và các giống hoa hồng dũ nhập được trồng lâu đời từ thế kỷ 19 và nửa đầu thế kỳ 20 thường được gọi là hồng cổ, được trồng rộng rãi để trang trí trong sân vườn do khả năng thích nghi tốt với khí hậu, chống chịu sâu bệnh và cho hoa đẹp. Đây là nguồn nguyên liệu quý để lai tạo giống và làm gốc ghép cho các giống hồng mới, cần được bảo tồn cũng như phân tích di truyền và các đặc điểm sinh thái, thực vật.

Nhân diên thực vật theo phương pháp truyền thống bằng các đặc điểm hình thái tuy có nhược điểm với một số nhóm thực vật nhưng khá hiệu quả đối với hoa hồng do sự đa dạng và đặc trưng của màu sắc, hình dạng hoa, hình thái lá và thân cây của từng giống. Tuy nhiên, chính sự đa dạng lớn khiến việc xác định mối quan hệ phát sinh bằng hình thái trở nên không chính xác. Với mục đích đánh giá đa dạng và xác định nguồn gốc, quan hệ phát sinh, chỉ thị phân tử được sử dụng rộng rãi do đây là phương pháp khách quan và chính xác. Các chỉ thị với độ đa hình cao như RAPD, AFLP, SSR, ISSR... từ lâu đã được sử dụng trong nghiên cứu đa dạng di truyền hoa hồng. Các chỉ thị này thường có mức độ tái lập thấp hoặc quy trình thực hiện phức tạp. Gần đây, với sự phát triển của kỹ thuật giải trình tự, mã vạch DNA trở thành chỉ thi phổ biến nhất trong nghiên cứu đa dạng và phân loai thực vật do nhiều ưu điểm như đô lặp lại, tính ốn định cao và chi phí ngày càng giảm. Trong các mã vach DNA phổ biến ở thực vật, ITS được sử dụng nhiều nhất trong phân biệt loài, đánh giá đa dạng di truyền trên hoa hồng. Trong đó, vùng ITS2 có nhiều ưu điểm hơn so với vùng ITS1 (EL-Banhawy et al., 2020). Tại Việt Nam, nghiên cứu của Tuong và đồng tác giả (2020) phân tích đa dạng di truyền trên các loài hoa hồng bản địa thông qua vùng trình tự ITS với 23 mẫu thu thập từ 5 tỉnh thành khác nhau. Nghiên cứu này cho thấy sự đa dạng trong chi Rosa và phân bố của chúng tùy thuộc vào vị trí sinh sống của các loài hoa hồng (Tuong et al., 2020). Ngoài ITS, hai mã vạch khác được đề xuất thường xuyên trong các nghiên cứu đa dạng di truyền trên thực vật là rbcL (ribulose-1,5-bisphosphate carboxylase) và matK (maturase K) cũng được sử dụng (EL-Banhawy et al., 2020). Tuy nhiên, hiệu quả của các mã vạch có thể khác nhau giữa các nhóm/loài thực vật với mức độ đa dạng cũng như khả năng nhận diện thực vật khác nhau. Vào năm 2012, Peng và đồng tác giả đã giới thiệu vùng đệm giữa 2 gene trnH và psbA trên hệ gene lục lạp như một mã vạch bổ sung để nhận dạng thực vật (Pang et al., 2012). Ở một số loài, vùng này có độ phân hóa cao do sư biến đổi của các nucleotide, có thể đem lai hiệu quả cao trong các nghiên cứu đánh giá đa dạng di truyền và phân loại (Štorchová, Olson, 2007). Trên thế giới, vùng trnH-psbA

đã được sử dụng trong các nghiên cứu phân tích đa dạng di truyền hoa hồng (EL-Banhawy et al., 2020; Maloupa et al., 2021). Trong nghiên cứu này, mã vạch trnH-psbA được sử dụng để nghiên cứu mối quan hệ di truyền một số giống hoa hồng cổ ở miền Bắc Việt Nam.

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Vật liệu

Mẫu hoa hồng được thu thập từ vườn tại Học viện Nông nghiệp Việt Nam (Hà Nội), tại tỉnh Hà Nam và tỉnh Yên Bái (Việt Nam) trong thời gian từ tháng 10 năm 2023 đến tháng 02 năm 2024. Danh sách 16 mẫu hoa hồng sử dụng trong nghiên cứu được trình bày ở Hình 1.



Hình 1. Các mẫu hoa hồng sử dụng trong nghiên cứu

(Tọa độ bên dưới mỗi mẫu chỉ vị trí thu thập mẫu)

Phương pháp ly trích DNA tổng số

Mẫu lá hoa hồng được rửa sạch, bảo quản ở 4 - 8°C trong 1 tháng và ở -80°C trong thời gian dài. DNA tổng số được tách chiết dựa trên quy trình được công bố bởi Sahu và đồng tác giả (2012) có cái tiến (Sahu et al., 2012). Cụ thể, 200 mg mẫu lá tươi được nghiền trong nitơ lỏng sau đó được rửa 2 lần bằng cách bổ sung 1500 µL đệm rửa (100 mM Tris-HCl, 10 mM EDTA, 0,35 M Glucose, PVP 2%, β-mercaptoethanol 4%), trộn đều và ủ trong đá ở 40 phút sau đó ly tâm ở 12.000 rpm trong 15 phút ở 4°C. Sinh khối lá sau khi rửa được bổ sung 1200 µL đệm ly trích (100 mM Tris-HCl, 1,5 M NaCl, 50 mM EDTA, 3% CTAB, 4% β-mercaptoethanol), ủ ở 60°C trong 40 phút sau đó ly tâm 12.000 rpm trong 15 phút ở 4°C, thu dịch nổi. DNA được chiết 2 lần với cùng thể tích hỗn hợp phenol : chloroform : isoamylalcohol (25:24:1). DNA trong pha nước thu nhận được sau ly tâm được kết tủa bằng ethanol qua đêm ở -20°C. Cặn DNA được thu nhận bằng ly tâm 13.000 rpm trong 15 phút ở 4°C và được rửa 2 lần với ethanol 70%. Sau khi được hòa tan trong 100 µL đệm TE (10 mM Tris-HCl, 1 mM EDTA, NaCl 0,5 M), DNA được chiết 1 lần nữa bằng cùng thể tích hỗn hợp chloroform : isoamylalcohol (24:1), ly tâm ở 13.000 rpm trong 15 phút ở 4°C và thu dịch nổi. DNA được kết tủa lần 2 bằng isopropanol với tỉ lệ 1:1, rửa 2 lần bằng ethanol 70% và làm khô ở nhiệt độ phòng. DNA được hòa tan trong 50 µL đệm TE, định tính bằng phương pháp điện di ngang trên gel agarose 0,8% và bảo quản ở -20°C để sử dụng cho phản ứng PCR.

Khuếch đại và giải trình tự vùng trnH-psbA

Cặp mồi sử dụng để khuếch đại vùng trnH-psbA được thiết kế bằng công cụ PrimerBLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/), dựa trên các vùng bảo thủ của gene trnH và gene psbA của chi Rosa L. tham khảo từ Genbank (NCBI, Mỹ). Mồi xuôi trnH-F có trình tự ACTGCCTTGATCCACTTGGC, mồi ngược psbA-R có trình tự GCGCTAACCTTGGTATGGAAG. Cặp mồi cho sản phẩm PCR dự kiến có kích thước 411 bp, với vị trí bắt cặp của 2 mồi được thể hiện trong Hình 2. Ngoài ra, cặp mồi ITS-F có trình tự GTTTCTTTTCCTCCGCT và ITS-R có trình tự AGGAGAAGTCGTAACAAG (Zuo et al., 2011) cùng với cặp mồi rbcL-F сó TGAAAACGTGAATTCCCAACCGTTTATGCG trình tư và rbcL-R СÓ trình tư GCAGCAGCTAGTTCCGGGCTCCA (Hasebe et al., 1994) được sử dụng để khuếch đại mẫu TXB1.





DNA mục tiêu được khuếch đại với thành phần phản ứng PCR bao gồm: 12 µL 2X My taq Mix (Bioline, Anh), 2 µL mồi xuôi (2 pmol/µL), 2 µL mồi ngược (2 pmol/µL), 2 µL DNA tổng số (100 ng/µL), 6 µL nước khử ion. Thiết bị luân nhiệt PCR Thermal Cycler Block (Thermo Fisher Scientific Oy, Mỹ) được sử dụng để điều nhiệt cho phản ứng theo chu kỳ nhiệt như sau: biến tính ban đầu 95°C trong 5 phút, sau đó là 30 chu kỳ với 30 giây ở 95°C, 30 giây ở 47,7°C, 40 giây ở 72°C, và cuối cùng 72°C trong 1 phút để hoàn thành các phản ứng. Sản phẩm PCR được kiểm tra bằng cách điện di trên gel agarose 1,2%, sau đó được giải trình tự Sanger theo 2 chiều tại Công ty TNHH Dịch vụ và Thương mại Nam Khoa (TP. Hồ Chí Minh, Việt Nam).

Phương pháp phân tích trình tự

Kết quả giải trình tự được kiểm tra tín hiệu peak, hiệu chỉnh và ghép kết quả giải 2 chiều để được trình tự hoàn chỉnh. Các trình tự được so sánh phân tích cùng với 10 trình tự tham khảo từ Genbank. Trong đó, 9 trình tự đại diện các loài hoa hồng phổ biến trong chi *Rosa* là FN687518.1 (*Rosa canina*), GU575153.1 (*Rosa chinensis*), AB043947.1 (*Rosa gallica*), LC769628.1 (*Rosa lucieae*), LC662405.1 (*Rosa multiflora*), AB043945.1 (*Rosa phoenicia*), JX989204.1 (*Rosa roxburghii*), MK862354.1 (*Rosa rugosa*) với AB048211.1 (*Rosa stellata*) và 1 trình tự ngoài nhóm (outgroup) thuộc chi *Rubus*, họ Rosaceae là MZ128738.1 (*Rubus coreanus*). Khoảng cách di truyền giữa các trình tự được đánh giá bằng công cụ p-distance trên phần mềm MEGA-X. Các trình tự được đánh giá mức độ phù hợp của mô hình bằng công cụ Find best DNA/protein Models (ML) từ phần mềm MEGA-X, các thông số tỉ lệ đồng hoán và dị hoán được tính toán dựa trên mô hình phù hợp từ phần mềm trước khi xây dựng cây phát sinh loài. Cây phát sinh loài được xây dựng theo phương pháp Maximum Likelihood bằng phần mềm MEGA-X với giá trị boostrap 500.

KẾT QUẢ VÀ THẢO LUẬN

Kết quả giải trình tự vùng trnH-psbA

DNA tổng số ly trích từ các mẫu hoa hồng cho các băng vạch sáng và rõ nét khi điện di trên gel agarose (Hình 3). Một số mẫu còn nhiễm tạp chất, thể hiện ở vệt sáng tại miệng giếng, một số mẫu khác có hiện tượng đứt gãy nhẹ DNA, tạo ra vệt sáng mờ bên dưới băng chính. Tuy nhiên, với phản ứng PCR truyền thống nhằm khuếch đại DNA cho mục đích giải trình tự với yêu cầu độ nhạy không quá cao, DNA tổng số thu được đủ chất lượng để thực hiện các phản ứng nghiên cứu tiếp theo.



Hình 3. Kết quả điện di DNA tổng số các mẫu hoa hồng

(L: Thang DNA chuẩn, HD1: Hồng điều 1, VK1: Vân khôi 1, TV1: Tường vi 1, BH1: Bạch ho 1, HQ1: Hồng quế 1, QCK1: Quế cánh kép 1, SCK1: Son cánh kép 1, SCD1: Quế son cánh đơn 1, HN1: Hồng nhung 1, HN2: Hồng nhung 2, SP1: Hồng Sapa 1, LTX1: Leo tầm xuân 1, BD: Bạch đào 1, TXB1: Tuần xuân bắc 1, BX1: Bạch xếp 1, HT1: Hồng trắng 1)

DNA tổng số thu nhận được từ các mẫu hoa hồng được sử dụng làm khuôn để khuếch đại vùng *trn*H-*psb*A bằng cặp mồi tự thiết kế trên vùng bảo thủ của các gene *trn*H và *psb*A. Tất cả các mẫu đều cho sản phẩm PCR với băng vạch sáng và rõ, không có sản phẩm khuếch đại không đặc hiệu và primer-dimer (Hình 4). Kích thước sản phẩm PCR của các mẫu đều nằm trong khoảng 400-600 bp khi so với thang DNA chuẩn 1.000 bp (Bioline, Mỹ), phù hợp với kích thước sản phẩm dự kiến khi thiết kế mồi. Như vậy, trình tự vùng *trn*H-*psb*A của các mẫu hoa hồng đều được khuếch đại DNA thành công.

Trình tự vùng *trn*H-*psb*A của các mẫu sau khi cắt xem phần của gene *trn*H và gene *psb*A có độ dài khoảng 270 bp. Kết quả so sánh trình tự với dữ liệu Genbank bằng công cụ NCBI BLAST cho thấy đa số các mẫu tương đồng 100% với các loài *Rosa chinensis* và *Rosa henryi*. BH1, HN1 và HN2 có mức độ tương đồng cao nhất 99,6% với loài *Rosa canina*. Kết quả này khá thống nhất với nghiên cứu phân tích ITS của Tuong và đồng tác giả (2020), cho thấy đa phần các giống hồng Việt Nam thuộc loài *Rosa chinensis*. Riêng mẫu TXB1 (tầm xuân bắc) có độ tương đồng 100% với các loài trong chi *Rubus* của họ Rosaceae. Tầm xuân bắc (*Rosa tunquinensis*) là một loài hồng bản địa của Việt Nam, thường mọc hoang dại ở các vùng núi phía bắc hoặc đôi khi được trồng thành bụi để làm cảnh. Cây có hình thái tương đối khác biệt với các giống hoa hồng cảnh phổ biến khác. Việc trình tự của nhiều gene mã vạch khác nhau của cây tương đồng lớn với một chi khác gợi ý rằng cần xem xét lại vị trí phân loại của cây. Do đó, mẫu này được giải trình tự thêm 2 mã vạch khác là ITS và *rbc*L. So sánh trình tự 2 mã vạch này cho kết quả tương đồng 100% với *Rosa lucieae, Rosa kwangtungenis, Rosa multiflora, Rosa odorata và Rosa soulieana*. Kết quả này khẳng định vị trí phân loại của nó trong chi *Rosa*. Như vậy, vùng *trn*H-

*psb*A có thể không phải là mã vạch đáng tin cậy và không nên sử dụng trong phân loại, định danh các loài của chi *Rosa* nói riêng và họ Rosaceae nói chung.



Hình 4. Sản phẩm PCR khuếch đại vùng *trn*H-*psb*A của các mẫu hoa hồng

(L: Thang DNA chuẩn, VK1: Vân khôi 1, HD1: Hồng điều 1, TV1: Tường vi 1, BH1: Bạch ho 1, HQ1: Hồng quế 1, QCK1: Quế cánh kép 1, SCK1: Son cánh kép 1, SCD1: Quế son cánh đơn 1, SP1: Hồng Sapa 1, HN1: Hồng nhung 1, HN2: Hồng nhung 2, HT1: Hồng trắng 1, LTX1: Leo tầm xuân 1, TXB1: Tuần xuân bắc 1, BX1: Bạch xếp 1, BD: Bạch đào 1)

Phân tích mối quan hệ di truyền của các giống hồng cổ nghiên cứu

Bảng 1. Khoảng cách di truyền giữa các mẫu hoa hồng cổ nghiên cứu

Mẫu	BD1	BH1	HD1	HN1	HN2	HT1	HQ1	QCK1	SCD1	SCK1	SP1	TV1	VK1	BX1	LTX1	TXB1
BD1																
BH1	0,004															
HD1	0,000	0,004														
HN1	0,004	0,000	0,004													
HN2	0,004	0,000	0,004	0,000												
HT1	0,000	0,004	0,000	0,004	0,004											
HQ1	0,000	0,004	0,000	0,004	0,004	0,000										
QCK1	0,000	0,004	0,000	0,004	0,004	0,000	0,000									
SCD1	0,000	0,004	0,000	0,004	0,004	0,000	0,000	0,000								
SCK1	0,000	0,004	0,000	0,004	0,004	0,000	0,000	0,000	0,000							
SP1	0,000	0,004	0,000	0,004	0,004	0,000	0,000	0,000	0,000	0,000						
TV1	0,000	0,004	0,000	0,004	0,004	0,000	0,000	0,000	0,000	0,000	0,000					
VK1	0,000	0,004	0,000	0,004	0,004	0,000	0,000	0,000	0,000	0,000	0,000	0,000				
BX1	0,000	0,004	0,000	0,004	0,004	0,000	0,000	0.000	0,000	0,000	0,000	0,000	0,000			
LTX1	0,009	0,013	0,009	0,013	0,013	0,009	0,009	0,009	0,009	0,009	0,009	0,009	0,009	0,009		
TXB1	0,136	0,145	0,136	0,145	0,145	0,136	0,136	0,136	0,136	0,136	0,136	0,136	0,136	0,136	0,141	

Khoảng cách di truyền giữa 16 mẫu hoa hồng nghiên cứu ở mức thấp, dao động trong khoảng từ 0,000 đến 0,013 giữa 15/16 mẫu bao gồm BD1, BH1, HD1, HN1, HN2, HT1, HQ1, QCK1, SCD1, SCK1, SP1, TV1, BX1 và LTX. Chỉ riêng mẫu TXB có hệ số sai khác với các mẫu còn lại khá lớn, từ 0,136 đến 0,145 (Bảng 1). Các nghiên cứu đa dạng di truyền trên hoa hồng trước đây cũng đã khẳng định, chi *Rosa* có mức độ đa dạng phân tử thấp ở cả hệ gene nhân và hệ gene lục lạp. Điều này cho thấy mặc dù khá đa dạng về hình thái, nhưng hầu hết các giống hoa hồng ngày nay có chung nguồn gốc gần (Anne *et al.*, 2007).

Sự thay đổi nucleotide xảy ra khắp các vị trí codon được thể hiện ở Bảng 2. Về mặt lý thuyết cho biết có 4 loại đồng hoán (A ↔ G và C ↔ T) và 8 loại dị hoán (giữa purine và pyrimidine) (Stoltzfus, Norris, 2016). Tỷ lệ đồng hoán/dị hoán được dự kiến là 0,5. Trong mô hình này, tỷ lệ đồng hoán chỉ chiếm 39,6% thấp hơn đáng kể so với tỷ lệ dị hoán là 60,4%, cho thấy rằng các đột biến dị hoán có thể ảnh hưởng đến cấu trúc của DNA, làm thay đổi tính chất và chức năng của gene do sự thay thế giữa purine và pyrimidine. Vì vậy, thông thường các đột biến đồng hoán được phổ biến hơn do tính ổn định trong các vùng bảo tồn qua các thế hệ. Qua đó, mô hình cũng cung cấp thông tin quan trọng cho việc phân tích phân loại trong việc xây dựng cây phát sinh loài.

Bảng 2. Mô hình thay thế nucleotide giữa các trình tự vùng t*rn*H-psbA của 16 mẫu hoa hồng (%)

Α	т	С	G
-	11,63	3,47	4,54
11,63	-	4,54	3,47
11,63	15,26	-	3,47
15,26	11,63	3,47	-
	A - 11,63 11,63 15,26	A T - 11,63 11,63 - 11,63 15,26 15,26 11,63	A T C - 11,63 3,47 11,63 - 4,54 11,63 15,26 - 15,26 11,63 3,47

(Mô hình và tỉ lệ thay thế được ước tính theo mô hình Tamura 3-parameter với tỷ lệ thay thế chuyển tiếp khác nhau và tỷ lệ thay thế chuyển đổi được thể hiện bằng chữ in đậm)

Trên cây phát sinh loài, các mẫu hoa hồng nghiên cứu phân thành 4 nhóm (Hình 5). Nhóm A bao gồm 11 trình tự tương đồng hoàn toàn với nhau (TV1, VK1, SP1, SCK1, SCD1, QCK1, HQ1, HD1, BD1, HT1 và BX1) và với loài

Rosa chinensis (GU575153.1). Trong nhóm B, ba mẫu HN1, HN2 và BH1 cho thấy mức độ tương đồng cao nhất với loài Rosa canina (FN687518.1). LTX1 nằm ở nhóm C cùng với hai loài Rosa multiflora (LC662405.1) và Rosa lucieae (LC769628.1). Riêng TXB1, với trình tự khác biệt các mẫu hoa hồng còn lại và nằm ở nhóm D, cùng với loài Rubus coreanus (MZ128738.1).



Hình 5. Cây phát sinh loài của các giống hoa hồng cổ nghiên cứu được xây dựng dựa trên các trình tự vùng *trn*H-psbA bằng phương pháp Maximum Likelihood

KÉT LUẬN

Kết quả giải trình tự các vùng *trn*H-*psb*A cho thấy 16 mẫu hoa hồng cổ phân thành 4 nhóm trên cây phát sinh loài. Các giống hồng cổ trồng lâu đời ở miền Bắc Việt Nam có quan hệ di truyền gần nhất với loài *Rosa chinensis* (GU575153.1) và *Rosa canina* (FN687518.1). Các giống hoa hồng cổ này có độ tương đồng di truyền cao và có khoảng cách di truyền khá lớn với loài hoang dại bản địa *Rosa tunquinensis*. Tuy nhiên, kết quả cũng cho thấy *trn*H-*psb*A không phải là mã vạch DNA thích hợp để sử dụng trong định danh, phân loại hoa hồng.

Lời cảm ơn: Nghiên cứu này được cấp kinh phí bởi Trường Đại học Công Thương Thành phố Hồ Chí Minh theo hợp đồng số 156/HĐ-DCT ngày 31 tháng 8 năm 2024.

TÀI LIỆU THAM KHẢO

Anne B, Julian RS, Simon J (2007). Phylogenetic Relationships in the Genus *Rosa*: New Evidence from Chloroplast DNA Sequences and an Appraisal of Current Knowledge. *Systematic Botany*, 32(2): 366-378.

Bendahmane M, Dubois A, Raymond O, & Bris M Le. (2013). Genetics and genomics of flower initiation and development in roses. *J Exper Bot*, 64(4): 847-857.

EL-Banhawy A, Acedo C, Qari S, Elkordy A (2020). Molecular Identification and Phylogenetic Placement of *Rosa arabica* Crép. (Rosaceae), a Critically Endangered Plant Species. *Life*, 10(12): 335.

Fougère-Danezan M, Joly S, Bruneau A, Gao XF, Zhang LB. (2015). Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann Bot*, 115(2): 275-291.

Hasebe M, Omori T, Nakazawa M, Sano T, Kato M, Iwatsuki K. (1994). *rbcL* gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. *PNAS*, 91(12): 5730-5734.

Maloupa E, Karapatzak E, Ganopoulos I, Karydas A, Papanastasi K, Kyrkas D, Yfanti P, Nikisianis N, Zahariadis A, Kosma I S, Badeka A V, Patakioutas G, Fotakis D, Krigas N. (2021). Molecular Authentication, Phytochemical Evaluation and Asexual Propagation of Wild-Growing *Rosa canina* L. (Rosaceae) Genotypes of Northern Greece for Sustainable Exploitation. *Plants*, 10(12): 2634.

Pang X, Liu C, Shi L, Liu R, Liang D, Li H, Cherny SS, Chen S. (2012). Utility of the *trn*H-*psb*A Intergenic Spacer Region and ITS Combinations as Plant DNA Barcodes: A Meta-Analysis. *PLoS ONE*, 7(11): e48833.

Sahu S K, Thangaraj M, & Kathiresan K. (2012). DNA Extraction Protocol for Plants with High Levels of Secondary Metabolites and Polysaccharides without Using Liquid Nitrogen and Phenol. *ISRN Mol Biol*, 2012, 1-6.

Stoltzfus A, Norris R W. (2016). On the Causes of Evolutionary Transition: Transversion Bias. Mol Biol Evol, 33(3): 595-602.

Štorchová H, Olson MS (2007). The architecture of the chloroplast *psb*A-*trn*H non-coding region in angiosperms. *Plant Sys Evol*, 268(1-4): 235-256.

Tuong LQ, Tam TV, Dong DV, Duong TD, Cuc DTK, Thu PTL, Luong DT, Tuyen VTM, Giang N, Tuan NT, Trung NT, Khanh TD, Trung KH (2020). Identification of Vietnamese native rose species by using internal transcribed spacers (ITS) sequencing. *Plant Cell Biotechnol Mol Biol*, 21(11 & 12): 1-10.

Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou . (2011). DNA Barcoding of Panax Species. Plant Med, 77(02): 182–187.

STUDY ON GENETIC RELATIONSHIP OF SOME OLD-FASHIONED ROSES (*Rosa* spp.) FROM NORTHERN VIETNAM BASED ON *trn*H-*psb*A SEQUENCE

Ngo Van Duc, Hoang Hai Dang, Trieu Phuong Mai, Pham Quoc Toan, Nguyen Thi Bich Huong, Le Thi Kieu Trinh, Nguyen Minh Phuong*

Ho Chi Minh City University of Industry and Trade

Summary

This study aims to determine and analyze the *trn*H-*psb*A region sequence to assess the genetic relationship of old-fashioned roses cultivated in Northern Vietnam. Primers for DNA amplification were designed based on referenced *Rosa* sequences obtained from the NCBI Genbank database. Total DNA was extracted from leaf samples using the CTAB method, and the *trn*H-*psb*A region was subsequently amplified through polymerase chain reaction and sequenced using the Sanger sequencing technique. Sixteen representative samples of old-fashioned roses were collected, and their DNA was successfully extracted, amplified, and sequenced. The sequences of the rose samples, and 10 reference sequences from Genbank were analyzed for similarity and used to construct a phylogenetic tree. The results showed that the examined rose samples exhibited the highest similarity to *Rosa chinensis*, *Rosa henryi*, and *Rosa maximowicziana* species. The genetic distance among the rose samples in Northern Vietnam was low, ranging from 0.000 to 0.145.

Keywords: DNA barcode, genetic relationship, Rosa, rose, trnH-psbA.

^{*} Author for correspondence: Tel: 0963372277; Email: phuongnguyen@huit.edu.vn

BƯỚC ĐẦU ĐÁNH GIÁ HIỆU QUẢ CỦA MÃ VẠCH *trn*H-*psb*A TRONG VIỆC XÁC ĐỊNH ĐA DẠNG DI TRUYỀN CỦA MỘT SỐ MẫU LAN KIM TUYẾN THU THẬP TẠI VIỆT NAM

Hoàng Hải Đăng^{1*}, Ngô Văn Đức¹, Trịnh Thị Hương¹, Hồ Viết Thế¹, Trần Trọng Tuấn²

¹Đại học Công Thương Thành phố Hồ Chí Minh

²Viện Sinh học Nhiệt đới, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Ở Việt Nam, lan kim tuyến được sử dụng làm dược liệu quan trọng để điều trị nhiều bệnh. Tuy nhiên, việc khai thác quá mức những cây lan này đã dẫn đến sự cạn kiệt dần tài nguyên thiên nhiên của chúng. Do đó, việc áp dụng các chỉ thị phân tử để nhận diện và phân biệt các chủng lan kim tuyến nhằm mục đích bảo tồn các loài lan này là vô cùng cần thiết. Việc sử dụng mã vạch DNA đã nổi lên như một phương pháp có độ tin cậy cao để phân loại thực vật. Trong nghiên cứu này, chúng tôi đã đánh giá hiệu quả của vùng mã vạch DNA *trnH-psbA* trong việc phân biệt chín mẫu lan kim tuyến được thu thập từ tỉnh Quảng Bình và Thành phố Hồ Chí Minh. Những phát hiện này cho thấy sự thay đổi đáng kể trong trình tự DNA của vùng *trnH-psbA* trong số các mẫu lan được thu thập. Những kết quả này nêu bật tiềm năng của mã vạch DNA *trnH-psbA* trong đặc tính di truyền và nhận diện tài nguyên của lan kim tuyến ở Việt Nam.

Từ khóa: Mã vạch DNA, lan kim tuyến, đa dạng di truyền, trnH-psbA.

MỞ ĐẦU

Lan kim tuyến là tên gọi được sử dung để miêu tả những loài hoặc nhóm cây có lá mềm mượt, như được làm từ gấm, với những đường gân đẹp mắt của một số chi trong họ Orchidaceae, cụ thể là Anoectochilus, Dossinia, Goodyera, Ludisia (De et al., 2018), phân bố rộng rãi ở các vùng nhiệt đới châu Á, bao gồm Trung Quốc, Ấn Độ, Lào, Indonesia và Việt Nam. Lan kim tuyến là một loại cây thảo được được sử dụng phổ biến làm thực phẩm bổ sung và y học dân gian ở châu Á (Kuan *et al.*, 2011). Với tác dụng chăm sóc sức khỏe, nó có thể được sử dụng để bảo vệ gan (Fang *et al*., 2008), giảm mệt mỏi (Ikeuchi *et al*., 2005), chống oxy hóa (Wang *et al*., 2005), hạ đường huyết (Ho *et al*., 2018), giảm huyết áp (Fang *et al*., 2008), ngăn ngừa loãng xương (Masuda *et al*., 2008), chống khối u (Tseng et al., 2006) và cải thiện chức năng miễn dịch (Kuan et al., 2011). Tại Việt Nam, lan kim tuyến đã được sử dụng rộng rãi trong y học cổ truyền để điều trị nhiều bệnh khác nhau như viêm màng phổi, bệnh tim mạch, tăng huyết áp, sốt, viêm thận, bầm tím và rắn độc cắn (Shao *et al.*, 2014). Ngoài ra, hoa của lan kim tuyến sở hữu một số đặc tính dược lý, bao gồm khả năng chống mệt mỏi, chống oxy hóa, giảm mỡ máu, phòng ngừa ung thư và tác dụng điều hòa miễn dịch. Theo truyền thống, việc phân loại lan kim tuyến thường dựa vào các đặc điểm hình thái như hình dạng lá, màu sắc và cấu trúc gân. Mặc dù phương pháp này mang lại lợi ích về tiết kiệm thời gian và chi phí, nhưng cũng gặp phải một số vấn đề. Ví dụ, các đặc điểm hình thái bên ngoài thường rất giống nhau, sự đa dạng trong hình thái giữa giai đoạn trưởng thành và giai đoạn non, cũng như ảnh hưởng của môi trường và giai đoạn phát triển của cây thường dẫn đến sự không chính xác trong việc nhận dạng. Ngoài ra, việc nhân dang dựa trên hình thái thường không chính xác nếu mẫu vật bị hỏng hoặc đã được xử lý trước đó. Sử dụng không đúng loại cây hoặc các hợp chất dược liêu khác nhau để chế biến thuốc thảo dược có thế làm giảm hiệu quả của thuốc và có thế gây hai cho người sử dụng (Ho et al., 2021).

Trong những năm gần đây, sự phát triển của các chỉ thị phân tử (RAPD, AFLP, SSR, ISSR, mã vạch DNA,...) mà các nghiên cứu về định danh loài, đánh giá phân loại và đa dạng di truyền được phân tích một cách khách quan, tăng độ chính xác và được sử dụng phổ biến. Hai dấu hiệu mã vạch DNA thường được đề xuất cho thực vật là ribulose-1,5-bisphosphate carboxylase (*rbcL*) và maturase K (*mat*K). Tuy nhiên, khả năng phân biệt của các dấu hiệu này có thể khác nhau giữa các loài thực vật (Ahmed *et al.*, 2020). Vào năm 2012, Pang và đồng tác giả đã giới thiệu vùng *trn*H-*psb*A như một mã vạch bổ sung để nhận dạng thực vật (Pang *et al.*, 2012). Mã vạch DNA đặc biệt này rất khác nhau giữa các vùng plastid và có thể để dàng khuếch đại thông qua phản ứng PCR ở nhiều loại thực vật trên đất liền (Ralitsa *et al.*, 2020). Vùng *trn*H-*psb*A đem lại hiệu quả cao trong việc xác định các loài thực vật khác nhau, bao gồm lúa mạch đen (Skuza *et al.*, 2019), hạt hạnh nhân và hạt mơ (Uncu *et al.*, 2020) và thực vật họ Poaceae, Panicoideae, Paspaleae (Delfini *et al.*, 2023). Dựa trên thành công này, nghiên cứu hiện tại được tiến hành nhằm đánh giá hiệu quả của mã vạch *trn*H-*psb*A trong việc phân biệt các loại lan kim tuyến khác nhau. Những phát hiện của nghiên cứu này có khả năng góp phần vào việc sử dụng các dấu hiệu mã vạch này để xác định chính xác các loài lan kim tuyến cụ thể. Điều này có thể hỗ trợ các nỗ lực liên quan đến nhân giống, bảo tồn và nghiên cứu về tính đa dạng của các loài cây thuốc có giá trị.

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Vật liệu

Chín mẫu lan kim tuyến được thu thập từ núi Trường Sơn thuộc tỉnh Quảng Bình (QB) và Thành phố Hồ Chí Minh (TP. HCM) được sử dụng trong nghiên cứu này (Bảng 1). Hình thái lá của chín mẫu này được thể hiện ở (Hình 1).

Bảng	1. Da	anh	sách	chín	mẫu	lan	kim	tuvến	được	Sử	duna	trona	nahiên	cứu
				•••••							~~			

STT	Ký hiệu mẫu	Nơi thu thập	Đặc điểm
1	HCM-1	Trung tâm thí nghiệm thực hành Trường Đại học Công Thương TP. HCM	Mẫu nuôi cấy mô
2	HCM-2	Khu Nông nghiệp Công nghệ cao TP. HCM	Mẫu nuôi cấy mô
3	HCM-4	Khu Nông nghiệp Công nghệ cao TP. HCM	Mẫu nuôi cấy mô
4	HCM-5	Viện Sinh học Nhiệt đới	Mẫu nuôi cấy mô
5	HCM-6	Viện Sinh học Nhiệt đới	Mẫu nuôi cấy mô
6	QB-4	Núi Trường Sơn thuộc tỉnh Quảng Bình	Mẫu cây tươi và lá
7	QB-5	Núi Trường Sơn thuộc tỉnh Quảng Bình	Mẫu cây tươi và lá
8	QB-6	Núi Trường Sơn thuộc tỉnh Quảng Bình	Mẫu cây tươi và lá
9	QB-8	Núi Trường Sơn thuộc tỉnh Quảng Bình	Mẫu cây tươi và lá

Trình tự mồi được sử dụng trong nghiên cứu này được trình bày ở (Bảng 2).

Bảng 2.	Trình t	ư mồi	khuếch	đai vùng	trnH-psbA

Tên mồi	Độ dài mồi	Kích thước sản phẩm PCR (bp)	Tài liệu tham khảo
<i>trn</i> H- <i>psb</i> A-F	CGCGCATGGTGGATTCACAATCC	910 950 bp	(Costion of al. 2011)
<i>tm</i> H- <i>psb</i> A-R	GTTATGCATGAACGTAATGCTC	910-930 bp	(Costion et al., 2011)

Phương pháp nghiên cứu

Phương pháp ly trích DNA tổng số

Trong nghiên cứu này, mẫu lan kim tuyến thu thập được rửa sạch, bảo quản ở -80°C. Quy trình trích ly DNA tổng số lan kim tuyến được tham khảo dựa trên nghiên cứu của Asif và đồng tác giả (2005): nghiền 300 mg mẫu lá tươi trong nitơ lỏng, bổ sung 800 μL đệm ly trích CTAB 3% (100 mM Tris-HCl, 3 M NaCl, 20 mM EDTA, 3% CTAB, 1% PVP) và 60 μL SDS 10%, ủ trong bể ổn nhiệt (Digital Water Bath, Hàn Quốc) ở 65°C trong 60 phút sau đó ly tâm 13.000 rpm trong 15 phút ở 4°C. Dịch nổi thu nhận được bổ sung cùng thể tích hỗn hợp Phenol:Chloroform:Isoamyl alcohol (25:24:1), ly tâm 13.000 rpm trong 15 phút ở 4°C và thu dịch nổi. Mẫu DNA được bổ sung cùng thể tích hỗn hợp Chloroform:Isoamyl alcohol (24:1), ly tâm ở 13.000 rpm trong 15 phút ở 4°C và thu dịch nổi. Dịch nổi được kết tủa bằng Isopropanol với tỉ lệ 1:1, đảo nhẹ và ly tâm 13.000 trong 15 phút ở 4°C và thu tủa. Rửa tủa 2 lần bằng EtOH 76%, ly tâm ở 13.000 rpm trong 15 phút và để khô ở nhiệt độ phòng. Sau đó được tái huyền phù trong 50 μL nước khử ion (Canada). Mẫu được đánh giá chất lượng tinh sạch bằng phương pháp đo quang phổ (UV-Vis 6600) ở bước sóng 260 nm với 280 nm và xác định bằng phương pháp điện dì ngang trên gel agarose 1,5% trong 100 ml TAE 1X (Thermo Fisher Scientific, Mỹ). Dung dịch DNA được bảo quản ở -20°C để sử dụng cho phản ứng PCR.

PCR khuếch đại vùng trình tự trnH-psbA

Các DNA mục tiêu được khuếch đại với thành phần phản ứng PCR bao gồm: 12 µL 2X Mytaq Mix (Bioline, Anh), 2 µL mồi (2 pmol/µL), 2 µL DNA template (50 ng/µL) và 6 µL nước khử ion (Canada). Bên cạnh đó, các mẫu DNA được khuếch đại bằng thiết bị luân nhiệt PCR Thermal Cycler Block (Thermo Fisher Scientific Oy, Mỹ) với các điều kiện cho phản ứng PCR như sau: biến tính ban đầu 95°C trong 2 phút, sau đó 35 chu kỳ với 30 giây 95°C, 30 giây 53,8°C, 1 phút ở 72°C, và cuối cùng 72°C trong 1 phút để hoàn thành các phản ứng. Sau đó, sản phẩm PCR được tinh sạch bằng ISOLATE II PCR và Gel Kit (Bioline, UK). Sản phẩm PCR được xác định bằng phương pháp điện di ngang trên gel agarose 1,2 % trong 100 ml TAE 1X (Thermo Fisher Scientific Oy, Mỹ). Sản phẩm sau đó được gửi giải trình tự tại Công ty TNHH Dịch Vụ Và Thương Mại Nam Khoa (Thành phố Hồ Chí Minh, Việt Nam).

Phương pháp xử lý số liệu

Sau khi sản phẩm giải trình tự được kiểm tra bằng tín hiệu peak thông qua phần mềm FinchTV, các trình tự sẽ được đưa lên cơ sở dữ liệu Genbank của NCBI (National Center for Biotechnology Information). Các trình tự sẽ được so sánh và phân tích bằng công cụ BLAST của NCBI. Mức độ khác biệt giữa các trình tự sẽ được đánh giá bằng công cụ p-distance trên phần mềm MEGA-X. Để đánh giá mức độ phù hợp của mô hình, các trình tự sẽ được phân tích bằng công cụ "Find best DNA/protein Models (ML)" trong phần mềm MEGA-X. Các thông số tỉ lệ

đồng vị và hoán vị sẽ được tính toán dựa trên mô hình phù hợp từ phần mềm trước khi xây dựng cây phát sinh loài. Phân tích tỷ lệ thay thế (%) của nucleotide trong vùng mã vạch *trn*H - *psb*A của chín mẫu lan kim tuyến theo mô hình Tamura 3-parameter. Cây phát sinh loài sẽ được xây dựng bằng phương pháp UPGMA. Trong phân tích phát sinh loài, hai trình tự từ hai chi khác nhau trong họ Orchidaceae, MN331616.1 thuộc loài *Cattleya forbesii* và KX298661.1 thuộc loài *Vanda pumila*, sẽ được sử dụng như nhóm ngoài (outgroup). Cách sử dụng nhóm ngoài trong việc xây dựng lại cây phát sinh loài sẽ được mô tả và đánh giá theo các tiêu chí phân tích. Bằng việc xem xét các trạng thái, đặc điểm và mối quan hệ của các nhóm ngoài, chúng ta có thể ước lượng các trạng thái tổ tiên của nhóm nghiên cứu với nhóm ngoài, người ta có thể xác định các đặc điểm chung và phân biệt giữa các đặc điểm dẫn đến sự hiểu biết sâu hơn về mối quan hệ phát sinh và tiến hóa giữa các thành viên trong nhóm nghiên cứu (Maddison *et al.*, 1984).

KẾT QUẢ VÀ THẢO LUẬN

Kết quả thu thập mẫu lan kim tuyến

Tổng cộng có chín mẫu lan kim tuyến đã được thu thập, trong đó có bốn mẫu được thu thập từ tỉnh Quảng Bình và năm mẫu được thu thập từ Thành phố Hồ Chí Minh. Các mẫu cây này đã thể hiện sự đa dạng đáng kể về hình dạng lá, như được mô tả trong hình 1. Tuy nhiên, trong mỗi nhóm mẫu không có kiểu gân đặc biệt nào được xác định. Những phát hiện này càng làm nổi bật thêm tính không đáng tin cậy của việc nhận dạng hình thái trong việc phân biệt các loài lan kim tuyến. Các trình tự sau khi loại bỏ các đoạn tín hiệu nhiễu và đã được đăng lên cơ sở dữ liệu Genbank NCBI với các mã số được thể hiện trong hình 1.



Hình 1. Hình thái lá của chín mẫu lan kim tuyến được sử dụng trong nghiên cứu này

Kết quả ly trích DNA

Khi xem xét kết quả ly trích DNA tổng số được thể hiện trong hình 2, chúng ta có thể thấy xuất hiện các băng vạch DNA và RNA sáng và rõ nét. Trong đó, các băng vạch DNA sắc nét có thể nhìn thấy kèm theo hai băng vạch rRNA 23S và 16S. Nguyên nhân có thể là do quá trình tách chiết DNA không đảm bảo loại bỏ hoàn toàn enzyme RNase, gây tồn đọng RNA trong mẫu. Bên cạnh đó, trên miệng giếng của gel, chúng ta cũng có thể thấy xuất hiện các dải sáng khác, có thể là do sự hiện diện của protein. Nhìn chung, kết quả này cho thấy một số hiện tượng không mong muốn trong quá trình ly trích DNA tổng số. Việc đánh giá và cải thiện phương pháp tách chiết là cần thiết để đảm bảo kết quả chính xác và đáng tin cậy trong nghiên cứu.

M	нсмі	HCM2	HCM4	HCM5	HCM6	QB4	QB5	QB6	QB8	
			1					1		Genomic DNA
			1	5 (1 (B))	=		100			238 rRNA 168 rRNA

Hình 2. Kết quả điện di DNA tổng số chín mẫu lan kim tuyến

Kết quả PCR khuếch đại vùng trình tự trnH-psbA

Sau khi thu thập DNA tổng số từ các mẫu lan kim tuyến, DNA này đã được khuếch đại bằng cặp mồi được mô tả trong Bảng 2, sử dụng nhiệt độ bắt cặp mồi phù hợp cho phản ứng PCR. Kết quả quan sát cho thấy rằng các mẫu lan kim tuyến đã thể hiện các dải DNA mục tiêu rõ ràng, có kích thước tương đương với dự kiến là 910-950bp, so với thang chuẩn HyperLadder™ 1kb được sử dụng trong nghiên cứu, do hãng Bioline thuộc Meridian cung cấp. Điều này cho thấy rằng quá trình PCR đã thành công trong việc khuếch đại và tạo ra sản phẩm mong muốn. Đáng chú ý, không có bất kỳ sản phẩm phụ nào được quan sát, bao gồm cả primer dimer, các sản phẩm không mong muốn được tạo ra khi cặp mồi tự bắt cặp với nhau trong quá trình PCR.



Hình 3. Sản phẩm PCR khuếch đại trình tự trnH-psbA của chín mẫu lan kim tuyến

Kết quả đánh giá đa dạng di truyền lan kim tuyến

Sau khi được kiểm tra và điều chỉnh bằng phần mềm FinchTV, kết quả giải trình tự DNA của vùng *trn*H - *psb*A đã được sử dụng để xác định thông qua công cụ NCBI BLAST. Kết quả cho thấy rằng đối với nhóm mẫu thu thập ở Thành phố Hồ Chí Minh bao gồm HCM-1, HCM-4, HCM-5, chúng tương đồng 100% với các loài *Anoectochilus roxburghii, Anoectochilus zhejiangensis, Anoectochilus chapaensis và Anoectochilus emeiensis.* Mẫu HCM-2 tương đồng 100% với loài *Anoectochilus formosanus* và mẫu HCM-6 tương đồng cao nhất là 99,39% với loài *Anoectochilus burmannicus.* Đối với nhóm mẫu thu thập ở tỉnh Quảng Bình bao gồm QB-1, QB-5, QB-6, hệ số tương đồng dao động từ 99,57-100% với loài *Ludisia discolor.* Mẫu QB-8 có hệ số tương đồng 99,6% với loài *Cheirostylis chinensis.* Phân tích trình tự cho thấy sự khác biệt giữa các vùng *trn*H - *psb*A nằm trong khoảng từ 0,000 đến 0,061 (Xem Bảng 3).

	HCM-1	HCM-2	HCM-4	HCM-5	HCM-6	QB-4	QB-5	QB-6	QB-8
HCM-1	0,000								
HCM-2	0,005	0,000							
HCM-4	0,000	0,005	0,000						
HCM-5	0,000	0,005	0,000	0,000					
HCM-6	0,002	0,008	0,002	0,002	0,000				
QB-4	0,009	0,011	0,011	0,009	0,011	0,000			
QB-5	0,009	0,010	0,010	0,009	0,010	0,000	0,000		
QB-6	0,013	0,015	0,016	0,013	0,016	0,004	0,004	0,000	
QB-8	0,039	0,041	0,044	0,039	0,036	0,057	0,047	0,061	0,000

Bảng 3. Khoảng cách di truyền của chín mẫu lan kim tuyến trong nghiên cứu

Kết quả được trình bày trong nghiên cứu này cho thấy sự thay thế của các nucleotide khác nhau trong vùng mã vạch DNA, như được mô tả trong bảng 4. Đặc biệt, các sự thay thế dị hoán giữa các phân tử purine và pyrimidine đã được quan sát, nhằm hiểu rõ hơn về mô hình tiến hóa của các loại cây thảo dược này. Về mặt lý thuyết, theo Tamura và đồng tác giả (1992, 1993) có 4 loại đồng hoán (A ↔ G và C ↔ T) và 8 loại dị hoán (giữa purine và pyrimidine). Tỷ lệ dự kiến của đồng hoán và dị hoán là 0,5. Tuy nhiên, trong nghiên cứu này, tỷ lệ đồng hoán chỉ đạt 41,47%, mức thấp hơn đáng kể so với tỷ lệ dị hoán là 58,53%. Thông thường đột biến đồng hoán được xác định là hình thức phổ biến nhất gây ra sự biến đổi và đa dạng hóa giữa các cá thể và nó đóng vai trò quan trọng trong quá trình tiến hóa của các loài. Mô hình này cũng cung cấp thông tin quan trọng để phân tích và phân loại trong việc xây dựng cây phát sinh loài.

Bảng 4. Phân tích tỷ lệ thay thế (%) của nucleotide trong vùng mã vạch <i>trn</i> H- <i>psb</i> A của chín mẫu lan kim tuyến							
Nitrogenous bases	Α	т	C	G			
Α	-	9,85	4.78	6,77			
т	9,85	-	6,77	4,78			
С	9,85	13,96	-	4,78			
G	13,96	9,85	4.78	-			

Mô hình và tỉ lệ thay thế được ước tính theo mô hình Tamura 3-parameter với tỷ lệ thay thế chuyển tiếp khác nhau và tỷ lệ thay thế chuyển đổi được thể hiện bằng chữ in đậm

Áp dụng phương pháp UPGMA, các mẫu lấy từ Thành phố Hồ Chí Minh (Tp. HCM) và tỉnh Quảng Bình (QB) được phân thành ba nhóm phân biệt rõ ràng, thể hiện giá trị bootstrap cao (Hình 4). Nhóm A bao gồm 5 trình tự (HCM-1, HCM-2, HCM-4, HCM-5, HCM-6) tương đồng đồng cao với các loài Anoectochilus roxburghii (NC 061758.1), Anoectochilus zhejiangensis (NC 054353.1), Anoectochilus chapaensis (MW589500.1), Anoectochilus formosanus (NC 061756.1)) và Anoectochilus emeiensis (NC 033895.1). Trong nhóm B, ba mẫu QB-4, QB-5, QB-6 cho thấy mức độ tương đồng cao với loài Ludisia discolor (OP719317.1). Trong nhóm C, có một mẫu QB-8 cho thấy mức đô tương đồng cao với loài Cheirostylis chinensis (MN641483.1). Ngoài ra, nhóm outgroup (MN331616.1 và KX298661.1) từ các loài khác nhau đã hình thành nên một nhánh riệng biệt, càng khẳng định độ tin cậy của phân tích phát sinh gene. Điều này phù hợp với nghiên cứu trước đây được thực hiện trên họ Combretaceae với chi Ficus, cũng chỉ ra rằng dấu hiệu mã vạch trnH-psbA vươt trội hơn các dấu hiệu mã vạch rbcL và matK (Li et al., 2012; Gere et al., 2013). Ngoài ra, kết quả của các nghiên cứu trước đây cho thấy rằng trên các loài thực vật khác nhau, khả năng nhận diện trên vùng trình tự trnH-psbA tương đối cao như lúa mạch đen (Skuza et al., 2019), hạt hạnh nhân và hạt mơ (Uncu et al., 2020) và thực vật họ Poaceae, Panicoideae, Paspaleae (Delfini et al., 2023). Mã vạch trnH-psbA trước đây cũng được coi là một locus có tiềm năng cao vì nó chứa trình tự ngắn tạo điều kiện thuận lợi cho việc khuếch đại và trình tự đơn giản hơn. Kết quả chính từ nghiên cứu này có thể là tiềm năng cho các nghiên cứu sâu hơn để điều tra nhiều hơn về lợi ích của trnH-psbA trong việc phân biệt các loài lan kim tuyến ở Việt Nam.



Hình 4. Cây phát sinh loài của chín mẫu lan kim tuyến sử dụng trình tự trnH-psbA

KÉT LUẬN

Nhóm nghiên cứu đã thu thập chín mẫu lan kim tuyến ở Thành phố Hồ Chí Minh và tỉnh Quảng Bình. Khuếch đại thành công và giải mã được trình tự vùng *trn*H-*psb*A của chín mẫu lan kim tuyến và phân tích trình tự thu được cho thấy khoảng cách di truyền giữa chín mẫu lan kim tuyến thu thập được ở mức thấp từ 0,000 đến 0,061. Kết quả chính từ nghiên cứu này có thể là tiềm năng cho nghiên cứu sâu hơn để điều tra lợi ích của *trn*H-*psb*A trong việc phân biệt và phân loại khác nhau.

Lời cảm ơn: Nhóm nghiên cứu trân trọng cảm ơn Trường Đại học Công Thương Thành phố Hồ Chí Minh và Viện Sinh học Nhiệt đới đã giúp đỡ trong quá trình thực hiện dự án.

TÀI LIỆU THAM KHẢO

Ahmed EB, Carmen A, Sameer Q, Ahmed E (2020). Molecular identification and phylogenetic placement of *Rosa arabica* Crép (Rosaceae), a critically endangered plant species. *Life*, 10(12), 335.

Asif MJ, Cannon CH (2005). DNA extraction from processed wood: A case study for the identification of an endangered timber species (*Gonystylus bancanus*). *Plant Mol Biol Rep*, 23(2), 185-192.

Costion C, Ford A, Cross H, Crayn D, Harrington M, Lowe A (2011). Plant DNA Barcodes Can Accurately Estimate Species Richness in Poorly Known Floras. *PLoS ONE*, 6(11), e26841.

De LC, Pathak P (2018). Conservation, management and utilization of orchid genetic resources. TOSI, 32, 81-91.

Delfini C, Acosta JM, Aliscioni S., Souza VC, Zuloaga FO (2023). Phylogenetic Relationships in the Group Caespitosa of *Paspalum* L (Poaceae, Panicoideae, Paspaleae). *Diversity 2023*, 15, 134.

Fang HL, Wu, JB, Lin, WL, Ho, HY, Lin, WC (2008). Further studies on the hepatoprotective effects of Anoectochilus formosanus. Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives, 22(3), 291-296.

Gere J, Yessoufou K, Daru BH, Mankga LT, Maurin O, van der Bank M (2013). Incorporating *trn*H-*psb*A to the core DNA barcodes improves significantly species discrimination within southern African Combretaceae. *ZooKeys*, (365), 129.

Ho VT, Tran TKP, Vu TTT, Widiarsih S (2021). Comparison of *mat*K and *rbcL* DNA barcodes for genetic classification of jewel orchid accessions in Vietnam. *JGEB*, 19(1), 93.

Ho Y, Chen YF, Wang LH, Hsu KY, Chin YT, Yang YCS, Lin SJ (2018). Inhibitory effect of *Anoectochilus formosanus* extract on hyperglycemia-related PD-L1 expression and cancer proliferation. *Front Pharmacol*, 9, 807.

Ikeuchi M, Yamaguchi K, Nishimura T, Yazawa K (2005). Effects of Anoectochilus formosanus on endurance capacity in mice. J Nutr Sci Vitaminol, 51(1), 40-44.

Kuan YC, Wu TJ, Kuo CY, Hsu JC, Chang WY, Sheu F (2011). Molecular cloning of a new immunomodulatory protein from *Anoectochilus formosanus* which induces B cell IgM secretion through a T-independent mechanism. *PloS one*, *6*(6), e21004.

Li HQ, Chen JY, Wang S, Xiong SZ (2012). Evaluation of six candidate DNA barcoding loci in *Ficus* (Moraceae) of China. *Mol Ecol Resour*, 12(5), 783-790.

Maddison WP, Donoghue MJ, Maddison, DR (1984). Outgroup analysis and parsimony. Sys Biol, 33(1), 83-103.

Masuda K, Ikeuchi M, Koyama T, Yamaguchi K, Woo JT, Nishimura T, Yazawa K (2008). Suppressive effects of *Anoectochilus* formosanus extract on osteoclast formation *in vitro* and bone resorption *in vivo*. *JBMM*, *26*, 123-129.

Pang X, Liu C, Shi L, Liu R, Liang D, Li H, ...Chen S (2012). Utility of the *trn*H–*psb*A intergenic spacer region and its combinations as plant DNA barcodes: a meta-analysis. *PloS One*, *7*(11), e48833.

Ralitsa B, Katerina S, Radostina SG, Maya I (2020). A preliminary assessment of *trn*H-*psb*A as DNA barcode for botanical identification of polyfloral honey samples and comparison with *rbc*L marker. *Bulg J Agric*, 26(1), 238–242.

Skuza L, Szućko I, Filip E, Adamczyk A (2019). DNA barcoding in selected species and subspecies of Rye (Secale) using three chloroplast loci (*mat*K, *rbcL*, *trn*H-*psbA*). *Not Bot Horti Agrobo*, 47(1), 54-62.

Shao Q, Deng Y, Liu H, Zhang A, Huang Y, Xu G, Li M (2014). Essential oils extraction from *Anoectochilus roxburghii* using supercritical carbon dioxide and their antioxidant activity. *Ind Crop Prod*, 60, 104-112.

Tamura K (1992). Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+ C-content biases. *Mol Biol Evol*, 9(4), 678-687.

Tamura K, Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *MBE*, 10(3), 512-526.

Tseng CC, Shang HF, Wang LF, Su B, Hsu CC, Kao HY, Cheng KT (2006). Antitumor and immunostimulating effects of *Anoectochilus formosanus Hayata. Phytomed*, *13*(5), 366-370.

Uncu AO (2020). A *trn*H-*psb*A barcode genotyping assay for the detection of common apricot (*Prunus armeniaca* L.) adulteration in almond (*Prunus dulcis* Mill.). *CyTA* - *J* Food, 18(1), 187–194.

Wang LF, Lin CM, Shih CM, Chen HJ, Su B, Tseng CC, ... Cheng KT (2005). Prevention of cellular oxidative damage by an aqueous extract of *Anoectochilus formosanus*. Ann N Y Acad Sci, 1042(1), 379-386.

INITIALLY ASSESS THE EFFECTIVENESS OF *trn*H-*psb*A BARCODE IN IDENTIFYING GENETIC DIVERSITY OF NINE JEWEL ORCHID SPECIES COLLECTED VIETNAM

Hai Dang Hoang^{1*}, Van Duc Ngo¹, Thi Huong Trinh¹, Viet The Ho¹, Trong Tuan Tran²,

¹Ho Chi Minh City University of Industry and Trade

²Institute of Tropical Biology, Vietnam Academy of Science and Technology

SUMMARY

Jewel orchids are used as important medicinal herbs to treat many diseases in Vietnam. However, overexploitation of these orchids has led to the gradual depletion of their natural resources. Therefore, it is extremely necessary to apply molecular markers to identify and differentiate the jewel orchid strains to preserve these orchid species. The use of DNA barcoding has emerged as a highly reliable method for classifying plants. In this study, we evaluated the effectiveness of the *trn*H-*psb*A DNA barcoding region in distinguishing nine jewel orchid accessions collected from Quang Binh province and Ho Chi Minh City. These findings revealed significant variations in the DNA sequence of the *trn*H-*psb*A region among the orchid samples collected. These results highlight the potential of *trn*H-*psb*A DNA barcoding in genetic characterization and resource identification of jewel orchids in Vietnam.

Keywords: DNA barcode, jewel orchids, phylogenetic analysis, trnH-psbA.

^{*} Author for correspondence: Tel: 0967921920; Email: Hoanghaidang09011996@gmail.com
ĐỊNH DANH CHỦNG NẤM HƯƠNG THU NHẬN TẠI VƯỜN QUỐC GIA BẠCH MÃ BẰNG SINH HỌC PHÂN TỬ VÀ KHẢO SÁT MÔI TRƯỜNG NHÂN GIỐNG NHẰM BẢO TÒN NGUÔN GEN

Lê Thị Như Ngọc¹, Nguyễn Vũ Linh¹, Trần Thiện Ân¹, Võ Đình Ba², Nguyễn Việt Thắng², Nguyễn Minh Trí^{2*}

¹Vườn Quốc gia Bạch Mã

²Trường Đại học Khoa học, Đại học Huế

TÓM TẮT

Mẫu nấm Hương (Ký hiệu BM01) được thu hái tại vùng rừng núi của Vườn Quốc gia Bạch Mã, huyện Phú Lộc, tỉnh Thừa Thiên Huế vào tháng 01/2024 khi đang phát triển trên các thân cây mục. Các đặc điểm hình thái bên ngoài của quả thể như: màu nâu sậm khi tiếp xúc với nước, chuyển sang nâu nhạt khi khô, đường kính mũ nấm trung bình từ 2,5-6,5 cm, mặt trên của mũ nấm có các vảy sợi thô màu trắng, lồi nhẹ khi còn non, phẳng dần khi trưởng thành, mép mũ mỏng dễ rách. Bào tử có màu vàng nhạt, kích thước: 3-5 μ m, có hình dạng từ bầu dục đến hình elip, vách mỏng nhẵn, với một vài hạt nhỏ bên trong. Các dẫn liệu phân tích đoạn gen ITS của mẫu nấm Hương trên cho thấy đây là loài *Lentinula edodes*. Giống thuần của chủng nấm này đã được phân lập từ mô thịt quả thể, phát triển tốt trên môi trường nhân giống cấp 1 là PGA (Potato Glucose Agar) và môi trường cấp 2 là thóc nấu chín có bổ sung 1% CaCO₃.

Từ khóa: Bạch Mã, nấm Hương, giải trình tự, ITS (Internal transcribed spacer), Lentinula edodes, nhân giống.

MỞ ĐẦU

Nấm Hương có thành phần dinh dưỡng cao và hương vị đặc biệt nên được xem là một loại nấm thực phẩm cao cấp (Hui *et al.*, 2004). Ngoài ra, nấm Hương còn có những giá trị về mặt y dược như: hoạt tính kháng sinh, hoạt tính chống ung thư, chống virus, tăng cường miễn dịch, khả năng loại trừ cholesterol... Do đó, tại các nước như Nhật Bản, Trung Quốc, Hàn Quốc... nấm Hương đã được nuôi trồng trên quy mô lớn và cho hiệu quả kinh tế cao (Nguyễn Lân Dũng, 2005).

Hiện nay, công nghệ sản xuất nấm Hương ở các nước như Việt Nam, Nhật Bản, Hàn Quốc... đang phụ thuộc nhiều vào điều kiện tự nhiên (Chan, 2005). Vì thế, việc tìm kiếm, sưu tập các nguồn gene bản địa và kiểm định danh pháp của chúng ở mức độ phân tử là rất quan trọng, đặt nền tảng cho công tác tạo giống và xây dựng công nghệ nuôi trồng nấm Hương.

Vào tháng 01/2024, chúng tôi đã phát hiện và thu hái một số quả thể nấm phát triển trên các thân cây lá rộng đang mục tại khu vực rừng của Vườn Quốc gia Bạch Mã, huyện Phú Lộc, tỉnh Thừa Thiên Huế. Dựa vào các đặc điểm hình thái bên ngoài cho thấy đây là mẫu nấm thuộc một loài trong chi Lentinula (Earle), chúng tôi nhận định sơ bộ đây là một chủng nấm hoang dại quý, cần phải định danh và nhân giống để tạo ra chủng nấm để nuôi trồng phù hợp với điều kiện tự nhiên của địa phương.

NGUYÊN LIỆU VÀ PHƯƠNG PHÁP

Nguyên liệu

Mẫu nấm ký hiệu BM01 được thu hái vào tháng 1/2024 trên thân cây gỗ lớn đang mục tại khu vực Vườn Quốc gia Bạch Mã ở vị trí độ cao 1370 m so với mặt nước biển có tọa độ X: 00592127; Y: 01791358 (Tọa độ GPS: VN 2000, KT 107, múi 3°).

Môi trường phân lập: PGA (Potato Glucose Agar): 200 g khoai tây; 20 g glucose; 20 g agar; nước cất đủ 1000 mL.

Môi trường nhân giống cấp 1:

- PGA: 200 g khoai tây; 20 g glucose; 20 g agar; nước cất đủ 1000 mL.
- PGA nước dừa: 200 g khoai tây; 20 g glucose; 20 g agar; nước dừa đủ 1000 mL.
- Raper: 2 g pepton; 2 g dịch chiết nấm men; 0,5 g MgSO₄.7H₂0; 0,46 g KH₂PO₄; 1 g K₂HPO₄, 20 g glucose, 20 g agar; nước cất đủ 1000 mL.

Môi trường nhân giống cấp 2: thóc (lúa) là nguyên liệu chính để nhân giống nấm cấp 2; sau khi rửa nhiều lần bằng nước sạch, thóc được nấu chín cho đến khi tách vỏ và để ráo nước, phối trộn theo các công thức sau:

- Công thức 1 (CT1): 99% thóc luộc + 1% CaCO3
- Công thức 2 (CT2): 69% thóc luộc + 30% mùn cưa + 1% CaCO3
- Công thức 3 (CT3): 39% thóc luộc + 60% mùn cưa + 1% CaCO3

Phân phối các công thức môi trường nhân giống cấp 2 vào các chai thủy tinh và khử trùng ở 121°C trong thời gian 30 phút.

Phương pháp

Phân lập và thuần khiết mẫu

Mẫu được đưa về phòng thí nghiệm và khử trùng bằng cồn 70° để phân lập bằng phương pháp cấy mô. Trong điều kiện vô trùng, mẫu nấm được cắt dọc từ cuống đến phần mũ và tách một mảnh mô có kích thước 0,5x0,5x0,1cm tại vị trí giữa mũ và cuống nấm rồi cho vào các đĩa petri có chứa môi trường PGA đã được khử trùng để tạo nguồn giống thuần dùng cho tách chiết DNA và các nghiên cứu về đặc điểm sinh trưởng của loài nấm này (Nguyễn Lân Dũng, 2005).

Phương pháp định danh khoa học

Phương pháp so sánh hình thái: Mẫu quả thể nấm thu từ tự nhiên được phân tích các đặc điểm hình thái theo phương pháp của Trịnh Tam Kiệt (2012). Cấu trúc hiển vi của sợi nấm, bào tử... được quan sát và chụp hình dưới kính hiển vi quang học Olympus BX51.

Phương pháp sinh học phân tử: Hệ sợi nấm được sử dụng cho tách chiết DNA theo phương pháp của Gardes và Bruns (1993): 50mg sợi nấm được nghiền trong 5 phút, thêm 500µl dung dịch trích ly, voxtex và để ở nhiệt độ phòng 10 phút, sau đó ly tâm rồi thu lấy dung dịch phía trên. DNA được tủa bằng cồn 96° và rửa 2 lần bằng cồn 70°. DNA sau đó được sấy chân không 10 phút ở 45°C rồi hòa tan trong 100 µL TE 0.1X. Cuối cùng DNA được kiểm tra chất lượng thông qua quá trình đo quang phổ và điện di trên gel agarose 0,8%. Những sản phẩm đạt yêu cầu được bảo quản ở -20°C cho những bước tiếp theo.

Phản ứng PCR được thực hiện với cặp mồi ITS1 và ITS4 (ITS1: 5'-TCCGTAGGTGAACCTGCGG-3'; ITS4: 5'-TCCTCC GCTTATTG ATATGC-3') theo phương pháp của White và đồng tác giả (1990). Sản phẩm PCR sẽ được phân tích trên gel agarose và sau đó giải trình tự. Trình tự này sau đó sẽ được dùng để dò tìm trình tự tương đồng trên ngân hàng dữ liệu gene bằng công cụ BLAST (Basic Local Alignment search tool) (Altschul *et al.*, 1990).

Các trình tự được sắp gióng bằng phần mềm ClustalW. Cây phả hệ thể hiện mối quan hệ di truyền giữa mẫu nghiên cứu và các loài thuộc chi Lentinula hiện có trên dữ liệu GenBank được xây dựng bằng phần mềm MEGA 11 theo phương pháp Maximum Parsimony với hệ số tin tưởng (bootstrap) 100%.

Các phân tích được thực hiện trong ba lần lặp lại, kết quả là giá trị trung bình ± độ lệch chuẩn. Tất cả các số liệu được xử lý bằng chương trình MS. Excel 2016 (Đặng Văn Giáp, 2000)

KẾT QUẢ VÀ THẢO LUẬN

Đặc điểm hình thái

Mẫu nấm Hương phân bố tại Vườn Quốc gia Bạch Mã mọc riêng lẻ trên các thân gỗ mục của nhóm cây lá rộng ở khu vực có độ cao 1300-1400 m so với mặt nước biển, nhiệt độ: 23° C và độ ẩm 86%.

Quả thể có dạng hình cái dù bao gồm mũ nấm và cuống nấm (Hình1). Mũ nấm có dạng tròn, khum, màu nâu sậm khi tiếp xúc với nước, chuyển sang nâu nhạt khi khô, đường kính: 2,5-6,5cm, mặt trên của mũ nấm có các vảy sợi thô màu trắng, lồi nhẹ khi còn non, phẳng dần khi trưởng thành, mép mũ mỏng dễ rách (Hình 2A). Phần thịt mũ nấm mỏng: 0,3-0,5cm, dai, cấu tạo bởi các lớp sợi kết hợp khá lỏng lẻo, thường phân nhánh và có khóa ở hầu hết các vách ngăn (Hình 2C), quả thể sau khi sấy khô có mùi thơm đặc trưng.

Bào tầng dạng phiến, đính rời, màu trắng, nằm ở mặt dưới của mũ nấm (Hình 2B). Cuống nấm có dạng hình trụ, đặc chắc, màu trắng ở phần trên gần mũ nấm, màu nâu ở phía dưới gần giá thể; dài: 3-5cm; đường kính: 0,5-1,5cm; cuống nấm đính vào mũ nấm ở giữa hoặc hơi lệch tâm; bề mặt cuống nấm có vảy thô, chất thịt màu trắng, mềm (Hình 2A, B).



Hình 1. Mẫu nấm Hương tại Vườn Quốc gia Bạch Mã



Hình 2. Giải phẫu quả thể nấm Hương Bạch Mã

A: Quả thể (mặt trên và mặt dưới); B: Bào tầng dạng phiến; C: Hệ sợi trong thịt mũ nấm; D: Bào tử đảm.

Bào tử có màu màu vàng nhạt, kích thước: 3-5µm, có hình dạng từ bầu dục đến hình elip, trong suốt không chứa tinh bột, vách mỏng nhẵn, với một vài hạt nhỏ bên trong (hình 2D).

Về hình thái và cấu tạo giải phẫu của mẫu nấm Hương thu tại Bạch Mã có các đặc điểm tương đồng với mẫu nấm Hương *Lentinula edodes* thu nhận Sa Pa của Lê Xuân Thám và đồng tác giả (2010) và của Lê Huyền Ái Thúy và đồng tác giả (2012).

Dựa vào hình dạng của bào tử quan sát được và các đặc điểm hình thái bên ngoài của nấm so với những mô tả của Trịnh Tam Kiệt (2012) cho thấy đây là mẫu nấm thuộc một loài trong chi Lentinula (Earle). Tuy nhiên cần định danh chính xác, chúng tôi tiến hành giải trình tự gen bằng phương pháp PCR để định danh khoa học.

Kết quả phân tích vùng gene ITS

Các ứng dụng sinh học phân tử để xác định nguồn gốc phát sinh của các loài nấm đã được tiến hành từ những năm 1990. Kết quả xác định trình tự vùng gene ITS sau khi loại bỏ trình tự mồi và các vùng tín hiệu nhiễu, chúng tôi đã thu được trình tự nucleotide như sau:

Hình 3. Kết quả giải trình tự vùng gen ITS

Bảng 1. Mức độ tương đồng của trình tự loài nấm thu thập với loài Lentinula edodes trên cơ sở dữ liệu NCBI

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Lentinula edodes strain HA-1 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene	1312	1312	98%	0.0	99.45%	MK940842.1
Lentinula edodes voucher 520521MF228-1 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene	1312	1312	98%	0.0	99.45%	MZ172422.1

Kết quả giải trình tự vùng gen ITS được so sánh với cơ sở dữ liệu trên Ngân hàng gen NCBI. Đoạn gen rRNA 729 bps vùng ITS của nấm có tỷ lệ tương đồng là 99,45% so với loài *Lentinula edodes* trên cơ sở dữ liệu NCBI (Acession number: MK940842.1).

Cây phát sinh chủng loại được xây dựng dựa trên trình tự ITS của chủng nghiên cứu và các chủng có quan hệ họ hàng gần thuộc loài *Lentinula edodes* (Berk) Pegler được thể hiện qua hình 4.



Hình 4. Cây phả hệ thể hiện mối quan hệ di truyền của loài Lentinula edodes (BM01)

(Ghi chú: Chủng nghiên cứu được đánh dấu bằng mũi tên)

Theo kết quả phân tích phả hệ dựa theo trình tự ITS1 của Lê Huyền Ái Thúy và đồng tác giả (2012) về mẫu nấm Hương thu tại Sa Pa thì chủng nấm Hương tại Bạch Mã mà chúng tôi phân tích vùng ITS1 có độ tương đồng với nhau và cùng tương đồng với loài Lentinula edodes STCL 125 với giá trị bootstrap lặp lại 1000 lần. Từ so sánh kết quả mô tả hình thái theo Trịnh Tam Kiệt (2012) và kết hợp phân tích rRNA vùng ITS, chúng tôi nhận thấy mẫu nấm Hương thu được ở Vườn Quốc gia Bạch Mã là loài *Lentinula edodes* (Berk) Pegler.

Khảo sát môi trường nhân giống

Kết quả phân lập trên môi trường PGA cho thấy sợi nấm sinh trưởng khá mạnh, hệ sợi đồng đều, phát triển dày đặc. Khi nuôi sợi nấm trên ba loại môi trường được khảo sát để nhân giống cấp 1 gồm PGA, PGA - nước dừa và Raper cho thấy giữa hai môi trường PGA và PGA - nước dừa không có sự khác biệt về mặt thống kê về quá trình phát triển của sợi nấm (bảng 2). Kết quả này cũng phù hợp với nghiên cứu của Phạm Nữ Kim Hoàng và đồng tác giả (2014) về khi tăng lượng peptone trong môi trường nuôi đã làm giảm tốc độ phát triển của sợi nấm Hương hoang dã Len 003 được thu tại rừng cây lá rộng núi Langbiang, Lâm Đồng.

STT	Công thức thí nghiệm	Độ lan sợi nấm sau 3 ngày (mm)	Độ lan sợi nấm sau 5 ngày (mm)
1	PGA	4,1 ^a	6,2 ^a
2	PGA - nước dừa	3,8ª	5,9 ^a
3	Raper	2,5 [°]	3,9°

Bảng 2. Kết quả khảo sát môi trường nhân giối	ng cấ	ip '	1
---	-------	------	---



Hình 5. Sợi nấm phát triển trên các loại môi trường dinh dưỡng sau 5 ngày

Kết quả khảo sát này cho thấy khi cấy chuyền nấm Hương thì môi trường PGA là phù hợp nhất vì môi trường này có đầy đủ chất dinh dưỡng cần thiết cho sự phát triển của sợi nấm. Đồng thời cũng dễ dàng chuẩn bị môi trường nuôi cấy, chi phí thấp hơn so với môi trường PGA có bổ sung nước dừa và Raper. Vì vậy chúng tôi sử dụng môi trường PGA để nhân giống nấm Hương cấp 1.

Trong công nghệ nuôi trồng nấm, giống cấp 2 được sử dụng để nhân nhanh số lượng giống cũng như giúp cho giống thích nghi dần với môi trường nuôi trồng (Đinh Xuân Linh và đồng tác giả, 2012). Hiện nay tại các cơ sở nghiên cứu và sản xuất giống nấm ở nước ta đang sử dụng các hạt ngũ cốc như thóc, lúa mì, gạo lức, ngô... có bổ sung CaCO₃ 1% để nhân giống cấp 2.

Tiến hành đánh giá sự sinh trưởng và phát triển của hệ sợi nấm trên các môi trường cấp 2 được phối trộn, kết quả nghiên cứu ở bảng 3 cho thấy: Thời gian sợi nấm mọc lan đều 100% trong chai nuôi ở CT1 là 5,8 ngày; cao hơn so với các công thức sử dụng thóc có bổ sung mùn cưa với các tỷ lệ khác nhau, theo chúng tôi trong thóc nấu chín có chứa các thành phần dinh dưỡng thích hợp cho sợi nấm Hương sử dụng để phát triển.

STT	Thành phần	Số ngày sợi nấm mọc lan 50%	Số ngày sợi nấm mọc lan 100%	Đặc điểm hệ sợi nấm					
CT1	99% thóc luộc + 1% CaCO $_3$	3,6ª	5,8 ^ª	Hệ sợi nấm dày, trắng, dạng bông xốp, đồng đều, mọc nhanh					
CT2	69% thóc luộc + 30% mùn cưa + 1% CaCO $_3$	4,3 ^b	6,4 ^b	Hệ sợi nấm nhiều, màu trắng, dạng bông xốp, mọc bình thường					
CT3	39% thóc luộc + 60% mùn cưa + 1% CaCO $_3$	2,6 ^c	3,5 ^c	Hệ sợi nấm ít, màu trắng ngà, dạng bông xốp, mọc chậm					

Dana J. Net uta kilao sat ilioi titrona ilian utona cap \mathbf{z}
--

KÉT LUẬN

Với các đặc điểm hình thái, giải phẫu và phân tích sinh học phân tử cho thấy mẫu nấm thu tại Vườn quốc gia Bạch Mã chính là chủng nấm Hương *(Lentinula edodes)*, thuộc chi *Lentinula* (Earle), họ Marasmiaceae, bộ Agaricales. Môi trường nhân giống cấp 1 tốt nhất là PGA (Potato Glucose Agar) cho hệ sợi nấm sinh trưởng và phát triển mạnh. Môi trường thóc luộc có bổ sung 1% CaCO₃ là môi trường nhân giống cấp 2 tối ưu từ 5-6 ngày cho hệ sợi nấm mọc nhanh, dày và đồng đều.

TÀI LIỆU THAM KHẢO

Hui FL, Wei MH, & Liu ZH (2004). Assay study on amino acid, trace elements and toxic heavy metals of the fruit bodies of *Lentinus edodes. Shipin Kexue* (Beijing, China), 25, 161-163.

Nguyễn Lân Dũng (2005). Công nghệ nuôi trồng nấm. NXB Nông Nghiệp Hà Nội.

Chan AW (2005). Shiitake bag cultivation. In Mushroom growers handbook 2. Seoul, Korea: MushWorld.

Trịnh Tam Kiệt (2012). Nấm lớn Việt Nam. Tập 1. NXB Khoa học và Kỹ thuật, Hà Nội.

Gardes M and TD Bruns (1993). ITS primers with enhanced specificity for basidiomycetes: Application to the identifi cation of mycorrhizae and rusts. *Mol Ecol* 2: 113-118.

White TJ, Bruns TD, Lee SB, and Taylor JW (1990). Amplilcation and direct sequencing of fungal ribosomal RNA Genes for phylogenetics. In: PCR Protocols: A Guide to Methods and Applications. Academic Press. US. 482pp.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol, 215*(3), 403-410 Đặng Văn Giáp (2000). *Phân tích dữ liệu khoa học bằng Microsoft Excel*. Hà Nội: NXB Giáo dục Hà Nội.

Lê Xuân Thám, Nguyễn Như Chương, Phạm Ngọc Dương, Bùi Hoàng Thiêm (2011). The biogeographical speciations of shiitake *Lentinula edodes* and a new species *Lentinula platinedodes* sp.nov. found in Cat Tien, South Vietnam. *Acad J Biol* 33 (3):29-39.

Lê Huyền Ái Thúy, Lao Đức Thuận, Nguyễn Hoàng Mai, Phan Hoàng Đại, Nguyễn Trương Kiến Khương, Trương Bình Nguyên (2021). Bổ sung dẫn liệu phân tử và khảo sát đặc điểm nuôi trồng của chủng nấm Hương Sapa *Lentinula edodes*. *HCMCOUJS-Kỹ thuật và Công nghệ*, 16(1), 102-111

Phạm Nữ Kim Hoàng, Trương Bình Nguyên, Lê Xuân Thám, Phan Hữu Hùng, Hoàng Ngọc Ánh, Đỗ Thị Thiên Lý (2014). Nghiên cứu một số đặc điểm sinh học của chủng nấm Hương lai từ núi Langbiang, Lâm Đồng và chủng thương mại Nhật Bản. *Tạp chí Sinh học*, 36(1se): 158-164

Đinh Xuân Linh Thân Đức Nhã, Nguyễn Hữu Đống, Nguyễn Thị Sơn (2012). Kỹ thuật trồng, chế biến nấm ăn và nấm dược liệu. NXB Nông Nghiệp Hà Nội.

IDENTIFICATION OF MUSHROOM SPECIES RECEIVED IN BACH MA NATIONAL PARK USING MOLECULAR BIOLOGY AND SURVEYING THE BREAKING ENVIRONMENT TO CONSERVE GENETIC RESOURCES

Le Thi Nhu Ngoc¹, Nguyen Vu Linh¹, Tran Thien An¹, Vo Đinh Ba², Nguyen Viet Thang², Nguyen Minh Tri^{2*}

¹Bach Ma National Park

²Hue University of Science, Hue University

SUMMARY

The sample (Symbol BM01) was collected in the mountainous areas of Bach Ma National Park, Phu Loc district, Thua Thien Hue province in January 2024 while growing on rotten tree trunks. The external morphological characteristics of the fruit body are dark brown when exposed to water, turning light brown when dry, the average diameter of the mushroom cap is from 2.5-6.5 cm, the upper surface of the mushroom cap has scales. The coarse fibers are white, slightly convex when young, gradually flatten when mature, the thin edge of the cap easily tears. Spores are light yellow in color, size: $3-5\mu$ m, shaped from oval to elliptical, with thin, smooth walls, with a few small seeds inside. Analysis of the ITS gene segment of the above Shiitake mushroom sample shows that this is *Lentinula edodes* species. Pure varieties of this mushroom strain have been isolated from fruiting body flesh tissue, growing well on the level 1 propagation medium, PGA (Potato Glucose Agar), and the level 2 medium, cooked rice supplemented with 1% CaCO₃.

Keywords: Bachma, mushroom, sequencing, ITS (Internal transcribed spacer), Lentinula edodes, breeding.

Author for correspondence: Tel: 0914031085; Email: nguyenminhtri@husc.edu.vn

BIÊU HIỆN PROTEIN TOLL-LIKE RECEPTOR 22 TỪ CÁ TRA Pangasianodon hypophthalmus VÀ ĐÁNH GIÁ KHẢ NĂNG TƯƠNG TÁC VỚI VI KHUẨN

Nguyễn Thanh Tấn^{1,3}, Trần Văn Hiếu^{1,2,3*}

¹Phòng thí nghiệm Cảm biến Sinh học, Khoa Sinh học – Công nghệ Sinh học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

²Bộ môn Công nghệ Sinh học Phân tử và Môi trường, Khoa Sinh học – Công nghệ Sinh học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh

³Đại học Quốc gia Thành phố Hồ Chí Minh, Việt Nam

TÓM TẮT

Cá tra (Pangasianodon hypophthalmus) là loài cá da tron được nuôi phổ biến nhất ở Đồng bằng sông Cửu Long, nổi bật với giá trị dinh dưỡng cao và nhu cầu tiêu thụ ngày càng tăng cả trong và ngoài nước. Để nâng cao sức cạnh tranh, cần tập trung phát triển, nâng cao chất lượng và sản lượng cá tra. Trong những năm gần đây, việc sản xuất và tiêu thụ cá tra gặp nhiều khó khăn như do giá cả không ổn định, dịch bệnh ở cá, trong đó bệnh gan thận mủ và bệnh xuất huyết có tỉ lệ chết cao (50-90%). Việc sử dụng kháng sinh để điều trị dịch bệnh ở cá dẫn đến dư lượng kháng sinh trong thịt cá và hiện tượng kháng thuốc. Hơn nữa, vaccine hiện tại còn gặp nhiều hạn chế về phân phối và bảo quản. Do đó, chúng tôi hướng tới việc phát triển chế phẩm sinh học mới nhằm phòng bệnh gan thận mủ và bệnh xuất huyết cho cá tra, cũng như các bệnh do vi khuẩn ở cá nói chung. Thụ thể Toll-like receptor (TLR) 22 chỉ được tìm thấy ở cá, có khả năng nhân diện, liên kết với dsRNA từ virus và lipopolysaccharide nhờ vào các trình tự lặp lại giàu leucine (LRR). Vì vậy, vùng LRR của TLR22 có thể được ứng dụng để bắt vi khuẩn. Trong nghiên cứu này, thụ thể TLR22 được tạo ra dưới dạng protein tái tổ hợp và tiến hành đánh giá tương tác với vi khuẩn. Plasmid pET22b-tlr22 được cấu trúc trong E. coli MC1061, và thụ thể TLR22 được biểu hiện ở E. coli SHuffle® T7 Express và xác nhận bằng Western blot với kháng thể kháng His-tag (6xHis). Khả năng bắt vi khuẩn được đánh giá bằng kỹ thuật Dot blot, với kết quả cho thấy thụ thể TLR22 có khả năng bắt vi khuẩn Gram dương. Nghiên cứu này là tiền đề cho việc phát triển chế phẩm sinh học bổ sung thức ăn cho cá, điều tri bênh ở cá tra và các loài cá khác.

Từ khóa: Gram dương, LRR, Pangasianodon hypophthalmus, Toll-like receptor 22, Streptococcus agalactiae.

MỞ ĐẦU

Cá tra (*Pangasianodon hypophthalmus*) là loài cá da tron được nuôi phổ biến nhất Đồng bằng sông Cửu Long. Thêm vào đó, cá tra là thực phẩm mang giá trị dinh dưỡng cao, được tiêu thụ ngày càng nhiều ở trong và ngoài nước. Ngoài vai trò cung cấp nguyên liệu trong lĩnh vực thực phẩm, da cá tra còn có thể dùng để tách chiết collagen, ứng dụng trong mỹ phẩm, thực phẩm chức năng. Ở ĐBSCL, các tỉnh/thành có diện tích nuôi và sản lượng cá tra lón gồm Đồng Tháp; An Giang; Cần Thơ, và được xuất khẩu sang hơn 180 quốc gia và vùng lãnh thổ trên thế giới (Thi, 2017). Theo báo cáo VASEP (Hiệp hội chế biến và xuất khẩu thuỷ sản Việt Nam) đến tháng 10 năm 2023 doanh thu xuất khẩu cá tra đạt 1,5 tỷ USD giảm 29% so với cùng kì năm 2022. Hiện nay, cá tra đang là mặt hàng được quan tâm và nước ta là một trong những quốc gia có lợi thế về ngành nuôi trồng, xuất khẩu cá tra. Ngoài Việt Nam, một số quốc gia khác như Indonesia, Trung Quốc cũng đang tập trung phát triển ngành này. Để nâng cao sức cạnh tranh, chúng ta phải không ngừng phát triển, nâng cao chất lượng và sản lượng. Tuy nhiên, trong những năm gần đây, việc sản xuất và tiêu thụ cá tra đang phải đối mặt với nhiều khó khăn và thách thức do giá cả bấp bênh, thị trường xuất khẩu không ổn định và dịch bệnh trên cá ngày càng tăng. Để đạt những tiêu chí trên thì chủ đầu tư cần phải giảm chi phí đầu tư bằng cách tăng hiệu suất sử dụng thức ăn và kiểm soát tốt dịch bệnh ở cá.

Hiện nay, thách thức lớn nhất của chúng ta đối với ngành hàng này là các vấn đề về dịch bệnh, cá dễ bị tấn công bởi nhiều loại vi sinh vật gây bệnh. Trong đó, vi khuẩn Gram âm là nguyên nhân chính gây bệnh gan thận mủ và bệnh xuất huyết là một trong số những bệnh có tỉ lệ chết cao (50-90%) (Thi, 2017). Hơn thế, cá tra còn bị nhiễm bệnh bởi các loài vi khuẩn Gram dương, đặc biệt là *Streptococcus agalactiae*, là một trong những vi khuẩn Gram dương gây bệnh ở cá, gây xuất huyết và tỷ lệ tử vong rất cao (45%) (Trân *et al.*, 2016; Meidong *et al.*, 2021). Bên cạnh gây bệnh trên cá, nhiều nghiên cứu đã cho thấy *S. agalactiae* có liên quan đến các trường hợp bùng phát bệnh Streptococcosis ở người sau khi ăn cá, cho thấy khả năng của nó như một tác nhân lây truyền từ động vật sang người (Kalimuddin *et al.*, 2017). Tính đến thời điểm hiện nay, sử dụng kháng sinh là một phương pháp điều trị hữu hiệu khi cá mắc bệnh này. Tuy nhiên, dư lượng kháng sinh trong thịt cá, và hiện tượng kháng thuốc đang là những vấn đề quan trọng cần phải giải quyết của phương pháp này. Tuy nhiên, tiêm hay ngâm cá giống trong vaccine gặp nhiều vấn đề về phân phối thuốc, cá giống có hệ miễn dịch chưa hoàn chỉnh nên hiệu quả vaccine

kém; vaccine uống tạo miễn dịch dịch thể không mạnh mẽ, gặp khó khăn trong việc bảo quản, vận chuyển (Wise *et al.*, 2015). Do đó, ngoài việc phòng ngừa bằng vaccine thì đề tài mong muốn hướng đến tìm ra một biện pháp an toàn hơn để phòng bệnh trên cá tra.

Ở cá, hệ miễn dịch bẩm sinh là tuyến phòng thủ ban đầu khi có sự xâm nhiễm. Trong đó, Toll-like Receptor (TLR) là một họ thụ thể có khả năng nhận diện các kiểu mẫu phân tử của vi sinh vật (PAMPs). TLR22 là thành viên của họ TLR, và chỉ được tìm thấy ở cá. Thông qua các trình tự lặp lại giàu leucine (Leucine-rich Repeat – LRR), TLR22 có khả năng nhận diện, và liên kết với dsRNA ở virus, và lipopolysaccharide (LPS). Do đó, vùng LRR này có thể được sử dụng để bắt những vi khuẩn gây bệnh trên cá. Với bản chất là protein, vùng LRR có thể được tạo ra dưới dạng protein tái tổ hợp, và tiến hành đánh giá tương tác với vi khuẩn. Kết quả này sẽ là tiền đề cho việc phát triển sản phẩm sinh học hỗ trợ điều trị bệnh ở cá do vi khuẩn gây ra sau này. Kết quả của nghiên này sẽ được ứng dụng làm tiền đề cho các nghiên cứu sâu hơn để có thể ứng dụng trong việc phát triển vaccine uống trong việc phòng ngừa và điều trị bệnh ở cá tra và các loài cá khác.

NGUYÊN VẬT LIỆU VÀ PHƯƠNG PHÁP

Chủng vi khuẩn và plasmid

Chủng *E. coli* MC1061 được sử dụng làm chủng chủ để nhân bản plasmid tái tổ hợp. Chủng *E. coli* SHuffle[®] T7 Express sử dụng làm chủng chủ biểu hiện protein tái tổ hợp. Plasmid pET22b được sử dụng làm plasmid dòng hóa, đồng thời là plasmid biểu hiện protein tái tổ hợp nhờ vào promoter T7 có trên plasmid giúp kiểm soát sự biểu hiện gen thông qua chất cảm ứng IPTG (isopropyl-ß-D-1-thiogalactopyranoside) (Biobasic). Môi trường nuôi cấy sử dụng cho cả hai chủng đều là Luria –Bertani (LB) chứa 100 µg/mL Ampicillin (Sigma Aldrich) ở 37°C. Các chủng vi sinh vật, plasmid, và hoá chất được cung cấp bởi Phòng thí nghiệm Cảm biến Sinh học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia Thành phố Hồ Chí Minh; Viện nuôi trồng Thuỷ sản 2; và Đại học Tài nguyên và Khoa học sự sống/BOKU, Viên, Cộng hoà Áo.

Cấu trúc plasmid tái tổ hợp pET22b-tlr22

Biểu hiện protein tái tổ hợp TLR22

Sau khi đã cấu trúc thành công plasmid tái tổ hợp pET22b-tlr22, tiến hành thu nhận plasmid bằng phương pháp Alkaline Lysis (Ehrt et al., 2003). Plasmid sau khi thu nhân được biến nap vào chủng biểu hiện E. coli SHuffle® T7 Express khả nạp bằng phương pháp sốc nhiệt và sản phẩm biến nạp được trải trên đĩa LB chứa kháng sinh Ampicillin đạt nồng cuỗi 100 µg/mL. Các dòng E. coli SHuffle® T7 Express/pET22b-tlr22 được sàng lọc thông qua kỹ thuật PCR khuẩn lạc với cặp mồi trên plasmid (T7Pro/T7Ter). Chủng E. coli SHuffle[®] T7 Express/pET22b-tlr22 được nuôi cấy lắc ở 30⁰C trong môi trường LB chứa kháng sinh Ampicillin (100 µg/mL). Sau 16 giờ nuôi cấy, vi khuẩn được cấy chuyền với tỉ lệ 1:10 (v/v) và tiếp tục nuôi cấy lắc ở 30°C. Đến khi OD_{600nm} của dịch vi khuẩn đạt giá trị 0,8–1,0, chất cảm ứng IPTG được bổ sung vào ống dịch vi khuẩn sao cho nồng độ cuối đạt 0,5 mM, tiếp tục lắc mẫu ở 16°C. Sau 6 giờ cảm ứng, tiến hành thu sinh khối tế bào (sinh khối trong 1,5 mL dịch nuôi cấy được hòa trong 300 µL PBS 1X) và phá màng tế bào bằng sóng siêu âm để thu được protein ở các pha tổng, tan và tủa. Sự biểu hiện của protein tái tổ hợp được kiểm tra bằng phương pháp điện di SDS-PAGE và nhuộm Coomassie Brilliant Blue G-250 (Biobasic). Thực hiện đồng thời với mẫu đối chứng âm là mẫu dịch pha tổng của E. coli SHuffle® T7 Express/pET22b-tlr22 không có cảm ứng IPTG. Sau đó protein mục tiêu được xác nhận bằng Western blot với kháng thể đặc hiệu kháng anti His-tag (ProteinTech, Mỹ). Cơ chất TMB (3,3',5,5'-tetra methyl benzidine, Thermo Scientific, Mỹ) được bổ sung vào sẽ phản ứng với HRP (Horseradish peroxidase enzyme) gắn trên kháng thể đặc hiệu tạo tín hiệu màu trên màng, từ đó xác nhận được sự có mặt của protein mục tiêu.

Đánh giá khả năng bắt vi khuẩn của protein TLR22 bằng kỹ thuật Dot blot

Để kiểm tra khả năng bắt vi khuẩn của TLR22 tái tổ hợp, các mẫu vi khuẩn *Escherichia coli, Aeromonas hydrophila, Streptococcus agalactiae, Lactiplantibacillus plantarum* được ủ với dịch protein tổng ở pha tan. Vi khuẩn *A. hydrophila* được tăng sinh trong môi trường TSB lỏng lắc qua đêm (16-24 tiếng), ở 30°C. Vi khuẩn *E. coli* DH5α được tăng sinh trong môi trường LB lỏng lắc qua đêm (16-24 tiếng), ở 37°C. Vi khuẩn *S. agalactiae* được tăng sinh trong môi trường BHI lỏng lắc qua đêm (16-24 tiếng), ở 37°C. Vi khuẩn *S. agalactiae* được tăng sinh trong môi trường BHI lỏng lắc qua đêm (16-24 tiếng), ở 37°C.

sinh trong môi trường MRS lỏng qua đêm (16-24 tiếng), ở nhiệt độ phòng, không lắc. Các loài vi khuẩn được dùng phương pháp đo độ đục OD_{600nm} để điều chỉnh dịch huyền phù. Đối với *E. coli*, $OD_{600nm} \approx 1,0$ (tương đương 10^8 CFU/mL) (Burns, Hull, 1999); *A. hydrophila*, $OD_{600nm} \approx 1,0$ (tương đương 10^9 CFU/mL) (Orozova *et al.*, 2012); *S. agalactiae*, $OD_{600nm} \approx 0,9$ (tương đương 10^9 CFU/mL) (Rajagopal *et al.*, 2005), *L. plantarum*, $OD_{600nm} \approx 0,22$ (tương đương 10^8 CFU/mL) (Hosiana *et al.*, 2020). Mật độ vi khuẩn ban đầu được xác định lại bằng phương pháp đếm khuẩn lạc. Pha loãng dịch vi khuẩn đồng nhất về 10^6 CFU/mL, ly tâm ở 5.000 rpm trong 5 phút để thu sinh khối và hoà lại bằng PBS 1X. Sau đó, các mẫu vi khuẩn được chuyển lên màng nitrocellulose đã được khoá bằng BSA 5%. Cuối cùng, kháng thể kháng His-tag gắn HRP được bổ sung lên màng. Ở thí nghiệm này chứng dương là các dịch tổng protein TLR22 ủ với kháng thể kháng His-tag gắn HRP. Chứng âm sẽ sử dụng dịch protein tổng ở pha tan của chủng *E. coli* SHuffle[®] T7 Express không mang plasmid tái tổ hợp.

KẾT QUẢ VÀ THẢO LUẬN

Cấu trúc plasmid tái tổ hợp pET22b-tlr22

Nhằm mục đích dòng hóa plasmid tái tổ hợp pET22b-*tlr22*, gen *tlr22* được thu nhận bằng phương pháp PCR với cặp mồi đặc hiệu 998FNde và 999RXho trên plasmid khuôn là G-Block. Kết quả điện di sản phẩm PCR trên bản gel agarose cho thấy vạch gen thu được (Hình 1, giếng 2), phù hợp với kích thước dự đoán ban đầu của gen *tlr22* là 1.436 bp so với thang phân tử lượng.





Hình 2. Kết quả thu nhận, xử lý plasmid pET22b với cặp enzyme *Ndel/Xho*l và phân tích bằng điện di trên gel agarose 1,5%

bằng điện di trên gel agarose 1,5%
M, thang DNA 1 kb (Thermo Fisher Scientific, Hoa Kỳ); 1, chứng âm; 2, sản phẩm PCR thu gen th22

Hình 1. Kết quả PCR thu nhận gen tlr22 và phân tích

M, thang DNA 1 kb; 1, plasmid chưa xử lý với enzyme; 2, plasmid đã xử lý với cặp enzyme Ndel/Xhol

Plasmid pET22b được thu nhận từ chủng *E. coli* DH5α/pET22b bằng phương pháp Alkaline Lysis. Tiến hành xử lí mẫu plasmid đã thu được với cặp enzyme cắt giới hạn *Nde*l và *Xho*l. So sánh với kết quả điện di plasmid ban đầu tồn tại ở ba dạng cấu hình (Hình 2, giếng 2), plasmid đã cắt còn một dạng cấu hình dạng thẳng do đã cắt mở vòng thành công.

Các khuẩn lạc *E. coli* MC1061 mọc trên đĩa môi trường được chọn ngẫu nhiên để thực hiện PCR khuẩn lạc với cặp mồi chéo (T7Pro/999RXho) nhằm kiểm tra việc mang plasmid tái tố hợp pET22b-*tlr22*. Hình ảnh bản gel điện di sản phẩm PCR khuẩn lạc xuất hiện vạch DNA ở giếng 3; 5; và 6, với kích thước xấp xỉ 1.636 bp, phù hợp với kích thước dự đoán khi PCR khuôn plasmid pET22b-*tlr22* (Hình 3). Kết quả cho thấy có 3 trong 4 khuẩn lạc dự tuyển mang plasmid tái tổ hợp mục tiêu pET22b-*tlr22*. Các khuẩn lạc dương tính được tách chiết, giải trình, và so sánh với trình tự được công bố. Kết quả cho thấy gen *tlr22* đã được dòng hóa đồng khung và đúng với trình tự công bố trên Ngân hàng gen (kết quả không trình bày).



Hình 3. Kết quả PCR sàng lọc thể biến nạp *E. coli* MC1061/pET22b-*tlr*22 bằng cặp mồi T7Pro/T7Ter và phân tích bằng điện di trên gel agarose 1,5%

M, thang DNA 1 kb (Thermo Fisher Scientific, Hoa Kỳ); 1, chứng âm; 2, chứng dương với khuôn pET22b; 3-6, các khuẩn lạc dự tuyển sàng lọc

Kiểm tra sự biểu hiện protein tái tổ hợp TLR22

Tương tự việc kiểm tra dòng hóa trên chủng *E. coli* MC1061. Bản gel điện di Hình 4 cho thấy, sản phẩm PCR khuẩn lạc *E. coli* SHuffle[®] T7 Express dự tuyển xuất hiện vạch DNA ở vị trí xấp xỉ vạch 1.736 bp của thang phân tử lượng, phù hợp kích thước dự đoán khi dòng tế bào dự tuyển mang plasmid pET22b-*tlr*22; đồng thời, chứng âm là sản phẩm phản ứng PCR không bổ sung khuôn không thấy xuất hiện vạch. Như vậy, có thể kết luận plasmid tải tổ hợp pET22b-*tlr*22 đã được biến nạp thành công vào chủng biểu hiện *E. coli* SHuffle[®] T7 Express sẵn sàng để được cảm ứng biểu hiện protein tái tổ hợp TLR22.



Hình 4. Kết quả PCR sàng lọc thể biến nạp *E. coli* SHuffle[®] T7 Express /pET22b-*tlr*22 bằng cặp mồi T7Pro/T7Ter và phân tích bằng điện di trên gel agarose 1,5%

M, thang DNA 1 kb (Thermo Fisher Scientific, Hoa Kỳ); 1, chứng âm; 2, plasmid pET22b; 3, khuẩn lạc dự tuyển sàng lọc.

Kết quả cảm ứng biểu hiện được trình bày ở hình điện di SDS-PAGE nhuộm Coomassie (Hình 5-A) và Western blot (Hình 5-B) cho thấy rằng có sự biểu hiện của protein, nhưng không thể hiện được sự biểu hiện vượt mức ở bản gel nhuộm Coomassie, vì thế không thể quan sát ở kết quả gel nhuộm Coomassie. Quan sát kết quả Western blot sau khi bổ sung cơ chất TMB cho thấy rằng, protein TLR22 biểu hiện đúng với khoảng kích thích dự đoán là 53 kDa. Bên cạnh đó chứng âm là *E. coli* SHuffle[®] T7 Express được cảm ứng IPTG không thấy xuất hiện vạch protein tương ứng, điều này cũng cho thấy bản thân chủng *E. coli* SHuffle[®] T7 Express không có trình tự gen mã hóa cho các protein mục tiêu. Cụ thể hơn, protein TLR22 biểu hiện ở dạng thể vùi và thể tan với mức độ biểu hiện rất thấp (Hình 5-B, giếng 5, giếng 6). Như vậy, đã biểu hiện thành công của thụ thể TLR22 tái tổ hợp trên chủng *E. coli* SHuffle[®] T7 Express protein với điều kiện biểu hiện là 0,5 mM IPTG ở 30°C.



Hình 5. Kết quả kiểm tra biểu hiện bằng điện di SDS-PAGE trên gel 12,5% nhuộm Coomassie (A) và khẳng định protein TLR22 bằng Western blot (B)

M, thang protein (GE Healthcare, Mỹ); 1 và 4, E. coli SHuffle[®] T7 Express; 2 và 5, E. coli SHuffle[®] T7 Express/pET22b-tlr22 ở pha tan; 3 và 6, E. coli SHuffle[®] T7 Express/pET22b-tlr22 ở pha tủa.

Thụ thể TLR22 tự nhiên ở cá tra có sự biến đổi sau dịch mã cùng với ba cầu nối disulfide, yếu tố ảnh hưởng đến sự gấp cuộn đúng của thụ thể (Quiniou *et al.*, 2013). Nghiên cứu trước, tiểu phần thụ thể TLR22(Irr8-11) đã được biểu hiện ở chủng *E. coli* BL21(DE3) hoàn toàn ở dạng thể vùi (Mai Quốc Gia *et al.*, 2022). Vì vậy, trong nghiên cứu này, sử dụng chủng biểu hiện *E. coli* SHuffle[®] T7 Express thay cho chủng *E. coli* BL21(DE3) nhằm tạo cầu nối disulfide trong việc hỗ trợ gấp cuộn đúng và làm tăng tính tan của các protein mục tiêu. Kết quả SDS-PAGE và Western blot cho thấy, thụ thể TLR22 (ba cầu nối disulfide) chỉ thu được ở dạng thể vùi. Điều này cho thấy, việc sử dụng chủng *E. coli* SHuffle® T7 Express thay vì *E. coli* BL21(DE3) đã cải thiện sự biểu hiện hòa tan của TLR22. Mặc khác, vai trò của cấu trúc cầu nối disulfide của TLR22 ở cá tra vẫn chưa được làm rõ, nên vẫn khảo sát thêm các hệ thống biểu hiện khác như nấm men hoặc tế bào động vật. Đối với nghiên cứu này, chúng tôi

bước đầu đánh giá khả năng tương tác của thụ thể TLR22 với vi khuẩn. Mặt khác, TLR22 sẽ được tối ưu cho việc biểu hiện protein ở các nghiên cứu tiếp theo.

Đánh giá khả năng bắt vi khuẩn của protein TLR22 bằng kỹ thuật Dot blot



Hình 6. Đánh giá khả năng tương tác của TLR22 với vi khuẩn bằng kỹ thuật Dot blot

Bảng 1. Mô tả các nghiệm thức Dot blot

Chấm	Vi khuẩn	Mẫu
1	A. hydrophila	
2	E. coli	Dich protein tổng ở pha tan của chủng <i>E. coli</i> SHuffle [®] T7 Express không
3	S. agalactiae	mang plasmid tái tổ hợp.
4	L. plantarum	
5	PBS	
6	A. hydrophila	
7	E. coli	Dịch protein tổng ở pha tan của chủng <i>E. coli</i> SHuffle [®] T7 Express biểu hiện protein TLR22
8	S. agalactiae	
9	L. plantarum	

Quan sát kết quả Dot blot cho thấy rằng, có tín hiệu màu ở chứng dương, cho thấy kháng thể His-tag gắn HRP bắt được protein TLR22 (Hình 6, chấm 5) không có hoặc có tín hiệu màu nhưng rất mờ ở các đối chứng âm (Hình 6, chấm 1-4). So với chứng âm kết quả cho thấy, có tín hiệu màu sau khi bổ sung cơ TMB ở hai mẫu là S. *agalactiae* (Hình 6, chấm 8) và *L. plantarum* (Hình 6, chấm 9), vì vậy có thể kết luận rằng TLR22 có khả năng bắt với vi khuẩn Gram dương. Điều này cho thấy có sự khác biệt với các công bố lý thuyết trước đây dự đoán thụ thể TLR22 có thể tương tác với vi khuẩn Gram âm (Zhang *et al.*, 2017). Tuy nhiên, với ở kết quả trong nghiên cứu này cho thấy TLR22 có khả năng tương tác với vi khuẩn Gram dương và tương tác yếu hay không tương tác với vi khuẩn Gram đương cũng như so sánh mức độ tương tác với Gram âm và Gram dương. Phát hiện này một lần nữa mở ra thêm ứng dụng cho TLR22 do cá tra vẫn bị nhiễm bệnh bởi vi khuẩn Gram dương như S. *agalactiae* nên nếu sử dụng thụ thể TLR22 có thể bắt được vi khuẩn Gram âm như các công bố trước đây và vừa có thể bắt được vi khuẩn Gram dương như trong nhiến cứu này.

KÉT LUẬN

Biểu hiện thành công thụ thể TLR22 tái tổ hợp trên chủng *E. coli* SHuffle[®] T7 Express ở điều kiện 30°C, nồng độ chất cảm ứng là 0,5mM IPTG. TLR22 có khả năng tương tác *in vitro* với vi khuẩn Gram dương thông qua kỹ thuật Dot blot, nhưng khả năng bắt vi khuẩn Gram dương còn thấp. Khó khăn hiện nay là chưa có tài liệu hoặc nghiên cứu nào đã và đang tìm hiểu sâu về thụ thể TLR22, nên việc hiểu rõ các đặc tính của TLR22 ở cá tra vẫn còn gặp còn nhiều hạn chế. Để có thể ứng dụng và phát triển chế phẩm sinh học bổ sung thức ăn cho cá thì đề tài này cần nghiên cứu sâu về các vùng chức năng/tiểu phần thụ thể, tối ưu việc biểu hiện, tinh sạch protein ở hiệu suất cao và cần sử dụng hệ thống biểu hiện protein có đủ những biến đổi sau dịch mã cần thiết cho việc biểu hiện các thụ thể TLR22.

Lời cảm ơn: Đề tài này được sự tài trợ bởi Sở Khoa học và Công nghệ tính An Giang cho chủ nhiệm đề tài Trần Văn Hiếu.

TÀI LIỆU THAM KHẢO

Burns SM, Hull SI (1999) Loss of resistance to ingestion and phagocytic killing by O- and K- mutants of a uropathogenic *Escherichia coli* O75: K5 strain. *Infect Immu* 67: (8) 3757-3762.

Hosiana N, Astuti D, Surono I (2020) Physio-chemical, Microbiology, and Preference of Probiotic Fresh Soft Cheese Using Lactobacillus plantarum IS-10506 and *Streptococcus thermophilus* as Mixed Starter Culture. Vol. 426 ed.^eds.), p.^pp. 012185. IOP Publishing.

Kalimuddin S, Chen SL, Lim CT, Koh TH, Tan TY, Kam M, Wong CW, Mehershahi KS, Chau ML, Ng LC (2017) 2015 epidemic of severe *Streptococcus agalactiae* sequence type 283 infections in Singapore associated with the consumption of raw freshwater fish: a detailed analysis of clinical, epidemiological, and bacterial sequencing data. *Clin Infect Dis* 64: (2) S145-S152.

Mai Quốc Gia, Lê Thị Xuân Trang, Trần Văn Hiếu (2022) Tạo dòng, biểu hiện và tái gấp cuộn Toll-like receptor 22 từ cá tra Pangasianodon hypophthalmus. Hội nghị Công nghệ sinh học toàn quốc 2022 Tiểu ban Công nghệ Sinh học Nông nghiệp.

Meidong R, Nakao M, Sakai K, Tongpim S (2021) *Lactobacillus paraplantarum* L34b-2 derived from fermented food improves the growth, disease resistance and innate immunity in *Pangasius bocourti*. *Aquacult* 531: 735878.

Orozova P, Sirakov I, Petkov I, Crumlish M, Austin B (2012) Recovery of *Aeromonas hydrophila* associated with bacteraemia in captive snakes. *FEMS Microbiol Let* 334: (1) 22-26.

Rajagopal L, Vo A, Silvestroni A, Rubens C (2005) Regulation of purine biosynthesis by a eukaryotic-type kinase in *Streptococcus agalactiae. Mol Microbiol* 56: (5) 1329-1346.

Thi QVC (2017) Nghiên cứu đặc điểm bệnh học và cơ chế đa kháng thuốc của hai loài vi khuẩn *Edwardsiella ictaluri* và *Aeromonas hydrophila* gây bệnh trên cá tra (*Pangasianodon hypophthalmus*) nuôi thâm canh ở Đồng bằng sông cửu long. Thesis.

Trân NTN, Nghĩa NT, Oanh ĐTH (2016) Xác định tính kháng khuẩn của vi khuẩn lactic với vi khuẩn (*Streptococcus agalactiae*) phân lập từ cá rô phi (*Oreochromis niloticus*) bệnh phù mắt và xuất huyết. *Tạp chí Khoa học Đại học Cần Thơ* 42: (42) 48-55.

Vo-Nguyen HV, Nguyen TT, Mai QG, Tran TT, Tran TL, Tran-Van H (2022) Recombinase-free cloning (RFC) protocol for gene swapping. *Mol Biol Res Comm* 11: (1) 21.

Wise DJ, Greenway TE, Byars TS, Griffin MJ, Khoo LH (2015) Oral vaccination of channel catfish against enteric septicemia of catfish using a live attenuated *Edwardsiella ictaluri* isolate. *J Aqua Animal Health* 27: (2) 135-143.

Zhang XT, Zhang GR, Shi ZC, Yuan YJ, Zheng H, Lin L, Wei KJ, Ji W (2017) Expression analysis of nine Toll-like receptors in yellow catfish (*Pelteobagrus fulvidraco*) responding to *Aeromonas hydrophila* challenge. *Fish Shell Immunol* 63: 384-393.

EXPRESSION OF TOLL-LIKE RECEPTOR 22 OF Pangasianodon hypophthalmus AND EVALUATION OF THE INTERACTION WITH BACTERIA

Thanh-Tan Nguyen^{1,3}, Hieu Tran-Van^{1,2,3*}

¹Laboratory of Biosensors, Faculty of Biology and Biotechnology, University of Science, Ho Chi Minh City, Vietnam

²Department of Molecular and Environmental Biotechnology, Faculty of Biology and Biotechnology, University of Science, Ho Chi Minh City, Vietnam

³Vietnam National University, Ho Chi Minh City, Vietnam

SUMMARY

Catfish (Pangasianodon hypophthalmus) is one of the most commonly farmed catfish species in the Mekong Delta. Both domestically and internationally, people are increasingly consuming catfish due to their high nutritional value. To enhance competitiveness, we need to focus on developing and improving the quality and quantity of catfish. In recent years, the production and consumption of catfish have faced numerous challenges, including unstable prices and fish diseases, particularly the high mortality rate (50-90%) due to liver, kidney, and hemorrhagic diseases. The use of antibiotics in disease treatment leads to antibiotic residues in fish meat and drug resistance. Vaccines for fish still have many limitations in distribution and preservation. As a result, we aimed to develop a new biological product capable of preventing liver, kidney, and hemorrhagic disease in fish, as well as bacterial diseases in fish in general. Only fish harbors the Toll-like receptor (TLR) 22 receptor, which recognizes and binds to viral dsRNA and lipopolysaccharides through leucine-rich repeat sequences (LRR), which can be exploited to capture bacteria. In this study, we produced the TLR22 receptor as a recombinant protein and evaluated its interaction with bacteria. We constructed the plasmid pET22b-tlr22 in E. coli MC1061, expressed the TLR22 receptor in E. coli SHuffle® T7 Express, and confirmed it by Western blot using anti-Histag (6xHis) antibodies. The Dot-blot technique assessed the bacterial binding ability. Results showed that the TLR22 receptor could bind to Gram-positive bacteria. This study is a foundation for further research on developing biological products to supplement fish feed for treating bacterial diseases in catfish, and other fish species.

Keyword: Gram-positive bacteria, Leucine-rich repeat, Pangasianodon hypophthalmus, TLR22, Streptococcus agalactiae.

^{*} Author for correspondence: Tel: 0983260781; Email: tvhieu@hcmus.edu.vn

NGHIÊN CỨU ĐẶC ĐIỀM HỆ GENE CỦA 243 MÃU VI KHUẦN Neisseria gonorrhoeae PHÂN LẬP TẠI VIỆT NAM

Nguyễn Tiến Đạt¹, Vương Thị Hương¹, Trịnh Thị Xuân², Nguyễn Cường^{3,*}

¹Công ty TNHH LOBI Việt Nam

²Khoa Công nghệ thông tin, Đại học Mở Hà Nội

³Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Tình trạng kháng kháng sinh ở loài vi khuẩn Neisseria gonorrhoeae gây bệnh lậu ở người, đang là một mối nguy hại sức khỏe toàn cầu theo WHO. Khu vực Đông Á và Đông Nam Á là nguồn chính xuất hiện và lây lan các chủng vi khuẩn này trong nhiều thập kỉ gần đây, trong đó có Việt Nam. Mặc dù vậy, các nghiên cứu về đặc điểm hệ gene của loài vi khuẩn này ở Việt Nam còn rất hạn chế. Đa số các nghiên cứu ở Việt Nam về vi khuẩn này trong 5 năm trở lai đây tập trung vào nghiên cứu xu hướng kháng kháng sinh của chúng dựa vào các kĩ thuật vi sinh và có rất ít nghiên cứu phân tích về đặc điểm hệ gene của N. gonorrhoeae. Do đó, chúng tôi xây dựng một quy trình bao gồm 4 nôi dung cần thiết để phân tích các đặc điểm di truyền của 243 trình tư thô của vi khuẩn N. gonorrhoeae ở Việt Nam bao gồm: Lắp ráp, dự đoán, chú giải hệ gene; dự đoán các gene độc lực và gene kháng kháng sinh; xác định ST và vẽ cây phân loài dựa vào các house-keeping gene; phân tích Pangenome. Kết quả thu được 241/243 hê gene lắp ráp và chú giải phù hợp (kích thước hê gene 2,074-2,307 Mb; số lượng gene 2.024-3.112; tỷ lệ GC 52,27%-52,70%) có độ tương đồng cao với hệ gene tham chiếu ASM1303007v1 (kích thước hệ gen 2,2Mb; số lượng gene 2.283; tỷ lệ GC 52,5%). Đồng thời, phát hiện được 43 gene độc lực và 11 gene kháng kháng sinh. Trong đó, có 39/43 gene đôc lực và 7/11 gene kháng kháng sinh xuất hiện ở tất cả các hệ gene lắp ráp. Kết quả phân tích pangenome bao gồm: 1600/4270 (37,47%) core gene, 1925/4270 (45,08%) accessory gene và 745/4270 (17,45%) unique gene. Kết quả phân nhóm ST cho biết 240/241 mẫu này thuộc 41 nhóm ST khác nhau với mẫu ST7371 chiếm tỷ lê cao nhất (11,11%). Kết quả xây dựng cây phân loài cho thấy có 2 nhánh được phân chia từ gốc, nhánh 1 với 117 hệ gene và nhánh 2 với 123 hệ gene.

Từ khóa: Cây phân loại, đặc điểm hệ gene, Neisseria gonorrhoeae, Tin sinh học, tình trạng kháng kháng sinh.

MỞ ĐẦU

Vào năm 2019, Tổ chức WHO đã đưa ra một cảnh báo về 10 nguy cơ ảnh hưởng lớn đến sức khỏe mà nhân loại đang phải đối mặt (WHO, 2019). Trong đó phải kể đến sự xuất hiện ngày càng nhiều của các chủng vi khuẩn kháng sinh do sự lạm dụng nhiều loại thuốc trong ngăn ngừa và điều trị các bệnh nhiễm trùng, nhiễm khuẩn. *N. gonorrhoeae* là một loài vi khuẩn gây bệnh lậu ở người, lây qua đường tình dục (STIs). Theo WHO, mỗi năm có khoảng 82,4 triệu ca nhiễm trùng do *N. gonorrhoeae* trên toàn thế giới, xếp thứ 3 trong nhóm STIs trên phạm vi toàn cầu (WHO, 2024). Tuy nhiên, việc quản lý và điều trị nhiễm trùng *N. gonorrhoeae* ngày càng trở nên phức tạp do sự xuất hiện tình trạng kháng kháng sinh (AMR). Hiện nay, *N. gonorrhoeae* đã phát triển đế kháng tất cả các nhóm kháng sinh truyền thống được sử dụng để điều trị nhiễm trùng bệnh lậu (Adamson *et al.*, 2020).

Adamson và các cộng sự Việt Nam đã thực hiện một nghiên cứu về 409 chủng *N. gonorrhoeae* phân lập từ 2017- 2019 và phát hiện chúng đã kháng với penicillin, tetracycline, ciprofloxacin và đang có xu hướng giảm dần độ nhạy cảm với ceftriaxone, azithromycin. Đây là hai loại kháng sinh được khuyến cáo sử dụng khi bị nhiễm khuẩn *N. gonorrhoeae* mang lại hiệu quả khi mà các kháng sinh khác đã không thể tiêu diệt được loài vi khuẩn này (Adamson *et al.*, 2020). Điều này chứng tỏ mức độ nguy hiểm của loài vi khuẩn này ngày càng tăng ở Việt Nam. Thêm vào đó, số lượng các nghiên cứu về *N. gonorrhoeae* trong những năm gần đây rất ít và thường chỉ tập trung vào kiểm tra độ nhạy cảm với kháng sinh bằng các kĩ thuật vi sinh chứ chưa đi sâu vào nghiên cứu đặc điểm hệ gene.

Vào năm 2023, Adamson và các cộng sự Việt Nam đã tiến hành một nghiên cứu khác về phát hiện các chủng *N. gonorrhoeae* kháng ceftriaxon. Kết quả nhấn mạnh rằng: Ceftriaxon, liệu pháp duy nhất còn lại trong điều trị lậu cầu đang bị đe dọa. Và việc phát hiện các chủng này là có thể sự xuất hiện của gene *penA-60.001* trong hệ gene của chúng (Adamson *et al.*, 2024).Vì vậy, các biện pháp can thiệp dựa vào đặc điểm hệ gene rất quan trọng nhằm phát hiện trình trạng kháng thuốc và giảm thiểu sự lây truyền thêm của *N. gonorrhoeae* một cách chính xác và nhanh chóng. Do đó, việc tìm hiểu về các đặc điểm di truyền để dự đoán và giám sát tình trạng kháng kháng sinh của *N. gonorrhoeae* là cần thiết.

Nhằm thực hiện mục tiêu này, chúng tôi lấy dữ liệu giải trình tự của 243 mẫu *N. gonorrhoeae* phân lập ở Việt Nam, được lưu trữ trên SRA và phân tích các đặc điểm di truyền của chúng bằng các quy trình Tin sinh học. Nội dung nghiên cứu quy trình bao gồm: Lắp ráp, dự đoán, chú giải hệ gene; dự đoán các gene độc lực và gene kháng kháng sinh; xác định ST và vẽ cây phân loài dựa vào các house-keeping gene và phân tích Pangenome. Các kết quả này có thể ứng dụng trong dự đoán, giám sát và các nghiên cứu liên quan về *N. gonorrhoeae* nói riêng cũng như nhiều loài vi khuẩn khác nói chung trong tương lai.

DỮ LIỆU VÀ PHƯƠNG PHÁP

Dữ liệu

Trong nghiên cứu này, chúng tôi sử dụng 243 bộ dữ liệu WGS của vi khuẩn *N. gonorrhoeae* được phân lập ở Việt Nam. Trong đó, 229 mẫu thu thập vào năm 2011 và 2015-2016 ở Bệnh viện Da liễu, Hà Nội (Lan et al., 2020). Còn lại 14 mẫu phân lập trong năm 2019-2020 ở Hồ Chí Minh và Đà Nẵng (Trinh et al., 2022). Toàn bộ dữ liệu WGS này được tải về từ cơ sở dữ liệu SRA với mã truy cập là PRJEB34425 và PRJEB45627.

Phương pháp

Lắp ráp, dự đoán và chú giải hệ gen

Dữ liệu WGS với định dạng FastQ được đánh giá bằng FastQC v0.12.1 và tiền xử lý bằng Trimmomatic v0.39 với các tham số như sau: SLIDINGWINDOW:4:20 (Loại bỏ các vị trí bases có chất lượng trung bình thấp hơn 20), MINLEN:100 (Loại bỏ các đoạn trình tự ngắn hơn 100bp) và loại bỏ các adapters. Dữ liệu trình tự sau tinh sạch được lắp ráp bằng phần mềm Shovill v1.1.0 với tham số mặc định. Kết quả lắp ráp được đánh giá chất lượng bằng phần mềm Shovill v1.1.0 với tham số mặc định. Kết quả lắp ráp được đánh giá chất lượng bằng phần mềm QUAST v5.0.2 (Gurevich *et al.*, 2013) với tham số mặc định và đánh giá độ hoàn thiện bằng BUSCO v5.6.1 (Manni *et al.*, 2021) với cơ sở dữ liệu bacteria_odb10. Các hệ gene có độ hoàn thiện cao (>90% Complete) được giữ lại và thực hiện dự đoán, chú giải gene bằng công cụ Prokka v1.14.6 với các tham số mặc định.

Dự đoán gene độc lực và gene kháng kháng sinh

Dựa vào các thông tin về gene độc lực và gene kháng kháng sinh trên cơ sở dữ liệu lần lượt là VFDB (http://www.mgc.ac.cn/VFs/main.htm) và CARD (https://card.mcmaster.ca/), công cụ Abricate v1.0.1 (Seemann, 2023) được sử dụng để phát hiện các đoạn trình tự có mức độ tương đồng cao (>80%) so với các gene trong cơ sở dữ liệu. Các gene này sau đó được thống kê và phân nhóm theo các nhóm độc lực trên VFDB và nhóm kháng sinh trên CARD.

Phân tích đại hệ gene Pan-Genome

Các kết quả chú giải (định dạng gff) từ Prokka v1.14.6 (Seemann, 2014) được tổng hợp lại và phân tích pangenome bằng công cụ Roary (Page *et al.*, 2015). Sau đó, số lượng các nhóm genes như: Core gene, Unique gene và new gene được thống kê theo số lượng hệ gene phân tích và trực quan hóa bằng python.

Xây dựng cây phân loài dựa trên ST

Các contigs thu được từ kết quả lắp ráp được sử dụng làm đầu vào cho công cụ mlst v2.23 dựa vào thông tin phân loại ST trên cơ sở dữ liệu pubMLST (https://pubmlst.org/neisseria). Phương pháp này dựa vào việc xác định sequence type của 7 house-keeping gene của *Neisseria* (*adk, abcZ, aroZ, fumC, pgm, gdh, pdhC*) để phân nhóm ST. Sau đó, công cụ MAFFT v7.526 (Katoh *et al.*, 2002) được dùng cho việc đóng hàng đa trình tự. Kết quả dóng hàng được đưa vào FastTree v2.1.11 để vẽ cây phân loài và iTOL v6 được sử dụng để trực quan hóa hình dạng cây.

KẾT QUẢ VÀ THẢO LUẬN

Kết quả lắp ráp, dự đoán và chú giải hệ gene

Hai trăm bốn mươi ba bộ dữ liệu hệ gene của *N. gonorrhoeae* được tinh sạch bằng Trimmomatic và lắp ráp bằng Shovill, thu được 243 hệ gen với kích thước từ 2,074-2,307Mb (trung bình 2,153 \pm 0,027) với số lượng contig từ 68-1.375 (trung bình 103 \pm 98) với chỉ số N50 từ 2.178-80.821 (trung bình 55.604 \pm 10.133).

Kết quả dự đoán gene xác định số lượng gene của 243 hệ gene vi khuẩn *N. gonorrhoeae* dao động từ 2.024-3.112 (trung bình 2.125 ± 79) gene. Thêm vào đó, các thống kê về số lượng RNA bao gồm: 3-4 rRNA, 48-61 tRNA và 1 tmRNA. Tỷ lệ GC khoảng 52,27%-52,70%. Kết quả lắp ráp, dự đoán và chú giải hệ gene này phù hợp với hệ gen tham chiếu (ASM1303007v1) có kích thước hệ gene là 2,2Mb, số lượng gene 2.283, tỷ lệ GC 52,5%.

Độ toàn vẹn hệ gene được đánh giá bằng phần mềm BUSCO cho thấy tỷ lệ hoàn thiện của 241/243 mẫu *N. gonorrhoeae* đều trên 90% (Complete). Hai mẫu còn lại là *ERR3623126* và *ERR3623149* có độ hoàn thiện thấp hơn 90%, lần lượt là 59.7% và 75% bị loại bỏ khỏi các phân tích tiếp theo.

Kết quả dự đoán gene độc lực và gene kháng kháng sinh

Các gene độc lực là các gene có khả năng mã hóa thành các protein giúp vi khuẩn xâm nhập, bám dính, sinh sản và gây hại cho tế bào vật chủ, dẫn đến nhiễm khuẩn. Kết quả xác định gene độc lực trong cơ sở dữ liệu VFDB cho thấy có 43 gene độc lực tồn tại trong 241 hệ gene *N. gonorrhoeae*. Trong đó, 39/43 gene được phát hiện ở hầu như tất cả các mẫu, chiếm 88%. Trong 39 gene chung này, có 18 gene thuộc nhóm Adherence (Bám dính), chiếm số lượng nhiều nhất. Ngoài ra, số lượng các gene thuộc nhóm khác bao gồm: 6 gene thuộc Stress survival, 9 gene thuộc nhóm Nutritional/metabolic factor, 5 gene thuộc Antimicrobial activity/Competitive advantage, 1 gene thuộc nhóm Immune modulation. 4/43 gene còn lại chỉ tồn tại ở một số mẫu (Tỷ lệ chưa tới 10%). Đặc biệt có xuất hiện gene *porB* là một gene được phân vào nhóm Invasion với tỷ lệ 6.6%, một nhóm mới so với các nhóm ở trên. Các kết quả thống kê chi tiết được trình bày ở Bảng 1.

Nhóm	Tên gene	Số lượng	Tỷ lệ
	pilX, pilK, pilJ, pilI, pilH, pilZ, pilT2, pilP, pilO, pilN, pilM, pilW, pilF, pilD, pilG, pilV, pilT	241	100%
Adherence	pilU	240	99.6%
	pilQ	2	0.8%
	pilE	9	3.7%
Stress Survival	recN, mntA, mntB, mntC, katA, msrA/B (pilB)	241	100%
Nutritional/matchalia factor	fbpA, fbpB, fbpC, lbpA, lbpB, hpuA, hpuB, hmbR, tbpA	241	100%
	tbpB	2	0.8%
Antimicrobial activity	mtrC, mtrD, mtrE, farA, farB	241	100%
Immune modulation	nspA	241	100%
Invasion	porB	16	6.6%

Bảng 1. Thống	kê số lượng	ác gene độc lự	rc trong 241 mẫu l	N. gonorrhoeae
---------------	-------------	----------------	--------------------	----------------

Trong 241 mẫu *N. gonorrhoeae*, chúng tôi phát hiện được 11 gene kháng kháng sinh theo cơ sở dữ liệu CARD. Đồng thời, có 7/11 (63.64%) gene có ở tất cả các mẫu với tỷ lệ xuất hiện là 100%. Các gene này kháng với macrolide, penam và free fatty acids. Các gene còn lại chia làm 2 nhóm, bao gồm: Nhóm kháng tetracycline có 97 mẫu; nhóm kháng cephalosporin, monobactam, penam, penem với 80 mẫu. Thông tin về tên gene và số lượng mẫu có ở Bảng 2.

Tên gene	Số lượng mẫu	Tỷ lệ (%)	Loại thuốc kháng
tetM	97	40,25%	Tetracycline
mtrC	241	100%	Macrolide, penam
mtrD	241	100%	Macrolide, penam
mtrE	241	100%	Macrolide, penam
farA	241	100%	Free fatty acids
farB	241	100%	Free fatty acids
macA	241	100%	Macrolide
macB	241	100%	Macrolide
TEM-1	43	17,84%	Cephalosporin, monobactam, penam, penem
TEM-135	36	14,94%	Cephalosporin, monobactam, penam, penem
TEM-206	1	0,41%	Cephalosporin, monobactam, penam, penem

Bảng 2. Thống kê số lượng các gene kháng kháng sinh trong 241 mẫu N. gonorrhoeae

Kết quả phân tích Pangenome

Pangenome là tập hợp toàn bộ gene của tất cả các cá thể trong một loài. Cấu trúc của pangenome bao gồm: Core genome (Các gene chung của tất cả các cá thể và có liên quan đến các chức năng cơ bản và thiết yếu cho sự sống), Accessory genome (Các gene hiện diện ở một số cá thể và có liên quan đến đặc điểm thích nghi cụ thể của từng nhóm nhỏ), Unique genome (Các gene tồn tại ở 1 chủng duy nhất, mang tính đặc trưng cho từng cá thể). Trong 4270 gene được phát hiện ở 241 mẫu *N. gonorrhoeae*, có 1675 (37,47%) core gene, 1925 (45,08%) accessory gene và 745 (17,45%) unique gene.



Hình 1. A-Mối quan hệ giữa core genes và số lượng mẫu; B-Mối quan hệ giữa unique genes và new genes với số lượng mẫu

Các mối quan hệ giữa các nhóm genes với số lượng hệ genes được thể hiện ở Hình 1. Khi số lượng hệ genes tăng thì số lượng các gene bảo tồn (Core genome) có xu hướng giảm dần trong khoảng 25 hệ gene đầu và hầu như không thay đổi khi phân tích thêm vào các hệ gene tiếp theo. Tương tự, số lượng unique gene (Tổng hợp các gene duy nhất trong các hệ gene phân tích) có xu hướng giảm ở các hệ gene đâu và tăng dần khi phân tích thêm càng nhiều hệ gene. Ngược lại, số lượng new genes (Số lượng gene mới so với các hệ gene phân tích trước đó) giảm dần theo cấp số nhân. Điều này chứng tỏ rằng các gene đặc trưng riêng cho từng cá thể *N. gonorrhoeae* xuất hiện ngày càng nhiều, giúp vi khuẩn thích nghi với nhiều loại kháng sinh khác nhau.

Kết quả xây dựng cây phân loài dựa trên ST

Trong 241 mẫu *N. gonorrhoeae* thực hiện phân tích, có 1 mẫu là *ERR3623233* không xác định được nhóm ST. Kết quả chí tiết về các nhóm ST và số lượng mẫu thuộc từng nhóm có ở Bảng 3. Đồng thời, các kết quả phân tích tỷ lệ của từng nhóm cũng được thể hiện ở Hình 2.

STT	Số lượng	Tỷ lệ	ST	Số lượng	Tỷ lệ	ST	Số lượng	Tỷ lệ
ST7371	27	11,11%	ST7360	5	2,06%	ST7367	2	0,82%
ST1588	23	9,47%	ST6714	5	2,06%	-	1	0,82%
ST7363	18	7,41%	ST1596	4	1,65%	ST11710	1	0,41%
ST1600	16	6,58%	ST8153	4	1,65%	ST12516	1	0,41%
ST1901	14	5,76%	ST7823	3	1,23%	ST12520	1	0,41%
ST13871	14	5,76%	ST8109	3	1,23%	ST1904	1	0,41%
ST7822	13	5,35%	ST1902	3	1,23%	ST16471	1	0,41%
ST11230	11	4,53%	ST8143	2	0,82%	ST1921	1	0,41%
ST7827	10	4,12%	ST10241	2	0,82%	ST1925	1	0,41%
ST1583	10	4,12%	ST9900	2	0,82%	ST1927	1	0,41%
ST10313	7	2,88%	ST6960	2	0,82%	ST11231	1	0,41%
ST10314	7	2,88%	ST11173	2	0,82%	ST1963	1	0,41%
ST8123	7	2,88%	ST12518	2	0,82%	ST1580	1	0,41%
ST1920	7	2,88%	ST12514	2	0,82%	ST1582	1	0,41%

Bảng 3. Thống kê số lượng	T trong 243 mẫu	N. gonorrhoeae
---------------------------	-----------------	----------------

Kết quả cây phân loài theo ST cho thấy các nhóm ST được phân theo từng cụm với đơn vị độ dài của cây là 0.001. Trong đó, Ngoài ra, có thể thấy được nhóm ST có khoảng cách xa nhất so với gốc là ST8153 (thuộc nhóm others) và nhóm ST8123 (màu vàng) có khoảng cách gần nhất so với gốc. Từ gốc, cây phân loài chia làm 2 nhánh: Nhánh 1 chứa 117 hệ gene với đa số các hệ gene thuộc 6 nhóm ST: ST1901, ST7371, ST10314, ST1583, ST7822 và ST11230; nhánh 2 chứa 123 hệ gene. Thêm vào đó, nhánh 1 cũng là nhánh có sự đa dạng về ST hơn so với nhánh 2 vì số lượng các hệ gene thuộc nhóm others (màu đen) chiếm số lượng nhiều hơn hẳn. Trong đó, nhóm ST7371 chiếm tỷ lệ nhiều nhất ở nhánh 1 có mối quan hệ gần nhất với nhóm ST7822 (màu xám) và nhóm ST1588 chiếm tỷ lệ nhiều nhất ở nhóm 2 có khoảng cách gần nhất với nhóm ST1596 (nhóm others).

CÔNG NGHỆ GEN



Hình 2. Cây phân loại theo MLST

KÉT LUÂN

243 bộ dữ liệu giải trình tự Illumina từ SRA của loài vi khuẩn N. gonorrhoeae được sử dụng trong nghiên cứu đặc điểm hệ gene thu được 241 hệ gene phù hợp sau khi lắp ráp. Các hệ gene này có kích thước từ 2,074-2,307 Mb, tỷ lệ GC khoảng 52,27%-52,70% và số lượng gene 2.024-3.112. Các số liệu này phù hợp với hệ gene tham chiếu của loài vi khuẩn này với mã ID là ASM1303007v1 trên NCBI. Số lượng gene độc lực và kháng kháng sinh phát hiện lần lượt là 43 và 11. Trong đó, có 39/43 và 7/11 gene xuất hiện ở tất cả các mẫu. Do đó, 241 hệ gene này đều có khả năng kháng các loại kháng sinh bao gồm: macrolide, penam, free fatty acids và có thể gây độc theo nhiều cơ chế khác nhau như: Adherence, stress survival, nutritional/metabolic factor, antimicrobial activity, immune modulation. Đồng thời, phát hiện được 1675/4270 (37,47%) core gene, 1925/4270 (45,08%) accessory gene và 745/4270 (17,45%) unique gene. Xu hướng của core gene và unique gene giảm dần ở các mẫu ban đầu và thay đối rất ít khi tăng số lượng mẫu lên. Mặt khác, số lượng new gene giảm theo cấp số nhân khi tăng lượng mẫu. Kết quả phân tích ST cho thấy có 240/241 hệ gene xác định được ST. Trong đó, tỷ lệ ST7371 là nhiều nhất với 11,11%. Cây phân loài xây dưng bằng sự khác biết giữa 7 house-keeping gene chia các mẫu thành 2 nhánh: nhánh 1 (117 hệ gene) với độ đa dạng về ST hơn so với nhánh 2 (123 hệ gene).

Lòi cảm on: Nghiên cứu này được tài trợ bởi Công ty TNHH LOBI Việt Nam.

TÀI LIÊU THAM KHẢO

Adamson PC, Le HV, Le HHL, Le GM, Nguyen TV, Klausner JD (2020) 'Trends in antimicrobial resistance in Neisseria gonorrhoeae in Hanoi, Vietnam, 2017-2019, BMC Infect Dis, 20(1), p. 809. Available at: https://doi.org/10.1186/s12879-020-05532-3.

Adamson PC, Hieu VN, Nhung PH, Whiley DM, Chau TM (2024) 'Ceftriaxone resistance in Neisseria gonorrhoeae associated with the penA-60.001 allele in Hanoi, Viet Nam', Lancet Infect Dis, 24(6), pp. e351-e352. Available at: https://doi.org/10.1016/S1473-3099(24)00230-5.

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) 'QUAST: quality assessment tool for genome assemblies', Bioinformatics, 29(8), pp. 1072-1075. Available at: https://doi.org/10.1093/bioinformatics/btt086.

Katoh K, Misawa K, Kuma K, Miyata T (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', Nucleic Acids Res, 30(14), pp. 3059–3066. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC135756/ (Accessed: 2 July 2024).

Lan PT, Golparian D, Ringlander J, Van HL, Van TN, Unemo M (2020) 'Genomic analysis and antimicrobial resistance of Neisseria gonorrhoeae isolates from Vietnam in 2011 and 2015–16', J Antimicrob Chemoth, 75(6), pp. 1432–1438. Available at: https://doi.org/10.1093/jac/dkaa040.

Manni M, Berkeley MR, Seppey M, Zdobnov EM (2021) 'BUSCO: Assessing Genomic Data Quality and Beyond', Current Protocols, 1(12), p. e323. Available at: https://doi.org/10.1002/cpz1.323.

Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S (2015) 'Roary: rapid large-scale prokaryote pan genome analysis', Bioinformatics, 31(22), pp. 3691-3693. Available at: https://doi.org/10.1093/bioinformatics/btv421.

Seemann T (2014) 'Prokka: rapid prokaryotic genome annotation', *Bioinformatics*, 30(14), pp. 2068–2069. Available at: https://doi.org/10.1093/bioinformatics/btu153.

Seemann T (2023) 'ABRicate'. Available at: https://github.com/tseemann/abricate (Accessed: 14 September 2023).

Trinh TM, Nguyen TT, Le TV, Nguyen TT, Ninh DT, Duong BH, Nguyen MV, Kesteman T, Pham LT, Doorn HR (2022) 'Neisseria gonorrhoeae FC428 Subclone, Vietnam, 2019–2020 - Volume 28, Number 2—February 2022 - Emerging Infectious Diseases journal - CDC'. Available at: https://doi.org/10.3201/eid2802.211788.

WHO (2019) Ten health issues WHO will tackle this year. Available at: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019 (Accessed: 10 July 2024).

WHO (2024) Gonorrhoea (Neisseria gonorrhoeae infection). Available at: https://www.who.int/news-room/fact-sheets/detail/gonorrhoea-(neisseria-gonorrhoeae-infection) (Accessed: 10 July 2024).

STUDY ON THE GENOMIC CHARACTERISTICS OF 243 NEISSERIA GONORRHOEAE ISOLATES IN VIETNAM

Nguyễn Tiến Đạt¹, Vương Thị Hương¹, Trinh Thi Xuan², Nguyễn Cường^{3,*}

¹LOBI Vietnam Ltd.

²Faculty of Information Technology, Hanoi Open University

³Institute of Information Technology, Vietnam Academy of Science and Technology

SUMMARY

Currently, antibiotic resistance Neisseria gonorrhoeae, which causes gonorrhea in humans, is a global health threat according to the WHO. The East Asia and Southeast Asia regions have been the main sources of emergence and spread of these bacterial strains in recent decades, including Vietnam. However, studies on the genetic characteristics of this bacteria in Vietnam are still very limited. Most studies in Vietnam on this bacteria in the past five years have focused on researching its antibiotic resistance trends using microbiological techniques, with very few studies analyzing the genetic characteristics of N. gonorrhoeae. Therefore, we have developed a protocol comprising four essential components to analyze the genetic characteristics of 243 N.gonorrhoeae raw sequences in Vietnam, including: genome assembly, gene prediction and gene annotation; prediction of virulence genes and antibiotic resistance genes; determination of ST and phylogenetic tree construction based on housekeeping genes; pangenome analysis. The results obtained 241/243 assembled and annotated N. gonorrhoeae genomes (genome size 2.074-2.307Mb; number of genes 2,024-3,112; GC content 52.27%-52.7%) with high similarity to reference genome ASM1303007v1 (genome size 2.2Mb; number of genes 2,283; GC content 52.5%). Additionally, 43 virulence genes and 11 antibiotic resistance genes were identified. Among them, 39/43 virulence genes and 7/11 antibiotic resistance genes were present in all assembled genomes. P¹angenome analysis results include: 1600/4270 (37.47%) core genes, 1925/4270 (45.08%) accessory genes, and 745/4270 (17.45%) unique genes. The ST grouping results showed that 240/241 samples belonged to 41 different ST groups, with ST7371 being the most prevalent (11.11%). The phylogenetic tree construction results revealed two distinct clades, with clade 1 comprising 117 genomes and clade 2 comprising 123 genomes.

Keywords: Phylogenetic tree, genomic characteristics, Neisseria gonorrhoeae, bioinformatics, antibiotic resistance.

^{*} Author for correspondence: Tel: +84-916110333; Email: ncuong@ioit.ac.vn

NGHIÊN CỨU ĐA DẠNG DI TRUYỀN CỦA *Pseudomonas aeruginosa* TẠI VIỆT NAM THÔNG QUA PHÂN TÍCH PAN GENOME, CORE GENOME MLST VÀ CORE GENOME SNP

Vương Thị Hương¹, Nguyễn Tiến Đạt¹, Trịnh Thị Xuân², Nguyễn Cường^{3*}

¹Công ty TNHH LOBI Việt Nam

²Khoa Công nghệ thông tin, Đại học Mở Hà Nội

³Viện Công nghệ thông tin, Viện Hàn lâm Khoa học và Công nghệ Việt Nam

TÓM TẮT

Pseudomonas aeruginosa, một vi khuẩn có khả năng thích nghi cao, là một trong các nguyên nhân chính gây nhiễm trùng bênh viên ở Việt Nam với sự phổ biến của các chủng kháng carbapenem, đặc biệt là ST235. Kiểm soát nhiễm trùng hiệu quả đòi hỏi phải hiểu rõ chi tiết về sự lây truyền và mô hình kháng thuốc của nó. Nghiên cứu này nhằm phân tích pan genome, core genome MLST (cgMLST) và core genome SNP (cgSNP) của các chủng P. aeruginosa được phân lập ở Việt Nam, so sánh và tích hợp các phương pháp này để nâng cao độ phân giải dịch tễ học. Chúng tôi đã kiểm tra 196 bộ gene của P. aeruginosa được phân lập ở Việt Nam và 956 hệ gene hoàn chỉnh từ cơ sở dữ liệu NCBI. Định dạng chủng ban đầu được thực hiện bằng cách sử dụng MLST. Phân tích pan genome được thực hiện bằng Roary, trong khi các sơ đồ cgMLST được thiết lập với ChewBBACA và phân tích SNP của core genome được thực hiện bằng Parsnp. Sau đó, cây phát sinh loài được xây dựng và hiển thị bằng GrapeTree và iTOL. Phân tích xác định 334 sequence type (ST) đã biết từ cơ sở dữ liệu NCBI, bao gồm 127 ST mới, và 25 ST đã biết từ bộ dữ liệu Việt Nam, với 8 ST mới. ST235 là phổ biến nhất, chiếm 77,04% mẫu Việt Nam. Phân tích bộ gene toàn phần tiết lộ 68.963 gene, với 4,16% là các gen lõi. Sơ đồ cgMLST bao gồm 3.289 locus, cung cấp đô phân giải cao hơn so với MLST truyền thống. Phân tích phát sinh loài trên các chủng ST235 cho thấy các mô hình phân cụm rõ ràng, thể hiện khả năng phân biệt vượt trội của cgMLST và phân tích cgSNP. Nghiên cứu này làm nổi bât hiêu quả của các công cu genome tiên tiến trong phân loai và so sánh các chủng P. aeruginosa, góp phần nâng cao hiểu biết về đa dang di truyền và hỗ trợ phát triển các biên pháp kiểm soát nhiễm trùng và kháng thuốc của P. aeruginosa hiệu quả tại Việt Nam.

Từ khóa: Core genome MLST, core genome SNP, đa dạng di truyền, pan genome, Pseudomonas aeruginosa, Việt Nam.

MỞ ĐẦU

Pseudomonas aeruginosa là một vi khuẩn có khả năng biến đổi di truyền và chuyển hóa vượt trội, cho phép nó sống sót và phát triển trong nhiều môi trường khác nhau. Khả năng thích nghi linh hoạt của *P. aeruginosa* khiến nó trở thành một trong những tác nhân gây bệnh cơ hội nguy hiểm nhất, được xếp vào nhóm các tác nhân gây bệnh quan trọng lâm sàng ESKAPE (Rice, 2008) (bao gồm *Enterococcus faecium, Staphylococcus aureus, Klebsiella pneumoniae, Acinetobacter baumannii, P. aeruginosa* và *Enterobacter* spp.). Ở người, *P. aeruginosa* là nguyên nhân gây ra nhiều loại nhiễm trùng trong cộng đồng và bệnh viện, từ nhiễm trùng da và mô mềm đến viêm phổi và nhiễm trùng máu phức tạp (Driscoll *et al.,* 2007). Đặc biệt, *P. aeruginosa* là tác nhân gây nhiễm trùng chủ yếu ở bệnh nhân xơ nang và là nguồn gốc chính của nhiễm trùng ở vết bỏng, dẫn đến tỷ lệ mắc bệnh và tử vong đáng kể (Kerr and Snelling, 2009).

Khu vực Đông Nam Á được coi là một "điểm nóng" về kháng kháng sinh, và *P. aeruginosa* đã được xác định là nguyên nhân phổ biến gây nhiễm trùng bệnh viện ở Việt Nam. Theo báo cáo giám sát kháng kháng sinh của Bộ Y tế Việt Nam năm 2020 (công bố tháng 11/2023), dữ liệu từ 16 bệnh viện tại 10 tỉnh cho thấy trong tổng số 69.715 chủng vi khuẩn được thu thập, *P. aeruginosa* đứng thứ năm với tỷ lệ 9,4% (6.564 chủng), là nguyên nhân hàng thứ ba khi phân lập bệnh phẩm từ đường hô hấp dưới (12,9% - 3.087 chủng), bệnh phẩm nước tiểu (8,5% - 826 chủng) và bệnh phẩm dịch ổ bụng (7,7% - 341 chủng). Theo dữ liệu từ Trung tâm Động lực học Bệnh tật, Kinh tế & Chính sách (CDDEP) năm 2016, 36% các chủng *P. aeruginosa* ở Việt Nam kháng carbapenems, đứng thứ hai toàn cầu chỉ sau Ấn Độ về tỷ lệ kháng thuốc. Một nghiên cứu từ năm 2016 tại một bệnh viện ở Hà Nội đã ghi nhận sự xuất hiện của các chủng *P. aeruginosa* thuộc kiểu ST235 mang các gen kháng carbapenemase như blaIMP-15, blaIMP-26, và blaIMP-51 (Tada *et al.*, 2016).

Để nhanh chóng và hiệu quả triển khai các biện pháp kiểm soát trong quản lý đợt bùng phát, việc hiểu rõ sự lây truyền của các tác nhân kháng thuốc là điều cần thiết. Việc xác định các cụm và phân loại các tác nhân gây

bệnh, bao gồm các đặc điểm kháng thuốc và gen độc lực, đóng vai trò quan trọng trong việc kiểm soát nhiễm trùng một cách hiệu quả (Leopold *et al.*, 2014). Điện di gel với trường điện từ (Pulsed-field gel electrophoresis-PFGE) và phân loại trình tự nhiều locus (multilocus sequence typing-MLST) đã được sử dụng như các tiêu chuẩn vàng trong nhiều năm để xác nhận các đợt bùng phát nghi ngờ. Tuy nhiên, PFGE thường gặp vấn đề về tính chủ quan, khó khăn trong việc giải thích kết quả và tốn thời gian. Ngược lại, MLST, được phát triển vào năm 2004 cho *P. aeruginosa*, dựa trên việc giải trình tự và đánh giá sự biến đổi alen của bảy gen bảo tồn, tạo ra các kiểu trình tự (sequence type - ST) để đặc trưng hóa các chủng (Curran *et al.*, 2004). Tuy nhiên, khả năng phân biệt của MLST không phải lúc nào cũng đủ để giải quyết các đợt bùng phát.

Gần đây, các phương pháp dựa trên giải trình tự toàn bộ hệ gen (WGS) đã trở nên khả thi nhờ chi phí giải trình tự giảm; cùng với các công cụ tin sinh học cho phép phân biệt cao giữa các chủng có cùng ST trong các nghiên cứu dịch tễ học (Tang *et al.*, 2017) như pan genome, core genome MLST, core genome SNP, ... Pan genome là tập hợp toàn bộ các gen trong một loài vi khuẩn, bao gồm cả gen lõi và gen phụ thuộc vào từng chủng. Phân tích Pan genome xác định gen lõi (có mặt trong tất cả các chủng) và gen phụ (chỉ có trong một số chủng), cung cấp cái nhìn về sự đa dạng di truyền và các yếu tố ảnh hưởng đến khả năng gây bệnh và kháng thuốc. Core Genome MLST (cgMLST) mở rộng MLST bằng cách sử dụng dữ liệu giải trình tự toàn bộ genome (WGS) để kiểm tra nhiều gen đích, tạo ra hệ thống đánh số alen có hệ thống, giúp phân giải cao hơn trong phát hiện và phân loại các đợt bùng phát. Phân tích cgSNP tập trung vào các SNPs trong các vùng bảo tồn của bộ gen lõi, xác định mối quan hệ tiến hóa và theo dõi sự lây lan của các chủng vi khuẩn, phát hiện khác biệt di truyền nhỏ và cung cấp cái nhìn chi tiết về sự phát tán của tác nhân gây bệnh.

Nghiên cứu này nhằm phân tích pan genome, core genome MLST và core genome SNP cho các chủng *P. aeruginosa* phân lập được tại Việt Nam. So sánh và kết hợp các phương pháp này có thể đạt được độ phân giải tối đa cho các phân tích dịch tễ học và giám sát *P. aeruginosa* (Zhou *et al.*, 2017), hỗ trợ xác định và theo dõi sự lây truyền, tiến hóa và kháng thuốc của các chủng vi khuẩn *P. aeruginosa* tại Việt Nam.

DỮ LIỆU VÀ PHƯƠNG PHÁP

Dữ liệu

Dữ liệu giải trình tự của các chủng phân lập tại Việt Nam

Các mã Sequence Read Archive (SRA) của các mẫu *P. aeruginosa* phân lập tại Việt Nam được lưu trữ trên cơ sở dữ liệu National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/). Tương ứng với các mã SRA này, 307 trình tự đọc đã được truy xuất. Chúng đều được phân lập từ bệnh phẩm của người trong khoảng thời gian từ năm 2013 đến năm 2019, thuộc bốn dự án PRJEB29424, PRJEB28400, PRJDB4025, và PRJDB2736. Các trình tự này đã được tinh sạch và lắp ráp theo các bước sau: 1. Đánh giá chất lượng dữ liệu ban đầu bằng FastQC với thông số mặc định; 2. Tinh sạch dữ liệu bằng Trimmomatic với các thông số: SLIDING WINDOW=25:4, MINLEN=100, ILLUMINACLIP=ON; 3. Đánh giá chất lượng dữ liệu sau tinh sạch bằng FastQC với thông số mặc định; 6. Đánh giá độ toàn vẹn hệ gen bằng BUSCO với thông số auto-lineage-prok; 7. Định danh taxonomy cho hệ gen lắp ráp bằng Kraken2 với database Minikraken2; 8. Lọc bỏ các contigs có độ dài nhỏ hơn 500bp. Các hệ gen sau lắp ráp thỏa mãn các điều kiện sau sẽ được đưa vào các phân tích phylogeny tiếp theo: kết quả định danh Loài (Species -S) gần nhất với *P. aeruginosa* và %S >=90%, độ hoàn thiện hệ gen >=90%, kích thước hệ gen sau lọc <=7,5M bp và số lượng contigs sau lọc <=400 (gọi tắt là tập dữ liệu VN).

Dữ liệu các hệ gen hoàn chỉnh của Pseudomonas aeruginosa có sẵn trên NCBI

Để phục vụ cho phân tích cgMLST và tìm hiểu sự liên hệ của các loài phân lập ở Việt Nam với các loài trên thế giới, tất cả 956 trình tự bộ gen hoàn chỉnh (tệp fasta) của *P. aeruginosa* có sẵn công khai tại cơ sở dữ liệu NCBI Reference Sequence (RefSeq) tính đến ngày 20 tháng 6 năm 2024 (gọi tắt là tập dữ liệu NCBI) đã được đưa vào các phân tích xây dựng cây phát sinh loài, cùng với các hệ gen đã được lắp ráp của *P. aeruginosa*. Thống kê cho thấy phần lớn các hệ gen này có từ 0 (738, 77,2%) hoặc 1 plasmid (156, 16,0%). Các hệ gen này được phân lập và lắp ráp trong khoảng thời gian từ năm 2006 đến năm 2024, với phần lớn được thực hiện từ năm 2018 đến 2024 (848, 88,7%). Các hệ gen *P. aeruginosa* này chủ yếu đến từ Trung Quốc (239, 25,0%), Mỹ (151, 15,8%), và Đan Mạch (87, 9,1%), và đa số được phân lập từ mẫu bệnh phẩm của người (767, 80,2%). Có duy nhất 1 hệ gen được phân lập tại Việt Nam nhưng từ mẫu đất.

Phương pháp

Phân tích MLST

956 hệ gen hoàn chỉnh được tải về từ cơ sở dữ liệu genome NCBI và 196 hệ gen được lắp ráp từ các trình tự của các mẫu phân lập tại Việt Nam đã được quét đối chiếu với các sơ đồ phân loại truyền thống của PubMLST (https://pubmlst.org/) bằng công cụ mlst. Các ST mới được gán khi không tìm thấy ST tương ứng trong cơ sở dữ liệu PubMLST trong khuôn khổ của nghiên cứu này.

Phân tích Pan Genome

956 hệ gen hoàn chỉnh và 196 hệ gen lắp ráp được chú giải bằng công cụ Prokka. Các tập tin GFF3 được tạo ra bởi Prokka được sử dụng để thực hiện phân tích pan genome bằng công cụ Roary (phiên bản 3.11.2). Theo đó, các gen được phân loại thành bốn nhóm khác nhau: nhóm "cốt lõi" (99% \leq số chủng \leq 100%), nhóm "cốt lõi mềm" (95% \leq số chủng < 99%), nhóm "vỏ" (15% \leq số chủng < 95%) và nhóm "mây" (0% \leq số chủng < 15%). Cây pan genome phản ánh mối quan hệ tiến hóa giữa các chủng *P. aeruginosa* dựa trên sự tương đồng và khác biệt trong toàn bộ gen của chúng, được xây dựng từ dữ liệu sự hiện diện và vắng mặt của các gen và trực quan hóa bằng phần mềm Phandango (Phandango.net).

Phân tích core genome MLST

Sơ đồ cgMLST của *P. aeruginosa* (scheme) được thiết lập với công cụ Comprehensive and High Efficient Workflow for a Blast Score Ratio Based Allele Calling Algorithm (ChewBBACA) (Silva *et al.*, 2018). Trong đó, trình tự genome của *P. aeruginosa* PAO1 (RefSeq assembly accession: GCF_000006765.1) chỉ được sử dụng làm genome tham chiếu để dự đoán các locus whole-genome MLST (wgMLST). Chúng tôi đã chọn các locus ứng viên cho sơ đồ cgMLST có mặt trong 99% các hệ gen hoàn chỉnh hiện có (956 hệ gen). Sau đó, 196 hệ gen lắp ráp từ các mẫu được phân lập tại Việt Nam đã được gọi alen theo sơ đồ cgMLST đã xác định. Cây phân loài cgMLST được xây dựng bằng phần mềm GrapeTree (phiên bản 1.5.0) (Zhou *et al.*, 2018) với các thông số được triển khai trong MSTree v2, và được trực quan hóa bằng phần mềm Interactive Tree Of Life (iTOL, phiên bản 4.2.3) (Letunic and Bork, 2016). Qua đó, khả năng phân loại các mẫu trình tự của *P. aeruginosa* bằng cgMLST và MLST được nhận xét và so sánh.

So sánh phân tích core genome MLST với core genome SNP cho các mẫu Pseudomonas aeruginosa ST235

Tất cả các hệ gen *P. aeruginosa* ST235 của Việt Nam và các quốc gia khác trong nghiên cứu này sẽ được chọn. Một cây phát sinh loài dựa trên các cgMLST đã được xây dựng bằng phần mềm GrapeTree với các thông số được triển khai trong NJ. Đối với phân tích SNP dựa trên bộ gen lõi, bộ gen lõi được căn chỉnh bằng Parsnp, là một phần của gói phần mềm Harvest, sử dụng NCGM2.S1 (RefSeq assembly accession: GCF_000284555.1) làm bộ gen tham chiếu. Cả hai cây cùng được trực quan hóa bằng iTOL. Sự tương đồng giữa hai phương pháp, cgMLST và phân tích SNP dựa trên bộ gen lõi, sẽ được thảo luận dựa trên những điểm tương đồng và khác biệt trong việc phân cụm của hai cây phát sinh loài.

KẾT QUẢ VÀ THẢO LUẬN

Lắp ráp các chủng phân lập tại Việt Nam

Sau khi áp dụng các tiêu chuẩn loại trừ, 196 hệ gen lắp ráp (PRJEB29424: 110; PRJEB28400: 37; PRJDB2736: 26; PRJDB4025: 23) được chọn để tiến hành các phân tích tiếp theo. Các hệ gen này có kích thước từ 6,39 Mbp đến 7,45 Mbp và đạt mức độ hoàn thiện rất cao, với giá trị trung vị đạt 100% và các giá trị dao động từ 96,8% đến 100%. Tỷ lệ phần trăm tương đồng với *P. aeruginosa* (%S_*P. aeruginosa*) có giá trị trung bình khoảng 94,36%. Các hệ gen có chất lượng lắp ráp tốt, với số lượng contigs dao động từ 35 đến 390, đảm bảo chất lượng cao cho các phân tích cây phát sinh loài tiếp theo.

Xác định MLST

Trong tập dữ liệu từ NCBI, tổng số 334 ST được xác định, trong đó 207 ST đã được ghi nhận trong cơ sở dữ liệu pubMLST và 127 ST là các ST mới. Các ST đã biết xuất hiện nhiều nhất trong tập dữ liệu này bao gồm ST235 (54 mẫu), ST111 (43 mẫu), ST463 (29 mẫu), ST309 (18 mẫu), ST357 (17 mẫu), ST253 (17 mẫu), ST277 (15 mẫu), ST233 (13 mẫu), ST395 (13 mẫu)... Đối với tập dữ liệu VN (**Bảng 1**), có 25 ST được xác định, trong đó 17 ST đã được ghi nhận và 8 ST là các ST mới. Tương tự như tập dữ liệu NCBI, ST235 là ST phổ biến nhất với 151 mẫu được định danh, chiếm 77.04% tổng số mẫu. Tiếp theo là ST357 với 10 mẫu, chiếm 19.60%. Các ST khác như ST644, ST360, ST310, ST1650, ST2332,... có số lượng mẫu dao động từ 1-4. Các ST mới chỉ xuất hiện ở mức thấp, với số lượng mẫu dao động từ 1-2 cho mỗi ST mới.

	-							-		
ST	Sheme	acsA	aroE	guaA	mutL	nuoD	ppsA	trpE	NCBI	VN
235	paeruginosa	38	11	3	13	1	2	4	54	151
357	paeruginosa	2	4	5	3	1	6	11	17	10
644	paeruginosa	28	3	94	13	1	4	10	3	4
360	paeruginosa	15	5	36	11	27	4	2	2	4
310	paeruginosa	5	59	60	3	1	6	4	0	4

Bảng 1. Kết quả định danh MLST của 1.152 hệ gen Pseudomonas aeruginosa

1650	paeruginosa	16	5	26	3	4	15	8	0	2
2332	paeruginosa	15	5	91	11	3	4	2	0	2
new_1	paeruginosa	38	~11	3	13	1	2	4	0	2
277	paeruginosa	39	5	9	11	27	5	2	15	1
773	paeruginosa	5	4	5	5	5	7	8	11	1
1971	paeruginosa	32	190	3	62	8	7	26	8	1
1212	paeruginosa	11	10	11	72	3	10	3	4	1
664	paeruginosa	9	5	11	3	4	40	18	2	1
2165	paeruginosa	87	42	114	37	53	97	147	1	1
830	paeruginosa	5	13	109	5	1	1	47	0	1
1410	paeruginosa	13	158	5	5	12	7	15	0	1
1649	paeruginosa	16	5	12	77	11	4	18	0	1
3359	paeruginosa	47	50	65	31	1	6	8	0	1
new_2	paeruginosa	38	~11	3	~13	1	2	4	0	1
new_3	paeruginosa	9	~5	~11	~21	4	40	18	0	1
new_4	paeruginosa	39	5	9	11	~27	5	2	0	1
new_5	paeruginosa	27	~14	25	~23	1	16	~46	0	1
new_6	paeruginosa	~5	4	~5	~5	5	7	8	0	1
new_7	paeruginosa	~28	3	~94	~13	1	4	~10	0	1
new_8	paeruginosa	11	~6	11	13	2	7	53	0	1

Phân tích Pan genome từ Roary

Phân tích Roary của pan genome của tổng số 1.152 chủng *P. aeruginosa* đã xác định được 2.868 gen lõi (4,16%), 1.619 gen lõi mềm (2,45%), 2.330 gen vỏ (3,38%) và 62.146 gen đám mây (90,01%), trong tổng số 68.963 gen. **Hình 1** (bên trái) cho thấy các hệ gen từ các mẫu phân lập tại Việt Nam (màu tím) có số lượng các gen phụ nhiều hơn so với các hệ gen hoàn thiện đã được công bố trên NCBI. Đặc biệt, có thể dễ dàng thấy tồn tại một nhánh nhỏ các hệ gen trong tập dữ liệu VN nằm trong số 1% không chứa đầy đủ một số lượng khá lớn các gen lõi nhưng lại cùng có một số lượng gen phụ đặc trưng ở vị trí khoảng 551,7kb. **Hình 1** (bên phải) cho thấy số lượng gen lõi được cân bằng khi số lượng hệ gen tăng cho thấy rằng các gen lõi là cần thiết và bảo tồn cao giữa các chủng khác nhau, phản ánh tính ổn định và thiết yếu của chúng trong bộ gen. Số lượng lớn các gen đám mây cho thấy có sự dị biệt lớn giữa tổng số 1.152 chủng *P. aeruginosa* được xem xét, nhấn mạnh tính chất 'mở' của pan genome *P. aeruginosa*.



Hình 1. Pan genome của Pseudomonas aeruginosa

Bên trái: cây phân loài và biểu đồ ma trận (bên trái) dựa trên sự hiện diện hay vắng mặt của các gen lõi và gen phụ cùng với thông tin về sequence type (ST) và tập dữ liệu (NCBI: màu cam, VN: màu tím); Bên phải: đường màu đỏ thể hiện tổng số gen và đường màu xanh thể hiện số gen bảo tồn (gen lõi) khi số lượng hệ gen Pseudomonas aeruginosa tăng lên.

Phân tích core genome MLST

Một sơ đồ cgMLST bao gồm 3.289 gen đích đã được xác định, bao phủ 73,58% trong số 5.570 ORF được dự đoán cho chủng tham chiếu PAO1. Kết quả này khá đồng nhất với kết quả tổng có 3.168 gen đích trong nghiên cứu phân tích cgMLST của *P. aeruginosa* của Romário Oliveira de Sales và cộng sự. Theo đó, 196 chủng phân lập tại Việt Nam được xác định sơ đồ cgMLST theo bộ gen đích trên. Kết quả của bộ sơ đồ cgMLST của 1.152 chủng *P. aeruginosa* sau đó được sử dụng để xây dựng cây phân loài (**Hình 2**). Tính từ trong ra ngoài, trong cùng là cây phân loài được xây dựng từ kết quả cgMLST. Vòng tròn đầu tiên phân biệt giữa các chủng được phân lập tại Việt Nam (màu xanh lá) và trên thế giới (màu tím). Vòng tròn thứ hai phân loại cụ thể nguồn gốc quốc gia của các chủng *P. aeruginosa*. Vòng tròn thứ ba thể hiện kích thước hệ gen của từng chủng. **Hình 2** đã cho thấy được sự vượt trội về khả năng phân loại các chủng *P. aeruginosa* và thể hiện chi tiết hơn về mối quan hệ di truyền giữa các chủng này của phân tích cgMLST so với phương pháp truyền thống MLST.



Hình 2. Cây phát sinh loài dựa vào phân tích cgMLST

Tính từ trong ra ngoài: Tập dữ liệu (xanh lá: VN, tím: NCBI); Quốc gia phân lập; Kích thước hệ gen; Sequence Type

So sánh phân tích cgMLST và cgSNP cho các chủng Pseudomonas aeruginosa ST235

Nghiên cứu sự đa dạng của *P. aeruginosa* ST235 ở Việt Nam là quan trọng vì đây là nhóm vi khuẩn đa kháng phổ biến tại Việt Nam (chiếm tới 77,14% tổng số mẫu trong nghiên cứu này), đòi hỏi sự quan tâm đặc biệt để hỗ trợ dịch tễ học và phát triển chiến lược kiểm soát. ST235 thể hiện sự thay đổi không đáng kể trong kiểu gen qua một thập kỷ, như được ghi nhận trong nghiên cứu của Wei Feng và cộng sự, cho phép khảo sát các phương pháp phân loại có độ phân giải cao hơn cấp độ gen như cgMLST và cgSNP, giúp đánh giá chính xác biến đổi di truyền và hiệu quả của các công cụ phân tích này.

Hình 3A và **Hình 3B** lần lượt là cây phân loài được vẽ từ phân tích cgMLST và cgSNP cho 205 chủng *P. aeruginosa* ST235 được xác định trong nghiên cứu của chúng tôi. Ở cả hai hình, chúng ta có thể thấy được sự tương đồng của đa số các chủng ở Việt Nam, và chúng có sự khác biệt với các chủng trên thế giới. Đặc biệt, chủng có ID DRR021817 (Việt Nam, 2013) có khoảng cách xa hẳn các chủng còn lại; các chủng còn lại này đều được phân lập năm 2013 và chúng tách thành hai nhóm rõ rệt ở cả hai hình. Tuy nhiên, vẫn có sự khác biệt giữa hai phương pháp phân loại này. Cụ thể, ở cây phân loài cgMLST, chúng tôi phát hiện hai nhóm trên lần lượt có mối quan hệ gần nhất với hai chủng được phân lập năm 2023 ở châu Á (Thái Lan - 2023, Trung Quốc - 2023); trong khi đó, ở cây phân loài cgSNP, hai nhóm này lần lượt có độ tương đồng cao nhất lần lượt với một chủng ở Trung Quốc năm 2023 và một chủng ở Ý năm 2024.



Hình 3. Cây phát sinh loài cho các chủng Pseudomonas aeruginosa ST235

A. Cây từ phân tích cgMLST; B. Cẩy từ cgSNP. Tính từ trong ra ngoài, trong cùng là cây phân loài với nhãn màu tím thuộc dữ liệu NCBI và nhãn màu xanh thuộc tập dữ liệu VN; vòng ở trong thể hiện vị trí địa lý; vòng ngoài cùng thể hiện năm phân lập của các chủng Pseudomonas aeruginosa ST235

Nghiên cứu này có một vài hạn chế. Thứ nhất, chỉ các hệ gen hoàn chỉnh có sẵn trên cơ sở dữ liệu NCBI được đưa vào đồng phân tích cây phát sinh loài, do đó thiếu các dữ liệu để kết luận về mối quan hệ của các hệ gen chưa hoàn thiện thuộc các ST được quan tâm và công bố ở thời điểm khác. Thứ hai, dựa trên kết quả tìm kiếm trên NCBI, chỉ có dữ liệu của các chủng *P. aeruginosa* Việt Nam được phân lập vào các năm 2013, 2014, 2018, và 2019, dẫn đến thiếu hụt dữ liệu hệ gen của vi khuẩn này trong những năm gần đây để đánh giá và nhận xét về mức độ phát triển của chúng. Bên cạnh ST235, ST357 cũng nằm trong top 10 sequence type đáng quan tâm về vấn đề kháng thuốc nhưng chưa được tìm hiểu sâu hơn về mức độ đa dạng. Trong các nghiên cứu tương lai, chúng tôi sẽ cố gắng thu thập dữ liệu và giải trình tự các chủng *P. aeruginosa* mới nhất, cũng như khám phá mức độ phân loại của các chủng sequence type khác để bổ sung vào hiểu biết hiện có.

KÉT LUẬN

Nghiên cứu này đã chứng minh được mức độ phân giải cao của các công cụ được khảo sát trong việc phân loại các chủng *P. aeruginosa* ở Việt Nam và so sánh chúng với các chủng trên thế giới. Các kết quả của nghiên cứu này sẽ góp phần vào việc nâng cao hiểu biết về sự đa dạng di truyền và sự lan truyền của *P. aeruginosa* nói riêng và các chủng vi khuẩn đa kháng nói chung, từ đó hỗ trợ phát triển các biện pháp kiểm soát và phòng ngừa hiệu quả hơn, từ đó bảo vệ sức khỏe cộng đồng trong tình hình kháng thuốc dữ dội tại Việt Nam.

TÀI LIỆU THAM KHẢO

Curran B, Jonas D, Grundmann H, Pitt T, Dowson CG, 2004. Development of a Multilocus Sequence Typing Scheme for the Opportunistic Pathogen *Pseudomonas aeruginosa. J Clin Microbiol* 42, 5644-5649. https://doi.org/10.1128/JCM.42.12.5644-5649.2004

Driscoll JA, Brody SL, Kollef MH, 2007. The Epidemiology, Pathogenesis and Treatment of Pseudomonas aeruginosa Infections. *Drugs* 67, 351-368. https://doi.org/10.2165/00003495-200767030-00003

Kerr KG, Snelling AM, 2009. *Pseudomonas aeruginosa*: a formidable and ever-present adversary. *J Hospital Infect*, Proceedings of The Lancet Conference on Healthcare-Associated Infections 73, 338-344. https://doi.org/10.1016/j.jhin.2009.04.020

Leopold SR, Goering RV, Witten A, Harmsen D, Mellmann A, 2014. Bacterial Whole-Genome Sequencing Revisited: Portable, Scalable, and Standardized Analysis for Typing and Detection of Virulence and Antibiotic Resistance Genes. *J Clin Microbiol* 52, 2365–2370. https://doi.org/10.1128/JCM.00262-14

Letunic I, Bork P, 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44, W242-W245. https://doi.org/10.1093/nar/gkw290

Rice LB, 2008. Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE. J Infect Dis 197, 1079-1081. https://doi.org/10.1086/533452

Silva M, Machado MP, Silva DN, Rossi M, Moran-Gilad J, Santos S, Ramirez M, Carriço JA, 2018. chewBBACA: A complete suite for gene-by-gene schema creation and strain identification. *Microb Genom* 4, e000166. https://doi.org/10.1099/mgen.0.000166

Tada T, Nhung PH, Miyoshi-Akiyama T, Shimada K, Tsuchiya M, Phuong DM, Anh NQ, Ohmagari N, Kirikae T, 2016. Multidrug-Resistant Sequence Type 235 Pseudomonas aeruginosa Clinical Isolates Producing IMP-26 with Increased Carbapenem-Hydrolyzing Activities in Vietnam. *Antimicrob Agents Chemother* 60, 6853-6858. https://doi.org/10.1128/AAC.01177-16

Tang P, Croxen MA, Hasan MR, Hsiao WWL, Hoang LM, 2017. Infection control in the new age of genomic epidemiology. *Amer J Infect Cont* 45, 170-179. https://doi.org/10.1016/j.ajic.2016.05.015

Zhou H, Liu W, Qin T, Liu C, Ren H, 2017. Defining and Evaluating a Core Genome Multilocus Sequence Typing Scheme for Whole-Genome Sequence-Based Typing of *Klebsiella pneumoniae. Front Microbiol* 8, 371. https://doi.org/10.3389/fmicb.2017.00371

Zhou Z, Alikhan NF, Sergeant MJ, Luhmann N, Vaz C, Francisco AP, Carriço JA, Achtman M, 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. *Genome Res* 28, 1395–1404. https://doi.org/10.1101/gr.232397.117

STUDY ON THE GENETIC DIVERSITY OF *Pseudomonas aeruginosa* IN VIETNAM THROUGH PAN-GENOME, CORE GENOME MLST, AND CORE GENOME SNP ANALYSES

Vuong Thi Huong¹, Nguyen Tien Dat¹, Trinh Thi Xuan², Nguyen Cuong^{3*}

¹LOBI Vietnam Ltd

²Faculity of Information Technology, Hanoi Open University

³Institute of Information Technology, Vietnam Academy of Science and Technology

SUMMARY

Pseudomonas aeruginosa, a highly adaptable pathogen, is one of the major causes of hospital infections in Vietnam, particularly with the prevalence of carbapenem-resistant strains, notably ST235. Effective infection control requires a detailed understanding of its transmission and resistance patterns. This study aims to analyze the pan-genome, core genome MLST (cgMLST), and core genome SNP (cgSNP) of P. aeruginosa strains isolated in Vietnam, comparing and integrating these methods to enhance epidemiological resolution. We examined 196 P. aeruginosa gene profiles from Vietnam and 956 complete genome sequences from the NCBI database. Initial strain typing was performed using MLST. Pan-genome analysis was conducted with Roary, while cgMLST schemas were developed with ChewBBACA, and core genome SNP analysis was performed using Parsnp. Phylogenetic trees were constructed and visualized using GrapeTree and iTOL. The analysis identified 334 known sequence types (ST) from the NCBI database, including 127 new STs, and 25 known STs from the Vietnamese dataset, with 8 new STs. ST235 was the most prevalent, accounting for 77.04% of the Vietnamese samples. The pan-genome analysis revealed 68,963 genes, with 4.16% being core genes. The cgMLST schema included 3,289 loci, providing higher resolution compared to traditional MLST. Phylogenetic analysis of ST235 strains demonstrated clear clustering patterns, showcasing the superior discriminative ability of cgMLST and cgSNP analysis. This study highlights the effectiveness of advanced genomic tools in classifying and comparing *P. aeruginosa* strains, enhancing understanding of genetic diversity, and supporting the development of effective infection control and resistance management strategies in Vietnam.

Keywords: Core genome MLST, core genome SNP, genetic diversity, pan-genome, Pseudomonas aeruginosa, Vietnam.

^{*} Author for correspondence: Tel: 0916110333; Email: ncuong@ioit.ac.vn