

PHÂN TÍCH MÃ VẠCH ADN CỦA NGUỒN GEN LÚA ĐỊA PHƯƠNG VIỆT NAM DỰA TRÊN TRÌNH TỰ GEN *rbcl*

Lã Tuấn Nghĩa¹, Lê Thị Thu Trang^{1,*}, Nguyễn Thanh Nhung², Đỗ Hà Thu¹,
Lã Hoàng Nhật Minh¹, Đàm Thị Thu Hà¹, Nguyễn Hoài Thu¹, Tạ Diệp Viễn Phương¹

¹ Trung tâm Tài nguyên thực vật

² Viện Di truyền Nông nghiệp

*Email: lethutrang2810@gmail.com

TÓM TẮT

Nghiên cứu này được thực hiện nhằm xác định đặc điểm mã vạch ADN vùng gen *rbcl* của 180 mẫu giống lúa địa phương thu thập từ các vùng sinh thái khác nhau tại Việt Nam. Sau khi khuếch đại bằng PCR, các sản phẩm gen *rbcl* được giải trình tự bằng phương pháp Sanger, dữ liệu trình tự được xử lý và phân tích bằng các phần mềm BioEdit và MEGA 7. Kết quả cho thấy, chiều dài trình tự dao động từ 541 - 551 bp. Thành phần nucleotide trung bình gồm: 26,67% A, 29,13% T, 23,24% G và 20,96% C. So sánh với cơ sở dữ liệu GenBank bằng công cụ BLAST ghi nhận mức độ tương đồng từ 99,45 - 99,82% với độ bao phủ đạt 100%, cho thấy vùng gen *rbcl* có tính bảo tồn cao ở cây lúa. Phân tích đa hình trình tự phát hiện 10 vị trí biến dị, bao gồm 4 SNP và 6 InDel. Các biến dị này hình thành các haplotype đặc trưng, cho phép nhận diện và phân biệt một số giống lúa như: Nàng Keo Sim, Nàng Hoa, Nghêu Tan Oi, Tép An Giang và Cẩm. Khoảng cách di truyền theo mô hình Kimura-2-parameter (K2P) dao động từ 0,0000 - 0,0037. Phân tích phát sinh loài bằng phương pháp UPGMA đã chia 180 mẫu giống thành hai nhóm chính. Kết quả nghiên cứu cho thấy, chỉ thị *rbcl* có thể hỗ trợ đánh giá đa dạng và phân nhóm sơ bộ nguồn gen lúa địa phương; cần kết hợp thêm các vùng gen có độ đa hình cao hơn để nâng cao khả năng phân biệt giữa các mẫu giống có quan hệ di truyền gần.

Từ khóa: Đa dạng di truyền, lúa địa phương, mã vạch ADN, trình tự nucleotide, *rbcl*.

1. ĐẶT VẤN ĐỀ

Lúa (*Oryza sativa* L.) là cây lương thực chủ lực, giữ vai trò quan trọng trong đảm bảo an ninh lương thực và phát triển nông nghiệp ở Việt Nam. Nguồn gen lúa địa phương có mức độ đa dạng di truyền cao, là nguồn vật liệu quan trọng phục vụ công tác chọn tạo giống và nâng cao khả năng thích nghi với các điều kiện môi trường bất lợi. Nhiều giống bản địa mang các đặc tính quý như: Chất lượng gạo tốt, khả năng chống chịu sâu, bệnh, chịu mặn, chịu hạn và chịu ngập úng, góp phần quan trọng trong chiến lược phát triển nông nghiệp bền vững và thích ứng với biến đổi khí hậu [1], [2].

Mã vạch ADN (DNA barcoding) được xem là công cụ hiệu quả trong nhận diện loài, đánh giá đa dạng di truyền và phân tích quan hệ phát sinh loài ở thực vật. Trong số các chỉ thị thuộc hệ gen lục lạp, gen *rbcl* được sử dụng phổ biến nhờ khả năng khuếch đại ổn định, mức độ bảo tồn cao và

có nguồn dữ liệu tham chiếu phong phú trên các ngân hàng gen quốc tế [3]. Bên cạnh ưu điểm về tính ổn định và khả năng khuếch đại cao, *rbcl* còn là chỉ thị có chi phí phân tích tương đối thấp, phù hợp cho nghiên cứu sàng lọc số lượng mẫu lớn. Ngoài ra, do được sử dụng rộng rãi trong hệ thống mã vạch ADN thực vật toàn cầu, gen *rbcl* cho phép chuẩn hóa dữ liệu và thuận lợi trong so sánh với các cơ sở dữ liệu quốc tế, góp phần hỗ trợ truy xuất nguồn gốc và quản lý nguồn gen cây trồng. Sự phát triển của công nghệ giải trình tự ADN đã cung cấp nhiều thông tin quan trọng về cấu trúc di truyền và tiến hóa của các nhóm giống lúa [4].

Tại Việt Nam, các nghiên cứu về đa dạng di truyền lúa hiện nay chủ yếu tập trung vào các chỉ thị phân tử như SSR và SNP toàn hệ gen. Trong khi đó, dữ liệu mã vạch ADN, đặc biệt đối với gen *rbcl* vẫn còn hạn chế và chưa được xây dựng một cách hệ thống trên quy mô tập đoàn giống

lớn. Phần lớn nguồn gen lúa địa phương mới chỉ được đánh giá dựa trên đặc điểm hình thái và nông sinh học, trong khi dữ liệu phân tử còn thiếu hụt. Điều này gây khó khăn trong công tác định danh, phân tích quan hệ di truyền và khai thác hiệu quả nguồn vật liệu cho chọn giống và bảo tồn. Nghiên cứu này được thực hiện nhằm phân tích trình tự gen *rbcL* của 180 mẫu giống lúa địa phương Việt Nam, qua đó bổ sung dữ liệu mã vạch ADN và cung cấp cơ sở khoa học cho công tác bảo tồn, khai thác và sử dụng hiệu quả nguồn

gen lúa địa phương phục vụ mục tiêu phát triển giống lúa bền vững.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

2.1. Vật liệu nghiên cứu

Vật liệu nghiên cứu gồm 180 mẫu giống lúa địa phương được thu thập từ 8 vùng sinh thái nông nghiệp khác nhau của Việt Nam và hiện đang được lưu giữ, bảo tồn tại Ngân hàng gen cây trồng Quốc gia thuộc Trung tâm Tài nguyên thực vật (Bảng 1).

Bảng 1. Phân bố 180 mẫu giống lúa địa phương theo vùng sinh thái nông nghiệp ở Việt Nam

Bảng 1.1 Phân bố 10 mẫu gốc của địa phương theo vùng sinh thái nông nghiệp ở Việt Nam									
Vùng sinh thái nông nghiệp	Nguồn gốc mẫu	Số mẫu	Tổng	Vùng sinh thái nông nghiệp	Nguồn gốc mẫu	Số mẫu	Tổng		
Tây Bắc	Tỉnh Sơn La	6	23	Nam Trung Bộ	Tỉnh Quảng Nam	4	23		
	Tỉnh Điện Biên	6			Tỉnh Quảng Ngãi	4			
	Tỉnh Lai Châu	6			Tỉnh Bình Định	3			
	Tỉnh Hoà Bình	5			Tỉnh Phú Yên	1			
Đông Bắc	Tỉnh Hà Giang	3	33		Tỉnh Khánh Hòa	4		13	
	Tỉnh Cao Bằng	2			Tỉnh Ninh Thuận	3			
	Tỉnh Bắc Kạn	2			Tỉnh Bình Thuận	4			
	Tỉnh Lạng Sơn	4			Tây Nguyên	Tỉnh Kon Tum			3
	Tỉnh Tuyên Quang	1		Tỉnh Gia Lai		3			
	Tỉnh Thái Nguyên	2		Tỉnh Đắk Lắk		4			
	Tỉnh Phú Thọ	5		Tỉnh Đắk Nông		1			
	Tỉnh Yên Bái	4		Tỉnh Lâm Đồng		2			
	Đồng bằng sông Hồng	Tỉnh Lào Cai		7	Đông Nam Bộ	Thành phố Hồ Chí Minh	3		42
		Tỉnh Bắc Giang		3		Tỉnh Bà Rịa - Vũng Tàu	2		
Tỉnh Nam Định		4	Tỉnh Tây Ninh	2					
Tỉnh Thái Bình		2	Tỉnh Bình Phước	3					
Tỉnh Hải Dương		1	Tỉnh Long An	4					
Bắc Trung Bộ	Thành phố Hải Phòng	3	14	Tây Nam Bộ	Tỉnh An Giang	7	42		
	Tỉnh Quảng Ninh	4			Tỉnh Tiền Giang	1			
	Tỉnh Thanh Hóa	4			Tỉnh Bến Tre	7			
	Tỉnh Nghệ An	6			Tỉnh Trà Vinh	4			
	Tỉnh Hà Tĩnh	3			Tỉnh Sóc Trăng	4			
	Tỉnh Quảng Bình	5			Tỉnh Bạc Liêu	4			
	Tỉnh Quảng Trị	1			Tỉnh Cà Mau	5			
	Tỉnh Thừa Thiên Huế	3		Tỉnh Kiên Giang	6				

Ghi chú: Tên tỉnh/thành phố theo đơn vị hành chính tại thời điểm thu thập mẫu trước khi sáp nhập năm 2025.

- Cặp mồi *RbcL*-F/ *RbcL*-R (5' - ATG TCA CCA CAA ACA GAG ACT AAA GC - 3'/5' - GTAAAATCAAGTCCACCRG - 3') [5].

2.2. Phương pháp nghiên cứu

- Tách chiết ADN:

ADN tổng số của các mẫu lúa được tách chiết và tinh sạch theo phương pháp của Doyle J J và Doyle J L (1987) [6], có cải tiến phù hợp với vật liệu nghiên cứu.

- Khuếch đại PCR và xác định trình tự nucleotide:

Phản ứng PCR được thực hiện với thể tích 25 μ l, bao gồm: 12,5 μ l PCR Master Mix, 1,0 μ l mỗi xuôi (10 pmol/ μ l), 1,0 μ l mỗi ngược (10 pmol/ μ l), 2 μ l ADN khuôn (nồng độ 50 ng/ μ l) và 8,5 μ l H₂O khử ion.

Chương trình nhiệt của phản ứng PCR gồm: Biến tính ban đầu 94°C trong 3 phút; tiếp theo là 35 chu kỳ gồm biến tính ở 94°C trong 45 giây, bắt cặp 58°C trong 45 giây và kéo dài ở 72°C trong 1 phút; kéo dài cuối ở 72°C trong 10 phút và bảo quản 4°C.

Sản phẩm PCR được điện di trên gel agarose 1% trong đệm TAE, nhuộm Ethidium bromide (0,5 μ g/ml) và quan sát dưới hệ thống GelDoc-It.

Sản phẩm PCR được tinh sạch bằng bộ kit GeneJET PCR Purification Kit (Thermo Fisher Scientific) theo hướng dẫn của nhà sản xuất. Giải

trình tự nucleotide được thực hiện theo phương pháp Sanger, sử dụng bộ BigDye Terminator v3.1 Cycle Sequencing và phân tích trên hệ thống ABI 3500 Genetic Analyzer.

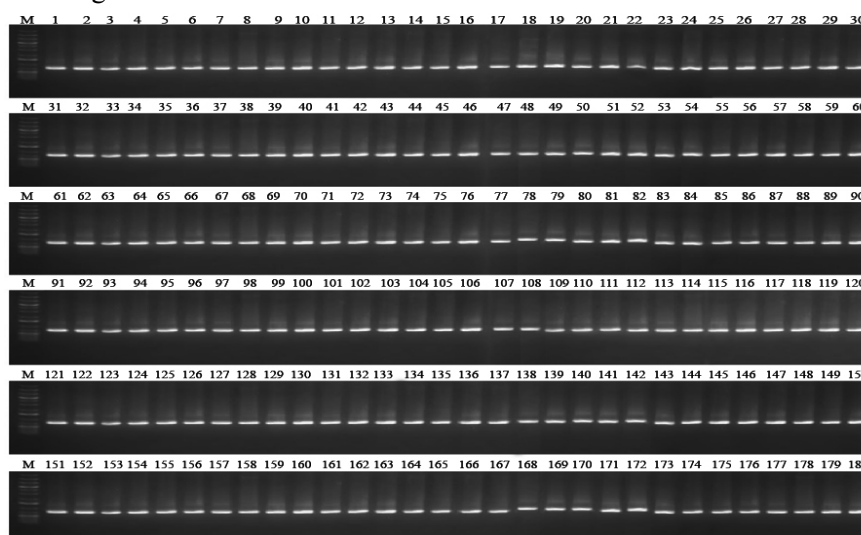
- Phân tích số liệu:

Các trình tự nucleotide được xử lý và phân tích bằng phần mềm BioEdit v.7.2.5 [7]. Trình tự sau khi xử lý được so sánh với cơ sở dữ liệu GenBank bằng công cụ BLASTn. Các biến dị di truyền, bao gồm đa hình nucleotide đơn (SNP) và đột biến chèn/xóa (InDel), được xác định dựa trên sự sai khác giữa các trình tự sau khi căn chỉnh. Khoảng cách di truyền giữa các mẫu được tính toán theo mô hình Kimura-2-parameter (K2P) bằng phần mềm MEGA7 [8]. Cây phát sinh loài được xây dựng theo phương pháp UPGMA với 1.000 lần lặp bootstrap nhằm đánh giá độ tin cậy của các nhánh.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

3.1. Kết quả khuếch đại vùng gen *rbcL* bằng kỹ thuật PCR

Vùng gen *rbcL* đã được khuếch đại thành công ở toàn bộ 180 mẫu giống lúa nghiên cứu. Kết quả điện di cho thấy, các băng ADN rõ nét, chỉ xuất hiện một băng duy nhất với kích thước khoảng 600 bp (Hình 1). Sản phẩm PCR sau đó được tinh sạch và sử dụng cho bước giải trình tự nucleotide tiếp theo.



Hình 1. Kết quả khuếch đại trình tự *rbcL* của 180 mẫu giống lúa địa phương

M: Marker 1kb; 1-180: 180 mẫu giống lúa nghiên cứu.

3.2. Phân tích trình tự nucleotide của vùng gen *rbcL*

Kết quả giải trình tự nucleotide vùng gen *rbcL* của 180 mẫu giống lúa địa phương Việt Nam cho thấy, chiều dài trình tự dao động từ 541 - 551 bp. Thành phần nucleotide trung bình gồm: 26,67% A, 29,13% T, 23,24% G và 20,96% C, trong đó tỷ lệ AT cao hơn GC. Kết quả này tương đồng với kết quả nghiên cứu của Mursyidin và cs (2021) [9] trên các giống lúa địa phương tại Indonesia, khi ghi nhận tỷ lệ AT (khoảng 55 - 56%) cao hơn GC. Sự tương đồng về thành phần nucleotide cho thấy, vùng gen *rbcL* có tính ổn định tương đối cao và có giá trị ứng dụng trong nghiên cứu mã vạch ADN (DNA barcoding), hỗ trợ hiệu quả công tác định danh và đánh giá quan hệ di truyền giữa các giống lúa địa phương tại khu vực Đông Nam Á.

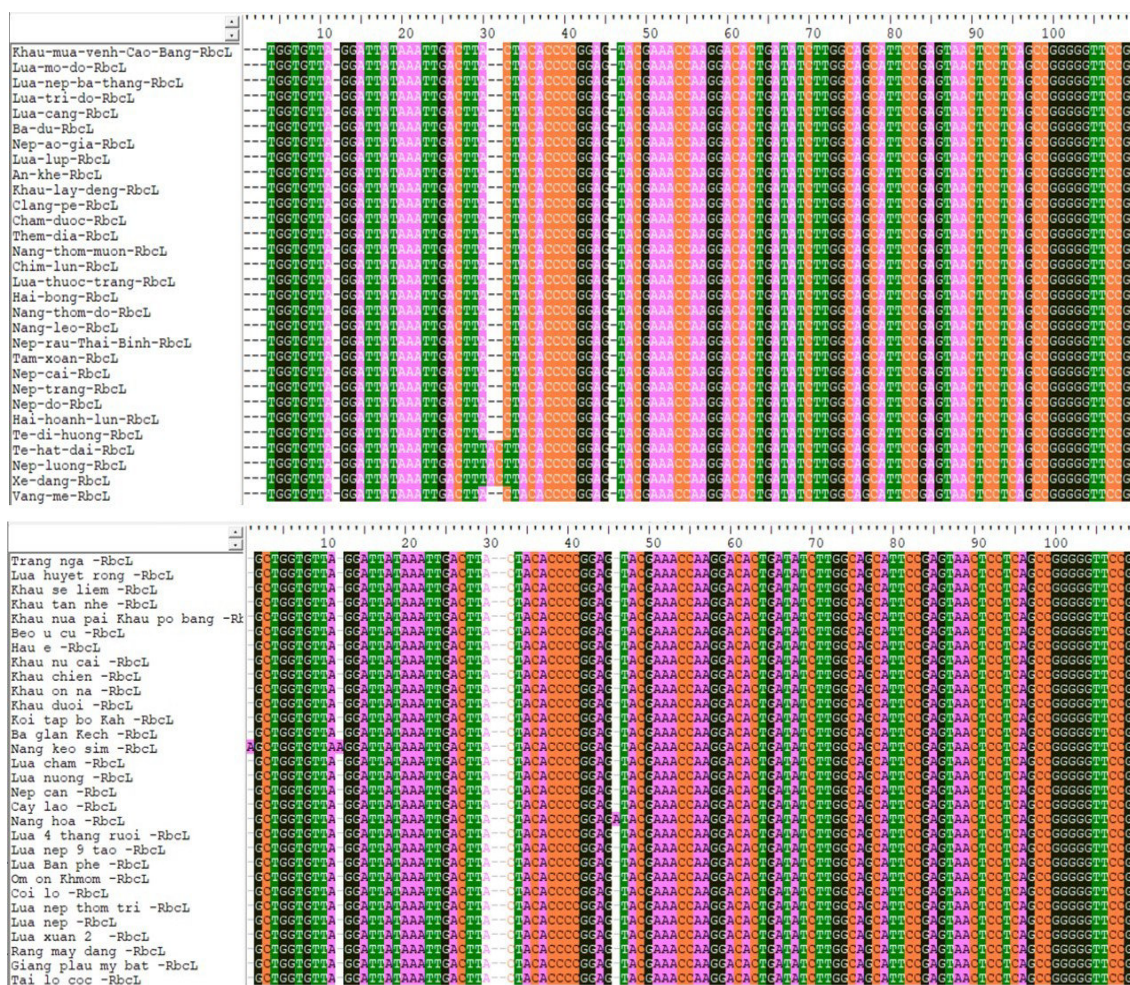
Kết quả so sánh bằng BLASTn với cơ sở dữ liệu GenBank cho thấy các trình tự *rbcL* của mẫu nghiên cứu có độ tương đồng từ 99,45 - 99,82% và

độ bao phủ đạt 100% so với các trình tự *rbcL* của loài *Oryza sativa* đã được công bố trên GenBank (HG997738.1, HG997272.1, HG997326.1). Hiện nay, cơ sở dữ liệu GenBank ghi nhận 152 trình tự *rbcL* thuộc chi *Oryza*, bao gồm: Các trình tự của *Oryza sativa*, *Oryza rufipogon*, *Oryza meridionalis*, *Oryza glaberrima*, *Oryza barthii* và *Oryza* sp.

Phân tích đa hình bằng phần mềm MEGA xác định 10 vị trí biến dị trên vùng gen *rbcL*, gồm 4 SNP và 6 InDel tại các vị trí 12, 22, 30, 31, 32, 33, 46, 495, 508 và 542 (Bảng 2, hình 2). Số lượng biến dị ghi nhận ở các giống lúa địa phương Việt Nam cao hơn so với kết quả nghiên cứu của Mursyidin và cs (2021) [9] trên các giống lúa địa phương tại Indonesia, chỉ phát hiện 8 vị trí đột biến. Sự khác biệt này có thể liên quan đến phạm vi sinh thái và điều kiện canh tác đa dạng của Việt Nam từ vùng núi cao, trung du đến đồng bằng ngập nước, góp phần thúc đẩy quá trình phân hoá di truyền của các quần thể lúa địa phương.

Bảng 2. Các vị trí đa hình nucleotide trên vùng gen *rbcL* của các mẫu giống lúa nghiên cứu

Số đăng ký	Tên giống	Vị trí đa hình nucleotide									
		12	22	30	31	32	33	46	495	508	542
7116	Tẻ Hạt Dài	-	A	T	A	C	T	-	T	A	-
7219	Nếp Luồng	-	A	T	A	C	T	-	T	A	-
7117	Xê Đắng	-	A	T	A	C	T	-	T	A	-
7811	Cầm	-	A	T	A		T	-	T	A	-
25302	Nàng Keo Sim	A	A	A	-	-	C	-	T	A	-
25316	Nàng Hoa	-	A	A	-	-	C	A	G	A	-
17874	Nghệ Tan Oĩ	-	-	A	-	-	C	-	T	A	-
25024	Một Bụi	-	A	A	-	-	C	-	G	A	-
17009	Khẩu Tẻ Thơm	-	A	A	-	-	C	-	G	A	-
9212	Tép An Giang	-	A	A	-	-	C	-	G	G	G
Tham chiếu		-	A	A	-	-	C	-	T	A	-



Hình 2. So sánh trình tự nucleotide vùng *rbcL* của một số mẫu giống lúa nghiên cứu

Dựa trên các vị trí biên dị thu được, các mẫu giống Tẻ Hạt Dài, Nếp Luồng và Xê Đăng có kiểu đa hình tương đồng tại các vị trí 30 - 33, trong khi giống Cẩm khác biệt tại vị trí 32. Mẫu giống Nàng Keo Sim ghi nhận biến dị tại vị trí 12, còn mẫu giống Nghêu Tan Ôi khác biệt tại vị trí 22. Mẫu giống Nàng Hoa mang các biến dị tại các vị trí 46 và 495, trong khi mẫu giống Tép An Giang có tổ hợp biến dị tại các vị trí 495, 508 và 542. Ngoài ra, hai mẫu giống Một Bụi và Khẩu Tẻ Thơm có kiểu đa hình tương đồng tại các vị trí khảo sát, cho thấy mối quan hệ di truyền gần gũi. Các tổ hợp biến dị SNP và InDel này góp phần hình thành các haplotype đặc trưng giữa các nhóm giống lúa địa phương và hỗ trợ nhận diện sơ bộ các nguồn gen có sự khác biệt về di truyền.

Về ý nghĩa sinh học, các biến dị InDel trên vùng gen *rbcL* về lý thuyết có thể gây đột biến

địch khung nêu số nucleotide thêm hoặc mất không phải bội số của 3. Trong nghiên cứu này, dữ liệu giải trình tự Sanger chỉ bao gồm đoạn trình tự một phần và đã được loại bỏ các tín hiệu nhiễu ở hai đầu, nên việc xác định đầy đủ khung đọc mở (ORF) cũng như phân loại toàn diện các biến dị đồng nghĩa (synonymous SNP) và không đồng nghĩa (non-synonymous SNP) còn hạn chế. Tuy nhiên, kết quả phân tích trong phạm vi trình tự khảo sát không ghi nhận mã kết thúc sớm hay các gián đoạn lớn, cho thấy các InDel phát hiện được chưa biểu hiện dấu hiệu ảnh hưởng rõ rệt đến chức năng của gen *rbcL*. Khi đối chiếu với trình tự tham chiếu của *Oryza sativa* trên GenBank, các SNP phát hiện được ghi nhận trong vùng gen có tính bảo tồn cao. Kết quả này phù hợp với các nghiên cứu của Kress và Erickson (2007) [5]; Hollingsworth và cs (2011) [10], cho

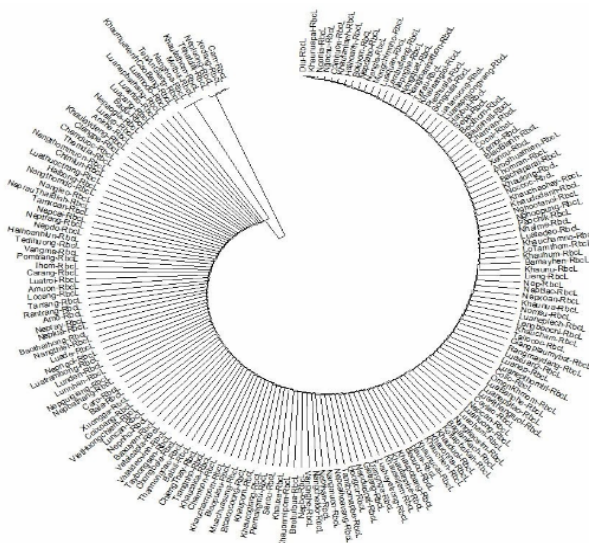
thấy vùng gen *rbcL* có tính bảo tồn cao và chịu áp lực chọn lọc nhằm duy trì chức năng sinh học của enzyme RuBisCO.

3.3. Phân tích quan hệ di truyền dựa trên vùng gen *rbcL*

Phân tích khoảng cách di truyền giữa các mẫu theo mô hình Kimura-2-parameter (K2P) cho thấy, giá trị khoảng cách giữa các mẫu dao động từ 0,0000 - 0,0037. Kết quả này tương đồng với

nhận định của Kumagai và cs (2010) [11], cho thấy trình tự gen lục lạp trong chi *Oryza* thường có mức sai khác thấp giữa các loài và giống có quan hệ gần.

Cây phát sinh loài được xây dựng theo phương pháp UPGMA dựa trên trình tự vùng gen *rbcL* của 180 mẫu giống lúa nghiên cứu được thể hiện ở hình 3.



Hình 3. Cây phân nhóm di truyền dựa trên trình tự vùng gen *rbcL* của 180 mẫu giống lúa

Kết quả phân tích cho thấy, 180 mẫu giống nghiên cứu được phân thành hai nhóm chính. Nhóm I bao gồm bốn mẫu: Tẻ Hạt Dài, Xê Đẳng thu thập tại tỉnh Kon Tum (cũ), Nếp Luồng thu thập tại tỉnh Thanh Hoá và Cầm thu thập tại tỉnh Quảng Trị, thuộc hai vùng sinh thái nông nghiệp là Tây Nguyên và Bắc Trung Bộ. Nhóm II gồm 176 mẫu giống còn lại, phân bố ở 8 vùng sinh thái nông nghiệp của Việt Nam. Kết quả cho thấy, sự phân bố các mẫu trong hai nhóm không theo vùng sinh thái nông nghiệp mà chủ yếu liên quan đến sự khác biệt về tổ hợp haplotype hình thành từ 10 vị trí đa hình trên vùng gen *rbcL*.

Các mẫu thuộc nhóm I có kiểu đa hình tương đồng tại nhiều vị trí nucleotide, trong đó các vị trí 30 - 33 có mức độ tương đồng cao và được xếp vào cùng một nhóm trên cây phân nhóm UPGMA. Trong khi đó, một số mẫu thuộc nhóm II có xu hướng phân tách nhánh rõ hơn do mang các biến dị đặc trưng, như Tép An Giang với biến dị tại các vị trí 508 và 542, hoặc Nàng Hoa với sai khác tại các vị trí 46 và 495. Đặc biệt, vùng

nucleotide 30 - 33 bao gồm cả SNP và InDel nên có thể đóng vai trò quan trọng trong cấu trúc phân nhóm của cây phát sinh loài. Các mẫu có sai khác nucleotide thấp thường được xếp cùng nhánh, trong khi các mẫu mang nhiều biến dị hơn có xu hướng phân tách thành các nhánh riêng, phản ánh đặc tính bảo tồn của vùng gen *rbcL*. Kết quả này phù hợp với nghiên cứu của Zhang và cs (2020) [12], trong đó *rbcL* được đánh giá là chỉ thị có mức độ biến dị thấp hơn so với các chỉ thị như *matK* hoặc ITS. Mặc dù có khả năng hỗ trợ đánh giá quan hệ di truyền và phân nhóm sơ bộ trong chi *Oryza*, *rbcL* vẫn còn hạn chế trong việc phân biệt các giống có quan hệ di truyền gần. Tuy nhiên, vùng gen này vẫn cung cấp thông tin hữu ích cho phân nhóm, truy xuất nguồn gốc và đánh giá quan hệ di truyền giữa các giống lúa địa phương Việt Nam.

4. KẾT LUẬN

Vùng gen *rbcL* của 180 mẫu giống lúa địa phương Việt Nam có thành phần nucleotide trung bình gồm: 26,67% A; 29,13% T; 23,24% G và

20,96% C, trong đó tỷ lệ A+T cao hơn G+C. Mức độ tương đồng trình tự nucleotide giữa các mẫu nghiên cứu và các trình tự tham chiếu thuộc chi *Oryza* trên GenBank dao động từ 99,45 - 99,82%, với độ bao phủ đạt 100%.

Phân tích đa hình trình tự xác định 10 vị trí biến dị SNP và InDel trên vùng gen *rbcL*, góp phần hình thành một số haplotype đặc trưng giữa các giống lúa địa phương. Khoảng cách di truyền giữa các mẫu giống theo mô hình K2P dao động từ 0,0000 - 0,0037, phản ánh mức độ sai khác nucleotide thấp do tính bảo tồn cao của vùng gen *rbcL*. Cây phát sinh loài xây dựng theo phương pháp UPGMA cho thấy các mẫu giống được phân thành hai nhóm chính.

Dựa trên các vị trí nucleotide đặc trưng, một số mẫu giống như: Nàng Keo Sim, Nàng Hoa, Tép An Giang và Cầm có sự khác biệt về SNP và InDel so với trình tự tham chiếu và các mẫu còn lại. Kết quả nghiên cứu cho thấy, chỉ thị *rbcL* phù hợp cho đánh giá đa dạng di truyền và phân nhóm sơ bộ nguồn gen lúa địa phương, đồng thời hỗ trợ xây dựng cơ sở dữ liệu phục vụ bảo tồn và truy xuất nguồn gốc nguồn gen lúa Việt Nam. Tuy nhiên, để nâng cao khả năng phân biệt các mẫu giống có quan hệ di truyền gần, cần kết hợp chỉ thị *rbcL* với các vùng gen khác hoặc các chỉ thị SSR và SNP.

LỜI CẢM ƠN

Nghiên cứu được thực hiện với sự hỗ trợ kinh phí của nhiệm vụ: “Nghiên cứu xây dựng cơ sở dữ liệu và mã vạch ADN (DNA barcode) của nguồn gen lúa địa phương/bản địa của Việt Nam” mã số: 20230923004 thuộc Đề án phát triển công nghiệp sinh học ngành nông nghiệp đến năm 2030, Bộ Nông nghiệp và Môi trường. Nhóm tác giả xin trân trọng cảm ơn!

TÀI LIỆU THAM KHẢO

[1]. Luong NH, Linh LH, Shim KC, Adeva C, Lee HS, Ahn SN (2021). Genetic structure and geographical differentiation of traditional rice (*Oryza sativa* L.) from Northern Vietnam. *Plants*, 10(10), 2094. <https://doi.org/10.3390/plants10102094>.

[2]. Higgins *et al.* (2021). Resequencing of 672 native rice accessions to explore genetic diversity and trait associations in Vietnam.

Rice, 14(52). <https://doi.org/10.1186/s12284-021-00481-0>.

[3]. CBOL Plant Working Group (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31): 12794 - 12797.

[4]. Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2020). Plant DNA barcoding: From gene to genome. *Biol Rev*, 95(1): 1 - 18.

[5]. Kress, W. J. & Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding trnH-psbA spacer region. *PLoS One*, 2(6), e508. <https://doi.org/10.1371/journal.pone.0000508>.

[6]. Doyle J J, Doyle J L (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*, 19(1): 11 - 15.

[7]. Hall TA. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*, 41, 95 - 98.

[8]. Kumar S, Stecher G, Tamura K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*, 33(7): 1870 - 1874.

[9]. Mursyidin D H, Nazari Y A, Badruzsaufari, Masmitra MRD (2021). DNA barcoding of the tidal rice (*Oryza sativa*) landraces from South Kalimantan, Indonesian. *Biodiversitas*, 22, 1593 - 1599

[10]. Hollingsworth, P. M., Graham, S. W. & Little, D. P. (2011). Choosing and using a plant DNA barcode. *PLoS One*, 6(5), e19254. <https://doi.org/10.1371/journal.pone.0019254>.

[11]. Kumagai M, Wang L, Ueda S. (2010). Genetic diversity and evolutionary relationships in genus *Oryza* revealed by using highly variable regions of chloroplast DNA. *Gene*, 462(1 - 2): 44 - 51. <https://doi.org/10.1016/j.gene.2010.04.013>.

[12]. Zhang W, Sun Y, Liu J, Xu C, Zou X, Chen X, Liu Y, Wu P, Yang X, Zhou S. (2020). DNA barcoding of *Oryza*: Conventional, specific, and super barcodes. *Plant Molecular Biology*, 105, 215 - 228. <https://doi.org/10.1007/s11103-020-01054-3>.

**DNA BARCODING ANALYSIS OF VIETNAMESE LOCAL RICE GERMPLASM
BASED ON *rbcL* GENE SEQUENCES**

**La Tuan Nghia¹, Le Thi Thu Trang¹, Nguyen Thanh Nhung², Do Ha Thu¹,
La Hoang Nhat Minh¹, Dam Thi Thu Ha¹, Nguyen Hoai Thu¹, Ta Diep Vien Phuong¹**

¹*Plant Resources Center*

²*Agricultural Genetics Institute*

Abstract

This study was conducted to characterize the DNA barcode profile of the *rbcL* gene region in 180 indigenous rice landraces collected from different ecological regions of Vietnam. Following PCR amplification, the *rbcL* amplicons were sequenced using the Sanger method and the sequence data were processed and analyzed using BioEdit and MEGA 7 software. The results showed that sequence lengths ranged from 541 to 551 bp. The average nucleotide composition comprised 26.67% A, 29.13% T, 23.24% G and 20.96% C. BLAST comparison against the GenBank database revealed sequence similarities ranging from 99.45% to 99.82% with 100% query coverage, indicating the highly conserved nature of the *rbcL* region in rice. Sequence polymorphism analysis identified 10 variable sites, including 4 SNPs and 6 InDels. These variations generated distinct haplotypes that enabled the identification and discrimination of several rice landraces, including Nang Keo Sim, Nang Hoa, Nghoo Tan Oi, Tep An Giang and Cam. Genetic distance based on the Kimura 2-parameter (K2P) model ranged from 0.0000 to 0.0037. Phylogenetic analysis using the UPGMA method classified the 180 rice landraces into two major groups. The findings indicate that the *rbcL* marker can support preliminary assessments of genetic diversity and the grouping of indigenous rice germplasm; however, additional highly polymorphic loci should be incorporated to improve discriminatory power among closely related rice accessions.

Keywords: *DNA barcoding, genetic diversity, local rice, nucleotide sequence, rbcL.*

Ngày nhận bài: 21/4/2026

Ngày chuyển phản biện: 27/4/2026

Ngày thông qua phản biện: 18/5/2026

Ngày duyệt đăng: 11/6/2026