

Nghiên cứu đa dạng di truyền của đoạn gen *matK* ở một số nguồn gen nhãn Việt Nam

Nguyễn Thị Ngọc Lan^{1*}, Nguyễn Thị Lan Hoa², Nguyễn Thị Thanh Thủy³, Lê Tuấn Nghĩa²

¹Viện Di truyền Nông nghiệp

²Trung tâm Tài nguyên Thực vật

³Bộ Nông nghiệp và Phát triển nông thôn

Ngày nhận bài 18/9/2018; ngày gửi phản biện 20/9/2018; ngày nhận phản biện 12/11/2018; ngày chấp nhận đăng 15/11/2018

Tóm tắt:

Kết quả nghiên cứu đa dạng trình tự đoạn gen *matK* gồm 829 nucleotid của tập đoàn 31 mẫu giống nhãn Việt Nam đã xác định được đột biến dị hoán (T>G) tại vị trí 939 của gen ở 11 giống (N10 - Nhãn Bản Nguyên, N14 - Long Gia Sắn, N16 - Tiêu Vung Tàu, N17 - Tiêu Da Me, N19 - Nhãn Sài Gòn, N22 - Cơm Vàng Bánh Xe, N26 - Xuồng Cơm Ráo, N28 - Long Tiêu, N29 - Xuồng Cơm Vàng Bà Rịa, N30 - Xuồng Cơm Trắng và N32 - Nhãn Vung Tàu). Những đột biến này có ý nghĩa trong việc nhận dạng các mẫu giống nhãn của nước ta. Các trình tự này đã được đăng ký NCBI với số đăng ký lần lượt là: KR073235, KR073239, KR073240, KR073241, KR073243, KR073245, KR073249, KR073251, KR073252, KR073253 và KR073255. Kết quả nghiên cứu cây phả hệ theo phương pháp Neighbour Joining cho thấy, các trình tự của chi *Dimocarpus* được nhóm thành công và phân biệt rõ ràng với trình tự của chi *Litchi*, *Arytera*, *Sapindoidaea* và *Cupaniopsis* trong họ Sapindaceae. 11 trình tự nhãn (N10, N14, N16, N17, N19, N22, N26, N28, N29, N30, N32) được tách biệt rõ ràng với các trình tự của nhãn Việt Nam và các nguồn gen đại diện khác.

Từ khóa: ADN mã vạch, giải trình tự, *matK*, nhãn.

Chỉ số phân loại: 4.6

Đặt vấn đề

Cây nhãn thuộc chi *Dimocarpus*, họ Sapindaceae, bộ Sapindales, là loài cây ăn quả nhiệt đới có giá trị kinh tế cao. Ở nước ta, nhãn được trồng phổ biến và tập trung tại một số vùng, tạo thành các vùng chuyên canh với nhiều giống nhãn nổi tiếng như nhãn lồng Hương chi, nhãn Đường phèn (Hưng Yên), nhãn Cùi (Lào Cai), Nhãn Tiêu Da Bò (Bà Rịa - Vũng Tàu), nhãn Xuồng Cơm Vàng (Tiền Giang)... Các giống nhãn Việt Nam với hương vị thơm ngon đã nổi tiếng trên thế giới và trở thành một trong những mặt hàng xuất khẩu chủ lực của nước ta. Do vậy, công tác bảo tồn và sử dụng bền vững tài nguyên di truyền cây nhãn là vô cùng quan trọng và cấp bách. Việc lưu giữ nguồn gen không chỉ phục vụ cho việc chọn giống mà còn nhằm tìm hiểu mối quan hệ di truyền gần gũi hay đặc trưng của các nguồn gen. Cũng giống như nhiều loài thực vật khác, những tiêu chí kiểu hình không phù hợp để phân biệt được các giống nhãn và việc áp dụng chỉ thị phân tử nhằm xác định chính xác các giống là cách tiếp cận tốt hơn nhằm nâng cao hiệu quả quản lý và sử dụng nguồn gen cây nhãn [1]. Trong khi chỉ

thị SNP dựa trên trình tự ADN cho cây nhãn bắt đầu được quan tâm ứng dụng [2] thì các mã vạch ADN lại chưa được phát triển và phổ biến rộng rãi trong phân tích xác định các giống nhãn. Gần đây, Hiệp hội Mã vạch cho cuộc sống (The Consortium for the barcode of life-CBOL) đã thành lập nhóm nghiên cứu với đại diện của các đơn vị, tổ chức trong lĩnh vực phát triển mã vạch ADN thực vật nhằm đánh giá các đoạn gen/trình tự tiềm năng [3]. Trình tự gen *matK* có tỷ lệ tiến hóa cao nhất trong các gen nên có khả năng phân biệt cao [4]. Chính vì vậy, việc lựa chọn đoạn gen *matK* để tiến hành khuếch đại và xác định trình tự nucleotid là cơ sở khoa học cần thiết nhằm phục vụ cho việc nhận dạng các nguồn gen nhãn, từ đó góp phần nâng cao hiệu quả quản lý, bảo tồn và chọn tạo các nguồn gen quý.

Vật liệu và phương pháp nghiên cứu

Vật liệu, địa điểm nghiên cứu

Đối tượng nghiên cứu: 31 mẫu giống nhãn được lưu giữ tại các vườn bảo tồn của các đơn vị nghiên cứu thuộc Viện Khoa học Nông nghiệp Việt Nam (bảng 1).

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

Genetic diversity of *matK* gene in Vietnam's longan germplasm

Thi Ngoc Lan Nguyen^{1*}, Thi Lan Hoa Nguyen²,
Thi Thanh Thuy Nguyen³, Tuan Nghia La²

¹Agricultural Genetics Institute

²Plant Resources Center

³Ministry of Agriculture and Rural Development

Received 18 September 2018; accepted 15 November 2018

Abstract:

The genetic diversity survey on the *matK* gene segment of 829 nucleotides among 31 Vietnam's longan samples has identified transversion mutation (T>G) at the 939 downstream position of the sequences in 11 longan samples, which might be the molecular identification to distinguish Vietnam longan varieties with others. These *matK* nucleotide sequences have been registered with NCBI codes (KR073235- Ban Nguyen, KR073239- Long Gia San, KR073240- Tieu Vung Tau, KR073241- Tieu Da Me, KR073243- Nhan Sai Gon, KR073245- Com Vang Banh Xe, KR073249- Xuong Com Rao, KR073251- Long Tieu, KR073252- Xuong Com Vang Ba Ria, KR073253- Xuong Com Trang and KR073255- Nhan Vung Tau). The phylogenetic tree analysed by the Neighbour Joining method based on the 829 nucleotides of *matK* gene has exactly grouped all surveyed sequences. In addition, this analysis has also clearly separated the *Dimocarpus* and the *Litchi*, *Arytera*, *Sapindoidaea*, and *Cupaniopsis* in Sapindaceae. Eleven longan sequences (N10, N14, N16, N17, N19, N22, N26, N28, N29, N30, N32) have been distinguished with the Vietnam's longan and other reference sequences.

Keywords: *Dimocarpus longan*, DNA barcode, *matK*, sequencing.

Classification number: 4.6

Bảng 1. Danh sách các mẫu giống nhãn nghiên cứu.

ID ¹	ID ²	Tên giống	ID ¹	ID ²	Tên giống
GBVNML1.131	N1	Hương chi	GBVNML18.572	N18	Tiểu lá bầu
GBVNML1.140	N2	Nhãn lồng chín sớm	GBVNML18.578	N19	Nhãn Sài Gòn
GBVNML1.126	N3	Nhãn lồng Hưng Yên chín muộn	GBVNML18.577	N20	Tiểu lá dài
GBVNML1.132	N4	HTM1	GBVNML18.569	N21	Nhãn Cùi
GBVNML1.144	N5	HTM2	GBVNML18.574	N22	Com Vàng Bánh Xe
GBVNML1.146	N6	Nhãn lồng Hưng Yên chính vụ	GBVNML18.570	N23	Cùi diếc
GBVNML1.129	N7	Nhãn Trư Lương	GBVNML18.571	N24	Long Hưng Yên
GBVNML1.127	N8	Nhãn Đoàn Hùng 3	GBVNML18.565	N25	Thái Long Tiêu
GBVNML1.125	N9	Nhãn Mỹ Đức	GBVNML18.585	N26	Xuong Com Rao
GBVNML1.137	N10	Nhãn Bàn Nguyên	GBVNML18.942	N27	Tiểu Da Bò
GBVNML1.138	N11	Nhãn Lập Trạch	GBVNML18.562	N28	Long Tiêu
GBVNML1.142	N12	Nhãn Việt Trì	GBVNML18.573	N29	Xuong Com Vàng Bà Rịa
GBVNML1.133	N13	Nhãn Phong Châu 5	GBVNML18.584	N30	Xuong Com Trắng
GBVNML18.566	N14	Long Gia Sản	GBVNML18.568	N31	Nhãn Đường phèn
GBVNML18.580	N16	Tiểu Vung Tàu	GBVNML18.579	N32	Nhãn Vung Tàu
GBVNML18.564	N17	Tiểu Da Me			

Ghi chú: ID¹: số đăng ký cơ quan mạng lưới, Ngân hàng gen cây trồng quốc gia (GBVNML), ID²: ký hiệu giống.

Bộ môi khuếch đại các vùng gen *matK* được thể hiện ở bảng 2.

Bảng 2. Danh sách môi khuếch đại vùng gen *matK*.

Gen	Môi	Trình tự 5'-3'	Tham khảo
<i>matK</i>	Kim3F	CGTACAGTACTTTTGTGTTTACGAG	[5]
	Kim1R	TAGAATTCCTCCGGTTCGCTCGCCGTTAC	

Địa điểm nghiên cứu: Phòng thí nghiệm của Bộ môn Đa dạng sinh học, Trung tâm Tài nguyên Thực vật (An Khánh - Hoài Đức - Hà Nội).

Phương pháp nghiên cứu

ADN tổng số được tách chiết và tinh sạch theo phương pháp dùng công nghệ màng lọc silica và cột lọc của QUIAGEN Dnaeasy Plan Kit.

Phản ứng PCR khuếch đại gen và *matK* được thực hiện với thành phần: 14,34 µl H₂O deion, 2 µl 10x PCR buffer, 0,16 µl dNTP mix, 0,2 µl môi Kim1R/Kim1F, 0,1 µl Taq polymerase (Fusion taq) và 0,1 µl ADN tổng số; thể tích phản ứng PCR là 20 µl; điều kiện nhiệt: biến tính ở 94°C trong 4 phút, 1 chu trình; 94°C trong 40 giây, 52°C trong 35 giây, 72°C trong 1 phút, 35 chu trình; 72°C trong 10 giây,

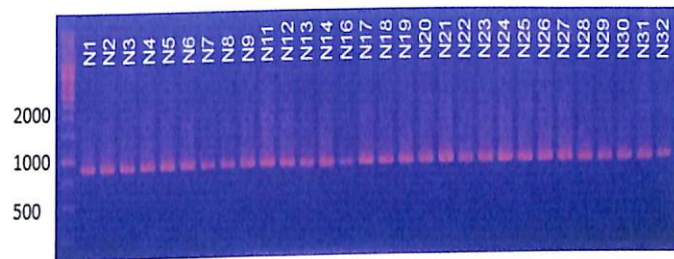
bảo quản tại 16°C. Sản phẩm PCR được điện di trên gel agarose 1,5% trong đệm TBE, nhuộm bằng EtBr. Kiểm tra sản phẩm PCR bằng máy đo quang phổ nanodrop 2000. Những mẫu có nồng độ sản phẩm khoảng 100 ng/μl được dùng để giải trình tự.

Chu trình và phản ứng cho giải trình tự: mẫu được khuếch đại với từng môi, mỗi mẫu lặp lại 5 lần với thành phần như sau: 5,5 μl water nuclease-free, 1 μl Bigdie, 2 μl X5 buffer, 0,5 μl mỗi xuôi hoặc ngược và 1 μl PCR template; tổng thể tích 20 ul; ở điều kiện 96°C trong 1 phút, 1 chu trình; 96°C trong 10 giây, 50°C trong 5 giây, 60°C trong 4 phút, 25 chu trình; 72°C trong 2 phút; bảo quản tại 4°C. Sản phẩm được đưa vào hệ thống giải trình tự bằng máy ABI3700 của 1st base Seq. company (Singapore).

Trình tự của các mẫu thu được đã được xử lý, hiệu chỉnh bằng phần mềm BioEdit 4.9, sắp xếp thẳng hàng trình tự bằng công cụ ClustalW và so sánh trên Ngân hàng gen bằng công cụ NCBI/BLAST được tích hợp trong phần mềm phân tích Geneious 7.0.

Kết quả nghiên cứu và thảo luận

Kết quả tách chiết mẫu lá của 31 mẫu giống nhãn nghiên cứu cho thấy sự đồng đều về kích thước có chiều dài 800 bp (hình 1), ADN có độ tinh sạch cao và không còn ARN lẫn tạp, đủ tiêu chuẩn để thực hiện các bước khuếch đại để giải trình tự tiếp theo.



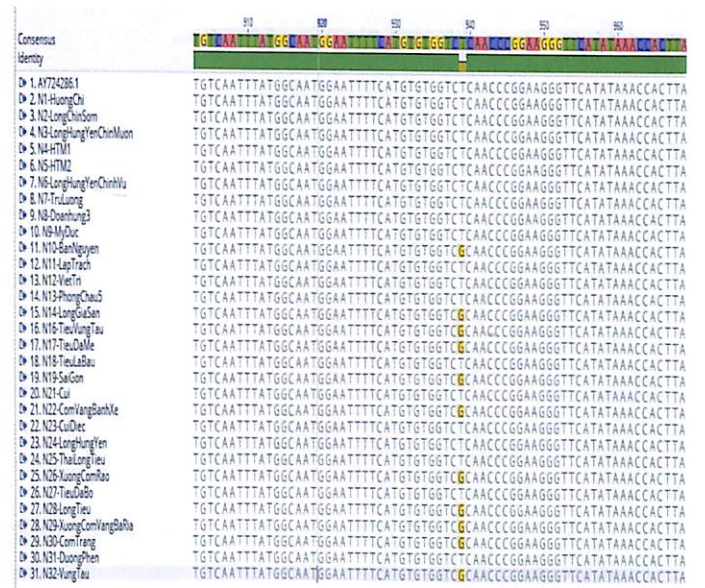
Hình 1. Khuếch đại vùng gen *matK* bằng cặp môi Kim3F/1R.

Kết quả kiểm tra tinh sạch mẫu sản phẩm PCR cho thấy, các mẫu đều có nồng độ cao >700 ng/μl và chất lượng đảm bảo. Sau khi tinh sạch và thổi gel, các mẫu này được kiểm tra định lượng lại bằng nanodrop 2000, kết quả cũng cho nồng độ đạt 100 ng/μl ở tất cả các mẫu với chỉ số tinh sạch đảm bảo cho giải trình tự.

Kết quả giải trình tự đoạn gen *matK* từ 31 mẫu giống nhãn thu được các trình tự với chiều dài biến thiên từ 677 bp (N10 - Nhân Bàn Nguyên) đến 829 bp (N8 - Nhân Đoàn Hùng 3). Tiến hành căn chỉnh và so sánh với trình tự của toàn bộ hệ gen lục lạp (BLAST với dữ liệu trình tự NCBI), kết quả BLAST hit cho thấy 829 bp nucleotide của các mẫu giống nhãn nằm trong vùng cấu trúc gen của gen *matK*. Vì

vậy, các trình tự này đã được phân tích để tìm khung đọc ORF cho việc dịch mã coden với các axit amin tương ứng.

Kết quả phân tích so sánh các trình tự đoạn gen *matK* thu được từ 31 mẫu giống nhãn cho thấy, các trình tự của đoạn gen *matK* gần như hoàn toàn tương đồng. Tuy nhiên, có sự khác biệt ở 1 nucleotit với cả tập đoàn nhãn nghiên cứu tại vị trí 939 T>G (vị trí của gen *matK*, tham chiếu từ trình tự có số đăng ký NCBI AY724286.1). 11 nguồn gen có sự khác biệt là N10 - Nhân Bàn Nguyên, N14 - Long Gia Sắn, N16 - Tiêu Vừng Tàu, N17 - Tiêu Da Me, N19 - Nhân Sài Gòn, N22 - Cơm Vàng Bánh Xe, N26 - Xuồng Cơm Ráo, N28 - Long Tiêu, N29 - Xuồng Cơm Vàng Bà Rịa, N30 - Xuồng Cơm Trắng, N32 - Nhân Vừng Tàu (hình 2).



Hình 2. So sánh trình tự gen *matK* của nguồn gen nhãn nghiên cứu.

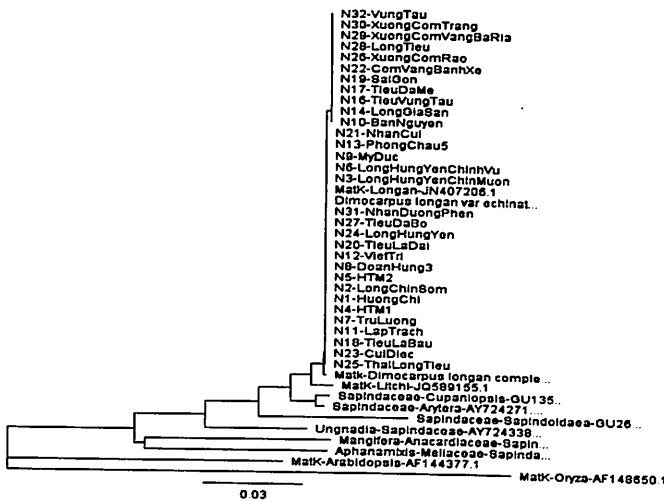
Tiến hành BLAST các trình tự này với ngân hàng dữ liệu trình tự của NCBI và CBOLD, kết quả thu được như sau:

- Mức độ tương đồng của trình tự đoạn gen *matK* giữa các loài nhãn trong chi *Dimocarpus* biến thiên từ 99 lên đến 100%.

- Đột biến dị hoán (T>G) tại vị trí 934 của đoạn trình tự ở cả 11 giống nhãn khác biệt hoàn toàn với toàn bộ các đoạn trình tự trên ngân hàng gen, chỉ xuất hiện trên 11 giống nhãn của Việt Nam đến thời điểm đăng ký. Các trình tự này có mã số trên NCBI lần lượt là KR073235 (N10 - Nhân Bàn Nguyên), KR073239 (N14 - Long Gia Sắn), KR073240 (N16 - Tiêu Vừng Tàu), KR073241 (N17 - Tiêu Da Me), KR073243 (N19 - Nhân Sài Gòn), KR073245 (N22 - Cơm Vàng Bánh Xe), KR073249 (N26 - Xuồng Cơm Ráo), KR073251 (N28 - Long Tiêu), KR073252 (N29 - Xuồng Cơm Vàng Bà Rịa), KR073253 (N30 - Xuồng Cơm Trắng)

và KR073255(N32 - Nhãn Vững Tàu).

Các trình tự của đoạn gen *matK* gồm 829 nucleotid và một số trình tự tham chiếu từ NCBI tương ứng đại diện cho chi *Dimocarpus*, họ Sapindaceae, bộ Sapindales được sử dụng để phân tích tương quan di truyền giữa các mẫu giống trong tập đoàn nghiên cứu, bằng phương pháp lập sơ đồ hình cây NJ [6], với outgroup là trình tự của loài *Oryza sativa* - đại diện nhóm cây 1 lá mầm. Các trình tự tham chiếu được chọn với tiêu chí là các trình tự có tương đồng từ gần nhất đến xa nhất trong danh sách BLAST-hit. Các trình tự này có phân loại *Dimocarpus longan* var. *echinatus*, *Paullinia pinnata* - Sapindoideae, *Arytera*, *Cupaniopsis*, họ Sapindaceae và bộ Sapindales (hình 3).



Hình 3. Cây phân loại dựa trên đoạn gen *matK* của các mẫu giống nhãn nghiên cứu.

Từ hình 3 cho thấy, trình tự cây *Arabidopsis thaliana* được nhóm riêng tách biệt với các nhóm trình tự thuộc bộ Sapindales. Trong bộ Sapindales, các trình tự của họ Sapindaceae được tách biệt với các họ Anacardiaceae và họ Meliaceae. Trong họ Sapindaceae, các trình tự của chi *Dimocarpus* được nhóm thành công và phân biệt rõ ràng với chi *Litchi*, *Arytera*, *Sapindoideae* và *Cupaniopsis*. 11 trình tự nhãn (N10, N14, N16, N17, N19, N22, N26, N28, N29, N30, N32) đứng tách biệt với các trình tự của nhãn Việt Nam và các nguồn gen đại diện khác. Như vậy, kết quả phân nhóm dựa vào đoạn 829 nucleotid này phù hợp với trình tự phân loại từ bộ, họ, chi theo khóa phân loại thông thường tham khảo từ NCBI. Kết quả này cho thấy đoạn gen *matK* có ý nghĩa trong phân loại đến mức nhận dạng chi.

Kết luận

Đánh giá đa dạng trình tự đoạn gen *matK* gồm 829 nucleotid của tập đoàn 31 mẫu giống nhãn nghiên cứu đã xác định được đột biến dị hoán (T>G) tại vị trí 939 của đoạn trình tự ở cả 11 giống nhãn: N10 - Nhãn Bản Nguyên,

N14 - Long Gia Sần, N16 - Tiêu Vững Tàu, N17 - Tiêu Da Me, N19 - Nhãn Sài Gòn, N22 - Cơm Vàng Bánh Xe, N26 - Xuồng Cơm Ráo, N28 - Long Tiêu, N29 - Xuồng Cơm Vàng Bà Rịa, N30 - Xuồng Cơm Trắng, N32 - Nhãn Vững Tàu. Các đột biến này khác biệt với toàn bộ các đoạn trình tự trên ngân hàng gen cho đến thời điểm đăng ký trên NCBI với mã số lần lượt là KR073235, KR073239, KR073240, KR073241, KR073243, KR073245, KR073249, KR073251, KR073252, KR073253 và KR073255.

Phân tích cây phả hệ cho thấy, trình tự của đoạn gen *matK* đã nhóm được các loài *Dimocarpus longan* với nhau và tách biệt chi *Dimocarpus* với các chi khác trong họ Sapindaceae, bộ Sapindales. Thông tin cả đoạn với 829 trình tự (từ 446 đến 1318 bp) đã không phân biệt chính xác các loài nhãn nghiên cứu và tham chiếu. Tuy nhiên, đoạn gen *matK* đã tách biệt được 11 giống nhãn có SNP đặc trưng của Việt Nam.

LỜI CẢM ƠN

Công trình là kết quả của Đề tài Xây dựng tiêu bản ADN (DNA barcode) cho các giống cây trồng đặc hữu có giá trị kinh tế của Việt Nam thuộc Chương trình Công nghệ sinh học nông nghiệp và thủy sản. Nhóm tác giả xin trân trọng cảm ơn.

TÀI LIỆU THAM KHẢO

- [1] Z.X. Lai, C.L. Chen, Z.G. Chen (2001), "Progress in biotechnology research in longan", *International Society for Horticultural Science (ISHS)*, **558**, pp.137-141, doi: 10.17660/ActaHortic.2001.558.18.
- [2] Boyi Wang, Hua-Wei Tan, Wanping Fang, Lyndel W. Meinhardt, Sue Mischke, Tracie Matsumoto and Dapeng Zhang (2015), "Developing single nucleotide polymorphism (SNP) markers from transcriptome sequences for identification of longan (*Dimocarpus longan*) germplasm", *Horticulture Research*, **2**, 14065, doi: 10.1038/hortres.2014.65.
- [3] CBOL Plant Working Group (2009), "A DNA barcode for land plants", *Proc. Natl. Acad. Sci. USA*, **106(31)**, pp.12794-12797, doi: 10.1073/pnas.0905845106.
- [4] K. Vijayan and C.H. Tsou (2010), "DNA barcoding in plants: taxonomy in a new perspective", *Current Science*, **99(11)**, pp.1530-1541.
- [5] W. John Kress and David L. Erickson (2012), "DNA barcodes: methods and protocols", *Methods in Molecular Biology*, **858**, doi: 10.1007/978-1-61779-591-6_11.
- [6] N. Saitou and M. Nei (1987), "The neighbor-joining method: a new method for reconstructing phylogenetic trees", *Molecular Biology and Evolution*, **4(4)**, pp.406-425, doi: 10.1093/oxfordjournals.molbev.a040454.