

XÁC ĐỊNH TRÌNH TỰ NUCLEOTIDE VÙNG GEN LỤC LẠP (*trnL - trnF*) VÀ KHẢ NĂNG SỬ DỤNG ĐỂ NHẬN DIỆN LOÀI GIỐI ĂN HẠT (*Michelia tonkinensis* A.Chev.)

VŨ ĐÌNH GIÁP ⁽¹⁾, NGUYỄN THANH TUẤN ⁽³⁾, NGUYỄN VĂN QUÝ ⁽³⁾, VŨ ĐÌNH DUY ⁽²⁾,
BÙI THỊ TUYẾT XUÂN ⁽⁴⁾, BÙI VĂN THẮNG ⁽⁵⁾, VŨ QUANG NAM ⁽⁵⁾

1. ĐẶT VĂN ĐỀ

Giổi ăn hạt (*Michelia tonkinensis* A.Chev.), họ Mộc lan (Magnoliaceae) đặc trưng với các đặc điểm: lá không có sẹo lá kèm trên cuống lá, các lá noãn ít (thường dưới 10), các đài trướng thành hình thuôn dài, có cuống quả và có các eo thắt hình cù lục. Giổi ăn hạt là loài cây gỗ đa tác dụng, có giá trị kinh tế và bảo tồn cao. Tại Việt Nam, loài cây này phân bố từ Lào Cai đến các tỉnh Bắc Trung Bộ và Tây Nguyên [1]. Đây là một trong những loài gỗ được ưa chuộng trong xây dựng nhà cửa, đóng đồ đạc. Hạt có tinh dầu, loại gia vị truyền thống của nhân dân vùng núi phía Bắc, dùng làm thuốc chữa đau bụng, ăn uống không tiêu, xoa bóp khi đau nhức, tê thấp [2, 3]. Hiện nay các quần thể giổi ăn hạt trong rừng tự nhiên đang bị suy giảm nghiêm trọng do bị khai thác cạn kiệt và số lượng cây tái sinh tự nhiên thấp do hạt bị thu hái quá mức.

Mã vạch DNA sử dụng đoạn DNA ngắn đã chuẩn hóa phân biệt giữa các loài [4, 5, 6] và trở thành công cụ hỗ trợ có hiệu quả cho công tác giám định, phân loại, đánh giá mối quan hệ di truyền và phát hiện loài mới [6 - 8]. Ở thực vật, một số vùng gen lục lạp (*matK*, *rbcL*, *psbA-trnH*, *trnL-trnF*, *atpF-atpH...*) và vùng gen nhân (*ITS-rDNA*) đang được ứng dụng rộng rãi trong các nghiên cứu mối quan hệ phát sinh chủng loại, phân loại và nhận dạng loài [9, 10]. Vùng gen *trnL-trnF* nằm trong vùng sao chép đơn lớn của bộ gen lục lạp, có tính đa hình cao và được sử dụng trong định danh, phân loại và xác định mối quan hệ phát sinh loài. Nghiên cứu mối quan hệ di truyền và xác định vùng mã vạch DNA của một số loài thuộc họ Mộc lan ở Việt Nam đã được tiến hành bởi Ha Van Huan và cộng sự [11]. Các tác giả chỉ ra trong số 4 vùng mã vạch DNA, vùng *trnH-psbA* là vùng mã vạch DNA hiệu quả nhất và các tổ hợp (*matK + trnH-psbA*, *rbcL + trnH-psbA*, *ycf1b + trnH-psbA*, *matK + ycf1b + trnH-psbA*, *matK + rbcL + ycf1b + trnH-psbA*) có thể được sử dụng làm mã vạch DNA để xác định loài và phân tích mối quan hệ di truyền giữa một số loài thuộc họ Mộc lan. Wang và cộng sự đã phân tích đặc điểm bộ gen lục lạp hoàn chỉnh cho bảy loài *Manglietia* (*Manglietia aromatica*, *M. calcarea*, *M. conifera*, *M. duclouxii*, *M. glaucifolia*, *M. insignis* và *M. megaphylla*) và một loài *Michelia* (*Michelia alba*) sử dụng công nghệ giải trình tự Illumina, với kích thước bộ gen lục lạp hoàn chỉnh từ 159.973 đến 160.106 bp [12]. Phân tích phát sinh loài đã hỗ trợ phân loại ở cấp độ phân họ hiện tại của họ Mộc lan, nhưng phân loại cấp độ gen của các loài phân họ Magnolioideae vẫn còn là một thách thức. Gần đây, Deng và cộng sự đã mô tả bộ gen lục lạp hoàn chỉnh của loài *Michelia shiliensis* [13]. Kết quả phân tích so sánh cho

thấy sự tương đồng cao giữa bộ gen lục lạp của loài *M. shiluensis*, bốn loài (*Michelia odora*, *Magnolia laevifolia*, *Magnolia insignis* và *Magnolia cathcartii*) và xây dựng cây phát sinh chủng loại cho thấy loài *M. shiluensis* có liên quan chặt chẽ nhất với *M. odora*. Thông tin bộ gen trong nghiên cứu này có giá trị để phân loại, nghiên cứu phát sinh chủng loại và để hỗ trợ các nỗ lực bảo tồn các loài này. Tuy nhiên, ở Việt Nam các nghiên cứu trước đây mới chỉ tập trung vào các đặc điểm sinh học, sinh thái và phân bố của loài Giổi ăn hạt [1, 3], các nghiên cứu ứng dụng DNA mã mạch trong nhận dạng loài Giổi ăn hạt vẫn rất hạn chế.

Trong nghiên cứu này chúng tôi giải trình tự nucleotide vùng gen (*trnL - trnF*) của loài Giổi ăn hạt nhằm xem xét khả năng sử dụng trong định danh loài và xây dựng mối quan hệ di truyền giữa các loài trong chi *Michelia*.

2. VẬT LIỆU VÀ PHƯƠNG PHÁP

Trong các chuyến khảo sát thực địa đã thu được 3 mẫu Giổi ăn hạt (*Michelia tonkinensis*) (ký hiệu 1L-3L) được dùng làm vật liệu trong nghiên cứu này (Hình 1, Bảng 1). Các mẫu được bảo quản trong túi nhựa dẻo có chứa silicagel ngay tại thực địa và chuyên đến phòng thí nghiệm giữ ở nhiệt độ phòng đến khi sử dụng tách chiết DNA.



Hình 1. Cành lá mang hoa với bộ nhụy mang ít lá noãn (bên trái) và quả trưởng thành với các đại mang cuồng ngắn và eo thắt (bên phải) của loài Giổi ăn hạt

Bảng 1. Địa điểm thu các mẫu Giổi ăn hạt trong nghiên cứu

Ký hiệu	Số hiệu tiêu bản	Noi thu mẫu	Vĩ độ	Kinh độ	Mã số GenBank gen <i>trnL - trnF</i>
1L	LN 01	Rừng núi luốt, Trường Đại học Lâm nghiệp, Hà Nội	20°51'13''	105°30'45''	MT176165
2L	LS 01	Lương Sơn, Hòa Bình	20°36'30''	105°41'25''	MT176166
3L	Nam 18818.4	Xuân Liên, Thanh Hóa	19°58'43''	104°59'43''	MT176167

Tách chiết DNA tổng số: DNA tổng số của loài Giổi ăn hạt được tách chiết bằng bộ hóa chất Plant DNA isolation Kit (Norgenbiotek, Canada). Các bước được thực hiện theo hướng dẫn của nhà sản xuất.

Nhân bản gen đích bằng kỹ thuật PCR: Nhân bản vùng gen *trnL - trnF* bằng kỹ thuật PCR sử dụng các cặp mồi: *trnL-F*: CGAAATCGGTAGACGCTACG; *trnF-R*: GGGGATAGAGGGACTTGAAC [14]. Mỗi phản ứng PCR có thể tích 25 µL với các thành phần: 7 µL H₂O deion, 12,5 µL PCR Master mix 2X (Thermo Scientific, Đức), 1,25 µL mồi xuôi (10 pmol/µL), 1,25 µL mồi ngược (10 pmol/µL), 3 µL DNA (10 - 20 ng). Phản ứng được thực hiện trên (hệ thống GeneAmp PCR System 9700 (Applied Biosystems, Mỹ). Chu trình nhiệt của PCR gồm: 94°C trong 3 phút; tiếp sau là 35 chu kỳ nối tiếp nhau với các bước: 94°C trong 45 giây, 55°C trong 45 giây, 72°C trong 45 giây; kết thúc phản ứng nhân gen ở 72°C trong 10 phút, giữ sản phẩm ở 4°C.

Giải trình tự và hiệu chỉnh trình tự: Quá trình xác định trình tự nucleotide được thực hiện tại Công ty Macrogen (Hàn Quốc). Trình tự DNA được hiệu chỉnh loại bỏ các tín hiệu nhiễu sử dụng phần mềm ChromasPro2.1.6 và so sánh với các trình tự đã công bố trên GenBank (sử dụng công cụ BLAST trong NCBI: <http://www.ncbi.nlm.nih.gov/BLAST>). Các trình tự phân tích được sắp xếp thẳng hàng bằng phần mềm Bioedit v7.0.5.2 [15]. Các vùng không có khả năng sắp xếp bị loại bỏ trước khi phân tích.

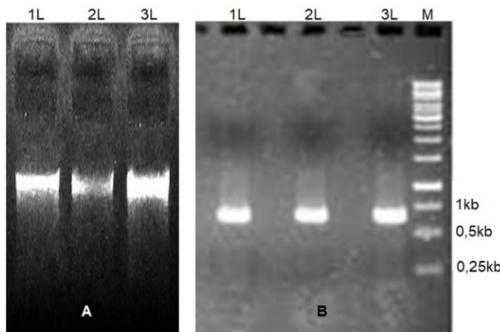
Xây dựng cây phát sinh chủng loại: Cây phát sinh chủng loại được xây dựng dựa trên phương pháp xác suất tối đa Maximum Likelihood (ML) sử dụng phần mềm Treefinder v 2011 [16]. Trước khi phân tích ML, dữ liệu trình tự nucleotide được khảo sát phân bố nucleotide, kiểm tra các giả thuyết và xác định mô hình tiến hóa tối ưu sử dụng Kakusan 4.0 [17] dựa trên thông tin Akaie được hiệu chỉnh (corrected AICc - Akaike Information Criterion). Mô hình tiến hóa tốt nhất được chọn cho ML là mô hình đảo chiều thời gian tổng thể (GTR) với giá trị tham số gamma (G: 3,66) trên vùng gen *trnL - trnF*. Thực hiện với 1000 lần lặp lại để xác định giá trị ủng hộ (bootstrap) trong cây ML (MLBS). Khoảng cách di truyền (*P*) giữa các loài trong chi được tính toán bằng Mega 7.0 [18].

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Trình tự nucleotide gen *trnL - trnF* của loài Giổi ăn hạt

DNA tổng số của 3 mẫu Giổi ăn hạt đã được tách chiết bằng bộ hóa chất Plant DNA isolation Kit (Norgenbiotek, Canada) (Hình 2A). Cặp mồi *trnL-F/trnF-R* đã nhân bản thành công vùng gen *trnL - trnF* ở nhiệt độ gắn mồi là 55°C với kích thước khoảng 900 bp chỉ có một băng duy nhất, sáng đậm, đủ tiêu chuẩn để giải mã trình tự nucleotide (Hình 2B). Sản phẩm PCR 3 mẫu Giổi ăn hạt sau khi được giải trình tự hai chiều (xuôi và ngược) của vùng gen *trnL - trnF* được hiệu chỉnh, ghép nối với sự trợ giúp của phần mềm Chromaspro 2.1.6 để loại bỏ các vùng tín hiệu nhiễu và các đinh màu không rõ ràng. Tổng số 3 trình tự nucleotide sau hiệu chỉnh đã thu được đoạn trình tự có kích thước tương ứng là 855, 921 và 928 nucleotide. Các trình tự

này được đăng ký trên ngân hàng gen thế giới (Bảng 1). Trình tự nucleotide thu được từ 3 mẫu được kiểm tra tính tương đồng với các trình tự sẵn có trên ngân hàng GenBank bằng công cụ BLAST. Kết quả tìm kiếm cho thấy trình tự nucleotide vùng gen *trnL - trnF* của 3 mẫu tương đồng cao 100% với loài *M. champaca* (AY009041) và *M. baillonii* (AY009042). Kết quả này đã xác định được chính xác đoạn gen *trnL - trnF* trong hệ gen lục lạp của loài nghiên cứu.



Hình 2. Hình ảnh DNA tổng số (A) và sản phẩm PCR (B) của mẫu Giổi ăn hạt *Michelia tonkinensis* điện di trên gel agarose 1,5% (A). Marker phân tử 1 kb

3.2. Khoảng cách di truyền (*P*) giữa loài nghiên cứu với một số loài trong chi Giổi

Cho đến nay trên GenBank mới chỉ lưu trữ trình tự nucleotide vùng gen *trnL - trnF* của 6 loài trong chi Giổi. Do đó, trình tự 3 mẫu Giổi ăn hạt được so sánh về khoảng cách di truyền với 6 loài trong cùng chi *Michelia* lấy trên Genbank (Bảng 2).

Bảng 2. Khoảng cách di truyền giữa các mẫu trong nghiên cứu và 6 loài cùng chi lấy trên Genbank trên cơ sở phân tích vùng gen *trnL - trnF*

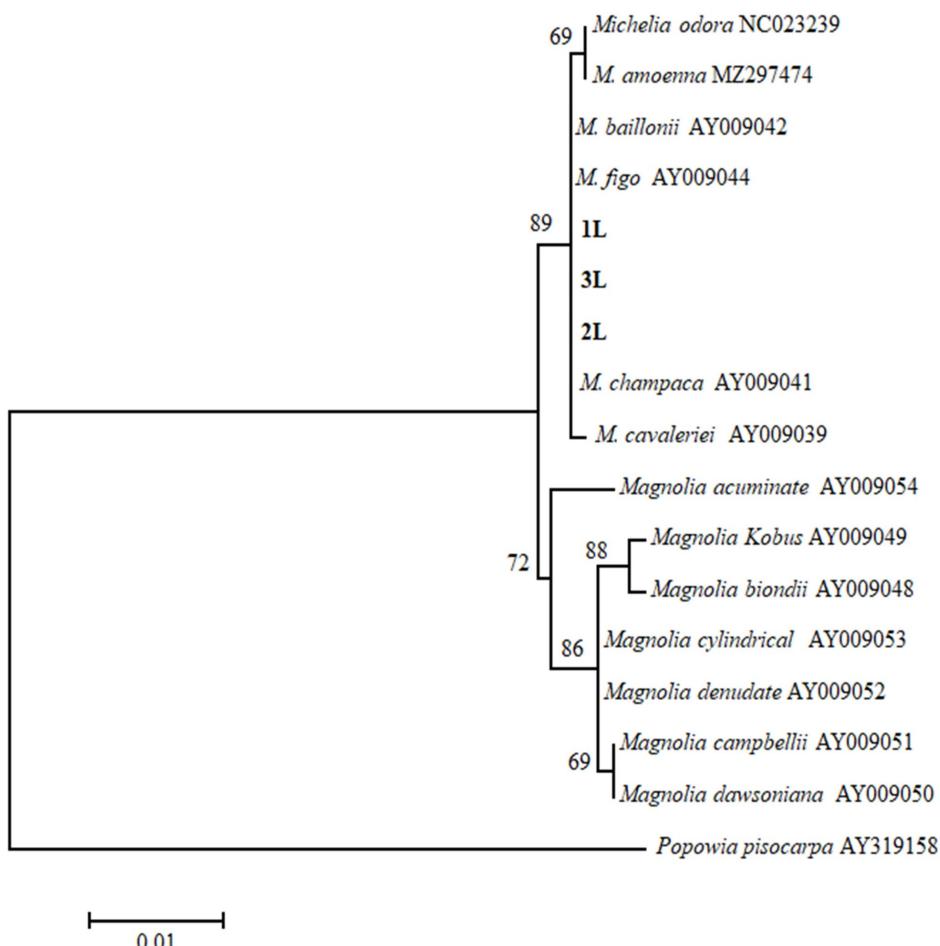
	1	2	3	4	5	6	7	8	9
1. 1L MT176165	-								
2. 2L MT176166	0,000	-							
3. 3L MT176167	0,000	0,000	-						
4. <i>Michelia figo</i> AY009044	0,000	0,000	0,000	-					
5. <i>M. baillonii</i> AY009042	0,000	0,000	0,000	0,000	-				
6. <i>M. champaca</i> AY009041	0,000	0,000	0,000	0,000	0,000	-			
7. <i>M. cavaleriei</i> AY009039	0,001	0,001	0,001	0,001	0,001	0,001	-		
8. <i>M. odora</i> NC023239	0,001	0,001	0,001	0,001	0,001	0,001	0,002	-	
9. <i>M. amoenna</i> MZ297474	0,001	0,001	0,001	0,001	0,001	0,001	0,002	0,000	-

Kết quả đã xác định được 3/886 vị trí nucleotide biến đổi, 2/886 vị trí nucleotide mang thông tin. Khoảng cách di truyền giữa các cặp loài trên cơ sở phân tích theo phương pháp p-distance đã chỉ ra mức độ khác nhau giữa các cặp loài trong chi *Michelia*. Kết quả cho thấy, đối với các loài trong chi *Michelia* hiện đã được công bố dựa trên vùng gen *trnL - trnF*, khoảng cách di truyền giữa các loài rất thấp dao động từ 0 đến 0,2%, điều này phản ánh mức độ tương đồng di truyền giữa các loài rất lớn. Không tìm thấy sự sai khác nào trong 3 mẫu nghiên cứu và giữa chúng với các loài *M. figo* (AY009044), *M. champaca* (AY009041) và *M. baillonii* (AY009042). Điều này sẽ rất khó để xác định loài nghiên cứu có quan hệ di truyền gần với loài nào nhất.

3.3. Mối quan hệ di truyền của 3 mẫu Giổi ăn hạt

Sơ đồ mối quan hệ di truyền của một số loài thuộc chi Giổi (*Michelia*) và chi Mộc lan (*Magnolia*) trong họ Mộc lan đã được xây dựng theo phương pháp ML (Hình 3). Phân tích ML đã tạo ra với thông số (-lnL) = 2713,778. Các mẫu trong nghiên cứu này và 06 loài trong chi *Michelia* tạo thành một nhóm riêng có mức độ tương đồng di truyền cao (98-100%) và có quan hệ mật thiết với nhau với giá trị bootstrap (MLBS = 89%). Do đó, khả năng sử dụng vùng gen này để nhận biết loài trong chi Giổi là không khả thi. Trong khi đó, các loài trong chi *Magnolia* cũng hình thành 1 nhánh tiến hóa riêng, tách riêng rẽ trên cây phân loại. Kết quả này cho phép nhận định 3 mẫu trong nghiên cứu có chung nguồn gốc với 6 loài *Michelia odora*, *M. amoenna*, *M. cavaleriei*, *M. figo*, *M. champaca* và *M. baillonii* trên thế giới.

Một số vùng gen được sử dụng cho mã vạch DNA của Magnoliaceae [19]. Huỳnh Thị Thu Huệ và cộng sự đã chỉ ra vùng gen *trnL-trnF* lục lạp có tính đa hình cao khi tiến hành phân tích trình tự DNA vùng *trnL-trnF* của loài Cà gai leo và xác định mối quan hệ di truyền giữa các loài trong chi *Solanum* [20]. Kết quả cho thấy khoảng cách di truyền dựa trên đoạn gen *trnL-trnF* của loài Cà gai leo với một số loài khác trong chi *Solanum* dao động khá lớn từ 0,82 tới 3,99%. Tuy nhiên, Deng và cộng sự đã giải mã toàn bộ vùng gen lục lạp của loài *Michelia shiluensis* và phân tích mối quan hệ di truyền giữa 26 loài của hai chi *Magnolia* và *Michelia* [13]. Tác giả đã chỉ ra hai loài *M. shiluensis* và *M. odora* trong chi *Michelia* có quan hệ rất gần gũi với nhau với bootstrap 100%. Điều đó khẳng định trên toàn bộ vùng lục lạp này cũng khó có thể nhận dạng được hai loài này. Tương tự, Wang và cộng sự chỉ ra rằng, hai loài *Manglietia cathcartii* và *Michelia odora* không tách biệt nhau về mặt di truyền khi phân tích hệ gen lục lạp của 26 loài thuộc họ Mộc lan [12]. Trong nghiên cứu này, kết quả của chúng tôi cũng đồng quan điểm với hai tác giả trên, trong vùng gen lục lạp (*trnL - trnF*) cũng không giải quyết được sự khác biệt di truyền giữa các loài trong chi *Michelia*.



Hình 3. Mối quan hệ họ hàng của mẫu nghiên cứu với các loài trong cùng chi trên cơ sở phân tích trình tự nucleotide vùng gen lục lạp (*trnL* - *trnF*) bằng phương pháp ML. Các số trên các nhánh là giá trị bootstrap

4. KẾT LUẬN

Vùng gen *trnL* - *trnF* của 3 mẫu Giổi ăn hạt được giải trình tự nucleotide có kích thước tương ứng là 855, 921 và 928 nucleotide. Các trình tự này đã được đăng ký trên GenBank góp phần xây dựng cơ sở dữ liệu mã vạch DNA cung cấp cho các nghiên cứu tiền hóa và hệ thống sinh học của loài.

Trong chi *Michelia* khoảng cách di truyền giữa một số loài rất thấp (0 - 0,2%) dựa trên vùng gen *trnL* - *trnF*. Hơn nữa, kết quả phân tích cũng chỉ ra các loài trong chi *Michelia* có cùng nguồn gốc tiền hóa và trình tự nucleotide vùng gen *trnL* - *trnF* không có khả năng nhận diện tên khoa học của các mẫu nghiên cứu và các loài trong chi *Michelia*.

Lời cảm ơn: Nghiên cứu này được tài trợ bởi Quỹ Phát triển khoa học và công nghệ Quốc gia (NAFOSTED) trong đề tài mã số 106.03-2017.16 cho thu mẫu tại một số sinh cảnh rừng tại Núi Luôt Trường ĐH Lâm Nghiệp Hà Nội và tại Lương Sơn - Hòa Bình; Nhóm tác giả xin chân thành cảm ơn Trung tâm nhiệt đới Việt - Nga, đề tài UBPH (E.1-2) nhiệm vụ số 03 đã hỗ trợ thu mẫu tại một số sinh cảnh rừng tại Xuân Liên, Thanh Hóa; Viện Công nghệ sinh học Lâm nghiệp, Trường Đại học Lâm nghiệp đã tạo điều kiện về cơ sở vật chất và phòng thí nghiệm cho thực hiện nghiên cứu này.

TÀI LIỆU THAM KHẢO

1. Vũ Quang Nam, Đào Ngọc Chuong, *Một số loài Giổi ăn hạt (Michelia sp.) ở Việt Nam*, Hội nghị quốc gia về Sinh thái và Tài nguyên sinh vật lần thứ 7, 2017, 283-288.
2. Do Tat Loi, *Medicinal plants and medicine of Vietnam*, Medical Publisher, 2006, Hanoi.
3. Vu Q. N, Xia N. H., *Notes on the type of Michelia tonkinensis (Magnoliaceae) from Vietnam*, J. Trop. Subtrop. Bot., 2011, **19**(6):549-553.
4. Hebert P. D. N., Penton E. H., Burns J. M., Janzen D. H., Hallwachs W., *Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astrape fulgerator*, PNAS, 2004, **101**:14812-14817.
5. Liu J, Provan J, Gao L. M, Li D. Z, *Sampling strategy and potential utility of indels for DNA barcoding of closely related plant species: A case study in Taxus*, Int. J. Mol. Sci., 2012, **13**:8740-8751.
6. Liu Z, Zeng X, Yang D, Ren G, Chu G, Yuan Z, Luo K, Xiao P, Chen S., *Identification of medicinal vines by ITS2 using complementary discrimination methods*, J. Ethnopharmacol., 2012, **141**:242-249.
7. Chen S. et al., *Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species*, PLoS One, 2010, **5**:e8613.
8. Gao T., Yao H., Song J., Liu C., Zhu Y., Ma X., Pang X., Xu H., Chen S., *Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2*, J. Ethnopharmacol., 2010, **130**:116-121.
9. Hollingsworth P. M. et al., *A DNA barcode for land plants*, PNAS, 2009, **106**(31):12794-12797.
10. Li D. Z. et al., *Comparative analysis of a large dataset indicates that the internal transcribed spacer (ITS) shoud be incorporated into the core barcode for seed plants*, PNAS, 2011, **108**:19641-19646.
11. Huan H. V., Trang H. M., Toan N. V., *Identification of DNA barcode sequence and genetic relationship among some species of Magnolia Family*, Asian J. Plant Sci., 2018, **17**:56-64.
12. Wang G., Hou N., Zhang S., Luo Y., *Characterization of the complete chloroplast genomes of seven Manglietia and one Michelia species (Magnoliales: Magnoliaceae)*, Conservation Genet. Resour., 2017, **10**:705-708.

13. Deng Y., Luo Y., He Y., Qin X., Li C., Deng X., *Complete chloroplast genome of Michelia Shiluensis and a comparative analysis with four Magnoliaceae species*, Forests, 2020, **11**(3):267.
14. Taberlet P., Gielly L., Pautou G., Bouvet J., *Universal primers for amplification of three non-coding regions of chloroplast DNA*, Plant Mol. Biol., 1991, **17**:1105-1109.
15. Hall T. A., *BioEdit v7.0.5.2: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*, Nucleic Acids Symp. Ser., 1999, **41**:95-98.
16. Jobb G., *TREEFINDER version March 2011*. <http://www.treefinder.de>.
17. Tanabe A. S., *Kakusan 4 and Aminosan: two programs for comparing nonpartitioned, proportional and separate models for combined molecular phylogenetic analyses of multilocus sequence data*, Mol. Ecol. Resour., 2011, **11**:914-921.
18. Kumar S., Stecher G., Tamura K., *MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets*, Mol. Biol. Evol., 2016, **33**(7):1870-1874.
19. Yu H. et al., *Expedient identification of Magnoliaceae species by DNA barcoding*, Plant Omics, 2014, **7**:47.
20. Huỳnh Thị Thu Huệ, Nguyễn Thị Thanh Hoa, Lê Thị Thu Hiền, Nguyễn Đăng Tôn, *Phân tích vùng gen trnL - trnL trên cây Cà gai leo (Solanum procumbens Lour.) của Việt Nam*, Tạp chí Công nghệ Sinh học, 2021, **19**(2):309-319.

Nhận bài ngày 30 tháng 7 năm 2022

Phản biện xong ngày 08 tháng 10 năm 2022

Hoàn thiện ngày 05 tháng 11 năm 2022

⁽¹⁾ Viện Công nghệ Hau, Đại học Công nghiệp Hà Nội

⁽²⁾ Viện Sinh thái nhiệt đới, Trung tâm Nhiệt đới Việt - Nga

⁽³⁾ Trường Đại học Lâm nghiệp, Phân hiệu Đồng Nai

⁽⁴⁾ Viện Sinh thái và Tài nguyên Sinh vật, Viện Hàn lâm KH&CNVN

⁽⁵⁾ Viện Công nghệ sinh học Lâm nghiệp, Trường Đại học Lâm nghiệp

Liên hệ: **Vũ Đình Duy**

Viện Sinh thái nhiệt đới, Trung tâm Nhiệt đới Việt - Nga

Số 63 Nguyễn Văn Huyên, Nghĩa Đô, Cầu Giấy, Hà Nội

Email: duyvu@vnmn.vast.vn