

SỬ DỤNG MÃ VẠCH DNA TRONG VIỆC ĐỊNH LOẠI LOÀI CÂY ĐƯỢC LIỆU THẮT DIỆP NHẤT CHI HOA Ở VIỆT NAM

Vũ Thị Thu Thủy, Nguyễn Thị Thu Nga*,
Hoàng Phú Hiệp, Chu Hoàng Mậu
Trường Đại học Sư phạm - ĐH Thái Nguyên

TÓM TẮT

Sự khác biệt về trình tự DNA barcode của đa số các loài thực vật là rất rõ ràng, do đó giải trình tự DNA ngắn được sử dụng làm mã vạch cho các loài thực vật hữu hạn sẽ cung cấp một công cụ giám định loài chính xác, hiệu quả và định loại được với cả các mẫu không nguyên vẹn, mẫu con non khó định loại bằng hình thái. Hiện nay, vùng gen *matK* được coi là một trong những vùng gen chuẩn trong xây dựng mã vạch DNA để nhận dạng loài và được công nhận bởi tổ chức barcode quốc tế (<http://www.barcodeoflife.org>). Trong bài báo này, chúng tôi trình bày kết quả sử dụng trình tự gen *matK* để định loại 2 mẫu cây Thắt diệp nhất chi hoa thu thập tại Hà Giang và Lào Cai.

Từ khóa: DNA barcode, gen *matK*, giám định loài, mã vạch, thắt diệp nhất chi hoa.

ĐẶT VẤN ĐỀ

Phương pháp phân loại hình thái có lịch sử phát triển lâu đời và đã xây dựng được một hệ thống phân loại sinh vật nói chung và thực vật nói riêng tương đối đầy đủ và toàn diện. Phương pháp phân loại này chủ yếu dựa vào sự khác biệt về hình thái của các cơ quan trong cơ thể thực vật, đặc biệt là cơ quan sinh sản (hoa). Tuy nhiên, phương pháp này cũng gặp rất nhiều khó khăn khi cần xác định những mẫu vật đang trong giai đoạn phát triển (chưa ra hoa), những mẫu có đặc điểm giống nhau do cùng thích nghi với điều kiện môi trường, hoặc khó nhận biết do có điểm tương đồng ở bậc phân loại thấp như dưới loài.

Từ giữa những năm 1990, với sự phát triển mạnh mẽ của sinh học phân tử, một phương pháp nghiên cứu mới trong lĩnh vực phân loại học đã hình thành và được gọi là phương pháp phân loại học phân tử. Phương pháp này dựa trên các dữ liệu thông tin về hệ gen trong và ngoài nhân hoặc các sản phẩm của chúng (protein). Tùy mục đích hoặc đối tượng nghiên cứu, người ta có thể lựa chọn các gen khác nhau hoặc các sản phẩm khác nhau của hệ gen [6]. Năm 2003, Paul và cs, nhà nghiên cứu tại Đại học Guelph ở Ontario, Canada, đề

xuất "mã vạch DNA" (DNA barcode) như là một cách để xác định loài. Mã vạch được sử dụng là một đoạn DNA ngắn từ một phần của hệ gen và được dùng giống như cách một máy quét ở siêu thị phân biệt được các sản phẩm bằng cách nhận diện được các sọc màu đen đặc trưng của từng sản phẩm. Trong công nghệ mã vạch, có thể hai mặt hàng trông rất giống nhau và không phân biệt được bằng mắt thường, song qua mã vạch, máy quét có thể phân biệt được. Một mã vạch DNA điển hình phải đáp ứng được các yêu cầu sau: (i) có tính phổ biến cao để có thể thực hiện trên nhiều loài thực vật; (ii) trình tự có tính đặc hiệu cao và có hiệu suất nhân bản cao; (iii) có khả năng phân biệt đồng thời được nhiều loài [5].

Nếu các gen ti thể như *COI* và *Cytb* được dùng rộng rãi cho động vật và một số loài tảo, thì khi chúng được áp dụng cho các loài thực vật trên cạn lại biểu hiện tính bảo thủ cao và vì vậy không phù hợp làm DNA mã vạch. Các vùng rời rạc trong hệ gen lập thể đã được dùng trong các nghiên cứu phát sinh loài (như các vùng exon của các gen *rbcL*, *atpB*, *ndhF* và *matK* và vùng intron của các gen *trnL* và *trnL-F*). Một vùng trình tự thông thường khác cho nghiên cứu phát sinh loài thực vật trên cạn là ribosome ITS nhân (vùng đệm của tiểu đơn vị lớn DNA ribosome). Tính đến năm

* Tel 0976714982; Email nguyenthuhung-ktsinh@dhspn.edu.vn

2009, đã có khoảng 8 locus gen được sử dụng làm mã vạch DNA ở các loài thực vật, bao gồm cả hệ gen nhân và hệ gen lục lạp [3].

Mã vạch DNA thực vật có khả năng cung cấp cái nhìn sâu vào phân loại cấp độ loài trong những nhóm có hình thái đơn giản, nhóm có phân bố rất rộng, nhóm có kích thước nhỏ, và/hoặc những nhóm đã được phân loại nhưng không đầy đủ, chưa tương xứng với đặc điểm đa dạng của chúng (ví dụ như một số trường hợp không thể dựa trên nguyên tắc phân loại hình thái học để đánh giá, phân loại, hoặc đã được phân loại nhưng chưa thực sự rõ ràng) [7].

Thực vật dùng làm thuốc luôn cần được xác định ở cấp độ loài, vì vậy, xác định chính xác là một bước quan trọng để có thể đảm bảo về chất lượng sản phẩm và có tầm quan trọng trong việc nghiên cứu các đặc tính của sự đa dạng di truyền, phát sinh loài và phát sinh vùng địa lý cũng như bảo vệ các loài có nguy cơ tuyệt chủng. Cùng với sự phát triển của thị trường thảo dược, sự giả mạo các nguyên liệu thuốc thảo dược cũng trở thành vấn đề toán câu. Các nguyên liệu thảo dược này có thể được thay thế bằng các loại thảo mộc khác có quan hệ họ hàng gần gũi, thậm chí từ những nguyên liệu giả mạo. Việc giả mạo các nguyên liệu thảo dược thường là do: (i) vật liệu không phân biệt được bằng đặc điểm hình thái (ii) những vật liệu có tên tương tự nhau, và (iii) việc thay thế những nguyên liệu có giá trị kinh tế bằng nguyên liệu khác rẻ tiền hơn [8], [9].

Tuy nhiên, việc xác định chính xác các nguyên liệu thảo dược làm thuốc theo phương pháp truyền thống như sự đánh giá cảm quan và phương pháp hóa học đôi khi gặp nhiều khó khăn, đặc biệt là những nguyên liệu có nguồn gốc từ thực vật đã được chế biến một phần hoặc ở dạng bột. Vì vậy, phương pháp sử dụng các dấu chuẩn phân tử rõ ràng là chính xác và phổ biến hơn. Việc nhận biết các nguyên liệu thảo dược sử dụng phương pháp mã vạch DNA có thể bảo vệ người dùng tránh khỏi tác dụng độc hại của các loại thuốc giả mạo, đặc

biệt trong nhiều trường hợp có thể nguy hiểm đến tính mạng [8].

Maturase K (*matK*) là một trong những trình tự được nghiên cứu và ứng dụng thành công trong việc xác định các loại thuốc thảo dược như "Dahuang" có nguồn gốc từ *Rheum palmatum* L (Polygonaceae), *R. tanguticum* (Maxim. ex Regel) Maxim. ex Balf, *R. officinale* Baill., và loài gần gũi *Rheum* L. với mức độ biến đổi nội bộ loài và giữa các loài khác nhau là cao. Vì vậy, nó thường được sử dụng để xác định các nguyên liệu thảo dược ở những vị trí địa lý khác nhau [6], [10].

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Vật liệu thực vật

Các mẫu cây Thất diệp nhất chi hoa sử dụng trong nghiên cứu do nhóm nghiên cứu thu thập trong các chuyến đi thực tế tại Hà Giang và Lào Cai.

Bảng 1. Địa điểm thu mẫu và thông tin mẫu sử dụng trong nghiên cứu

Nhóm mẫu	Địa điểm thu mẫu	Kí hiệu mẫu
Mẫu 6 lá	Thôn Phìn Hồ, tỉnh Hà Giang	P9.1 <i>matK</i>
Mẫu 7 lá	Vườn quốc gia Hoàng Liên, tỉnh Lào Cai	P9.2 <i>matK</i>

Hóa chất

Các hóa chất sử dụng trong nghiên cứu được đặt từ các hãng uy tín đảm bảo về chất lượng và độ tin cậy như Qiagen, Fermentas, Sigma...

Thiết bị sử dụng trong nghiên cứu

Một số thiết bị chính được sử dụng trong nghiên cứu này gồm: cân điện tử Satorius (Thụy Sĩ), bể ổn nhiệt, máy đo quang phổ, máy đo pH, máy li tâm, máy PCR, máy xác định trình tự gen.

Cập mỗi sử dụng trong nghiên cứu

Cập mỗi được sử dụng để nhân bản đoạn gen đích là gen *matK*, trình tự và thông tin về cập mỗi được thể hiện ở bảng 2.

Bảng 2. Thông tin về cặp mồi sử dụng trong nghiên cứu

Tên mồi	Trình tự mồi	Nhiệt độ gắn mồi	Kích thước dự kiến
matK-F	5'- CGATCTATTTCATTC AATATTTTC-3'	54°C	800bp
matK-R	5'- TCTAGCACACGAAAGTCGAAGT-3'		

Phương pháp nghiên cứu**Tách chiết DNA tổng số**

Các mẫu thất diệp nhất chi hoa được tách chiết DNA tổng số bằng quy trình mini-CTAB và kit DNeasy® Plant Mini Kit của hãng Qiagen.

Kiểm tra sản phẩm DNA sau tách chiết

DNA tổng số của các mẫu thất diệp nhất chi hoa sau khi tách chiết được điện di trên gel agarose 1%, trong đệm TBE 1X, ở hiệu điện thế 90V với marker 1kb. Nồng độ và độ tinh sạch của DNA tổng số được kiểm tra qua phương pháp đo quang phổ hấp phụ ở bước sóng 260 và 280nm. Pha loãng dịch chiết 100 lần (5µl mẫu + 450µl DDW), mỗi mẫu đo ba lần và giá trị TB của ba lần đo được lấy làm kết quả cuối cùng. Độ tinh sạch của mẫu thể hiện qua thông số A260/280 (1,6-2,0 được coi là mẫu tinh sạch).

Phương pháp nhân gen bằng kỹ thuật PCR

Phản ứng PCR được thực hiện trên máy PCR 9700.

Thành phần phản ứng PCR bao gồm: master mix (2X) – 7,5 µl, mỗi xuôi (10pmol/µl) - 0,5 µl, mỗi ngược (10 pmol/µl) - 0,5µl, DNA khuôn (10ng/µl)- 0,5 µl, H₂O - 6 µl. Tổng thể tích phản ứng là 15µl.

Chu trình nhiệt cho gen matK: 94°C/4 phút; lặp lại 35 chu kỳ với (94°C/30giây, 54°C/40giây, 72°C/40 giây); 72°C/10 phút và giữ ở 4°C.

Trình tự nucleotide của đoạn gen matK được xác định bằng máy giải trình tự ABI PRISM® 3100 Avant Genetic Analyzer, sử dụng bộ Kit BigDye® Terminator v3.1 Cycle Sequencing với cặp mồi đặc hiệu. Trình tự gen được phân

tích, so sánh và xây dựng cây phát sinh chủng loại bằng các chương trình Bioedit và DNASTar.

KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN**Kết quả tách chiết và tinh sạch DNA tổng số**

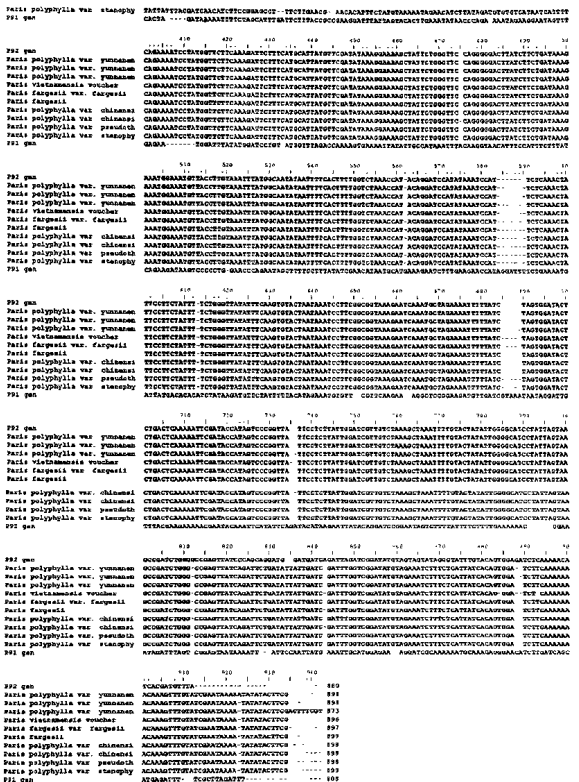
Để tách chiết DNA có độ tinh sạch nhất thì việc lựa chọn một phương pháp tách chiết phù hợp với đối tượng nghiên cứu là rất quan trọng. Hiện nay, phương pháp tách chiết DNA sử dụng CTAB là phương pháp khá phổ biến để tách chiết DNA từ các mẫu có nguồn gốc thực vật. Phương pháp sử dụng lần đầu tiên được Murray and Thompson mô tả vào năm 1980, tới nay đã được sử dụng rất phổ biến có hiệu quả với các mẫu khó tách chiết như thực vật giàu polysaccharides hay các sản phẩm có nguồn gốc từ thực vật.

Chúng tôi đã tách chiết thành công DNA của 2 mẫu lá từ cây Thất diệp nhất chi hoa thu tại Hà Giang và Lào Cai. Kết quả thí nghiệm được thể hiện trên hình 1.



Hình 1. Kết quả điện di DNA tổng số của 2 mẫu lá cây Thất diệp nhất chi hoa nghiên cứu

Chất lượng DNA thu được sau tách chiết và tinh sạch ảnh hưởng rất lớn đến kết quả phân tích nghiên cứu. Do vậy sau khi tách chiết DNA, chúng tôi kiểm tra chất lượng DNA thu được bằng phương pháp điện di và đo quang phổ hấp phụ. Kết quả sau khi điện di trên gel agarose 0,8% (hình1) cho thấy DNA thu được



Hình 3. Kết quả phân tích trình tự gen *matK* mẫu P91, P92 và so sánh với một số trình tự được công bố trên ngân hàng gen

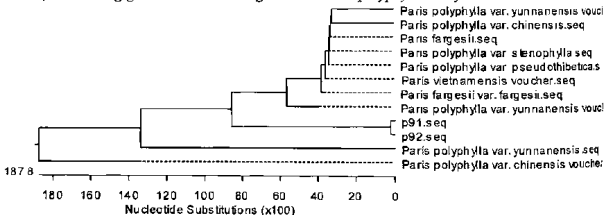
Phân tích sự đa hình trình tự DNA trên vùng gen *matK*

Align trình tự DNA vùng gen *matK* của mẫu nghiên cứu P9.1 cùng với 10 trình tự DNA vùng gen *matK* của các loài thuộc chi *Paris* đã được công bố trên GenBank bằng phần mềm Bioedit nhận thấy, chiều dài vùng gen *matK* giữa các loài dao động từ 802 đến 900 bp, GC trong vùng này dao động trong khoảng từ 31,85% đến 33,17%, có 101 điểm sai khác giữa mẫu nghiên cứu với các loài đem so sánh, chiếm 12,50% trên toàn bộ trình tự. Trong đó sai khác trong nội loài từ 0,0-0,37%. Có thể thấy sự đa hình trong trình tự nucleotide là khá cao, tại đây xảy ra các biến đổi như thêm hoặc mất một vài nucleotide. Vùng từ nucleotide 32 đến 62; 64 đến 121; 123 đến 785 trên trình tự gen đem so sánh (P9.1) là các vùng bảo thủ cao, với sự biến đổi rất ít trong trình tự ở tất cả các loài, điều này cũng phù hợp với các nghiên cứu trước đó [6], [8], [9]. Tương tự, chúng tôi so sánh trình tự DNA vùng gen *matK* của mẫu nghiên

cứ (P9.2) cho thấy, chiều dài vùng gen *matK* giữa các loài dao động từ 801 đến 900 bp, GC trong vùng này dao động trong khoảng từ 31,85% đến 32,96%, có 51 điểm sai khác giữa mẫu nghiên cứu với các loài đem so sánh, chiếm 6,70% trên toàn bộ trình tự. Trong đó sai khác trong nội loài từ 0,0-0,25%. Vùng từ nucleotide 50 đến 740 trên trình tự gen P9.2 là vùng bảo thủ cao, với sự biến đổi rất ít trong trình tự ở tất cả các loài [6], [8], [9].

Xây dựng cây phát sinh chủng loại

Cây phát sinh chủng loại cho thấy mẫu nghiên cứu P9.1; P9.2 thuộc cùng một loài, nằm cùng nhánh gần hơn với *Paris polyphylla* var. *yunnanensis* voucher, và nằm xa hơn một chút với nhánh *Paris polyphylla* var. *yunnanensis* seq, điều này phù hợp với kết quả phân tích tính đa dạng trong trình tự vùng gen *matK*. Với kết quả so sánh trình tự ở trên, có thể tạm xác định và xếp các mẫu nghiên cứu của chúng tôi (P9.1; P9.2) thuộc loài *Paris polyphylla* var. *yunnanensis*.



Hình 4. Sơ đồ phân loại hình cây được xây dựng dựa trên trình tự gen *matK* của mẫu P9.1; P9.2 bằng phần mềm DNASTar

KẾT LUẬN

Từ các kết quả thu được, chúng tôi đưa ra một số kết luận chính như sau:

1. Đã khuếch đại thành công vùng gen *matK* của 2 mẫu Thất diệp nhất chi hoa thu thập được tại Hà Giang và Lào Cai. Kết quả phân tích sự đa hình trên mỗi vùng gen cho thấy vùng gen *matK* ở mẫu P9.1 có kích thước 808

nucleotide, mẫu P9.2 có kích thước 806 nucleotide.

2. Từ kết quả so sánh và phân tích trình tự DNA vùng gen *matK* có thể xác định các mẫu nghiên cứu (P9.1; P9.2) thuộc loài *Paris polyphylla* var. *yunnanensis*. Xây dựng sơ đồ phân loại hình cây dựa trên trình tự gen *matK* bằng phần mềm DNASTar về kết quả tương

đồng cho thấy độ đáng tin cậy của cây thu được. Xác định được các mẫu nghiên cứu thuộc loài *Paris polyphylla* var. *yunnanensis* (đối với mẫu P9.1 thu được ở Hà Giang, mẫu P9.2 thu được ở Lào Cai).

4. Từ kết quả phân tích trình tự DNA và sơ đồ phân loại hình cây, kết luận vùng gen *matK* có thể được sử dụng để phân biệt các dưới loài của chi *Paris*. Nghiên cứu này cùng các kết quả nghiên cứu trước đây trên thế giới khẳng định vùng gen *matK* có thể giúp nhận diện loài và dưới loài như một mã vạch phân tử, đồng thời công nghệ này có thể được dùng cho các nghiên cứu tiếp theo về tiến hóa phân tử và nghiên cứu bảo tồn nguồn gen của các loài được liệt kê.

TÀI LIỆU THAM KHẢO

1. Avise J. C. (1995), *Conservation Biology*, 9 (3): pp. 686-690.
2. Banks R. C., Cicero C., Dunn J. L., Rasmussen P. C., (2003), *The Auk*, 120, pp. 923-931.
3. Cucnoud P. S. V., Chatrou L. W., (2002), "Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid *rbcL*,

atpB, and *matK* DNA sequences", *Amer. J. of Bot.*, 89, pp. 132-144.

4. Gill F. B., Slikas B., (1992), *The Condor*, 94: pp. 20-28.
5. Hebert P. D. N., Cywinska A., Ball S. L., Ward J. R., (2003), "Biological identifications through DNA barcodes", *Proc. R. Soc. Lond. B Biol. Sci.*, 270, pp. 313-321.
6. Hilu K. W. L.H. (1997), "The *matK* gene: sequence variation and application in plant systematics", *Amer. J. of Bot.*, 84, pp. 830-839.
7. Hollingsworth P. M., Little D.P. (2011), "Choosing and using a Plant DNA Barcode", *PLoS*, 6(5).
8. Li M., Cao H., But P. P. H., and Shaw P.C. (2011), "Identification of herbal medicinal materials using DNA barcodes", *J. of System. and Evolution*, 49(3), pp. 271-283.
9. Mark Y. S., H.P.D.N. (2008), "Barcode of life", *Scientific American*, pp. 82-88.
10. Yang D. Y., F. H., Cai S. Q., Komatsu K. (2004), "Molecular analysis of Rheum species used as Rhei Rhizoma based on the chloroplast *matK* gene sequence and ITS application for identification", *Biological & Pharmaceutical Bulletin*, 27, pp. 375-383.
11. Web: www.barcodeoflife.org; <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

SUMMARY

USING *matK* DNA BARCODES TO IDENTIFY SPECIES OF MEDICINAL PLANT *Paris polyphylla* Sm. IN VIETNAM

Vu Thi Thu Thuy, Nguyen Thi Thu Nga*,
Hoang Phu Hiep, Chu Hoang Mau
College of Education - TNU

Differences in DNA barcode sequences of the majority of plants are very clear, so short DNA sequences used as a plant barcode promises to provide an tool to identify species correctly, effectively and can be applied with all kinds of patterns that are not complete and old enough to be defined by morphology. Currently, the *matK* area is considered one of the standard genetic area in building the correct gene DNA barcoding to identify species and recognized by international organizations barcode (<http://www.barcodeoflife.org>). In this paper, we present results using the *matK* gene sequence to determine two samples of *Paris polyphylla* Sm. Collected in Ha Giang and Lao Cai.

Keywords: *Barcodes, DNA barcode, matK genes, Paris polyphylla Sm., species identification.*

Ngày nhận bài: 05/12/2016; Ngày phản biện: 15/12/2016, Ngày duyệt đăng: 24/01/2017

Phản biên khoa học: TS. Nguyễn Thị Hải Yến - Trường ĐH Khoa học - ĐHTN