

ỨNG DỤNG MỘT SỐ PHẦN MỀM PHÂN TÍCH MÃ VẠCH DNA ĐỂ ĐỊNH DANH CÂY DƯỢC LIỆU

Vũ Thị Như Trang

Trường Đại học Y Dược - ĐH Thái Nguyên

TÓM TẮT

Định danh cây dược liệu bằng mã vạch DNA ngày càng phổ biến do tính chính xác cao. Để định danh dược bằng mã vạch DNA, các nhà nghiên cứu thường sử dụng các phần mềm máy tính giúp phân tích, so sánh trình tự DNA của mẫu nghiên cứu với các trình tự có sẵn trên Genbank. Do dữ liệu trình tự gen được công bố tăng nhanh về số lượng nên để tiến hành phân tích hiệu quả các dữ liệu này, các phần mềm cần phải dễ sử dụng, có khả năng tính toán nhanh, chính xác và cung cấp nhiều phương pháp thống kê khác nhau. Trong nhóm các phần mềm này thì ba phần mềm BLAST trong NCBI, BioEdit v7.0.5.2, MEGA 7 được sử dụng phổ biến nhất do đáp ứng được các yêu cầu trên. Bài báo giới thiệu ba phần mềm để định danh cây dược liệu với ví dụ minh họa là định danh mẫu Thổ nhân sâm (*Talinum paniculatum*) bằng mã vạch vùng ITS. Kết quả nghiên cứu của bài báo cho thấy quy trình phối hợp ba phần mềm để định danh cây dược liệu bằng DNA cho kết quả chính xác nhất.

Từ khóa: Phần mềm BLAST; BioEdit v7.0.5.2; MEGA 7; Cây dược liệu; mã vạch DNA; Thổ nhân sâm; vùng ITS

Ngày nhận bài: 14/4/2020; Ngày hoàn thiện: 08/6/2020; Ngày đăng: 11/6/2020

APPLICATION OF SOME DNA BARCODE ANALYSIS SOFTWARE TO IDENTIFY MEDICINAL PLANTS

Vu Thi Nhu Trang

TNU - University of Medicine and Pharmacy

ABSTRACT

The identification of medicinal plants with DNA barcodes is increasingly popular due to its high accuracy. To identify using DNA barcodes, researchers often use computer software to analyze and compare DNA barcodes of research samples with barcodes available on Genbank. Because the published gene sequence data increases rapidly in number, in order to conduct an effective analysis of these data, the software needs to be easy to use, able to calculate quickly, accurately and provide many statistical methods. In this group of software, three software programs, including BLAST software in NCBI, BioEdit v7.0.5.2, MEGA 7, are most commonly used due to meeting the above requirements. The article introduces three software programs to identify medicinal plants with illustrative examples of identifying the Jewels of Opar (*Talinum paniculatum*) pattern by ITS region barcodes. The research results of the paper show that the process of combining three software programs to identify medicinal plants by DNA gives the most accurate results.

Keywords: BLAST software; BioEdit; MEGA 7; medicinal plant; DNA barcodes; Jewels of Opar; ITS region

Received: 14/4/2020; Revised: 08/6/2020; Published: 11/6/2020

1. Mở đầu

Thực vật dùng làm thuốc cần được định danh chính xác ở cấp độ loài để có thể đảm bảo về chất lượng sản phẩm cũng như bảo tồn các nguồn dược liệu quý. Hiện nay, phương pháp định danh cây dược liệu cho mức độ chính xác cao, đặc biệt đối với các loài có quan hệ gần gũi, khắc phục được hạn chế của phương pháp hình thái so sánh, đó là phương pháp phân loại học phân tử [1]. Có nhiều công trình nghiên cứu trên thế giới sử dụng mã vạch DNA để định danh cây dược liệu. Chen và cs (2010) đã so sánh hiệu quả sử dụng 7 mã vạch DNA (*psbA-trnH*, *matK*, *rbcL*, *rpoC1*, *ycf5*, *ITS2* và *ITS*) để nhận diện một số loài cây dược liệu như cây xuyên tâm liên (*Andrographis paniculata*), cây thanh táo (*Gendarussa vulgaris*), cây Thổ nhân sâm (*Talinum paniculatum*)... [2]. Các loài trong chi nhân sâm được định danh bằng mã vạch *ITS*, *MatK* do Zuo và cs (2011) nghiên cứu [3],... Ở Việt Nam, hiện có một số công trình nghiên cứu định danh mẫu cây dược liệu bằng mã vạch DNA. Một số loài thuộc chi nhân sâm *Panax* L. đã được định danh bằng sử dụng mã vạch *ITS*, *matK*, *psbA-trnH* và *rbcL* [4]. Nguyễn Tiến Dũng và cs (2018) đã sử dụng mã vạch vùng *ITS* và *psbA-trnH* để định danh cây bầy lá một hoa [5]. *Hay công trình của Vũ Thị Như Trang và cộng sự định danh cây Thổ nhân sâm bằng mã vạch ITS, matK, rpoC1, rpoB* [6]...

Khi sử dụng mã vạch DNA để định danh cây dược liệu nói riêng và các loài khác nói chung, ta cần thực hiện các bước sau:

Bước 1: So sánh trình tự mã vạch DNA của mẫu cần định danh với các trình tự tương ứng lưu giữ trong Ngân hàng dữ liệu nhằm tìm kiếm chuỗi tương đồng và mức độ phân ly về cấu trúc của chuỗi phân tích với các chuỗi khác để xác định tên loài của mẫu đang nghiên cứu;

Bước 2: So sánh trình tự mã vạch DNA của mẫu cần định danh với một số trình tự cùng

loài, cùng chi trên genBank nhằm tìm ra điểm sai khác giữa các trình tự;

Bước 3: Xây dựng sơ đồ hình cây xác định mối quan hệ di truyền giữa các mẫu nghiên cứu với các loài cùng chi, họ.

Để hỗ trợ các bước của việc định danh loài bằng DNA, có nhiều phần mềm khác nhau như ClustalW2, BLAST, FASTA3x, DNASTART Lasergene, Geneious, PatternHunter, Bioedit, MEGA,... Do dữ liệu trình tự gen được công bố tăng nhanh về số lượng nên để tiến hành phân tích hiệu quả các dữ liệu này, các phần mềm phải đảm bảo yêu cầu dễ sử dụng, có khả năng tính toán nhanh, chính xác và cung cấp nhiều phương pháp thống kê khác nhau. Thực tế cho thấy, trong hầu hết các nghiên cứu về định danh, các nhà nghiên cứu thường dùng một phần mềm cho mỗi bước, cụ thể như: phần mềm BLAST thường được dùng trong bước 1, phần mềm Bioedit thường được dùng trong bước 2, phần mềm MEGA thường được dùng trong bước 3. Tuy nhiên, các phần mềm thường chỉ được nêu tên trong các nghiên cứu mà chưa được trình bày, hướng dẫn khai thác chi tiết để thu được kết quả mong muốn. Điều này gây khó khăn rất lớn cho các nhà nghiên cứu mới muốn thực hiện theo các bước trong nghiên cứu cũ để kiểm chứng hoặc thực hiện các bước cho nghiên cứu mới. Chính vì vậy, bài báo trình bày chi tiết về cách khai thác 3 phần mềm BLAST trong NCBI, BioEdit v7.0.5.2, MEGA 7 nhằm thực hiện đầy đủ 3 bước của quy trình định danh các loài bằng DNA, đồng thời minh chứng hiệu quả của các thuật toán trong việc định danh loài bằng DNA. Bố cục bài báo gồm 4 phần như sau: Phần 1 là mở đầu, phần 2 là giới thiệu về các phần mềm, phần 3 là ví dụ minh họa khai thác các phần mềm để định danh loài của mẫu Thổ nhân sâm dựa vào vùng *ITS*, cuối cùng là phần 4 kết luận.

2. Các phần mềm định danh loài bằng DNA

2.1. Phần mềm BLAST

Phần mềm BLAST (Basic Local Alignment Search Tool) được phát triển bởi Stephen Altschul, Warren Gish, David Lipman tại U.S. National Center for Biotechnology Information (NCBI), Webb Miller tại Đại học Bang Pennsylvania, và Gene Myers tại Đại học Arizona vào năm 1990 [7]. Phần mềm BLAST có thể được tìm thấy trên web (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) hoặc cài đặt trên máy tính. BLAST có chức năng tìm kiếm những bắt cặp trình tự có điểm số cao giữa chuỗi truy vấn và các chuỗi trong cơ sở dữ liệu bằng cách sử dụng phương pháp dựa trên kinh nghiệm (heuristic). Tốc độ và sự chính xác tương đối của BLAST là những cải tiến kỹ thuật quan trọng của các chương trình BLAST và điều đó cho thấy lý do vì sao công cụ này lại là công cụ tìm kiếm phổ biến nhất trong tin sinh học [8].

BLAST trả ra kết quả so sánh các trình tự nucleotide hoặc protein với cơ sở dữ liệu trên GenBank và mức độ trùng khớp có ý nghĩa về mặt thống kê. Nhờ có phần mềm BLAST mà ta có thể kết luận mối quan hệ về chức năng và tiến hóa giữa các trình tự cũng như xác định các thành viên trong chi, họ [9].

BLAST là một tập hợp các chương trình gồm: Nucleotide-nucleotide BLAST (blastn); Protein-protein BLAST (blastp); Position-Specific Iterative BLAST (PSI-BLAST); Nucleotide-protein 6-frame translation (blastx); Nucleotide-nucleotide 6-frame translation (tblastx); Protein-nucleotide 6-frame translation (tblastn); Large numbers of query sequences (MEGAbast) [8].

Tùy vào kết quả mong muốn mà nhà nghiên cứu có thể lựa chọn một hoặc một số phiên bản của phần mềm BLAST để sử dụng, trong đó biến thể được sử dụng phổ biến nhất là Nucleotide-nucleotide BLAST (blastn), đây cũng là biến thể được sử dụng trong bài báo này.

2.2. Phần mềm BioEdit

Phần mềm BioEdit được phát triển bởi Hall vào năm 1999. Phần mềm BioEdit được thiết

kế nhằm hỗ trợ các nhà khoa học và các kỹ thuật viên thí nghiệm xử lý các trình tự sinh học. Phần mềm tích hợp nhiều tính năng nâng cao và tùy chọn xuất, nhập dữ liệu liên quan đến các tác vụ thực hiện hay bắt cặp trình tự. Hầu hết các tính năng của BioEdit đều nhằm cải thiện hiệu suất của người dùng và tự động hóa các tác vụ nhất định như bắt cặp ClustalW hoặc tìm kiếm Blast. Bên cạnh đó, BioEdit còn cấp quyền truy cập tới GenBank để tải hoặc gửi các trình tự DNA. BioEdit sẽ giúp giảm bớt thời gian cần thiết để chỉnh sửa hay bắt cặp các trình tự, đồng thời còn cung cấp nhiều phương pháp bắt cặp thủ công [10].

Để phân tích mã vạch DNA định danh loài thì chức năng phân tích dữ liệu trình tự DNA của phần mềm BioEdit được nhiều nhà nghiên cứu sử dụng nhất, đây cũng là chức năng được giới thiệu trong bài báo này.

2.3. Phần mềm MEGA

Phần mềm MEGA (Molecular Evolutionary Genetics Analysis) cung cấp các công cụ để khám phá và phân tích chuỗi DNA và protein từ góc độ tiến hóa. Phần mềm MEGA được xây dựng phiên bản 1.0 đầu tiên bởi Sudhir Kumar, Koichiro Tamura và Masatoshi Nei tại Đại học bang Pennsylvania năm 1990 [11]. Phần mềm MEGA cũng liên tục được cập nhật và nâng cấp từ MEGA 1 đến MEGA X nhằm thực hiện các chức năng chính như sau: xử lý và chỉnh sửa dữ liệu trình tự, sắp xếp trình tự, cây phát sinh suy luận và ước tính khoảng cách tiến hóa. Phần mềm cho phép người dùng duyệt, chỉnh sửa, tóm tắt, xuất và tạo chú thích cho kết quả của nghiên cứu hay phân tích Realtime. Trong các phiên bản hiện có của phần mềm MEGA thì phần mềm MEGA 7 được nhiều nhà nghiên cứu lựa chọn sử dụng nhất. Phiên bản MEGA 7 cho phép người dùng lựa chọn nhiều cách để xây dựng cây phát sinh, bao gồm cây UPGMA, Maximum Parsimony, Neighbor-Joining và Maximum Likelihood. Mỗi lựa chọn có ưu và nhược điểm riêng và phù hợp cho mục đích cụ thể của người sử dụng [12]. Chức năng

được sử dụng nhiều nhất trong MEGA 7 là xây dựng sơ đồ hình cây xác định mối quan hệ di truyền của mẫu nghiên cứu với các loài cùng chi, họ. Đây cũng là chức năng được giới thiệu trong bài báo này.

3. Hướng dẫn sử dụng phần mềm định danh loài bằng mã vạch DNA

Để khai thác 3 phần mềm BLAST trong NCBI, BioEdit v7.0.5.2, MEGA 7 nhằm phân tích mã vạch DNA để định danh cây được liệu nói riêng và các loài nói chung, chúng tôi giới thiệu chi tiết quy trình 3 bước định danh mẫu Thổ nhân sâm bằng mã vạch vùng *ITS* như sau:

3.1. Vật liệu và phương pháp nghiên cứu

Phần mềm BLAST trong NCBI, phần mềm Bioedit v7.0.5.2, phần mềm MEGA 7 [13] và trình tự vùng *ITS* của 5 mẫu Thổ nhân sâm (*ITS-TN1*, *ITS-TN2*, *ITS-BG*, *ITS-HT*, *ITS-QN*) [6].

Sau khi giải trình tự vùng *ITS* của các mẫu Thổ nhân sâm, vùng *ITS* sẽ được xử lý bằng phần mềm BLAST trong NCBI, phần mềm Bioedit v7.0.5.2, phần mềm MEGA 7 để định danh các mẫu Thổ nhân sâm.

Trình tự nucleotide vùng *ITS* của các mẫu Thổ nhân sâm được so sánh với các trình tự đã có sẵn trên GenBank bằng phần mềm BLAST trong NCBI theo các bước sau:

Bước 1: Vùng *ITS* được lấy trình tự dưới dạng FASTA được lưu trong notepad.

Bước 2: Vào địa chỉ đường link: <https://www.ncbi.nlm.nih.gov/>, sau đó chọn BLAST thể hiện ở hình 1.



Hình 1. Chọn BLAST theo địa chỉ web

Bước tiếp theo, ta chọn Nucleotide Blast → chọn tập tin (chọn file trình tự vùng *ITS* đã lưu trong notepad) → Blast thể hiện ở hình 2.



(a)



(b)

Hình 2. Chọn Nucleotide Blast (a) và chọn tập tin (b) trên web NCBI

Trình tự vùng *ITS* của mẫu Thổ nhân sâm được hiệu chỉnh và so sánh với trình tự cùng loài, chi trên GenBank nhờ phần mềm Bioedit v7.0.5.2 theo các bước sau:

Bước 1: Vùng *ITS* và các trình tự cùng loài, chi có sẵn trên GenBank từ kết quả xử lý của phần mềm Blast được lấy dưới dạng FASTA và lưu trong notepad;

Bước 2: Tải các trình tự lên phần mềm Bioedit v7.0.5.2

Dùng lệnh File → New Alignment → File → Import → Sequence alignment file → Chọn trình tự đã lưu trong notepad, thể hiện ở hình 3.



Hình 3. Chọn trình tự trong phần mềm Bioedit v7.0.5.2

Bước 3: Chọn tất cả các trình tự dùng lệnh Accessory Application → ClustalW Multiple alignment → Run ClustalW → Chọn biểu tượng

(thay những vị trí nucleotide giống nhau bằng dấu "...") thể hiện ở hình 4.

Bước 4: Copy các trình tự đã so sánh sang file word dùng lệnh File → Graphic View → Edit → Copy Page as a Bitmap → Paste vào file word mới thể hiện ở hình 5.



(a)



(b)

Hình 4. Cách chọn (a) và so sánh các trình tự (b) trong phần mềm Bioedit v7.0.5.2



Hình 5. Cách lưu trình tự đã so sánh sang file word trong phần mềm Bioedit v7.0.5.2

Xây dựng cây phát sinh chủng loại bằng phương pháp Neighbor-Joining nhờ phần mềm MEGA 7 theo các bước sau:

Bước 1: Vùng *ITS* của mẫu nghiên cứu và các trình tự cùng loài, chi có sẵn trên GenBank từ kết quả xử lí của phần mềm Blast được lấy dưới dạng FASTA và lưu file dưới dạng có đuôi “fas” hoặc “fasta”.

+ Kích chuột phải → New → Text document → Ghi tên file;

+ Copy trình tự dưới dạng FASTA vào file theo mẫu “>Tên trình tự

Trình tự” ;

+ Sau khi copy xong trình tự thì đổi tên file từ đuôi “txt” sang “fas”.

Bước 2: Chạy phần mềm MEGA 7

+ Dùng lệnh Align → Edit/BuildAlignment → Create a new alignment → OK → DNA;

+ Dùng lệnh Insert sequences from MEGA/FASTA/Text/Sequencer files → Chọn file có đuôi “fas” hoặc “fasta” → Open;

+ Dùng lệnh Edit → Select all;

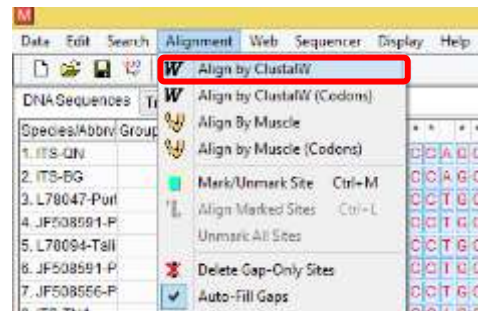
Kết quả các bước trên được thể hiện ở hình 6.



Hình 6. Kết quả tải các trình tự lên phần mềm MEGA 7 + Dùng lệnh Alignment → Align by ClustalW → OK;

+ Dùng lệnh Chọn biểu tượng Save để lưu file dưới dạng “max”;

Các bước trên được thể hiện ở hình 7.

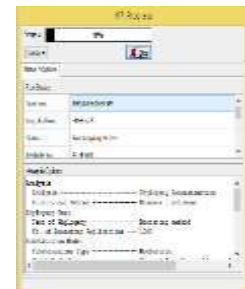


Hình 7. Cách so sánh các trình tự DNA

+ Dùng lệnh Phylogeny → Construct/Test Maximum Likelihood Tree → Chọn file đã lưu dưới dạng đuôi “max” → Compute, được thể hiện trên hình 8.



a. Bảng thông số M7: Analysis Preferences



b. Quá trình xây dựng sơ đồ hình cây

Hình 8. Bước thiết lập xây dựng sơ đồ hình cây trong MEGA 7

3.2. Kết quả và thảo luận

Sau khi sử dụng phần mềm BLAST trong NCBI với ví dụ minh họa là trình tự nucleotide vùng *ITS* của mẫu Thổ nhân sâm sẽ thu được kết quả là tỷ lệ tương đồng giữa các mẫu nghiên cứu với các mẫu có sẵn trên GenBank. Dựa vào kết quả này, có thể định danh loài đang nghiên cứu (Tên loài, chi).

Cụ thể là: bằng phần mềm BLAST trong NCBI cho thấy vùng *ITS* phân lập từ 5 mẫu nghiên cứu (*ITS-TN1*, *ITS-TN2*, *ITS-BG*, *ITS-HT*, *ITS-QN*) có tỷ lệ tương đồng là 99% với ba trình tự vùng *ITS* cùng loài *T. paniculatum*, mang mã số *JF508608*, *L78094*, *EU410357* trên GenBank; kết quả này đã khẳng định trình tự nucleotide phân lập được là vùng *ITS* thuộc loài *T. Paniculatum* [6].

Đồng thời, vùng *ITS* phân lập từ 5 mẫu nghiên cứu có tỷ lệ tương đồng 91% với trình tự vùng *ITS* của loài *T. fruticosum* cùng chi *Talinum*, mang mã số *KJ380908* trên GenBank; tương đồng 87% với trình tự vùng *ITS* của loài *Portulaca oleracea* mang mã số *L78047* trên GenBank trong cùng họ Rau sam (Portulacaceae) (Hình 9).



Hình 9. Kết quả xác định vùng *ITS* của các mẫu Thổ nhân sâm bằng BLAST trong NCBI

Trình tự nucleotide vùng *ITS* của mẫu Thổ nhân sâm sau khi xử lý bằng phần mềm Bioedit v7.0.5.2 sẽ thu được kết quả là tỷ lệ và vị trí sai khác nucleotide giữa các mẫu nghiên cứu với trình tự cùng loài, chi trên GenBank. Dựa vào kết quả này, có thể suy luận nguyên nhân của sự sai khác có thể do một dạng đột biến gen nào đó (Hình 10).

Từ kết quả ở hình 10 cho thấy, có 4 vị trí nucleotide sai khác giữa các trình tự của vùng *ITS* phân lập từ 5 mẫu Thổ nhân sâm (*ITS-TN1*, *ITS-TN2*, *ITS-BG*, *ITS-HT*, *ITS-QN*) với

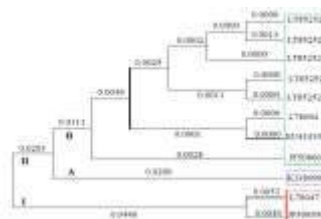
trình tự cùng loài mang mã số *EU410357*, đó là các vị trí nucleotide thứ 139, 148, 219 và 455, chiếm 0,62%, các vị trí sai khác có thể do đột biến thay thế cặp nucleotide. Tương tự khi so sánh trình tự vùng *ITS* phân lập từ 5 mẫu Thổ nhân sâm với loài *T. fruticosum* mang mã số *KJ380908* (cùng chi *Talinum*) cho thấy, các mẫu *ITS-BG*, *ITS-QN*, *ITS-TN2*, *ITS-TN1* có 37 vị trí sai khác nucleotide (chiếm 5,75%), mẫu *ITS-HT* có 38 vị trí sai khác nucleotide (chiếm 5,9%). Tất cả các vị trí nucleotide sai khác có thể do đột biến thay thế [6].



Hình 10. Kết quả một đoạn so sánh trình tự nucleotide của vùng *ITS* phân lập từ năm mẫu Thổ nhân sâm (*BG*, *TN2*, *QN*, *TN1*, *HT*) và mẫu đối sánh *EU410357* của loài *T. paniculatum* trên GenBank

Kết quả xác định mối quan hệ di truyền của trình tự vùng *ITS* phân lập từ các mẫu nghiên cứu và các trình tự vùng *ITS* mang mã số *EU410357*, *L78094*, *JF508608* cùng loài *T. paniculatum*, *KJ380908* (*T. fruticosum*) cùng chi *Talinum* và các trình tự gen cùng họ Rau sam như *L78047* (*P. oleracea*), *JF508556* (*Portulaca intraterranea*) được thể hiện ở hình 10.

Kết quả ở sơ đồ hình cây, các nhà nghiên cứu phải xác định số lượng nhánh được phân chia, tính toán khoảng cách di truyền giữa các chi trong họ, giữa các loài trong chi và trong cùng loài.



Hình 11. Sơ đồ hình cây thiết lập dựa trên trình tự nucleotide của vùng *ITS*

Sơ đồ hình cây ở hình 11 cho thấy, các đối tượng nghiên cứu phân bố trên 2 nhánh lớn, nhánh I có 2 trình tự vùng *ITS* và nhánh II có 9 trình tự vùng *ITS*, tất cả các trình tự thuộc hai chi *Talinum* và *Portulaca* cùng thuộc họ Rau sam (Portulacaceae) có độ tương đồng 88,4 - 89,2%, khoảng cách di truyền giữa hai chi *Talinum* và *Portulaca* là 4,4%. Nhánh II lại chia thành 2 nhánh phụ A và B. Nhánh A có một trình tự vùng *ITS* mang mã số *KJ380908* của loài *T. fruticosum* và nhánh B có 8 trình tự *ITS* cùng loài *T. paniculatum* với độ tương đồng là 94,1-100% và khoảng cách di truyền giữa hai loài *T. paniculatum* và *T. fruticosum* khoảng 2,37% [6].

4. Kết luận

Định danh cây dược liệu bằng mã vạch DNA ngày càng phổ biến do tính chính xác cao. Do dữ liệu trình tự gen được công bố tăng nhanh về số lượng nên để tiến hành phân tích hiệu quả các dữ liệu này, các phần mềm cần phải dễ sử dụng, có khả năng tính toán nhanh, chính xác và cung cấp nhiều phương pháp thống kê khác nhau. Bài báo đã giới thiệu chi tiết việc khai thác ba phần mềm BLAST trong NCBI, BioEdit v7.0.5.2, MEGA 7 nhằm thực hiện đầy đủ 3 bước của quy trình định danh cây dược liệu với ví dụ minh họa là các mẫu Thổ nhân sâm bằng mã vạch *ITS*. Đồng thời, kết quả của bài báo giúp các nhà nghiên cứu mới có thể sử dụng ba phần mềm để dễ dàng kiểm chứng các nghiên cứu khác hoặc thực hiện các bước cho nghiên cứu mới về định danh loài dựa trên DNA.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1]. H. Ledford, "Botanical identities: DNA barcoding for plants comes a step closer," *Nature*, vol. 451, pp. 451-616, 2008.
- [2]. S. Chen, H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao, X. Pang, K. Luo, Y. Li, X. Li, X. Jia, Y. Lin, and C. Leon, "Validation of the *ITS2* region as a novel DNA barcode for identifying medicinal plant species," *PLoS One*, vol. 5, no. 1, p. e8613, 2010.
- [3]. Y. Zuo, Z. Chen, K. Kondo, T. Funamoto, J. Wen, and S. Zhou, "DNA barcoding of *Panax* species," *Planta Med*, vol. 77, no. 2, pp. 182-187, 2011.
- [4]. T. H. Le, N. L. Nguyen, M. M. Bui, H. H. Ha, T. T. H. Huynh, V. H. Nong, V. H. Ha, and T. T. H. Le, "Application of DNA barcodes in identification of ginseng samples in the genus *Panax* L.," *Viet Nam Journal of Biotechnology*, vol. 15, no. 1, pp. 63-72, 2017.
- [5]. T. D. Nguyen, Q. N. Nguyen, N. L. Tran, T. T. Nguyen, T. P. Ninh, T. T. N. Doan, T. T. H. Le, and N. L. Nguyen, "Morphological characteristics and DNA barcodes of one-leaved seven-flowered tree, *Paris vietnamensis* (Takht.) H.li, in Vietnam," *Vietnam Journal of Agricultural Sciences- VJAS*, vol. 16, no. 4, pp. 282-289, 2017.
- [6]. T. N. T. Vu, M. T. Ho, V. S. Le, T. T. Nguyen, and H. M. Chu, "Morphological characteristics of *Talinum paniculatum*, and nucleotide sequences of *ITS* region, *rpoC1* and *rpoB* genes," *Viet Nam Journal of Biotechnology*, vol. 16, no. 3, pp. 451- 458, 2018.
- [7]. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, "Basic local alignment search tool," *J. Mol Biol*, vol. 215, no. 3, pp. 403-410, 1990.
- [8]. Wikipedia, "BLAST," 2020. [Online]. Available: <https://vi.wikipedia.org/wiki/BLAST> [Accessed May 30, 2020].
- [9]. Biological Journal, "Compare sequences - BLAST online", 2020. [Online]. Available: <https://tapchisinhhoc.com/so-sanh-trinh-tu-blast-online.html/>. [Accessed April 09, 2020].
- [10]. T. A. Hall, "BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT," *Nucl. Acids. Symp. Ser.*, vol. 41, pp. 95-98, 1999.
- [11]. S. Kumar, K. Tamura, and M. Nei, "MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers," *Comput Appl Biosci.*, vol. 10, no. 2, pp. 189-191, 1994.
- [12]. Biology of Vietnam, "Build phylogenetic tree easily with BLAST-MEGA", 2020. [Online]. Available: <https://sinhhocvietnam.com/xay-dung-cay-phat-sinh-loai-de-dang-voi-blast-MEGA/>. [Accessed April 09, 2020]
- [13]. S. Kumar, G. Stecher, and K. Tasma, "MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Mol Biol Evol*, vol. 33, pp. 1870 – 1874, 2016.