

# DỮ LIỆU HOÀN CHỈNH HỆ GEN LỤC LẠP CÂY SÂM ĐẤT CÔN ĐẢO ĐỊNH HƯỚNG ỨNG DỤNG NGHIÊN CỨU ĐA DẠNG SINH HỌC VÀ BẢO TỒN NGUỒN GEN

Trần Thị Thu Hồng<sup>1</sup>, Trần Thị Mộng Khoa<sup>2</sup>, Nguyễn Ngọc Trai<sup>3</sup>,  
Đỗ Hoàng Đăng Khoa<sup>4</sup>, Nguyễn Nhật Nam<sup>5\*</sup>

## A COMPLETE CHLOROPLAST GENOME OF POUZOLZIA SP. AND IMPLICATIONS FOR BIODIVERSITY AND CONSERVATION OF MEDICINAL PLANTS

Tran Thi Thu Hong<sup>1</sup>, Tran Thi Mong Khoa<sup>2</sup>, Nguyen Ngoc Trai<sup>3</sup>,  
Do Hoang Dang Khoa<sup>4</sup>, Nguyen Nhat Nam<sup>5\*</sup>

**Tóm tắt** – Sâm đất Côn Đảo (*Pouzolzia* sp.), còn được gọi là sâm rừng, sâm quy bầu và sâm nam, có giá trị dược liệu cao. Trong nghiên cứu này, hệ gen lục lạp của cây sâm đất Côn đảo được giải trình tự và phân tích. Hệ gen lục lạp có kích thước 153.176 cặp nucleotide được chia làm bốn vùng gồm: một vùng trình tự đơn lớn (kích thước 83.928 cặp nucleotide), một vùng trình tự đơn nhỏ (kích thước 18.648 cặp nucleotide), và hai vùng trình tự lặp đảo với kích thước mỗi vùng là 25.300 cặp nucleotide. Hệ gen lục lạp của cây sâm đất Côn Đảo bao gồm 79 gen liên quan mã hóa protein, 30 gen liên quan mã hóa tRNA và bốn gen liên quan mã hóa rRNA. Trong đó, tám gen liên quan mã hóa protein, bảy gen liên quan mã hóa tRNA và bốn gen liên quan mã hóa rRNA có thêm một bản sao do nằm trong vùng trình tự lặp đảo. Các gen có một intron cũng được ghi nhận. Trong đó, *clpP1* và *pafl* gen có hai intron. Các dữ liệu cơ bản về hệ gen lục lạp của cây sâm đất Côn Đảo làm tiền đề cho các nội dung thực hiện tiếp

theo về phát triển các chỉ thị phân tử, di truyền quần thể, bảo tồn giống và tiến hóa phân tử.  
**Từ khóa:** bảo tồn nguồn gen, dữ liệu hệ gen lục lạp, sâm đất Côn Đảo.

**Abstract** – ‘Sâm đất Côn Đảo’ (*Pouzolzia* sp.), also known as forest ginseng, sand ginseng, and southern ginseng, has high medicinal value. In this study, the chloroplast genome of *Pouzolzia* sp was sequenced and analyzed. Consequently, a quadripartite circular genome was completed and was 153,176 bp in length, including an LSC region of 83,928 bp, an SSC region of 18,648 bp, and two IR regions of 25,300 bp. The structure of CCG was analyzed and discussed. This genome contained 79 protein-coding genes, 30 tRNA genes, and four rRNA genes. Among the genes, eight protein-coding genes, seven tRNA genes, and four rRNA genes were duplicated because of being located in the inverted repeat regions. Genes with a single intron were also noted. Notably, the *clpP1* and *pafl* genes contain two introns. The results of this study provided fundamental information about the chloroplast genome of *Pouzolzia* sp. and initial data for further studies examining molecular markers, population genetics, conservation, and molecular evolution.

**Keywords:** chloroplast genomic data, genomic conservation, *Pouzolia*.

<sup>1,2,3,5</sup>Trường Đại học Trà Vinh, Việt Nam  
<sup>4</sup>Trường Đại học Nguyễn Tất Thành, Việt Nam  
Ngày nhận bài: 02/4/2024; Ngày nhận bài chỉnh sửa: 04/5/2024; Ngày chấp nhận đăng: 07/5/2024  
\*Tác giả liên hệ: [nnnam@tvu.edu.vn](mailto:nnnam@tvu.edu.vn)  
<sup>1,2,3,5</sup>Tra Vinh University, Vietnam  
<sup>4</sup>Nguyen Tat Thanh University, Vietnam  
Received date: 02<sup>nd</sup> April 2024; Revised date: 04<sup>th</sup> May 2024; Accepted date: 07<sup>th</sup> May 2024  
\*Corresponding author: [nnnam@tvu.edu.vn](mailto:nnnam@tvu.edu.vn)

## I. GIỚI THIỆU

Côn Đảo là huyện trực thuộc tỉnh Bà Rịa – Vũng Tàu, ở bờ biển Nam Bộ của Việt Nam [1]. Sự xuất hiện và phân bố các loài thực vật tại Côn Đảo rất đa dạng và phong phú với giống loài thực vật tự nhiên có giá trị. Vườn quốc gia Côn Đảo có hơn 800 loài thực vật rừng bậc cao thuộc hơn 560 chi và 160 họ. Trong đó, sâm đất Côn Đảo là loài thực vật có đặc tính thảo dược quý được quan tâm đầu tư và khai thác phát triển kinh tế.

Sâm đất Côn Đảo (SĐCĐ) thuộc chi thuộc vôi (*Pouzolzia*) của họ Tầm ma (*Urticaceae*), có nguồn gốc hoang dã [2]. Đây là loại sâm đã được phát hiện bởi những người tù chính trị vì thấy được những công dụng cũng như được tính của nó với khả năng chữa bệnh và hỗ trợ hiệu quả trong các vấn đề sức khỏe của người dân bản địa. Về đặc điểm hình thái, SĐCĐ có thân cao dài khoảng 50 cm, lá mọc đối xứng, có hình dạng là hình bầu dục, xung quanh mặt lá có lông tơ. Hoa của SĐCĐ có màu tím. SĐCĐ xuất hiện và được trồng rộng khắp xung quanh các hòn đảo ở đây. SĐCĐ được người dân sử dụng để bồi bổ cơ thể, giúp tăng cường miễn dịch. SĐCĐ được ghi nhận có khả năng hỗ trợ điều trị trong nhiều vấn đề sức khỏe khác nhau. Trong nghiên cứu y học, các hoạt chất chiết xuất từ họ này bao gồm các hoạt chất polyphenol, flavonoid, tanin, carotene, carotenoids, ascorbic và nhiều loại khoáng chất khác có khả năng kháng khuẩn và được nghiên cứu ứng dụng trong hỗ trợ điều trị nhiều căn bệnh khác nhau [3–5]. Tại Việt Nam, vấn đề bảo tồn nguồn gen tự nhiên của cây SĐCĐ chưa được quan tâm nghiên cứu. Dưới tác động của biến đổi khí hậu và việc khai thác lai tạo các giống cây bản địa dẫn đến nguy cơ bị mất hoặc thay đổi nguồn gen của một số giống vật nuôi, cây trồng. Những tác động này làm tăng nguy cơ các nguồn gen tự nhiên của giống cây bị thay đổi. Vì vậy, các dự án về bảo tồn nguồn gen, đặc biệt là các giống cây bản địa có giá trị văn hóa và kinh tế, cần được thực hiện nghiên cứu. Nội dung giải trình tự hệ gen lục lạp của SĐCĐ là cơ sở dữ liệu đầu tiên cung cấp các thông tin cơ sở cho công tác bảo tồn và các nghiên cứu chuyên sâu về tiến hóa, di truyền, và chỉ thị phân tử.

## II. TỔNG QUAN NGHIÊN CỨU

Hệ gen lục lạp (HGLL) là cấu trúc gen đặc biệt được quan tâm do có nhiều ứng dụng trong cơ sở di truyền khác như chuyển gen nội bào, nhân và lai tạo giống. Ngoài ra, HGLL còn là cơ sở quan trọng cho nhiều nội dung ứng dụng thực tiễn. Lục lạp là một trong hai bào quan có vật liệu di truyền riêng trong tế bào thực vật. Bộ gen lục lạp là DNA sợi đôi mạch vòng kín, kích thước dao động từ 120–220 kb [6, 7]. HGLL có cấu trúc ổn định và thành phần gen nhất quán, chứa thông tin di truyền phong phú. Ngoài ra, tính bảo tồn của các vùng gen lục lạp tương đối cao do tốc độ tiến hóa thấp và áp thấp trong chọn lọc tự nhiên. Sự khác biệt ở các vùng mã hóa và vùng không mã hóa trong sự tiến hóa phân tử đa dạng ở nhiều mức độ và trình tự mã hóa được bảo tồn nhiều hơn trình tự không mã hóa cho phép việc giải trình tự hệ gen lục lạp thuận tiện cho việc thực hiện phân tích và so sánh. Dựa trên cơ sở dữ liệu hệ gen này, lịch sử tiến hóa và mối quan hệ phát sinh giữa các loài thực vật đã được khám phá [8, 9]. Bên cạnh đó, các con đường chuyển hóa các chất cũng đã được tìm hiểu thông qua các dữ liệu hệ gen [12]. Các trình tự gen cũng đã cung cấp dữ liệu cần thiết cho các nghiên cứu phát triển các dấu phân tử để định danh các loài [10, 11].

Việt Nam là nơi phân bố của hơn 4.000 loài thực vật mang các nguồn hoạt chất thiên nhiên quý. Vì vậy, nhiều đơn vị nghiên cứu tiến hành khám phá và khai thác nguồn tài nguyên này bao gồm các công tác định danh loài được liệu mới, các nghiên cứu về hình thái, điều kiện sinh trưởng, thành phần và công dụng của các hoạt chất chính nguồn thảo dược. Tuy nhiên, nội dung nghiên cứu về HGLL không được triển khai nhiều tại Việt Nam. Năm 2020, một báo cáo về HGLL của một loài lan hài (*Paphiopedilum delenatii* Guillaumin) do nhóm nghiên cứu ở Việt Nam thực hiện [12]. Gần đây, các nghiên cứu liên quan về gen của sâm Ngọc Linh được triển khai bởi Viện Nghiên cứu hệ Gen – Viện Hàn lâm Khoa học Việt Nam [13]. Các nghiên cứu di truyền quần thể và chỉ thị phân tử của sâm Ngọc Linh được tập trung nghiên cứu [14]. Dữ liệu đưa ra các thông tin quan trọng về các nghiên cứu công nghệ gen trong công tác bảo tồn, xây dựng chỉ

thị phân tử và các nghiên cứu di truyền. Thông tin dữ liệu HGLL rất cần thiết cho các nghiên cứu mang tính chuyên sâu về sinh học phân tử và ứng dụng cho các cây dược liệu. Trong nghiên cứu này, nhóm tác giả tập trung xây dựng dữ liệu HGLL của cây SĐCĐ làm tiền đề và là cơ sở cho các nội dung thực hiện về hệ gen của SĐCĐ và các loài liên quan.

III. PHƯƠNG PHÁP NGHIÊN CỨU

A. Thu thập mẫu, tách chiết ADN và giải trình tự

SĐCĐ được thu mẫu nguyên cây, sau đó được trồng tại Trường Đại học Trà Vinh để lưu trữ mẫu và sử dụng. Mẫu lá tươi không bị sâu bệnh của cây SĐCĐ được thu thập. Mẫu lá được làm khô bằng cách để trong túi kín có chứa các hạt silica hút ẩm. Mẫu lá sau đó được dùng để tách DNA tổng bằng quy trình CTAB [15]. Mẫu DNA sau khi tách chiết sẽ được kiểm tra chất lượng bằng gel-agarose điện di và máy NanoDrop OneC (Thermo Fisher Scientific, USA) để đo nồng độ. Các mẫu DNA có băng rõ trên gel điện di và không có vệt kéo dài cùng với nồng độ ít nhất 100 ng/ $\mu$ L được đem giải trình tự bằng hệ thống máy Miniseq (Illumina, USA) với dữ liệu trình tự hai chiều với kích thước 151 bp.

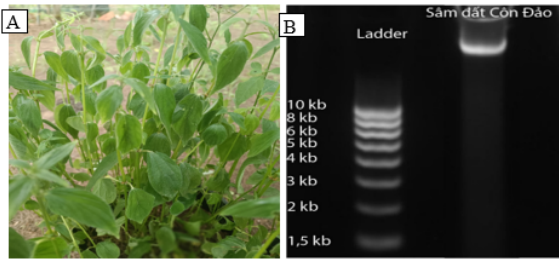
B. Lắp ráp và phân tích

Dữ liệu trình tự thô từ hệ thống Illumina sẽ được kiểm tra chất lượng (Q-score > 20, không chứa kí tự N và chiều dài lớn hơn 100 bp) và sàng lọc bằng chương trình fastp với các thiết lập mặc định [16]. Các trình tự đạt chất lượng sau khi lọc được dùng để lắp ráp HGLL bằng chương trình NOVOPlasty [17]. Sau đó, thành phần gen của HGLL của cây mù u được xác định bởi chương trình Geseq [18]. Các thành phần gen được kiểm tra lại bằng chương trình Geneious Prime v2023.2.

IV. KẾT QUẢ VÀ THẢO LUẬN

Mẫu DNA có băng rõ trên gel agarose, không có vệt kéo dài và kích thước lớn hơn 10 kb (Hình 1). Ngoài ra, kết quả đo nồng độ cho thấy mẫu DNA có nồng độ là 371 ng/ $\mu$ L, đạt chuẩn để áp dụng cho quá trình giải trình tự. Sản phẩm

của quá trình giải trình tự cho thấy tổng cộng có 29.616.018 trình tự với kích thước 151 bp được tạo ra. Các trình tự đều đạt về mặt chất lượng (Q-score > 20; không có chứa kí tự N và kích thước lớn hơn 100 bp) để sử dụng lắp ráp HGLL. Kết quả lắp ráp cho thấy có một DNA với trình tự hoàn chỉnh có cấu trúc tròn và kích thước 153.176 bp. Hệ gen này có bốn phần bao gồm một trình tự đơn lớn (kích thước 83.928 bp), một trình tự đơn nhỏ (kích thước 18.648 bp) và hai trình tự lặp đảo với kích thước 25.300 bp cho mỗi vùng. Thành phần GC của bộ gen này là 36,1%. Trong đó, thành phần GC của trình tự đơn lớn là 33,7%, trình tự đơn nhỏ là 29%, và trình tự lặp đảo là 42,6%.



Hình 1: (A) Hình ảnh cây sâm đất Côn Đảo được chọn thu nhận mẫu; (B) Kết quả điện di trên gel-agarose từ mẫu thu nhận (DNA tổng số)

Kết quả chú giải thành phần gen cho thấy có 79 gen mã hóa liên quan protein, 30 gen mã hóa liên quan RNA vận chuyển (tRNA) và bốn gen mã hóa liên quan RNA ribosome (rRNA) biểu thị ở Bảng 1. Trong số các gen, có 15 gen mang một intron, 02 gen mang hai intron, và 19 gen có hai bản sao do nằm trong vùng trình tự lặp đảo.

Quá trình tiến hành giải trình tự và phân tích HGLL của cây SĐCĐ cho thấy rằng hệ gen của loài này có độ tương đồng với các loài khác trong họ Tầm ma về cấu trúc, thành phần và vị trí các gen [19]. Các nghiên cứu trước đây cũng cho thấy tiềm năng khai thác và sử dụng HGLL để phát triển các chỉ thị phân tử như ở loài *Crepidiastrum denticulatum* và *Cenchrus longispinus* của họ Cúc và họ Hòa bản [20, 21]. Cây SĐCĐ là một loài dược liệu tiềm năng phát triển trong tương lai. Do đó, dữ liệu HGLL góp phần quan trọng trong quá trình xây dựng chỉ thị phân tử đặc trưng riêng cho cây sâm Côn Đảo.

Ngoài ra, dữ liệu HGLL có liên quan mật thiết đến di truyền của thực vật [22, 23] nên cung cấp thông tin thiết yếu cho quá trình lai tạo và chọn giống của cây SĐCĐ sau này.

Bảng 1: Thành phần các nhóm gen trong hệ gen lục lạp cây sâm đất Côn Đảo

| Nhóm gen                                | Tên gen                                                                                                                                                                                                                                                                                                                                                               |
|-----------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ARN ribosome                            | <i>rrn4.5(2x), rrn5(2x), rrn16(2x), rrn23(2x)</i>                                                                                                                                                                                                                                                                                                                     |
| ARN vận chuyển                          | <i>trnA-UGC, trnA-UGC*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-MCAU, trnG-GCC (2x), trnG-UCC*, trnH-GUG, trnI-CAU (2x), trnI-GAU* (2x), trnK-UUU*, trnL-AUU*, trnL-CAA (2x), trnL-UAG, trnM-CAU, trnN-GUU (2x), trnP-UGG, trnQ-UUG, trnR-ACG (2x), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC (2x), trnV-UAC*, trnW-CCA, trnY-GUA</i> |
| Hệ thống quang I                        | <i>psaA, psaB, psaC, psal, psaj, pafI*, pafII</i>                                                                                                                                                                                                                                                                                                                     |
| Hệ thống quang II                       | <i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbZ, pbfI, psbT, psbZ</i>                                                                                                                                                                                                                                                                 |
| Cytochrome                              | <i>petA, petB*, petD*, petG, petL, petN</i>                                                                                                                                                                                                                                                                                                                           |
| Enzyme ATP synthases                    | <i>atpA, atpB, atpE, atpF*, atpH, atpI</i>                                                                                                                                                                                                                                                                                                                            |
| Enzyme Rubisco                          | <i>rbcL</i>                                                                                                                                                                                                                                                                                                                                                           |
| Enzyme NADH dehydrogenase               | <i>ndhA*, ndhB(2x)*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>                                                                                                                                                                                                                                                                                         |
| Enzyme ATP-dependent protease subunit P | <i>clpP*</i>                                                                                                                                                                                                                                                                                                                                                          |
| Protein màng                            | <i>cemA</i>                                                                                                                                                                                                                                                                                                                                                           |
| Đơn vị lớn của ribosome                 | <i>rpl2(2x)*, rpl14, rpl16*, rpl20, rpl22, rpl23(2x), rpl32, rpl33, rpl36</i>                                                                                                                                                                                                                                                                                         |
| Đơn vị nhỏ của ribosome                 | <i>rps3, rps4, rps7(2x), rps8, rps11, rps12(2x)*, rps14, rps15, rps16*, rps18, rps19(2x)</i>                                                                                                                                                                                                                                                                          |
| Enzyme RNA polymerase                   | <i>rpoA, rpoB, rpoC1*, rpoC2</i>                                                                                                                                                                                                                                                                                                                                      |
| Yếu tố khởi đầu                         | <i>infA</i>                                                                                                                                                                                                                                                                                                                                                           |
| Các loại protein khác                   | <i>accD, ccsA, matK</i>                                                                                                                                                                                                                                                                                                                                               |
| Vùng mã hóa tiềm năng                   | <i>yefI(2x), yefJ2(2x)</i>                                                                                                                                                                                                                                                                                                                                            |

Ghi chú: \*: gen có intron; 2x: gen thuộc vùng trình tự lặp đảo.

V. KẾT LUẬN

HGLL hoàn chỉnh đầu tiên của cây SĐCĐ đã được giải trình tự và phân tích thành phần của gen. Các thông tin là cơ sở dữ liệu quan trọng, là tiền đề hữu ích cho các nghiên cứu tiếp theo nội dung về đa dạng di truyền cũng như di truyền quần thể và mối liên hệ phát sinh loài. Đồng thời, dữ liệu HGLL giúp phát triển chỉ thị phân tử đặc trưng riêng cho SĐCĐ, phục vụ công tác xây dựng chỉ dẫn địa lí của SĐCĐ thông qua dữ liệu HGLL để xác định chính xác nguồn gốc và kiểm định chất lượng sản phẩm thô từ cây SĐCĐ.

LỜI CẢM ƠN

Kết quả nghiên cứu được tài trợ bởi Trường Đại học Trà Vinh thông qua Hợp đồng số 149/2023/HĐ.HĐKH&ĐT-DHTV.

TÀI LIỆU THAM KHẢO

[1] Hoàng Tuyết Ngán. *Thông tin tổng quan huyện Côn Đảo*. <https://condao.com.vn/vi/news/Tin-tuc/thong-tin-tong-quan-huyen-con-dao-1676.html>. [Ngày truy cập: 24/3/2024]. [Hoang Tuyet Ngan. General information of Con Dao. <https://condao.com.vn/vi/news/Tin-tuc/thong-tin-tong-quan-huyen-con-dao-1676.html>. [Accessed 24<sup>th</sup> March 2024]]].

[2] Plants of the World Online. *Pouzolzia*. <https://powo.science.kew.org/taxon/urn:lsid:ipni.org:names:326038-2#children>. [Accessed 24<sup>th</sup> March 2024]].

[3] Lujun W, Die G, Qifeng F, Kai Z, Xhinng X. Bioactivity-guided isolation of antioxidant compounds from *Pouzolzia zeylanica* (L.) benn. *Pharmacognosy Magazine*. 2018;14(56): 444–450. [https://doi.org/10.4103/pm.pm\\_233\\_17](https://doi.org/10.4103/pm.pm_233_17).

[4] Nguyen DT, Vo TXT, Nguyen MT. Bioactive compounds, pigment content and antioxidant activity of *Pouzolzia zeylanica* plant collected at different growth stages. *Can Tho University Journal of Science*. 2018;54: 54–61. <https://doi.org/10.22144/ctu.jsi.2018.095>.

[5] Le THN, Nguyen HH, Bui HT, Nguyen XN, Phan VK. Norlignans from the aerial parts of *Pouzolzia sanguinea*. *Vietnam Journal of Chemistry*. 2020;58(4): 554–562. <https://doi.org/10.1002/vjch.202000034>.

[6] Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biology*. 2016; 17(1): 134–163. <https://doi.org/10.1186/s13059-016-1004-2>.

[7] Dobrogojski J, Adamiec M, Luciński R. The chloroplast genome: a review. *Acta Physiologiae Plant*. 2020;42(6): 98–111. <https://doi.org/10.1007/s11738-020-03089-x>.

[8] Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR, Soltis DE. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *Botanical Society of America*. 2018;105(3): 291–301. <https://doi.org/10.1002/ajb2.1048>.

[9] One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 2019;574(7780): 679–685. <https://doi.org/10.1038/s41586-019-1693-2>.

[10] Chakraborty P. Herbal genomics as tools for dissecting new metabolic pathways of unexplored medicinal plants and drug discovery. *Biochimie Open*. 2018;6: 9–16. <https://doi.org/10.1016/j.biopen.2017.12.003>.

[11] Grover A, Sharma PC. Development and use of molecular markers: past and present. *Critical Reviews in Biotechnology*. 2016;36(2): 290–302. <https://doi.org/10.3109/07388551.2014.959891>.

[12] Vu HT, Tran N, Nguyen TD, Vu QL, Bui MH, Le MT, et al. Complete Chloroplast Genome of *Paphiopedilum delenatii* and Phylogenetic Relation-

- ships among Orchidaceae. *Plants*. 2020;9(1): 61–78. <https://doi.org/10.3390/plants9010061>.
- [13] Vu D, Muhammad SS, Pham MP, Bui VT, Nguyen MT, Nguyen TPT. De novo assembly and transcriptome characterization of an endemic species of Vietnam, *Panax vietnamensis* Ha et Grushv., including the development of EST-SSR markers for population genetics. *BMC Plant Biology*. 2020;20(1): 358–374. <https://doi.org/10.1186/s12870-020-02571-5>.
- [14] Pham TN, Pham TH, Nguyen QN, Phan VT, Nguyen MK, Dinh DL. A molecular phylogeny of *Panax* L. Genus (Araliaceae) based on ITS-rDNA and matK support for identification of *Panax* species in Vietnam. *VNU Journal of Science: Medical and Pharmaceutical Sciences*. 2020;36(2): 91–99. <https://doi.org/10.25073/2588-1132/vnumps.4198>.
- [15] Schenk JJ, Becklund LE, Carey SJ, Fabre PP. What is the ‘modified’ CTAB protocol? Characterizing modifications to the CTAB DNA extraction protocol. *Applications in Plant Science*. 2023;11(3): 11517–11528. <https://doi.org/10.1002/aps3.11517>.
- [16] Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17): 884–974. <https://doi.org/10.1093/bioinformatics/bty560>.
- [17] Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research*. 2016;45(4): 18–27. <https://doi.org/10.1093/nar/gkw955>.
- [18] Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Ralph Bock R, et al. GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*. 2017;45: 6–11. <https://doi.org/10.1093/nar/gkx391>.
- [19] Wang RN, Milne RI, Du XY, Liu J, Wu ZY. Characteristics and mutational hotspots of plastomes in *Debregeasia* (Urticaceae). *Frontiers in Genetics*. 2020;11: 729–742. <https://doi.org/10.3389/fgene.2020.00729>.
- [20] Hyun JY, Do HDK, Jung J, Kim JH. Development of molecular markers for invasive alien plants in Korea: a case study of a toxic weed, *Cenchrus longispinus* L., based on next generation sequencing data. *Plant Biology*. 2019;7: 7965–7976. <https://doi.org/10.7717/peerj.7965>.
- [21] Do HDK, Jung J, Hyun JY, Yoon SJ, Lim C, Park K, et al. The newly developed single nucleotide polymorphism (SNP) markers for a potentially medicinal plant, *Crepidiastrum denticulatum* (Asteraceae), inferred from complete chloroplast genome data. *Molecular Biology Reports*. 2019;46(3): 3287–3484. <https://doi.org/10.1007/s11033-019-04789-5>.
- [22] Park HS, Lee WK, Lee SC, Lee HO, Joh HJ, Park JY, et al. Inheritance of chloroplast and mitochondrial genomes in cucumber revealed by four reciprocal F1 hybrid combinations. *Scientific Reports*. 2021;11(1): 2506–2517. <https://doi.org/10.1038/s41598-021-81988-w>.
- [23] Armbrust EV. Uniparental inheritance of chloroplast genomes. In: Rochaix JD, Goldschmidt-Clermont M, Merchant S. (eds.) *The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas. Advances in Photosynthesis and Respiration*, vol 7. Dordrecht: Springer Netherlands. 2013; p.93–113. [https://doi.org/10.1007/0-306-48204-5\\_6](https://doi.org/10.1007/0-306-48204-5_6).

