

Nghiên cứu xác định các vùng EST-SSR đặc trưng của loài sâm Ngọc Linh (*Panax vietnamensis* Ha & Grushv) bằng phương pháp giải trình tự gen thế hệ mới

Nguyễn Thị Phương Trang^{1,2*}, Nguyễn Hùng Mạnh¹, Bùi Thu Hà³

¹Viện Sinh thái và Tài nguyên Sinh vật, Viện Hàn lâm KH&CN Việt Nam

²Học viện KH&CN, Viện Hàn lâm KH&CN Việt Nam

³Khoa Sinh học, Trường Đại học Sư phạm Hà Nội

Ngày nhận bài 20/9/2021; ngày chuyển phản biện 24/9/2021; ngày nhận phản biện 22/10/2021; ngày chấp nhận đăng 27/10/2021

Tóm tắt:

Sâm Ngọc Linh (*Panax vietnamensis* Ha & Grushv) là loài đặc hữu, có phân bố hẹp. Do có nhiều công dụng nên giá thành sâm Ngọc Linh cao và hiện đã xuất hiện nhiều mẫu làm giả loài sâm này (mẫu giả được sử dụng là loài hoặc phân loài khác cùng trong chi sâm có phân bố ở Việt Nam). Vì vậy, việc nhận biết chính xác loài sâm Ngọc Linh là rất quan trọng. Trong nghiên cứu này, các tác giả thực hiện giải mã trình tự hệ gen phiên mã (cDNA) của sâm Ngọc Linh bằng công nghệ giải trình tự thế hệ mới (Next-generation sequencing) trên hệ thống Illumina HiSeq™ 4000 nhằm xác định các vùng SSR (Simple sequence repeat) đặc trưng của loài sâm này. Kết quả đã giải mã được 23.741.783 đoạn đọc với tổng khối lượng là 7,08 Gb và tổng hợp thành 262.999 gen chức năng (133 Mb). Trong đó, đã xác định được 101 vùng SSR lặp đặc trưng của sâm Ngọc Linh.

Từ khóa: ETS-SSR, giải trình tự, ITS, sâm Ngọc Linh (*Panax vietnamensis* Ha & Grushv).

Chỉ số phân loại: 1.6

Bật vấn đề

Sâm Ngọc Linh là loài đặc hữu, chỉ phân bố ở núi Ngọc Linh thuộc tỉnh Quảng Nam và Kon Tum của Việt Nam [1]. Các nghiên cứu đã có đều cho thấy, Sâm Ngọc Linh là loài có tính đa dạng di truyền cao và có hàm lượng saponin nhiều hơn hẳn so với các loài *Panax* khác [2]. Do phạm vi phân bố hẹp, tình trạng khai thác quá mức, khả năng phục hồi tự nhiên thấp nên các quần thể sâm Ngọc Linh tự nhiên hầu như đã không còn, hiện nay trên thị trường đang lưu hành đều là sâm Ngọc Linh trồng [3]. Do có nhiều công dụng nên giá thành sâm Ngọc Linh cao và hiện đã xuất hiện nhiều mẫu làm giả loài sâm này (mẫu giả được sử dụng thường là loài hoặc phân loài khác cùng trong chi *Panax* có phân bố ở Việt Nam) [4]. Vì vậy, việc nhận biết chính xác loài sâm Ngọc Linh trở nên khó khăn hơn. Phân loại học truyền thống chỉ dựa vào hình thái là khó thực hiện trong thực tế do việc thu mẫu không phải lúc nào cũng thu được mẫu đạt tiêu chuẩn (có đủ hoa, quả), vì vậy, cần có sự hỗ trợ của các kỹ thuật sinh học phân tử. Đặc biệt, đối với loài quý hiếm và thường xuyên bị làm giả như sâm Ngọc Linh.

SSR là những trình tự nucleotide ngắn, được lặp lại nhiều lần. Ưu điểm của SSR là chỉ thị đồng trội, có tính đặc hiệu thường được dùng trong các nghiên cứu về đa dạng di truyền ở cấp độ loài [5]. Các SSR phát triển từ hệ gen

phiên mã (EST-SSRs) có thể giúp tăng cường khả năng nhận dạng loài. Ngoài ra, các chỉ thị EST-SSR được thiết kế dành riêng cho một loài thường biểu thị mức độ đa hình đặc trưng do vị trí của chúng trong các vùng gen được xác định là đặc thù. Với tính đặc hiệu ưu việt đó mà đã có một số lượng lớn các EST-SSR được phát triển để giúp đánh giá cho công tác chọn giống, phân loại ở nhiều loài có giá trị kinh tế/xuất khẩu như khoai lang (*Ipomoea batatas*), củ cải (*Raphanus sativus*), địa lan (*Cymbidium sinense*), sâm Nga (*Panax ginseng*)...

Trong nghiên cứu này, chúng tôi thực hiện giải mã trình tự hệ gen biểu hiện của sâm Ngọc Linh bằng phương pháp giải trình tự gen thế hệ mới nhằm xác định các vùng SSR đặc trưng của loài sâm này. Việc xác định được các vùng SSR đặc trưng cho loài sẽ giúp công tác định danh chính xác hơn, thời gian thực hiện nhanh chóng và có chi phí rẻ hơn (ước tính chỉ bằng 1/10 so với phương pháp đọc trình tự gen hiện nay).

Vật liệu và phương pháp nghiên cứu

Vật liệu

Mẫu nghiên cứu là 1 cá thể sâm Ngọc Linh được thu tại vườn sâm gốc Tắc Ngo thuộc sự quản lý của UBND xã Trà Linh, huyện Nam Trà My, tỉnh Quảng Nam ở tọa độ:

*Tác giả liên hệ: Email: nprang@gmail.com

Identifying special EST-SSR regions of *Panax vietnamensis* Ha & Grushv based on Next-generation sequencing

Thi Phuong Trang Nguyen^{1,2*}, Hung Manh Nguyen¹,
Thu Ha Bui³

¹Institute of Ecology and Biological Resources, VAST

²Graduate University of Science and Technology, VAST

³Faculty of Biology, Hanoi National University of Education

Received 20 September 2021; accepted 27 October 2021

Abstract:

Panax vietnamensis var. *vietnamensis* Ha & Grushv is an endemic species with a narrow distribution. Because of having many benefits and high prices, there is an increase in faked Ngoc Linh ginseng. Most of the faked Ngoc Linh ginseng samples are other *Panax* species or subspecies distributed in Vietnam. Thus, it is very important to authenticate the Ngoc Linh ginseng samples. In this study, the author sequenced the transcriptomes genome of *Panax vietnamensis* var. *vietnamensis* based on the Next-generation sequencing method using Illumina HiSeq™ 4000 system to detect the special SSR regions of this ginseng. A total of 23,741,783 raw reads were obtained and assembled with a total volume of 7.08 Gb. From these raw reads, 262,999 unigenes with an average length of 133 Mb were generated and 101 typical SSR regions of Ngoc Linh ginseng were successfully identified.

Keywords: ETS-SSR, ITS, Ngoc Linh ginseng (*Panax vietnamensis* Ha & Grushv), sequencing.

Classification number: 1.6

15°1'51.92" vĩ độ bắc và 107°58'46.44" kinh độ đông, độ cao 1.920 m so với mực nước biển. Mẫu nghiên cứu trước khi tiến hành giải trình hệ gen biểu hiện được phân loại bằng hình thái kết hợp đọc trình tự vùng ITS để khẳng định chính xác mẫu thu là sâm Ngọc Linh.

Kiểm tra mẫu nghiên cứu bằng giải trình tự gen ITS với cặp mồi có trình tự sau: mồi xuôi (F): 5'-CCTTATCAGTTAGAGGAAGGAG-3'; mồi ngược (R): U4: 5'-CCGCTTAGTGATATGCTTAAA-3'.

Phương pháp nghiên cứu

Phương pháp so sánh hình thái: phân tích các đặc điểm hình thái theo Hoàng Thị Sản (2009) [6].

Xác định chính xác các mẫu thu được là sâm Ngọc Linh bằng đọc trình tự gen ITS.

Tách chiết DNA từ lá sâm Ngọc Linh theo phương pháp của J.J. Doyle và J.L. Doyle (1991) [7] có hiệu chỉnh theo điều kiện phòng thí nghiệm của Việt Nam.

Tách chiết mRNA tổng số và tổng hợp cDNA: mẫu lá sâm Ngọc Linh tươi thu tại vườn sâm Tắc Ngo sau khi đã được định danh chính xác là loài sâm Ngọc Linh được làm sạch bằng cồn 96° và sử dụng để tiến hành tách chiết mRNA tổng số. mRNA tổng số được tách bằng kit RNAeasy Mini Kit 250 của Qiagen. Dung dịch nước sử dụng trong tách chiết đều được xử lý với DEPC. Số lượng và chất lượng của RNA thu được được kiểm tra trên máy quang phổ Nanodrop ND-2000 (Thermo Electron Corporation, Mỹ).

cDNA sợi đơn và sợi đôi được tổng hợp từ RNA tổng số bằng kit Access RT-PCR với các thành phần gồm 4 µl 5X RT-Buffer, 2 µl dNTP (10 mM), 1 µl Inhibitor Rnase (20 u/µl), 1 µl Reverse transcriptase (20 u/µl), chu trình nhiệt được thực hiện trên máy PCR 9700 ở 25°C trong 5 phút, 45°C trong 60 phút và 70°C trong 5 phút.

Giải trình tự cDNA: sản phẩm cDNA sau đó được giải trình tự trên hệ thống máy Illumina HiSeq™ 4000 được thực hiện bởi Công ty First Base, Malaysia. Các lần đọc được hiệu chỉnh và lắp ráp bằng phần mềm Trinity [8].

Xác định các vùng lặp SSR trong hệ gen biểu hiện và tổng hợp mồi SSR: các SSR thích hợp là những đoạn lặp nucleotide ngắn (2-6 nucleotide), với đơn vị lặp lại tối thiểu được xác định là 6 lần cho di-nucleotide và 5 lần cho các trình tự cao theo phương pháp của Jurka và Pethiyagoda (1995) [9]. Phát hiện vùng SSR trong các khu vực chưa được dịch mã hoặc ở những vùng ORF trong hệ gen bằng công cụ SSRIT. Các đoạn mồi ETS-SSR được thiết kế bằng phần mềm Primer 5.0 [10] với các cài đặt mặc định để thiết kế mồi tạo ra các sản phẩm PCR có kích thước 100-300 bp. Các sản phẩm PCR được phân tách bằng hệ thống điện di

Sequi-Gen® GT trong gel polyacrylamide 8% trong đệm TAE và nhuộm Gelred™ 10000x.

Kết quả và bàn luận

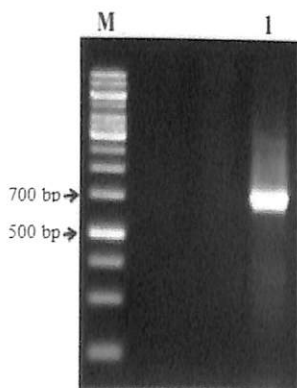
Định loại mẫu sâm Ngọc Linh thu được bằng đặc điểm hình thái

Sau khi phân tích những đặc điểm hình thái, so sánh với các tài liệu chuyên khảo về phân loại và với mẫu tiêu bản của loài được công bố chuẩn đã xác định được một số đặc điểm hình thái đặc trưng như sau: i) Cây thảo, có lá kèm, cuống lá không có bẹ, cụm hoa tán, hoa lưỡng tính, mẫu 5, bầu dưới, đĩa mật ở đỉnh bầu, quả hạch. Đây là các đặc điểm đặc trưng của họ nhân sâm (*Araliaceae*); ii) Rễ củ mập, thân không gai, lá kép chân vịt, cánh hoa xếp lợp. Đây là các đặc điểm đặc trưng của chi sâm (*Panax*); iii) Dựa vào các kết quả khi phân tích cấu tạo, đặc điểm hình thái các thành phần của hoa đã xây dựng công thức hoa của mẫu nghiên cứu là *K5C5A5G(2).

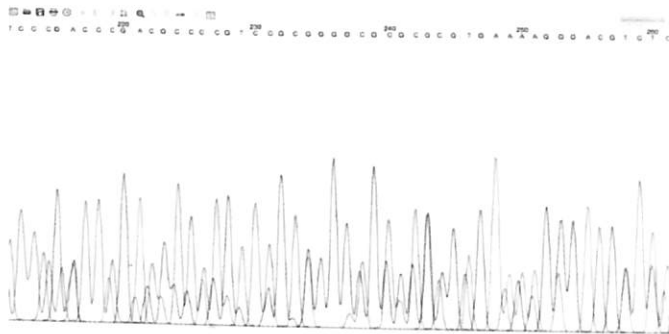
Những đặc điểm nhận biết loài sâm Ngọc Linh với các loài khác trong cùng chi là: quả khi chín có đốm đen ở đỉnh, rễ nằm ngang, dạng đốt, các đốt xếp so le, củ có vỏ ngoài mỏng, màu vàng đến nâu nhạt, lát cắt củ thường có viền màu vàng nhạt, mùi thơm đặc trưng.

Định loại mẫu sâm Ngọc Linh bằng chỉ thị ITS

Sau khi định loại bằng các đặc điểm hình thái, mẫu sâm Ngọc Linh tiếp tục được tiến hành giải mã trình tự vùng gen *ITS* để kiểm tra lại chính xác mẫu đã thu là sâm Ngọc Linh. Kết quả PCR khuếch đại trình tự vùng gen *ITS* của mẫu nghiên cứu cho một vạch sáng nét với kích thước 700 bp, tương ứng đúng với chiều dài lý thuyết của gen *ITS* (hình 1). Sản phẩm PCR sau đó được thực hiện giải trình tự với mỗi xuôi (*ITS-F*). Kết quả giải trình tự thu được cho hình ảnh sắc nét (hình 2).



Hình 1. Kết quả điện di sản phẩm PCR khuếch đại trình tự vùng gen *ITS* trên agarose 1%. Giếng M: DNA 1 kb plus; giếng 1: mẫu nghiên cứu.



Hình 2. Kết quả trình tự sản phẩm PCR (gen *ITS*).

Kết quả thu được đem so sánh với trình tự gen *ITS* sâm Ngọc Linh trên GenBank cho kết quả trùng khớp 100% với loài có mã số MK979391, đây là mẫu sâm Ngọc Linh được thu và nghiên cứu bởi Viện Dược liệu nên kết quả so sánh là rất đáng tin cậy (hình 3). Ngoài ra, kết quả giải trình tự gen *ITS* này cũng được so sánh với trình tự gen *ITS* của mẫu sâm Ngọc Linh chuẩn đang lưu giữ tại Viện Sinh thái và Tài nguyên Sinh vật (mã số GenBank MW931744), kết quả cũng cho thấy độ trùng khớp là 100%. Điều đó khẳng định mẫu sâm nghiên cứu chính xác là loài sâm Ngọc Linh.

Sequences producing significant alignments		Download	Select columns	Show	100				
select all	100 sequences selected	GenBank	GenBank	GenBank	GenBank				
	Description	Scientific Name	Max Bits	Total Score	Query Cover	E	Per. Ident	Per. Len	Accession
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1
<input checked="" type="checkbox"/>	<i>Panax vietnamensis</i> voucher P10, ITS, 700, bases, internal transcribed spacer 1, partial sequence, S.8.	<i>Panax vietnamensis</i>	1194	1194	100%	0.0	100.00%	668	MG979391.1

Hình 3. Kết quả so sánh trình tự gen *ITS* mẫu nghiên cứu với trình tự gen *ITS* sâm Ngọc Linh trên GenBank cho thấy trùng khớp 100%.

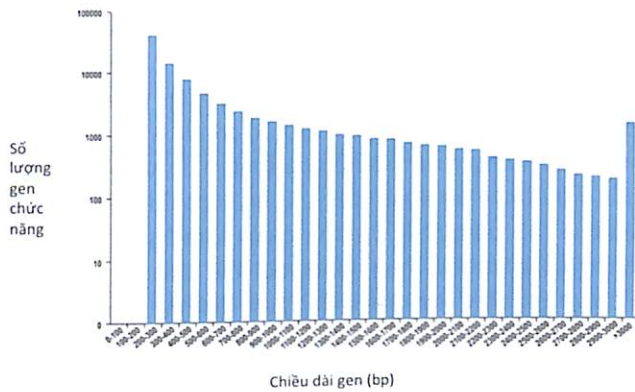
Tách chiết mRNA tổng số

Kết quả kiểm tra OD260/280 của mẫu RNA tổng số trên máy đo quang phổ Nanodrop ND-2000 (Thermo Electron Corporation, Mỹ) cho thấy, OD260/280 là 2,14, điều đó chứng tỏ số lượng RNA thu được là tốt và đạt độ tinh sạch cao.

Tổng hợp và giải mã trình tự cDNA từ RNA tổng số

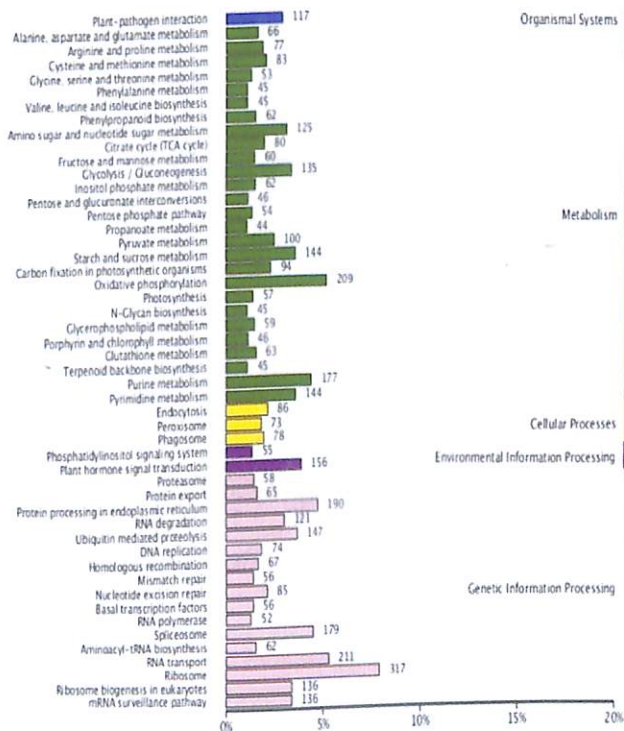
cDNA sợi đơn và sợi đôi được tổng hợp từ RNA tổng số bằng kit Access RT-PCR trong thời gian 70 phút. Sản phẩm cDNA thu được được đọc trình tự trên hệ thống máy Illumina HiSeq™ 4000. Kết quả giải trình tự hệ gen phiên mã của sâm Ngọc Linh sau khi kiểm tra chất lượng nghiêm ngặt và lọc dữ liệu đã cho ra 23.741.783 đoạn đọc (7,08 Gb) và tổng hợp thành 262.999 gen cố hữu (133 Mb), trong đó có 153.074 bản sao có chiều dài trung bình 770.572 bp, 48.314 bản sao (31,56%) nằm ở khoảng 200-300 bp, 35.174 bản sao (22,98%) nằm trong khoảng 301-500 bp, 32.031

(20,93%) bản sao nằm trong khoảng 501-1.000 bp, 25.800 bản sao (16,85%) nằm trong khoảng 1.001-2.000 bp và 11.755 bản sao (7,68%) dài hơn 2.000 bp (hình 4).



Hình 4. Tổng hợp sự phân bố (về độ dài) của các đoạn DNA tập hợp được từ giải trình tự hệ gen biểu hiện của sâm Ngọc Linh.

Kết quả so sánh với bộ dữ liệu NR, GO, Swiss-Prot, KOG và KEGG cho thấy, 80,86% các gen cổ hữu được phân loại thành công, theo đó các gen cổ hữu được phân thành 3 nhóm chức năng chính gồm: thành phần tế bào, chức năng phân tử và quá trình sinh học. Trong đó, gen cổ hữu tham gia vào quá trình sinh học là lớn nhất (39,69%), tiếp theo là các gen tham gia vào quá trình trao đổi chất (27,98%), đáp ứng kích thích (24,40%), hoạt động xúc tác, bảo quan, điều hoà dịch mã... (hình 5).

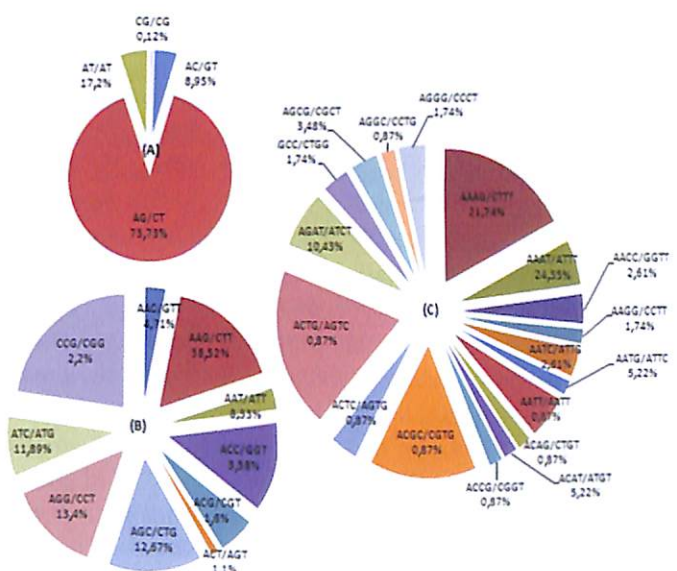


Hình 5. Tổng hợp sự phân bố các gen theo chức năng.

Xác định các vùng lặp SSR trên hệ gen biểu hiện của sâm Ngọc Linh

Các SSR thích hợp được lựa chọn là những đoạn lặp nucleotide ngắn (2-6 nucleotide) với các lần lặp tối thiểu là 6 đối với di-nucleotide, 5 lần đối với tri-nucleotide và 3-4 lần đối với các đoạn lặp cao hơn. Để đánh giá chất lượng lắp ráp và phát triển các dấu SSR, 89.271 gen cổ hữu được lắp ráp và sử dụng để khai thác các dấu phân tử nhỏ tiềm năng được xác định là các vùng lặp đến 6 nucleotide. Sử dụng công cụ SSRIT đã xác định được tổng cộng 11.343 EST-SSR tiềm năng. Tần số EST-SSR là 12,71% và mật độ phân phối của một EST-SSR được tìm thấy chiếm tổng số khoảng 5,98 kb ở trong các gen cổ hữu.

Loại tần số và phân phối của EST-SSR tiềm năng cũng được phân tích. Theo đó, motif lặp lại phổ biến nhất là mono-nucleotide (5.004, 44,12%), tiếp sau là di-nucleotide (4.648, 40,98%), tri-nucleotide (1.563, 13,78%), tetra-nucleotide (66, 0,58%), hexa-nucleotide (29, 0,26%) và penta-nucleotide (32, 0,28%) (hình 6, bảng 1). Kết quả thống kê cho thấy, EST-SSR với 10 mô đun lặp lại (2.040, 20,50%) là phổ biến nhất, theo sau là 6 (1.363, 12,02%), 5 (925, 8,15%), 7 (862, 7,6%), 8 (594, 5,24%) và 9 (428, 3,77%) (bảng 2). Motif chủ đạo trong lặp lại di-nucleotide là AG/TC (73,73%), tiếp theo là AT/TA (17,2%) và AC/TG (8,95%). Trong 10 loại lặp lại tri-nucleotide hầu hết được phân phối AAG/TTC (38,52%). Loại mô đun phổ biến nhất của lặp lại tetra-nucleotide là AAAT/TTTA (24,35%) (hình 6).



Hình 6. Tỷ lệ % các loại lặp nucleotide (SSR) trên hệ gen biểu hiện của sâm Ngọc Linh.

Bảng 1. Tổng hợp tỷ lệ các SSR ghi nhận được (dựa trên số lượng lặp của các nucleotide - N).

Số lần lặp	Lặp 1N	Lặp 2N	Lặp 3N	Lặp 4N	Lặp 5N	Lặp 6N	Tổng số	Tỷ lệ (%)
5	0	0	851	34	20	20	925	8,15
6	0	965	363	24	2	9	1.363	12,02
7	0	705	147	2	5	3	862	7,60
8	0	485	105	3	1	0	594	5,24
9	0	394	32	2	0	0	428	3,77
10	2.040	260	24	1	0	0	2.325	20,50
>10	2.964	1.839	41	1	1	0	4.846	42,72
Tổng	5.004	4.648	1.563	66	29	32	11.343	100
%	44,12	40,98	13,78	0,58	0,26	0,28		

Để tối ưu hóa tính đặc hiệu, theo Jurka và Pethiyagoda (1995) [9] chỉ lựa chọn locus SSR với tối thiểu là 10 lần lặp lại của di-nucleotide, 7 của tri-nucleotide và 5 của tetra-nucleotide. Kết quả nghiên cứu đã xác định được 101 vùng SSR thỏa mãn điều kiện trên (bảng 2).

Bảng 2. Bảng kết quả 101 vùng lặp SSR được lựa chọn.

TT	SRR	TT	SRR	TT	SRR	TT	SRR	TT	SRR
1	(CTCAGG)5	21	(GTTGGC)6	41	(ATG)17	61	(TAT)11	81	(ATT)11
2	(CATCAC)5	22	(AAAAGA)5	42	(TGC)10	62	(TAA)11	82	(ATT)10
3	(TTCATT)6	23	(CCTCTC)5	43	(TGT)10	63	(GTG)12	83	(ATC)12
4	(AACCT)5	24	(CGACCC)5	44	(TTC)11	64	(GCT)10	84	(AGA)23
5	(TGAAGA)5	25	(AACCT)6	45	(GGC)11	65	(GCG)11	85	(AGA)12
6	(CCTCCA)6	26	(CCCTAA)5	46	(GAA)12	66	(GAT)11	86	(AGA)10
7	(GAAAAA)7	27	(TTTTTC)7	47	(CTT)12	67	(GAT)10	87	(AAG)17
8	(GGGAGC)7	28	(GAAGA)6	48	(TCT)17	68	(GAG)12	88	(AAC)12
9	(GGCGGT)5	29	(CTTTT)11	49	(TCC)11	69	(GAG)10	89	(TC)35
10	(AAGGAA)5	30	(GAGGA)5	50	(AGC)10	70	(GAA)18	90	(TC)32
11	(TGAAT)5	31	(TTTGT)7	51	(CAG)10	71	(GAA)11	91	(TC)30
12	(TTTGCT)5	32	(CTCCT)6	52	(AAC)10	72	(GAA)10	92	(GA)31
13	(GCCTCC)6	33	(GGGTC)7	53	(TCT)14	73	(CTT)15	93	(GA)31
14	(ATACAA)5	34	(TAGA)10	54	(GAA)17	74	(CTT)11	94	(CT)38
15	(GAGACC)6	35	(ATAC)9	55	(TTC)13	75	(CTC)10	95	(CT)36
16	(CGGAGC)5	36	(TGTA)9	56	(TTC)11	76	(CGG)10	96	(CT)35
17	(GGAGCC)5	37	(TCTA)8	57	(TTC)11	77	(CCG)10	97	(CT)31
18	(GGTGGC)6	38	(GAA)16	58	(TGT)10	78	(CAG)15	98	(CT)30
19	(CATCAC)7	39	(AAG)10	59	(TGG)14	79	(CAG)11	99	(AG)34
20	(AGCGGC)5	40	(AGA)11	60	(TGA)10	80	(CAC)10	100	(AG)31
								101	(AC)30

Kết luận

Trong nghiên cứu này, hệ gen biểu hiện của sâm Ngọc Linh đã được giải mã thành công và xác định được tổng cộng 11.343 EST-SSR tiềm năng. Tần số EST-SSR là 12,71% và mật độ phân phối của một EST-SSR được tìm thấy chiếm tổng số khoảng 5,98 kb ở trong các gen cổ hữu. Đã lựa chọn được 101 vùng SSR có độ lặp tối thiểu là 10 lần của di-nucleotide, 7 của tri-nucleotide và 5 của tetra-nucleotide trong toàn bộ hệ gen biểu hiện của sâm Ngọc Linh. Đây là cơ sở quan trọng trong việc thiết kế mồi đặc hiệu cho các vùng SSR đặc trưng của sâm Ngọc Linh, giúp ích cho công tác xác định chính xác loài sâm này, cũng như để khám phá các gen mới liên quan đến sự hình thành và phát triển của loài sâm Ngọc Linh đặc hữu của Việt Nam.

LỜI CẢM ƠN

Nghiên cứu được tài trợ bởi Quỹ Phát triển KH&CN Quốc gia (NAFOSTED) thông qua đề tài mã số 106.06-2018.310. Các tác giả xin chân thành cảm ơn.

TÀI LIỆU THAM KHẢO

[1] Le Thanh Sơn, Nguyen Tap (2006), "Ecological characteristics of *Panax vietnamensis* Ha et Grushv", *Journal of Medicinal Material*, **14(4)**, pp.145-147.

[2] Nguyễn Thị Hồng Mai, Lê Thanh Sơn, Nguyễn Thị Phương Trang (2020), "Đặc điểm di truyền quần thể sâm Ngọc Linh (*Panax vietnamensis* Ha and Grushv) bằng phương pháp SSR", *Tạp chí Sinh học*, **42(1)**, tr.11-19.

[3] Bộ Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam (2007), *Sách đỏ Việt Nam, phần II*, Nhà xuất bản Khoa học Tự nhiên và Công nghệ.

[4] <https://baodantoc.vn/san-pham-sam-ngoc-linh-dang-bi-gia-mao-1578627504977.htm>.

[5] Nguyễn Đức Thành (2014), "Các kỹ thuật chi thị DNA trong nghiên cứu và chọn lọc thực vật", *Tạp chí Sinh học*, **36(3)**, tr.265-294.

[6] Hoàng Thị Sân (2009), *Thực hành phân loại thực vật*, Nhà xuất bản Giáo dục.

[7] J.J. Doyle, J.L. Doyle (1991), "A rapid DNA isolation procedure from small quantities of fresh leaf tissues", *Phytochem. Bull.*, **19**, pp.11-15.

[8] M.G. Grabherr, et al. (2011), "Full-length transcriptome assembly from RNA-seq data without a reference genome", *Nature Biotechnology*, **29**, pp.644-652.

[9] J. Jurka, C. Pethiyagoda (1995), "Simple repetitive DNA sequences from primates: compilation and analysis", *J. Mol. Evol.*, **40**, pp.120-126.

[10] <https://primer-premier-5.software.informer.com>.