

Phân lập và phân tích *in silico* các đặc điểm chức năng của một biến thể bản sao terpene synthase (*HbTPS6L-X1*) hiện diện ở vỏ thân cây cao su *Hevea brasiliensis* RRIV 209

Trần Thị Diễm Hương^{1,2}, Nguyễn Ngọc Tuyết^{1,2}, Nguyễn Thị Hồng Thương^{1,2*}

¹Khoa Sinh học - Công nghệ Sinh học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP Hồ Chí Minh, 227 Nguyễn Văn Cừ, phường 4, quận 5, TP Hồ Chí Minh, Việt Nam

²Đại học Quốc gia TP Hồ Chí Minh, quốc lộ 1A, khu phố 6, phường Linh Trung, TP Thủ Đức, TP Hồ Chí Minh, Việt Nam

Ngày nhận bài 17/1/2023; ngày chuyển phản biện 19/1/2023; ngày nhận phản biện 7/2/2023; ngày chấp nhận đăng 10/2/2023

Tóm tắt:

Cây cao su (*Hevea brasiliensis*) là loài cây duy nhất được trồng để sản xuất cao su tự nhiên. Phương pháp thu hoạch mủ cao su thông qua cạo vỏ có tác động gây tổn thương cơ học cho cây. Ngoài ra, cây cao su có thể dễ bị sâu bệnh tấn công qua vết thương ở vùng bị rạch cạo. Terpenoid là nhóm hợp chất thứ cấp có vai trò quan trọng trong tương tác giữa thực vật và môi trường, đặc biệt là trong phản ứng của cây với điều kiện stress. Sự đa dạng của terpenoid chủ yếu được quyết định bởi các enzyme thuộc họ terpene synthase (TPS). Kết quả phân tích dữ liệu RNA-seq của *H. brasiliensis* cho thấy có nhiều gen TPS hiện diện ở loài này nhưng chức năng của chúng vẫn chưa được nghiên cứu. Bài báo trình bày kết quả phân lập từ vỏ thân cây cao su *H. brasiliensis* RRIV 209 một biến thể bản sao của gen TPS (ký hiệu *HbTPS6L-X1*). Thông qua kết quả phân tích cây tiến hóa, sự hiện diện của các motif bảo tồn, vùng trình tự peptide tín hiệu của protein và mức độ tương đồng về cấu trúc với các sesquiterpene synthase đã được nghiên cứu, *HbTPS6L-X1* được dự đoán thuộc phân họ TPS-a, định vị trong tế bào chất và xúc tác chuyển hóa FPP thành các sesquiterpene. Những đặc điểm chức năng *in silico* này là cơ sở để thiết kế các khảo sát thực nghiệm nhằm khẳng định chức năng *in planta* của *HbTPS6L-X1*.

Từ khóa: *Hevea brasiliensis*, phân họ TPS-a, terpene synthase, terpenoid.

Chỉ số phân loại: 4.6

1. Đặt vấn đề

Cây cao su *H. brasiliensis* là cây công nghiệp nhiệt đới lâu năm, được trồng rộng rãi ở Đông Nam Á và là nguồn tổng hợp cao su tự nhiên chính trên thế giới [1]. Cây cao su thường xuyên bị cạo vỏ trong quá trình thu hoạch mủ và hệ quả là cây có thể phải chịu thêm nhiều tác động stress ngoại sinh khác như sự tấn công của nấm, vi khuẩn... thông qua vết cạo ở thân. Thực vật nói chung và cây cao su nói riêng có thể đáp ứng với các tác nhân stress bằng cách điều hòa sự biểu hiện các gen liên quan đến các con đường phòng vệ, bao gồm cả các gen tham gia vào quá trình sinh tổng hợp một số hợp chất thứ cấp.

Terpenoid là nhóm hợp chất thứ cấp đa dạng về mặt hóa học và đóng vai trò quan trọng trong các tương tác giữa thực vật với những tác động sinh học và phi sinh học của môi trường xung quanh chúng. Ở thực vật, terpenoid được hình thành từ những đơn vị 5 nguyên tử carbon, isopentenyl diphosphate (IPP) và dimethylallyl diphosphate (DMAPP), thông qua con đường 2-C-methyl-D-erythritol-4-phosphate (MEP) diễn ra ở lục thể (plastid) hoặc con đường mevalonate (MEV) diễn ra ở tế bào chất (cytosol) [2, 3]. Sự kết hợp đầu-đuôi (head-to-tail) giữa IPP và DMAPP dưới sự xúc tác của các enzyme cis-prenyltransferase hoặc trans-prenyltransferase tạo ra các hợp chất trung gian cis- hoặc trans-prenyl diphosphate có số nguyên tử carbon là bội số của 5 như neryl diphosphate và geranyl diphosphate (NPP và GPP, C10), farnesyl diphosphate (FPP, C15), nerylgeranyl diphosphate và geranylgeranyl diphosphate (NNPP và GGPP, C20)... [4]. Các prenyl diphosphate này được

chuyển hóa thành monoterpene, sesquiterpene và diterpene tương ứng bởi các enzyme thuộc họ (TPS).

TPS được phân thành hai loại chính: TPS loại I và TPS loại II, tùy theo trình tự motif bảo tồn hiện diện trong cấu trúc và cơ chế xúc tác của protein. TPS loại I chứa motif DDXXD trên miền α ở đầu C, trong khi TPS loại II chứa motif DXDD trên miền β ở đầu N của phân tử protein. Những TPS có cả 2 motif giàu aspartate này được xem là TPS có chức năng kép (bifunctional TPS) [5]. Quá trình sinh tổng hợp hemiterpene (C5), monoterpene (C10), sesquiterpene (C15) và sesterterpene (C25) thường chỉ bao gồm sự tham gia của TPS loại I với vai trò loại bỏ nhóm diphosphate để tạo ra chất trung gian mang điện tích dương (carbocation) [6]. Tương tự, một số diterpene (C20) được tổng hợp từ geranylgeranyl diphosphate (GGPP) hoặc nerylgeranyl diphosphate (NNPP) chỉ thông qua một chất trung gian dạng "carbocation" bởi các TPS loại I [7-9]. Ngược lại, một số diterpene khác được tổng hợp thông qua con đường phức tạp hơn, trong đó TPS loại II hoặc miền chứa motif DDXD của TPS chức năng kép xúc tác quá trình chuyển hóa (GGPP) thành chất trung gian prenyl diphosphate mang cấu trúc vòng đôi, sau đó TPS loại I hoặc miền chứa motif DDXXD của TPS chức năng kép loại bỏ nhóm diphosphate và hình thành chất trung gian "carbocation" [10, 11].

Phân tích cây tiến hóa dựa trên so sánh trình tự của các gen/protein TPS từ thực vật đã chia họ enzyme này thành 7 phân họ, với những TPS thuộc cùng một phân họ thường có hoạt tính xúc tác tương tự nhau: TPS-a (sesquiterpene synthase), TPS-b (hemiterpene synthase và các monoterpene synthase tổng hợp

*Tác giả liên hệ: Email: nththuong@hcmus.edu.vn

Isolation and *in silico* analysis of a transcript variant for a putative terpene synthase (*HbTPS6L-X1*) from the bark tissues of rubber *Hevea brasiliensis* RRIV 209

Thi Diem Huong Tran^{1,2}, Ngoc Tuyet Nguyen^{1,2},
Thi Hong Thuong Nguyen^{1,2*}

¹Faculty of Biology and Biotechnology, University of Science,
Vietnam National University - Ho Chi Minh City,

277 Nguyen Van Cu Street, Ward 4, District 5, Ho Chi Minh City, Vietnam

²Vietnam National University - Ho Chi Minh City,

1A Highway, Quarter 6, Linh Trung Ward, Thu Duc City, Ho Chi Minh City, Vietnam

Received 17 January 2023; revised 7 February 2023; accepted 10 February 2023

Abstract:

Hevea brasiliensis is the only tree species grown to produce natural rubber. The method of harvesting latex by slicing a groove into the bark of the tree mimics a mechanical wounding. In addition, rubber trees could be frequently exposed to pest and pathogen attacks through the scraped bark region. Terpenoids are a group of specialised compounds that play an important role in plant-environment interactions, especially in plant response to stress factors. The diversity of terpenoids is mainly determined by enzymes of the terpene synthase (TPS) family. Screening the RNA-seq data of *H. brasiliensis* revealed the presence of many putative TPS genes expressed in the bark tissues of rubber trees; however, their function has not been studied. This paper presents the identification of a TPS gene transcript variant (designated as *HbTPS6L-X1*) from the bark tissues of *H. brasiliensis* RRIV 209. Through the analysis of the phylogenetic tree, the presence of conserved motifs, and the signal peptide sequence region of the protein, as well as structural homology to known sesquiterpene synthases from other plants, *HbTPS6L-X1* was predicted to belong to the TPS-a subfamily, be localised in the cytoplasm and catalyse the conversion of farnesyl diphosphate (FPP) to sesquiterpenes. These *in silico* functional characteristics are the basis for designing experiments to confirm the *in planta* biological function of *HbTPS6L-X1* protein.

Keywords: *Hevea brasiliensis*, terpene synthase, terpenoids, TPS-a subfamily.

Classification number: 4.6

monoterpene vòng), TPS-c (copalyl diphosphate synthase và entkaurene synthase), TPS-d (các terpene synthase đặc trưng cho thực vật hạt trần), TPS-e/f (ent-kaurene synthase và các diterpene synthase khác, cùng một số monoterpene synthase và sesquiterpene synthase), TPS-g (các monoterpene synthase tổng hợp monoterpene dạng mạch hở hoặc nhánh) và TPS-h (các terpene synthase đặc trưng cho nhóm cây không hạt Selaginella) [5].

Khi nghiên cứu kỹ cấu trúc của TPS, các nhà khoa học đã xác định được một số motif quan trọng đối với hoạt động xúc tác của enzyme [7]. Trong trường hợp của các sesquiterpene synthase (STS), ngoài motif giàu aspartate DDXXD đặc trưng của TPS loại I, các STS còn có các motif bảo tồn khác tham gia vào phản ứng xúc tác như NSE/DTE và RXR [12-14]. Phần kỵ nước (chuỗi alkyl) của cơ chất FPP được hướng vào khoang tâm hoạt động. Phần ưa nước (nhóm diphosphate) của FPP được cố định giữa motif RXR và các ion kim loại hóa trị II Mg²⁺ hoặc Mn²⁺, những cofactor này được neo bởi hai motif DDXXD và NSE/DTE ở lõi vào tâm hoạt động. Góc Asp (D) thứ nhất cùng với Asp cuối cùng của motif DDXXD liên kết với 2 nguyên tử Mg²⁺, trong khi bộ ba Asn (N) - Ser (S) - Glu (E) hoặc Asp (D) - Thr (T) - Glu (E) của motif NSE/DTE liên kết với nguyên tử Mg²⁺ còn lại. 3 nguyên tử Mg²⁺ và 2 gốc Arg (R) bảo tồn của motif RXR cố định cơ chất FPP trong tâm hoạt động của STS [15]. Motif RXR được tìm thấy ở cách motif DDXXD khoảng 35 axit amin về phía đầu N. Nó định vị trên một vòng có cấu trúc linh hoạt nhưng cấu trúc vòng này trở nên trật tự và ổn định hơn khi enzyme gắn với cơ chất [14]. 2 gốc Arg (R) trong motif RXR được cho là tham gia tạo phức với nhóm diphosphate tách ra sau bước ion hóa cơ chất FPP và hướng nó ra khỏi túi kỵ nước - nơi “carbocation” tiếp tục trải qua quá trình đóng vòng hoặc sắp xếp lại cấu trúc để tạo thành các sản phẩm sesquiterpene [14].

Hiện nay, các dữ liệu trình tự bộ gen (genome) và hệ gen biểu hiện (transcriptome) của một số dòng cao su *H. brasiliensis* đã được công bố và ngày càng được hoàn thiện, mang lại nhiều thuận lợi cho việc phát hiện và nghiên cứu chức năng các gen liên quan đến quá trình chuyển hóa các hợp chất thứ cấp, trong đó có các gen TPS. Phân tích dữ liệu trình tự bộ gen và hệ gen biểu hiện của *H. brasiliensis* RRIM 600 (<http://matsui-lab.riken.jp/rubber/search.html>) [16] cho thấy có nhiều gen TPS hiện diện ở vỏ thân cây cao su nhưng chức năng của các gen này vẫn chưa được nghiên cứu và công bố.

Trong nghiên cứu này, chúng tôi phân lập từ vỏ thân cây cao su *H. brasiliensis* RRIV 209 một biến thể bản sao của gen TPS (ký hiệu *HbTPS6L-X1*). *HbTPS6L-X1* tương ứng với "hit" trình tự có mã số Hb_001271_040 trong cơ sở dữ liệu trình tự bộ gen và hệ gen biểu hiện của *H. brasiliensis*. Dựa trên trình tự *HbTPS6L-X1* đã phân lập, nhóm nghiên cứu dự đoán một số đặc điểm chức năng của protein được mã hóa bởi gen này, bao gồm vị trí định vị trong tế bào và hoạt tính xúc tác, thông qua phân tích phân họ TPS, xác định sự hiện diện của các motif bảo tồn, dự đoán vùng trình tự peptide tín hiệu của protein và mức độ tương đồng trong cấu trúc không gian của nó so với các sesquiterpene synthase đã nghiên cứu.

2. Đối tượng và phương pháp nghiên cứu

2.1. Đối tượng

Cây cao su *H. brasiliensis* RRIV 209 được trồng tại Viện Nghiên cứu Cao su Việt Nam (huyện Bàu Bàng, tỉnh Bình Dương, Việt Nam).

2.2. Phân lập trình tự biến thể bản sao HbTPS6L-X1

Trình tự mã hóa protein HbTPS6L-X1 được phân lập bằng PCR với khuôn là cDNA chuẩn bị từ mô vỏ thân cây cao su *H. brasiliensis* và cặp mồi cho phép nhân bản đặc hiệu cả 2 trình tự bản sao dự đoán của gen. Mồi xuôi HbTPS6L-F và mồi ngược HbTPS6L-R lần lượt có trình tự là: 5'-ATGGAAGTGCAACCTCATTAC-3' và 5'-GGGAAAGACATATATACGAAACTCAT-3'. PCR được thực hiện dưới sự xúc tác của Phusion DNA Polymerase (Thermo Scientific) với chu trình nhiệt như sau: 98°C/30 giây; 38 x (98°C/10 giây, 58°C/30 giây, 72°C/90 giây); 72°C/10 phút; 4°C/∞. Sản phẩm của phản ứng PCR được điện di song song với thang chuẩn DNA trên gel agarose 1%, sau đó được tinh sạch từ gel bằng bộ kit EZ-10 Spin column DNA gel Extraction (Thermo Scientific) theo hướng dẫn của nhà sản xuất.

2.3. Tạo dòng HbTPS6L-X1 vào vector pJET1.2/blunt

Sản phẩm PCR đặc hiệu (*HbTPS6L-X1*) được chèn vào vector pJET1.2/blunt thông qua phản ứng nối. Sản phẩm nối được biến nạp vào tế bào *E. coli* TOP10 khả nạp. Sau đó, sản phẩm biến nạp được trải trên đĩa LB có chứa kháng sinh ampicillin 100 µg/ml và ủ ở 37°C trong 14-16 giờ. Các thể biến nạp mọc trên đĩa môi trường LB bổ sung ampicillin được sàng lọc bằng PCR khuẩn lạc với mồi HbTPS6L-R bắt cặp đặc hiệu với gen và mồi pJET1.2-F (5'-CGACTCACTATAGGGGAGAGCGGC-3') bắt cặp đặc hiệu với vector. Khuẩn lạc đơn của dòng tế bào *E. coli* TOP10 cho kết quả sàng lọc dương tính được nuôi nhân sinh khối để tách chiết plasmid. Trình tự phân lập đã đồng hóa được giải mã và được so sánh với trình tự dự đoán ban đầu bằng các phần mềm tin-sinh học chuyên dụng như: ClustalW, Blast...

2.4. Phân tích trình tự của HbTPS6L-X1

2.4.1. Dự đoán phân họ của HbTPS6L-X1

Để xác định phân họ của protein HbTPS6L-X1, trình tự axit amin của HbTPS6L-X1 được so sánh với các trình tự protein TPS đại diện cho 5 phân họ TPS khác nhau của cả chua *Solanum lycopersicum* - loài thực vật duy nhất có toàn bộ hệ gen TPS đã được phân lập và khảo sát chức năng [8, 17, 18]. Cụ thể, các trình tự TPS tham chiếu được sử dụng cho phân tích phân họ bao gồm SITPS12 và SITPS31 (TPS-a), SITPS3 và SITPS27 (TPS-b), SITPS40 và SITPS41 (TPS-c), SITPS37 và SITPS39 (TPS-g), SITPS20 và SITPS24 (TPS-e/f). Cây phả hệ được thiết lập thông qua bộ công cụ phân tích ETE3 v3.1.2 (<http://www.genome.jp/tools/ete/>) [19], trong đó các trình tự được giống cột bởi phần mềm MUSCLE v3.8.31 [20] và cây phả hệ được dựng bởi phần mềm IQ-TREE 1.5.5 [21].

2.4.2. Dự đoán vùng trình tự peptide tín hiệu (transit peptide)

Vùng trình tự “transit peptide” của HbTPS6L-X1 được dự đoán bằng 2 phương pháp khác nhau. Phương pháp thứ nhất dựa trên phân tích kết quả sắp giống cột trình tự protein này với những TPS thuộc cùng phân họ và có vùng “transit peptide” đã được xác định. Phương pháp thứ hai dựa trên kết quả dự đoán trình tự peptide vận chuyên đầu N của protein bởi chương trình TargetP 2.0 [22].

2.4.3. Mô hình hóa tương đồng cấu trúc protein

Mô hình cấu trúc của HbTPS6L-X1 được dự đoán bởi AlphaFold 2 [23, 24]. Interpro được sử dụng để dự đoán các miền bảo tồn có thể có trong cấu trúc của protein [25]. Sự chồng lấp cấu trúc dự đoán của HbTPS6L-X1 với cấu trúc mẫu 5-epi-aristolochene synthase (NtEAS4) (PDB id: 5EAT) được thực hiện thông qua chương trình ChimeraX [26]. Cấu trúc mẫu (template) được chọn từ thư viện cấu trúc PDB dựa trên kết quả đánh giá bởi chương trình phân tích xâu chuỗi (threading) LOMETS [27]. Giá trị căn bậc hai của độ lệch bình phương trung bình (Root Mean Square Deviation, RMSD) giữa cấu trúc nghiên cứu (HbTPS6L-X1) và cấu trúc mẫu (NtEAS4) được tính theo tọa độ của nguyên tử Ca bởi ChimeraX.

3. Kết quả và bàn luận

3.1. Phân tích in silico sự hiện diện của trình tự mã hóa terpene synthase 6 (HbTPS6L CDS) ở cây cao su *H. brasiliensis*

“Hb_001271_040” là một trong những “hit” trình tự mã hóa terpene synthase giả định được ghi nhận từ kết quả tra cứu thông tin bộ gen cao su *H. brasiliensis* RRIM 600 (<http://matsui-lab.riken.jp/rubber/search.html>) với từ khóa “terpene synthase” và “terpene cyclase”. Kết quả BLAST cơ sở dữ liệu “nucleotide collection” của NCBI với trình tự truy vấn “Hb_001271_040” cho thấy xuất hiện đầu tiên trong danh sách các “hit” có mức độ tương đồng cao với “Hb_001271_040” là 2 biến thể bản sao của gen với chú giải chức năng dự đoán là “*Hevea brasiliensis* probable terpene synthase 6” (E-value = 0). Hai biến thể này có mã số truy cập lần lượt là XM_021794026.1 (*HbTPS6L-X1_insilico*) và XM_021794027.1 (*HbTPS6L-X2_insilico*).

Để phân lập CDS *HbTPS6* dựa trên trình tự nucleotide của 2 biến thể dự đoán, mồi xuôi HbTPS6L-F được thiết kế bắt cặp bổ sung với vùng trình tự có chứa codon mở đầu của gen *HbTPS6L* và mồi ngược HbTPS6L-R bắt cặp bổ sung với vùng 3'UTR của gen (hình 1). Kết quả kiểm tra tính đặc hiệu của mồi bằng chương trình Primer-BLAST cho thấy, cặp mồi đã chọn bắt cặp bổ sung với cả 2 trình tự biến thể *HbTPS6L-X1_insilico* và *HbTPS6L-X2_insilico*. Sản phẩm nhân bản được dự đoán có kích thước lần lượt là 1707 (*HbTPS6L-X1_insilico*) và 1458 bp (*HbTPS6L-X2_insilico*).

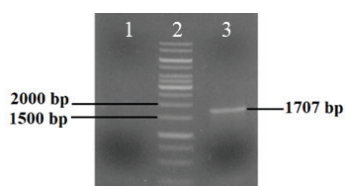
```
>XM_021794026.1 PREDICTED: Hevea brasiliensis probable terpene synthase 6
(LOC110642084), transcript variant X1, mRNA
TTCCAAGCATTGGACAAGTATACAAGCACCTAAGTTACGTACTTCTCCTTCCCTTCCCTCCTG
HbTPS6L-F
GGGAAAATGGAAAGTGCACCTCATTTACGCCAAACACTGACCAAAACCAAAATTCGTCAGTGTTCCT
CCTCCTGTGGGAAATCTACAGCTTTACTCTTTTCTTTGCTGGACTCAGAATATGAATCGTGGACTA
GACGGGTGGAAACCGCTCAAAGAAAAGGTCAAAGACATGTTAATGGCGTATCAAGTAAATCCGATCAAGAA
AATGAAATGATTAACTTATTTATGCGCTCTGGTATTATCATATCATTTGAGAGGAAATGAAAGACAA
TAAATTCATATTTTAAATGGGTCTCACTCTTTGATGAAATGATATGATCTTACACTACAGCTC
TCCATTTTCAAGTCTTGAGACAACCGGATTCAAAATGAGCTGTGATGTGTTCAAGAAATTCAGGACAG
CAATGGAGAGTTCAGAAAGAGCATCAAAAGATGTGAGAGGATCTTCAGCTTGTATGAAAGCAAGTTTT
ATGAGTGTGCATGGAGAAATATTTCTGATGAAGCTCTTGAATTTACAAAGTTCACACTGGAGTCATTTG
CTATGCAATCAAAACCCATTTATGACCTATAAAGAAATGCTCTGATCTCCCTTTCAGAAAGGAGT
ACCCAGAGTAGAGGCTGGCAATACATCTCTTTCTATGAAAAGAAAGAAATCTCCAAATGAATTCCTACTC
GAATTTGCAAAATGATTTCATTCAGTGAATATTTGCAAGAGAGGAGCTCAAGGTGATGGCAAGT
GGTGGAAAGACTTAAATCTTGACAGAAACTTCCCTATGCAAGAGACAGACTTGGAAATCTATTCTGT
GGGTGTTCCATTGAGCTTGGAGCTCACTGCTCTTCAAGGATGATGGCCAAATTTGAAATG
ATATCACTAGCTGATGACACATACGATGATATGCTAAAATGAAGGAAGTCAATGCTTCAAGATGCTA
TCGAGAGGTGCATATTGATGCTACTGACGCAACTGCCAGATTACATGAAAAGTTTTCTACAGAAATTA
GATGTTTTGGAGAAACAGAGAATATTTGCGATGAAGCTCTTGAATTTACAAAGTTCACACTGGAGTCAT
GAGAAATTTAAACTGTTTGAATCTCATGCGCATGCGCATGGAGCCACACTGGTTTAAAGAGGTATTTG
CAATTTGAGTGTATTGGCAAAATGGTTCTGAGTCAAGCGGCTATAGCTTAGTTGAGCAGCGGCTTAT
TGGAAATGAAAGATTTGACGGATCAAAAGATATCAAGTGGCTTCAAGTACCCAAAATTTGCAAGCT
CAAGGATCTTGGGCTCTTGAGAAGCAGATAGTATCCACCAGGCTGAGCAAAAGAGAGGAGACTCTG
CTTCACTGGTAGAGTCTACATGAAAGAACATGATGTTTATAGAGAAAGGCTGGAAAGAGATTTCTTAA
AATTTGTGCAAAATGATGAAAGGACATCAACGAAAGAAATGATGAAGCCAACTGCTCTGCCAAGTCTTAT
CTCCAAATGTTGGTTAACTTGTCTGCAATACAGAGGTTTTTACAGATTTGACGACTCTTACACCAACC
CATCAAGTCTGAAAGATAAAATACAGCATGCTTCTGGAGCCACTCTTCTGCAAGACCAAGTAAATAT
ATAATACCTGAGTTGGTATTTGCTTCCCTGCTTGGCTTGGATTGCTGCTGGCCGGAAGGCA
HbTPS6L-R
AGAATAAATGTTGGCTTGGCTTGGCTATATATCAGTTGTTGAAATAAACAATAATGATCGAGAGGCA
AACACTGTGCTCCTCTGCTCATCTTAATATGAAATTAAGGTGCTATGCTTGTGATTTGATA
ATGTGGCAATTTTATGTTCTTAAAAAATAAATAATGTAATAGATGATA
```

Hình 1. Vị trí bắt cặp bổ sung của mồi xuôi HbTPS6L-F và mồi ngược HbTPS6L-R trên trình tự bản sao dự đoán *HbTPS6L-X1_insilico* và *HbTPS6L-X2_insilico*.

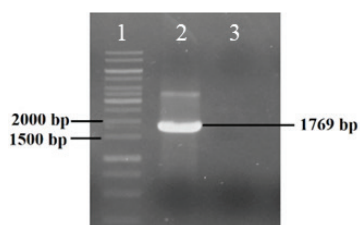
Phần được in đậm tương ứng vị trí codon mở đầu và codon kết thúc. Mũi tên chỉ vị trí bắt cặp của mỗi xuôi (HbTPS6L-F) và mỗi ngược (HbTPS6L-R). Đoạn phía trước codon mở đầu và phía sau codon kết thúc lần lượt tương ứng với vùng 5' UTR và 3'UTR. Đoạn được gạch dưới tương ứng với vùng trình tự chỉ có ở *HbTPS6L-X1_insilico* nhưng không có ở *HbTPS6L-X2_insilico*.

3.2. Kết quả phân lập trình tự mã hóa HbTPS6L (CDS HbTPS6L)

PCR nhân bản vùng trình tự mã hóa của gen *HbTPS6L* được thực hiện với khuôn là cDNA chuẩn bị từ mô vỏ của cây cao su *H. brasiliensis* và cặp mồi HbTPS6L-F và HbTPS6L-R. Kết quả điện di sản phẩm PCR (hình 2) cho thấy, ở giếng 3 xuất hiện một vạch băng có kích thước ở khoảng giữa vạch 1500 và 2000 bp so với thang chuẩn DNA 1 kb, phù hợp với kích thước dự đoán của phân đoạn *HbTPS6L-X1_insilico* (1707 bp), còn đối chứng âm không xuất hiện vạch băng tương tự (hình 2, giếng 1).



Hình 2. Kết quả điện di sản phẩm của PCR với cặp mồi HbTPS6L-F và HbTPS6L-R. 1: đối chứng âm không bổ sung khuôn cDNA; 2: thang DNA chuẩn 1 kb; 3: sản phẩm của phản ứng PCR với khuôn cDNA từ mô vỏ thân cây cao su *H. brasiliensis*.



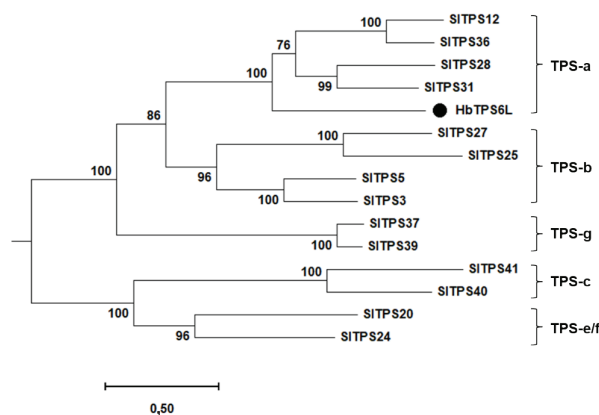
Hình 3. Kết quả sàng lọc dòng tế bào *E. coli* TOP10/pJET1.2/blunt-HbTPS6L bằng PCR khuẩn lạc. 1: thang chuẩn DNA 1kb; 2: sản phẩm PCR khuẩn lạc của dòng tế bào *E. coli* TOP10/ pJET1.2/blunt-HbTPS6L; 3: đối chứng âm.

Sản phẩm PCR đặc hiệu (dự đoán là phân đoạn *HbTPS6L-X1*) được tinh sạch và chèn vào vector pJET1.2/blunt. Kết quả sàng lọc bằng PCR khuẩn lạc dòng tế bào *E. coli* TOP10 nghi ngờ mang plasmid pJET1.2/blunt-*HbTPS6L-X1* được mô tả ở hình 3. Giếng 2 xuất hiện một vạch băng tương ứng với sản phẩm nhân bản 1769 bp trong trường hợp plasmid có mang đoạn chèn *HbTPS6L-X1*. Ở giếng 3, đối chứng âm không xuất hiện vạch băng tương tự. Kết quả này cho thấy, khuẩn lạc đơn được sử dụng trong PCR có thể là của dòng tế bào *E. coli* TOP10 chứa plasmid pJET1.2/blunt-*HbTPS6L-X1*. Vì vậy, khuẩn lạc này được nuôi nhân sinh khối để tách chiết plasmid. Plasmid được gửi giải trình tự với cặp mồi pJET1.2-F và pJET1.2-R nhằm xác định trình tự của đoạn gen được chèn.

Khi BLAST cơ sở dữ liệu “nucleotide collection” của NCBI với trình tự *HbTPS6L-X1* được phân lập, chúng tôi ghi nhận trình tự có mã số truy cập P(*HbTPS6L-X1_insilico*) xuất hiện đầu tiên trong danh sách các trình tự tương đồng với trình tự truy vấn, với tỷ lệ phần trăm trình tự giống nhau là 98,63% (phụ lục 1). Trình tự *HbTPS6L-X1* phân lập từ *H. brasiliensis* RRIV 209 sai khác 23 nucleotide so với trình tự biến thể dự đoán XM_021794026.1 từ *H. brasiliensis* Reyan7-33-97. Sự sai khác này có thể đã phản ánh biến dị di truyền giữa 2 dòng cao su. Bên cạnh đó, khả năng có lỗi phát sinh liên quan đến quá trình giải và lắp ráp trình tự bộ gen cây cao su cũng không được loại trừ.

3.3. HbTPS6L-X1 thuộc phân họ TPS-a và được dự đoán định vị trong tế bào chất

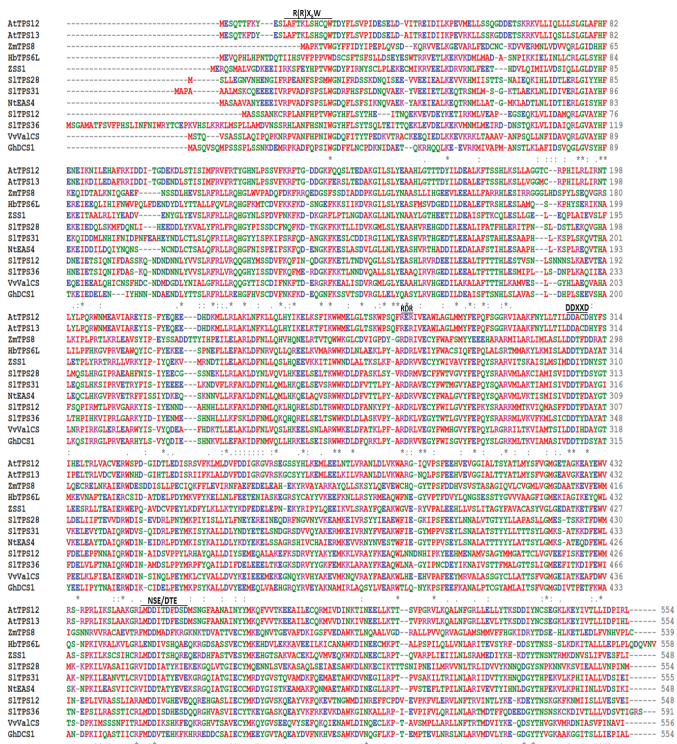
Một số đặc điểm chức năng của protein TPS có thể được dự đoán thông qua phân tích trình tự của chúng. Kết quả phân tích cây tiến hóa dựa trên so sánh trình tự axit amin của protein mã hóa bởi *HbTPS6L-X1* với các trình tự TPS thuộc 5 phân họ TPS khác nhau ở thực vật hạt kín (với cà chua *Solanum lycopersicum* là đại diện có toàn bộ hệ gen TPS đã được nghiên cứu chức năng) cho thấy, HbTPS6L-X1 có mức độ bảo thủ trình tự cao nhất với các terpene synthase thuộc phân họ TPS-a: SITPS12 (39,22%), SITPS28 (42,67%), SITPS31 (43,47%) và SITPS36 (43,92%) (hình 4). Các TPS trong phân họ TPS-a thường có hoạt tính xúc tác chính là sesquiterpene synthase [5].



Hình 4. Cây tiến hóa được thiết lập dựa trên so sánh các trình tự axit amin của HbTPS6L-X1 với các trình tự TPS đại diện cho các phân họ TPS khác nhau ở thực vật hạt kín. Phân tích tiến hóa được thực hiện theo phương pháp Maximum Likelihood dựa trên 14 trình tự SITPS của cà chua *Solanum lycopersicum* và trình tự HbTPS6L-X1 được phân lập từ vỏ thân cây cao su *H. brasiliensis*. Các số ở các nhánh cây thể hiện chỉ số bootstrap được phân tích từ 1000 lần lặp lại. Mã số truy cập của các trình tự SITPS như sau: SITPS3 (G1JUH1), SITPS5 (Q1XBU5), SITPS12 (D5KXD2), SITPS20 (C1K5M3), SITPS24 (NP_001307929.1), SITPS25 (G5CV49), SITPS27 (G5CV48), SITPS28 (G5CV47), SITPS31 (G5CV46), SITPS36 (G5CV42), SITPS37 (G5CV41), SITPS39 (NP_001306121.1), SITPS40 (G5CV38), SITPS41 (XP_010324500.1).

Kết quả sắp giống cột trình tự protein HbTPS6L-X1 với các trình tự TPS thuộc phân họ TPS-a ở những loài thực vật khác cho thấy, HbTPS6L-X1 có đủ 3 motif chức năng DDXXD, NSE/DTE và RXR của TPS loại I (hình 5) [5]. Các motif DDXXD và NSE/

DTE thường định vị trên 2 xoắn α đối diện nhau ở lõi vào tâm hoạt động của TPS loại I và liên kết với các đồng yếu tố (Mg^{2+} hoặc Mn^{2+}) để hỗ trợ ion hóa cơ chất prenyl diphosphate thành “carbocation” - chất trung gian mang điện tích dương trong phản ứng tổng hợp terpene [12, 13, 28]. Motif NSE/DTE thường ít được bảo tồn hơn với trình tự “consensus” là (L,V)(V,L,A)(N,D)D(L,I,V)X(S,T)XXXE [29] và motif này hiện diện trong HbTPS6L-X1 với trình tự là LEN450DIVS454HQAE458. Ngoài ra, motif giàu arginine RXR nằm cách motif DDXXD 33 axit amin về phía đầu N cũng có mặt trong trình tự protein HbTPS6L-X1 (hình 5). Đáng lưu ý, HbTPS6L-X1 không chứa R(R)_XW - motif tham gia trong quá trình đồng phân hóa tạo các sản phẩm terpene mạch vòng [30]. Vì vậy, HbTPS6L-X1 được dự đoán là một sesquiterpene synthase với khả năng xúc tác tạo sản phẩm sesquiterpene mạch hở.



Hình 5. Kết quả sắp giống cột trình tự protein HbTPS6L-X1 và các trình tự TPS thuộc phân họ TPS-a.

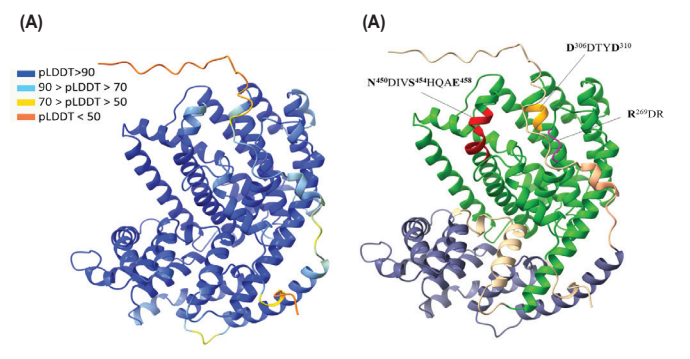
Hầu hết các TPS thuộc phân họ TPS-a không chứa đoạn peptide tín hiệu ở đầu N và chúng thường định vị ở tế bào chất, trừ một số ít trường hợp ngoại lệ như SITPS36 [18]. Những TPS-a không có peptide tín hiệu ở đầu N thường chỉ chứa một Met cách motif bảo tồn R(R)_XW không quá 18 axit amin về phía đầu N, còn các TPS-a có peptide tín hiệu sẽ chứa thêm vùng trình tự bắt đầu với Met ở trước Met vừa nêu [17].

Kết quả sắp giống cột trình tự ở hình 5 cho thấy, HbTPS6L-X1 chỉ có một Met mở đầu cách chuỗi trình tự căn chỉnh (align) với motif R(R)_XW 14 axit amin về phía đầu N. Bên cạnh đó, trình tự protein HbTPS6L-X1 có chiều dài tương đương với các protein TPS-a không có vùng peptide tín hiệu ở đầu N (AtTPS12

[31], AtTPS13 [31], GhDCS1 [32], SITPS12 [17], SITPS28 [18], SITPS31 [17], VvValCS [33], ZmTPS8 [34] và ZSS1 [35]). Chương trình TargetP 2.0 [22] cũng cho ra kết quả dự đoán tương tự. Tổng hợp các kết quả phân tích trình tự, HbTPS6L-X1 được dự đoán không chứa peptide tín hiệu ở đầu N và định vị ở tế bào chất - nơi quá trình sinh tổng hợp farnesyl diphosphate (FPP) xảy ra. Dự đoán này cũng phù hợp với nhận định rằng HbTPS6L-X1, một thành viên của phân họ TPS-a, có thể hoạt động như một sesquiterpene synthase, giúp chuyển hóa FPP hiện diện trong tế bào chất thành các sesquiterpene mạch hở.

3.4. HbTPS6L-X1 có mô hình cấu trúc dự đoán tương đồng chặt chẽ với các STS đã xác định

Do cấu trúc protein thường được bảo tồn hơn so với trình tự của nó, sự chồng lắp cấu trúc protein (superimposition) cho phép đánh giá mức độ tương đồng về cấu trúc và chức năng giữa protein nghiên cứu với protein có cấu trúc và chức năng đã được xác định bằng thực nghiệm. Dựa trên dữ liệu đầu vào là trình tự cấu trúc bậc một của HbTPS6L-X1 và các phân tử protein tương đồng đã được giải mã cấu trúc, phần mềm AlphaFold 2 dự đoán mô hình cấu trúc của HbTPS6L-X1 và hiển thị cấu trúc này theo độ tin cậy pLDDT cho từng axit amin trong chuỗi (hình 6). Kết quả cho thấy, ngoại trừ đoạn có cấu trúc xoắn ngắn chứa 20 axit amin đầu tiên ở đầu N và đoạn chứa 6 axit amin cuối cùng ở đầu C có giá trị pLDDT < 50, các axit amin ở phần còn lại của cấu trúc đều có pLDDT > 70, vị trí của các axit amin đó trong cấu trúc dự đoán có độ tin cậy cao.

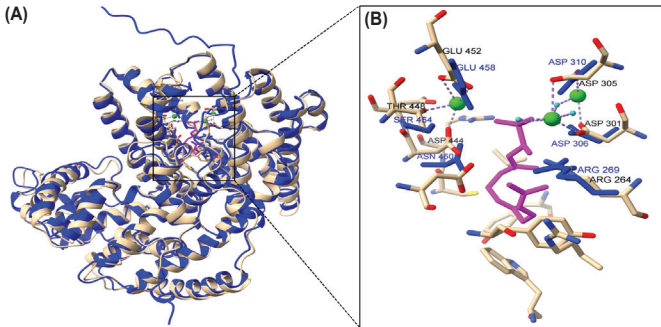


Hình 6. Mô hình cấu trúc dự đoán bởi AlphaFold 2 của HbTPS6L-X1. (A) Cấu trúc dự đoán của HbTPS6L-X1 được hiển thị theo độ tin cậy pLDDT cho từng axit amin trong chuỗi. Dãy màu biến thiên từ xanh đến đỏ tương ứng với độ tin cậy rất cao (pLDDT>90), cao (90>pLDDT>70), thấp (70>pLDDT>50) và rất thấp (pLDDT<50); (B) Cấu trúc của HbTPS6L-X1 với các motif bảo tồn đặc trưng và các miền chức năng được dự đoán bởi Interpro: miền β đầu N (màu xanh khaki), miền α đầu C (màu xanh lá) chứa các motif DDXXD (màu vàng), NSE/DTE (màu đỏ) và RXR (màu hồng).

Kết quả phân tích sự hiện diện các miền (domain) có thể có trong cấu trúc protein bằng Interpro [25] cho thấy, HbTPS6L-X1 chứa một miền cấu trúc ở đầu N (axit amin 37-200) (Interpro: IPR001906) và một miền ở đầu C (axit amin 227-550) (Interpro: IPR034741) (hình 6). Dự đoán này phù hợp với đặc điểm chung của đa số các terpene synthase với cấu trúc bao gồm các xoắn α và được tổ chức thành 2 miền: miền α ở đầu C chứa tâm hoạt động và miền β ở đầu N.

Theo chương trình phân tích xâu chuỗi (threading) LOMETS [27] 5-*epi*-aristolochene synthase từ cây thuốc lá (NtEAS4) có cấu trúc đứng đầu trong danh sách các mẫu cấu trúc được chọn làm khuôn (template) để mô hình hóa cấu trúc dự đoán của HbTPS6L-X1. Mẫu NtEAS4 được chọn (PDB id: 5EAT) có cấu hình tâm hoạt động đóng bao lấy Mg²⁺ (cofactor) và farnesyl hydroxyphosphonate - hợp chất tương đồng về cấu trúc với cơ chất FPP. NtEAS4 có trình tự tương đồng 64% với HbTPS6L-X1 (hình 5) và thuộc phân họ TPS-a với khả năng chuyển hóa FPP thành (+)-5-*epi*-aristolochene (sesquiterpene) [36].

Sự chồng lấp mô hình cấu trúc dự đoán bởi AlphaFold 2 của HbTPS6L-X1 với khuôn 5-*epi*-aristolochene synthase (NtEAS4) (PDB id: 5EAT) bằng ChimeraX cho giá trị căn bậc hai của độ lệch bình phương trung bình (Root Mean Square Deviation, RMSD) tính theo tọa độ của nguyên tử C_α là 1,388 Å. Giá trị RMSD này nhỏ hơn giá trị khuyến nghị [37] và phản ánh mức độ tương đồng cao giữa hai cấu trúc được chồng lấp (hình 7A).



Hình 7. HbTPS6L-X1 có mô hình cấu trúc dự đoán tương đồng chặt chẽ với 5-*epi*-aristolochene synthase (PDB id: 5EAT). (A) Sự chồng lấp mô hình cấu trúc dự đoán bởi AlphaFold của HbTPS6L-X1 (màu xanh dương đậm) và cấu trúc của 5-*epi*-aristolochene synthase (màu nâu nhạt) khi gắn với farnesyl hydroxyphosphonate - hợp chất có cấu trúc tương đồng với cơ chất FPP (màu hồng) (PDB id: 5EAT) thông qua phần mềm ChimeraX. Giá trị RMSD tính theo tọa độ của nguyên tử C_α là 1,388 Å; (B) Hình phóng đại vùng tâm hoạt động để hiển thị rõ các axit amin thiết yếu thuộc motif DDXXD, NSE/DTE và RXR tham gia vào hoạt động xúc tác của hai enzyme. Hình tròn màu xanh lá cây và xanh dương nhạt lần lượt mô phỏng Mg²⁺ (cofactor) và H₂O.

Khi quan sát kỹ vùng tâm hoạt động của hai protein trong mô hình chồng lấp (hình 7B), có thể thấy các axit amin trực tiếp tham gia vào hoạt động xúc tác thuộc 3 motif bảo tồn (DDXXD, NSE/DTE và RXR của HbTPS6L-X1) định vị chồng lấp hoặc sát với các axit amin đã được chứng minh có vai trò thiết yếu với hoạt tính xúc tác của STS nói chung và NtEAS4 nói riêng. Cụ thể, Asp³⁰⁶ và Asp³¹⁰ trong motif D³⁰⁶DTYD³¹⁰ của HbTPS6L-X1 có mặt ở cùng vị trí với Asp³⁰¹ và Asp³⁰⁵ trong motif (D³⁰¹DTFD³⁰⁵) của NtEAS4, có vai trò liên kết với Mg²⁺. Tương tự, ở motif NSE/DTE, bộ ba Asn⁴⁵⁰ - Ser⁴⁵⁴ - Glu⁴⁵⁸ (N⁴⁵⁰S⁴⁵⁴E⁴⁵⁸) của HbTPS6L-X1 gần như chồng lấp với bộ ba Asp⁴⁴⁴ - Thr⁴⁴⁸ - Glu⁴⁵² (D⁴⁴⁴T⁴⁴⁸E⁴⁵²) của NtEAS4 - chúng tạo liên kết phối trí với ion Mg²⁺ còn lại (hình 7B). Ngoài ra, Arg²⁶⁹ (motif R²⁶⁹DR) của HbTPS6L-X1 cũng định vị chồng lấp hoàn toàn với Arg²⁶⁴ (motif R²⁶⁴DR) của protein mẫu, và cùng với các Mg²⁺ được neo giữ bởi hai motif DDXXD và NSE/DTE cố định cơ chất FPP trong tâm hoạt động của enzyme và tạo điều kiện cho hoạt động xúc tác diễn ra.

4. Kết luận

Biến thể bản sao HbTPS6L-X1 phân lập từ vỏ thân của cây cao su *H. brasiliensis* RRIV 209 có kích thước 1677 bp và mã hóa chuỗi polypeptide dài 558 axit amin. Protein HbTPS6L-X1 chứa đủ 3 motif chức năng của TPS loại I là DDXXD, NSE/DTE và RXR. Các phân tích *in silico* cho thấy, HbTPS6L-X1 thuộc phân họ TPS-a, không chứa trình tự mã hóa peptide tín hiệu ở đầu N và không chứa motif R(R)X₈W tham gia trong quá trình đồng phân hóa tạo các sản phẩm terpene mạch vòng. Cấu trúc dự đoán bởi AlphaFold 2 của HbTPS6L-X1 chồng lấp với cấu trúc của 5-*epi*-aristolochene synthase (PDB id: 5EAT) (là một STS) - với giá trị RMSD phản ánh mức độ tương đồng cao giữa 2 cấu trúc được chồng lấp (1,388 Å). Những dữ liệu phân tích *in silico* đều ủng hộ khả năng HbTPS6L-X1 là một sesquiterpene synthase định vị trong tế bào chất, xúc tác chuyển hóa FPP thành các sesquiterpene mạch hở. Những đặc điểm chức năng *in silico* vừa nêu là cơ sở để thiết kế các khảo sát thực nghiệm nhằm khẳng định chức năng *in vitro* và *in planta* của HbTPS6L-X1.

Phụ lục

Kết quả sắp giống cột trình tự HbTPS6L-X1 đã phân lập với trình tự HbTPS6L-X1_ *in silico* (mã số truy cập XM_021794026.1).

```

XM_021794026.1 ATGGGAAGTGCACCTCATTTACAGCCAAACACTGACCAAAACAAAATCGCTCAGTGTCTC
HbTPS6L-X1 ATGGGAAGTGCACCTCATTTACAGCCAAACACTGACCAAAACAAAATCGCTCAGTGTCTC
*****
XM_021794026.1 CCTCCTCTGTGGGGATCTTACAGCTTACTCTTTTTCTTGTGCGGACTCAGAATAT
HbTPS6L-X1 CCTCCTCTGTGGGGATCTTACAGCTTACTCTTTTTCTTGTGCGGACTCAGAATAT
*****
XM_021794026.1 GAATCGTGGACTAGACGGTGGAAACCGCTCAAGAAAGGTCAAAGACATGTTAATGGCT
HbTPS6L-X1 GAATCGTGGACTAGACGGTGGAAACCGCTCAAGAAAGGTCAAAGACATGTTAATGGCT
*****
XM_021794026.1 GATGCAAGTAAATCCGATCAAGAAATTAATGATTAATGATTAATGATGCGCTCTGGTFTA
HbTPS6L-X1 GATGCAAGTAAATCCGATCAAGAAATTAATGATTAATGATTAATGATGCGCTCTGGTFTA
*****
XM_021794026.1 TCATATCATTTGAGAGGGAAATGAGAACAATTAATTCATATTTTAAATTTGGGTCTCT
HbTPS6L-X1 TCATATCATTTGAGAGGGAAATGAGAACAATTAATTCATATTTTAAATTTGGGTCTCT
*****
XM_021794026.1 CAACTCTTGATGAAATGATTAATGATCTACACTACAGCTCTCCCTTTCAAGTCTGTG
HbTPS6L-X1 CAACTCTTGATGAAATGATTAATGATCTACACTACAGCTCTCCCTTTCAAGTCTGTG
*****
XM_021794026.1 AGACAACAGGATCAAAATGACTGTGATGTGTCAAGAAATCAAGAGCAGCAATGGGA
HbTPS6L-X1 AGACAACAGGATCAAAATGACTGTGATGTGTCAAGAAATCAAGAGCAGCAATGGGA
*****
XM_021794026.1 GAGTTCAGAAGAGCATCATCAAGAGTGTGAGAGGATTTCTCAGTGTGATGAGCAAGT
HbTPS6L-X1 GAGTTCAGAAGAGCATCATCAAGAGTGTGAGAGGATTTCTCAGTGTGATGAGCAAGT
*****
XM_021794026.1 TTTATGAGTGTGGATGGAGAAGATATTCGATGAAGCTCTGAATTTACAAAGTCCACAC
HbTPS6L-X1 TTTATGAGTGTGGATGGAGAAGATATTCGATGAAGCTCTGAATTTACAAAGTCCACAC
*****
XM_021794026.1 CCGGACACCTGGCTCAAGTCAAAACCCATATCTCAAGTCAAGTCAAGTCAAGTCAAGT
HbTPS6L-X1 CCGGACACCTGGCTCAAGTCAAAACCCATATCTCAAGTCAAGTCAAGTCAAGTCAAGT
*****
XM_021794026.1 ATCCTCCGTTTCAAGGGAGTACCAGAGTGAAGGTTGGCAATACATCTTCTTCTAT
HbTPS6L-X1 ATCCTCCGTTTCAAGGGAGTACCAGAGTGAAGGTTGGCAATACATCTTCTTCTAT
*****
XM_021794026.1 GAAAAGAGAAATCCCAAATGAATTCCTACCGAATTTGCAAAATGATTTCAATCGA
HbTPS6L-X1 GAAAAGAGAAATCCCAAATGAATTCCTACCGAATTTGCAAAATGATTTCAATCGA
*****
XM_021794026.1 GTGCAATATTCCAGAGAGGAGCTCAAGTGTGATGGAGGTTGGGAAGACTAAAT
HbTPS6L-X1 GTGCAATATTCCAGAGAGGAGCTCAAGTGTGATGGAGGTTGGGAAGACTAAAT
*****
XM_021794026.1 CTGCAAGAAACTTCCTATGCAAGAGACAGACTGTGGAATATCTATCTGGCGTGT
HbTPS6L-X1 CTGCAAGAAACTTCCTATGCAAGAGACAGACTGTGGAATATCTATCTGGCGTGT
*****
XM_021794026.1 TCCATTCAAGTGTGAGCTCAACTGTCTTCAAGAGTATGATGGCAAGTATTTGAAA
HbTPS6L-X1 TCCATTCAAGTGTGAGCTCAACTGTCTTCAAGAGTATGATGGCAAGTATTTGAAA
*****
XM_021794026.1 ATGATATCACTAGCTGATGACACATACGATGCAATGCTAAATGAAGAGTCAATGCC
HbTPS6L-X1 ATGATATCACTAGCTGATGACACATACGATGCAATGCTAAATGAAGAGTCAATGCC
*****
XM_021794026.1 TFCAGAGTCTGATCGAGAGTGCAGATATGATGCTACTGACGAACGCCAGATPACATG
HbTPS6L-X1 TFCAGAGTCTGATCGAGAGTGCAGATATGATGCTACTGACGAACGCCAGATPACATG
*****
XM_021794026.1 AAAGTTTCTACAAGAAATTAAGTAAATTTTGGGAACAGAGAATATGGCGAGTAG
HbTPS6L-X1 AAAGTTTCTACAAGAAATTAAGTAAATTTTGGGAACAGAGAATATGGCGAGTAG
*****
XM_021794026.1 GAAGGAAGTCCCTACTGTGCTATACGATAAAGAGAAATTAAGAACTTTGTGATGCTC
HbTPS6L-X1 GAAGGAAGTCCCTACTGTGCTATACGATAAAGAGAAATTAAGAACTTTGTGATGCTC
*****
XM_021794026.1 TATCGATGGAGCACAGTGGTAAATGAAGGTTATGGCCCAACATTTGATGTTATTTG
HbTPS6L-X1 TATCGATGGAGCACAGTGGTAAATGAAGGTTATGGCCCAACATTTGATGTTATTTG
*****
XM_021794026.1 CAAAATGGTCTGATCAAGCGCTATAGCTAGTGTGATGAGCAGAGCTTATTTGGAATG
HbTPS6L-X1 CAAAATGGTCTGATCAAGCGCTATAGCTAGTGTGATGAGCAGAGCTTATTTGGAATG
*****
XM_021794026.1 GAAAAGATTCAGAGGATCAAGAAATATCAGTGGCTCAAGTCAACCAAAATTTGCAAA
HbTPS6L-X1 GAAAAGATTCAGAGGATCAAGAAATATCAGTGGCTCAAGTCAACCAAAATTTGCAAA
*****
XM_021794026.1 GCTCTCAAGTACTGGGGCTTGTGAGAACGACATAGTATCCCAAGGCTGAGCAAAAG
HbTPS6L-X1 GCTCTCAAGTACTGGGGCTTGTGAGAACGACATAGTATCCCAAGGCTGAGCAAAAG
*****
XM_021794026.1 AGGGAGACTGCTCTCAAGTGTGAGGCTATGAAAGACAGATGTTTGAAGAGAG
HbTPS6L-X1 AGGGAGACTGCTCTCAAGTGTGAGGCTATGAAAGACAGATGTTTGAAGAGAG
*****
XM_021794026.1 TGCATGAGCCCAACTGCTCTGCAAGTCTTATCTCAATTTGTTGGTTAACCTTGTCTGA
HbTPS6L-X1 TGCATGAGCCCAACTGCTCTGCAAGTCTTATCTCAATTTGTTGGTTAACCTTGTCTGA
*****
XM_021794026.1 ATACAGAGGTTTCTTACAGATTTGACGACTCTCACCAACCCATCAAGTCTGAAAGAT
HbTPS6L-X1 ATACAGAGGTTTCTTACAGATTTGACGACTCTCACCAACCCATCAAGTCTGAAAGAT
*****
XM_021794026.1 AAAATACAGCATGCTTCTGGAGCCTCTCTTGGCAAGCAGTAAATATATA
HbTPS6L-X1 AAAATACAGCATGCTTCTGGAGCCTCTCTTGGCAAGCAGTAAATATATA
*****

```

LỜI CẢM ƠN

Nhóm nghiên cứu chân thành cảm ơn TS Trần Thanh - Viện Nghiên cứu Cao su Việt Nam đã cung cấp mẫu vỏ thân cây cao su *H. brasiliensis* RRIV 209, ThS Phạm Thị Mỹ Bình và sinh viên Võ Thị Kim Thảo - Bộ môn Sinh hóa, Khoa Sinh học - Công nghệ sinh học, Trường Đại học Khoa học Tự nhiên, Đại học Quốc gia TP Hồ Chí Minh đã hỗ trợ quá trình thu và bảo quản mẫu.

TÀI LIỆU THAM KHẢO

- [1] C. Wu, L. Lan, Y. Li, et al. (2018), "The relationship between latex metabolism gene expression with rubber yield and related traits in *Hevea brasiliensis*", *BMC Genomics*, **19**(1), DOI: 10.1186/s12864-018-5242-4.
- [2] D.J. McGarvey, R. Croteau (1995), "Terpenoid metabolism", *Plant Cell*, **7**(7), pp.1015-1026, DOI: 10.1105/tpc.7.7.1015.
- [3] M.R. Concepción, A. Boronat (2002), "Elucidation of the methylerythritol phosphate pathway for isoprenoid biosynthesis in bacteria and plastids: A metabolic milestone achieved through genomics", *Plant Physiology*, **130**(3), pp.1079-1089, DOI: 10.1104/pp.007138.
- [4] J.D. Connolly, R.A. Hill (1991), *Dictionary of Terpenoids*, CiNii, DOI: 10.1007/978-1-4899-4513-6, <http://link.springer.com/content/pdf/10.1007/978-1-4899-4513-6.pdf>, accessed 25 January 2023.
- [5] F. Chen, D. Tholl, J. Bohlmann, et al. (2011), "The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom", *Plant J.*, **66**, pp.212-229, DOI: 10.1111/j.1365-3113X.2011.04520.x.
- [6] D.W. Christianson (2017), "Structural and chemical biology of terpenoid cyclases", *Chem. Rev.*, **117**(17), pp.11570-11648, DOI: 10.1021/acs.chemrev.7b00287.
- [7] V. Falara, J.M. Alba, M.R. Kant, et al. (2014), "Geranylinalool synthases in solanaceae and other angiosperms constitute an ancient branch of diterpene synthases involved in the synthesis of defensive compounds", *Plant Physiol.*, **166**(1), pp.428-4411, DOI: 10.1104/pp.114.243246.
- [8] Y. Matsuba, T.T.H. Nguyen, K. Wiegert, et al. (2013), "Evolution of a complex locus for terpene biosynthesis in *Solanum*", *The Plant Cell*, **25**(6), pp.2022-2036.
- [9] Y. Matsuba, J. Zi, A.D. Jones, et al. (2015), "Biosynthesis of the diterpenoid lycosantalol via nerylerythritol diphosphate in *Solanum lycopersicum*", *PLOS ONE*, **10**(3), DOI: 10.1371/journal.pone.0119302.
- [10] K. Hayashi, H. Kawaide, M. Notomi, et al. (2006), "Identification and functional analysis of bifunctional ent-kaurene synthase from the moss *Physcomitrella patens*", *FEBS Lett.*, **580**, pp.6175-6181, DOI: 10.1016/j.febslet.2006.10.018.
- [11] F. Zhou, E. Pichersky (2020a), "More is better: The diversity of terpene metabolism in plants", *Curr. Opin. Plant Biol.*, **55**, pp.1-10, DOI: 10.1016/j.pbi.2020.01.005.
- [12] C.A. Lesburg (1997), "Crystal structure of pentalene synthase: Mechanistic insights on terpenoid cyclization reactions in biology", *Science*, **277**(5333), pp.1820-1824, DOI: 10.1126/science.277.5333.1820.
- [13] M.J. Rynkiewicz, D.E. Cane, D.W. Christianson (2001), "Structure of trichodiene synthase from *Fusarium sporotrichioides* provides mechanistic inferences on the terpene cyclization cascade", *Proc. Natl. Acad. Sci.*, **98**(24), pp.13543-13548, DOI: 10.1073/pnas.23131309.
- [14] C.M. Starks, K. Back, J. Chappell, et al. (1997), "Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-Aristolochene Synthase", *Science*, **277**, pp.1815-1820, DOI: 10.1126/science.277.5333.1815.
- [15] J. Durairaj, A.Di. Girolamo, H.J. Bouwmeester, et al. (2019), "An analysis of characterized plant sesquiterpene synthases", *Phytochemistry*, **158**, pp.157-165, DOI: 10.1016/j.phytochem.2018.10.020.
- [16] N.S. Lau, Y. Makita, M. Kawashima, et al. (2016), "The rubber tree genome shows expansion of gene family associated with rubber biosynthesis", *Sci. Rep.*, **6**, DOI: 10.1038/srep28594.
- [17] V. Falara, T.A. Akhtar, T.T.H. Nguyen, et al. (2011), "The tomato terpene synthase gene family", *Plant Physiol.*, **157**(2), pp.770-789, DOI: 10.1104/pp.111.179648.
- [18] F. Zhou, E. Pichersky (2020b), "The complete functional characterisation of the terpene synthase family in tomato", *New Phytol.*, **226**(5), pp.1341-1360, DOI: 10.1111/nph.16431.
- [19] J.H. Cepas, F. Serra, P. Bork (2016), "ETE 3: Reconstruction, analysis, and visualization of phylogenomic data", *Mol. Biol. Evol.*, **33**(6), pp.1635-1638, DOI: 10.1093/molbev/msw046.
- [20] R.C. Edgar (2004), "MUSCLE: A multiple sequence alignment method with reduced time and space complexity", *BMC Bioinformatics*, **5**, DOI: 10.1186/1471-2105-5-113.
- [21] L.T. Nguyen, H.A. Schmidt, A.V. Haeseler, et al. (2015), "IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies", *Mol. Biol. Evol.*, **32**(1), pp.268-274, DOI: 10.1093/molbev/msu300.
- [22] J.J.A. Armenteros, M. Salvatore, O. Emanuelsson, et al. (2019), "Detecting sequence signals in targeting peptides using deep learning", *Life Sci. Alliance*, **2**(5), DOI: 10.26508/lsa.201900429.
- [23] M. Mirdita, K. Schütze, Y. Moriwaki, et al. (2022), "ColabFold: Making protein folding accessible to all", *Nat. Methods*, **19**(6), pp.679-682, DOI: 10.1038/s41592-022-01488-1.
- [24] J. Jumper, R. Evans, A. Pritzel, et al. (2021), "Highly accurate protein structure prediction with AlphaFold", *Nature*, **596**(7873), pp.583-589, DOI: 10.1038/s41586-021-03819-2.
- [25] T.P. Lafosse, M. Blum, S. Chuguransky, et al. (2023), "InterPro in 2022", *Nucleic Acids Res.*, **51**(D1), pp.418-427, DOI: 10.1093/nar/gkac993.
- [26] E.F. Pettersen, T.D. Goddard, C.C. Huang, et al. (2021), "UCSF ChimeraX: Structure visualization for researchers, educators, and developers", *Protein Sci.*, **30**(1), pp.70-82, DOI: 10.1002/pro.3943.
- [27] W. Zheng, Q. Wuyun, X. Zhou, et al. (2022), "LOMETS3: Integrating deep learning and profile alignment for advanced protein template recognition and function annotation", *Nucleic Acids Res.*, **50**(W1), pp.W454-W464, DOI: 10.1093/nar/gkac248.
- [28] D.E. Cane, Q. Xue, B.C. Fitzsimons (1996), "Trichodiene synthase. probing the role of the highly conserved aspartate-rich region by site-directed mutagenesis", *Biochemistry*, **35**(38), pp.12369-12376, DOI: 10.1016/S0959-440X(98)80088-2.
- [29] D.W. Christianson (2006), "Structural biology and chemistry of the terpenoid", *Cyclases Chem. Rev.*, **106**(8), pp.3412-3442, DOI: 10.1021/cr050286w.
- [30] D.C. Williams, D.J. McGarvey, E.J. Katahira, et al. (1998), "Truncation of limonene synthase preprotein provides a fully active 'Pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair", *Biochemistry*, **37**(35), pp.12213-12220, DOI: 10.1104/pp.120.3.879.
- [31] D.K. Ro, J. Ehrling, C.I. Keeling, et al. (2006), "Microarray expression profiling and functional characterization of AtTPS genes: Duplicated *Arabidopsis thaliana* sesquiterpene synthase genes At4g13280 and At4g13300 encode root-specific and wound-inducible (Z)- γ -bisabolene synthases", *Arch. Biochem. Biophys.*, **448**(1-2), pp.104-116, DOI: 10.1177/1934578X1000500507.
- [32] X.Y. Chen, Y. Chen, P. Heinstein, et al. (1995), "Cloning, expression, and characterization of (+)- δ -cadinene synthase: A catalyst for cotton phytoalexin biosynthesis", *Arch. Biochem. Biophys.*, **324**(2), pp.255-266, DOI: 10.1006/abbi.1995.0038.
- [33] J. Lückner, P. Bowen, J. Bohlmann (2004), "*Vitis vinifera* terpenoid cyclases: Functional identification of two sesquiterpene synthase cDNAs encoding (+)-valencene synthase and (-)-germacrene D synthase and expression of mono- and sesquiterpene synthases in grapevine flowers and berries", *Phytochemistry*, **65**(19), pp.2649-2659, DOI: 10.1016/j.phytochem.2004.08.017.
- [34] A. Fontana, M. Held, C.A. Fantaye, et al. (2011), "Attractiveness of constitutive and herbivore-induced sesquiterpene blends of maize to the parasitic wasp *Cotesia marginiventris* (Cresson)", *J. Chem. Ecol.*, **37**(6), pp.582-591, DOI: 10.1007/s10886-011-9967-7.
- [35] F. Yu, S. Okamoto, K. Nakasone, et al. (2008), "Molecular cloning and functional characterization of α -humulene synthase, a possible key enzyme of zerumbone biosynthesis in shampoo ginger (*Zingiber zerumbet* Smith)", *Planta*, **227**(6), pp.1291-1299, DOI: 10.1007/s00425-008-0700-x.
- [36] P.E. O'Maille, J. Chappell, J.P. Noel (2006), "Biosynthetic potential of sesquiterpene synthases: Alternative products of tobacco 5-epi-aristolochene synthase", *Arch. Biochem. Biophys.*, **448**(1-2), pp.73-82, DOI: 10.1016/j.abb.2005.10.028.
- [37] C. Chothia, A.M. Lesk (1986), "The relation between the divergence of sequence and structure in proteins", *EMBO J.*, **5**(4), pp.823-826, DOI: 10.1002/j.1460-2075.1986.tb04288.x.