

IDENTIFICATION, STRUCTURAL AND PHYLOGENY ANALYSES OF THE SUCROSE TRANSPORTER FAMILY IN *CERATOPTERIS RICHARDII*

Chu Duc Ha^{1*}, Ha Thi Quyen¹, Pham Chau Thuy¹, Tran Dang Khoa¹

La Viet Hong², Nguyen Thi Yen Linh³, Tran Van Tien⁴, Trinh Thi Thu Thuy³

¹University of Engineering and Technology – VNU, ²Hanoi Pedagogical University 2

³Vietnam National University of Agriculture, ⁴National Academy of Public Administration

ARTICLE INFO	ABSTRACT
Received: 01/4/2022	The purpose of this research was to identify and characterize the sucrose transporters (SWEET) in <i>Ceratopteris richardii</i> , one of the major ornamental fern houseplants. By screening the conserved domain, a total of 20 members of the CericSWEET family has been identified in the assembly of <i>C. richardii</i> . Our comparison revealed that the numbers of members in the SWEET family were variable between plant species. Subcellular localization showed that CericSWEET proteins were localized in most essential organelles. Next, the CericSWEET family exhibited the variable physic-chemical properties, while a large number of genes contained 6 exons like reported in other plant species. Based on the Neighbor-Joining phylogenetic tree, the CericSWEET family could be classified into 3 sub-groups. Two genes, <i>CericSWEET07</i> and <i>12</i> were hypothesized to function in the drought response. Additionally, numerous drought- and abscisic acid-responsive <i>cis</i> - regulatory elements have been found in the promoter regions of the <i>CericSWEET</i> genes. Taken together, our results could provide an important foundation for the CericSWEET family in <i>C. richardii</i> .
Revised: 24/6/2022	
Published: 24/6/2022	
KEYWORDS	
Structure	
<i>Ceratopteris richardii</i>	
Characteristic	
Bioinformatics	
Sucrose transporter	

XÁC ĐỊNH, PHÂN TÍCH CẤU TRÚC VÀ PHÂN NHÓM PROTEIN VẬN CHUYỂN SUCROSE TRONG CÂY *CERATOPTERIS RICHARDII*

Chu Đức Hà¹, Hà Thị Quyên¹, Phạm Châu Thùy¹, Trần Đăng Khoa¹

La Việt Hồng², Nguyễn Thị Yên Linh³, Trần Văn Tiến⁴, Trịnh Thị Thu Thủy³

¹Trường Đại học Công nghệ - ĐH Quốc gia Hà Nội, ²Trường Đại học Sư phạm Hà Nội 2

³Học viện Nông nghiệp Việt Nam, ⁴Học viện Hành chính Quốc gia

THÔNG TIN BÀI BÁO	TÓM TẮT
Ngày nhận bài: 01/4/2022	Mục tiêu của nghiên cứu nhằm xác định và phân tích các đặc điểm cơ bản của protein vận chuyển đường sucrose (SWEET) ở cây <i>Ceratopteris richardii</i> , một trong số những loài cây dương xỉ cảnh được trồng trong nhà. Bằng việc sàng lọc vùng bảo thủ đặc trưng, tổng số 20 thành viên của họ protein CericSWEET đã được xác định trong hệ tham chiếu của <i>C. richardii</i> . Kết quả so sánh cho thấy số lượng thành viên của họ SWEET đa dạng giữa các loài thực vật. Phân tích vị trí cư trú nội bào cho thấy các protein CericSWEET cư trú tại hầu hết bào quan thiết yếu. Tiếp theo, nhóm CericSWEET có tính chất lý hóa đa dạng, trong khi hầu hết các gen đều có 6 exon tương tự như ở các loài thực vật khác. Dựa vào sơ đồ hình cây Neighbor-Joining, nhóm CericSWEET gồm 3 phân nhóm. Trong đó, hai gen, <i>CericSWEET07</i> và <i>12</i> được dự đoán có thể có chức năng đáp ứng với điều kiện hạn. Bên cạnh đó, nhiều yếu tố điều hòa <i>cis</i> - đáp ứng hạn và abscisic acid đã được phân tích vùng promoter của các gen <i>CericSWEET</i> . Tóm lại, kết quả của nghiên cứu này đã cung cấp những dẫn liệu quan trọng về nhóm CericSWEET ở <i>C. richardii</i> .
Ngày hoàn thiện: 24/6/2022	
Ngày đăng: 24/6/2022	
TỪ KHÓA	
Cấu trúc	
<i>Ceratopteris richardii</i>	
Đặc tính	
Tin sinh học	
Vận chuyển sucrose	

DOI: <https://doi.org/10.34238/tnu-jst.5793>

* Corresponding author. Email: cd.ha@vnu.edu.vn

1. Giới thiệu

Phát triển cây công trình và cây cảnh trong nhà tại các khu vực đô thị được xem là một trong những giải pháp xanh hóa và bảo vệ môi trường trước tác động của tình trạng biến đổi khí hậu và sự nóng lên toàn cầu [1]. Trong đó, các loài cây cảnh đặt trong nhà được chứng minh là có tác dụng giảm hiệu ứng nhà kính, đồng thời hỗ trợ làm sạch và tăng chất lượng không khí, góp phần xanh hóa khuôn viên [1]. Trong số các loại cây cảnh trong nhà hiện nay, *Ceratopteris richardii*, một loài dương xỉ trong họ Pteridaceae, được đánh giá là một trong những cây xử lý môi trường có thể giải quyết các vấn đề liên quan đến cải thiện không khí trong nhà. Bên cạnh đó, *C. richardii* cũng được báo cáo là loài thực vật có mạch mô hình điển hình cho nghiên cứu sự biến đổi của thành tế bào [2], với thông tin di truyền mới được giải mã hoàn chỉnh gần đây [3]. Chính vì vậy, tìm hiểu về sinh trưởng và phát triển của cây *C. richardii* đang thu hút được sự quan tâm, nhất là ở cấp độ phân tử.

Trong quá trình sinh trưởng và phát triển ở thực vật nói chung, rất nhiều nhóm protein điều hòa (như nhân tố phiên mã) và protein chức năng (như enzyme hoặc protein vận chuyển) tham gia. Trong đó, nhóm protein vận chuyển đường sucrose, ‘Sugars Will Eventually be Exported Transporters’ (SWEET) có chức năng vận chuyển các phân tử đường monosaccharide và disaccharide tại các cơ quan/bộ phận, từ đó kiểm soát các quá trình sinh học diễn ra trong cây trồng [4]. Đến nay, các nghiên cứu về nhóm SWEET đã được báo cáo trên một số đối tượng cây quan trọng như lúa gạo (*Oryza sativa*) [5], cà chua (*Solanum lycopersicum*) [6], đậu tương (*Glycine max*) [7], bông vải (*Gossypium* spp.) [8], lúa mì (*Triticum aestivum*) [9]. Tuy nhiên, nghiên cứu về nhóm protein SWEET trên cây *C. richardii* vẫn chưa được ghi nhận.

Nghiên cứu này được thực hiện nhằm xác định và phân tích nhóm protein SWEET ở cây *C. richardii*. Theo đó, các mã định danh gen và vị trí phân bố của các thành viên được phân tích bằng công cụ tin sinh học. Dữ liệu về trình tự gen và protein được khai thác để phân tích tính chất lý hóa, vị trí cư trú nội bào, cấu trúc gen và phân nhóm. Cuối cùng, sơ đồ hình cây giữa protein SWEET ở *C. richardii* và protein SWEET đã biết chức năng đã được thiết lập nhằm đưa ra giả thuyết về vai trò của gen liên quan đến cơ chế đáp ứng hạn ở thực vật.

2. Vật liệu và phương pháp nghiên cứu

2.1. Vật liệu nghiên cứu

Nghiên cứu này đã sử dụng nguồn dữ liệu là genome, proteome của cây *C. richardii* được khai thác dựa theo báo cáo trước đây [3] trên cổng Phytozome [10] và NCBI.

2.2. Phương pháp nghiên cứu

- Phương pháp xác định và chú giải SWEET: Vùng bảo thủ đặc trưng của protein SWEET ở thực vật (mã Pfam: PF03083) [4] được truy vấn trên proteome của *C. richardii* [3] trên Phytozome [10] theo mô tả trong nghiên cứu trước đây [11]. Theo đó, giá trị cut-off được lựa chọn là 1e-10 nhằm sàng lọc tất cả kết quả tìm kiếm hữu hiệu [11]. Trình tự amino acid đầy đủ tiếp tục được khai thác để đối chiếu bằng công cụ BlastP trên NCBI để xác định mã định danh gene của từng ứng viên. Mã định danh protein, vị trí gene trên nhiễm sắc thể (NST), trình tự đoạn mã hóa (coding DNA sequence, CDS) và trình tự vùng gene (genomic DNA sequence, gDNA) được thu thập cho các phân tích tiếp theo.

- Phương pháp dự đoán vị trí phân bố dưới tế bào của SWEET: Sử dụng trình tự amino acid của các protein SWEET ở *C. richardii* phân tích trên công cụ Yloc [12] dựa theo mô tả trong nghiên cứu trước đây [11]. Hai chỉ tiêu, gồm mức độ tin cậy (%) và xác suất phân bố (%) của từng protein SWEET được đánh giá.

- Phương pháp dự đoán tính chất lý hóa của SWEET: Sử dụng trình tự amino acid của các protein SWEET ở *C. richardii* để xử lý trên ExPASy ProtParam [13] dựa theo mô tả trong nghiên cứu trước đây [11]. Trong đó, ba tính chất của protein, bao gồm kích thước, điểm đẳng điện và độ

ura nước trung bình được tính toán dựa trên tính chất của các amino acid. Protein thể hiện tính acid, trung tính, base với giá trị điểm đẳng điện tương ứng < 7 , $= 7$ và > 7 , trong khi protein có tính ưa nước và kỵ nước nêu giá trị độ ưa nước trung bình tương ứng < 0 và ≥ 0 [13].

- Phương pháp xây dựng sơ đồ hình cây của SWEET: Trình tự amino acid của các protein SWEET ở *C. richardii* được sử dụng để xây dựng sơ đồ hình cây trên MEGA [14] dựa theo mô tả trong nghiên cứu trước đây [11] với các tham số: thuật toán Neighbor-Joining, với giá trị bootstrap 1000. Bên cạnh đó, trình tự amino acid của các protein SWEET ở loài *Cicer arietinum* mã hóa bởi gen được chứng minh có đáp ứng với điều kiện thiếu nước khai thác trong nghiên cứu trước đây [11] được kết hợp với nhóm protein SWEET ở *C. richardii* để xây dựng sơ đồ hình cây với giá trị cut-off 50% [14].

- Phương pháp phân tích cấu trúc gen của SWEET: Trình tự CDS và gDNA của các gen mã hóa cho protein SWEET ở *C. richardii* được sử dụng để mô hình hóa cấu trúc gen (trật tự exon/intron) bằng GSDS [15]. Theo đó, thứ tự của các gen được sắp xếp theo phân loại của sơ đồ cây Neighbor-Joining [14]. Kết quả được minh họa bằng Adobe Illustrator.

- Phương pháp phân tích vùng promoter của gen mã hóa SWEET: Trình tự 1000 bp trong vùng promoter của gen mã hóa cho protein SWEET ở *C. richardii* được sử dụng để phân tích các yếu tố điều hòa *cis*- đáp ứng với hạn, bao gồm MYB-binding site (MBS), dehydration responsive element (DRE), MYC-binding site (MYCR), T/G Box, evening element (EE), binding site of drought-inducible NAC transcription factors (NACR), coupling element 3 (CE3), induction of CBF expression region 2 (ICER2) và TC-rich repeats và yếu tố cảm ứng abscisic acid (abscisic acid responsive element, ABRE) theo mô tả trong nghiên cứu trước đây [11].

3. Kết quả và bàn luận

3.1. Xác định và phân tích vị trí phân bố của nhóm gene mã hóa protein vận chuyển đường sucrose ở cây *Ceratopteris richardii*

Kết quả sàng lọc ($E\text{-value} \leq 1e-10$) đã xác định được tổng số 20 thành viên của nhóm SWEET ở *C. richardii*. Theo đó, các thông tin chú giải, bao gồm mã định danh gen và protein, và vị trí phân bố được khai thác từ dữ liệu của *C. richardii*. Dựa trên vị trí phân bố trên nhiễm sắc thể, các thành viên của nhóm SWEET ở *C. richardii* được đặt tên từ CericSWEET01 đến 20, với ‘Ceric’ đại diện cho tên khoa học *Ceratopteris richardii*, ‘SWEET’ đại diện cho nhóm protein vận chuyển đường sucrose và số thứ tự đại diện cho thứ tự xuất hiện trên các NST của gen. Thông tin định danh cơ bản của nhóm SWEET được thể hiện ở Bảng 1.

Theo đó, nhóm gen CericSWEET phân bố rải rác trên khắp các NST ở hệ gen của *C. richardii*. Cụ thể, NST số 11 chứa nhiều gen CericSWEET nhất, với sáu thành viên (tương ứng 30%), lần lượt tiếp theo là NST số 4 (ba thành viên, tương ứng 15%) và NST số 23 (hai thành viên, tương ứng 10%) (Bảng 1). Các NST số 1, 2, 6, 13, 18, 24, 31, 32 và 37 lần lượt chỉ chứa một thành viên (tương ứng 3,33%), trong khi không có gen CericSWEET nào phân bố trên 27 NST còn lại trong hệ gen của *C. richardii* (Bảng 1).

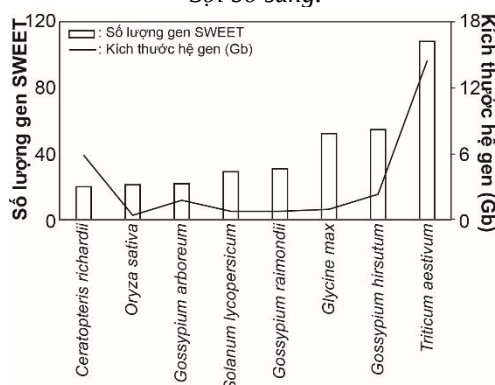
Đối chiếu với các nghiên cứu trước đây, số lượng thành viên của nhóm protein SWEET giữa các loài thực vật tương đối đa dạng (Hình 1). Cụ thể, tổng số 21 thành viên của nhóm SWEET đã được báo cáo ở lúa gạo [5], trong khi 29 và 52 protein SWEET đã được ghi nhận ở cà chua [6] và đậu tương [7]. Ở các cây bông, 22, 31 và 55 thành viên đã được lần lượt xác định ở nhóm protein SWEET của *G. arboreum*, *G. raimondii*, *G. hirsutum* [8]. Đáng chú ý, 108 protein SWEET đã được tìm hiểu và phân tích trên hệ tham chiếu của lúa mì [9]. Nhóm SWEET ở *C. richardii* (kích thước hệ gen = 5,85 Gb) có số lượng thành viên lần lượt ít hơn so với lúa gạo (kích thước hệ gen = 0,39 Gb) [5], *G. arboreum* (kích thước hệ gen = 1,78 Gb) [8], cà chua (kích thước hệ gen = 0,81 Gb) [6], *G. raimondii* (kích thước hệ gen = 0,76 Gb) [8], đậu tương (kích thước hệ gen = 0,99 Gb) [7], *G. hirsutum* (kích thước hệ gen = 2,31 Gb) [8] và lúa mì (kích thước hệ gen = 14,45 Gb) [9] (Hình 1). Có thể thấy rằng, nhóm protein SWEET ở thực vật tương đối đa dạng và số lượng thành viên không phụ thuộc vào kích thước hệ gen cũng như số lượng NST giữa các loài.

Bảng 1. Thông tin về nhóm SWEET ở cây *Ceratopteris richardii*

STT	Tên gene	GeneID	ProteinID	Vị trí phân bố
1	CericSWEET01	Ceric.01G117600	KAH7447702	NST01:3'-222956586..222959957-5'
2	CericSWEET02	Ceric.02G033800	KAH7443425	NST02:5'-65785421..65790015-3'
3	CericSWEET03	Ceric.04G043900	KAH7439069	NST04:3'-73063695..73065752-5'
4	CericSWEET04	Ceric.04G091100	KAH7440096	NST04:5'-170959268..170964738-3'
5	CericSWEET05	Ceric.04G045000	KAH7439095	NST04:3'-76095797..76097460-5'
6	CericSWEET06	Ceric.06G033300	KAH7434756	NST06:3'-89235784..89240132-5'
7	CericSWEET07	Ceric.11G041300	KAH7425137	NST11:3'-78491517..78495760-5'
8	CericSWEET08	Ceric.11G057700	KAH7425498	NST11:3'-112703081..112705138-5'
9	CericSWEET09	Ceric.11G057800	KAH7425499	NST11:3'-113071004..113115910-5'
10	CericSWEET10	Ceric.11G057300	KAH7425494	NST11:5'-112352898..112355102-3'
11	CericSWEET11	Ceric.11G057100	KAH7425492	NST11:5'-112258390..112259620-3'
12	CericSWEET12	Ceric.11G041200	KAH7425136	NST11:5'-78339226..78343982-3'
13	CericSWEET13	Ceric.13G035200	KAH7420992	NST13:5'-68146481..68154545-3'
14	CericSWEET14	Ceric.18G081100	KAH7366493	NST18:3'-171375045..171380360-5'
15	CericSWEET15	Ceric.23G039500	KAH7301734	NST23:5'-84488389..84493304-3'
16	CericSWEET16	Ceric.23G037700	KAH7301674	NST23:5'-79403060..79436928-3'
17	CericSWEET17	Ceric.24G040500	KAH7300002	NST24:3'-76766300..76789750-5'
18	CericSWEET18	Ceric.31G056600	KAH7289084	NST31:5'-108297596..108310334-3'
19	CericSWEET19	Ceric.32G024800	KAH7286841	NST32:3'-46734764..46738697-5'
20	CericSWEET20	Ceric.37G021000	KAH7279467	NST37:3'-48027674..48034566-5'

GeneID: Mã định danh gen, ProteinID: Mã định danh protein, NST: Nhiễm sắc thể, 3'-5': Sợi xuôi, 5'-3':

Sợi bổ sung.



Hình 1. So sánh số lượng thành viên của nhóm SWEET và kích thước hệ gen ở một số loài thực vật

3.2. Dự đoán vị trí cư trú dưới tế bào và dự đoán tính chất lý hóa của protein vận chuyển đường sucrose ở cây *Ceratopteris richardii*

Phân tích vị trí phân bố dưới tế bào của protein có thể cung cấp những dẫn liệu định hướng cho nghiên cứu chức năng gen. Trong nghiên cứu này, trình tự amino acid đầy đủ của 20 thành viên của nhóm SWEET ở *C. richardii* được khảo sát trên công cụ Yloc [12]. Kết quả phân tích cho thấy, phần lớn thành viên (18 trên tổng số 20) của nhóm SWEET nằm trên màng sinh chất. Trong đó, 12 (trên tổng số 18) protein SWEET đều có xác suất phân bố tại màng sinh chất lớn hơn 90%, với mức độ tin cậy lớn hơn 0,70, ngoại trừ CericSWEET13 và 15 được dự đoán nằm ở màng sinh chất với độ tin cậy thấp, đạt lần lượt là 0,18 và 0,44 (Bảng 2). Kết quả cũng chỉ ra hai thành viên còn lại của nhóm SWEET ở *C. richardii*, bao gồm CericSWEET09 và 10 có thể phân bố tại lục lạp và peroxisome, tuy nhiên xác suất phân bố và mức độ tin cậy chỉ đạt mức trung bình (Bảng 2). Sự phân bố của nhóm protein SWEET tại màng sinh chất được giải thích do cấu trúc chuỗi amino acid của các thành viên này đều có tính kỵ nước, nhất là ở đầu N- [12], tương tự như mô tả trong nghiên cứu trước đây [4]. Kết quả này cũng được kiểm chứng bởi các nghiên

cứu về nhóm SWEET trước đây. Cụ thể, 5 SWEET ở cà chua được dự đoán phân bố trên màng sinh chất [6], trong khi 22 (trên tổng số 25) protein SWEET ở cây *Juglans regia* cũng được xác định nằm ở màng sinh chất [16]. Gần đây, tổng số 168 protein SWEET từ các loài *Gossypium* spp. đã được báo cáo cư trú tại màng sinh chất [8].

Bảng 2. Phân tích vị trí cư trú nội bào của nhóm SWEET ở cây *Ceratopteris richardii*

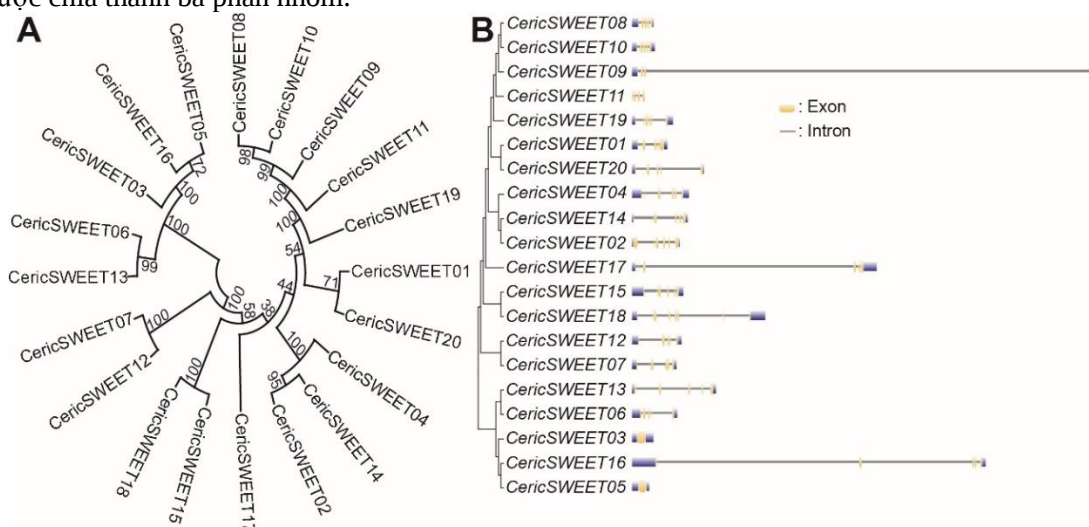
STT	Tên protein	Vị trí cư trú	Xác suất phân bố (%)	Mức độ tin cậy	Kích thước	Điểm đẳng điện	Độ ưa nước
1	CericSWEET01	Màng sinh chất	98,41	0,98	329	9,44	0,46
2	CericSWEET02	Màng sinh chất	99,87	0,95	428	9,72	0,38
3	CericSWEET03	Màng sinh chất	95,07	0,90	277	7,67	0,63
4	CericSWEET04	Màng sinh chất	99,02	0,99	233	9,68	0,86
5	CericSWEET05	Màng sinh chất	53,79	0,81	267	9,53	0,57
6	CericSWEET06	Màng sinh chất	86,75	0,06	259	9,42	0,70
7	CericSWEET07	Màng sinh chất	94,05	0,72	259	9,05	0,58
8	CericSWEET08	Màng sinh chất	99,76	0,82	255	9,39	0,49
9	CericSWEET09	Lục lạp	64,54	0,09	218	9,98	0,71
10	CericSWEET10	Peroxisome	33,81	0	255	9,59	0,47
11	CericSWEET11	Màng sinh chất	51,25	0,04	202	9,69	0,34
12	CericSWEET12	Màng sinh chất	64,84	0,11	255	9,01	0,69
13	CericSWEET13	Màng sinh chất	99,30	0,18	300	6,98	0,61
14	CericSWEET14	Màng sinh chất	57,36	0,82	273	9,02	0,71
15	CericSWEET15	Màng sinh chất	94,37	0,44	242	9,52	0,81
16	CericSWEET16	Màng sinh chất	96,80	0,79	274	8,68	0,39
17	CericSWEET17	Màng sinh chất	85,21	0,12	267	9,33	0,49
18	CericSWEET18	Màng sinh chất	100	0,96	273	8,57	0,60
19	CericSWEET19	Màng sinh chất	99,81	0,97	253	9,55	0,75
20	CericSWEET20	Màng sinh chất	95,32	0,92	258	9,20	0,58

Kích thước, điểm đẳng điện và độ ưa nước trung bình của nhóm CericSWEET được trình bày ở Bảng 2. Kết quả cho thấy, kích thước của các phân tử dao động từ 202 (CericSWEET11) đến 428 (CericSWEET02) amino acid, trong đó, hầu hết các protein (14 trên tổng số 20) có kích thước tương đối đồng đều (240 - 280 amino acid) (Bảng 2). Đa số thành viên (19 trên tổng số 20) của nhóm CericSWEET đều có tính base (điểm đẳng điện > 7,0), ngoại trừ CericSWEET13 có tính trung tính với giá trị điểm đẳng điện đạt 6,98 (Bảng 2). Đáng chú ý, tất cả các thành viên của nhóm protein vận chuyển đường sucrose ở *C. richardii* đều thể hiện tính kỵ nước, với độ ưa nước trung bình đạt giá trị dương (Bảng 2).

Trong các nghiên cứu trước đây, tính chất của protein SWEET ở một số loài thực vật cũng đã lần lượt được mô tả [4]. Cụ thể, nhóm protein SWEET ở đậu tương có kích thước 155 (GmSWEET3) đến 316 (GmSWEET2) amino acid, với 37 trên tổng số 52 thành viên có kích thước ổn định trong khoảng 240 - 280 amino acid [7]. Ở cà chua, nhóm protein SWEET có kích thước dao động 233 - 308 amino acid [6]. Trong khi đó, các thành viên của nhóm protein SWEET ở loài *G. arboreum* có kích thước tương đối ổn định, từ 138 - 300 amino acid, tương ứng 14,84 - 33,53 kDa, với 20 trên tổng số 22 thành viên thể hiện tính base [8]. Tương tự, nhóm protein SWEET ở loài *G. raimondii* có kích thước trong khoảng 230 - 311 amino acid, với 25 trên tổng số 31 thành viên có giá trị điểm đẳng điện > 7 [8]. Ở hai loài *G. hirsutum* và *G. barbadense*, kích thước của nhóm protein SWEET lần lượt đạt từ 116 - 374 amino acid (48 trên tổng số 55 thành viên có tính base) và 88 - 331 amino acid (52 trên tổng số 60 thành viên có tính base) [8].

3.3. Phân tích cấu trúc gen và phân nhóm theo chức năng của protein vận chuyển đường sucrose ở cây *Ceratopteris richardii*

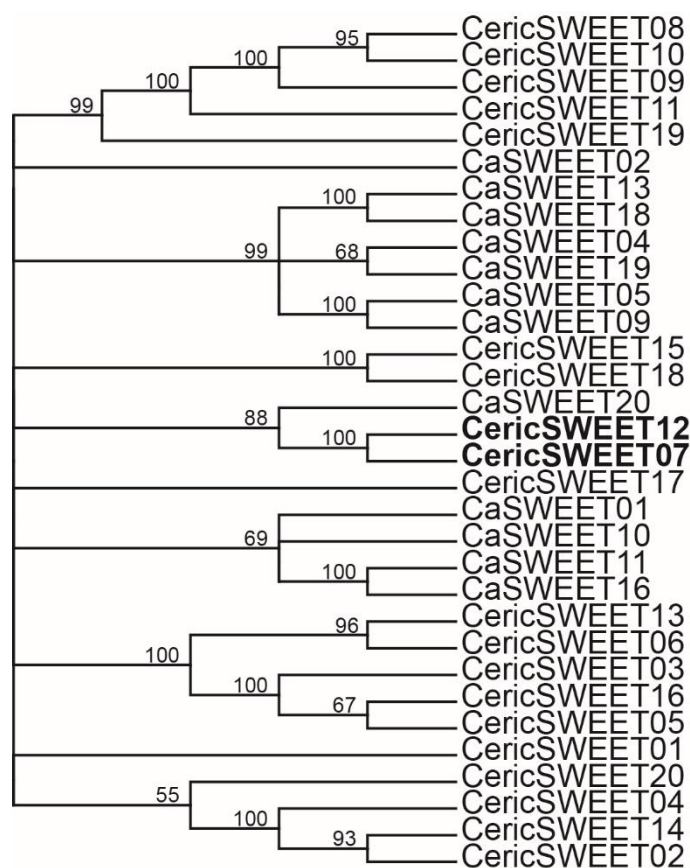
Trong nghiên cứu này, mức độ tương đồng của nhóm protein SWEET ở *C. richardii* được xác định bằng cách thiết lập sơ đồ hình cây Neighbor-Joining [14]. Kết quả cho thấy, nhóm CericSWEET có thể được chia thành 3 phân nhóm (Hình 2A). Cụ thể, phân nhóm 1 có 5 thành viên, tương ứng 25%, bao gồm CericSWEET03, 05, 06, 13 và 16; trong khi chỉ có 2 thành viên (chiếm 10%) thuộc phân nhóm 2 (CericSWEET07 và 12) (Hình 2A). Phân nhóm 3 chứa nhiều thành viên của nhóm SWEET nhất, 13 thành viên (tương ứng 65%) (Hình 2A). Trước đây, sự phân chia của nhóm protein SWEET thành 3 phân nhóm cũng được ghi nhận ở các loài thực vật khác như *O. sativa* [5], *S. lycopersicum* [6], *G. max* [7], *Gossypium* spp. [8], *T. aestivum* [9], *C. arietinum* [11]. Những kết quả này đã chỉ ra rằng, nhóm protein SWEET ở thực vật nhìn chung được chia thành ba phân nhóm.



Hình 2. Phân nhóm (A) và cấu trúc gen (B) của nhóm SWEET ở cây *Ceratopteris richardii*

Phân tích cấu trúc gen cho thấy nhóm gen mã hóa protein SWEET ở cây *C. richardii* có cấu trúc tương đối đồng nhất. Cụ thể, phần lớn gen (14 trên tổng số 20) có 6 exon, trong khi chỉ có 2 gen, lần lượt là CericSWEET02 và 20 có 5 exon. Bên cạnh đó, gen CericSWEET11 chỉ chứa 4 exon, trong khi 2 gen, CericSWEET03 và 05 chỉ có 1 exon (không có intron). Cuối cùng, CericSWEET18 được xác định là gen có chứa nhiều exon nhất, tương ứng 9 exon (Hình 2B). Trong nghiên cứu trước đây, cấu trúc phân mảnh ở nhóm gen mã hóa protein SWEET cũng đã được ghi nhận ở các loài thực vật khác [4]. Ví dụ, 16 trên tổng số 17 gen mã hóa protein SWEET ở loài *Arabidopsis thaliana* có cấu trúc phân mảnh (chứa ít nhất 1 intron), trong đó cấu trúc của 15 gen đều chứa 6 exon. Ở các loài bông, hầu hết các gen mã hóa protein SWEET đều chứa 3 - 7 exon, ngoại trừ 2 gen ở loài *G. barbadense*, bao gồm GbSWEET10a-A và 10a-D chứa 9 exon [8]. Tương tự, ở đậu tương, 31 trên tổng số 52 gen mã hóa protein SWEET đã được báo cáo có chứa 6 exon [7].

Để dự đoán vai trò cũng như định hướng cho nghiên cứu chức năng gen, sơ đồ hình cây dựa trên thuật toán Neighbor-Joining [14] đã được xây dựng giữa nhóm protein SWEET ở *C. richardii* với các protein SWEET ở *C. arietinum* đã biết chức năng [11] (Hình 3). Cụ thể, 12 protein SWEET mã hóa bởi các gen có đáp ứng với điều kiện hạn ở *C. arietinum* mô tả trong nghiên cứu trước đây [11] đã được khai thác. Với nguyên lý các protein có mức độ tương đồng cao về cấu trúc được sắp xếp cùng nhánh có thể chia sẻ chức năng tương tự nhau [17], nghiên cứu này đã chỉ ra được 2 protein SWEET ở *C. richardii*, bao gồm CericSWEET07 và 12 nằm cùng phân nhánh với CaSWEET20 ở *C. arietinum* (Hình 3). Hơn nữa, đặc tính lý hóa và cấu trúc gen của CericSWEET07 và 12 ở *C. richardii* đều tương tự với CaSWEET20 ở *C. arietinum* [11]. Kết quả này đã gợi ý rằng, 2 gen mã hóa cho protein CericSWEET07 và 12 ở *C. richardii* có thể có vai trò trong đáp ứng hạn, tương tự như gen CaSWEET20 ở *C. arietinum* [11].



Hình 3. Phân nhóm theo chức năng của nhóm protein SWEET ở cây *Ceratopteris richardii* với các protein SWEET đã biết chức năng ở cây *Cicer arietinum*

Bảng 3. Phân tích vị trí cư trú nội bào của nhóm SWEET ở cây *Ceratopteris richardii*

STT	Tên gen	MBS	DRE	MYCR	T/G Box	EE	NACR	CE3	ICEr2	TC-rich repeats	ABRE
1	CericSWEET01			1		1					2
2	CericSWEET02	1						1			1
3	CericSWEET03			1							
4	CericSWEET04			1			1				
5	CericSWEET05	2									1
6	CericSWEET06	1		1							
7	CericSWEET07			2		1					
8	CericSWEET08	1		1	1	1	2				3
9	CericSWEET09	1			1	1	1				1
10	CericSWEET10					1	1				1
11	CericSWEET11			1	1		2				1
12	CericSWEET12	2		1	1	1					1
13	CericSWEET13			1		1					
14	CericSWEET14	2				1	1				
15	CericSWEET15	3		1		1	2				3
16	CericSWEET16			1		1					2
17	CericSWEET17						1				
18	CericSWEET18	1					1				
19	CericSWEET19			1							1
20	CericSWEET20			2		1					1

Để củng cố giả thuyết về chức năng của các gen mã hóa protein SWEET ở cây *C. richardii*, các yếu tố điều hòa *cis*- đáp ứng hạn và cảm ứng với ABA đã được phân tích trên vùng promoter dựa theo nghiên cứu trước đây [11]. Kết quả minh họa ở Bảng 3 cho thấy, vùng promoter của 20 gen *CericSWEET* đều chứa ít nhất một nhóm yếu tố điều hòa *cis*- đáp ứng hạn. Số lượng yếu tố đáp ứng hạn ở vùng promoter dao động từ 1 (*CericSWEET03*, 17 và 19) đến 7 (*CericSWEET15*), trung bình đạt 2,85 yếu tố/gen (Bảng 3). Trong đó, 4 yếu tố đáp ứng hạn, bao gồm MBS, MYCR, EE và NACR xuất hiện phổ biến ở vùng promoter của các gen *CericSWEET* (Bảng 3). Đáng chú ý, vùng promoter của 12 gen được xác định có sự phân bố của ABRE, chứng tỏ các gen này liên quan đến con đường tín hiệu phụ thuộc ABA.

4. Kết luận

Nghiên cứu này đã xác định được 20 protein *CericSWEET* ở *C. richardii*. Trong đó, 6 gen nằm ở NST số 11, 3 gen nằm ở NST số 4 và 2 gen nằm ở NST số 23, các gen còn lại phân bố rải rác ở hệ gen. Dự đoán vị trí cho thấy, 18 protein *CericSWEET* cư trú trên màng sinh chất, trong khi *CericSWEET09* và 10 nằm tại lục lạp và peroxisome. Phân tích tính chất lý hóa cho thấy kích thước của protein *CericSWEET* đạt 202 - 428 amino acid, 19 protein *CericSWEET* có tính base. Tất cả các protein *CericSWEET* có tính kỵ nước. Phân tích cấu trúc gen cho thấy 1 gen chứa 9 exon, 14 gen có 6 exon, 2 gen chứa 5 exon, 1 gen chứa 4 exon và 2 gen chứa 1 exon. Nhóm protein *CericSWEET* được chia làm 3 phân nhóm chính, tương tự như ở các loài thực vật khác. Dựa trên mức độ tương đồng với protein *CaSWEET* ở *C. arietinum*, nghiên cứu đã đề xuất nghiên cứu chức năng của *CericSWEET07* và 12 liên quan đến cơ chế đáp ứng hạn ở *C. richardii*. Vùng promoter của các gen *CericSWEET* có ít nhất một yếu tố điều hòa *cis*- đáp ứng hạn.

Lời cảm ơn

Công trình này được hỗ trợ bởi Trường Đại học Công Nghệ thông qua Đề tài khoa học và công nghệ có mã số CN21.25.

TÀI LIỆU THAM KHẢO/ REFERENCES

- [1] F. Aram, E. H. García, E. Solgi, and S. Mansournia, "Urban green space cooling effect in cities," *Heliyon*, vol. 5, no. 4, pp. e01339, 2019.
- [2] O. Leroux, S. Eeckhout, R. Viane, and Z. Popper, "*Ceratopteris richardii* (C-fern): A model for investigating adaptive modification of vascular plant cell walls," *Front Plant Sci*, vol. 4, p. 367, 2013.
- [3] D. Marchant, E. Sessa, P. Wolf, K. Heo, W. Barbazuk, P. Soltis, and D. Soltis, "The C-Fern (*Ceratopteris richardii*) genome: insights into plant genome evolution with the first partial homosporous fern genome assembly," *Sci Reps*, vol. 9, no. 1, p. 18181, 2019.
- [4] J. Ji, L. Yang, Z. Fang, Y. Zhang, M. Zhuang, H. Lv, and Y. Wang, "Plant SWEET family of sugar transporters: Structure, evolution and biological functions," *Biomolecules*, vol. 12, no. 2, p. 205, 2022.
- [5] M. Yuan, and S. Wang, "Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms," *Mol Plant*, vol. 6, no. 3, pp. 665-674, 2013.
- [6] C. Feng, J. Han, X. Han, and J. Jiang, "Genome-wide identification, phylogeny, and expression analysis of the SWEET gene family in tomato," *Gene*, vol. 573, no. 2, pp. 261-272, 2015.
- [7] G. Patil, B. Valliyodan, R. Deshmukh, S. Prince, B. Nicander, M. Zhao, H. Sonah, L. Song, L. Lin, J. Chaudhary, Y. Liu, T. Joshi, D. Xu, and H. Nguyen, "Soybean (*Glycine max*) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis," *BMC Genomics*, vol. 16, p. 520, 2015.
- [8] L. Zhao, J. Yao, W. Chen, Y. Li, Y. Lü, Y. Guo, J. Wang, L. Yuan, Z. Liu, and Y. Zhang, "A genome-wide analysis of SWEET gene family in cotton and their expressions under different stresses," *J Cotton Res*, vol. 1, p. 7, 2018.
- [9] T. Gautam, G. Saripalli, V. Gahlaut, A. Kumar, P. Sharma, H. Balyan, and P. Gupta, "Further studies on sugar transporter (SWEET) genes in wheat (*Triticum aestivum* L.)," *Mol Biol Rep*, vol. 46, no. 2, pp. 2327-2353, 2019.

-
- [10] D. Goodstein, S. Shu, R. Howson, R. Neupane, R. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, and D. Rokhsar, "Phytozome: a comparative platform for green plant genomics," *Nucleic Acids Res*, vol. 40, pp. D1178 - D1186, 2012.
- [11] V. H. La, D. H. Chu, D. C. Tran, H. K. Nguyen, N. Q. Le., M. C. Hoang, P. B. Cao, A. T. Pham, D. B. Nguyen, Q. T. Nguyen, V. L. Nguyen, V. C. Ha, T. H. Le, H. H. Le, D. T. Le, and T. P. Lam-Son, "Insights into the gene and protein structures of the CaSWEET family members in chickpea (*Cicer arietinum*), and their gene expression patterns in different organs under various stress and abscisic acid treatments," *Gene*, vol. 819, p. 146210, 2022.
- [12] S. Briesemeister, R. Jörg, and O. Kohlbacher, "YLoc - An interpretable web server for predicting subcellular localization," *Nucleic Acids Res*, vol. 38, pp. W497-W502, 2010.
- [13] E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. Appel, and A. Bairoch, "ExPASy: the proteomics server for in-depth protein knowledge and analysis," *Nucleic Acids Res*, vol. 31, no. 13, pp. 3784-3788, 2003.
- [14] S. Kumar, G. Stecher, and K. Tamura, "MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets," *Mol Biol Evol*, vol. 33, no. 7, pp. 1870-1874, 2016.
- [15] B. Hu, J. Jin, A. Guo, H. Zhang, J. Luo, and G. Gao, "GSDS 2.0: an upgraded gene feature visualization server," *Bioinformatics*, vol. 31, no. 8, pp. 1296-1297, 2015.
- [16] S. Jiang, B. Balan, R. Assis, C. Sagawa, X. Wan, S. Han, L. Wang, L. Zhang, P. Zaini, S. Walawage, A. Jacobson, S. Lee, L. Moreira, C. Leslie, and A. Dandekar, "Genome-wide profiling and phylogenetic analysis of the SWEET sugar transporter gene family in walnut and their lack of responsiveness to *Xanthomonas arboricola* pv. *juglandis* infection," *Int J Mol Sci*, vol. 21, no. 4, p. 1251, 2020.
- [17] H. D. Chu, K. H. Nguyen, Y. Watanabe, D. T. Le, T. L. T. Pham, K. Mochida, and L. -S. P. Tran, "Identification, structural characterization and gene expression analysis of members of the Nuclear factor-Y family in chickpea (*Cicer arietinum* L.) under dehydration and abscisic acid treatments," *Int J Mol Sci*, vol. 19, p. 3290, 2018.