HNUE JOURNAL OF SCIENCE DOI: 10.18173/2354-1059.2021-0082
Natural Sciences 2021, Volume 66, Issue 4F, pp. 188-195
This paper is available online at http://stdb.hnue.edu.vn

XÁC ĐỊNH MỘT SỐ ĐẶC ĐIỂM CẤU TRÚC, VỊ TRÍ VÀ PHÂN NHÓM CỦA CÁC PROTEIN VẬN CHUYỂN ĐƯỜNG SUCROSE HỌ SWEET Ở CÂY DIÊM MẠCH (*Chenopodium quinoa*) BẰNG CÁC CÔNG CỤ TIN SINH HỌC

Phạm Phương Thu[1], Chu Đức Hà[2], Nguyễn Sông Thao[3] và Trần Thị Thanh Huyền[3]
[1]*Khoa Sinh Kĩ thuật Nông nghiệp, Trường Đại học Sư phạm Hà Nội 2*
[2]*Khoa Công nghệ Nông nghiệp, Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội*
[3]*Khoa Sinh học, Trường Đại học Sư phạm Hà Nội*

Tóm tắt. Trong nghiên cứu này, 29 thành viên của nhóm protein CqSWEET trên cây diêm mạch (*Chenopodium quinoa*) đã được phân tích bằng các công cụ tin sinh học một cách toàn diện. Theo đó, các protein CqSWEET có kích thước đạt từ 161 - 428 axít amin (17,7 - 47,7 kDa). Giá trị điểm đẳng điện của các protein CqSWEET dao động từ ngưỡng axít (4,9) đến bazơ (9,7), trong khi đa số thành viên của nhóm CqSWEET có tính ổn định trong ống nghiệm. Phân tích cho thấy nhóm CqSWEET có tính kỵ nước và rất nhiều protein có thể cư trú ở hệ thống bao gói. Dựa theo sơ đồ hình cây, các protein CqSWEET được phân loại thành 3 nhóm, tương tự như ở các loài thực vật khác. Phân tích cấu trúc cho thấy phần lớn các protein CqSWEET đều mang 6 hoặc 7 cấu trúc xoắn xuyên màng. Nghiên cứu đã ghi nhận 4 motif bảo thủ xuất hiện ở hầu hết các thành viên của nhóm CqSWEET. Kết quả của nghiên cứu này đã cung cấp những dẫn liệu quan trọng cho việc phân tích chức năng của gen *CqSWEET* ứng viên liên quan đến điều hòa sinh trưởng và phát triển của cây diêm mạch.

Từ khóa: *Chenopodium quinoa*, diêm mạch, SWEET, tính chất lí hóa, tin sinh học.

1. Mở đầu

Diêm mạch (*Chenopodium quinoa*) được biết đến là một trong những loài thân thảo quan trọng có giá trị kinh tế cao và được trồng trọt phổ biến trên thế giới. Có nguồn gốc từ khu vực Nam Mỹ, diêm mạch bắt đầu được nhập nội và thuần hóa tại Việt Nam nhằm tận dụng đặc tính thích ứng cao, không kén đất, chống chịu hạn và mặn tốt. Với thành tựu giải mã toàn bộ hệ gen của diêm mạch được công bố gần đây [1] đã cho phép đi sâu tìm hiểu về cơ chế chống chịu các điều kiện bất lợi ở cấp độ phân tử.

Nghiên cứu trước đây đã ghi nhận về chức năng của các protein vận chuyển đường sucrose, bao gồm nhóm SUT (viết tắt cho protein Sucrose Transporter) và SWEET (viết tắt cho protein Sugar Will Eventually be Exported Transporter) liên quan đến cơ chế đáp ứng bất lợi ở thực vật [2]. Trong đó, nhóm protein SWEET, đặc trưng bởi vùng domain MtN3/saliva (PF03083), đóng vai trò trong các quá trình sinh học quan trọng trong suốt giai đoạn sinh trưởng, phát triển và năng suất

Ngày nhận bài: 7/9/2021. Ngày sửa bài: 18/10/2021. Ngày nhận đăng: 26/10/2021.
Tác giả liên hệ: Trần Thị Thanh Huyền. Địa chỉ e-mail: tranthanhuyen@hnue.edu.vn

của cây trồng [2]. Trong nghiên cứu trước đây, nhóm SWEET đã được ghi nhận trên cây diêm mạch [3], như đậu gà (*Cicer arietinum*) [4], sắn (*Manihot esculenta*) [5] và lạc (*Arachis hypogaea*) [6]. Theo đó, nghiên cứu đã xác định được tổng số 29 thành viên trong nhóm CqSWEET trên hệ gen tham chiếu của diêm mạch. Cấu trúc gen và mức độ biểu hiện của các gen bước đầu đã được phân tích. Tuy nhiên, các đặc tính cơ bản của protein CqSWEET vẫn chưa được ghi nhận cụ thể.

Nghiên cứu này được thực hiện nhằm phân tích những tính chất cơ bản của nhóm CqSWEET ở cây diêm mạch bằng các công cụ tin sinh học. Bên cạnh đó, các thành viên trong nhóm CqSWEET cũng đã được sắp xếp theo nhóm dựa trên mức độ tương đồng trong trình tự protein. Cuối cùng, cấu trúc đặc trưng của protein CqSWEET đã được tìm hiểu nhằm làm rõ về mức độ bảo thủ của nhóm SWEET ở các loài thực vật.

2. Nội dung nghiên cứu

2.1. Dữ liệu nghiên cứu

Hệ gen, hệ protein, hệ phiên mã tham chiếu của giống diêm mạch *C. quinoa* Willd. (accession PI614886) [1] được khai thác trên Phytozome [7].

Dữ liệu về nhóm CqSWEET ở diêm mạch [3], CaSWEET ở đậu gà [4], MeSWEET ở sắn [5], AhSWEET ở lạc [6], bao gồm trình tự protein đầy đủ và mã định danh được khai thác cho phân tích *in silico*.

2.2. Phương pháp nghiên cứu

Phương pháp phân tích tính chất lí hóa của protein SWEET: Trình tự protein đầy đủ của các phân tử CqSWEET được phân tích trên Expasy Protparam (http://web.expasy.org/protparam/) [8] theo mô tả trong nghiên cứu trước đây. Cụ thể, tính chất của protein, bao gồm kích thước (aa), trọng lượng phân tử (kDa), điểm đẳng điện, độ bất ổn định (< 40 và >40 thể hiện tính ổn định và không ổn định trong điều kiện ống nghiệm) [9], chỉ số béo, chỉ số GRAVY (<0 và >0 thể hiện tính ưa nước và kỵ nước) [10].

Phương pháp dự đoán vị trí cư trú của protein SWEET: Trình tự protein đầy đủ (.fasta) của các phân tử CqSWEET được phân tích trên TargetP v2.0 (http://www.cbs.dtu.dk/services/TargetP/) [11] theo mô tả trong nghiên cứu trước đây. Theo đó, các protein đặc trưng cho một số bào quan, bao gồm lục lạp (cTP), ty thể (mTP) và hệ thống bao gói (SP) được xác định trên đầu N-terminal [12].

Phương pháp xây dựng sơ đồ hình cây của protein SWEET: Cây phân loại Neighbor-Joining được sử dụng để thiết lập mối quan hệ giữa trình tự đầy đủ của các protein SWEET mục tiêu trên MEGA v7.0 [13] với giá trị bootstrap đạt 1000 lần nhắc lại theo mô tả trong nghiên cứu trước đây [4-6]. Cụ thể, các thông số được cài đặt như sau, giá trị gap opening penalty 10, gap extension penalty 0,2. Các sơ đồ hình cây (.netwick) được minh hoạ bằng Adobe Illustrator.

Phương pháp phân tích cấu trúc xuyên màng của protein SWEET: Trình tự protein đầy đủ (.fasta) của nhóm CqSWEET được phân tích trên TMHMM v2.0 (http://www.cbs.dtu.dk/services/TMHMM-2.0) để xác định các đoạn xuyên màng (độ tin cậy > 50%) [14].

Phương pháp phân tích motif đặc trưng của protein SWEET: Trình tự protein đầy đủ (.fasta) của nhóm CqSWEET được truy vấn trên MEME (https://meme-suite.org/meme/) [15] để xác định các đoạn đặc trưng. Theo đó, tổng số 10 motif, có kích thước từ 6 - 50 aa, được rà soát trên toàn bộ chuỗi truy vấn [15]. Kết quả được phân nhóm theo sơ đồ hình cây Neighbor-Joining [13].

2.3. Kết quả và thảo luận

2.3.1. Phân tích tính chất lí hóa và vị trí cư trú nội bào của nhóm protein CqSWEET ở cây diêm mạch

Để tìm hiểu về các tính chất lí hóa của nhóm CqSWEET ở cây diêm mạch, trình tự protein đầy đủ của từng thành viên đã được khảo sát trên Expasy Protparam [8]. Kết quả cho thấy, kích thước và trọng lượng phân tử của nhóm CqSWEET dao động từ 161 aa - 17,7 kDa (CqSWEET09) đến 428 aa - 47,7 kDa (CqSWEET18). Trong đó, đa số các protein ($^{22}/_{29}$) có giá trị điểm đẳng điện ở khoảng base (>7), ngoại trừ 6 protein ở khoảng acid (<7) và 1 protein (CqSWEET28) ở khoảng trung tính (~7). Dựa trên giá trị bất ổn định [9], $^{18}/_{29}$ protein được dự đoán có cấu trúc ổn định trong điều kiện ống nghiệm, trong khi $^{11}/_{29}$ protein không có tính ổn định. Ngoài ra, chỉ số béo của các protein dao động từ 93,4 (CqSWEET03) đến 131,9 (CqSWEET26), trung bình đạt 114,3. Đáng chú ý, tất cả thành viên trong nhóm CqSWEET đều có chỉ số GRAVY đạt giá trị dương, chứng tỏ các phân tử này có tính kỵ nước (Bảng 1).

Bảng 1. Thông tin về tính chất lí hóa và vị trí cư trú của họ protein CqSWEET ở cây diêm mạch

Stt	CqSWEET	L (aa)	mW (kDa)	pI	II	AI	GRAVY	Vị trí cư trú
1	CqSWEET01	246	27,6	9,1	37,5	119,3	0,6	-
2	CqSWEET02	242	26,9	9,3	32,2	115,5	0,7	-
3	CqSWEET03	400	45,0	6,3	40,84	93,4	0,3	-
4	CqSWEET04	250	27,7	9,3	26,56	115,8	0,5	-
5	CqSWEET05	319	35,9	8,1	30,7	108,5	0,4	S
6	CqSWEET06	242	27,2	9,2	42,2	117,9	0,7	S
7	CqSWEET07	205	22,8	4,9	40,6	98,3	0,2	S
8	CqSWEET08	236	27,2	9,5	40,8	124,2	0,8	-
9	CqSWEET09	161	17,7	9,0	49,5	117,5	0,9	S
10	CqSWEET10	245	27,5	9,3	48,0	120,9	0,6	S
11	CqSWEET11	205	22,7	6,2	32,9	126,4	0,8	-
12	CqSWEET12	253	28,6	9,3	33,9	115,1	0,6	S
13	CqSWEET13	332	36,8	9,2	28,0	100,5	0,3	S
14	CqSWEET14	246	27,5	9,1	35,2	121,7	0,6	M
15	CqSWEET15	215	24,2	9,4	33,7	107,4	0,6	S
16	CqSWEET16	188	21,6	6,0	32,1	114,5	0,7	-
17	CqSWEET17	185	20,4	9,2	33,8	115,9	0,7	S
18	CqSWEET18	428	47,7	9,0	40,9	105,9	0,3	S
19	CqSWEET19	266	29,6	6,6	52,4	112,1	0,4	S
20	CqSWEET20	260	28,6	9,7	29,2	113,6	0,6	S
21	CqSWEET21	250	27,7	9,5	31,6	115,0	0,4	-
22	CqSWEET22	247	28,2	8,0	43,7	103,4	0,4	-
23	CqSWEET23	287	31,8	8,1	51,0	115,8	0,4	S
24	CqSWEET24	245	27,3	9,2	46,9	120,1	0,6	S
25	CqSWEET25	253	28,6	9,2	36,7	120,2	0,7	S
26	CqSWEET26	226	24,8	5,1	32,1	131,9	0,9	S
27	CqSWEET27	318	35,6	9,6	36,0	112,5	0,7	S
28	CqSWEET28	309	34,5	6,9	35,1	116,3	0,5	S
29	CqSWEET29	277	31,6	9,4	37,8	115,0	0,6	-

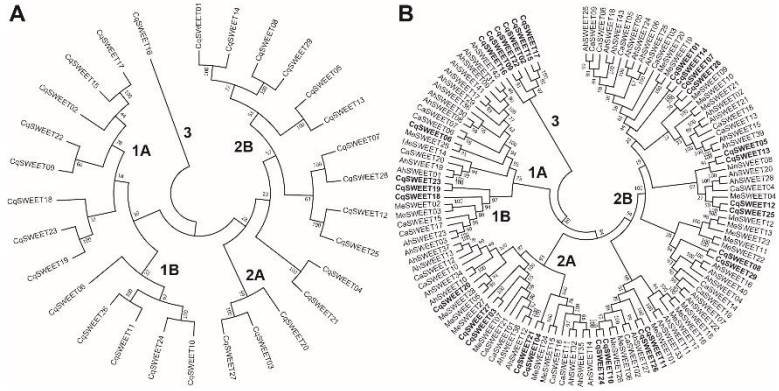
Ghi chú: L - Kích thước phân tử (aa), mW - Trọng lượng phân tử (kDa), pI - Điểm đẳng điện, II- Độ bất ổn định, AI - Chỉ số béo, S - Hệ thống bao gói, M - Ty thể.

Trong các nghiên cứu trước đây, nhóm protein SWEET cũng thể hiện mức độ đa dạng trong tính chất lí hóa, tương tự như kết quả trên nhóm CqSWEET ở cây diêm mạch. Cụ thể, nhóm protein CaSWEET ở đậu gà có kích thước dao động từ 230 - 296 aa, tương ứng 25,67 - 33,47 kDa [4]. Tương tự, đặc tính lí hóa nhóm protein MeSWEET ở cây sắn cũng được ghi nhận với kích thước đạt 234 (25,85 kDa) - 302 aa (33,2 kDa) [5]. Gần đây, nhóm protein AhSWEET trên cây lạc cũng được báo cáo có kích thước phân tử đạt 104 - 320 aa (trung bình đạt 254,3 aa), trọng lượng phân tử đạt 11,33 - 35,26 kDa (trung bình đạt 35,92 kDa) [6]. Đa số các protein SWEET trên một số đối tượng cây trồng khác đều có giá trị điểm đẳng điện ở ngưỡng bazơ [4-6], tương tự như nghiên cứu này. Tương tự, phần lớn các protein SWEET ở các loài thực vật, ví dụ như $^{32}/_{43}$ AhSWEET ở lạc và $^{15}/_{21}$ CaSWEET ở đậu gà, đều có tính ổn định với giá trị bất ổn định <40 [4, 6]. Các kết quả trước đây cũng cho thấy đa số các protein SWEET đều có tính kỵ nước [4-6].

Tiếp theo, nghiên cứu đã chỉ ra rằng $^{18}/_{29}$ phân tử CqSWEET cư trú trên hệ thống bao gói (với sự có mặt của đoạn SP) trong CqSWEET14 phân bố ở ty thể (với sự có mặt của đoạn mTP). Kết quả này cũng đồng thuận với những nghiên cứu trước đây về vị trí của protein SWEET liên quan đến hoạt động vận chuyển sucrose qua màng [2].

2.3.2. Xây dựng sơ đồ hình cây của nhóm protein CqSWEET ở cây diêm mạch

Để tìm hiểu về mối tương quan giữa các thành viên trong nhóm CqSWEET ở cây diêm mạch, sơ đồ hình cây đã được thiết lập dựa trên thuật toán Neighbor-Joining bằng MEGA [13]. Kết quả cho thấy, nhóm CqSWEET ở diêm mạch có thể được chia làm 3 nhóm riêng biệt. Trong đó, nhóm 1 gồm 2 phân nhóm phụ, 1A (chứa 8 thành viên, CqSWEET02/09/15/17/18/19/22 và 23) và 1B (chứa 5 thành viên, CqSWEET06/10/11/24 và 26), nhóm 2 cũng bao gồm 2 phân nhóm phụ, 2A (chứa 3 thành viên, CqSWEET03/20 và 27) và 2B (chứa 12 thành viên, CqSWEET01/04/05/07/08/12/13/14/21/25/28 và 29) và nhóm 3 chỉ chứa 1 thành viên (CqSWEET16) (Hình 1A). Dựa theo phân tích tính chất lí hóa (Bảng 1) và sơ đồ hình cây (Hình 1A), các protein CqSWEET xếp trên cùng nhánh của phân nhóm đều có các đặc tính tương tự nhau. Ví dụ, CqSWEET15 và 17, CqSWEET09 và 22, CqSWEET19 và 23 nằm cùng nhánh trong phân nhóm 1A (Hình 1A) và có đặc tính về giá trị pI, II, AI và GRAVY tương tự nhau (Bảng 1).

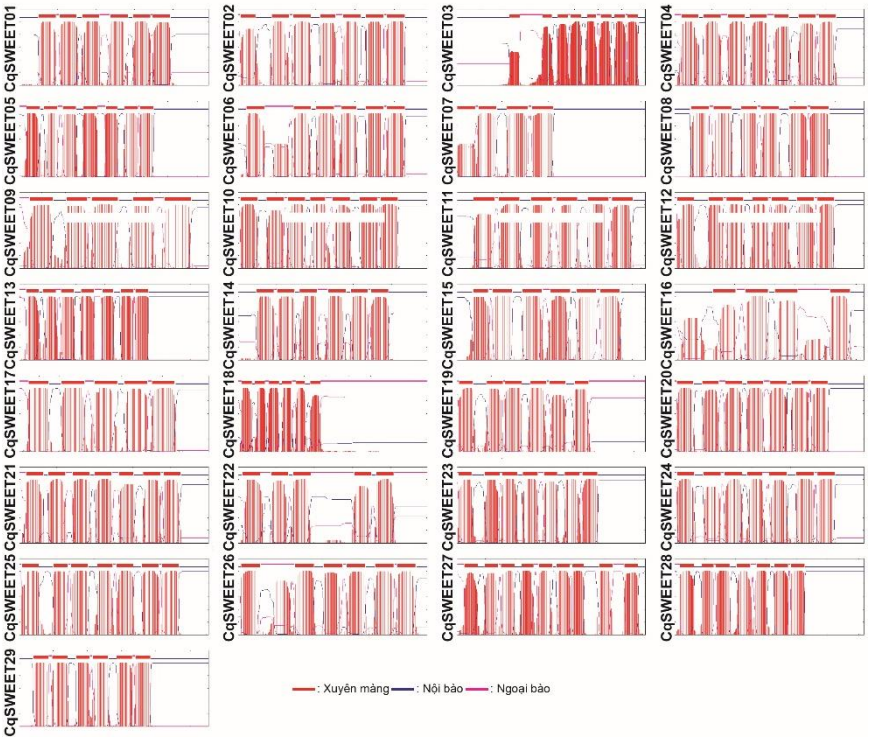


Hình 1. Sơ đồ hình cây của nhóm CqSWEET (A) và nhóm SWEET ở 4 loài thực vật (B)

Bên cạnh đó, để xác định mối tương quan giữa các thành viên của họ SWEET ở các loài thực vật khác nhau, một sơ đồ hình cây giữa nhóm CaSWEET ở đậu gà [4], MeSWEET ở sắn [5], AhSWEET ở lạc [6] và CqSWEET ở diêm mạch đã được xây dựng. Kết quả cho thấy, các thành viên của 4 họ SWEET lần lượt được sắp xếp theo 3 nhóm, 1 (gồm 2 phân nhóm 1A và 1B), 2 (gồm 2 phân nhóm 2A và 2B) và 3 (Hình 2B). Sự phân nhóm này cũng đồng thuận với kết quả trên họ CqSWEET ở diêm mạch (Hình 2A). Như vậy, nhìn chung, họ SWEET ở bốn loài thực vật này có thể được chia làm 3 nhóm riêng biệt với số lượng thành viên đa dạng.

2.3.3. Phân tích cấu trúc tương đồng của nhóm CqSWEET của cây diêm mạch

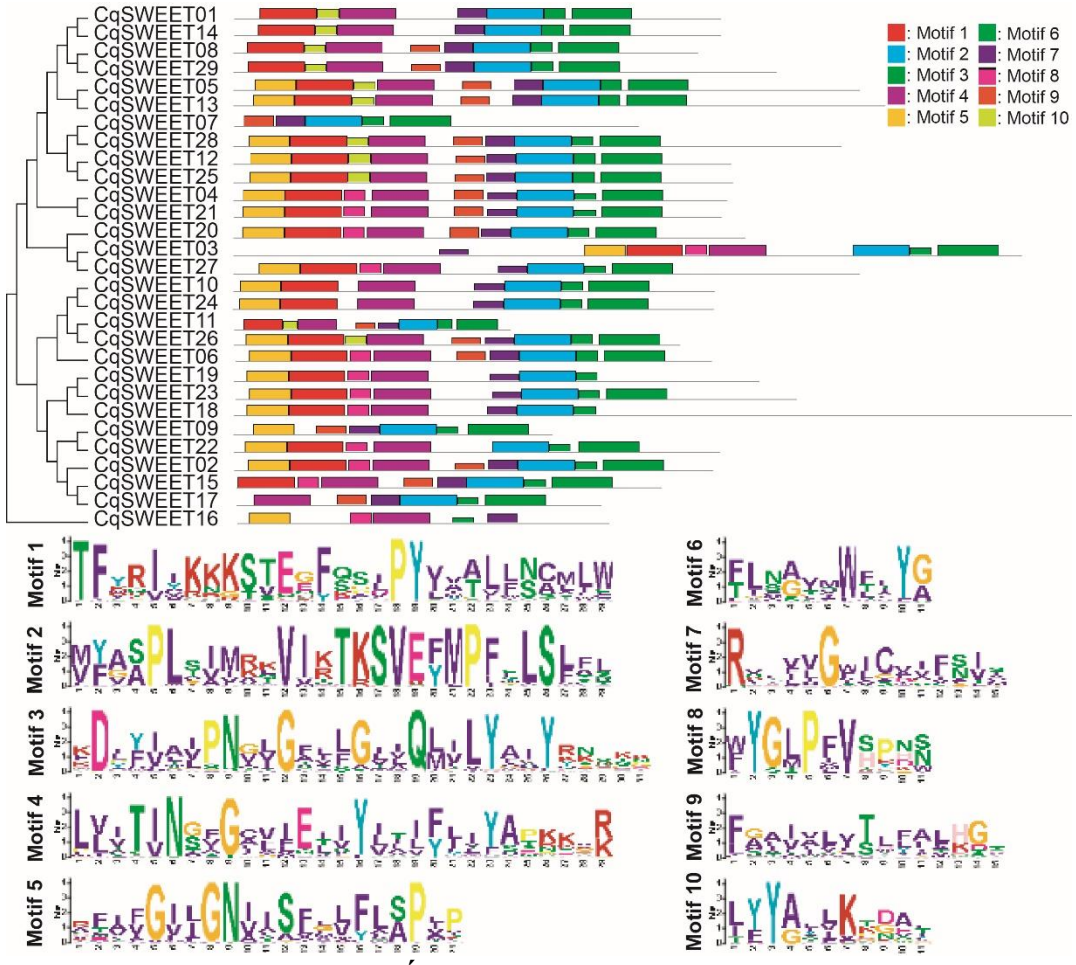
Để tìm hiểu cấu trúc xuyên màng của các phân tử CqSWEET ở cây diêm mạch, trình tự protein đầy đủ đã được truy vấn trên TMHMM [14]. Kết quả minh hoạt ở Hình 2 cho thấy số lượng cấu trúc xuyên màng tương đối bảo thủ giữa các thành viên trong nhóm CqSWEET. Cụ thể, phần lớn ($^{22}/_{29}$) protein cùng chia sẻ 6 (10 thành viên) và 7 (12 thành viên) cấu trúc xuyên màng. Kết quả chỉ ghi nhận 2 (CqSWEET07 và 16) và 3 (CqSWEET09/17 và 22) protein lần lượt có 4 và 5 cấu trúc xuyên màng, trong khi CqSWEET03 và 27 chứa 8 và 9 cấu trúc xuyên màng.



Hình 2. Cấu trúc xoắn xuyên màng trong nhóm protein CqSWEET ở cây diêm mạch

Thông thường, các protein SWEET ở thực vật được đặc trưng bởi 7 đoạn xoắn xuyên màng, chứa 2 vùng bảo thủ MtN3/saliva (gồm 3 cấu trúc xuyên màng) được xen kẽ bởi 1 đơn vị cấu trúc xuyên màng riêng rẽ [3, 16]. Ở sinh vật nhân sơ, cấu trúc của phân tử SWEET được xác định với chỉ 3 cấu trúc xoắn xuyên màng. Trong nghiên cứu trước đây, nhóm CaSWEET ở đậu gà cũng ghi nhận 7 cấu trúc xuyên màng tương tự như những báo cáo trước đây [3]. Sự có mặt của cấu trúc xuyên màng đảm bảo cho các phân tử SWEET có thể bám trên các màng bào quan để thực hiện chức năng vận chuyển sucrose nội bào [16, 17]. Có thể thấy rằng, sự đa dạng về cấu trúc xoắn xuyên màng ở họ CqSWEET ở cây diêm mạch có thể liên quan đến vai trò của các gen trong quá trình vận chuyển sucrose nội bào nhằm điều hòa áp suất thẩm thấu khi chịu tác động của bất lợi ngoại cảnh. Nghiên cứu tiếp theo có thể được tiếp tục nhằm kiểm chứng khả năng đáp ứng bất lợi của các gen mã hóa CqSWEET có liên quan đến cấu trúc xuyên màng.

Tiếp theo, để phân tích mức độ đặc trưng của nhóm CqSWEET, các motif bảo thủ đã được xác định bằng MEME [15]. Kết quả đánh giá sự phân bố của 10 motif đặc trưng trên 29 thành viên dựa theo sự phân nhóm của thuật toán Neighbor-Joining [13] được minh họa ở Hình 3. Theo đó, 5 motif (1 – 4 và 6) xuất hiện ở hầu hết các thành viên của nhóm CqSWEET, ngoại trừ CqSWEET07 (không có motif 1), 09 (không có motif 1 và 4), 17 (không có motif 1) và CqSWEET16 (không có motif 1 - 3). Motif 5 cũng xuất hiện ở đa số ($^{21}/_{29}$) thành viên của nhóm CqSWEET. Các motif còn lại (5 và 7 - 10) phân bố rải rác ở trong những phân nhóm trên sơ đồ.



Hình 3. Motif đặc trưng trong cấu trúc của nhóm protein CqSWEET ở cây diêm mạch

3. Kết luận

Nhóm CqSWEET ở cây diêm mạch có kích thước, trọng lượng phân tử, điểm đẳng điện, chỉ số béo đa dạng. Đa số protein CqSWEET có cấu trúc ổn định trong điều kiện ống nghiệm, trong khi tất cả các phân tử đều có tính kỵ nước. Phần lớn nhóm CqSWEET phân bố trên màng của các hệ thống bao gói nội bào.

Nhóm CqSWEET ở cây diêm mạch có thể được chia làm 3 nhóm tách biệt, với nhóm 1 và 2 đều gồm 2 phân nhóm phụ. Xây dựng sơ đồ hình cây giữa các nhóm SWEET ở 4 loài cũng ghi nhận sự phân nhóm tương tự.

Nhóm CqSWEET ở cây diêm mạch chủ yếu gồm 6 và 7 cấu trúc xuyên màng. Đa số các protein trong nhóm CqSWEET đều chứa ít nhất 4 motif đặc trưng, trong khi 6 motif còn lại phân bố rải rác trong cấu trúc của các thành viên.

Nghiên cứu này sẽ được tiếp tục nhằm đánh giá mức độ phiên mã của các gen ứng viên mã hóa CqSWEET có tiềm năng trong đáp ứng bất lợi ở cây diêm mạch.

***Lời cảm ơn*:** Nghiên cứu này được tài trợ bởi Trường Đại học Sư phạm Hà Nội 2 qua Đề tài có mã số HPU2.CS-2021.14.

TÀI LIỆU THAM KHẢO

[1] D.E. Jarvis, Y.S. Ho, D.J. Lightfoot, S.M. Schmöckel, B. Li, T.J. Borm, H. Ohyanagi, K. Mineta, C.T. Michell, N. Saber, N.M. Kharbatia, R.R. Rupper, A.R. Sharp, N. Dally, B.A. Boughton, Y.H. Woo, G. Gao, E.G. Schijlen, X. Guo, A.A. Momin, S. Negrão, S. Al-Babili, C. Gehring, U. Roessner, C. Jung, K. Murphy, S.T. Arold, T. Gojobori, C.G. Linden, E.N. Loo, E.N. Jellen, P.J. Maughan, M. Tester, 2017. The genome of *Chenopodium quinoa*. *Nature*, Vol. 542, No. 7641, pp.307-312.

[2] Y.H. Xuan, Y.B. Hu, L.Q. Chen, D. Sosso, D.C. Ducat, B.H. Hou, W.B. Frommer, 2013. Functional role of oligomerization for bacterial and plant SWEET sugar transporter family. *Proc Natl Acad Sci U S A*, Vol. 110, No. 39, pp.E3685-3694.

[3] Phạm Phương Thu, Trần Thị Phương Liên, La Việt Hồng, Tạ Hồng Lĩnh, Chu Đức Hà, Nguyễn Thị Ngọc Ánh, Bùi Thị Thu Hương, Nguyễn Văn Lộc, Lê Thị Ngọc Quỳnh, Nguyễn Sông Thao, Trần Thị Thanh Huyền, 2021. Nghiên cứu nhóm gen quy định protein vận chuyển đường sucrose ở cây diêm mạch (*Chenopodium quinoa*). *Tạp chí Khoa học Công nghệ Nông nghiệp Việt Nam*, Vol. 7, No. 128, pp. 1-5.

[4] Chu Đức Hà, Phùng Thị Vượng, Chu Thị Hồng, Phạm Thị Lí Thu, Phạm Phương Thu, Trần Thị Phương Liên, La Việt Hồng, 2018. Định danh và phân tích cấu trúc của họ gen mã hóa protein vận chuyển đường sucrose ở cây đậu gà (*Cicer arietinum*). *Tạp chí Khoa học và Công nghệ Đại học Thái Nguyên*, Vol. 194, No. 01, pp. 133-138.

[5] Chu Đức Hà, Phạm Thị Quỳnh, Phạm Thị Lí Thu, Nguyễn Văn Cương, Lê Tiến Dũng, 2018. Xác định họ gen mã hóa protein vận chuyển SWEET trên cây sắn (*Manihot esculenta* Crantz). *Tạp chí Khoa học Trường Đại học Sư phạm Hà Nội*, Vol. 63, No. 3, pp. 140-149.

[6] Chu Đức Hà, Hoàng Thị Hải Yến, Lê Thị Ngọc Quỳnh, Phạm Thị Ngọc Bích, Hoàng Minh Chính, La Việt Hồng, Phạm Phương Thu, Nguyễn Thị Minh Nguyệt, Lê Hùng Lĩnh, Lê Huy Hàm, Nguyễn Quốc Trung, Phạm Xuân Hội, 2020. Phân tích hiện tượng lặp ở họ gen mã hóa protein vận chuyển sucrose liên quan đến đáp ứng mặn ở lạc. *Hội nghị Công nghệ Sinh học Toàn quốc 2020*, tr. 80-85.

[7] D.M. Goodstein, S. Shu, R. Howson, R. Neupane, R.D. Hayes, J. Fazo, T. Mitros, W. Dirks, U. Hellsten, N. Putnam, D.S. Rokhsar, 2012. Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res*, Vol. 40, pp.D1178-D1186.

[8] E. Gasteiger, C. Hoogland, A. Gattiker, M. Wilkins, R. Appel, A. Bairoch. (2005). Protein identification and analysis tools on the ExPASy server. *In The proteomics protocols handbook*, pp. 571-607.

[9] K. Guruprasad, B.V. Reddy, M.W. Pandit, 1990. Correlation between stability of a protein and its dipeptide composition: A novel approach for predicting in vivo stability of a protein from its primary sequence. *Protein Eng*, Vol. 4, No. 2, pp.155-161.

[10] J. Kyte, R.F. Doolittle, 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol*, Vol. 157, No. 1, pp.105-132.

[11] O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*, Vol. 2, No. 4, pp.953-971.

[12] J.J. Almagro Armenteros, M. Salvatore, O. Emanuelsson, O. Winther, G. von Heijne, A. Elofsson, H. Nielsen, 2019. Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, Vol. 2, No. 5, pp. e201900429.

[13] S. Kumar, G. Stecher, K. Tamura, 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*, Vol. 33, No. 7, pp. 1870-1874.

[14] N. Chaturvedi, S. Shanker, V.K. Singh, D. Sinha, P.N. Pandey, 2011. Hidden markov model for the prediction of transmembrane proteins using MATLAB. *Bioinformation*, Vol. 7, No. 8, pp. 418-421.

[15] T.L. Bailey, J. Johnson, C.E. Grant, W.S. Noble, 2015. The MEME suite. *Nucleic Acids Res*, Vol. 43, No. W1, pp. W39-49.

[16] B.T. Julius, K.A. Leach, T.M. Tran, R.A. Mertz, D.M. Braun, 2017. Sugar transporters in plants: New insights and discoveries. *Plant Cell Physiol*, Vol. 58, No. 9, pp. 1442-1460.

[17] 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, H.T. Nguyen, 2015. Soybean (*Glycine max*) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis. *BMC Genomics*, Vol. 16, No. pp. 520.

ABSTRACT

Classification and analysis of conserved motifs in the sucrose transporter family in quinoa (*Chenopodium quinoa*) by bioinformatics approach

Pham Phuong Thu[1], Chu Duc Ha[2], Nguyen Song Thao[3] and Tran Thi Thanh Huyen[3]

[1]*Faculty of Biology - Agricultural Technology, Hanoi Pedagogical University 2*

[2]*Faculty of Agricultural Technology, University of Engineering Technology, VNU*

[3]*Faculty of Biology, Vietnam National University of Education*

In this study, a total of 29 members of the CqSWEET family in quinoa (*Chenopodium quinoa*) has been comprehensively investigated by various bioinformatics methods. As the result, we indicated that the CqSWEET proteins were varied from 161 - 428 amino acid residues in size (17,7 - 47,7 kDa in weight). The isoelectric point of these proteins ranged from acid (4.9) to base (9,7), while the majority of the CqSWEET proteins are stable in the test tube. Our analyses also revealed that whole members of the CqSWEET family are hydrophobic, and a large amount of the CqSWEET proteins may localize in the secretory pathways. According to the unrooted phylogenetic tree, the CqSWEET proteins could be clearly classified into 3 distinct groups, as similar as in other higher plant species. Interestingly, the majority of the CqSWEET proteins contained 6 or 7 transmembrane helices. Our results also showed that a total of 4 conserved motifs has occurred in the majority of the CqSWEET family. Take together, our study could provide critical data for further functional characterization of the *CaSWEET* candidate genes related to the regulation of the growth and development in quinoa plants.

***Keywords*:** *Chenopodium quinoa*, quinoa, SWEET, physic-chemical property, bioinformatics.