

PHÂN TÍCH ĐẶC ĐIỂM *IN SILICO* CÁC GENE MÃ HÓA PROTEIN SWEET Ở CÂY CA CAO (*Theobroma cacao* L.)

Cao Phi Bằng^{1,*}, Nguyễn Văn Đỉnh², Trần Thị Thanh Huyền³,
Lê Thị Mận¹, Vũ Xuân Dương¹,

Tóm tắt. SWEET (sugars will eventually be exported transporter) là một trong các nhóm protein vận chuyển đường quan trọng ở thực vật. Trong nghiên cứu này, 21 gene mã hóa protein SWEET trong hệ gene cây cao đã được xác định và phân tích. Các gene này có chiều dài từ 1379 đến 2700 gốc nucleotide, hầu hết gene có năm intron. Các protein suy diễn có từ 232 tới 306 gốc axit amin và có mang các xoắn xuyên màng đặc trưng đã biết cho các SWEET. Hầu hết các protein này có tính kiềm. Phân tích cây phả hệ cho thấy các gene *SWEET* của cây cao được phân chia thành bốn nhóm, nhóm I (4 gene), nhóm II (4 gene), nhóm III (6 gene) và nhóm IV (7 gene). Các gene này phân bố không đồng đều trong hệ gene của cây cao. Nhiễm sắc thể số 3 và số 5 mang 13 trong tổng số 21 gene SWEET. Một số hiện tượng nhân gene ở trên hai nhiễm sắc thể này có liên quan đến số lượng nhiều các gene *SWEET* ở nhóm III và nhóm IV. Mười hai trong tổng số 21 gene *SWEET* có các mã phiên EST được phát hiện ở một số loại mô khác nhau của cây cao, dưới các điều kiện khác nhau. Phần lớn trong số chúng có các EST được thu từ các mô sinh sản hoặc các mô chịu tác động của stress vô sinh và hữu sinh, gợi ý rằng các gene này giữ vai trò quan trọng trong sự phát triển cũng như tính chống chịu của cây cao.

Từ khóa: Biểu hiện gene, cao, cây di truyền đặc, điểm gene, SWEET.

1. MỞ ĐẦU

Ca cao (*Theobroma cacao* L.) là loài cây nhiệt đới thường xanh có nguồn gốc ở vùng đất thấp rừng mưa nhiệt đới Amazon, được con người thuần hóa trên 1.500 năm trước (Motamayor et al., 2002). Hiện nay, loài cây này đã được trồng ở trên 50 quốc gia trên thế giới, trong đó có Việt Nam. Hạt cao chủ yếu được sử dụng để làm socola, mứt và trong công nghiệp mỹ phẩm (Figueira et al., 2005). Tuy cao được coi như một đối tượng có nhiều hạn chế trong nghiên cứu (Figueira et al., 2005), nhưng hệ gene của loài này là nguồn tài nguyên tốt cho phép tăng cường các tiến bộ trong chọn giống cũng như trồng trọt và các đặc tính hóa sinh của nó (Motamayor et al., 2013).

SWEET là một trong các protein tham gia vào quá trình vận chuyển đường sucrose ở thực vật (Jeena et al., 2019). Các protein SWEET có cấu trúc gồm 7 vùng xoắn xuyên màng. Nhóm protein này có chức năng vận chuyển sucrose, liên quan đến sự phát triển hoa, quả và hạt, đồng thời có chức năng vận chuyển gibberellin cũng như sự phân bố

¹Trường Đại học Hùng Vương

²Trường Đại học Sư phạm Hà Nội 2

³Trường Đại học Sư phạm Hà Nội

*Email: phibang.cao@hvu.edu.vn

đường dưới các điều kiện stress thẩm thấu khác nhau, sự dinh dưỡng của sinh vật gây bệnh và liên quan đến sự điều hòa của các stress vô sinh (Jeena *et al.*, 2019). Với vai trò quan trọng như vậy, họ gene mã hóa các SWEET đã được nghiên cứu ở nhiều loài như *Arabidopsis thaliana* (Chen *et al.*, 2010) và lúa (Yuan & Wang, 2013), sắn (Chu Đức Hà và *nnk*, 2018) và nhiều loài thực vật khác (Li *et al.*, 2018). Tuy nhiên, nghiên cứu về họ gene SWEET ở cây ca cao đến nay chưa được thực hiện. Công trình này có mục tiêu xác định các gene mã hóa SWEET trong hệ gene của cây ca cao, phân tích các đặc điểm cấu trúc, sự phân bố của gene cũng như các đặc điểm hóa lí lí thuyết của các protein SWEET ở loài cây này và sự biểu hiện gene. Những kết quả nghiên cứu này bước đầu cung cấp các thông tin khoa học về cấu trúc, chức năng của các SWEET của loài cây công nghiệp quan trọng này.

2. NGUYÊN LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Cơ sở dữ liệu

Trình tự hệ gene của cây ca cao được lấy từ website phytozome (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Tcacao) (Argout *et al.*, 2011). Dữ liệu EST (expressed sequence tags) của cây ca cao (txid3641) được lấy từ cơ sở dữ liệu NCBI (<https://www.ncbi.nlm.nih.gov/nucleotide>).

Xác định các gene thuộc họ SWEET ở cây ca cao

Các protein SWEET của cây *Arabidopsis* (Chen *et al.*, 2010) được sử dụng làm khuôn dò để tìm kiếm các gene tương đồng trên dữ liệu hệ gene của cây ca cao nhờ chương trình TBLASTN, giúp tìm kiếm cả các gene chưa được chú giải tự động.

Xây dựng cây phả hệ

Trình tự protein suy diễn SWEET của cây ca cao, và một số cây có các SWEET đã được xác định trên quy mô hệ gene như nho (đại diện thân gỗ hai lá mầm), *A. thaliana* (hai lá mầm), lúa (một lá mầm) và của rêu (thực vật bậc cao) được sắp dãy bằng MAFFT (Kato & Standley, 2013), cây phả hệ được xây dựng nhờ phần mềm MEGA X (Kumar *et al.*, 2018).

Phân tích các đặc điểm hóa - lí

Các đặc điểm vật lí, hóa học của các gene cũng như protein được phân tích bằng các công cụ của ExPASy (Gasteiger *et al.*, 2005). Cấu trúc exon/intron được xây dựng nhờ GSDS 2.0 (Guo *et al.*, 2007).

Nghiên cứu sự biểu hiện gene

Sự biểu hiện của các gene được khảo sát nhờ phân tích hệ EST (expressed sequence tags) của cây ca cao có trên ngân hàng dữ liệu NCBI.

3. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

Xác định họ gene SWEET và đặc điểm các gene SWEET ở cây ca cao

Tổng số 21 gene mã hóa cho các SWEET ở cây ca cao đã được xác định (Bảng 1). Họ SWEET của cây ca cao lớn hơn so với của cây *A. thaliana* (17 gene) (Chen *et al.*,

2010), cây nho (16 gene) (Afoufa-Bastien et al., 2010), tương đương với cây lúa, nhưng nhỏ hơn so với cây sắn (28 gene) (Chu Đức Hà và nnk., 2018). Phân tích cấu trúc cho thấy các protein suy diễn của các gene này đều mang vùng bảo tồn đặc trưng (*MtN3_slv* (PF03083)) (Chen et al., 2010).

Bảng 1. Các gene thuộc họ *SWEET* của cây ca cao và đặc điểm của chúng

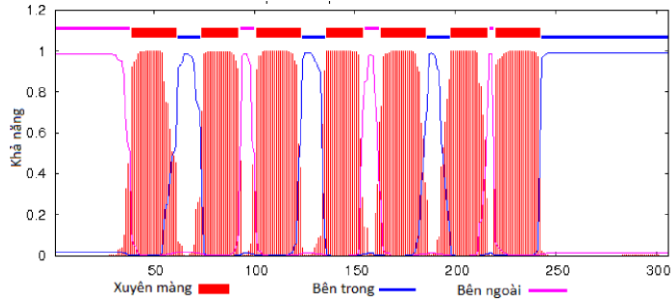
Gene	Tên locus	Nhóm	GL (bp)	PL (aa)	MW (kD)	pI	NST	IN	TMH	SCL
<i>TcSWEET01</i>	Thecc1EG004545	IV	1488	306	34,41	9,47	1	5	7	PM
<i>TcSWEET02</i>	Thecc1EG008493	IV	2080	287	32,16	8,75	2	5	7	PM
<i>TcSWEET03</i>	Thecc1EG012041	II	1456	232	26,37	9,36	3	5	7	PM
<i>TcSWEET04</i>	Thecc1EG014707	IV	1474	281	31,71	9,18	3	5	7	PM
<i>TcSWEET05</i>	Thecc1EG014709	IV	2482	289	32,36	8,66	3	5	7	PM
<i>TcSWEET06</i>	Thecc1EG014710	IV	1766	285	31,84	9,42	3	5	7	PM
<i>TcSWEET07</i>	Thecc1EG015352	I	1450	255	28,59	9,10	3	5	7	PM
<i>TcSWEET08</i>	Thecc1EG016513	II	2102	255	28,26	9,52	3	4	7	PM
<i>TcSWEET09</i>	Thecc1EG016865	II	1870	247	27,37	8,97	4	5	7	PM
<i>TcSWEET10</i>	Thecc1EG016866	II	2449	235	26,81	9,01	4	5	7	PM
<i>TcSWEET11</i>	Thecc1EG021237	III	2524	237	25,82	7,70	4	5	7	PM
<i>TcSWEET12</i>	Thecc1EG026042	IV	1379	254	28,68	8,45	5	5	7	PM
<i>TcSWEET13</i>	Thecc1EG026043	IV	1527	301	34,58	9,28	5	5	7	PM
<i>TcSWEET14</i>	Thecc1EG026675	III	2700	245	27,23	7,72	5	5	6	PM
<i>TcSWEET15</i>	Thecc1EG026676	III	1785	302	33,73	6,20	5	5	7	PM
<i>TcSWEET16</i>	Thecc1EG026677	III	1715	265	29,56	9,13	5	5	6	PM
<i>TcSWEET17</i>	Thecc1EG026679	III	1727	299	32,97	9,11	5	5	6	PM
<i>TcSWEET18</i>	Thecc1EG026680	III	1628	293	32,11	9,45	5	5	7	PM
<i>TcSWEET19</i>	Thecc1EG029586	I	1704	233	25,75	9,18	6	5	7	PM
<i>TcSWEET20</i>	Thecc1EG032142	I	2671	287	31,96	8,53	7	7	7	PM
<i>TcSWEET21</i>	Thecc1EG035405	I	2325	252	27,81	9,23	8	5	7	PM

Ghi chú: GL: Kích thước gene, PL: Chiều dài protein, MW: Khối lượng protein, NST: Nhiễm sắc thể, IN: Số lượng intron, TMH: Số lượng xoắn xuyên màng, SCL: Khu trú dưới tế bào, PM: Plasma membrane.

Các gene mã hóa *SWEET* ở cây ca cao có chiều dài từ 1379 - 2700 nucleotide (Bảng 1). Các gene này đều mã hóa không liên tục, hầu hết các gene (19 gene) có năm intron, chỉ *TcSWEET08* có bốn intron và *TcSWEET20* có bảy intron (Bảng 1). Các protein suy diễn có từ 232 (*TcSWEET03*) tới 306 (*TcSWEET01*) gốc axit amin, khối lượng phân tử lí thuyết từ 25,75 kDa (*TcSWEET19*) tới 34,58 kDa (*TcSWEET13*). Hầu hết protein này có tính kiềm với giá trị *pI* từ 7,70 - 9,52, chỉ duy nhất *TcSWEET15* có *pI* lí thuyết bằng 6,20. Các protein *SWEET* của cây ca cao có giá trị GRAVY nằm trong khoảng 0,32 - 0,93 (Bảng 1). Như vậy các *SWEET* của cây ca cao có khối lượng cũng như *pI* khá tương đồng với *SWEET* của cây sắn (Chu Đức Hà và nnk., 2018).

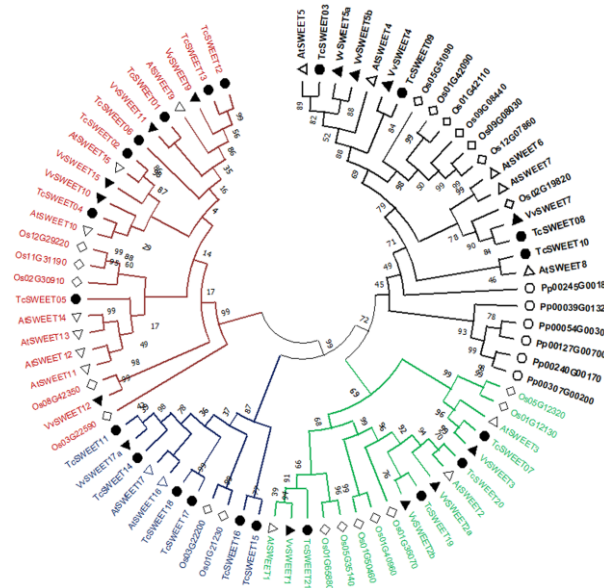
Hầu hết các *SWEET* của cây ca cao (18/21) có mang bảy vùng xoắn xuyên màng với cấu trúc đặc trưng gồm hai vùng 3 xoắn xuyên màng theo cấu trúc 3+1+3 (Hình 1). Riêng ba phân tử *TcSWEET14*, *TcSWEET16* và *TcSWEET17* chỉ có 6 xoắn xuyên màng

(Bảng 1). Cấu trúc đặc trưng này của các TcSWEET tương đồng với các SWEET của các loài đã biết như *A. thaliana*, lúa (Chen et al., 2010) và ở nhiều loài khác (Jeena et al., 2019). Phân tích lí thuyết về vị trí định khu dưới tế bào cho thấy tất cả các TcSWEET đều gắn với hệ thống màng (Bảng 1). Kết quả này phù hợp với cấu trúc của các protein SWEET có chứa các xoắn xuyên màng.



Hình 1. Mô hình cấu trúc với các xoắn xuyên màng điển hình của TcSWEET1 được xây dựng nhờ TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM>).

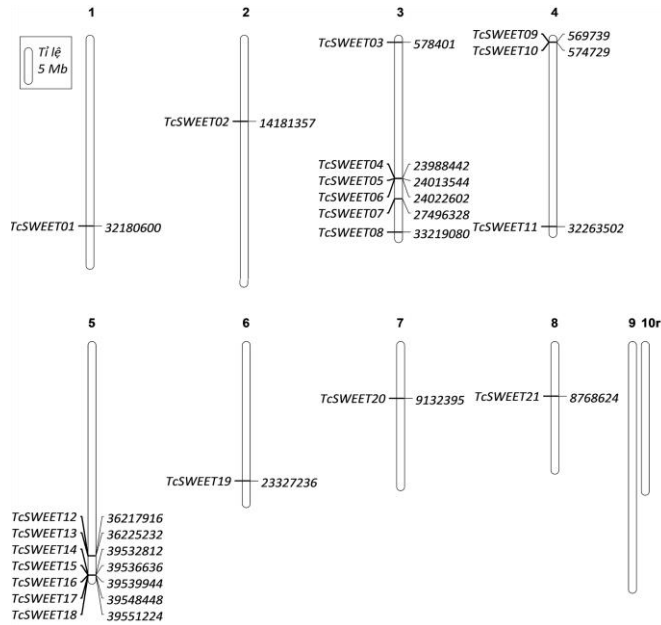
Phân tích cây di truyền



Hình 2. Cây di truyền được xây dựng từ các SWEET của cây cao (Tc), cây *A. thaliana* (At), cây nho (Vv), cây lúa (Os) và cây rêu (Pp)

Phân tích cây di truyền (Hình 2) chỉ ra rằng các SWEET của cây cao thuộc về bốn nhóm khác nhau, tương tự như cây phủ hệ được xây dựng từ các protein SWEET của các loài (Li et al., 2018). Nhóm I và nhóm II có bốn gene, nhóm III có sáu gene và nhóm IV có bảy gene. Các gene SWEET của cây cao phân bố không đồng đều trên toàn bộ hệ gene. Trong đó các NST số 1; 2; 6; 7 và 8 chỉ có một gene, NST số 4 có ba gene, NST số 3 có sáu gene trong khi NST số 5 có tới bảy gene (Hình 3). Sự phân bố không đồng đều của các gene SWEET ở cây cao tương đồng với ở cây sắn (Chu Đức Hà và nnk., 2018).

Vị trí sắp xếp của các *SWEET* trên các NST của cây ca cao cũng như mức độ tương đồng của các gene này cho thấy rằng có nhiều sự kiện nhân gene *SWEET* đã xảy ra, góp phần làm tăng số lượng các gene trong họ này. Căn cứ vào vị trí các gene trong cây di truyền và vị trí các gene trong hệ gene, có ba sự kiện nhân gene trước sau (tandem duplication) phát sinh ra các cặp gene tương đồng *TcSWEET12* và *TcSWEET13* (nhóm 4), *TcSWEET15* và *TcSWEET16*; *TcSWEET17* và *TcSWEET18* (nhóm III). Sự nhân gene này cũng được quan sát ở nhiều loài thực vật khác (Li et al., 2018).



Hình 3. Cây phả hệ được xây dựng từ các *SWEET* của cây ca cao (*Tc*), cây *A. thaliana* (*At*), cây nho (*Vv*), cây lúa (*Os*) và cây rêu (*Pp*).

Khảo sát sự biểu hiện của các gene *TcSWEET*

Sự biểu hiện của các gene *TcSWEET* được khảo sát từ hệ mã phiên EST được xây dựng từ các loại mô, các giai đoạn phát triển và điều kiện khác nhau của cây ca cao trên cơ sở dữ liệu NCBI (txid3641) (Bảng 2).

Trong số 21 gene *SWEET* của ca cao, 9 gene *TcSWEET03*, *TcSWEET10- TcSWEET13* và *TcSWEET16-TcSWEET18* không có EST nào được phát hiện. Các gene còn lại có số EST được phát hiện từ 1 tới 10. Gene *TcSWEET09* chỉ có một EST từ thân, tương tự *TcSWEET14* có một EST từ trụ mầm của cây mầm 2-3 tuần và *TcSWEET19* có một EST từ hỗn hợp hạt và lá. *TcSWEET01* có bốn EST được phát hiện trong đó có hai EST từ hạt 3-3,5 tháng sau thụ phấn và hai EST từ vỏ và thịt hạt trưởng thành. *TcSWEET04* có tới tám EST từ vỏ và thịt hạt trưởng thành và một EST từ vỏ quả lên men trong 6 tới 40 giờ. *TcSWEET05* có một EST thu được từ rễ bị stress hạn trong khi *TcSWEET08* ngoài một EST thu được từ rễ stress hạn còn có một EST thu được từ nõn 2-3 tháng sau thụ phấn. *TcSWEET07* chỉ có một EST từ chồi non nhiễm *Sahlbergella singularis*. Các gene *TcSWEET02*, *TcSWEET15*, *TcSWEET20* và *TcSWEET21* có các EST thu được từ nhiều loại mô khác nhau, cả trong điều kiện thường cũng như bị nhiễm nấm (tác nhân sinh học). Sự biểu hiện của các gene *TcSWEET* bước đầu

gợi ý chúng có vai trò nhất định trong sự phát triển cũng như trong sự chống chịu điều kiện bất lợi vô sinh hoặc hữu sinh. Trong đó có tám gene có các EST ở các cơ quan sinh sản như noãn, hạt và vỏ quả. Sự biểu hiện của các gene *SWEET* liên quan đến sự vận chuyển đường liên tế bào hoặc sự phân bố dinh dưỡng của tác nhân gây bệnh (Chen et al., 2010). Kết quả nghiên cứu này khẳng định các kết quả nghiên cứu về sự biểu hiện cũng như phân tích vai trò của các gene *SWEET* ở một số loài thực vật khác (Jeena et al., 2019).

Bảng 2. Các EST của các gene *TcSWEET* được phát hiện trong hệ mã phiên của cây cao

Gene	EST	Mô/điều kiện	Gene	EST	Mô/điều kiện	
<i>TcSWEET01</i>	CU534764.1, CU534696.1	Vỏ và thịt hạt trưởng thành	<i>TcSWEET14</i>	CU573462.1	Trụ mầm của cây mầm 2-3 tuần	
	CU592828.1, CU592918.1	Hạt 3-3,5 tháng sau thụ phấn.		<i>TcSWEET15</i>	CU477250.1, CU505531.1	Vỏ quả nhiễm <i>Phytophthora palmivora</i>
<i>TcSWEET02</i>	CU605659.1	Chồi non không nhiễm <i>Sahlbergella singularis</i> .	<i>TcSWEET16</i>		CU474050.1	Bầu quả giai đoạn 1 tuần tới 1 tháng .
	ES442624.1	Đỉnh sinh trưởng từ 24h đến 90 ngày ủ với <i>M. Perniciosa</i> .		nd	nd	
	CU495103.1	Hoa ở nhiều giai đoạn phát triển.		<i>TcSWEET17</i>	nd	nd
	CU533205.1	Vỏ và thịt hạt trưởng thành.		<i>TcSWEET18</i>	nd	nd
	CU614537.1	Vỏ lên men 6 tới 40 giờ.		<i>TcSWEET19</i>	CA798159.1	Hạt và lá
	CU525166.1	Bầu quả non 7-10 ngày sau thụ phấn .		<i>TcSWEET20</i>	ES442178.1	Hỗn hợp mô phân sinh không ủ và ủ với <i>M. perniciosa</i> từ 24h đến 90 ngày
<i>TcSWEET03</i>	nd	nd	CU497935.1		Lá nhiễm <i>Phytophthora megakarya</i>	
<i>TcSWEET04</i>	CU534075.1, CU534100.1, CU534128.1, CU534080.1, CU534743.1, CU534144.1, CU533939.1, CU534583.1	Vỏ và thịt hạt trưởng thành.	<i>TcSWEET20</i>	CU578190.1	Vỏ quả nhiễm <i>Moniliophthora rozeri</i> .	
	CU614935.1	Vỏ quả lên men 6 tới 40 giờ.		CU479189.1, CU480292.1	Vỏ quả nhiễm <i>Phytophthora palmivora</i> .	
<i>TcSWEET05</i>	CU490577.1	Rễ bị stress hạn	<i>TcSWEET21</i>	CU628934.1, CU628691.1	Lá mầm của cây mầm 1-3 tuần.	
<i>TcSWEET06</i>	nd	nd		CU541498.1	Phần mô phía trong của vỏ thân với ống lignin hóa.	
<i>TcSWEET07</i>	CU569284.1	Chồi non nhiễm <i>Sahlbergella singularis</i> .		CU541052.1	Phần mô phía trong của vỏ thân với ống lignin hóa.	
<i>TcSWEET08</i>	CU585235.1	Noãn 2-3 tháng sau thụ phấn.		CU517604.1	Hạt 2-5 tháng sau thụ phấn.	
	CU491018.1	Rễ bị stress hạn.		CU596845.1	Hạt 4-5 tháng sau thụ phấn.	
<i>TcSWEET09</i>	CU620334.1	Thân		CU485950.1	Lớp đệm quả non.	
<i>TcSWEET10</i>	nd	nd		CU503719.1	Noãn 1-7 ngày sau thụ phấn.	
<i>TcSWEET11</i>	nd	nd		CU534434.1, CU533859.1	Vỏ và thịt hạt trưởng thành.	
<i>TcSWEET12</i>	nd	nd		FC072071.1	Vỏ quả ủ với <i>M. perniciosa</i> từ 24h đến 120 ngày.	
<i>TcSWEET13</i>	nd	nd				

Ghi chú: nd = không xác định được.

4. KẾT LUẬN

Trong công trình này, 21 gene mã hóa SWEET đã được phát hiện ở trong hệ gene cây ca cao. Các protein suy diễn SWEET của cây ca cao có cấu trúc đặc trưng cho các SWEET đã được biết. Hầu hết các TcSWEET có năm intron. Các protein suy diễn có từ 232 tới 306 axit amin, hầu hết chúng có tính kiềm. Các protein TcSWEET được sắp xếp trong bốn nhóm SWEET điển hình ở thực vật. Các gene SWEET phân bố ở tám trong tổng số 10 nhiễm sắc thể của cây ca cao. Một số hiện tượng nhân gene SWEET sau quá trình biệt hóa loài được phát hiện có liên quan đến số lượng lớn gene SWEET nhóm III và IV ở loài cây này. Mười hai trong số 21 gene biểu hiện cả trong một số loại mô của cây ca cao, ở một số điều kiện khác nhau.

TÀI LIỆU THAM KHẢO

- Afoufa-Bastien D., Medici A., Jeuffre J., et al., 2010. The *Vitis vinifera* sugar transporter gene family: phylogenetic overview and microarray expression profiling. *BMC Plant Biology*, 10(1), 245. doi:10.1186/1471-2229-10-245.
- Argout X., Salse J., Aury J. M., et al., 2011. The genome of *Theobroma cacao*. *Nat Genet*, 43(2), 101-108. doi:10.1038/ng.736.
- Chen L. Q., Hou B. H., Lalonde S., et al., 2010. Sugar transporters for intercellular exchange and nutrition of pathogens. *Nature*, 468(7323), 527-532.
- Figueira A., Alemanno L., Litz R. E., 2005. *Theobroma cacao*. *Biotechnology of fruit and nut crops*, 639-669.
- Gasteiger E., Hoogland C., Gattiker A., et al., 2005. Protein identification and analysis tools on the ExPASy server. In *The proteomics protocols handbook*, Springer, 571-607.
- Guo A. Y., Zhu Q. H., Chen X., Luo J. C., 2007. GSDS: a gene structure display server. *Yi Chuan*, 29(8), 1023-1026.
- Chu Đức Hà, Phạm Thị Quỳnh, Phạm Thị Lý Thu, Nguyễn Văn Cường, 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*, 63(3), 140-149.
- Jeena G. S., Kumar S., Shukla R. K., 2019. Structure, evolution and diverse physiological roles of SWEET sugar transporters in plants. *Plant Mol Biol*, 100(4-5), 351-365. doi:10.1007/s11103-019-00872-4.
- Katoh K., Standley D. M., 2013, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*, 30(4), 772-780.
- Kumar S., Stecher G., Li M., et al., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*, 35(6), 1547-1549.
- Li X., Si W., Qin Q., et al., 2018. Deciphering evolutionary dynamics of SWEET genes in diverse plant lineages. *Scientific Reports*, 8(1), 13440. doi:10.1038/s41598-018-31589-x.
- Motamayor J. C., Mockaitis K., Schmutz J., et al., 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. *Genome Biol*, 14(6), r53. doi:10.1186/gb-2013-14-6-r53.
- Motamayor J. C., Risterucci A. M., Lopez P. A., et al., 2002. Cacao domestication I: the origin of the cacao cultivated by the Mayas. *Heredity*, 89(5), 380-386.

Yuan M., Wang S, 2013. Rice MtN3/saliva/SWEET family genes and their homologs in cellular organisms. *Mol Plant*, 6(3), 665-674. doi:10.1093/mp/sst035.

***IN SILICO* CHARACTERISATION OF GENES ENCODING SWEET PROTEIN IN COCOA (*Theobroma cacao* L.)**

**Cao Phi Bang^{1,*}, Nguyen Van Dinh², Tran Thi Thanh Huyen³,
Le Thi Man¹, Vu Xuan Duong¹**

Abstract: SWEET (sugars will eventually be exported transporter) is one of the important sugar transport protein groups in plants. In this study, 21 genes encoding the SWEET protein in the cocoa genome were identified and analyzed. The genomic full-length of these genes ranged from 1379 to 2700 nucleotides, most of which have five introns. The predicted proteins had 232 to 306 amino acids and contained the conserved transmembrane helix regions of known SWEETs. Most of these proteins were alkaline. Phylogeny analysis showed that the TcSWEETs were divided into four groups, group I (4 genes), group II (4 genes), group III (6 genes), and group IV (7 genes). These genes are not equally distributed in the cocoa genome. Thirteen of the 21 *SWEET* genes presented in two chromosomes, 3 and 5. Some of the gene duplication events that occurred in these two chromosomes were related to the expansion of *SWEET* genes of group III and group IV. Twelve of the 21 *SWEET* genes had EST which were detected in several different types of cocoa tissue, under different conditions. Most of them had ESTs obtained from reproductive tissues or tissues affected by abiotic and biotic stress. The expression of these genes suggested their important role in the development and stress resistance of the cocoa tree.

Keywords: SWEET, gene characterization, gene expression, phylogeny, cocoa.

¹Hung Vuong University

²Hanoi Pedagogical University 2

³Hanoi National University of Education

*Email: phibang.cao@hvu.edu.vn