

Dự đoán cơ chế lặp của họ gen mã hóa protein vận chuyển đường sucrose ở loài đậu gà (*Cicer arietinum*)

Chu Đức Hà^{1*}, Chu Thị Bích Thủy^{1,2}, Phạm Thị Lý Thu¹,
Phạm Phương Thu², La Việt Hồng²

¹Viện Di truyền nông nghiệp, VAAS

²Khoa Sinh - Kỹ thuật nông nghiệp, Trường Đại học Sư phạm Hà Nội 2

Ngày nhận bài 18/3/2019; ngày chuyển phản biện 22/3/2019; ngày nhận phản biện 2/5/2019; ngày chấp nhận đăng 3/6/2019

Tóm tắt:

Ở thực vật, SWEET (sugars will eventually be exported transporter) là họ protein vận chuyển đường sucrose đóng vai trò quan trọng, tham gia vào nhiều quá trình sinh học thiết yếu trong tế bào. Trong nghiên cứu này, hiện tượng lặp của họ gen mã hóa SWEET ở cây đậu gà (*Cicer arietinum*) đã được phân tích thông qua các công cụ tin sinh học. Kết quả đã xác định được 6 sự kiện lặp trong họ gen *CaSWEET* ở đậu gà. 3/6 cặp gen lặp đều nằm trên các nhiễm sắc thể khác nhau, cho thấy hiện tượng lặp trên các vùng lớn của nhiễm sắc thể là cơ chế chính để nhân rộng họ gen *CaSWEET* tương tự như các loài thực vật khác. Phân tích cũng đã dự đoán rằng, sự sai khác trong cấu trúc của các gen lặp đã bị kìm hãm dưới áp lực của chọn lọc tự nhiên. Các sự kiện lặp được tính toán một cách tương đối xảy ra cách đây từ 23,06 đến 38,31 triệu năm trước. Vùng bảo thủ của họ *CaSWEET* đặc trưng bởi 7 domain TM và 7 trình tự motif. Kết quả của nghiên cứu đã cung cấp những dẫn liệu quan trọng về sự nhân rộng của họ gen *CaSWEET*, từ đó có thể đưa ra các giả thuyết về vai trò của các gen lặp đối với loài đậu gà.

Từ khóa: đậu gà, lặp gen, *SWEET*, tin sinh học.

Chỉ số phân loại: 4.6

Mở đầu

Hiện tượng lặp gen (gene duplication) được giả thuyết là cơ chế chính để nhân rộng các họ đa gen (multigene family) ở thực vật [1]. Cụ thể, khoảng 65% gen được giải mã trong genome thực vật đều có thể tìm thấy gen lặp của chúng [1]. Hiện tượng này được giải thích chủ yếu do sự trao đổi bất thường giữa những alen tương đồng tạo nên cặp gen lặp ở khoảng cách gần nhau và sự lặp của các vùng nhiễm sắc thể lớn (segmental duplication) [2]. Vì vậy, nghiên cứu về cơ chế nhân rộng của các họ gen có thể cho phép tìm hiểu về phân hóa chức năng cũng như cấu trúc của các gen này, đây là những dẫn liệu quan trọng nhằm đánh giá vai trò của gen liên quan đến đáp ứng bất lợi của môi trường.

Ở thực vật, đường sucrose (đơn vị dự trữ năng lượng cơ bản trong tế bào) được vận chuyển chủ yếu nhờ nhóm protein SWEET. Trong đó, các gen mã hóa SWEET ở thực vật được chứng minh đóng vai trò trong nhiều quá trình

quan trọng như trao đổi chất, đáp ứng bất lợi... [3]. Trước đây, họ gen gồm 21 thành viên, mã hóa protein SWEET đã được xác định ở đậu gà (*Cicer arietinum*) [3] - cây trồng quan trọng thứ hai trong họ Đậu được sử dụng phổ biến làm thực phẩm, thức ăn gia súc, nhiên liệu sinh học và cải tạo môi trường (cố định N₂) [4-6]. Trong đó, hầu hết các gen *CaSWEET* đều có 6 exon [3]. Vì vậy, câu hỏi được đặt ra là họ gen *CaSWEET* được nhân rộng ra sao, và có tương tự như ở các loài thực vật khác hay không? Trong nghiên cứu này, cơ chế nhân rộng của họ gen *CaSWEET* ở *Cicer arietinum* đã được phân tích *in silico*. Cụ thể, các sự kiện lặp gen đã được dự đoán dựa trên sự tương đồng giữa các gen *CaSWEET* gần gũi. Sau đó, áp lực của chọn lọc tự nhiên và thời điểm xảy ra hiện tượng lặp gen đã được đánh giá. Cuối cùng, trình tự bảo thủ và các motif của họ SWEET đã được tìm hiểu. Kết quả của nghiên cứu này góp phần cung cấp những hiểu biết cơ bản về họ gen *CaSWEET* trong suốt quá trình tiến hóa của đậu gà.

*Tác giả liên hệ: Email: hachu_amser@yahoo.com

Prediction of the gene duplication mechanism in the genes encoding sucrose transporters in chickpea (*Cicer arietinum*)

Duc Ha Chu^{1*}, Thi Bich Thuy Chu^{1,2},
Thi Ly Thu Pham¹, Phuong Thu Pham², Viet Hong La²

¹Agricultural Genetics Institute (VAAS)

²Faculty of Biology - Agricultural Technology,
Hanoi Pedagogical University 2

Received 18 March 2019; accepted 3 June 2019

Abstract:

In plants, SWEETs (sugars will eventually be exported transporters) are known as the major sucrose transporters that involve in various important biological processes. In this study, the gene duplication of the gene family encoding SWEETs in chickpea (*Cicer arietinum*) has been analysed based on the bioinformatics approach. As the results, a total of six duplication events occurring in the *CaSWEET* gene family in chickpea has been determined. Three (out of six) duplicated pairs were distributed on the distinct chromosomes, suggesting that the segmental duplication is considered as the major reason to explain the expansion of the *CaSWEET* gene family, as confirmed in many plant species. Our results also predicted that the purifying selection was appeared to maintain the preservation of SWEET proteins. These duplication events were calculated to approximately occur from 23.06 to 38.31 million years ago. Additionally, the conserved domain of *CaSWEET* consisted of seven TM units and seven specific motifs. Our study would provide a critical understanding of the expansion of *CaSWEET* genes, this could suggest the role of duplicated genes in chickpea plants.

Keywords: bioinformatics, chickpea, gene duplication, *SWEET*.

Classification number: 4.6

Vật liệu và phương pháp nghiên cứu

Vật liệu

Genome và proteome của giống đậu gà kabuli 'CDC Frontier' [7] được khai thác trên Phytozome (<https://phytozome.jgi.doe.gov/>) [8].

Trình tự protein, đoạn exon mã hóa (coding DNA sequence, CDS) và vùng gen (genomic sequence) của 21 *CaSWEET* được khai thác từ nghiên cứu trước đây [3].

Phương pháp nghiên cứu

Phương pháp dự đoán sự kiện lặp gen: trình tự CDS (định dạng .fasta) của họ *CaSWEET* được sử dụng để so sánh trình tự tương đồng bằng công cụ ClustalX [9]. Mức độ tương đồng (%) của các trình tự CDS được tính toán bằng phần mềm BioEDIT [10]. Các bước tiến hành và thông số cài đặt được mô tả tương tự như trong nghiên cứu gần đây [11].

Phương pháp xây dựng cây phân loại: trình tự protein (định dạng .fasta) của họ *SWEET* được khai thác để phân tích trình tự tương đồng và thiết lập cây quan hệ phát sinh bằng phần mềm MEGA [12]. Trong đó, cây phân loại được xây dựng trên phương pháp Neighbor-Joining (N-J), phân tích bootstrap với 1.000 lần lấy lại mẫu như trong nghiên cứu gần đây [11].

Phương pháp xác định trị số thay thế đồng nghĩa và trái nghĩa: trị số thay thế đồng nghĩa (the number of synonymous substitutions per synonymous site - Ks) và trị số thay thế trái nghĩa (the number of nonsynonymous substitutions per non-synonymous site - Ka) của các sự kiện lặp gen của họ *CaSWEET* được dự đoán bằng cách truy vấn cặp CDS đã căn trình tự vào phần mềm DNAsp [13]. Thời điểm xảy ra sự kiện lặp gen (Y) được tính toán theo công thức $Y = Ks/2\lambda$, với $\lambda = 6,5 \times 10^{-9}$ (năm) [14].

Phương pháp phân tích vùng bảo thủ và motif: vùng bảo thủ SWEET ở đậu gà được kết xuất từ phần mềm MEGA [12]. Kết quả minh họa trình tự tương đồng của họ *CaSWEET* được biểu diễn trên phần mềm BioEDIT [10]. Các motif đặc trưng được xác định bằng cách truy vấn trình tự protein (.fasta) trên phần mềm MEME [15]. Kích thước motif được giới hạn trong khoảng 6÷50 amino acid với tối đa 29 motif.

Kết quả và thảo luận

Kết quả dự đoán hiện tượng lặp gen của họ gen *CaSWEET* ở đậu gà

Để dự đoán hiện tượng lặp trong họ gen *CaSWEET* ở đậu gà, trình tự CDS của các gen được căn trình tự trên

ClustalX [9] để đánh giá mức độ tương đồng bằng BioEDIT [10]. Trong nghiên cứu này, sự kiện gen lặp được định nghĩa là hai (hoặc nhiều gen) có trình tự nucleotide (CDS) tương đồng lớn hơn 70% như trong mô tả gần đây [16]. Kết quả đã xác định được 6 cặp gen lặp, có mức độ tương đồng cao hơn 70% (bảng 1). Cụ thể, *CaSWEET15* và *CaSWEET17* là cặp gen lặp có mức độ tương đồng thấp nhất (đạt 72,6%), trong khi *CaSWEET08* và *CaSWEET09* là cặp gen lặp có tương đồng ở cấp độ nucleotide cao nhất (77,3%) (bảng 1). Trước đó, 6 cặp gen lặp đã được dự đoán trong họ *MeSWEET* ở sắn (*Manihot esculenta*) với mức độ tương đồng lớn hơn 50% [11].

Bảng 1. Hiện tượng lặp của họ gen *CaSWEET* ở đậu gà.

TT	Cặp gen lặp	Chr	E	H (%)	Ka	Ks	Ka/Ks	Y (mya)
1	<i>CaSWEET01/CaSWEET21</i>	Ca1/US	-	73,5	0,246	0,450	0,547	34,58
2	<i>CaSWEET06/CaSWEET07</i>	Ca2/Ca2	T	75,6	0,199	0,382	0,520	29,41
3	<i>CaSWEET08/CaSWEET09</i>	Ca3/Ca3	T	77,3	0,231	0,300	0,770	23,06
4	<i>CaSWEET10/CaSWEET12</i>	Ca4/Ca5	S	74,8	0,243	0,388	0,626	29,83
5	<i>CaSWEET11/CaSWEET16</i>	Ca4/Ca5	S	73,3	0,244	0,498	0,490	38,31
6	<i>CaSWEET15/CaSWEET17</i>	Ca5/Ca6	S	72,6	0,292	0,420	0,695	32,33

TT: Số thứ tự; Chr: nhiễm sắc thể; US: vùng chưa chú giải; E: sự kiện lặp; T: lặp trong vùng lân cận trên cùng nhiễm sắc thể; S: lặp trong vùng lớn trên các nhiễm sắc thể; -: không xác định; H: mức độ tương đồng; Ka: giá trị thay thế trái nghĩa; Ks: giá trị thay thế đồng nghĩa; Y: thời điểm xảy ra hiện tượng lặp gen; mya: triệu năm về trước.

Phần lớn sự kiện lặp được dự đoán trong họ gen *CaSWEET* ở *C. arietinum* đều xảy ra trên các nhiễm sắc thể khác nhau (bảng 1). Cụ thể, 3 cặp gen lặp '*CaSWEET10* và *CaSWEET12*', '*CaSWEET11* và *CaSWEET16*' lần lượt phân bố trên nhiễm sắc thể Ca4 và Ca5, trong khi '*CaSWEET15* và *CaSWEET17*' nằm trên Ca5 và Ca6 (bảng 1). 2 cặp gen khác là '*CaSWEET06* và *CaSWEET07*' và '*CaSWEET08* và *CaSWEET09*' lần lượt cùng định vị trên nhiễm sắc thể Ca2 và Ca3, trong khi *CaSWEET21* trong cặp gen lặp '*CaSWEET01* và *CaSWEET21*' chưa được chú giải trên genome của *C. arietinum* [7] như trong ghi nhận gần đây [3] (bảng 1). Có thể thấy rằng, cơ chế lặp của họ gen *CaSWEET* được diễn ra chủ yếu do hiện tượng lặp trên các vùng nhiễm sắc thể khác nhau (3/6 sự kiện lặp) (bảng 1). Trước đó, hiện tượng lặp trên các nhiễm sắc thể cũng được ghi nhận đóng vai trò chính trong sự nhân rộng của họ gen mã hóa SWEET ở sắn [11], đậu tương (*Glycine max*) [17], cao lương (*Sorghum bicolor*) [18], mía đường (*Saccharum spp.*) [19]. Điều này cho thấy hiện tượng lặp trên các nhiễm sắc thể là nguyên nhân chính cho sự nhân rộng của họ gen *SWEET* ở thực vật nói chung.

Kết quả đánh giá vai trò của đột biến và chọn lọc tự nhiên đến quá trình lặp ở họ gen *CaSWEET*

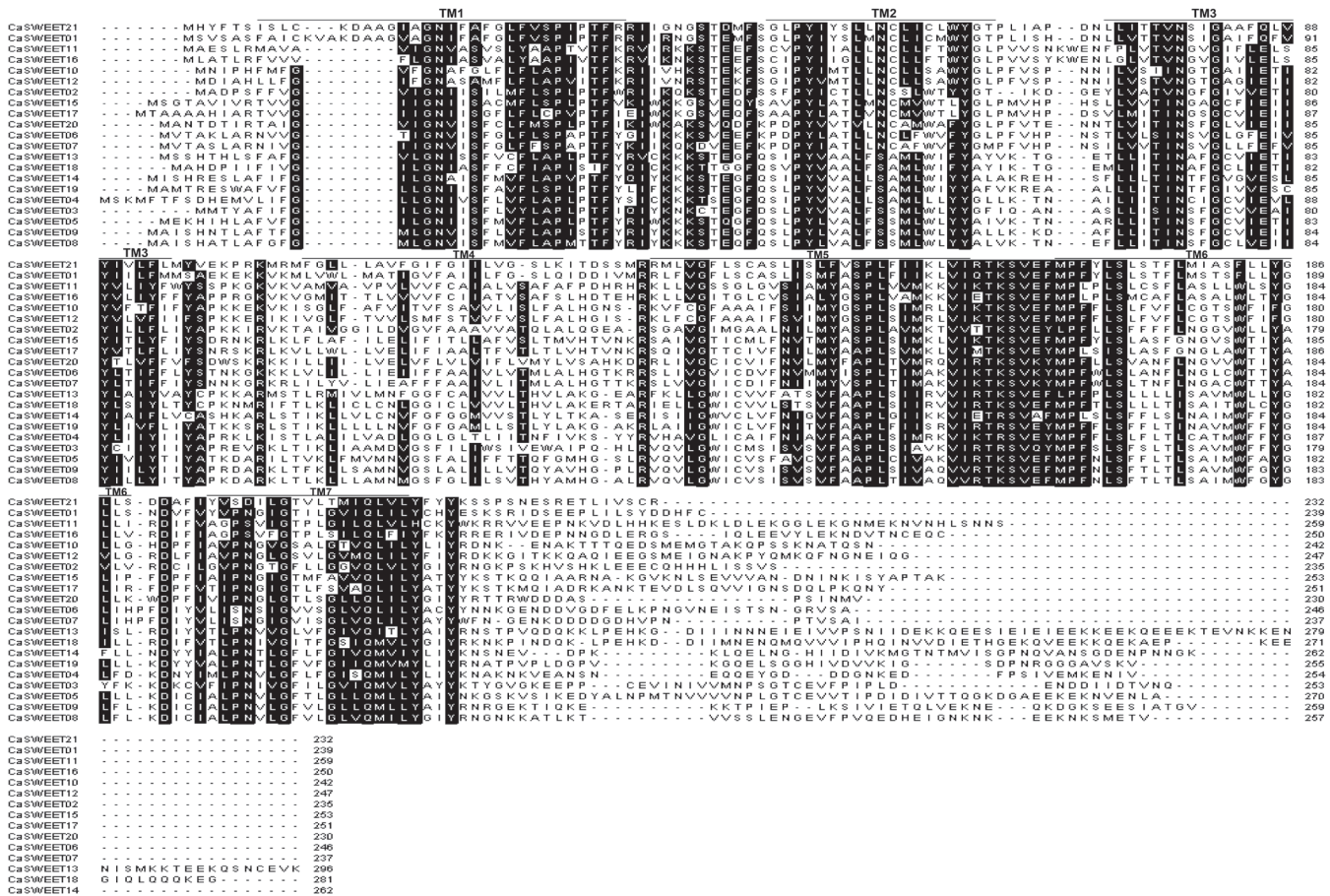
Để đánh giá vai trò của đột biến và áp lực của chọn lọc tự nhiên đến quá trình lặp của họ gen *CaSWEET*, trị số Ka, Ks và tỷ lệ Ka/Ks giữa 6 cặp gen lặp đã được nghiên cứu. Cụ thể, tỷ lệ Ka/Ks < 1, nghĩa là các đột biến điểm xảy ra trong cấu trúc gen bị kìm hãm, cặp gen có mức độ tương đồng cao, trong khi tỷ lệ này > 1, cho thấy chọn lọc tự nhiên đã thúc đẩy sự hình thành gen mới. Kết quả phân tích giá trị Ka/Ks của 6 cặp gen lặp trên phần mềm DNAsp [13] được thể hiện ở bảng 1.

Tất cả giá trị Ka/Ks của 6 sự kiện đều < 1, cho thấy số lượng đột biến thay thế trái nghĩa (Ka) nhỏ hơn đột biến thay thế đồng nghĩa (Ks) trên các gen lặp (bảng 1). Kết quả cho thấy, hiện tượng đột biến điểm gây sai khác xảy ra trên *CaSWEET* đã bị kìm hãm dưới áp lực của chọn lọc tự nhiên, từ đó đảm bảo sự toàn vẹn trong cấu trúc của SWEET ở *C. arietinum*. Trước đây, giá trị Ka/Ks < 1 cũng được ghi nhận ở 3/6 sự kiện lặp của họ *MeSWEET* ở sắn [11]. Ở đậu tương, vai trò kìm hãm sai khác của tiến hóa cũng được dự đoán ở phần lớn các sự kiện lặp (20/21) trong họ gen *GmSWEET* [17]. Những kết quả này đã chỉ ra rằng, trải qua quá trình tiến hóa, chọn lọc tự nhiên có thể đã gìn giữ sự toàn vẹn trong cấu trúc của họ gen *SWEET* nhằm đảm bảo chức năng của chúng ở phần lớn các loài thực vật.

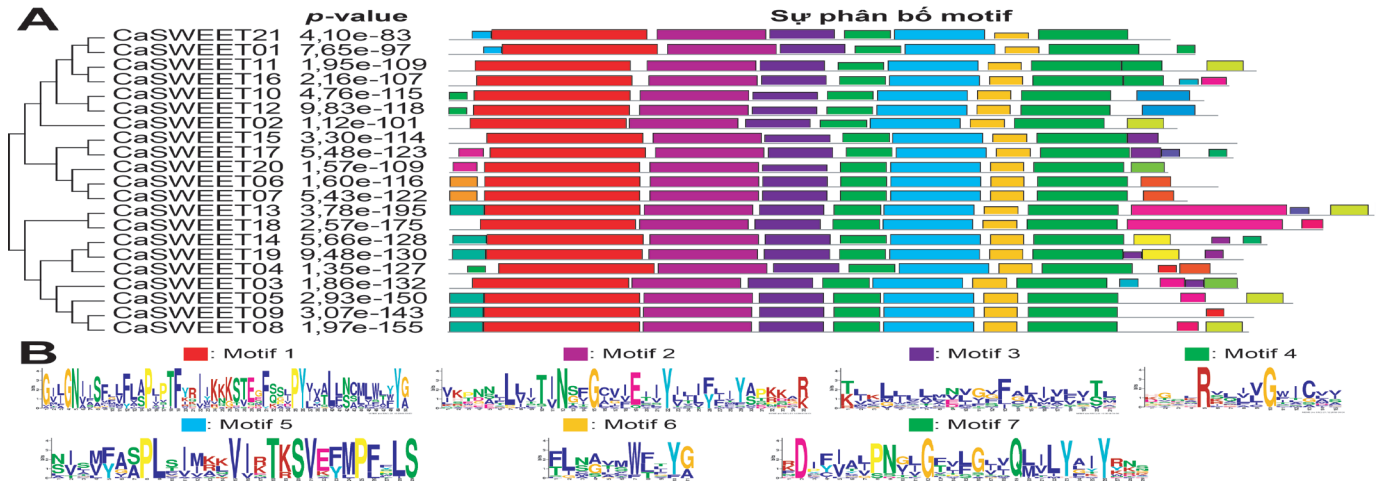
Cuối cùng, thời điểm xảy ra sự kiện lặp của họ gen *CaSWEET* được dự đoán một cách tương đối dựa trên công thức đã mô tả trước đây [14]. Các cặp gen *CaSWEET* lặp được xảy ra cách đây từ khoảng 23,06 ('*CaSWEET08* và *CaSWEET09*') đến 38,31 triệu năm về trước ('*CaSWEET11* và *CaSWEET16*') (bảng 1). Điều này cho thấy, các cặp gen lặp của họ *CaSWEET* đều xảy ra sau sự kiện lặp genome ở phân họ đậu (Papilionoideae), được cho là diễn ra cách đây khoảng 58 triệu năm về trước [20]. Trong đó, nhóm *Medicago truncatula*, *Lotus japonicus* và *C. arietinum* tách biệt với nhánh *Glycine max* và đậu triều (*Cajanus cajan*) cách đây khoảng 54 triệu năm về trước [21].

Kết quả phân tích vùng bảo thủ và trình tự motif đặc trưng của protein SWEET ở đậu gà

Trong nghiên cứu này, cấu trúc vùng bảo thủ và trình tự các motif đặc trưng của họ SWEET ở đậu gà lần lượt được phân tích trên MEGA [12] và MEME [15]. Kết quả xác định vùng bảo thủ và sự phân bố motif của họ SWEET được minh họa ở hình 1 và 2.



Hình 1. So sánh trình tự tương đồng của họ protein SWEET ở đậu gà.



Hình 2. Phân tích trình tự motif đặc trưng của họ SWEET ở đậu gà.

Dựa trên thuật toán căn trình tự tương đồng, vùng bảo thủ của họ SWEET ở đậu gà gồm 7 domain TM, trong đó có hai cặp lập của 3 domain TM ('TM1-TM2-TM3' và 'TM5-TM6-TM7') được xen kẽ bởi đoạn TM4 (hình 1). Đây là dạng cấu trúc đặc trưng của protein vận chuyển SWEET, đã được báo cáo trên đậu tương [17] và ở thực vật nói chung [22].

Phân tích cho thấy, họ SWEET ở đậu gà chứa rất đa dạng các vùng motif đặc trưng (hình 2A). Cụ thể, 7 motif đã được tìm thấy ở tất cả các SWEET ở đậu gà (hình 2A, B), trong khi một số motif khác xuất hiện ở các CaSWEET trong cùng nhánh phân loại (hình 2A). Trong khi đó, ở cải dầu (*Brassica napus*) chỉ ghi nhận 5 motif đặc trưng cho họ

BnSWEET [23]. Sự tương đồng về vùng bảo thủ (hình 1) và các trình tự motif đặc trưng (hình 2) của họ SWEET ở đậu gà có thể được giải thích do sự toàn vẹn về số lượng cũng như cấu trúc của các gen mã hóa (6 exon) đã được ghi nhận trong nghiên cứu trước đây [3].

Kết luận

Đã xác định được 6 sự kiện lặp trong họ gen *CaSWEET* ở đậu gà. Phần lớn các cặp gen lặp nằm trên nhiễm sắc thể khác nhau, chứng tỏ quá trình lặp trên các nhiễm sắc thể là cơ chế chính giải thích cho sự nhân rộng của họ gen *CaSWEET* ở đậu gà.

Tỷ lệ Ka/Ks của 6 sự kiện lặp ở họ gen *CaSWEET* nhỏ hơn 1, chứng tỏ chọn lọc tự nhiên đã kìm hãm sự sai khác trong cấu trúc của các gen lặp. Các sự kiện lặp được dự đoán xảy ra cách đây từ 23,06 đến 38,31 triệu năm về trước.

Cấu trúc vùng bảo thủ của họ *CaSWEET* ở đậu gà chứa 7 domain TM đặc trưng. Phân tích cũng đã chỉ ra 7 trình tự motif đặc trưng nhất cho các *CaSWEET*.

TÀI LIỆU THAM KHẢO

- [1] N. Panchy, M. Lehti-Shiu, S.H. Shiu (2016), "Evolution of gene duplication in plants", *Plant Physiol.*, **171(4)**, pp.2294.
- [2] J.W. Clark, P.C.J. Donoghue (2018), "Whole-genome duplication and plant macroevolution", *Trends Plant. Sci.*, **23(10)**, pp.933-945.
- [3] Chu Đức Hà, Chu Thị Hồng, Phùng Thị Vượng, Phạm Thị Lý Thu, Phạm Phương Thu, La Việt Hồng (2019), "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*, **194(1)**, tr.133-138.
- [4] H.S. El-Beltagi, N.A. El-Senousi, Z.A. Ali, A.A. Omran (2017), "The impact of using chickpea flour and dried carp fish powder on pizza quality", *PLoS ONE*, **12(9)**, pp.e0183657.
- [5] M.N. Esfahani, et al. (2014), "Mechanisms of physiological adjustment of N_2 fixation in *Cicer arietinum* L. (chickpea) during early stages of water deficit: single or multi-factor controls", *Plant J.*, **79(6)**, pp.964-980.
- [6] M. Nasr Esfahani, et al. (2017), "Comparative transcriptome analysis of nodules of two *Mesorhizobium*-chickpea associations with differential symbiotic efficiency under phosphate deficiency", *Plant J.*, **91(5)**, pp.911-926.
- [7] R.K. Varshney, et al. (2013), "Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement", *Nat. Biotech.*, **31(3)**, pp.240-246.
- [8] D.M. Goodstein, et al. (2012), "Phytozome: A comparative platform for green plant genomics", *Nucleic Acids Res.*, **40(Database issue)**, pp.D1178-D1186.
- [9] J. Thompson, et al. (1997), "The ClustalX windows interface:

flexible strategies for multiple sequence alignment aided by quality analysis tools", *Nucleic Acids Res.*, **25(24)**, pp.4876-4882.

- [10] T.A. Hall (1999), "BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT", *Nucleic Acids Symp. Ser.*, **41**, pp.95-98.
- [11] Chu Duc Ha, Le Xuan Dac, Tran Thi Thanh Huyen, Pham Thi Ly Thu (2017), "Evolutionary analysis and expression profiling of the SWEET sugar transporter gene family in cassava (*Manihot esculenta* Crantz)", *HNUE Journal of Science*, **62(10)**, pp.91-99.
- [12] S. Kumar, G. Stecher, K. Tamura (2016), "MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets", *Mol. Biol. Evol.*, **33(7)**, pp.1870-1874.
- [13] P. Librado, J. Rozas (2009), "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data", *Bioinformatics*, **25(11)**, pp.1451-1452.
- [14] B.S. Gaut, et al. (1996), "Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*", *Proc. Natl. Acad. Sci. USA*, **93(19)**, pp.10274-10279.
- [15] T.L. Bailey, J. Johnson, C.E. Grant, W.S. Noble (2015), "The MEME suite", *Nucleic Acids Res.*, **43(W1)**, pp.W39-W49.
- [16] H.D. Chu, et al. (2018), "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.*, **19(11)**, p.E3290.
- [17] G. Patil, et al. (2015), "Soybean (*Glycine max*) SWEET gene family: insights through comparative genomics, transcriptome profiling and whole genome re-sequence analysis", *BMC Genomics*, **16**, p.520.
- [18] H. Mizuno, S. Kasuga, H. Kawahigashi (2016), "The sorghum SWEET gene family: stem sucrose accumulation as revealed through transcriptome profiling", *Biotechnol. Biofuels*, **9**, p.127.
- [19] Q. Zhang, et al. (2016), "Structure, phylogeny, allelic haplotypes and expression of sucrose transporter gene families in *Saccharum*", *BMC Genomics*, **17**, p.88.
- [20] S.B. Cannon, et al. (2010), "Polyploidy did not predate the evolution of nodulation in all legumes", *PLoS ONE*, **5(7)**, pp.e11630.
- [21] M. Lavin, P.S. Herendeen, M.F. Wojciechowski (2005), "Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary", *Syst. Biol.*, **54(4)**, pp.575-594.
- [22] Y.H. Xuan, et al. (2013), "Functional role of oligomerization for bacterial and plant SWEET sugar transporter family", *Proc. Natl. Acad. Sci. USA*, **110(39)**, pp.E3685-E3694.
- [23] H. Jian, et al. (2016), "Genome-wide analysis and expression profiling of the SUC and SWEET gene families of sucrose transporters in oilseed rape (*Brassica napus* L.)", *Front. Plant Sci.*, **7**, p.1464.