

ĐỊNH DANH VÀ PHÂN TÍCH CẤU TRÚC CỦA HỌ GEN MÃ HÓA PROTEIN VẬN CHUYỂN ĐƯỜNG SUCROSE Ở CÂY ĐẬU GÀ (*Cicer arietinum*)

Chu Đức Hà^{1*}, Phùng Thị Vương^{2,3}, Chu Thị Hồng^{1,2}, Phạm Thị Lý Thu¹,
Phạm Phương Thu², Trần Thị Phương Liên², La Việt Hồng²

¹ Viện Di truyền Nông nghiệp (VAAS), ² Trường Đại học Sư phạm Hà Nội 2,
³ Trường THPT Ngô Quyền - Ba Vì, Hà Nội

TÓM TẮT

Trong nghiên cứu này, 21 protein vận chuyển sucrose (sugars will eventually be exported transporter, SWEET) đã được xác định ở giống đậu gà kabuli 'CDC Frontier'. Phần lớn gen *SWEET* nằm rải rác trên hệ gen của đậu gà, ngoại trừ gen *CaSWEET21*. Đáng chú ý, một số gen *CaSWEET* nằm ở vùng cận đầu mút của các nhiễm sắc thể. So sánh cho thấy không có sự tương quan giữa số lượng gen *SWEET* với kích thước hệ gen, số lượng nhiễm sắc thể ở các loài thực vật. Bằng công cụ PIECE, kết quả đã cho thấy họ gen mã hóa *SWEET* ở đậu gà có cấu trúc phân mảnh. Trong đó, hầu hết các gen *CaSWEET* đều có sáu exon. Kết quả của nghiên cứu này đã cung cấp những thông tin cơ bản về họ *SWEET* ở đậu gà, đặc tính cơ bản của *SWEET* và mức độ đáp ứng của các gen *SWEET* với điều kiện bất lợi sẽ được tìm hiểu trong nghiên cứu tiếp theo.

Từ khóa: Đậu gà, sucrose, vận chuyển, *SWEET*, tin sinh học

Ngày nhận bài: 17/12/2018; Ngày hoàn thiện: 03/01/2019; Ngày duyệt đăng: 31/01/2019

ANNOTATION AND STRUCTURAL ANALYSIS OF THE SUCROSE TRANSPORTER GENE FAMILY IN CHICKPEA (*Cicer arietinum*)

Chu Duc Ha^{1*}, Phung Thi Vuong^{2,3}, Chu Thi Hong^{1,2}, Pham Thi Ly Thu¹,
Pham Phuong Thu², Tran Thi Phuong Lien², La Viet Hong²

¹Agricultural Genetics Institute (VAAS), ²Hanoi Pedagogical University 2,
³Ngo Quyen Highschool - Ba Vi, Ha Noi

ABSTRACT

In this study, a total of 21 members of the sucrose transporter family (sugars will eventually be exported transporter, SWEET) has been identified in the chickpea kabuli 'CDC Frontier' cultivar. The majority of genes encoding SWEET, excluding *CaSWEET21* is located on the chromosomes with an uneven ratio. Interestingly, several *CaSWEET* genes were found in the subtelomeric regions in the chromosomes. Our comparisons showed that no direct correlation between the number of *SWEET* genes, genome size and the number of chromosomes in plant species. By using the PIECE tool, we found that the fragmented structure in the *CaSWEET* gene family. Among them, most of the *CaSWEET* genes had six exons. Our study would provide an initial understanding of the SWEET family in chickpea. The general characteristics of SWEETs and the expression profiles of *SWEET* genes in various stress conditions will be carried out in further studies.

Keywords: Chickpea, sucrose, transporter, SWEET, bioinformatics

Received: 17/12/2018; Revised: 03/01/2019; Approved: 31/01/2019

* Corresponding author: Tel: 0983 766070, Email: hachu_amser@yahoo.com

MỞ ĐẦU

Ở thực vật, đường sucrose được tổng hợp ở lá thông qua quá trình quang hợp, sau đó được phân phối đến toàn bộ mô cơ quan trong cây. Có hai nhóm protein chức năng, SUT (sucrose transporter) và SWEET (sugars will eventually be exported transporter) đã được chứng minh tham gia vào quá trình vận chuyển đường sucrose ở thực vật [10]. Trong đó, các gen mã hóa họ protein SWEET được quan tâm nhiều do chúng tham gia vào các quá trình sinh học, như trao đổi chất, và đáp ứng bất lợi ở cây trồng [5]. Đến nay, họ gen mã hóa protein SWEET đã được ghi nhận trên nhiều đối tượng cây trồng quan trọng, có thể kể đến như lúa gạo (*Oryza sativa*) [16], đậu tương (*Glycine max*) [13], cải dầu (*Brassica napus*) [9], cao lương (*Sorghum bicolor*) [12] và sắn (*Manihot esculenta*) [1]. Tuy nhiên, chưa có ghi nhận về họ gen mã hóa SWEET trên đậu gà (*Cicer arietinum*), một trong những cây họ Đậu quan trọng nhất trên thế giới.

Chứa nhiều chất dinh dưỡng, hàm lượng protein cao, đậu gà có thể được sử dụng làm thực phẩm, nguyên liệu cho chế biến thức ăn gia súc và nhiên liệu sinh học [2]. Bên cạnh đó, cây đậu gà có khả năng cố định N₂ tự do, vì vậy được khuyến cáo trồng luân canh với cây trồng khác nhằm cải thiện năng suất và bù đắp độ phì nhiêu cho đất [3]. Vì vậy, *C. arietinum* là đối tượng cây trồng được quan tâm nghiên cứu trên thế giới nhằm phát triển nền nông nghiệp bền vững.

Trong nghiên cứu này, họ gen mã hóa protein SWEET đã được xác định một cách có hệ thống trên hệ gen của đậu gà dựa trên cách tiếp cận tin sinh học. Cấu trúc gen và một số đặc tính cơ bản của protein đã được phân tích bằng các phần mềm sinh học. Kết quả của nghiên cứu này sẽ cung cấp những dẫn liệu khoa học quan trọng cho nghiên cứu cơ chế vận chuyển đường sucrose ở đậu gà, đồng thời tạo ra tiền đề cơ bản cho phát triển đậu gà tại Việt Nam.

VẬT LIỆU VÀ PHƯƠNG PHÁP NGHIÊN CỨU

Dữ liệu nghiên cứu

Hệ gen và hệ protein của đậu gà kabuli 'CDC Frontier' [14] trên cơ sở dữ liệu Phytozome v. 12.1 (<https://phytozome.jgi.doe.gov/>) [7].

Phương pháp nghiên cứu

Phương pháp tìm kiếm SWEET ở đậu gà: Các thành viên của họ SWEET được xác định bằng cách BlastP vùng bảo thủ PF03083 đặc trưng cho protein vận chuyển sucrose ở thực vật [10] vào hệ protein của *C. arietinum* trên Phytozome [7]. Tất cả kết quả tìm kiếm tương đồng có giá trị E-value > 1×10^{-6} được thu thập để kiểm chứng trên Pfam [6].

Phương pháp chú giải gen mã hóa SWEET ở đậu gà: Mã định danh và các thông tin về chú giải của gen SWEET được thu thập thông qua tìm kiếm BlastP trên NCBI (<https://www.ncbi.nlm.nih.gov/>) [14]. Vị trí phân bố trên nhiễm sắc thể được mô hình hóa trên Illustrator dựa trên kích thước hệ gen của *C. arietinum* [14].

Phương pháp xây dựng cây phân loại cho họ SWEET ở đậu gà: Công cụ MEGA v. 7.0 [11] được dùng để thiết lập cây phân loại họ SWEET bằng thuật toán Neighbor-Joining với giá trị bootstrap 1000 lần lặp lại. Cây phân loại được mô hình hóa trên Illustrator.

Phương pháp mô phỏng cấu trúc của gen mã hóa SWEET ở đậu gà: Trình tự genomic (.fasta), vùng mã hóa (.fasta) của gen SWEET và thứ tự của các SWEET (.nwk) được truy vấn vào công cụ PIECE (<https://wheat.pw.usda.gov/piece/index.php>) [15]. Cấu trúc exon/intron của gen SWEET được mô hình hóa trên Illustrator.

KẾT QUẢ VÀ BÀN LUẬN

Kết quả tìm kiếm protein vận chuyển sucrose SWEET ở *C. arietinum*

Đầu tiên, thuật toán BlastP được sử dụng để sàng lọc toàn bộ protein có vùng bảo thủ PF03083 đặc trưng cho SWEET ở thực vật [10] trên hệ protein của *C. arietinum* [14].

Đối chiếu các kết quả trên Pfam [6] đã xác định được tổng số 21 protein SWEET ở đậu gà (Bảng 1). Trình tự amino acid của các SWEET được khai thác để chú giải thông tin vào hệ gen của *C. arietinum* [14] trên NCBI. Toàn bộ thông tin về họ SWEET ở đậu gà được cung cấp ở Bảng 1.

Bảng 1. Thông tin của họ SWEET ở đậu gà

TT	Tên gen	Mã phiên mã	Mã protein	Mã locus
1	<i>CaSWEET01</i>	XM_004488926.2	XP_004488983.1	LOC101509458
2	<i>CaSWEET02</i>	XM_004487778.2	XP_004487835.1	LOC101509872
3	<i>CaSWEET03</i>	XM_012719392.1	XP_012574846.1	LOC101498095
4	<i>CaSWEET04</i>	XM_004489049.2	XP_004489106.1	LOC101497133
5	<i>CaSWEET05</i>	XM_004490445.2	XP_004490502.1	LOC101497351
6	<i>CaSWEET06</i>	XM_004489239.2	XP_004489296.1	LOC101506045
7	<i>CaSWEET07</i>	XM_004489238.2	XP_004489295.1	LOC101505723
8	<i>CaSWEET08</i>	XM_004491623.2	XP_004491680.1	LOC101511936
9	<i>CaSWEET09</i>	XM_004491624.2	XP_004491681.1	LOC101512270
10	<i>CaSWEET10</i>	XM_004498321.2	XP_004498378.1	LOC101498274
11	<i>CaSWEET11</i>	XM_004498340.2	XP_004498397.1	LOC101504169
12	<i>CaSWEET12</i>	XM_004502518.2	XP_004502575.1	LOC101515250
13	<i>CaSWEET13</i>	XM_004501669.2	XP_004501726.1	LOC101510607
14	<i>CaSWEET14</i>	XM_004501759.2	XP_004501816.1	LOC101488880
15	<i>CaSWEET15</i>	XM_004502010.1	XP_004502067.1	LOC101512545
16	<i>CaSWEET16</i>	XM_004502557.2	XP_004502614.1	LOC101499800
17	<i>CaSWEET17</i>	XM_004503532.2	XP_004503589.1	LOC101488443
18	<i>CaSWEET18</i>	XM_004503722.2	XP_004503779.1	LOC101491370
19	<i>CaSWEET19</i>	XM_004503721.1	XP_004503778.1	LOC101491054
20	<i>CaSWEET20</i>	XM_004508799.2	XP_004508856.1	LOC101491395
21	<i>CaSWEET21</i>	XM_004515143.1	XP_004515200.1	LOC101489507

Trước đó, hệ gen của giống *C. arietinum* kabuli 'CDC Frontier' được dự đoán có kích thước thực tế ~738 Mb, tuy nhiên chỉ 530,894 Mb (bao phủ khoảng 71,94%) được ghi nhận trong bản mô tả hệ gen gần đây (BioProject: PRJNA175619) [14]. Hơn nữa, một giống *C. arietinum* khác là desi 'ICC 4958' cũng đã được giải trình tự gần đây, với kích thước hệ gen đạt 510,877 Mb (BioProject: PRJNA78951) [8]. Vì vậy, trong những bản mô tả tiếp theo về hệ gen của *C. arietinum*, số lượng cũng như thông tin của họ SWEET sẽ được bổ sung và cải thiện.

Kết quả xác định vị trí phân bố của gen mã hóa SWEET ở *C. arietinum* và phân tích mối tương quan về họ gen SWEET ở thực vật

Minh họa vị trí phân bố của họ gen mã hóa SWEET trên hệ gen của *C. arietinum* cho thấy phần lớn các gen đều nằm trên nhiễm sắc thể, ngoại trừ *CaSWEET21* chưa được chú giải (unplaced scaffold) (Hình 1). Họ gen SWEET phân bố rải rác trên hầu hết các nhiễm sắc thể, trong khi không có gen nào được ghi nhận trên nhiễm sắc thể Ca8 (Hình

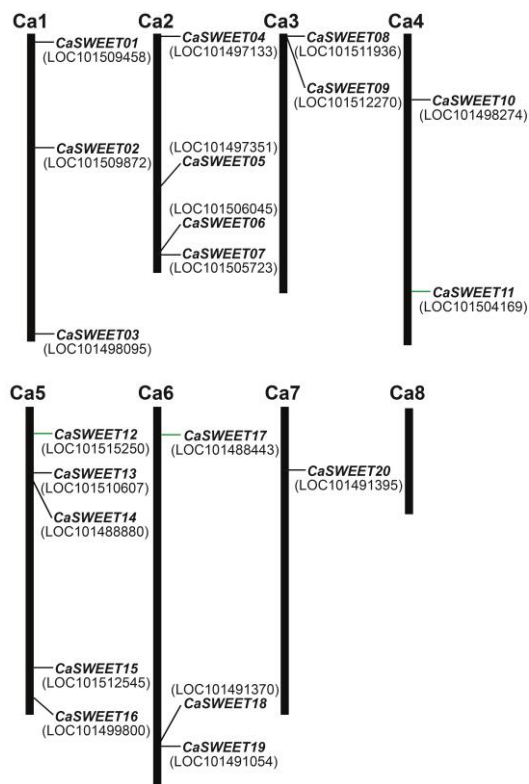
1). Nhiễm sắc thể Ca5 chứa nhiều gen *CaSWEET* nhất (5 gen) (Hình 1). Đáng chú ý, một số gen SWEET nằm ở vị trí cận đầu mút của nhiễm sắc thể (subtelomere) (Hình 1). Cụ thể bao gồm các gen *CaSWEET01* và *CaSWEET03* (Ca1), *CaSWEET04* (Ca2), *CaSWEET08* và *CaSWEET09* (Ca3) và *CaSWEET16* (Ca5) (Hình 1). Vị trí cận đầu mút của nhiễm sắc thể được cho là rất bảo thủ và đặc trưng cho loài, đóng vai trò quan trọng trong cơ chế nhận biết và bắt cặp trong nguyên phân sớm [4].

Tiếp theo, số lượng gen mã hóa SWEET ở đậu gà được so sánh với một số loài thực vật một và hai lá mầm khác (Bảng 2). Cụ thể, 52 gen *GmSWEET* đã được tìm thấy trên đậu tương [13], trong khi ở cải dầu và sắn đã xác định được lần lượt 68 và 28 gen mã hóa họ SWEET [1], [9] (Bảng 2). Trên đối tượng thực vật một lá mầm, 21 và 23 gen thuộc họ SWEET đã được nghiên cứu trên lúa gạo và cao lương [12], [16] (Bảng 2). Có thể thấy rằng, số lượng thành viên trong họ gen SWEET (21) ở đậu gà tương đương với lúa

gạo (21) và cao lương (23), mặc dù kích thước hệ gen của đậu gà (530,894 Mb) lớn hơn lúa gạo (382,627 Mb) nhưng nhỏ hơn cao lương (687,750 Mb) (Bảng 2). Điều này cho thấy không có sự tương quan giữa số lượng gen *SWEET* và kích thước hệ gen ở các loài thực vật. Tương tự, thành viên của họ gen *SWEET* không phụ thuộc vào số lượng nhiễm sắc thể cũng như giữa thực vật một lá mầm và hai lá mầm (Bảng 2).

Kết quả phân tích cấu trúc của họ gen mã hóa *SWEET* ở đậu gà

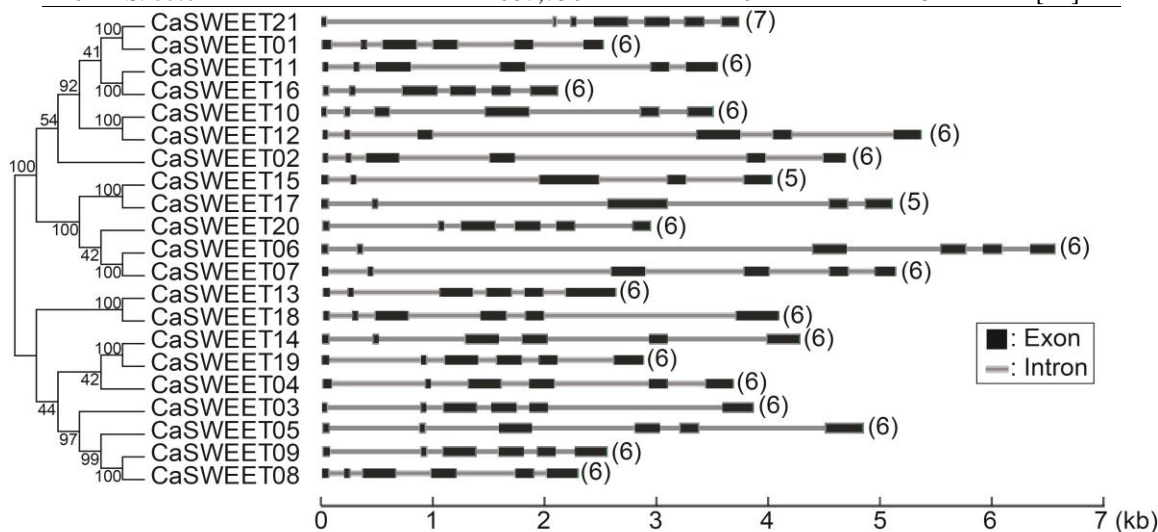
Phân tích trên PIECE [15] cho thấy họ gen *CaSWEET* đều có cấu trúc phân mảnh (Hình 2). Hầu hết các gen (18) mã hóa *SWEET* ở đậu gà rất bảo thủ, đều chứa sáu exon, chỉ có gen *CaSWEET21* chứa bảy exon và hai gen (*CaSWEET15* và *CaSWEET17*) chứa năm exon (Hình 2). Cấu trúc phân mảnh với kích thước intron rất dài (so với đoạn exon) cho thấy họ gen *CaSWEET* rất bền vững, có thể ít bị tác động do đột biến gen xảy ra trong quá trình chọn lọc tự nhiên.



Hình 1. Vị trí phân bố của họ gen *CaSWEET* ở đậu gà

Bảng 2. Tóm tắt họ gen *SWEET* ở một số loài thực vật

TT	Tên loài	Lá mầm	Hệ gen (Mb)	Nhiễm sắc thể	<i>SWEET</i>	Nguồn
1	<i>C. arietinum</i>	2	530,894	8	21	
2	<i>G. max</i>	2	984,880	20	52	[13]
3	<i>B. napus</i>	2	912,196	19	68	[9]
4	<i>M. esculenta</i>	2	390,836	18	28	[5]
5	<i>O. sativa</i>	1	382,627	12	21	[16]
6	<i>S. color</i>	1	687,750	10	23	[12]



Hình 2. Cấu trúc của họ gen *CaSWEET* ở đậu gà

Gần đây, hầu hết các thành viên thuộc họ *MeSWEET* ở sắn (26 trên 28 gen) cũng đều chứa sáu exon [1]. Số lượng exon này cũng đã được ghi nhận gần đây ở họ gen mã hóa *SWEET* ở cải dầu [9] và đậu tương [13]. Những kết quả này cho thấy cấu trúc của họ gen *SWEET* rất bảo thủ ở thực vật nói chung, trong đó hầu hết các gen đều chứa sáu exon. Kết quả của nghiên cứu này đã cung cấp những hiểu biết cơ bản về họ gen *SWEET* ở đậu gà. Trong nghiên cứu tiếp theo, đặc tính cơ bản của *SWEET* và đặc tính của gen *SWEET* liên quan đến tính chống chịu bất lợi sẽ được tìm hiểu và phân tích.

KẾT LUẬN

Đã xác định được 21 thành viên thuộc họ *SWEET* ở đậu gà. Đối chiếu trên hệ gen của đậu gà cho thấy số lượng gen *CaSWEET* có thể thay đổi trong các phiên bản cập nhật chú giải gen tiếp theo của các giống đậu gà.

Các gen *CaSWEET* phân bố không đồng đều trên hệ gen của đậu gà. Trong đó, không có mối tương quan nào được xác định giữa số lượng gen *SWEET* với kích thước hệ gen, số lượng nhiễm sắc thể giữa các loài thực vật một và hai lá mầm.

Họ gen *CaSWEET* ở đậu gà có cấu trúc phân mảnh với kích thước các đoạn intron dài để đảm bảo tính toàn vẹn của exon trong suốt quá trình tiến hóa. Đối chiếu với các loài khác cho thấy số lượng exon trong gen *SWEET* ở thực vật chủ yếu là sáu.

LỜI CẢM ƠN: Nghiên cứu này được thực hiện từ kinh phí của đề tài nghiên cứu cơ bản mã số 08/HĐUT-KHCN do Đại học Sư phạm Hà Nội 2 tài trợ.

TÀI LIỆU THAM KHẢO

1. 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), tr. 140 -149.
2. Acharjee S., Sarmah B. K. (2013), "Biotechnologically generating 'super chickpea'

for food and nutritional security", *Plant Sci.*, 207, pp. 108-116.

3. Aslam M., Mahmood I. A., Peoples M. B., Schwenke G. D., Herridge D. F. (2003), "Contribution of chickpea nitrogen fixation to increased wheat production and soil organic fertility in rain-fed cropping", *Biol. Fertil. Soils*, 38(1), pp. 59-64.

4. Calderón M. C., Rey M. D., Cabrera A., Prieto P. (2014), "The subtelomeric region is important for chromosome recognition and pairing during meiosis", *Sci. Rep.*, 4(6488), pp. 1-6.

5. Chen L. Q. (2014), "SWEET sugar transporters for phloem transport and pathogen nutrition", *New Phytol.*, 201(4), pp. 1150-1155.

6. El-Gebali S., Mistry J., Bateman A., Eddy S. R., Luciani A., Potter S. C., Qureshi M., Richardson L. J., Salazar G. A., Smart A., Sonnhammer E. L. L., Hirsh L., Paladin L., Piovesan D., Tosatto S. C. E., Finn R. D. (2018), "The Pfam protein families database in 2019", *Nucleic Acids Res.*, 47(D1), pp. D427-D432.

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

8. Jain M., Misra G., Patel R. K., Priya P., Jhanwar S., Khan A. W., Shah N., Singh V. K., Garg R., Jeena G., Yadav M., Kant C., Sharma P., Yadav G., Bhatia S., Tyagi A. K., Chattopadhyay D. (2013), "A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.)", *Plant J.*, 74(5), pp. 715-729.

9. Jian H., Lu K., Yang B., Wang T., Zhang L., Zhang A., Wang J., Liu L., Qu C., Li J. (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, pp. 1464.

10. Julius B. T., Leach K. A., Tran T. M., Mertz R. A., Braun D. M. (2017), "Sugar transporters in plants: New insights and discoveries", *Plant Cell Physiol*, 58(9), pp. 1442-1460.

11. Kumar S., Stecher G., Tamura K. (2016), "MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets", *Mol. Biol. Evol.*, 33(7), pp. 1870-1874.

12. Mizuno H., Kasuga S., Kawahigashi H. (2016), "The sorghum *SWEET* gene family: Stem sucrose accumulation as revealed through transcriptome profiling", *Biotechnol Biofuels*, 9, pp. 127.

13. Patil G., Valliyodan B., Deshmukh R., Prince S., Nicander B., Zhao M., Sonah H., Song L., Lin L., Chaudhary J., Liu Y., Joshi T., Xu D., Nguyen H. T. (2015), "Soybean (*Glycine max*) SWEET gene family: Insights through comparative genomics, transcriptome profiling and whole genome re-sequencing analysis", *BMC Genomics*, 16, pp. 520.
14. Varshney R. K., Song C., Saxena R. K., Azam S., Yu S., Sharpe A. G., Cannon S., Baek J., Rosen B. D., Taran B., Millan T., Zhang X., Ramsay L. D., Iwata A., Wang Y., Nelson W., Farmer A. D., Gaur P. M., Soderlund C., Penmetsa R. V., Xu C., Bharti A. K., He W., Winter P., Zhao S., Hane J. K., Carrasquilla-Garcia N., Condie J. A., Upadhyaya H. D., Luo M.-C., Thudi M., Gowda C. L. L., Singh N. P., Lichtenzveig J., Gali K. K., Rubio J., Nadarajan N., Dolezel J., Bansal K. C., Xu X., Edwards D., Zhang G., Kahl G., Gil J., Singh K. B., Jackson S. A., Wang J., Cook D. R. (2013), "Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement", *Nat. Biotechnol.*, 31(3), pp. 240-246.
15. Wang Y., You F. M., Lazo G. R., Luo M.-C., Thilmony R., Gordon S., Kianian S. F., Gu Y. Q. (2013), "PIECE: A database for plant gene structure comparison and evolution", *Nucleic Acids Res.*, 41(Database issue), pp. D1159-D1166.
16. Yuan M., Zhao J., Huang R., Li X., Xiao J., Wang S. (2014), "Rice MtN3/saliva/SWEET gene family: Evolution, expression profiling, and sugar transport", *J. Integr. Plant Biol.*, 56(6), pp. 559-570.