LẬP BẢN ĐỒ VÀ HỆ VÔ TÍNH CỦA QTL PROTEIN HẠT CHÍNH TRÊN CHROMOSOME 20 CÂY ĐẬU TƯƠNG

Christina E. Fliege[1], Russell A. Ward[1], Pamela Vogel[2], Hanh Nguyen[3], Truyen Quach[3], Ming Guo[2], João Paulo Gomes Viana[1], Lucas Borges dos Santos[1], James E. Specht[2], Tom E. Clemente[2], Matthew E. Hudson[1] and Brian W. Diers[1]

Võ Như Câm biên dịch

1. Khoa Khoa học Cây trồng, Đại học Illinois, 1101 W. Peabody Dr., Urbana, IL 61801, Hoa Kỳ.
2. Khoa Nông học và Làm vườn, Đại học Nebraska-Lincoln, Lincoln, NE 68583, Hoa Kỳ.
3. Trung tâm Đổi mới Khoa học Thực vật, Đại học Nebrasaka-Lincoln, Lincoln, NE 68583, Hoa Kỳ

TÓM TẮT

Đậu tương [*Glycine max* (**L.**) **Merr.**] là một loài cây trồng độc đáo vì nó có hàm lượng protein và dầu cao trong hạt của nó. Trong số nhiều locus tính trạng số lượng (QTL) kiểm soát hàm lượng protein hạt đậu tương, các alen của protein cqSeed-003 QTL trên nhiễm sắc thể 20 có tác dụng phụ lớn nhất. Loại alen có hàm lượng protein cao tồn tại trong mầm đậu tương trồng và đậu tương hoang (*Glycine soja* **Siebold & Zucc**.). Mục tiêu của chúng tôi là lập bản đồ tốt QTL này để cho phép nhân bản dựa trên vị trí của (các) gen gây bệnh cơ bản của nó. Việc lập bản đồ tốt đạt được bằng cách phát triển và thử nghiệm một loạt các quần thể trong đó vùng nhiễm sắc thể xung quanh các alen phân ly cao và protein thấp dần dần bị thu hẹp, sử dụng phương pháp phát hiện dựa trên dấu hiệu của các sự kiện tái tổ hợp. Khoảng 77,8kb kết quả được giải trình tự trực tiếp từ nguồn *G. soja* và so sánh với bộ gen tham chiếu để xác định các đa hình về cấu trúc và trình tự. Một biến thể chèn/xóa được phát hiện trong *Glyma.20G85100* được phát hiện có sự phù hợp +/− gần như hoàn hảo với kiểu gen alen protein cao/thấp được suy ra cho QTL này ở bố mẹ của các quần thể lập bản đồ đã công bố. Cấu trúc indel phù hợp với sự tiến hóa gần đây chèn một TIR vận chuyển vào gen thuộc dòng protein thấp. Protein hạt lớn hơn đáng kể trong đậu tương biểu hiện một yếu tố điều hòa downpin RNAi trong hai sự kiện độc lập liên quan đến các dòng phân ly không đối chứng. Chúng tôi kết luận rằng sự chèn transposon trong protein miền CCT được mã hóa bởi gen *Glyma.20G85100* chiếm các alen protein hạt cao / thấp của protein cqSeed-003 QTL.

GIỚI THIỆU

Đậu tương là một nguồn cung cấp protein và dầu quan trọng cho cả động vật và con người. Nó có hàm lượng protein hạt cao nhất (trung bình 400 g/kg, 40% trên cơ sở trọng lượng khô) khi so sánh với các loài cây họ đậu chính khác (khoảng trung bình 200–300 g/kg protein hạt), lần lượt vượt qua các loài cây ngũ cốc (trung bình khoảng 80–150 g/kg) (Liu, 1997). Hầu hết protein đậu tương trên thị trường là đồng sản phẩm của quá trình chiết xuất dầu đậu tương. Bột đậu tương này là một loại thức ăn chăn nuôi dễ tiêu hóa, cung cấp các axit amin thiết yếu cần thiết cho sự phát triển của động vật (Cromwell, 2012). Mặc dù được sử dụng như một nguồn cung cấp dầu và protein, giá trị thị trường đậu tương chủ yếu được thúc đẩy bởi phần khô đậu tương hơn là phần dầu. Năm 2018, 70% lượng tiêu thụ bột đạm trên thế giới là từ đậu tương, tổng cộng lên tới 235,4 triệu tấn (Hiệp hội Đậu tương Hoa Kỳ, 2019).

Hàm lượng protein hạt trong đậu tương được di truyền theo định lượng (Burton, 1985; Wilcox, 1985). Nhiều QTL kiểm soát protein của hạt hiện đã được xác định (Grant và cs, 2010). Trang web Soybase (Grant và cs, 2010) liệt kê 248 mối liên kết marker với hàm

lượng protein hạt mà kể từ đầu những năm 1990 đến nay (tháng 6 năm 2021) đã được phát hiện trong quần thể hai bố con. Những QTL này đã được lập bản đồ một cách lỏng lẻo đến nhiều vùng chồng chéo trên mỗi nhiễm sắc thể trong số 20 nhiễm sắc thể của hạt đậu tương, nhưng phần lớn có khả năng là phát hiện thừa của nhiều QTL ít thuần chủng hơn. Vì lý do này, các nhà lai tạo đã sử dụng từ viết tắt "cq" (QTL đã được xác nhận) để làm nổi bật những QTL có vị trí trên bản đồ đã được xác nhận bằng thực nghiệm và được Ủy ban Di truyền Đậu tương phê duyệt. Một QTL protein hạt trên nhiễm sắc thể 20 của đậu tương (trước đây là nhóm liên kết I), hiện được gọi là cqSeed Protein-003, là một trong những QTL protein được nghiên cứu rộng rãi nhất do tác dụng phụ lớn của nó đối với protein hạt. Khi QTL này được lập bản đồ trong một quần thể, QTL đối với dầu hạt (cqSeed Oil-004), năng suất hạt (cqSeed Yield-001) và khối lượng hạt (cqSeed weight-003) thường được xác định trong cùng một vùng di truyền có khả năng là do đa hướng (Nichols và cs, 2006). Thực vật đồng hợp tử về alen QTL protein cao đã được chứng minh là có hàm lượng protein tăng >20 g/kg và giảm lượng dầu khoảng 10 g/kg so với thực vật có alen thay thế (Diers và cs, 1992; Brummer và cs, 1997; Sebolt và cs, 2000; Csanádi và cs, 2001). Để nâng cao hiểu biết của chúng ta về khả năng kiểm soát di truyền cơ bản của protein cqSeed-003, lập bản đồ, tiếp theo là xác định gen ứng cử viên và nhân bản là cần thiết (Salvi và Tuberosa 2007).

Cả cqSeed protein-003 và cqSeed oil-004 lần đầu tiên được lập bản đồ với các marker RFLP bởi Diers và cs (1992). QTL này được phát hiện trong một quần thể có nguồn gốc từ việc lai dòng *Glycine max* A81-356022 với cây *Glycine soja* giới thiệu PI 468916. Dựa trên phân tích marker, alen từ *G. soja* có liên quan đến việc tăng protein 24 g/kg (Diers và cs, 1992). Sau khi QTL này được xác định, alen protein cao từ *G. soja* đã được xâm nhập vào các nền tảng di truyền khác nhau, điều này xác nhận rằng alen này có thể được sử dụng để tăng hàm lượng protein của hạt, nhưng nó cũng được phát hiện có liên quan đến năng suất thấp hơn (Sebolt và cs, 2000). Bản đồ của protein cqSeed-003 được khởi xướng bởi Nichols và cs (2006) trong một nghiên cứu thu hẹp vị trí bản đồ QTL thành vùng 3 cM. Bolon và cs (2010) sau đó đã sử dụng một tập hợp lớn các dấu lặp lại trình tự đơn giản (SSR) để thu hẹp khoảng QTL xuống vùng 8,4 Mbp giữa các dấu Sat_174 và ssrpqtl_38.

Cho đến nay, các QTL cqSeed protein-003 và cqSeed oil-004 đã được lập bản đồ trong nhiều phép lai giữa cha và mẹ (Brummer và cs, 1997; Chung và cs, 2003; Kim và cs, 2016; Lu và cs, 2012; Phansak và cs, 2016; Reinprecht và cs, 2006; Tajuddin và cs, 2003; Wang và cs, 2014; Warrington và cs, 2015). Các nghiên cứu lập bản đồ liên kết toàn bộ gen (GWAS) cũng được sử dụng để thu hẹp khoảng cặp cơ sở trong bản đồ QTL. Hwang và cs (2014) đã tiến hành GWAS sử dụng 42.368 nucleotide đơn đa hình (SNP) trong một tập hợp đa dạng về mặt di truyền gồm 298 dòng. Họ thu hẹp vùng gen ứng cử viên thành khoảng 2,4Mbp nằm ở 28,7–31,1Mbp (tập hợp Gmax2.0). Trong một GWAS do Vaughn và cs (2014), vùng gen ứng cử viên được định vị trong vùng khoảng 1Mbp giữa 32,1 và 33,1Mbp (tập hợp Gmax2.0) bằng cách sử dụng phần lớn các sự gia nhập nhóm trưởng thành (MG) V từ Hàn Quốc. Bandillo và cs (2015) đã sử dụng GWAS để phân tích 12.000 lượt gia nhập (tất cả MG) từ Bộ sưu tập mầm đậu tương USDA sử dụng 36.513 SNP và cung cấp bằng chứng cho thấy gen ứng cử viên nằm trong khoảng 2,4Mbp giữa 30,7 và 33,1Mbp (tập hợp Gmax2.0).

Để phát hiện QTL đậu tương có ảnh hưởng lớn đến protein hạt, Phansak và cs (2016) đã sử dụng chiến lược tạo kiểu gen chọn lọc đa quần thể bằng cách giao phối 48 con lai (từ MG 000 đến IV) với các cây trồng có năng suất cao của cùng MG và đánh giá thế hệ con

cháu $F_{2:3}$. Tất cả các giống cây trồng đều có hàm lượng protein cao (412–458 g/kg trên cơ sở 13% độ ẩm) và các giống cây trồng có hàm lượng protein bình thường (332–374 g/kg). Sau khi phân loại thế hệ con cái $F_{2:3}$ về hàm lượng protein hạt, chỉ các deciles trên và dưới được định kiểu gen bằng các markers SNP. Một QTL protein được phát hiện ở cùng vị trí với protein cqSeed-003 trong 27 trong số 48 lần giao phối, cho thấy rằng trong mầm protein cao, alen protein cao của QTL này chiếm ưu thế.

Trong nghiên cứu hiện tại của chúng tôi, việc lập bản đồ vi hơn nữa của protein cqSeed-003 đã được tiến hành với các dòng gần isogenic có nguồn gốc từ việc lai ngược alen protein *G. soja* với nền A81-356022. Vùng ứng cử viên bị thu hẹp đã được sắp xếp và lắp ráp theo trình tự bằng công nghệ Illumina. Sự đa hình trong các gen ứng cử viên từ khoảng thời gian đã được kiểm tra dựa trên một nhóm các kiểu gen đậu tương được biết là có hoặc không có alen protein cao trong khoảng protein cqSeed-003. Cuối cùng, tính đa hình chèn/xóa trong gen *Glyma.20G85100* mã hóa protein miền CCT được xác định là ứng cử viên có nhiều khả năng nhất. Vai trò của gen này trong việc kiểm soát hàm lượng protein trong hạt sau đó đã được khẳng định bằng sự biến nạp ổn định của các dòng đậu tương.

KẾT QUẢ

Lập bản đồ khoảng QTL

Vòng đầu tiên của bản đồ được thực hiện bằng cách sử dụng một tập hợp các quần thể BC_5F_7 được phát triển từ các cây BC_5F_6 đã được chọn để tái tổ hợp giữa các markers Satt239 và ssrpqtl_18 (Bảng 1). Nghiên cứu trước đây trong phòng thí nghiệm của chúng tôi chỉ ra rằng hai markers này có tên khác là cq-Seed protein-003 (Sebolt và cs, 2000; Nichols và cs, 2006). Các cây BC5F7 trong các quần thể này đã được kiểm tra với một markers tách biệt và được đánh giá trên thực địa về hàm lượng protein trong hạt vào năm 2008, sau đó là đánh giá các dòng $BC_5F_{7:8}$ vào năm 2009. Kết quả protein và markers được phân tích để kiểm tra mối liên hệ thống kê giữa protein và markers dữ liệu để xác định xem QTL nằm trong khoảng phân ly hay không phân ly trong mỗi quần thể. Ví dụ, một mối liên hệ đáng kể đã được tìm thấy trong quần thể 4 BC_5F_7, cho thấy rằng QTL là khoảng phân ly dưới ssrpqtl_17 (Bảng 1). Ngược lại, không tìm thấy mối liên quan nào trong quần thể 2, điều này cho thấy rằng QTL nằm trong khoảng không phân ly dưới ssrpqtl_17. Bằng cách kết hợp các kết quả trên 13 quần thể, chúng tôi kết luận rằng QTL nằm trong khoảng 5,5Mbp giữa ssrqtl_17 và ssrqtl_34 trên nhiễm sắc thể 20 (Gmax2.0) (Bảng 1, Bảng S1). Khoảng này mới lạ ở chỗ nó nằm ở phía xa của nơi mà ban đầu chúng tôi đưa ra giả thuyết về vị trí của QTL dựa trên dữ liệu sơ bộ của chúng tôi khi chúng tôi bắt đầu nghiên cứu.

Bảng 1. Danh sách 19 markers nhiễm sắc thể 20 của đậu tương (Bolon và cs, 2010), được sắp xếp theo vị trí cặp cơ sở (bp) của chúng (tập hợp Williams 82 - xem www.soybase.org), được sử dụng để đặc trưng cho 13 quần thể BC_5F_7 thực vật và dòng $BC_5F_{7:8}$ con cháu của chúng trong vòng đầu tiên của bản đồ (Hình S1). Mỗi quần thể trong số 13 quần thể BC_5F_7 được phát triển từ một cây BC_5F_6 riêng biệt và kiểu gen của mỗi cây bố mẹ này đối với các markers nhiễm sắc thể 20 được trình bày dưới đây. Nếu quần thể cây BC_5F_6 được phát triển từ đồng hợp tử với alen protein cao (*G. soja*) của bố mẹ cho (*G. soja*), mã kiểu gen B được sử dụng, A được sử dụng khi cây là đồng hợp tử với alen bố mẹ (*G. max*) và H. khi cây dị hợp tử. Xác suất được đưa ra để biết liệu sự phân ly marker trong mỗi quần thể có liên quan đáng kể với protein dựa trên các thử nghiệm thực địa hay không và một mũi tên được chỉ theo hướng, liên quan đến các phép

lai chéo, để biểu thị vị trí của QTL. Khu vực nơi các bài kiểm tra này cho thấy QTL được đặt được tô màu xám.

Tên marker	Vị trí Gmax1.01 bp	Vị trí Gmax1.01 bp	Số quần thể BC_5F_7												
			2	3	10	14	22	4	6	13	17	20	11	15	19
Satt239	24.129.682	25.275.083	H	H	H	H	H	A	A	B	A	A	B	B	A
															↓
Ssrpqtl_4	24.812.334	25.971.714	H	H	H	H	H	A	A	B	A	A	B	B	H
Ssrpqtl_8	25.751.901	26.920.157	H	H	H	H	H	A	A	B	A	A	B	B	H
Ssrpqtl_11	26.270.814	27.439.056	H	H	H	H	H	A	A	B	A	A	B	B	H
Ssrpqtl_13	26.444.803	27.606.228	H	H	H	H	H	A	A	B	A	A	B	B	H
														↓	
Ssrpqtl_14	26.538.403	27.699.841	H	H	H	H	H	A	A	B	A	A	B	H	H
Ssrpqtl_15	26.542.454	27.703.952	H	H	H	H	H	A	A	B	A	A	B	H	H
Ssrpqtl_16	26.609.299	27.770.740	H	H	H	H	H	A	A	B	A	A	H	H	H
Ssrpqtl_17	26.649.308	27.810.743	H	H	H	H	H	A	A	B	B	A	H	H	H
			↓	↓	↓	↓	↓	↓	↓	↓	↓	↓			
Ssrpqtl_18	26.958.336	28.124.804	B	A	B	A	B	H	H	H	H	H	H	H	H
Ssrpqtl_25	30.489.918	31.627.304	B	A	B	A	B	H	H	H	H	H	H	H	H
Ssrpqtl_29	31.787.239	32.934.791	B	A	B	A	B	H	H	H	H	H	H	H	H
Ssrpqtl_32	31.992.972	33.141.346	B	A	B	A	B	H	H	H	H	H	H	H	H
Ssrpqtl_33	32.022.042	33.170.565	B	A	B	A	B	H	H	H	H	H	H	H	H
Ssrpqtl_34	32.178.223	33.326.612	A	A	A	A	B	A	A	A	A	H	A	A	H
Ssrpqtl_35	32.216.450	33.359.151	A	A	A	A	B	A	A	A	A	H	A	A	H
Ssrpqtl_36	32.384.780	33.526.075	A	A	A	A	B	A	A	A	A	H	A	A	H
Ssrpqtl_37	32.717564	33.858.592	A	A	A	A	B	A	A	A	A	H	A	A	H
												↑			↑
Ssrpqtl_38	32.910.185	34.049.358	A	A	A	A	A	A	A	A	A	A	A	A	A
2008 Prob >F†			NS	NS	NS	NS	NS	*	**	*	**	*	**	**	**
2008 Prob >F			NS	NS	NS	NS	NS	**	**	-	-	**	-	-	-

† Mức độ quan trọng của kiểm tra liên kết chỉ thị trong từng quần thể dựa trên kiểm tra thực địa đối với từng quần thể cây BC_5F_7 trên thực địa năm 2008 và từng quần thể của dòng $BC_5F_{7:8}$ trên thực địa vào năm 2009. NS không có ý nghĩa, * Có ý nghĩa ở xác suất 0,05, ** Đáng kể ở xác suất 0,01 và dấu gạch ngang (-) không được kiểm tra.

Vòng thứ hai của phát triển quần thể và lập bản đồ đã được bắt đầu để thu hẹp khoảng QTL hơn nữa (Hình S1). Mười một quần thể đã được phát triển từ các cây BC_5F_8 có các

sự kiện tái tổ hợp trong khoảng QTL và các tập hợp con của các quần thể này đã được trồng trên đồng ruộng như cây BC_5F_9 vào năm 2011 và như dòng $BC_5F_{9:10}$ vào năm 2012 và dòng $BC_5F_{9:11}$ năm 2013 (Bảng 2). Như trong vòng đầu tiên, mỗi cây trong quần thể được định kiểu gen với một marker phân ly và hạt thu hoạch từ cây và dòng từ quần thể được phân tích hàm lượng protein để xác định quần thể nào đang phân ly về QTL (Bảng S2). Kết quả trên các quần thể, ngoại trừ một thử nghiệm, đã đặt protein cqSeed-003 trên nhiễm sắc thể 20 giữa BARCSOYSSR_20_0670 và BARCSOYSSR_20_0674 (Bảng 2). Điều này tương ứng với vùng 77,8 kb giữa 31 744 150 và 31 821 947 bp dựa trên hợp ngữ Wm82.a2.v1 (Gmax2.0) (SoyBase, https://soybase.org).

Bảng 2. Danh sách 17 marker nhiễm sắc thể 20 của đậu tương, được sắp xếp theo vị trí bp của chúng (tập hợp Williams 82 - xem www.soybase.org), được sử dụng để mô tả đặc điểm của 11 quần thể cây BC_5F_9 được trồng vào năm 2011 và thế hệ sau của chúng là dòng $BC_5F_{9:10}$ và $BC_5F_{9:11}$ được trồng lần lượt trong năm 2012 và 2013, trong vòng thứ hai của bản đồ. Mỗi quần thể trong số 11 quần thể BC_5F_9 được phát triển từ một cây BC_5F_8 riêng biệt và kiểu gen của mỗi cây bố mẹ này đối với các marker nhiễm sắc thể 20 được trình bày dưới đây. Nếu cây BC_5F_9 mà quần thể được phát triển là đồng hợp tử đối với alen protein cao (*G. soja*) của bố mẹ cho (*G. soja*), mã kiểu gen B được sử dụng, A được sử dụng khi cây đồng hợp tử với alen bố mẹ (*G. max*), và H khi cây dị hợp tử. Xác suất được đưa ra để biết liệu sự phân ly marker trong mỗi quần thể có liên quan đáng kể với protein dựa trên các thử nghiệm thực địa hay không và một mũi tên được chỉ theo hướng, liên quan đến các phép lai chéo, mà QTL nằm. Khu vực nơi các thử nghiệm này cho thấy QTL được đặt được tô màu xám. Các marker là của Bolon và cs (2010) và Song và cs (2010) và để ngắn gọn, tiền tố BARCSOYSSR đã bị loại bỏ khỏi tên của các marker từ Song và cs.

Tên marker	**Vị trí Gmax1.01 bp**	**Vị trí Gmax1.01 bp**	**Số quần thể BC_5F_7**										
			3	**5**	**7**	**9**	**10**	**15**	**26**	**27**	**28**	**31**	**344**
Ssrpqtl_17	26.649.365	Not avail.	H	H	H	H	H	H	A	A	B	A	A
													↓
20_0599	26.829.294	27.991.409	H	H	H	H	H	H	A	A	B	A	H
20_0616	27.811.875	28.974.676	H	H	H	H	H	H	A	A	B	A	H
												↓	
20_0617	27.877.620	29.040.539	H	H	H	H	H	H	A	A	B	H	H
20_0636	28.972.334	30.134.877	H	H	H	H	H	H	A	A	B	H	H
			↓										
20_0647	29.643.301	30.793.572	A	H	H	H	H	H	A	A	B	H	H
20_0650	29.758.405	30.909.346	A	H	H	H	H	H	A	A	B	H	H
20_0655	30.052.089	31.198.164	A	B	H	H	H	H	A	A	B	H	H
					↓								
20_0657	30.187.698	31.333.773	A	B	A	H	H	H	A	A	B	H	H

20_0667	30.489.863	31.627.414	A	B	A	H	H	H	A	A	B	H	H
											↓		
20_0668	30.517.621	31.655.159	A	B	A	H	H	H	A	A	H	H	H
20_0670	30.606.609	31.744.150	A	B	A	H	H	H	A	A	H	H	H
						↓				↑			
20_0674	30.684.404	31.821.947	A	B	A	B	H	H	A	H	H	H	H
							↑		↑				
20_0678	30.754.018	31.891.560	A	B	A	B	A	H	H	H	H	H	H
20_0715	32.030.122	33.178.717	A	B	A	B	A	H	H	H	H	H	H
								↑					
20_0718	32.178.222	33.326.648	A	B	A	B	A	A	H	H	H	H	H
ssrpqtl_34	32.178.303	Not avail.	A	B	A	B	A	A	H	H	H	H	H
2011 Prob >F†			NS	NS	NS	-	**	**	-	-	**	**	**
2012 Prob >F					NS	NS	**		NS	NS	**	-	-
2013 Prob >F					**	NS	-	-	-	NS	**	-	-

† Mức độ quan trọng của thử nghiệm liên kết chỉ thị dựa trên thử nghiệm thực địa đối với từng quần thể cây BC_5F_9 trên thực địa năm 2011 và từng quần thể của các dòng $BC_5F_{9:10}$ và $BC_5F_{9:11}$ trên thực địa vào năm 2012 và 2013, tương ứng. NS biểu thị là không quan trọng, * Có ý nghĩa với xác suất 0,05, ** Có ý nghĩa ở xác suất 0,01 và dấu gạch ngang (-) không được kiểm tra.

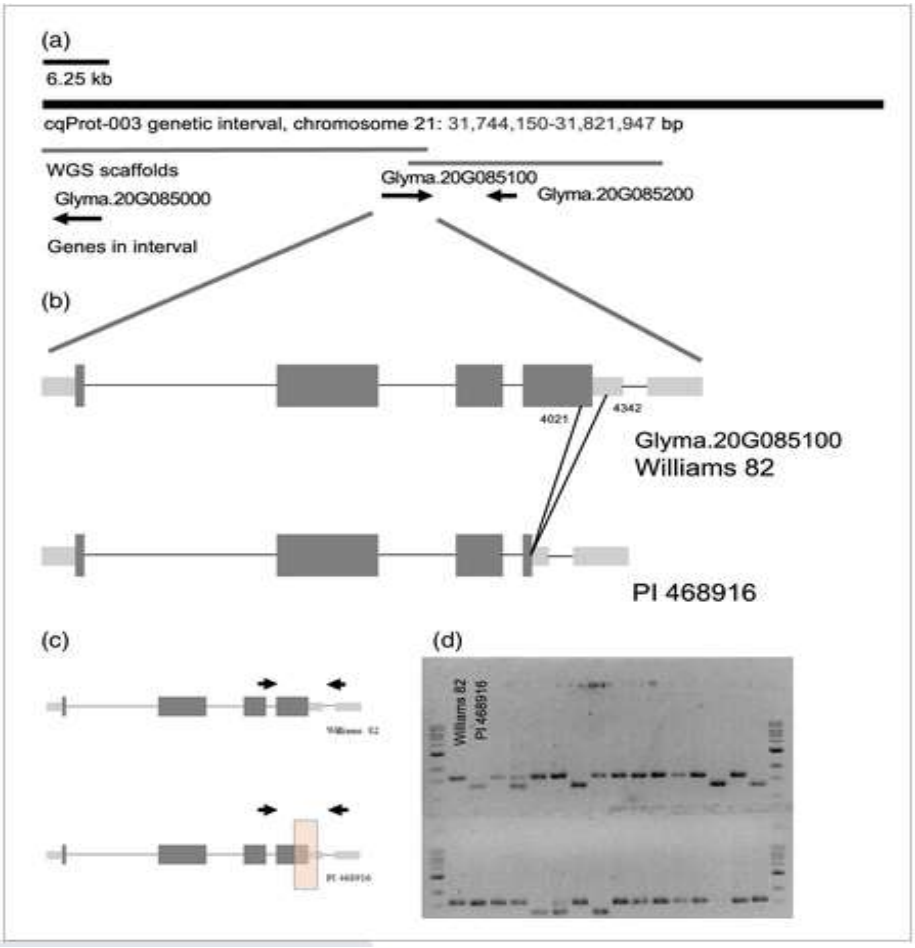
Các kết quả không nhất quán duy nhất là từ BC_5F_9 Quần thể 7 được đánh giá vào năm 2013 (Bảng 2 và Bảng S2). Mối liên kết protein-marker có ý nghĩa trong thử nghiệm này chỉ ra rằng QTL cao hơn marker 20_0657, điều này không phù hợp với 2 năm khác mà nó đã được thử nghiệm và các quần thể khác được đánh giá trong vòng này. Để giải quyết sự mâu thuẫn này, bảy quần thể con đã được phát triển từ Quần thể 7 có cùng điểm dừng tái tổ hợp như cây ban đầu được sử dụng để phát triển ProI-7 và tách biệt trong cùng một khoảng thời gian. Không tìm thấy mối liên kết marker đáng kể nào trong các thử nghiệm được thực hiện vào năm 2015 của bất kỳ quần thể nào trong số bảy quần thể con (Bảng S3), do đó bác bỏ sự mâu thuẫn ban đầu và hỗ trợ chắc chắn cho phát hiện rằng QTL thấp hơn BARCSOYSSR_20_0670.

Xác định các ứng cử viên đa hình trong khoảng

Ba gen ứng cử viên đã được xác định trong vùng 77,8kb dựa trên tập hợp bản đồ Gmax2.0 (SoyBase, https://soybase.org). Các gen này là: *Glyma.20g085000*, mã hóa một thành phần của phức hợp Golgi oligomeric; *Glyma.20g085100*, một gen mã hóa protein mô típ CCT; và *Glyma.20g085200*, mã hóa protein miền LIM liên kết ion kẽm. Để phân tích sâu vùng này, chúng tôi đã tận dụng toàn bộ hệ thống gen de novo của DNA PI 468916 được mô tả trước đó bởi Butler và cs (2021). Tập hợp trong khoảng bao gồm tất cả các gen được chú thích trong các đường viền lớn. Cùng với các lần đọc bổ sung và các sản phẩm phản ứng chuỗi polymerase (PCR), tập hợp được phân tích và so sánh với trình tự bộ gen tham chiếu Williams 82 bằng cách căn chỉnh và kiểm tra trực quan để xác định các đa hình về cấu trúc cũng như trình tự. Khi trình tự bộ gen liên kết của PI 468916 và

Williams 82 được so sánh với ba gen ứng cử viên, đột biến chèn - xóa lớn (indels) được xác định trong *Glyma.20g085100* và *Glyma.20g085200*, nhưng chỉ một đột biến đồng nghĩa ngoại lai được tìm thấy trong *Glyma.20g085000*.

Trong *Glyma.20g085100* (Hình 1a), phép so sánh cho thấy một indel 321-bp đã loại bỏ một exon được chú thích khỏi PI 468916 và đối với *Glyma.20g085200*, có một indel 8 kb dẫn đến việc loại bỏ gần như toàn bộ gen được chú thích trong PI 468916 (Hình 1b). Các marker kích thước dựa trên PCR được phát triển cho cả indels và các marker này được sử dụng để sàng lọc một bảng gồm 53 bố mẹ có các alen đã biết cho protein cqSeed-003 dựa trên kết quả lập bản đồ QTL trước đó (Bảng 3). Đối với sự mất đoạn 8 kb trong *Glyma.20g085200*, không có mối liên hệ nào giữa nó và sự hiện diện hay vắng mặt của alen protein cao trong bảng cha mẹ cho thấy rằng *Glyma.20g085200* có khả năng không phải là gen nhân quả nằm dưới protein cqSeed-003 (Bảng 3).



Hình 1. Đặc điểm phân tử của khoảng di truyền của khoảng protein cqSeed-003.
(a) Sơ đồ thể hiện vị trí của các giàn được lắp ráp trong khoảng và chú thích các gen. Vùng ở bên phải của khoảng chứa một trình tự lặp lại không mã hóa cho protein.
(b) Chế độ xem mở rộng của *Glyma.20G085100,* hiển thị phần chèn 321 bp tương ứng trong Williams 82 so với trình tự PI 468916.
(c) Sơ đồ cho thấy tổ hợp mồi được sử dụng để phát triển một marker đồng trội cho tính đa hình này.
(d) Gel agarose thể hiện kiểu gen của một phần quần thể cây đậu tương trong Bảng 3 bằng cách sử dụng marker.

Bảng 3. Danh sách các trường hợp tiếp cận nguồn mầm đậu tương đã được báo cáo là đồng hợp tử đối với alen protein cao (H) hoặc thấp (L) đối với QTL trong vùng QTL của nhiễm sắc thể 20 cqSeed protein-003. Đặc điểm này dựa trên việc sử dụng từng gia nhập với tư cách là cha mẹ trong quần thể hai cha con trong nghiên cứu lập bản đồ QTL (để biết chi tiết, xem báo cáo được trích dẫn trong cột Nguồn). Một xét nghiệm chẩn đoán marker dựa trên phản ứng chuỗi polymerase được sử dụng để xác định đặc điểm của từng gia nhập được liệt kê xem liệu nó sở hữu alen PI 468916 *Glycine soja* (+) hay alen Williams 82 *Glycine max* (-) ở mỗi hai gen *Glyma. 20* được suy ra trong nghiên cứu này là ứng cử viên tiềm năng cho QTL. Các kiểu đơn bội được xác định bằng cách sử dụng marker từ *Glyma.20G085100* và hai marker gần nhất nằm ở mỗi bên của gen. Các ký hiệu haplotype được xác định trong phần Quy trình thử nghiệm

Tên gia nhập	**Số gia nhập**	**Nguồn**	**Alen tại cqSeed protein-003**	***Glyma. 20g085100***	***Glyma. 20g085200***	**Haplot**
-	PI 468916	Diers và cs (1992)	H	+	+	5
-	PI 326582A	Chaky và cs (2003)	H	+	+	6
Kosodiguri Extra Early	FC 30867	P5	H	+	+	5
Akazu	PI 917254	P34	H	+	+	5
N-34	PI 153293	P6	H	+	+	5
V-4	PI 153296	P1	H	+	+	5
V-6	PI 153297	P14	H	+	-	5
V-14	PI 153301	P12	H	+	-	5
V-16	PI 153302	P9	H	+	+	5
No.51	PI 154196	P23	H	+	+	5
-	PI159764	P10	H	+	+	5
No.58	PI 181571	P20	H	+	+	2
Bitterhof	PI 189880	P13	L	-	+	1
Geant Vert	PI 189963	P2	H	+	-	5
Kariho-takiya	PI 243532	P36	H	+	+	2
No.17	PI253666A	P40	H	+	+	2
Wasedaizu No. 1	PI 261469	P19	H	+	-	2
-	PI 340011	P35	H	+	-	2
Oshimashirome	PI 360843	P39	L	-	-	1
Ronset 4	PI 372423	P4	H	+	-	5
KAERI.GNT 310-1	PI 398516	P33	L	-	-	1
KAS 33—9.1	PI 398704	P44	H	+	-	2
-	PI 398881	Dier và cs (chuẩn bị)	L	-	-	1
KLS 630-1	PI 398970	P45	L	-	-	2
Huaj an si er dian	PI 404188A	Dier và cs (chuẩn bị)	L	-	-	1
KAS 330-9.2	PI 407773B	P47	H	+	+	2
ORD 8113	PI 407788A	P41	H	+	-	2
-	PI 407823	P46	L	-	-	1
KAERI 511-11	PI 407877B	P43	(L)	(+)	-	2
KAS 640-7	PI 408138C	P37	H	+	-	2
Saikai 1	PI 423942	P28	H	+	-	2
Saikai 18	PI 423948A	P29	H	+	-	2
Saikai 20	PI 423949	P25	H	+	-	2
Shirome	PI 423954	P22	(L)	(+)	-	2
Shirome	PI 424148	P21	(L)	(+)	-	2
KAS 239-4	PI 424286	P42	L	-	-	2
Backchung No. 42	PI 427136	Dier và cs (chuẩn bị)	L	-	-	1
Choseng No. 1	PI 427138	P18	H	+	+	2
Seuhae No. 20	PI 427141	P26	H	+	+	2
DV-147	PI 437088A	P24	H	+	-	2

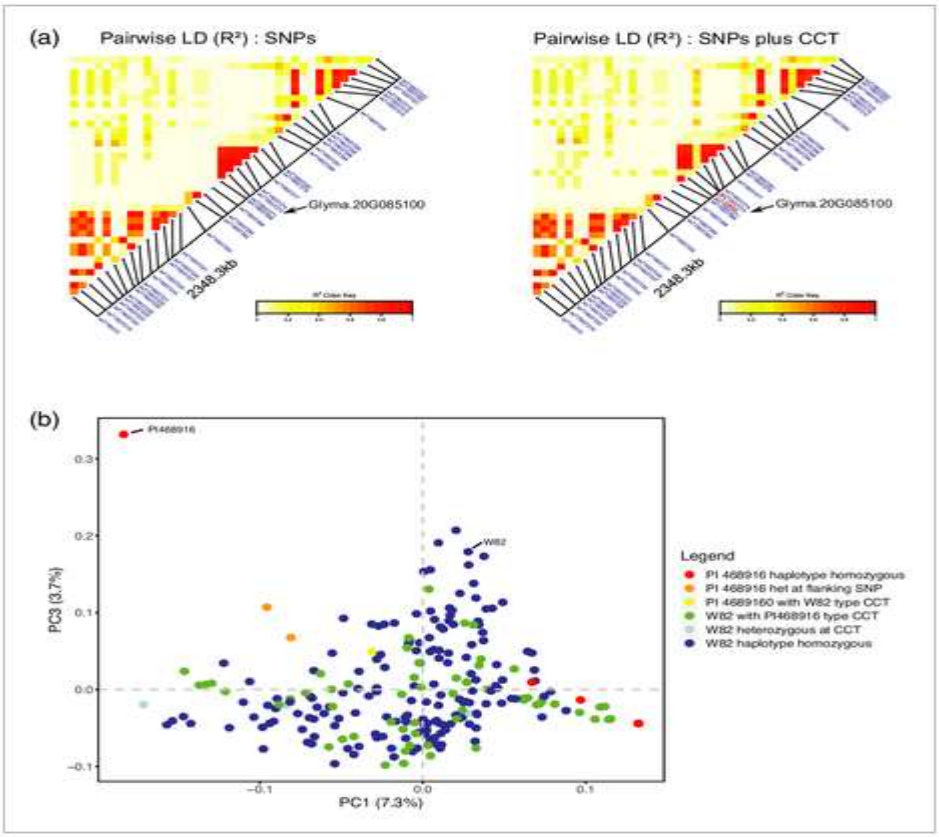
VIR 249	PI 437112A	P30	L	-	-	1
VNIISC-4	PI 437169B	Dier và cs (chuẩn bị)	L	-	-	1
Sjuj-dja-pyn-da-do	PI 437716A	P27	H	+	-	2
Ronest 4	PI 438415	P11	H	+	+	5
Szu yueh pa	PI 445845	P32	H	+	-	2
KAS 578-1	PI 458256	P48	L	-	+	1
NS-20	PI 518751	Dier và cs (chuẩn bị)	L	-	-	1
Provar	PI 548608	P31	L	-	-	1
Fen dou 14	PI 561370	Dier và cs (chuẩn bị)	L	-	-	1
Williams 82		Kim và cs (2016)	L	-	-	
A81-355012	A81-355012	Dier và cs (1992)	L	-	-	
Ina	PI 606749	Chưa công bố	L	-	-	
Danbaekkong	PI 619083	Warrington và cs (2015)	H	+	Chưa biết	

a: nghiên cứu trong đó một QTL quan trọng được lập bản đồ trong khoảng cqSeed protein-003 và kiểu gen là bố mẹ. Diers và cs (đang chuẩn bị) đề cập đến các kết quả chưa được công bố trong quần thể SoyNAM (Diers và cs, 2018), Chaky đề cập đến Chaky và cs (2003), Warrington đề cập đến Warrington và cs (2015), Kim đề cập đến Kim và cs (2016), và P theo sau là một số đề cập đến Phansak và cs (2016) với con số tương ứng với số quần thể trong nghiên cứu đó.

Ngược lại, mối liên hệ được tìm thấy giữa indel *Glyma.20g085100* và sự hiện diện hoặc vắng mặt của alen protein cao trong bảng điều khiển. Trong số 32 cặp bố mẹ đã được xác định là có alen protein cao tại cqSeed protein-003 dựa trên các nghiên cứu lập bản đồ trước đây (Bảng 3), tất cả đều có phiên bản ngắn hơn của đoạn PCR (alen PI 468916) tại *Glyma.20g085100*. Trong số 21 cặp bố mẹ đã được báo cáo có phiên bản protein cqSeed protein-003 thấp trong các nghiên cứu lập bản đồ (Bảng 3), 18 có đoạn PCR dài hơn (trình tự kiểu Williams 82), trong khi 3 có đoạn ngắn hơn (Hình 1c, d). Ba trường hợp ngoại lệ (PI 407877B, PI 423954 và PI 424148; Bảng 3) nằm trong số các cặp bố mẹ có hàm lượng protein cao được giao phối với bố mẹ có hàm lượng protein thấp trong nghiên cứu lập bản đồ QTL kiểu gen chọn lọc đa bố mẹ được thực hiện bởi Phansak và cs (2016). Lưu ý rằng những dòng này không được chỉ ra rõ ràng là có kiểu gen protein thấp, nhưng các tác giả đã không phát hiện ra sự phân li có ý nghĩa thống kê của một QTL protein ở nhiễm sắc thể 20 trong ba trường hợp ngoại lệ đó.

Sử dụng dữ liệu SNP 50-K từ khu vực, cùng với việc xác định kiểu gen PCR tại marker trong một bộ mẫu gồm các cách tiếp cận đa dạng (Bảng 3 và S4), chúng tôi đã thực hiện phân tích haplotype, mất cân bằng liên kết (LD) và các thành phần chính (PCA). Khi so sánh marker *Glyma.20g085100* với 4 marker bên cạnh gen của 238 điểm tiếp cận đơn bội, 167 người có alen Williams 82 cho tất cả 5 marker (haplotype 1), 53 là đồng hợp tử với alen Williams 82 đối với bốn marker SNP nhưng alen PI 468916 cho *Glyma.20g085100* (haplotype 2), ba đồng hợp tử với alen Williams 82 cho các marker bên và dị hợp tử cho *Glyma.20g085100* (haplotype 3), một đồng hợp tử với alen PI 468916 cho bốn marker hai bên và Williams 82 alen đối với *Glyma.20g085100* (haplotype 4), 13 đồng hợp tử PI 468916 cho tất cả năm marker (haplotype 5) và một đồng hợp tử PI 468916 cho tất cả trừ một marker bên sườn là dị hợp tử (haplotype 6). Số lượng gia nhập haplotype 2 cao cho thấy indel *Glyma.20g085100* có mức LD thấp bất ngờ với các marker bên sườn. Các marker bên sườn bổ sung đã được sử dụng trong phân tích LD, điều này cho thấy thêm rằng *Glyma.20g085100* không ở trạng thái cân bằng mạnh với SNP xung quanh (Hình 2a) hoặc với tính đa hình indel trong *Glyma.20g085200*, mặc dù gen này chỉ nằm ở vị trí 5160bp từ *Glyma .20g085100* trên bản đồ Gmax2.0 (SoyBase, https://soybase.org). Do sự thiếu tương quan này với các SNP rất gần kề mà thường được cho là có tương quan mạnh mẽ nếu không tương quan hoàn

hảo với bất kỳ tính đa hình nào trong gen, indel 321-bp liên kết mạnh mẽ hơn với đặc điểm cqSeed Protein-003 so với các đặc điểm đa hình xung quanh khác, giải thích các đỉnh liên kết không nhất quán được tìm thấy trong các nghiên cứu liên kết trước đây. Sự thiếu tương quan được giải thích là do thiếu mối liên hệ giữa sự hiện diện của indel và mức độ liên quan tổng thể của các dòng được nghiên cứu, theo đánh giá của PCA (Hình 2b).

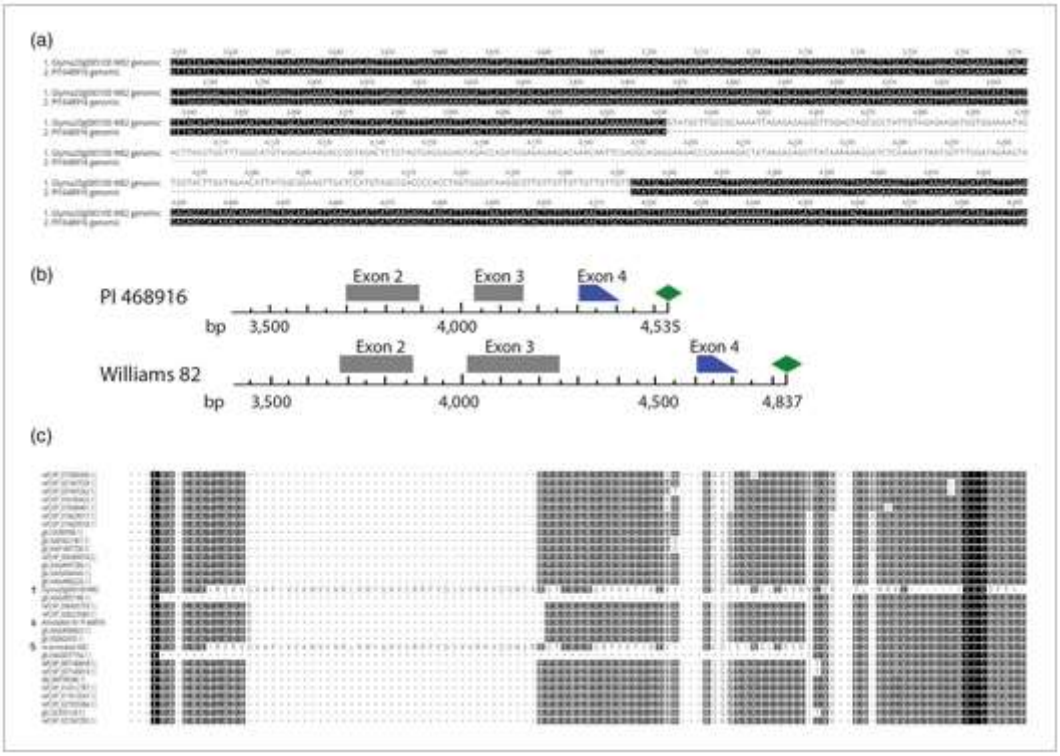


Hình 2. Sinh học quần thể của đa hình chèn/xóa (indel) tại *Glyma.20G085100*.

(a) Mất cân bằng liên kết (LD) xung quanh quỹ tích *Glyma.20G085100*. Bảng điều khiển bên trái: sử dụng tập hợp các gia nhập đa dạng (Bảng 3 và S4), các giá trị R2 LD xung quanh quỹ tích được tính toán bằng cách sử dụng dữ liệu từ mảng SoySNP50k (Song và cs, 2013); lưu ý rằng có một vùng LD bao quanh gen *Glyma.20G085100* được biểu thị bằng mũi tên (bảng điều khiển bên trái). Bảng bên phải: kiểu gen indel được xác định bằng marker dựa trên phản ứng chuỗi polymerase phát hiện tính đa hình chèn trong gen miền CCT (marker CCT) cũng được thêm vào. Rõ ràng là locus indel không có LD mạnh với các marker xung quanh. (b) Phân tích các thành phần chính của các phụ gia được nghiên cứu ở đây về hàm lượng protein. Các thành phần chính đầu tiên và thứ ba được tính toán từ thông tin marker mảng SoySNP50k toàn bộ bộ gen cho mỗi lần truy cập và được vẽ biểu đồ, và các điểm đại diện cho các phần tiếp cận được tô màu bởi kiểu gen tại bốn nucleotide đa hình (SNP) trong (a) trong LD với *Glyma.20G085100* locus cộng với marker indel *Glyma.20G085100*. Các phần tiếp cận với chuỗi PI 468916 trên quỹ tích được biểu thị bằng màu đỏ và chuỗi Williams 82 màu xanh lam. Các dòng cũng được phát hiện là đồng hợp tử hoặc dị hợp tử đối với phiên bản PI 468916 của indel trong *Glyma.20G085100*, nhưng mang các marker haplotype bên sườn giống hệt với Williams 82, được biểu thị bằng màu xanh lam nhạt hoặc xanh lục. Sự phân bố rộng rãi của các điểm màu xanh lá cây cho thấy khả năng đảo chiều bằng cách cắt bỏ đoạn chuyển vị.

Những kết quả này chỉ ra rằng *Glyma.20g085100* mã hóa protein CCT là locus có khả năng xảy ra nhất đối với protein cqSeed-003 và indel 321-bp do đó có khả năng là đa

hình dễ xảy ra nhất. Indel 321-bp xảy ra ở đầu Exon 3 (Hình 3a), làm cho vị trí mối nối được chú thích bị thay đổi và kích thước của exon thay đổi (Hình 3b). Tổng cộng, 73 axit amin đã bị thay đổi trong protein dự đoán vì indel này, với protein thấp, alen Williams 82 có 37 gốc bổ sung cũng như một số thay thế và loại bỏ trong vùng được bảo tồn của protein (Hình 3c). Bằng cách chú thích lại phiên bản Williams 82 của gen này, chúng tôi có thể tìm thấy một vùng mã hóa có khả năng khác của bộ gen, mã hóa cho vùng đầu c được bảo tồn của miền CCT, sau khi vùng của protein bị gián đoạn bởi 321-bp chèn được bảo tồn với các protein miền CCT khác nhưng không có trong chú thích hiện tại (a2.v1). Tuy nhiên, cả hai phiên bản của chú thích đều cho thấy phiên bản PI 468916 của gen mã cho một protein miền CCT được bảo tồn cao với các protein liên quan khác, nhưng việc chèn khiến phiên bản Williams 82 mã hóa cho một protein phân kỳ cao (Hình 3c).



Hình 3. Trình tự đa hình chèn/xóa.

Đa hình cảm ứng 321-bp trong gen *Glyma.20g085100* chịu trách nhiệm tạo ra các QTL cqSeed protein-003.

(a) Trình tự 321-bp có sự tương đồng về transposon có trong bộ gen Williams 82 chứ không phải bộ gen của sự gia nhập protein cao PI 468916.

(b) Tác động dự đoán của indel được hiển thị trong A lên cấu trúc intron – exon của bản sao *Glyma.20g085100*. Các exon mã hóa bên trong được hiển thị dưới dạng khối màu xám (Exon 1 giống hệt nhau và không được hiển thị), exon đầu cuối là khối màu xanh lam và tín hiệu polyadenyl hóa là một viên kim cương màu xanh lá cây. Các vị trí nằm trong các cặp bazơ từ vị trí bắt đầu phiên mã, như trong (A). Hộp màu cam hiển thị vùng chèn.

(c) Nhiều liên kết của các protein miền CCT liên quan. Trình tự protein đã xuất bản của *Glyma.20g085100* được hiển thị (†) cùng với chú thích của chúng tôi về trình tự từ PI 468916 (‡) và phiên bản được chú thích lại của trình tự Williams 82 bằng cách sử dụng các phương pháp tương tự như được sử dụng cho trình tự PI 468916 (§) liên kết với các protein liên quan trong Ngân hàng gen. Lưu ý rằng trình tự của PI 468916 được bảo tồn chặt chẽ với các protein liên quan.

Điều thú vị là, 321bp hiện diện trong alen Williams 82, nhưng không có trong alen PI 468916, cũng cho thấy sự tương đồng mạnh mẽ với các yếu tố có thể chuyển vị, bao gồm cả một transposon DNA loại TIR (Bảng S5). Phát hiện này chỉ ra rằng alen protein thấp được tìm thấy trong đậu tương *G. max* có thể có nguồn gốc tiến hóa gần đây và làm tăng khả năng trình tự có khả năng hoàn nguyên từ kiểu gen protein thấp thành allele *G. soja* protein cao bằng cách cắt bỏ mảnh vỡ. Bằng chứng về khả năng đảo ngược kiểu gen protein cao được đưa ra trong Hình 2b, trong đó một số trường hợp tiếp cận với alen protein cao, nhưng biểu hiện một loại đơn bội Williams 82 khác trong toàn vùng (haplotype 2), được quan sát thấy rải rác trong số protein thấp sự gia nhập.

Ba sự tiếp cận nguồn mầm, PI 407877B, PI 423954 và PI 424148, được xác định với phiên bản ngắn hơn (PI 468916) của gen *Glyma.20g085100*, nhưng trước đây được liên kết với phiên bản protein thấp của locus cqSeed protein-003. Vì điều này không phù hợp với cách giải thích rằng phiên bản ngắn của *Glyma.20g085100* chịu trách nhiệm về kiểu hình protein cao, chúng tôi đã đánh giá cách giải thích này trong một quần thể mới. Ba cách tiếp cận không nhất quán là từ bản đồ QTL của Phansak và cs (2016), bao gồm các quần thể tương đối nhỏ của các cây F_2. Do đó, một giải thích cho sự mâu thuẫn này là xác suất lỗi Loại II (không bác bỏ giả thuyết không có QTL) có thể khá cao trong các quần thể nhỏ này. Để đánh giá khả năng này, PI 423954 đã được lai với cây trồng Williams 82, được biết là có alen protein thấp. Một quần thể F2 được phát triển từ phép lai này đã được trồng trên đồng ruộng và được kiểm tra cùng với bố mẹ về khả năng xóa 321-bp trong *Glyma.20g085100* và được xét nghiệm tìm protein hạt bằng quang phổ phản xạ hồng ngoại gần (NIR). Người ta đã phát hiện ra mối liên hệ có ý nghĩa mạnh mẽ ($P < 0,01$) giữa sự đa hình về kích thước trong marker và hàm lượng protein, với các cá thể dị hợp tử có kiểu hình trung gian. Bằng chứng mới này cho thấy sự gia nhập này không có alen protein cao, chỉ ra sự thất bại bởi Phansak và cs (2016) để phát hiện nó có khả năng là do lỗi Loại II và cho thấy rằng điều này có thể xảy ra do không phát hiện được nó trong hai lần truy cập còn lại.

Thiết kế, xây dựng và thử nghiệm các sự kiện điều tiết giảm tải chuyển gen RNAi

Để xác nhận rằng gen ứng cử viên *Glyma.20G085100* điều chỉnh chức năng hàm lượng protein hạt bằng cách trùng hợp với tính đa hình gây bệnh trong protein cqSeed-003, các cây chuyển gen đã được tạo ra với nỗ lực loại bỏ phiên mã của gen bản địa. Do mức độ tương đồng về trình tự trong phần lớn vùng mã hóa protein của *Glyma.20G085100* và vùng tự nguyên của nó trên nhiễm sắc thể số 10 và độ tương đồng rất cao giữa vùng được chèn và nhiều phần khác của bộ gen, vùng 300-bp đã được chọn trong 5′ UTR của *Glyma.20G085100* cho cấu trúc RNAi, có một trình tự duy nhất có độ tương đồng thấp với các vùng khác của bộ gen đậu tương. Một cấu trúc vòng kẹp tóc được thiết kế để loại bỏ sự biểu hiện của gen *Glyma.20G085100*, được thúc đẩy bởi promoter CaMV 35S, đã được biến đổi thành giống đậu tương Thorne, là dòng protein thấp với alen loại Williams 82 ở *Glyma.20G085100*.

Sự kiện chuyển gen sơ cấp (T_0) cây trồng được thiết lập trong nhà kính và được phép tự thụ phấn. Các mô hình tích hợp trong các sự kiện độc lập đã được xác định thông qua phân tích Southern blot trên các cá thể T_1 được chọn (Hình S2). Quần thể cây T_2 có nguồn gốc từ hai sự kiện 1146-5 và 1157-1, chứa một và hai alen chuyển gen, tương ứng (Hình S2), được chọn lọc để xác định đặc điểm kiểu hình. Alen chuyển gen được theo dõi ở thế hệ T_2 bằng PCR và xét nghiệm khả năng chống chịu thuốc diệt cỏ để xác định ảnh hưởng của cấu trúc RNAi lên kiểu hình protein hạt trong điều kiện nhà kính. Mức độ sao chép tương đối của *Glyma.20G085100* trong các sự kiện được theo dõi trong mô lá (giai

đoạn phát triển V5) và trong phôi chưa trưởng thành (giai đoạn phát triển R5) (Hình S3). Việc theo dõi các cá thể T_2 và T_3 qua bảy sự kiện cho thấy mức giảm tương đối trong bảng điểm *Glyma.20G085100* ở cả hai giai đoạn V5 và R5 trong các sự kiện bao gồm 1146-5 và 1157-1 (Hình S3).

Hàm lượng protein của hạt giống T_3 thu hoạch từ các cây T_2 được tuốt riêng lẻ được đo bằng NIR. Tương đối ít các phân tách null được phát hiện, có thể là do gen chuyển có mặt trong quần thể ở nhiều bản sao không liên kết. Sự khác biệt về hàm lượng protein giữa hạt từ cây có và không có alen chuyển gen là có ý nghĩa thống kê (Bảng 4). Đối với một sự kiện, các cây biến đổi gen cho thấy lượng protein nhiều hơn trung bình > 3% so với các cây phân ly đối chứng. Do đó, một cấu trúc thấp nhất để giảm sự biểu hiện của phiên bản protein thấp của gen *Glyma.20G85100* có khả năng làm tăng hàm lượng protein của hạt. Những kết quả này hỗ trợ thêm cho kết luận của chúng tôi rằng biến thể chèn/xóa 321 bp trong gen mã hóa protein CCT *Glyma.20G85100* có khả năng là nguyên nhân của kiểu hình protein hạt và do đó tạo thành cơ sở gen phân tử của protein cqSeed-003.

Bảng 4. Mối liên hệ giữa gen vận chuyển RNAi kẹp tóc và tỷ lệ phần trăm protein hạt tính theo trọng lượng khô tính bằng g/kg đối với ba quần thể cây T_2 được trồng trong nhà kính

	Có gen vận chuyển		**Không có gen vận chuyển**		
Quần thể	**n[a]**	**Mean**	**n[a]**	**Mean**	**Pr>F*-**
1157-1-T1-3	18	417	5	399	0,05
1146-5-T1-4	25	444	5	415	0,04
1146-5-T1-5	20	441	5	409	0,02

a: số cây được thử nghiệm có hoặc không có gen vận chuyển.
** Mức độ quan trọng của thử nghiệm để phát hiện sự khác biệt giữa các nhóm hiện tại và không có mặt.*

THẢO LUẬN

Trong nghiên cứu này, chúng tôi đã thu hẹp vị trí của protein cqSeed-003 thành vùng 77,8 kb trên nhiễm sắc thể 20 nằm hai bên bởi các dấu hiệu di truyền BARCSOYSSR_20_0674 và BARCSOYSSR_20_0670. Vị trí vật lý của khoảng này là từ 31,74 đến 31,82Mbp dựa trên tổ hợp Gmax2.0, đây là khoảng hẹp nhất mà cqSeed protein-003 đã được lập bản đồ cho đến nay. Vùng thu hẹp này trùng với vùng gen ứng cử viên được xác định trong GWAS gần đây nhất của Bandillo và cs (2015), nơi họ lập bản đồ QTL với khoảng 2,4Mbp nằm trong khoảng từ 30,7 đến 33,1Mbp. Tuy nhiên, vùng được xác định trong nghiên cứu của chúng tôi không đồng nhất với các vùng được chỉ định cho QTL này trong hai nghiên cứu GWAS khác gần đây. Hwang và cs (2014) lập bản đồ QTL với khoảng 2,4Mbp ở 28,7–31,1, và Vaughn và cs (2014) lập bản đồ QTL tới vùng khoảng 1Mbp ở 32,1–33,1Mbp. Không có gì ngạc nhiên khi QTL này khó lập bản đồ trong các nghiên cứu liên kết vì *Glyma.20G85100* chèn có LD thấp với các marker xung quanh (Hình 2a).

Chúng tôi kết luận, dựa trên cả phân tích tương quan marker và nghiên cứu thực vật chuyển gen, rằng sự mất đoạn 321-bp trong alen PI 468916, so với alen tham chiếu Williams 82, trong *Glyma.20g085100* là vị trí phân tử gây bệnh cho protein cqSeed-003. Gen này mã hóa protein họ mô típ CCT (CONSTANS, CONSTANS-like, TOC1), một mô típ đã được chứng minh là có trong các gen kiểm soát sự ra hoa ở cây Arabidopsis (Masaki và cs, 2005), ở cây trồng (Zhang và cs, 2015), và nó đã được chứng minh là có

mặt trong protein kẽm loại GATA hoạt động như các mô-típ nhận dạng protein (Liew và cs, 2005). Mặc dù phiên bản Williams 82 của protein mã hóa miền CCT không bình thường, phiên bản PI 468916 của mã gen mã hóa miền CCT được bảo tồn cao (Hình 3c). Mặc dù liên quan chặt chẽ đến thời gian sinh học, nhưng protein miền CCT cũng có liên quan đến việc điều chỉnh một số quá trình khác trong cây trồng, bao gồm quang hợp, hiệu quả sử dụng chất dinh dưỡng và khả năng chống chịu căng thẳng (Liu và cs, 2020). Tuy nhiên, gen này đủ khác biệt so với trình tự chuẩn của mô típ CCT mà nó không được liệt kê trong một đánh giá gần đây về các protein vùng CCT của đậu tương (Mengarelli và Zanor, 2021). Vai trò có thể có của gen này đối với thời gian sinh học đang được quan tâm, vì vai trò có thể có trong thời điểm làm đầy hạt cũng như các quá trình khác liên quan đến sự tích tụ protein trong hạt đậu tương.

Có bằng chứng cho thấy *Glyma.20g085100* ảnh hưởng đến sự trưởng thành của thực vật ngoài protein và dầu. Người ta đã chỉ ra rằng các dòng đồng hợp tử đối với alen cho protein cao hơn cũng trưởng thành sớm hơn 0–5 ngày so với các dòng đồng hợp tử đối với alen thay thế (Sebolt và cs, 2000; Brzostowski và cs, 2017; Prenger và cs, 2019) nhưng không tài liệu đã được tìm thấy đánh giá ảnh hưởng của nó đối với ngày ra hoa. Ảnh hưởng của gen này lên protein và sự trưởng thành đã được tìm thấy trong các quần thể lập bản đồ từ MG II, được trồng ở miền bắc Hoa Kỳ (Sebolt và cs, 2000) đến MG VII, được trồng ở miền nam Hoa Kỳ (Prenger và cs, 2019). Ví dụ, Brzostowski và cs (2017) đã đánh giá hai quần thể MG IV phân ly về alen *Glyma.20g085100* và nhận thấy ở một quần thể có hiệu quả trưởng thành đáng kể trong 1 ngày giữa các dòng đồng hợp tử về hai alen khác nhau và ở quần thể thứ hai chênh lệch 3 ngày. Tác động của alen lên protein gần như giống hệt nhau giữa hai quần thể (22 và 23 g/kg). Những kết quả này cho thấy rằng gen tác động đến protein và dầu trên các vùng trồng trọt và nguồn gốc di truyền và những ảnh hưởng này không nhất thiết phải liên quan đến sự khác biệt lớn về thời gian trưởng thành. Bất kỳ tác động nào chủ yếu qua trung gian của quá trình trưởng thành sẽ được kỳ vọng sẽ cho thấy kiểu gen × ảnh hưởng môi trường lớn trên các vùng trồng khác nhau, khiến chúng tôi suy đoán rằng ảnh hưởng đến mức độ protein hạt của protein CCT được mã hóa tại *Glyma.20g085100* có thể liên quan đến một hệ thống điều tiết khác chẳng hạn như điều chỉnh và cảm nhận nồng độ đường (Masaki và cs, 2005). Nghiên cứu sâu hơn là cần thiết để nâng cao hiểu biết của chúng tôi về cách *Glyma.20g085100* ảnh hưởng đến sự trưởng thành của cây và điều này có liên quan như thế nào đến thành phần hạt giống.

Trình tự PI 468916 của *Glyma.20g085100* ngắn hơn trình tự từ Williams 82, có alen protein thấp hơn cho QTL (Kim và cs, 2016). Điều thú vị là việc xóa/chèn có LD thấp với các SNP và marker khác trong khu vực (Hình 2) và indel 321-bp cho thấy sự tương đồng với các chuyển vị DNA loại TIR (Bảng S5), cho thấy rằng lời giải thích có khả năng nhất là đột biến thực sự là một sự chèn vào transposon trong dòng dõi protein thấp. Để ủng hộ giả thuyết này, đậu tương hoang dã (*G. soja*) được sắp xếp theo trình tự và các họ hàng của chúng với các trình tự sẵn có tại thời điểm viết bài, được tìm thấy có phiên bản protein cao PI 468916 của gen này, cho thấy rằng protein thấp alen là một đột biến tương đối gần đây có thể đã được chọn trong các tế bào mầm ưu tú (*G. max*) trong quá khứ. Ngoài ra, phiên bản protein thấp dường như quay trở lại kiểu gen protein cao, bằng chứng là sự hiện diện của nhiều kiểu gen protein cao với các kiểu gen đơn bội có protein thấp được bảo tồn hoàn toàn (haplotype 2) trong số các kiểu tiếp cận protein thấp tương tự nhau về mặt di truyền (Hình 2b). Những chất hoàn nguyên này có thể được tạo ra bởi các sự kiện cắt bỏ. Cuối cùng, phiên bản protein cao của gen chứa miền CCT được bảo

tồn tốt cùng với nhiều gen liên quan khác (Hình 3c), cung cấp thêm bằng chứng rằng đó là phiên bản cũ hơn của trình tự.

Chúng tôi nhận thấy rằng bằng cách giảm sự biểu hiện của gen *Glyma.20g085100* bằng cách sử dụng RNAi, chúng tôi có thể tăng mức độ protein trong nền tảng di truyền Thorne (loại Williams 82, protein thấp). Kết quả này ngụ ý rằng phiên bản này của protein được mã hóa bởi alen với sự chèn 321bp có chức năng và làm giảm sự biểu hiện của nó bằng cách sử dụng RNAi, và do đó, có lẽ số lượng và hoạt động của sản phẩm protein của nó, làm tăng mức protein. Bất thường đối với đột biến tăng chức năng, alen này ảnh hưởng đến vùng mã hóa protein hơn là vùng điều hòa của gen. Transposon được biết là nguyên nhân gây ra xáo trộn exon (Quesneville, 2020), có thể đã góp phần vào việc tăng chức năng của protein được mã hóa tại vị trí này, thông qua sự phá vỡ một phần chính của miền CCT (Hình 3c). Chúng tôi kết luận rằng việc chèn có thể dẫn đến một chức năng mới gây ra sự phân bổ lại các nguồn lực từ protein hạt và thành carbohydrate và lipid tương tự với protein chính QTL khác trong đậu tương trên nhiễm sắc thể 15 đã được nhân bản gần đây (Wang và cs, 2020). Sự tăng lên của chức năng phân tử phù hợp với sự thay đổi và mở rộng đáng kể của trình tự protein được mã hóa bởi gen gây ra bởi sự chèn (Hình 3).

Kết luận, chúng tôi đã lập bản đồ cqSeed protein-003 thành vùng 77,8 kb trên nhiễm sắc thể 20 giữa các marker SSR BARCSOYSSR_20_0674 và BARCSOYSSR_20_0670, với mỗi vị trí tương ứng với các vị trí vật lý của 31 821 947 bp đến 31 744 150 bp, tương ứng, dựa trên bản đồ lắp ráp Gmax2.0. Vùng gen ứng cử viên bị thu hẹp này cho phép chúng tôi xác định alen nguyên nhân đối với QTL protein hiệu ứng lớn, được nghiên cứu rộng rãi nhất ở đậu tương. Bản thân alen này dường như là kết quả của sự chèn vào gen *Glyma.20g085100,* dẫn đến alen có hàm lượng protein thấp tại vị trí này, dường như đã hoàn nguyên thành alen có hàm lượng protein cao trong một số lần gia nhập. Gen mã hóa protein vùng CCT, mã hóa vùng CCT không bình thường trong alen Williams 82, có thể liên quan đến thời gian sinh học của quá trình hình thành và phát triển hạt giống.

Các nhà lai tạo đậu tương đã phải vật lộn với việc cân bằng giữa việc chọn giống để tăng hàm lượng protein trong hạt với các mối tương quan nghịch của nó với năng suất và dầu. Các nhà chọn giống chủ yếu tập trung vào việc tăng năng suất trong các chương trình chọn giống của họ và điều này đã dẫn đến việc giảm hàm lượng protein một cách tương xứng. Rincker và cs (2014) cho thấy trong hơn 80 năm chọn giống chủ yếu tập trung vào việc tăng năng suất mà hàm lượng protein giảm khoảng 2% hoặc 20 g/kg hạt. Zhang và cs (2020) gần đây đã báo cáo việc xác định tính đa hình trong gen GmSWEET39, gen này có liên quan chặt chẽ với hàm lượng protein và dầu trong hạt đậu tương và có khả năng là cơ sở của protein QTL quan trọng của nhiễm sắc thể 15. Việc xác định các gen kiểm soát hàm lượng protein và dầu của hạt có thể mở ra các phương pháp mới để sửa đổi các gen này theo những cách có thể làm tăng protein mà không làm giảm đáng kể năng suất và dầu.

Quy trình thử nghiệm

Lập bản đồ

Vị trí cqSeed protein-003 đã được lập bản đồ thông qua một quá trình lặp đi lặp lại phát triển và kiểm tra các quần thể lai tạo tách biệt cho các phần khác nhau của khu vực mà nó lập bản đồ. PI 468916 là nguồn cung cấp alen protein cao và alen này được lai ngược với nền A81-356022, một dòng thí nghiệm của Đại học Bang Iowa. Các quần thể được sử

dụng trong nghiên cứu hiện tại được phát triển từ các dòng lai ngược được mô tả bởi Nichols và cs (2006).

Dưới đây là phần giải thích ngắn gọn về sự phát triển và thử nghiệm các tế bào mầm được sử dụng trong lập bản đồ. Xem Phương pháp S1 để biết giải thích chi tiết về quy trình và Hình S1 để biết phác thảo hàng năm về các bước được thực hiện trong việc lập bản đồ. Việc lập bản đồ được thực hiện trong hai vòng với các nỗ lực của vòng thứ hai tập trung vào khoảng QTL được xác định trong vòng 1. Qua hai vòng, 7603 cây trồng phân ly QTL trong các quần thể lai ngược đã được sàng lọc bằng các marker ở hai bên khoảng thời gian mà bản đồ QTL xác định cây có giao nhau giữa các marker. Các cây lai được chọn sau đó được thử nghiệm với tất cả các marker sẵn có (Bolon và cs, 2010; Song và cs, 2010) trong khoảng QTL để lập bản đồ vị trí của các cây lai. Các cây có điểm giao chéo được lập bản đồ đến các vị trí duy nhất đã được chọn và một quần thể cây hoặc dòng sau đó được phát triển từ mỗi cây được chọn và trồng trên thực địa và được kiểm tra bằng marker phân tách bằng các phương pháp của Cregan và Quigley (1997), Keim và Shoemaker (1988)), Wang và cs (2003). Hạt giống thu hoạch từ cây và dòng được kiểm tra hàm lượng protein và hàm lượng dầu bằng cách sử dụng phương pháp truyền hồng ngoại gần. Kết quả marker và thành phần từ mỗi quần thể sau đó được phân tích bằng cách sử dụng hàm PROC GLM của SAS (SAS, 2016). Việc phân tích từng quần thể này dẫn đến việc thu hẹp khoảng QTL vì một mối liên kết có ý nghĩa chỉ ra rằng QTL nằm trong khoảng cách ly trong quần thể mà không có ý nghĩa chỉ ra rằng nó không nằm trong khoảng cách ly.

Đánh giá gen ứng cử viên

Để kiểm tra các gen ứng cử viên trong khoảng cqSeed protein-003 được lập bản đồ, một bảng gồm 53 kiểu gen đã được tập hợp. Những kiểu gen này đã từng là bố mẹ có hàm lượng protein cao hoặc thấp của nhiều quần thể bố con đã được sử dụng trong các nghiên cứu lập bản đồ QTL protein hạt và dầu, trong đó suy ra sự phân li alen của protein cqSeed-003 QTL (Bảng 3). Trong số 53 bố mẹ của quần thể lập bản đồ, 32 con được ghi nhận là đồng hợp tử về alen protein cao trong khoảng cqSeed protein-003, trong khi có 21 đồng hợp tử với alen protein thấp. DNA được chiết xuất từ các bố mẹ này và thử nghiệm với các marker được phát triển từ các đa hình trong *Glyma.20G85100* và *Glyma.20G85200*, hai gen ứng cử viên trong khoảng mà protein cqSeed-003 được lập bản đồ. Đối với *Glyma.20G85100*, một marker cộng hưởng được thiết kế để tạo ra các dải có sự khác biệt về kích thước 321 bp giữa hai alen bằng cách sử dụng trình tự mồi ACTGCATCAACCAAGCCTTATGC và TGTACGTTTCTAACTCACTTAACTTATTGG. Để có sự khác biệt về kích thước lớn hơn nhiều trong *Glyma.20G85200*, cần phải sử dụng hai marker chi phối riêng lẻ. Trình tự mồi được sử dụng cho biến thể alen dài hơn là CATGGGTAGTTTCTGAAAGCA và CGAGTCTTTCAAAGCATACCA và đối với biến thể ngắn hơn, trình tự mồi là TAGTGTCTACTGTACGTAAGTT và CGATATCCAAGTGAACGC. Cùng với nhau, dữ liệu được thu thập từ biến thể dài hơn và ngắn hơn đã cho kết quả marker đồng ưu thế.

Phân tích marker gen ứng cử viên được thực hiện bằng cách sử dụng giao thức PCR 20µl chứa 1µl mồi thuận và nghịch 5µm, cùng với các thành phần từ bộ TaKaRa DNA polymerase (TBSUSA, Mountain View, CA, USA): 0,1µl TaKaRa ExTaq DNA polymerase, 1,6µl dNTP, và 2µl 10 × đệm với 1µl DNA khuôn mẫu được pha loãng năm lần. Các amplicon PCR được hình dung trên gel điện di agarose 1% chạy ở 100 V trong 30 phút (Bio-Rad Hercules, CA, USA).

Phân tích và lắp ráp trình tự *De novo*

Việc giải trình tự toàn bộ bộ gen của sự gia nhập protein cao PI 468916 được thực hiện bằng cách sử dụng bộ trình tự DNA HiSeq v.1 (Illumina, San Diego, CA, Hoa Kỳ) tại Trung tâm Công nghệ Sinh học Carver tại Đại học Illinois. DNA được tách chiết bằng phương pháp mô tả ở trên. Các lần đọc thô được lắp ráp bằng ABySS (Simpson và cs, 2009) để tạo ra các giàn giáo. Tổng cộng, 182.255.190 lần đọc kết thúc được ghép nối đã được lắp ráp, mỗi đoạn dài 150 nucleotide, với kích thước đoạn gDNA trung bình là 580 nucleotide. K-mer trong số 64 được chọn là hiệu suất tốt nhất (ABySS được biên dịch cho kích thước k-mer tối đa là 64). Trình tự tham chiếu Williams 82 (a2.v1) tương ứng với vùng giữa các marker bên sườn, được so sánh với đầu ra khung từ tập hợp các trình tự PI 468916 bằng cách sử dụng chương trình BLAST để xác định các khung tiềm năng tương ứng với khoảng QTL được nhắm mục tiêu. Sau đó, các khung này được tải vào phần mềm phân tích DNA phổ biến 11.1.5 (Geneious, Auckland, New Zealand). Các trình tự từ PI 468916 được lập bản đồ đến tham chiếu Williams 82 và kiểm tra sự biến đổi nucleotide đơn và những thay đổi cấu trúc trong các vùng gen và được căn chỉnh với tham chiếu Williams 82. Tất cả sự khác biệt đã được ghi nhận và nghiên cứu về chức năng tiềm ẩn, mặc dù chỉ có ba đa hình được tìm thấy ảnh hưởng đến các gen mã hóa protein trong khoảng QTL cạnh marker cuối cùng.

Phân tích LD

LD được tính toán dưới dạng giá trị R2 theo từng cặp từ các marker SNP ở hai bên của marker *Glyma.20G85100* cho nhóm bố mẹ được mô tả ở trên bằng cách sử dụng di truyền gói R (Warnes, 2003). Dữ liệu SNP cho bảng được tạo bằng cách sử dụng mảng SoySNP50k để định kiểu gen cho bộ sưu tập mầm đậu tương (Song và cs, 2013) (dữ liệu có tại https://soybase.org/snps/). Các giá trị R2 này được hình dung dưới dạng bản đồ nhiệt và được vẽ biểu đồ để hiển thị mối quan hệ của từng SNP với từng SNP khác bằng cách sử dụng Sơ đồ LDheatmap của gói R.

Phân tích PCA và haplotype

Phân tích PCA được thực hiện và trực quan hóa bằng cách sử dụng gói SNPrelate trong Bioconductor cho R (Zheng và cs, 2012). Dữ liệu SNP cho số lượng truy cập được tạo từ dữ liệu SNP của SoySNP50k (Song và cs, 2013) (dữ liệu có sẵn tại https://soybase.org/snps/). Chi tiết về các phép truy cập bao gồm kiểu gen của SNP được trích xuất và marker indel được đưa ra trong Bảng 2 và S4.

Các kiểu gen Haplotype được xác định cho khoảng *Glyma.20G85100* bằng cách sử dụng marker đồng trội cho gen này và hai marker gần nhất (ss715637268 và ss715637271) ở một bên của indel trong gen và hai ở phía bên kia (ss715637273 và ss715637274). Các kiểu đơn bội được xác định cho năm marker này cho các loại đậu tương được liệt kê trong Bảng 3 và S4 và giống với các kiểu trong Hình 2b, được mã hóa như sau cho các bảng: 1 = Giống Williams 82 trên tất cả năm marker; 2 = Giống Williams 82 cho bốn marker SNP và giống PI 468916 cho marker *Glyma.20G085100*; 3 = Giống Williams 82 đối với bốn marker SNP và dị hợp tử đối với marker *Glyma.20G085100*; 4 = PI 468916 giống cho bốn marker SNP khác Giống như Williams đối với marker *Glyma.20G085100*; 5 = PI 468916 giống cho tất cả năm marker; và 6 = PI 468916 giống cho tất cả các marker ngoại trừ dị hợp tử đối với ss715637268.

Cấu trúc vector RNAi và chuyển đổi đậu tương

Một phần tử kẹp tóc được lắp ráp bằng trình tự 299 bp từ phiên bản bộ gen Williams 82 a2.v1 bắt đầu ở vị trí 31 774 770 và kết thúc ở 31775069 trên Chr 20 và đại diện cho hầu hết 5′ UTR của *Glyma.20G85100*. Phần tử được tổng hợp (GenScript Hoa Kỳ, Piscataway, NJ, Hoa Kỳ), với sự kết hợp của các vị trí *Spe*I và *Bgl*II ở các đầu, để cho phép các phần tử đảo ngược được ép dòng thành vectơ pUC58-RNAi (quà tặng từ Đại học H. Cerutti của Nebraska) (Hình S4). Kết quả là các phần tử đảo ngược được tách ra bởi intron thứ hai của một protein nucleolar nhỏ từ cây Arabidopsis (Tại 004G02840) để tạo điều kiện cho việc gấp kẹp tóc (Wesley và cs, 2001). Phần tử kết quả được phụ dòng giữa promoter 35S CaMV và trình kết thúc T35S từ virus khảm súp lơ. Băng biểu hiện RNAi tiếp theo sau đó được nhân bản thành một vectơ nhị phân chứa một băng chọn lọc gen thanh và vectơ nhị phân cuối cùng để biến nạp đậu tương được chỉ định là pPTN1379 (Thompson và cs, 1987) (Hình S4). Vectơ nhị phân pPTN1379 được huy động vào chủng vi khuẩn *Agrobacterium tumefaciens* EHA101 (Hood và cs, 1986) bằng cách giao phối ba bên và chất chuyển hóa có nguồn gốc được sử dụng để biến đổi giống đậu tương Thorne (McBlain và cs, 1993) theo quy trình đã được truyền đạt trước đó (Zhang và cs, 1999).

Phân tích Southern blot

Phân tích Southern blot đã được thực hiện như đã mô tả trước đây (Eckert và cs, 2006). Mười microgam tổng số DNA bộ gen được tiêu hóa giới hạn bằng EcoRI và được tách trên gel agarose 0,8%. Các DNA đã tách được chuyển sang màng nylon và lai với vùng 5′ UTR đánh dấu dCT 32P được sử dụng trong thiết kế kẹp tóc.

Vật liệu thực vật chuyển gen và tách chiết DNA để định kiểu gen PCR

Hạt giống chuyển gen được khử trùng bề mặt trong chất tẩy trắng 10% và được gieo trong giấy nảy mầm. Cây giống sạch bệnh đã nảy mầm được cấy vào bầu đất có đường kính 30cm trong nhà kính. Cây được trồng với độ dài 14,75 giờ ngày với nhiệt độ ban ngày dao động từ 28–30^{0}C và ban đêm từ 23–26^{0}C. Tách chiết DNA được thực hiện trên mô lá non ba lá bằng cách sử dụng một quy trình sửa đổi từ Keim và Shoemaker (1988) .

Sự phát triển của cây chuyển gen, phân tích kiểu gen và protein

Ba quần thể hạt giống T_2 từ hai sự kiện biến nạp T_0 đã được phát triển và trồng trong nhà kính. Hai quần thể T_2 (T_{1-4} và T_{1-5}) được phát triển từ sự kiện T_1 1146–5 và quần thể khác (T_{1-3}) là từ sự kiện 1157–1. Các dòng T_2 này được nuôi trong nhà kính với các điều kiện được mô tả ở trên và DNA được tách chiết theo Keim và Shoemaker (1988). Những cây này được định kiểu gen để xác định sự có mặt của alen chuyển gen ở mỗi cây. Sự hiện diện của alen chuyển gen được kiểm tra bằng cách sử dụng marker PCR được thiết kế với các đoạn mồi bên trong cấu trúc gen chuyển. Các đoạn mồi có trình tự 5′-CGAGGAGGTTTCCGGATATTAC-3′ và 5′-GCACGACACACTTGTCTACT-3′; Điều kiện PCR là ủ ban đầu 1 phút ở 95^{0}C, sau đó là 30 chu kỳ 30 giây ở 95^{0}C, 30 giây ở 52^{0}C, 1 phút ở 72^{0}C, tiếp theo là 65 phút ở 72^{0}C kéo dài cuối cùng. Marker đã phát hiện ra sự có mặt của gen chuyển và đưa ra kiểu hình trội và do đó không thể phân biệt được gen có ở trạng thái đồng hợp tử hay dị hợp tử hay không. Hạt giống từ mỗi cây được đập riêng lẻ, hàm lượng protein và dầu được đo bằng NIR (Perten DA 7250, Stockholm, Thụy Điển). Những hạt có nấm mốc nhìn thấy được, bắt đầu nảy mầm hoặc bị hỏng sẽ bị loại khỏi phép đo. Hạt được đóng gói lại và hàm lượng protein được đo lần thứ hai và tính trung bình.

Phân tích marker có thể lựa chọn của cây chuyển gen

Sự phân ly của alen chuyển gen trong thế hệ con cháu của dòng T_2 cũng được theo dõi bằng cách sử dụng thử nghiệm chống chịu thuốc diệt cỏ (Zhang và cs, 1999). Để lựa chọn tính kháng thuốc diệt cỏ, pha loãng 100 lần của Finale, một công thức thương mại của glufosinate đã được sử dụng (BASF, Ludwigshafen, Đức). 0,25% (v/v) chất hoạt động bề mặt không ion (Activator 90; Loveland Products Inc., Loveland, CO, Hoa Kỳ) đã được thêm vào để tăng cường sự thấm ướt của lá. Những chiếc lá có ba lớp đầu tiên được quét thuốc diệt cỏ trên khắp các lá. Các dải lá bị úa và hoại tử được ghi nhận sau 5 ngày.

Phân tích phiên mã ngược-PCR các sự kiện RNAi để theo dõi những thay đổi biểu hiện trong *Glyma.20G85100.*

Là một phương tiện để đánh giá sự thay đổi phiên mã tương đối giữa các sự kiện RNAi và cây Thorne kiểu hoang dã, RNA tổng số được phân lập từ mô sinh dưỡng ở giai đoạn phát triển V5 và phôi chưa trưởng thành ở giai đoạn phát triển R5, trong điều kiện thực địa. RNA tổng số được chiết xuất bằng RNeasy Plant Mini Kit (QIAGEN Corp. cat. No./ID: 74904, Hilden, Đức) theo hướng dẫn của nhà sản xuất. Hai microgam RNA tổng số được sử dụng để tổng hợp cDNA với Bộ phiên mã ngược cDNA dung lượng cao (Thermo Fisher Scientific, Waltham, MA, Hoa Kỳ). PCR được thực hiện với chu kỳ nhiệt 94^0C/1 phút và 30 chu kỳ 94^0C/30 giây, 55^0C/30 giây và 72^0C/30 giây, sau đó thêm 72^0C/5 phút để kéo dài.

Mồi Gm20P_osystem (5′-TCCACCACTTCTCCCAATCTCAAC-3′) và Gm20P_reverse (5′-CCCGTCAAAATTGAACCTGCTG-3′) để khuếch đại các đoạn cụ thể Gm20P và Gm-Actin6_L2 (5′-TGGTGTGTGATGGTTGTTGtin6-GAGG-3′) và Gm-Actin6_R2 (5′-GGGTTAAGAGGGGCCTCAGT-3′) để khuếch đại các đoạn *GmActin6* làm điều khiển bằng cách sử dụng GoTaq® Master Mixes (Promega Corporation, Madison, WI, Hoa Kỳ).

TÀI LIỆU THAM KHẢO

- American Soybean Association (2019) 2019 SoyStats Available at https://soygrowers.com/wp-content/uploads/2019/10/Soy-Stats-2019_FNL-Web.pdf (Verified 1 June 2021).
  Google Scholar
- Bandillo, N., Jarquin, D., Song, Q., Nelson, R., Cregan, P., Specht, J. et al. (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome, 8, 1– 13.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Bolon, Y.T., Joseph, B., Cannon, S.B., Graham, M.A., Diers, B.W., Farmer, A.D. et al. (2010) Complementary genetic and genomic approaches help characterize the linkage group I seed protein QTL. BMC Plant Biology, 10, 41.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Brummer, E.C., Graef, G.L., Orf, J., Wilcox, J.R. & Shoemaker, R.C. (1997) Mapping QTL for seed protein and oil content in eight soybean populations. Crop Science, 37, 370– 378.
  Wiley Online LibraryWeb of Science®Google Scholar
- Brzostowski, L.F., Pruski, T.I., Specht, J.E. & Diers, B.W. (2017) Impact of seed protein alleles from three soybean sources on seed composition and agronomic traits. Theoretical and Applied Genetics, 130, 2315– 2326.
  CrossrefCASPubMedWeb of Science®Google Scholar

- Burton, J.W. (1985) Breeding soybeans for improved oil quantity and quality. In: R. Shibles (Ed.) World soybean research conference III: proceedings. Boulder, CO: Westview Press, pp. 361– 367.
  Web of Science®Google Scholar
- Butler, K.J., Fliege, C., Zapotocny, R., Diers, B., Hudson, M., and Bent, A.F. (2021) Soybean cyst nematode resistance quantitative trait locus cqSCN-006 alters the expression of a γ-SNAP protein. Molecular Plant-Microbe Interactions, 34, 1433– 1445.
  CrossrefPubMedWeb of Science®Google Scholar
- Chaky, J.M., Specht, J.E. & Cregan, P.B. (2003) Advanced backcross QTL analysis in a mating between Glycine max and Glycine soja [abstract]. Plant and Animal Genome Abstracts, P545.
  Google Scholar
- Chung, J., Babka, H.L., Graef, G.L., Staswick, P.E., Lee, D.J., Cregan, P.B. et al. (2003) The seed protein, oil, and yield QTL on soybean linkage group I. Crop Science, 43, 1053– 1067.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Cregan, P.B. & Quigley, C.V. (1997) Simple sequence repeat DNA marker analysis. In: G. Caetano-Anolles & P.M. Gresshoff (Eds.) DNA markers: Protocols, applications and overviews. New York: John Wiley & Sons, pp. 173– 185.
  Google Scholar
- Cromwell, G.L. (2012) Soybean meal- an exceptional protein source. Available at http://www.soymeal.org/ReviewPapers/SBMExceptionalProteinSource.pdf.
  Google Scholar
- Csanádi, G., Vollmann, J., Stift, G. & Lelley, T. (2001) Seed quality QTLs identified in a molecular map of early maturing soybean. Theoretical and Applied Genetics, 103, 912– 919.
  CrossrefCASWeb of Science®Google Scholar
- Diers, B.W., Keim, P., Fehr, W.R. & Shoemaker, R.C. (1992) RFLP analysis of soybean seed protein and oil content. Theoretical and Applied Genetics, 83, 608– 612.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Diers, B.W., Specht, J., Rainey, K.M., Cregan, P., Song, Q., Ramasubramanian, V. et al. (2018) Genetic architecture of soybean yield and agronomic traits. G3: Genes, Genomes, Genetics, 8, 3367– 3375.
  CrossrefCASWeb of Science®Google Scholar
- Eckert, H., LaVallee, B., Schweiger, B.J., Kinney, A.J., Cahoon, E.B. & Clemente, T. (2006) Co-expression of the borage Delta 6 desaturase and the Arabidopsis Delta 15 desaturase results in high accumulation of stearidonic acid in the seeds of transgenic soybean. Planta, 224, 1050– 1057.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Grant, D., Nelson, R.T., Cannon, S.B., and Shoemaker, R.C. (2010) SoyBase, the USDA-ARS soybean genetics and genomics database. Nucleic Acids Research, 38 (Database issue), D843– D846.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Hood, E.E., Helmer, G.L., Fraley, R.T. & Chilton, M.-D. (1986) The hypervirulence of Agrobacterium tumefaciens A281 is encoded in a region of pTiBo542 outside of T-DNA. Journal of Bacteriology, 168, 1291– 1301.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Hwang, E.Y., Song, Q., Jia, G., Specht, J.E., Hyten, D.L., Costa, J. et al. (2014) A genome-wide association study of seed protein and oil content in soybean. BMC Genomics, 15, 1. https://doi.org/10.1186/1471-2164-15-1.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Keim, P. & Shoemaker, R.C. (1988) A rapid protocol for isolating soybean DNA. Soybean Genetics Newsletter, 15, 150– 152.
  Google Scholar

- Kim, M., Schultz, S., Nelson, R.L. & Diers, B.W. (2016) Identification and fine mapping of a soybean seed protein QTL from PI 407788A on chromosome 15. Crop Science, 56, 219– 225.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Liew, C.K., Simpson, R.J.Y., Kwan, A.H.Y., Crofts, L.A., Loughlin, F.E., Matthews, J.M. et al. (2005) Zinc fingers as protein recognition motifs: structural basis for the GATA-1/Friend of GATA interaction. Proceedings of the National Academy of Sciences of the United States of America, 102, 583– 588.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Liu, K. (1997) Soybeans: chemistry, technology, and utilization. Noew York: Chapman & Hall.
  CrossrefGoogle Scholar
- Liu, H., Zhou, X., Li, Q., Wang, L. & Xing, Y. (2020) CCT domain-containing genes in cereal crops: flowering time and beyond. Theoretical and Applied Genetics, 133, 1385– 1396. https://doi.org/10.1007/s00122-020-03554-8.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Lu, W., Wen, Z., Li, H., Yuan, D., Li, J., Zhang, H. et al. (2012) Identification of the quantitative trait loci (QTL) underlying water soluble protein content in soybean. Theoretical and Applied Genetics, 128, 425– 433.
  Google Scholar
- Masaki, T., Tsukagoshi, H., Mitsui, N., Nishii, T., Hattori, T., Morikami, A. et al. (2005) Activation tagging of a gene for a protein with novel class of CCT-domain activates expression of a subset of sugar-inducible genes in Arabidopsis thaliana. The Plant Journal, 43, 142– 152.
  Wiley Online LibraryCASPubMedWeb of Science®Google Scholar
- McBlain, B.A., Fioritto, R.J., St. Martin, S.K., Calip-Dubois, A.J., Schmitthenner, A.F., Cooper, R.L. et al. (1993) Registration of 'Thorne' soybean. Crop Science, 33, 1406.
  Wiley Online LibraryWeb of Science®Google Scholar
- Mengarelli, D.A. & Zanor, M.I. (2021) Genome-wide characterization and analysis of the CCT motif family genes in soybean (Glycine max). Planta, 253, 15.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Nichols, D.M., Glover, K.D., Carlson, S.R., Specht, J.E. & Diers, B.W. (2006) Fine mapping of a seed protein QTL on soybean linkage group I and its correlated effects on agronomic traits. Crop Science, 46, 834– 839.
  Wiley Online LibraryWeb of Science®Google Scholar
- Phansak, P., Soonsuwon, W., Hyten, D.L., Song, Q., Cregan, P.B., Graef, G.L. et al. (2016) Multi-population selective genotyping to identify soybean (Glycine max (L.) Merr.) seed protein and oil QTLs. G3-Genes Genom. Genetics, 6, 1635– 1648.
  CASWeb of Science®Google Scholar
- Prenger, E.M., Yates, J., Rouf Mian, M.A., Buckley, B., Boerma, H.R. & Li, Z. (2019) Introgression of a high protein allele into an elite soybean cultivar results in a high-protein near-isogenic line with yield parity. Crop Science, 59, 2498– 2508.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Quesneville, H. (2020) Twenty years of transposable element analysis in the Arabidopsis thaliana genome. Mobile DNA, 11, 28.
  CrossrefPubMedWeb of Science®Google Scholar
- Reinprecht, Y., Poysa, V., Yu, K., Rajcan, I., Ablett, G. & Pauls, K. (2006) Seed and agronomic QTL in low linolenic acid, lipoxygenase-free soybean (Glycine max (L.) Merrill) germplasm. Genome, 49, 1510– 1527.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Rincker, K., Nelson, R., Specht, J., Sleper, D., Cary, T., Cianzio, S.R. et al. (2014) Genetic improvement of soybean in maturity groups II, III, and IV. Crop Science, 54, 1– 14.
  Wiley Online LibraryGoogle Scholar
- Salvi, S. & Tuberosa, R. (2007) Cloning QTLs in plants. In: R. Varshney & R. Tuberosa (Eds.) Genomics-Assisted Crop Improvement. Netherlands: Springer, pp. 207– 225.

CrossrefGoogle Scholar
- SAS Institute. (2016) The SAS system for Microsoft Windows. Release 9.4, Cary, SAS Inst. Google Scholar
- Sebolt, A.M., Shoemaker, R.C. & Diers, B.W. (2000) Analysis of a quantitative trait locus allele from wild soybean that increases seed protein concentration in soybean. Crop Science, 40, 1438– 1444.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J. & Birol, I. (2009) ABySS: a parallel assembler for short read sequence data. Genome Research, 19, 1117– 1123. https://doi.org/10.1101/gr.089532.108.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Song, Q., Jia, G., Zhu, Y., Grant, D., Nelson, R.T., Hwang, E.-Y. et al. (2010) Abundance of SSR motifs and development of candidate polymorphic SSR markers (BARCSOYSSR_1.0) in soybean. Crop Science, 50, 1950– 1960.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Song, Q., Hyten, D.L., Jia, G., Quigley, C.V., Fickus, E.W., Nelson, R.L. et al. (2013) Development and evaluation of SoySNP50K, a high-density genotyping array for soybean. PLoS One, 8, E54985. https://doi.org/10.1371/journal.pone.0054985.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Tajuddin, T., Watanabe, S., Yamanaka, N. & Harada, K. (2003) Analysis of quantitative trait loci for protein and lipid contents in soybean seeds using recombinant inbred lines. Breeding Science, 53, 133– 140.e.
  CrossrefCASWeb of Science®Google Scholar
- Thompson, C.J., Movva, N.R., Tizard, R., Crameri, R., Davies, J.E., Lauwereys, M. et al. (1987) Characterization of the herbicide-resistance gene bar from Streptomyces hygroscopicus. EMBO, 6, 2519– 2523.
  Wiley Online LibraryCASPubMedWeb of Science®Google Scholar
- Vaughn, J.N., Nelson, R.L., Song, Q., Cregan, P.B. & Li, Z. (2014) The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3-Genes Genomes Genetics, 4, 2283– 2294.
  CrossrefWeb of Science®Google Scholar
- Wang, D., Shi, J., Carlson, S.R., Cregan, P.B., Ward, R.W. & Diers, B.W. (2003) A low-cost, high-throughput polyacrylamide gel electrophoresis system for genotyping with microsatellite DNA markers. Crop Science, 43, 1828– 1832.
  Wiley Online LibraryCASWeb of Science®Google Scholar
- Wang, X., Jiang, G., Green, M., Scott, R., Song, Q., Hyten, D., Cregan, P. (2014) Identification and validation of quantitative trait loci for seed yield, oil and protein contents in two recombinant inbred line populations of soybean. Molecular Genetics and Genomics, 2014, 289, 935– 949
  CrossrefCASPubMedWeb of Science®Google Scholar
- Wang, S., Liu, S., Wang, J., Yokosho, K., Zhou, B., Liu, Z. et al. (2020) Simultaneous changes in seed size, oil content and protein content driven by selection of SWEET homologues during soybean domestication. National Science Review, 7, 1776– 1786.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Warnes, G.R. (2003). The genetics package. R News, 3, 9– 13. Retrieved from https://ci.nii.ac.jp/naid/10030730040/#cit
  Google Scholar
- Warrington, C., Abdel-Haleem, H., Hyten, D., Cregan, P., Orf, J., Killam, A. et al. (2015) QTL for seed protein and amino acids in the Benning Danbaekkong soybean population. Theoretical and Applied Genetics, 128, 839– 850.
  CrossrefCASPubMedWeb of Science®Google Scholar

- Wesley, S.V., Helliwell, C.A., Smith, N.A., Wang, M., Rouse, D.T., Liu, Q. et al. (2001) Construct design for efficient, effective and high-throughput gene silencing in plants. The Plant Journal, 27, 581– 590.
  Wiley Online LibraryCASPubMedWeb of Science®Google Scholar
- Wilcox, J.R. (1985) Breeding soybeans for improved oil quantity and quality. In: R. Shibles (Ed.) World soybean research conference III: proceedings. Boulder: Westview Press, pp. 380–386.
  Google Scholar
- Zhang, Z.Y., Xing, A.Q., Staswick, P. & Clemente, T.E. (1999) The use of glufosinate as a selective agent in Agrobacterium-mediated transformation of soybean. Plant Cell Tiss. Org., 56, 37– 46.
  CrossrefCASWeb of Science®Google Scholar
- Zhang, L., Li, Q., Dong, H., He, Q., Liang, L., Tan, C. et al. (2015) Three CCT domain-containing genes were identified to regulate heading date by candidate gene-based association mapping and transformation in rice. Scientific Reports, 5, 7663.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Zhang, H., Goettel, W., Song, Q., Jiang, H., Hu, Z., Wang, M.L. et al. (2020) Selection of GmSWEET39 for oil and protein improvement in soybean. PLoS Genetics, 16, e1009114. https://doi.org/10.1371/journal.pgen.1009114.
  CrossrefCASPubMedWeb of Science®Google Scholar
- Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C. & Weir, B.S. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28, 3326– 3328. https://doi.org/10.1093/bioinformatics/bts606.
  CrossrefCASPubMedWeb of Science®Google Scholar